Genomics and regulatory functions of microRNAs and small silencing RNAs in *Arabidopsis thaliana*

by

Ramya Rajagopalan
B.A., Biological Sciences
Cornell University (2002)

Submitted to the Department of Biology in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

at the

Massachusetts Institute of Technology

September 2007

Signature of Author .....
Department of Biology
August 30, 2007

Certified by ...
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by .....
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

Genomics and regulatory functions of microRNAs and small silencing RNAs in *Arabidopsis thaliana*

Ramya Rajagopalan

Submitted to the Department of Biology in
Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

ABSTRACT

Small RNA-mediated gene silencing is a mechanism widely employed by eukaryotes to repress many loci including some involved in critical developmental transitions. In plants, endogenous small RNAs consist of two broad classes, the ~21-nucleotide (nt) microRNAs (miRNAs) and the diverse ~22–24-nt *trans*-acting and heterochromatic short interfering RNAs (siRNAs). In order to more extensively characterize the small RNA landscape in plants and to identify undiscovered miRNAs and siRNAs, we performed high-throughput sequencing of small RNAs. We generated a large dataset consisting of >340,000 unique sequences expressed in several representative plant organs and developmental stages including wild-type seedlings, flowers, leaves and siliques. Application of enhanced miRNA annotation criteria gleaned from the data revealed the existence of at least 38 apparently recently-evolved miRNAs that were much less abundant in plant tissues than the 26 conserved miRNA families, and had a greater diversity of predicted target genes. We characterized several of these miRNAs more closely. Our results supported a homeostatic auto-regulatory loop for DCL1 via the intron-embedded miR838, and elaborated on the prevailing model of DCL1-mediated miRNA biogenesis with the finding that at least two miRNAs (miR839 and miR822) are processed exclusively by DCL4. Several microRNA target sites were experimentally validated, including the miR823-directed cleavage of the DNA cytosine methylation factor *CHROMOMETHYLASE3*. We also identified a *trans*-acting siRNA-generating locus which we called *TAS4*, and confirmed that miR828 triggers phased siRNA production by specifying targeted cleavage of *TAS4* transcripts. The evolving miRNAs described in this work may have lineage-specific roles, and their discovery potentiates future functional investigation of recently-emerged miRNAs and their evolution in *Arabidopsis*. Finally, we discovered thousands of endogenous candidate heterochromatic siRNAs of unknown function, the majority of which mapped to unannotated regions of the genome especially prone to generating siRNAs ("hotspots") or to repetitive or transposable elements. Our small RNA study suggests that a significant proportion of the genome is primed for the emergence of new miRNA families or for siRNA production, and expands the roles of these small RNAs in shaping regulatory circuits and transcriptome output.

Thesis Supervisor: David P. Bartel
Title: Professor of Biology

Table of Contents

ACKNOWLEDGMENTS

I feel the inadequacy of words at this moment when I feel they should most matter, as I try to compile the list of all of the people who have helped me on my journey thus far. I would first like to thank my advisor, David Bartel, for allowing me to join his lab and pursue my interests in plant biology, computational biology and gene regulation. His scientific rigour and intuition are inspiring, and his attention to detail is admirable. I am grateful for his guidance, mentorship and support throughout my time in his lab.

I also thank the entire Bartel lab who have with humour and warmth created an amazingly collegial and vibrant place to do science. I have benefited from the guidance of several among them, in particular Mike Axtell, who has been a collaborator, mentor, interlocutor and friend, and shared with me reagents, advice, and expertise in plant handling. I am grateful to Hervé Vaucheret for offering perspectives on my work and sharing insights from his vast experience in small RNA biology, and to both Hervé and Jerry Trejo, whose collaboration in the work described in Chapters 1 and 2 was invaluable. I would also like to thank Allison Mallory and Matthew Jones-Rhoades for discussions and for their generosity with reagents/scripts. I will remember fondly the many stimulating conversations on science, biology and life I have had the privilege to share with current and former members of the lab, including: Mike Axtell, Daehyun Baek, Sarah Bagby, Scott Baskerville, Sue-Jean Hong, Laura Resteghini (thanks for keeping the lab running so smoothly and efficiently!), Uli Muller, Graham Ruby, Erik Schultes, I-hung Shih, Jerry Trejo and Soraya Yekta. I want to also thank Ndubuisi "Dubi" Azubuine for his generosity, Frank Solomon for edifying discussions, and Prakash Rao and Chris Hug for their kind encouragement especially during the writing of this Thesis.

I am deeply indebted to many others for their support during my time in graduate school. To my friends, thanks for laughing and making me laugh, for indulging my quirks and giving true meaning to that precious entity, "friendship." Vinay, your sincerity and thoughtfulness has truly been appreciated. Many others in various walks of life at MIT have been instrumental in making my time here so enjoyable, and I thank you all for your company.

Mom, Dad, Radha, and Ranga, I cannot thank you enough, for listening, for quelling my doubts, for motivating me, for cherishing all my idiosyncrasies, and most of all for giving me love to trilokam and back. Thanks for all your support, your encouragement, for late-night dindas, comforting hilarity, and for believing in me with all your heart—I thank you in return with all of mine.

Introduction

An overview of small RNA-directed gene silencing

The spatial and temporal precision of gene expression is essential for the proper progression of events throughout an organism's growth and development. To this end, the regulatory networks which fine-tune gene expression from cell to cell and from stage to stage in multicellular eukaryotes play central coordinating roles. Such networks encompass molecular events occurring temporally upstream of transcription, which determine where and when transcription initiates, and downstream of transcription, which affect the fates of resulting protein-coding and non-protein-coding RNAs.

The functions of proteins in turning on and off genes and modulating the accumulation of RNA transcripts have been under study for decades. However, it is only in recent years that the role of small silencing RNAs in both of these regulatory processes has gained increasing prominence (reviewed in (Bartel 2004; Lindbo and Dougherty 2004; Chen 2005; Jones-Rhoades et al. 2006). Transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) mediated by small RNAs is now known to function as a widespread and important form of regulatory control in both plants and metazoans, with significant consequences for the composition of cellular transcriptomes. RNA-based gene silencing is a system that is evolutionarily conserved throughout the plant, animal and fungal kingdoms, with several key components in common, attesting to its functional significance in eukaryotic biology. These components include class III endoribonucleases, binding proteins specific for double-stranded RNA (dsRNA), and the Argonaute family of proteins, comprising the Argonaute and Piwi subfamilies.

Endogenous small silencing RNAs, which downregulate genes by sequence-specific interactions with complementary transcripts or loci, are ~20–24 nucleotides (nt) in length and fall into two broad classes, the ~21-nt microRNAs (miRNAs) (Ambros 2004; Bartel 2004), and endogenous small interfering RNAs (siRNAs), which range between ~21–24 nt. Though indistinguishable biochemically, they can be differentiated by the biogenetic pathways through which they originate, and they can differ in the kinds of loci they target and the downregulatory mechanisms they activate.

In the plant kingdom, siRNAs differ from miRNAs in at least six aspects (Table I.1). First, they differ in the range of their lengths as described above. Second, they are

excised from different types of double-stranded precursor RNA structures. While miRNAs originate from genome-encoded transcripts that fold back on themselves to generate an imperfectly paired precursor molecule known as a hairpin or stemloop, siRNAs are typically derived from cleavage of bimolecular double-stranded RNA (dsRNA) formed when two synthesized RNA strands hybridize; unlike the miRNA precursor stemloop, siRNA precursor duplexes are sometimes perfectly complementary along their entire length (Ambros 2004; Bartel 2004). Third, siRNA accumulation often requires the activity of special polymerases to generate dsRNA, whereas the miRNA precursor is itself capable of forming a basepaired structure. Second strand synthesis at siRNA-generating loci is typically mediated by an RNA-dependent RNA polymerase (RDR) protein. Fourth, the processing of the miRNA hairpin transcript is typically very precise, liberating primarily a single small RNA duplex containing the mature miRNA, while many sequence-unrelated siRNAs can accumulate from processive cleavage of a single siRNA precursor molecule (Jones-Rhoades et al. 2006). Fifth, siRNAs can direct either PTGS or TGS of target loci, but studies to date have confirmed miRNA activities solely in PTGS pathways. Finally, miRNAs typically act to silence transcripts from loci that are distinct from their own locus of origin (acting in *trans*), whereas siRNAs can silence both parental and unrelated but highly homologous loci (acting in *cis* or *trans*) (Bartel 2004). Despite these distinctions, miRNAs and siRNAs share two general features in common: both are derived from double-stranded RNA intermediates, and both act sequence-specifically as negative regulators of gene expression.

MicroRNAs in *Arabidopsis*

Biogenesis of miRNAs

A. Expression of MIR loci

MicroRNAs are pervasive riboregulators in both animals and plants, ~21 nt in length, and greatly enriched in sequences beginning with a 5′ uracil nucleotide (Lau et al. 2001; Reinhart et al. 2002). Though several aspects of miRNA biology differ between the plant and animal kingdoms, the general sequence of events governing miRNA biogenesis is roughly similar. In both, microRNAs are expressed as precursor transcripts, usually

from non-protein coding *MIRNA* (*MIR*) loci in the genome, and expression is usually highly strand-specific in plants. The majority of *Arabidopsis* microRNA loci are autonomous transcription units expressed under the control of their own promoters (Xie et al. 2005a). In contrast, moss miRNAs often overlap protein-coding genes (Axtell et al. 2007). At least five non-conserved miRNAs in *Arabidopsis* map to the introns of protein-coding genes (Rajagopalan et al. 2006), and a number of miRNAs in fruitflies (Ruby et al. 2007) and mammals (Rodriguez et al. 2004) are encoded in introns. An intronic locus presumably allows the hitchhiking miRNA to circumvent the need for its own promoter, a possibility corroborated by the observation that intronic miRNAs in humans are frequently coexpressed with their host genes (Baskerville and Bartel 2005), and miRNA and spliced mRNA can both be derived from a common precursor transcript (Kim and Kim 2007). Intronic miRNAs may evolve in cases when coordinate expression of host gene and miRNA confers a beneficial outcome.

Polycistronic miRNA loci are far more numerous in animals than in plants (Bartel 2004), and while the former often encode several miRNAs unrelated in sequence, the latter generally comprise paralogous miRNAs in the same family. Examples include several clusters of both highly conserved and bryophyte-specific miRNAs in moss (Axtell et al. 2007), and clustered miR395 loci in the *A. thaliana* genome that could derive from a single primary transcript (Jones-Rhoades and Bartel 2004).

It is thought that both metazoan and plant miRNA genes possessing their own promoters are transcribed by RNA polymerase II (RNA pol II) (Lee et al. 2004; Xie et al. 2005a), or in the case of some repeat-associated miRNAs in animals, by RNA polymerase III (Borchert et al. 2006). Canonical promoter motifs consistent with RNA pol II-dependent transcription, such as a TATA box, have been observed upstream of 45 of the plant *MIRNA* loci, including many of those most highly conserved to other plant species (Xie et al. 2005a). In addition, precursor transcripts from 52 *Arabidopsis MIRNA* loci have been confirmed as present in poly(A^+)-selected and 5′-capped RNA pools, suggesting that like other RNA pol II products, plant miRNA transcripts are capped at their 5′ ends and polyadenylated at their 3′ ends (Xie et al. 2005a). The identification of putative miRNA precursor transcripts in expressed sequence tag (EST) datasets in *Arabidopsis* and other species further supports their probable polyadenylation and

derivation from RNA pol II transcription (Jones-Rhoades and Bartel 2004; Zhang et al. 2006a). It is possible that the promoters of non-ubiquitous or tissue-specific miRNAs are also populated with binding sites for specific transcription factors to ensure spatially- or temporally-restricted expression domains. In addition to being polyadenylated, transcripts from at least two plant miRNA loci undergo splicing, *MIR172a* (Aukerman and Sakai 2003) and *MIR163* (Kurihara and Watanabe 2004).

Plant miRNA precursors are much more variable in predicted structure than those of animals. They often have much more extensive loops and a greater distance separating the loop from the miRNA sequence on the stem, ranging from two basepairs (bp) to more than 40 (Reinhart et al. 2002). Animal precursors tend to be ~70 nt in length, with about 2-10 bp separating the miRNA sequence from the terminal loop (Lau et al. 2001).

B. Processing of miRNA precursors

The miRNA precursor transcript contains two arms with imperfect complementarity, and adopts a stem-loop secondary structure with a terminal loop adjacent to these sequences paired along the stem (Ambros 2004; Bartel 2004). This is the region from which the mature miRNA will be liberated. Typically, basepairing in the stem extends beyond the region of the miRNA (Jones-Rhoades et al. 2006), and in animals, flanking single-stranded segments at the base of the stem have been shown to be critical recognition elements for processing of the stemloop (Han et al. 2006). The structured form of the precursor, called the primary miRNA (pri-miRNA), undergoes two endonucleolytic processing steps to yield the mature miRNA. The required cleavages are performed by two proteins which are members of a class of ribonuclease III (RNAseIII) enzymes that specifically recognize dsRNA substrates (Fig. I.1). RNAseIII enzymes catalyze the release of two strands of small RNA as a duplex from precursor dsRNA structures, leaving two characteristic modifications at the termini of cleavage products, a 5′ monophosphate and a 3′ hydroxyl.

In metazoans, the first cleavage event occurs in the nucleus and the second in the cytoplasm. The pri-miRNA hairpin initially encounters the nuclear-localized Class 2 RNAseIII enzyme known as Drosha (Lee et al. 2003), for which there is no known plant homolog. Drosha acts in a multiprotein complex called the Microprocessor, in concert

with a double-stranded RNA binding domain (dsRBD) substrate recognition protein called Pasha (partner of Drosha), also known as *DGCR8* in humans (Denli et al. 2004; Gregory et al. 2004). Drosha makes the first set of staggered cuts on each arm near the base of the hairpin stem, cleaving off flanking hairpin sequence and establishing defined 5′ and 3′ termini (the latter with a 2-nt overhang) to generate the precursor miRNA (pre-miRNA) (Lee et al. 2003). Pre-miRNAs in animals tend to be ~60 nt long, with roughly 10 bases in the loop, and 10-20 bases in the loopside region of the stem separating the fated miRNA from its imperfectly complementary sequence, known as the miRNA*. Plant pri-miRNAs (and pre-miRNAs) are far more heterogeneous in the size of the loop and the distance along the paired stem between the miRNA and the loop.

Mammalian pre-miRNAs are exported to the cytoplasm via the nuclear export receptor Exportin-5 (Exp5) (Yi et al. 2003; Lund et al. 2004). Exp5 homologs also mediate nuclear-to-cytoplasmic transport of pre-miRNAs in *Xenopus laevis* (Bohnsack et al. 2004) and the fruit fly *Drosophila melanogaster* (Shibata et al. 2006). Exportin-5 may have been coopted for pre-miRNA export because its other cargoes include tRNA (Calado et al. 2002) and other double-stranded RNAs (Gwizdek et al. 2003), suggesting a specificity for small, minihelix-containing molecules. The export of pre-miRNAs to the cytoplasm is critical for the next and final step of animal miRNA maturation, catalyzed by the bidentate Class 3 RNAseIII enzyme Dicer. Like Drosha, Dicer makes staggered cuts that leave a 2-nt 3′ overhang (Bernstein et al. 2001; Grishok et al. 2001; Hutvagner et al. 2001). In metazoans, Dicer completes the processing of the pre-miRNA by catalyzing the second (loopside) cut, releasing the miRNA:miRNA* duplex from the hairpin, with 2-nt 3′ overhangs on both ends. Similar to the activity of Drosha endonuclease, proper cleavage is facilitated by a dsRBD protein called Loquacious in *Drosophila* (whose humans homolog is TRBP) (Forstemann et al. 2005; Saito et al. 2005).

In contrast to animal miRNA biogenesis, the majority of plant miRNA hairpins are processed by DICER-LIKE1 (DCL1) (Park et al. 2002; Reinhart et al. 2002), one of four DCL proteins encoded in the *Arabidopsis* genome. Indeed, no Drosha homologs have been identified in land plants. As with Drosha and metazoan Dicer proteins, precise processing of miRNA hairpins requires an interaction between DCL1 and a partnering nuclear-localized dsRNA-binding protein, HYPONASTIC LEAVES 1 (HYL1) (Han et

al. 2004; Vazquez et al. 2004a; Hiraguri et al. 2005; Kurihara et al. 2006) (Fig. I.1). It has been shown that *Arabidopsis* DICER-LIKE1 (DCL1) is primarily nuclear-localized and is itself competent to make both cuts on the miRNA hairpin prior to export of the mature miRNA to the cytoplasm (Kurihara and Watanabe 2004). Consistent with the idea that plant miRNAs are processed almost entirely in the nucleus, single-stranded mature miRNA has been observed in nuclear fractions of *Arabidopsis* leaf cells, though it is much more abundant in the cytoplasm (Park et al. 2005). In addition, a nuclear-localized variant of the P19 protein of tomato bushy stunt virus, encoding a viral silencing suppressor protein that binds and sequesters Dicer duplexes, reduced accumulation of miR159 whereas a cytoplasmically-localized P19 variant did not (Papp et al. 2003), suggesting that miRNA duplexes are produced in the nucleus. More recent live-cell cytological data indicates that a fraction of DCL1 colocalizes with HYL1 in nuclear dicing bodies (D-bodies) along with pri-miRNA transcripts (Fang and Spector 2007). D-bodies may be a site for pre-miRNA (and possibly miRNA) maturation, or they may be centers for storage or assembly of miRNA processing complexes (Fang and Spector 2007).

The temporal order of the stepwise cuts remains uncertain, although pre-miRNAs may be short-lived because they are detected very infrequently in *Arabidopsis* compared to animals, suggesting that if the cuts are staggered in time the second follows quickly after the first (Jones-Rhoades et al. 2006). The precision of DCL1 cleavage along miRNA hairpins is fairly high, in that processing preferentially centers around the region of the fated miRNA:miRNA* duplex. Heterogeneous length variants of the mature miRNA do appear in cloning datasets or on RNA gel blots, but are generally limited (Reinhart et al. 2002; Vaucheret et al. 2004; Xie et al. 2004; Rajagopalan et al. 2006). Given the sequence diversity of mature miRNAs, local helix geometry probably plays a larger role than primary sequence in DCL1 cleavage site recognition. Supporting this, mutations on both arms of a hairpin that alter the sequence of the processed duplex but maintain the local pattern of paired and unpaired nucleotides give rise to artificial miRNAs cleaved from an analogous position as the wild-type miRNA, but composed of a different sequence (Parizotto et al. 2004). Subtle heterogeneities in length, however, underline the

importance of the precise sequence of paired and unpaired residues in defining the cleavage site (Vaucheret et al. 2004).

The zinc-finger protein SERRATE (SE) also acts in the miRNA biogenetic pathway, although its role is still unclear (Grigg et al. 2005; Lobbes et al. 2006; Yang et al. 2006a). In *se* mutants, levels of several miRNAs are reduced while target mRNAs overaccumulate (Lobbes et al. 2006). The *se* mutant phenotypes overlap with those of *hyl1* (Lobbes et al. 2006), and the *se hyl1* double mutant is embryonically lethal, indicating that *se* might act in the same pathway (Yang et al. 2006a). Microscopy and biochemical studies reveal that SE is nuclear-localized and interacts with HYL1 in yeast two-hybrid assays, which is confirmed by glutathione S-transferase pull-down assays for either protein (Yang et al. 2006a). The accumulation of pri-miRNAs in *se* plants suggests that SE plays a role in specifying efficient cleavage of miRNA precursor hairpins (Lobbes et al. 2006; Yang et al. 2006a)

C. Covalent Modification

The events in plant miRNA biogenesis described thus far are thought to occur in the nucleus. Once liberated from the hairpin, the maturation of plant miRNA duplexes is completed by the HUA ENHANCER1 (HEN1) methyltransferase (Yu et al. 2005; Yu et al. 2006). HEN1 deposits a methyl group on the 2′ hydroxyl group of the 3′ terminal nucleotide (Fig 1). Whether the miRNA alone or both miRNA and miRNA* sequences in a single duplex are methylated remains to be resolved (Yang et al. 2006b), although the latter is more likely given that methylated miR172 and miR172* are detectable in control plants but not in plants expressing viral silencing suppressors that bind and sequester small RNA duplexes (Yu et al. 2006). Given that *hen1* mutants display reduced accumulation of miRNAs and a concomitant increase in size heterogeneity of miRNA species (Park et al. 2002; Boutet et al. 2003), HEN1 methylation likely serves to stabilize miRNAs in the cell. The additional nucleotides on extended miRNA species in *hen1* mutants are due to 3′-polyuridylation by an uncharacterized enzyme, while shorter miRNAs may be generated by non-specific degradation or a putative 3′ exonuclease, an activity that may also explain the reduced accumulation of miRNAs in *hen1* plants (Li et al. 2005). Thus, the methylation of plant miRNAs may protect the 3′ end from the action

of polyuridylation enzymes, exonucleases or other small RNA destabilizing pathways. The methyl group might also be a necessary structural feature for recognition of the miRNA by the RNA-induced silencing complex (RISC), given that the 2′ OH group is a contact point for Argonaute proteins in the RISC (Lingel et al. 2004; Ma et al. 2004; Yu et al. 2005). The methyl tag does not appear to enhance the ability of the guide RNA to direct RISC cleavage activity in *in vitro* assays (Qi et al. 2005). Methylation does not appear to function in animal miRNA biogenesis, though recently it has been shown that animal siRNAs and often repeat-associated piRNAs (discussed below) are methylated (Horwich et al. 2007; Saito et al. 2007).

Whether miRNA methylation happens in the nucleus or in the cytoplasm remains an open question, although microscopy suggests that HEN1 localizes to the nucleus (including nucleoplasm and D-bodies) and the cytoplasm (Fang and Spector 2007). Many miRNAs appear to exit the nucleus in order to repress target messages in the cytoplasm, a process facilitated by the Exp5 ortholog, HASTY. The accumulation of some miRNAs in both nuclear and cytoplasmic compartments is decreased in *Arabidopsis* mutants defective for HASTY, suggesting that HASTY plays a role in stabilizing at least a subset of plant miRNAs (Park et al. 2005) and assists in their export to the cytoplasm (Fig 1). Nevertheless, functional redundancy among RNA nucleotransporters may allow other components to compensate in *hasty* mutants (Park et al. 2005).

MicroRNA identification

The discovery of microRNAs was initiated by the identification of a regulatory circuit in *C. elegans* involving the genes *lin-4*, *lin-14*, and *lin-28*, which control the larval-to-adult developmental switch in post-embryonic hypodermal cells (Ambros 1989). The heterochronic *lin-4* locus did not code for a protein, but instead gave rise to a transcript with a predicted stemloop secondary structure (Lee et al. 1993). Furthermore, a small RNA presumably derived from this transcript displayed sequence complementarity to the 3′ untranslated region (UTR) of *lin-14*, a heterochronic developmental protein (Lee et al. 1993; Wightman et al. 1993). The evolutionary conservation of both the sequence and temporal expression of the *let*-7 gene, a locus encoding another small RNA with complementarity to *lin-14* (among other genes) (Reinhart et al. 2000), compelled the

notion that the two small RNAs were critical in early developmental events throughout bilaterally symmetric animals (Pasquinelli et al. 2000; Reinhart et al. 2000), and might represent a much larger set of small regulatory RNAs in animal genomes.

Since then, hundreds of miRNAs have been identified in vertebrates, invertebrates, mammals and plants (Ambros 2004; Bartel 2004; Jones-Rhoades et al. 2006). The majority of newly discovered miRNA loci have emerged from two complementary approaches, small RNA cloning experiments or computational prediction, in part because of the amenability of these high-throughput approaches to miRNA discovery.

Experimental cloning and sequencing of small RNA products in total RNA libraries often exploit the characteristic terminal covalent structure of RNAse III products to enrich for Dicer products containing a 5′ monophosphate and a 3′ hydroxyl group (Lau et al. 2001), or select solely for a 3′ hydroxyl group (Lagos-Quintana et al. 2001). Following size-selection of small RNAs from total RNA samples by gel purification, directional, sequential ligation of known-sequence adaptors to the 3′ and 5′ ends allows preferential enrichment of Dicer products within small RNA libraries (Fig. I.2). Once the small RNA sequences have been obtained, they are analyzed via bioinformatics tools that map them to loci of origin in the genome. MicroRNA identification is initiated by pinpointing a cloned small RNA within a predicted stemloop precursor, the genomic footprint of a miRNA locus. Large-scale microRNA cataloging efforts and novel miRNA discovery are therefore enormously facilitated in organisms with sequenced genomes.

The most confidently identified miRNAs tend to be among the most abundant sequences in small RNA cloning projects. To date, several hundred miRNAs have been identified in animals as diverse as zebrafish, worms, flies, mice, rats, chimps and humans. Many of these miRNAs are apparently conserved through the deepest branches of metazoans, while others appear to be lineage-specific (Kim and Nam 2006). In contrast to the comparatively large numbers of animal miRNAs discovered by the first small RNA cloning experiments in *C. elegans* (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001), initial cloning studies in *Arabidopsis* revealed only 18 miRNAs matching 43 loci, the vast majority of which had detectable homologs in the only other available sequenced plant genome at the time, *Orzya sativa* (rice) (Park et al. 2002; Reinhart et al.

2002). Several of the loci gave rise to miRNAs with identical or related sequences, implying that plant miRNAs were typically multigenic; additional paralogous loci corresponding to this first cohort of plant miRNAs were identified computationally (Jones-Rhoades and Bartel 2004), yielding a total of 15 families. The duplication processes (including genome-wide, tandem and segmental) driving the diversification of some of these conserved families have been reconstructed, pointing to accelerated evolutionary trajectories (Maher et al. 2006).

Until recently, reverse transcribed and amplified cDNA pools from small RNA libraries were concatenated and cloned by conventional Sanger sequencing techniques. However, increased coverage of these libraries can be obtained via high-throughput deep sequencing technologies such as MPSS (Brenner et al. 2000) and 454 (Margulies et al. 2005). The 454 strategy involves monitoring nucleotide addition upon clonal amplification of templates singly-conjugated to beads loaded individually into separate wells in a microfluidic chamber. Sequential flooding and clearing of the chamber with each of the four DNA nucleotides in turn allows basecalling by monitoring the excitation of luciferase activity upon release of pyrophosphate during DNA synthesis (Margulies et al. 2005). Both platforms yield datasets of approximately the same order of magnitude, on the order of several hundred thousand sequences per run. However, the MPSS technology is limited to obtaining 17-nt "signature" sequences, which preempts retrieval of important information regarding the length of the of ~20–24 nt small RNAs. Nevertheless, small RNA sequencing projects have employed both MPSS (Lu et al. 2005; Lu et al. 2006) and 454 (Axtell et al. 2006; Henderson et al. 2006; Lau et al. 2006; Qi et al. 2006; Rajagopalan et al. 2006; Ruby et al. 2006; Axtell et al. 2007; Brennecke et al. 2007; Fahlgren et al. 2007; Kasschau et al. 2007) technologies. The vast majority of sequences in these studies are only retrieved once, indicating that even high-throughput sampling is far from saturating the enormous diversity of endogenous small RNA populations.

The identification of miRNAs by experimental cloning is biased towards those species that are highly expressed, abundant, and spatially ubiquitous, a hurdle that was especially felt prior to the advent of deep sequencing by MPSS or 454 technology. Though this limitation is attenuated by deep sequencing, which can potentially uncover

less abundantly expressed or spatially-restricted miRNAs even when sampling from whole organs, high-throughput approaches are still far from saturating the endogenous small RNA pools and are likely to miss miRNAs that are specifically expressed in certain contexts, for example in a few cells or in response to environmental stimuli. In addition, some *bona fide* miRNAs may not emerge in small RNA libraries due to sequence biases incurred in the cloning process, for example during adaptor ligation (Michael Axtell, personal communication). Bioinformatics methods which mine the genome for candidate miRNAs can overcome some of these limitations and thus provide a complementary approach to *de novo* miRNA discovery.

The first computationally predicted, confidently annotated miRNAs emerged from vertebrates (Lim et al. 2003a), and was closely followed by the prediction of miRNAs in other species including six previously unidentified miRNA families in *Arabidopsis*, all of which are conserved in rice (Jones-Rhoades and Bartel 2004). To hone in on likely candidate loci, several pairing characteristics of the predicted stemloop structure at known miRNA loci are parametrized, and the genome is scanned for conforming loci. These structural criteria include the degree of pairing in the region of the miRNA/miRNA* duplex, the degree of pairing in the rest of the stem, and in animals, the loop length (Lim et al. 2003b; Jones-Rhoades and Bartel 2004). However, these descriptors alone cannot distinguish authentic miRNAs from the many thousands of false-positive loci evincing hairpin-folding potential in sequenced genomes such as *Arabidopsis* (Jones-Rhoades and Bartel 2004). Therefore, computationally predicted candidates must rely heavily on other criteria for accurate annotation as miRNAs, especially in the absence of strong experimental evidence for miRNA-like biogenetic dependencies. The easiest and perhaps most compelling filter to implement is a requirement for phylogenetic conservation of the candidate miRNA and its predicted precursor hairpin. This methodology has proven successful for the identification of previously unknown, well-conserved plant miRNAs, at least one of which would be missed in most wild-type cloning experiments because it is stress-induced (Jones-Rhoades and Bartel 2004). Further refinements of the candidate set can be obtained by exploiting the high degree of complementarity between miRNAs and targets in plants, and implementing a filter that requires conservation to at least one other species of the

predicted target gene and the target site (Jones-Rhoades and Bartel 2004). Thus, the accuracy of bioinformatic prediction of miRNAs depends on the availability of other reference genomes, making this methodology less suited to identifying apparently lineage- or species-specific miRNAs, unless very close evolutionary relatives of the organism under study have also had their genome sequences elucidated; in practice, this has not yet been the case in the plant kingdom.

Compared to cloning and computational methods, forward genetic screens for loss-of-function mutants have played a comparatively minor role in miRNA discovery efforts, identifying only a handful of previously uncharacterized animal miRNA loci, including *lsy-6* in *C. elegans* (Johnston and Hobert 2003) and *bantam* in *D. melanogaster* (Hipfner et al. 2002; Brennecke et al. 2003). This may be due to the smaller number of nucleotides that miRNA gene loci present as potential targets for mutagenesis leading to observable phenotypic outcomes, compared to other loci. In plants, loss-of-function screens have not identified any new miRNA families, although in one case the developmental function of a previously validated miRNA has been confirmed: the *early extra petals1* (*eep1*) loss-of-function allele is caused by a transposon insertion upstream of the *MIR164c* locus (Baker et al. 2005). On the other hand, activation tagging in *Arabidopsis*, in which viral enhancers are used to constitutively drive endogenous promoters, has successfully uncovered one novel miRNA, miR319, which forms a family of related sequences with the miR159 loci (Palatnik et al. 2003). Overexpression screens isolated gain-of-function mutant alleles for another two plant miRNAs from previously-identified miRNA families, miR172 (Aukerman and Sakai 2003), and miR166 (Kim et al. 2005; Williams et al. 2005), again providing evidence for the developmental consequences of perturbing plant miRNA expression. Perhaps forward genetic screens have proven less successful in plants because of the genetic redundancy of miRNA loci (Jones-Rhoades et al. 2006), and in animals because of the functional redundancy of members of miRNA gene families with identical seed sequences, both of which may help to buffer against mutation or loss at any one miRNA-encoding locus.

MicroRNA classification and nomenclature

Typically, miRNAs comprise only a fraction of the sequenced small RNAs in cloning projects, expecially in *Arabidopsis* where miRNAs are thought to be vastly outnumbered by endogenous siRNAs and other fragments from known and highly expressed non-coding RNA classes including ribosomal RNA (rRNA) and transfer RNA (tRNA) (Lu et al. 2005; Rajagopalan et al. 2006; Fahlgren et al. 2007; Kasschau et al. 2007). Due to the overwhelming diversity of small RNAs that emerge from such projects despite protocols that favour the representation of Dicer products, and the number of endogenous loci with predicted stemloop-like secondary structures (Jones-Rhoades and Bartel 2004), it has become necessary to establish and refine guidelines for confidently annotating miRNA loci. Candidate miRNAs typically must satisfy a subset of criteria in order to receive miRNA designation. The expression and biogenetic criteria are especially important for computationally-predicted miRNAs, and include experimental cloning of the ~21-nt sequence, and/or detection of the appropriately-sized mature miRNA sequence on an RNA gel blot (Jones-Rhoades et al. 2006). Further evidence can be furnished by demonstrating miRNA-like biogenetic requirements, including dependence on components known to be involved in miRNA but not siRNA biogenesis, as evidenced by reduced or unaffected accumulation, respectively, in a panel of silencing pathway mutants (Jones-Rhoades et al. 2006).

The structural criteria are aimed at establishing with reasonable confidence that the candidate originated from a miRNA stemloop. At the level of primary sequence, *Arabidopsis* miRNAs typically feature a uridine residue at their 5' terminus, though this is not a definitive characteristic. More importantly, *in silico* folding of the miRNA precursor sequence by RNAfold, MFOLD, or a similar RNA structure prediction program should yield a lowest-energy conformer that resembles a hairpin, with the miRNA sequence on one arm of the stem. This energetically-stable foldback structure should not contain any large loops or any asymmetric bulges (Ambros et al. 2003a). Specific guidelines for plant miRNAs recommend that candidate hairpins have no more than 7 unpaired nucleotides in the 25 nucleotides centered on the miRNA, of which no more than 3 are consecutive and no more than 2 are bulged, with identical pairing constraints for the 25 bases centered on the miRNA* (Jones-Rhoades et al. 2006). In

addition, the evolutionary conservation of the candidate sequence and the hairpin-forming potential of its predicted precursor transcript in other sequenced genomes (including Expressed Sequence Tag (EST) datasets) can support its candidacy as a miRNA and attest to its probable biological significance. The presence of orthologs in other genomes implies conservation of the secondary structure of the locus and thus negative selection against mutations that might disrupt pairing in the stemloop, all of which points to the small RNA arising from the stemloop itself, rather than a long dsRNA.

It is important to note that a small RNA candidate need not have a predicted or experimentally-verified target gene in order to be classified as a miRNA. Following the precedent established with the first usage of the term microRNA to describe this class of molecules (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001), miRNA designation is conferred based on biogenetic evidence of the candidate's genomic and molecular origins, and not on functional evidence for its regulatory activity. In practice, the above guidelines are sometimes loosely interpreted and loci which initially appear to be miRNAs are later shown to encode non-miRNA transcripts as the criteria for miRNA designation are refined.

MicroRNAs with the same number in different animal species typically imply shared ancestry, and the same is true for plant miRNA genes; for example, ath-miR165 and osa-miR165 represent orthologs of miR165 in *A. thaliana* and *O. sativa* respectively (Griffiths-Jones et al. 2006). Furthermore, in both animals and plants, several miRNAs appear to be represented by multiple genes. Animal miRNAs with identical seed sequences (encompassing positions ~2-7 from the 5′ end of the miRNA) are classified as belonging to the same family. In plants, many of the highly conserved miRNAs are represented by more than one locus in the genome which produce identical or highly sequence similar miRNAs. Paralogous *MIRNA* genes whose mature miRNA sequences differ at 0-3 nucleotide positions are grouped into a single family; in many cases family members are distinguished by a shared number but unique lettered suffixes, for example, miR165a and miR165b (Bartel 2004). With one notable example (described below), *Arabidopsis* miRNAs in the same family target the same set of genes (Jones-Rhoades et al. 2006).

The RNA-induced silencing complex (RISC)

The effector complex of small RNA-based gene silencing is the RNA-induced silencing complex (RISC). Several different kinds of RISC complexes are thought to be present in any given organism, each one specialized for a different silencing activity or pathway (Filipowicz 2005). At the core of all RISC complexes is an Argonaute family protein, of which there are ten in *A. thaliana* and four in mammals. The Argonaute proteins are distinguished by two conserved motifs. One is a PAZ domain that binds specifically to the 3′ end of single-stranded RNA (ssRNA); thus the 2-nt 3′ overhangs of the siRNA duplex act as a required specificity determinant for RISC recognition and loading (Lingel et al. 2003; Song et al. 2003; Yan et al. 2003; Lingel et al. 2004; Ma et al. 2004). The other characteristic Argonaute motif is a PIWI domain, which adopts an RNAseH-like tertiary fold (Song et al. 2004). The PIWI active site in Argonautes which direct target cleavage (known as "Slicers"), contains an atypical Aspartate-Aspartate-Histidine catalytic triad involved in coordinating metal ions to facilitate nucleophilic attack and subsequent substrate cleavage. Many AGO proteins characterized to date appear to function in RISC-mediated silencing. For example, the catalytic engine of the plant miRNA-programmed RISC complex is AGO1, which cleaves target mRNA transcripts with complementarity to the RISC-bound miRNA (Baumberger and Baulcombe 2005; Qi et al. 2005), and the AGO2 protein in mammals executes the analogous "Slicer" function within the siRNA RISC complex, cleaving siRNA targets *in vitro* (Liu et al. 2004).

MicroRNAs are released from pre-miRNAs as miRNA:miRNA* duplexes, of which only the miRNA appears to be functional in downstream gene regulatory activities. Furthermore, miRNAs generally accumulate to much higher levels than miRNA* strands, as demonstrated by cloning and expression data (Bartel 2004). This difference in accumulation and function is attributed to the preferential incorporation of the miRNA strand into the RISC, where it guides sequence-specific interactions with target transcripts, while the miRNA* strand is excluded from RISC (Khvorova et al. 2003; Schwarz et al. 2003). The selectivity of the RISC likely protects the miRNA from cellular nucleolysis, because the single-stranded miRNA* sequence is subject to degradation, which may account for its short half-life in cells and its decreased abundance in cloning

datasets compared to the miRNA. The rapid turnover of miRNA* sequences, coupled with the imperfect pairing between the miRNA and miRNA* in the precursor hairpin, may protect newly-transcribed plant miRNA precursors from autosilencing by the miRNA* strand.

A model for the asymmetry of RISC loading emerged from *in vitro* biochemical studies of functional siRNA duplexes, in which it was determined that the strand with the least stably paired 5′ terminus was preferentially incorporated into RISC to guide silencing (the guide strand) while the strand with more thermodynamically favourable 5′ end pairing was excluded (the passenger strand) (Schwarz et al. 2003). This thermodynamic asymmetry model seems to explain strand selection for the majority of plant miRNAs (Rajagopalan et al. 2006). In *Drosophila*, the RISC-loading complex component R2D2 associates with the siRNA duplex at the more stably paired end, orienting its binding partner Dicer on the other end of the duplex such that unwinding proceeds from this terminus and asymmetric AGO2-containing RISC loading can be accomplished (Tomari et al. 2004). Furthermore, the guide strand specifies AGO2-mediated cleavage of the passenger strand, allowing the passenger strand to dissociate (Matranga et al. 2005). The binding partners DCL1/HYL1 in plants might similarly form a RISC-loading complex to orient the miRNA duplex for proper RISC programming, but if so, this leaves open the possibility that the plant miRNA RISC is assembled in the nucleus (where DCL1 and HYL1 are much more abundant) rather than in the cytoplasm. Supporting this hypothesis, at least some AGO1 protein co-localizes with HYL1 in nuclear D-bodies in addition to being present in the cytoplasm (Fang and Spector 2007).

Once assembled, the miRNA-programmed RISC becomes catalytically competent (Filipowicz 2005). The miRNA guides the RISC complex to messenger RNAs with a high degree of complementarity to the miRNA, and tethers the RISC via binding interactions with the mRNA target site so that AGO-directed cleavage at a single phosphodiester bond (typically opposite nucleotides 10 and 11 from the 5′ end of the miRNA) can occur within the dsRNA region (reviewed in (Bartel 2004). The RISC is thought to be catalytic, such that cleaved mRNA fragments are released but the miRNA remains bound for multiple turnover cleavage of other target substrates (Tang et al. 2003). RISC cleavage activity is present in wheat germ lysates and endogenous

miR165/miR166 can guide cleavage of an exogenous *Arabidopsis PHV* target mRNA (Tang et al. 2003). Furthermore, *in vitro* reconstituted *Arabidopsis* AGO1 immunoprecipitates and *Arabidopsis* whole cell extracts can be programmed with exogenous siRNAs and are cleavage competent (Qi et al. 2005).

miRNA targeting in plants

The rationale for grouping miRNAs into families based on extended sequence homology in plants and seed sequence identity in animals is closely tied to the canonical sequence specificity requirements for target repression in each kingdom. In *Arabidopsis* and other plant species, basepairing between the miRNA and the target transcript generally extends throughout the length of the miRNA, with few mismatches (Rhoades et al. 2002). When *Arabidopsis* miRNAs were first cloned and sequenced, it was noted that all mapped to strictly intergenic loci in the genome, with the exception of miR171, which matched one potential intergenic locus of origin with stemloop forming potential, and three loci that were perfectly antisense to protein-coding *SCARECROW-LIKE* (*SCL*) genes (Jones-Rhoades et al. 2006). This prompted the idea that plant miRNAs might direct cleavage of target transcripts bearing miRNA recognition sites with near-perfect complementarity (Llave et al. 2002a; Reinhart et al. 2002).

A genome-wide computational scan for ungapped, antisense miRNA target sites in protein-coding cDNAs, with a maximum of three mismatches to the miRNA, predicted targets for 11 of the 13 miRNA families then known to exist in *Arabidopsis* (Rhoades et al. 2002). Such stringent basepairing requirements have the potential to miss authentic sites, but minimize false positives. Evolutionary conservation of miRNA:target interactions in plants can be used as a filter to sensitize target prediction algorithms to sites with more relaxed pairing. For miRNAs that are conserved to *Oryza sativa* or other plant genomes, allowing additional mismatches or gaps can reveal miRNA sites when orthologous mRNA targets with intact complementary sites exist in both species; these would be missed with more stringent pairing rubrics (Jones-Rhoades and Bartel 2004). Generally, evolutionary conservation enhances the identification of authentic sites compared to false positive sites, dubbed the signal:noise ratio. However, target site prediction for non-conserved miRNAs and target genes is possible with more stringent

scoring rubrics that encapsulate position-specific parameters for mismatches observed in authentic miRNA target sites. For example, it has been noted for confidently predicted miRNA targets that miRNA:target pairing tends to be stronger at the 5′ end of the miRNA than at the 3′ end (Mallory et al. 2004b). Therefore, more stringent scoring rubrics can increase the signal:noise ratio for *bona fide* sites complementary to non-conserved miRNAs by including penalties for mismatches occurring between nucleotides 2-12 at the 5′ end of the miRNA, dubbed the "core" of the interaction (Allen et al. 2005; Schwab et al. 2005).

Plant miRNA target sites are typically internal to the open reading frame of the transcript. Unsurprisingly, the *SCL* genes matching the antisense of miR171 were the first to be experimentally validated as miRNA targets, and the scissile phosphate was mapped to the middle of the 20-22nt miRNA:target interaction site (Llave et al. 2002b). Cleavage of miRNA targets typically occurs between nucleotides in the mRNA that pair to positions 10 and 11 as counted from the 5′ end of the miRNA (Llave et al. 2002b; Kasschau et al. 2003; Tang et al. 2003). Many plant miRNA targets identified *in silico* have subsequently been confirmed by detection of cleavage fragments whose 5′ ends terminate within the complementary site, using 5′ RNA-Ligation Mediated Rapid Amplification of cDNA Ends (5′-RLM-RACE) (Llave et al. 2002b; Rhoades et al. 2002; Kasschau et al. 2003; Palatnik et al. 2003; Tang et al. 2003; Xie et al. 2003; Jones-Rhoades and Bartel 2004; Mallory et al. 2004a; Mallory et al. 2004b; Allen et al. 2005; Kim et al. 2005; Mallory et al. 2005; Lu et al. 2006; Rajagopalan et al. 2006)

Further proof-of-principle support for bioinformatic prediction of miRNA targets has come from genome-wide expression profiling studies (Allen et al. 2005; Schwab et al. 2005), often in genetic backgrounds in which components of the silencing machinery are compromised (Allen et al. 2005). Microarray profiling has also been used to identify new targets (Palatnik et al. 2003; Allen et al. 2005). In addition, when used to monitor mRNA levels in plants overexpressing one of four miRNAs, transcriptome profiling revealed that plant miRNA targeting is highly specific for transcripts with extensive sequence complementarity, with few to no interactions with non-cognate, partially complementary messages, (so-called "off-targets") (Schwab et al. 2005; Schwab et al. 2006). A similar degree of target selection specificity was observed for artificial miRNAs

rationally designed to target genes with easily-scored loss-of-function phenotypes (Schwab et al. 2006). This suggests that the silencing machinery is optimized for a high level of miRNA:target sequence complementarity, and that empirically-derived targeting determinants gleaned from known miRNA:target interactions are sufficient to guide *de novo* prediction of a substantial fraction of *bona fide* miRNA target genes.

Although in general the members of a family target the same set of genes for downregulation, miR159 and miR319 comprise a miRNA family because they share 17 identical nucleotides, but have subfunctionalized to downregulate non-overlapping sets of unrelated target genes (Palatnik et al. 2003; Parizotto et al. 2004; Palatnik et al. 2007; Reyes and Chua 2007). Combined mutations of miR159 at positions 7 and at the extreme 3′ end allow it to target the *TCP4* gene, which is the primary target of miR319, indicating that sequence accounts for most of the reason that *TCP* genes escape miR159 regulation *in vivo*; conversely, though miR319 is sufficiently complementary to direct cleavage of miR159 *MYB* targets, its abundance is surpassed by that of miR159 in tissues where these *MYB* genes are expressed and therefore cleavage of these targets is predominantly effected by miR159 (Palatnik et al. 2007). Together these results suggest that the action of related miRNAs in plants can be modulated by both expression effects and sequence evolution to allow precise discrimination of regulatory targets.

Generally, miRNA regulatory functions at the protein level, such as translational repression or other mechanisms inhibiting protein production or accumulation, appear restricted to animal systems. The possibility of animal-like miRNA targeting sites in plants formally exists, but there is little evidence that sites with limited complementary are functional in plants. Computational prediction of animal-like seed sites containing conserved Watson-Crick basepairing between 3′ UTRs and the 5′ region of miRNAs does not yield more matches than expected by chance (Jones-Rhoades et al. 2006), even though similar algorithmic approaches in animals have yielded thousands of metazoan target predictions above the level of noise (Lewis et al. 2003; Brennecke et al. 2005; Krek et al. 2005; Lewis et al. 2005).

The extensive miRNA:target basepairing is thought to be intimately tied to the downregulation of plant miRNA targets by mRNA cleavage. However, at least one example suggests the potential for other forms of target repression: both mRNA and

protein levels appear to be under the control of miRNA-directed regulation for the miR172 *AP2* transcription factor targets *APETALA2*, *TOE1* and *TOE2*. In wild-type and miR172 overexpressing plants, *APETALA2* (Aukerman and Sakai 2003; Kasschau et al. 2003; Schwab et al. 2005), *TOE1* and *TOE2* cleavage fragments (Schwab et al. 2005) consistent with mirR172-directed cleavage have been detected, as expected for plant miRNA targets. However, the mRNA levels of *APETALA2* (Aukerman and Sakai 2003; Chen 2004), *TOE1* and *TOE2* (Aukerman and Sakai 2003) remain relatively unchanged in miR172 overexpressing plants, and APETALA2 protein levels decrease (Aukerman and Sakai 2003; Chen 2004), suggesting that miR172 regulation can also specify translational repression of target mRNAs or destabilization of target proteins. A confounding factor derives from APETALA2-mediated repression of its own transcription, which may explain why *APETALA2* levels appear to remain constant (Schwab et al. 2005), and points to the importance of multiple regulatory mechanisms at the mRNA and protein level which govern the accumulation of miR172 targets.

MicroRNA targeting and regulation in metazoans

Whereas plant miRNAs direct cleavage of target transcripts, animal miRNAs typically interfere with protein synthesis without inducing cleavage (Jackson and Standart 2007; Pillai et al. 2007). The sequence complementarity between the miRNA and the target transcript necessary to trigger translational repression is relatively limited (Doench and Sharp 2004). The 5′-proximal ~7-nt region known as the "seed" of the miRNA perfectly basepairs with a site in the 3′ untranslated region (UTR) of the target transcript, and this seed appears to be the primary specificity determinant for targeting, though it is sometimes bolstered by additional basepairing at the 3′ end of the miRNA (Lewis et al. 2003; Doench and Sharp 2004). The importance of the seed region for miRNA targeting has been demonstrated by genome-wide prediction of hundreds of target gene UTRs with seed complementary sites and experimental validation of the efficacy of a handful of these sites in mediating repression (Lewis et al. 2003). Subsequent analyses have identified other sequence determinants that contribute to the efficacy of miRNA targeting, including an "adenosine anchor" nucleotide at position 1 of the miRNA (Lewis et al. 2005). Local sequence context surrounding the target site also plays a role, with

several features enhancing miRNA target recognition including AU-rich nucleotide composition, pairing to residues 13–16 of the miRNA, and location of the site at a distance from the stop codon and from the center of long UTRs (Grimson et al. 2007).

Animal miRNAs regulate target genes by directing RISC-mediated translational repression of transcripts, with the notable exception of mirR196, the only animal miRNA known to be capable of directing the cleavage of an mRNA target, *HOXB8* (Yekta et al. 2004). The complementarity between miR196 and *HOXB8* is quite extensive and resembles the magnitude of basepairing seen for most plant miRNAs and their targets. It also resembles synthetic siRNAs which have been widely employed in animal systems to turn off genes by acting through RNAi pathways; they are perfectly antisense to their targets mRNAs, whose cleavage they direct. Supporting the idea that the extent of target complementarity is the main determinant of small RNA function, miRNAs can enter the RNAi pathway and direct cleavage of perfectly complementary targets (Hutvagner and Zamore 2002; Zeng et al. 2003) and conversely, siRNAs can specify translational repression of messages with partial complementarity (Doench et al. 2003; Zeng et al. 2003).

The limited basepairing potential exhibited by most predicted target sites to animal miRNAs appears inadequate to promote efficient target cleavage. Instead, the activity of most animal miRNAs appears to result in reduced protein accumulation, (Jackson and Standart 2007; Pillai et al. 2007), though at least some mRNA transcripts are destabilized, leading to reduced levels (Bagga et al. 2005; Giraldez et al. 2005; Lim et al. 2005; Chendrimada et al. 2007). Evidence supporting several different mechanisms of translational repression has accumulated, including inhibition of translational initiation via RISC interactions with the mRNA 5′-meG cap binding proteins and interfering with ribosome/polysome elongation and thus inhibiting the productive translation of mRNA to protein (Jackson and Standart 2007; Pillai et al. 2007). These activities may be highly localized, since Argonautes, miRNAs and target messages localize to cytoplasmic processing bodies (P-bodies) (Liu et al. 2005b; Pillai et al. 2005; Sen and Blau 2005), which appears to be critical for miRNA-dependent translational repression of target mRNAs (Liu et al. 2005a; Rehwinkel et al. 2005).

About 1/3 of human genes are predicted to be targeted by miRNAs, and many of them likely in a combinatorial fashion employing several miRNAs binding seed sites simultaneously in a single 3′ UTR (Bartel and Chen 2004; Krek et al. 2005; Lewis et al. 2005). The range of biological processes in which these genes are involved suggests a much broader role for miRNA regulatory networks in animals than in plants. It is postulated that animals employ miRNA activity to achieve the network architecture provided by a "combinatorial rheostat", in which each miRNA binding site contributes a subtle downregulatory effect to dampen protein expression below a threshold level needed for a particular molecular phenotypic outcome (Bartel and Chen 2004).

Evolutionary conservation of plant miRNAs

Only two plant genomes had been fully sequenced at the time that miRNAs were first described in plants, those of the angiosperms *Arabidopsis thaliana* (a eudicotyledonous plant, or dicot) and *Oryza sativa* (a monocotyledonous plant, or monocot). The genome of another dicot, *Populus trichocarpa* (black cottonwood) and the draft genome of the moss *Physcomitrella patens*, a basal plant lacking both seeds and vasculature, followed after. These four genomes are considered representative of core eudicot (eurosids I), monocots, core eudicot (eurosids II), and bryophytes, respectively. The split of seed plants, (including *Arabidopsis*, rice, and poplar) from basal land plants (including moss) is dated at ~450 million years (Quatrano et al. 2007). Recent molecular estimates using 12 fully sequenced chloroplast genomes from land plants peg the monocot split from dicots at 140-150 million years ago. The split of the eurosids I and II is dated at ~90 million years ago (Wikstrom et al. 2001). Therefore, these four genomes together represent almost half a billion years of evolution.

Of the 97 miRNA genes identified prior to this work, the vast majority have been retained along both of the major angiosperm lineages. 92 miRNA genes (21 families) are conserved to the eudicot poplar, and 20 families are conserved to *O. sativa*, a monocot. Large miRNA gene families in *A. thaliana* tend to be large in rice and poplar, and small families in *Arabidopsis* are likewise similarly represented across other available genomes. Some of the larger *Arabidopsis* gene families are greatly expanded in rice or poplar (Jones-Rhoades et al. 2006). Within a family of paralogs, the miRNA sequence

and to a lesser degree the miRNA* sequence is highly conserved, while the flanking sequence in the precursor, including the loop, can vary widely between family members in length, sequence and predicted structure. However, consistent with shared evolutionary origins, the miRNA is always found on the same arm of the hairpin in all paralogs and orthlogs (Jones-Rhoades et al. 2006). The pattern of mismatches between the miRNA and miRNA* along the stem is also often identical between orthologous miRNAs; it is postulated that they may orient proper DCL1-guided processing(Jones-Rhoades et al. 2006). Three apparently monocot-specific miRNAs identified in rice, miR437, miR444, and miR445, have orthologs in a related cereal crop, *Zea mays* (maize). In addition, miR403 appears to be a eudicot-specific miRNA, with orthologs in *A. thaliana* and *Populus trichocarpa*.

Other data establishes conservation of some miRNAs to phylogenetic branches of land plants that predate the emergence of angiosperms. At least 11 miRNA families have orthologs in gymnosperms, seven appear conserved to lycopods, and at least seven have orthologs in moss (Jones-Rhoades and Bartel 2004; Arazi et al. 2005; Axtell and Bartel 2005; Axtell et al. 2007). Furthermore, target sites for at least 10 miRNA families are found in orthologous target genes in gymnosperms or more basal plants (Jones-Rhoades and Bartel 2004) and miRNA:target interactions in gymnosperms, ferns, lycopods or mosses have been experimentally verified for miR160, miR166, miR167, miR171 and miR172; the targets are orthologous to validated *Arabidopsis* targets for these families (Floyd and Bowman 2004; Axtell and Bartel 2005; Axtell et al. 2007). These ancient regulatory circuits appear to have been preserved over several hundred million years of evolution, even though the developmental programs and architecture of these basal plants differs considerably from that of the angiosperms.

The conservation of miRNA target sites in orthologous mRNAs for the 20 miRNA families present in *A. thaliana*, *P. trichocarpa*, and *O. sativa* suggests that these miRNAs are performing similar functions in these evolutionarily divergent species. Still, some miRNA families have greatly expanded in certain lineages and may have diversified for species-specific functions. Such an example is presented by the miR397 family, which is predicted to target over 26 laccases in *P. trichocarpa* compared to three in *A. thaliana*, and which is presumably involved in lignification processes more

important in woody species than herbaceous ones (Jones-Rhoades et al. 2006). Interestingly, sequence-unrelated miRNAs can perform similar regulatory functions, suggesting convergent evolution. For instance, sequence-unrelated miRNAs embedded in the introns of DCL1 genes in *Arabidopsis* and moss establish an auto-regulatory feedback loop by which DCL1 can keep its own mRNA levels in check by short-circuiting proper splicing (Rajagopalan et al. 2006; Axtell et al. 2007).

Non-conserved miRNAs in plants

Prior to this work, there was minimal precedent for confidently-annotated non-conserved miRNAs. Such miRNAs lack orthologs in public sequence databases and are postulated to have evolved more recently, but are abundant enough in wild-type plants to emerge among the most likely miRNA candidates in cloning experiments (Park et al. 2002; Reinhart et al. 2002). Of the five miRNAs in *A. thaliana* characterized before this work that do not appear to be conserved to other plants (miR158, miR161, miR163, miR173, and miR447), the latter four have experimentally-confirmed cleavage targets (Allen et al. 2004; Allen et al. 2005), significantly bolstering their miRNA designation. It remains to be seen whether these miRNAs have orthologs within the *Arabidopsis* genus or other closely related eurosids II not represented in the public sequence databases. Two of these non-conserved miRNAs (miR161 and miR163) appear to have recently evolved from inverted duplication of their target genes, a mechanism that might be expected to give rise to novel miRNA hairpins with relative ease, and allow the plant to sample regulatory circuits which may prove advantageous and be retained, or prove of no use or even deleterious, and drift or be selected against (Allen et al. 2004). Homology extending into the hairpin arms beyond the region encoding the miRNA exists between the *MIR163* locus and three nearby S-adenosyl-$_L$-Met: salicylic acid carboxyl methyltransferase-like genes as well as the *MIR161* locus and four pentatricopeptide repeat (*PPR*) genes. The retention of such homology suggests that these two miRNAs represent young genes in *Arabidopsis* that have not had sufficient time to diverge to the point where they would be unrecognizeable from their parental/target genes, a hypothesis consistent with the lack of detectable orthologs for these miRNAs in other plant genome sequences or expressed sequence tag (EST) datasets.

Conservation is a criterion hardwired into algorithmic approaches to miRNA discovery, but its applicability is limited to the identification of miRNAs presumably central to the definition of plant body plans and thus preserved over evolutionary timescales. The possibility that genomes encode many apparently recently-emerged but weakly expressed miRNAs has long been considered, given the comparatively simple anatomy of a *MIR* locus. However, the challenge of identifying young non-abundant miRNAs cannot be easily surmounted using the work-intensive and time-consuming conventional sequencing technologies, especially against the background of the enormous diversity of endogenous small RNAs that are not miRNAs but are perhaps as equally abundant as feebly expressed miRNAs. Apparently non-conserved miRNAs remain formally indistinguishable from endogenous siRNAs in the absence of miRNA-like biogenetic evidence, conservation of target genes and targets sites, or experimental confirmation of target cleavage, even if the predicted stem-loop precursor satisfies the structural and pairing constraints observed among the known miRNAs (Jones-Rhoades et al. 2006). As described in Chapter 1, the sequencing depth provided by 454 high-throughput technology allowed us to revisit the issue of miRNA annotation criteria and expand the lines of evidence that can be used to confidently classify miRNAs for those which are infrequently retrieved from small RNA libraries.

Cellular distribution of miRNAs

MicroRNAs can be highly abundant in cellular contexts, with abundance estimates as high as 50,000 copies per cell for some *C. elegans* miRNAs (Lim et al. 2003b). Though similar quantification in plants has not been performed, plant miRNA localization has been assayed by tissue-specific cloning and gel blotting (Reinhart et al. 2002; Rajagopalan et al. 2006; Fahlgren et al. 2007), microarray (Axtell and Bartel 2005), *in situ* hybridization (Chen 2004; Kidner and Martienssen 2004; Williams et al. 2005; Valoczi et al. 2006), sensor constructs comprised of reporters bearing miRNA complementary sites (Parizotto et al. 2004), and miRNA promoter:reporter constructs (Parizotto et al. 2004; Vaucheret et al. 2006; Wu et al. 2006). Some miRNAs are responsive to hormone treatments, growth conditions, or stresses, or induced by environmental stimuli (discussed below).

miRNA-based regulation during plant development

Studies of *Arabidopsis* mutants with pleiotropic developmental defects identified the *DCL1*, *AGO1*, and *HEN1* genes long before their gene products were implicated in small RNA biogenesis and function. This is because mutating factors involved in miRNA biogenesis often results in dramatic developmental consequences. Severe *dcl1* mutants have early embryonic lethal phenotypes. The partial loss-of-function mutant *dcl1-9* with a disrupted dsRNA-binding domain exhibits defects in floral and embryonic patterning and axillary meristem initiation, in addition to abnormal proliferation of floral meristems (Jacobsen et al. 1999). Hypomorphic alleles of *ago1* (Bohmert et al. 1998), *hen1* (Chen et al. 2002), *hyl1* (Lu and Fedoroff 2000), and *hasty* (Telfer and Poethig 1998) all display developmental phenotypes that partially overlap with those of *dcl1* hypomorphs, in agreement with prevailing models implicating these genes in overlapping small RNA biogenetic pathways presumably debilitated in these mutants. In addition, mutants in *ago1*, *hen1* and *hyl1* all exhibit upregulated levels of target mRNAs (Boutet et al. 2003; Vaucheret et al. 2004; Vazquez et al. 2004a), consistent with miRNAs playing a role in downregulating target messages.

Many of the target genes predicted to be regulated by the first identified plant miRNAs are key transcription factors involved in specifying plant morphology and differentiation during development. Of the 95 conserved targets for the 21 miRNA families conserved to *Populus trichocarpa*, 65 (68%) are transcription factors (Jones-Rhoades et al. 2006). It has been proposed that miRNAs are critical to clearing regulatory gene transcripts during differentiation and cell-fate decisions (Rhoades et al. 2002). Other gene families known to be under miRNA control prior to this study include F-box proteins or uniquitin conjugating enzymes involved in proteosome-mediated protein degradation, as well as ATP sulfurylases, pentatricopeptide repeat proteins, superoxide dismutases, and laccases. In addition, plant miRNAs are such potent architects of cell fate that the silencing machinery components AGO1 and DCL1 are themselves targets of miRNA-directed downregulation (Rhoades et al. 2002; Xie et al. 2003; Vaucheret et al. 2004), giving rise to internal negative feedback loops that maintain levels of miRNAs and miRNA-associated factors at homeostatic levels.

Several miRNA:target site interactions have been validated by *in vivo* studies, uncovering anomalous phenotypes and biological defects associated with miRNA misregulation. Two reverse genetic approaches have become the *de facto* standard for revealing molecular functions and growth abnormalities that place miRNAs at the center of developmental regulatory circuits. The first is to over-express the miRNA as a transgene under a strong promoter like the Cauliflower Mosaic Virus (CaMV) 35S promoter. The second is to mimic disruption of miRNA recognition of target transcripts by expressing a predicted or confirmed cleavage target as a transgene harboring a mutated miRNA complementary site, under a strong 35S promoter or a native promoter. Silent mutations are made to preserve the underlying amino acid sequence while significantly disrupting the miRNA:target basepairing interaction. The transcripts of these miRNA-resistant transgenes are unresponsive to miRNA-mediated regulation, in many cases displaying an array of morphological defects that strongly support the pivotal role that miRNAs play in development. Sometimes these mutants phenocopy plants overexpressing the target gene beyond the regulatory capacity of endogenous levels of miRNA.

miRNAs involved in leaf and vascular development

The class III HD-ZIP transcription factor genes *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), and *REVOLUTA* (*REV*) specify abaxial-adaxial polarity in lateral organs such as leaves and central-peripheral polarity in vascular patterning. Gain-of-function *phb* and *phv* mutants with disrupted miR166 complementary sites exhibit abaxial (lower) to adaxial (upper) leaf transformations in *Arabidopsis* (McConnell et al. 2001; Kim et al. 2005; Williams et al. 2005), which are recapitulated in transgenic plants expressing miR166-resistant versions of *PHB* (Mallory et al. 2004b). Similarly, adaxialized leaf primordia result from dominant mutations in the miR166 complementary site of the maize *RLD1* gene (Juarez et al. 2004). Dominant *rev* mutations result in radialized vascular bundles in the stem, disrupting the wild-type ring-like collateral structure of central (adaxial) xylem surrounded by peripheral (abaxial) phloem in *Arabidopsis* (Emery et al. 2003; Zhong and Ye 2004).

The overexpression of miR319 results in crinkly leaf morphologies. Overexpression of miR319-resistant TCP4 results in arrested seedlings with fused cotyledons and missing apical meristems (Palatnik et al. 2003). Overexpression of miR159 results in male sterility (Achard et al. 2004), while overexpression of a miR159-resistant target gene, MYB33, results in abnormally rounded and upturned leaves (Palatnik et al. 2003), and stunted petioles (Millar and Gubler 2005).

The vegetative developmental transition from juvenile to adult in plants is attenuated in miR156 overexpressing plants, and accelerated by constitutive expression of the three miR156 targets, *SQUAMOSA PROMOTER LIKE3* (*SPL3*), *SPL4*, and *SPL5*. These *SPL* overexpressers display even more precocious development, such as earlier abaxial trichome production and reduced petiole length, when the miR156 complementary sites are mutated (Wu and Poethig 2006).

miRNAs involved in flower development

Many miRNAs appear to be involved in patterning of flowers and floral organ identity. Sepal-to-carpel floral homeotic transformations result from overexpression of miR172, as well as a loss of petals in developing floral primordia (Aukerman and Sakai 2003; Chen 2004). In addition, many of the *DCL1 sin1* and *caf1* mutants display delayed conversion of inflorescence meristems to floral growth, and floral meristem overproliferation (Ray et al. 1996; Jacobsen et al. 1999), suggesting that the global disruption of miRNA production has negative consequences for the formation of floral organs. The miR164 targets *NAC*, *CUC1*, and *CUC2* are important for specifying boundaries of adjacent embryonic, vegetative and floral organs (Laufs et al. 2004; Mallory et al. 2004a).

miRNAs and hormone sensing

The phytohormone auxin plays a central role in *Arabidopsis* development, and several miRNA regulatory circuits are integrated within the auxin signaling pathway. Examples include three F-box proteins targeted by miR393 which are auxin receptors (Dharmasiri et al. 2005a; Dharmasiri et al. 2005b; Kepinski and Leyser 2005) that target repressors of AUXIN-RESPONSE FACTOR (ARF) transcriptional activators for

ubiquitin-mediated proteolysis upon induction by auxin. Many *ARF* genes are themselves miRNA targets (Mallory et al. 2005; Wu et al. 2006), such as miR160-targeted ARF17, whose miR-resistant phenotypes include upward curling of leaf margins, leaf serration, reduced vegetative size, and abnormal phyllotaxy along the primary and lateral stems (Mallory et al. 2005).

miRNAs induced by nutrient stress

At least two miRNAs participate in nutrient uptake pathways under starvation conditions and are specifically upregulated in response to environmental shifts related to the availability of these nutrients. Under standard growth conditions, miR395 is weakly expressed (Jones-Rhoades and Bartel 2004; Rajagopalan et al. 2006), but sulfate deprivation activates its expression to levels that allow potent downregulation of *ATP SULFURYLASE*, a sulfate-metabolizing enzyme (Jones-Rhoades and Bartel 2004) and a sulfate transporter, *ATSULTR2;1* (Allen et al. 2005). Thus miR395 suppresses genes involved in sulphate assimilation under conditions in which sulphate is deficient. Conversely, miR399 enhances phosphate uptake when plants are starved for this macronutrient. miR399 is induced by low-phosphate stress, leading to cleavage of transcripts encoding the ubiquitin-conjugating enzyme *UBC24*, and enhancing phosphate accumulation and retention in the plant by the induction of a phosphate transporter gene (Fujii et al. 2005).

Endogenous short interfering RNAs (siRNAs) in *Arabidopsis*

A brief history of siRNAs and RNA silencing

Post-transcriptional gene silencing (PTGS) of endogenous genes was first associated not with miRNAs but with siRNAs. It was discovered in petunia plants overexpressing a *chsA* transgene encoding chalcone synthase A, involved in the biosynthesis of purple anthocyanin pigment in petals. Strongly-expressed transgenes led surprisingly not to more purple flowers, but to loss of anthocyanin in nonclonal sectors throughout the petals, and a mosaic pattern that included white sections of cells in which the *chsA* gene was not expressed. The phenomenon was observed to be reversible,

coordinate (involving a *trans* interaction between the transgene and the endogene that resulted in lowered expression of both), and heritable, and was termed cosuppression (Napoli et al. 1990). A similar phenomenon called quelling was noted in the filamentous fungus *Neurospora crassa* (Romano and Macino 1992), and in animals, the introduction of double-stranded RNA, rather than single-stranded RNA, was shown to be the most effective substrate for a process called RNA interference (RNAi) in which expression of homologous endogenous genes was silenced by clearing of mRNA transcripts (Fire et al. 1998). It took several years before it was discovered that small RNA production was concomitant with transgene introduction and co-suppression (Hamilton and Baulcombe 1999) and that small RNAs were also the silencing signals mediating RNAi in animals (Zamore et al. 2000; Elbashir et al. 2001), and quelling in fungi (Catalanotto et al. 2002). These small RNAs are now known as short interfering RNAs (siRNAs).

In addition to being produced from exogenously introduced sequences, siRNAs also arise from endogenous loci. During early cloning projects aimed at identifying plant miRNAs, many other small RNA sequences were recovered which did not correspond to loci with hairpin-coding or protein-coding potential (Llave et al. 2002a; Reinhart et al. 2002). However, it was unclear if any of them played a role in silencing. In *C. elegans*, small RNAs apparently derived from non-coding loci were called tiny non-coding RNAs (tncRNAs) (Ambros et al. 2003b). In addition, a large proportion of cloned endogenous small RNAs in *C. elegans* matched the antisense of protein-coding genes and were postulated to specify the silencing of the complementary mRNAs arising from these loci in a fashion analogous to RNAi (Lau et al. 2001; Ambros et al. 2003b; Lim et al. 2003b). The genic yet antisense origins of these small RNAs inspired provisional designation as endogenous short interfering RNAs (siRNAs) (Ambros et al. 2003b). More recent large-scale characterization of the small RNA complement in various species suggests that endogenous small RNA production is much more extensive in plants than in animals (Lu et al. 2005; Henderson et al. 2006; Rajagopalan et al. 2006; Ruby et al. 2006).

Biogenesis of endogenous siRNAs

The mechanisms of siRNA biogenesis in both plants and animals are much less well-defined at the molecular level than they are for miRNAs. However, siRNAs are

distinct in a number of key biogenetic steps, not least of which is the nature of the precursor RNA itself: siRNAs are derived not from stemloop structures but from long bimolecular dsRNA. This dsRNA can arise in one of three ways: from the action of an RNA-dependent RNA polymerase (RdRP) on a primary transcript to generate the reverse complement, generating the complementary strand; from convergent transcription of loci generating overlapping transcripts with complementarity at their mutual ends; and from the expression of an inverted repeat locus with self-complementarity within the primary transcript. Many siRNA-spawning dsRNAs undergo subsequent processing by a DCL protein.

The three characterized classes of endogenous siRNAs in *A. thaliana* are the *trans*-acting siRNAs, the natural antisense siRNAs (nat-siRNAs), and the heterochromatic siRNAs. Different biogenetic signatures have been associated with each class of siRNA in *A. thaliana*, according to their dependence on different DCL and RdRP proteins during processing of the precursor transcript. In addition to these biogenetic differences, the three classes are also distinguished by their primary mode of action. At least two classes of endogenous siRNAs in plants have been shown to inhibit gene expression by PTGS rather than TGS, the *trans*-acting siRNAs (tasiRNAs) (Peragine et al. 2004; Vazquez et al. 2004b) and the nat-siRNAs (Borsani et al. 2005; Katiyar-Agarwal et al. 2006). The tasiRNAs were so named because, though their biogenesis is more closely related to that of other siRNAs, their activity is reminiscent of miRNAs, in that they direct sequence-specific silencing of loci unrelated to their locus of origin.

Trans-acting siRNAs

Perhaps the most well-characterized of the endogenous plant siRNA pathways is the *trans*-acting siRNA (tasiRNA) pathway, which was independently converged upon by two groups studying PTGS mutants in *Arabidopsis* (Peragine et al. 2004; Vazquez et al. 2004b). Previous genetic screens for mutants impaired in transgene PTGS in plants had identified several candidate loci that were shown to be involved in the production of siRNAs at endogenous loci. Two of these were *SUPPRESSOR OF GENE SILENCING3* (*SGS3*), a plant-specific coiled-coil protein of unknown function (Mourrain et al. 2000) and *SUPPRESSOR OF GENE SILENCING2* (*SGS2*)/*SILENCING DEFECTIVE1*

(*SDE1*)/*RNA-DEPENDENT POLYMERASE6* (*RDR6*), which encodes an RdRP (Elmayan et al. 1998; Dalmay et al. 2000). Loss-of-function mutations in both genes lead to accelerated juvenile-to-adult phase change in *Arabidopsis*, the most prominent morphological features of which are slightly pointed, elongated leaves and downward-curled leaf margins (Peragine et al. 2004).

A closer analysis of *rdr6* (*sgs2-1*) molecular phenotypes revealed several upregulated genes, among them the locus *At2g27400*, which contained a single open reading frame of 21 amino acids and therefore did not appear to encode a protein (Peragine et al. 2004; Vazquez et al. 2004b). Polyadenylated and capped transcripts from the locus were detected, and the gene had canonical splice-site donor and acceptor sequences. However ESTs mapping to the gene were truncated within the single intron, suggesting that full-length transcript does not accumulate or that the primary transcripts undergo immediate processing (Vazquez et al. 2004b). Small RNAs matching both strands of the intron with almost perfect 21-nt phasing were observed, but the host locus was not predicted to fold into stable hairpin structures, suggesting that non-miRNA mechanisms were giving rise to the siRNAs (Peragine et al. 2004; Vazquez et al. 2004b). Furthermore, the majority of the small RNA sequences that were cloned obeyed the asymmetry rules for RISC incorporation; in conceptual duplexes containing the cloned strand and the perfectly antisense siRNA with 2-nt 3′ overhangs, the strand with less energetically stable 5′ pairing was most frequently the cloned strand, suggesting that these small RNAs had been protected within RISC complexes and thus might be cleavage-targeting siRNAs (Vazquez et al. 2004b). In an independent study, *rdr6* and *sgs3* mutants were found to have increased expression of *At5g18040*, which displayed sequence complementarity to several previously cloned small RNAs (Peragine et al. 2004). The small RNAs identified by these two groups were named *trans*-acting siRNAs because they exhibited complementarity to protein-coding genes and targeted these genes for PTGS, as evidenced by cleavage products with 5′ ends mapping to the complementary sites (Peragine et al. 2004; Vazquez et al. 2004b).

The accumulation of the tasiRNAs depends on the PTGS proteins *SGS3* and *RDR6* (Peragine et al. 2004; Vazquez et al. 2004b), providing the first evidence of a role for these proteins in silencing endogenous genetic elements. Though the function of

SGS3 in tasiRNA biogenesis is unclear, RDR6 apparently generates the antisense strand of the primary transcript, generating a long dsRNA (Fig. I.3). In addition, the *trans*-acting siRNAs were shown to be reduced in *DCL1* (Peragine et al. 2004; Vazquez et al. 2004b), *HYL1*, *AGO1*, and *HEN1* mutants (Vazquez et al. 2004b), but not in *AGO7/ZIPPY* mutants which display phenotypes similar to *sgs3* and *rdr6* (Peragine et al. 2004; Vazquez et al. 2004b). A previously uncharacterized protein known as SDE5 has also been implicated in the biogenesis of *trans*-acting siRNAs, since their levels are reduced in *sde5* mutants (Hernandez-Pinzon et al. 2007).

The dependence on *DCL1* suggested that the tasiRNA dsRNA precursor was processed by DCL1. However, it was subsequently noted that at least four other *trans*-acting siRNA loci were encoded in the *A. thaliana* genome, and that all five so-called *TAS* loci possessed miRNA complementary sites (Allen et al. 2005). In a computational scan of EST datasets for miRNA complementary sites, it was found that *TAS1a*, *TAS1b*, and *TAS1c* (all expressing one siRNA sequence in common, siR480(+)/siR255) and *TAS2* all have miR173 sites at the 5′ end of their transcripts that are targeted by miR173 for cleavage (Allen et al. 2005; Yoshikawa et al. 2005), while cleavage of *TAS3* transcripts occurred at a miR390 complementary site at the 3′ end (Allen et al. 2005). The accumulation of tasiRNAs was shown to be miRNA-dependent in a *Nicotiana benthemiana* transient expression system. MicroRNA-directed cleavage of the *TAS* primary transcripts establishes a defined 5′ terminus that sets the phasing register for successive cleavage by DCL in 21-nucleotide intervals, giving rise to tasiRNAs with specific sequences (Allen et al. 2005). Presumably, this defined end acts as a termination or initiation signal (depending on the *TAS* locus) for RDR6-guided synthesis of the antisense strand. Without such a precisely defined end, each transcript from a *TAS* locus could presumably give rise to tasiRNAs varying slightly in sequence, depending on the nucleotide position at which DCL cleavage initiated. This in turn would decrease the efficacies of tasiRNAs in guiding sequence-specific cleavage of cognate target transcripts, and might even lead to disadvantageous off-target downregulation.

The dependence of tasiRNA accumulation on *DCL1*, *HYL1*, *HEN1* and *AGO1* could be rationalized by the requirement for miRNA-mediated cleavage to generate the proper 5′ terminus for faithful tasiRNA production, although it remains formally possible

that some of these proteins act later in the tasiRNA pathway. For example, HEN1 may be involved in methylating the 3′ end of tasiRNAs in a fashion analogous to miRNA modification (Li et al. 2005) (Fig. I.3). The abundant siR480(+)/siR255 *trans*-acting siRNA from *TAS1* co-immunoprecipitates with AGO1 (Borevitz et al. 2000; Baumberger and Baulcombe 2005) and AGO1 immunoprecipitates catalyze cleavage of synthetic *At4g29770* target mRNA *in vitro* (Qi et al. 2005). Though this suggests that *TAS1* siRNAs are recruited by AGO1-RISC, genetic data (described below) implicate AGO7 as the best candidate for *TAS3* tasiRNA-mediated target cleavage.

A few of the initially reported tasiRNAs were almost perfectly antisense to their targets (Peragine et al. 2004; Vazquez et al. 2004b), suggesting that tasiRNA target prediction might be amenable to the same computational approaches employed for miRNAs (Allen et al. 2005). Indeed, transcripts for *AUXIN RESPONSE FACTOR3* (*ARF3*) and *ARF4* are upregulated in *rdr6* mutant plants, and contain complementary sites corresponding to two tasiRNAs from the *TAS3* locus which were shown to be cleaved using 5′ RACE (Allen et al. 2005). Though the *TAS1* and *TAS2* loci, as well as miR173, do not appear to be conserved in any other sequenced plant genomes, ESTs from other monocot and dicot species align well with the *TAS3* miR390 complementary site and the two ARF-targeting tasiRNAs in *A. thaliana* (Allen et al. 2005).

TAS2 and *TAS3* mutants have no visible phenotypes. Almost all of the mild vegetative phenotypes of *rdr6*, *sgs3* and *ago7/zippy* mutants have been attributed to the disabled activity of the two tasiRNAs from *TAS3* which downregulate ARF3 and ARF4, because the elongated rosette leaf morphology is phenocopied in *TAS3* mutants. *In vivo* studies of the biological interactions between ARF3 and *TAS3* implicate both in rosette leaf patterning and establishing leaf polarity (Adenot et al. 2006; Fahlgren et al. 2006; Garcia et al. 2006; Hunter et al. 2006), and curiously, *AGO7/ZIPPY* is required for the accumulation of *TAS3* tasiRNAs (Adenot et al. 2006; Fahlgren et al. 2006; Hunter et al. 2006). The precocious vegetative phase change phenotypes associated with *sgs3*, *rdr6*, and *ago7* mutants are almost identically recapitulated in plants expressing a nontargeted version of *ARF3* whose regulation by *TAS3* tasiRNAs is disrupted (Fahlgren et al. 2006), or overexpressing *ARF3* (Hunter et al. 2006). Perhaps unsurprisingly, *arf3* mutants suppress the *zip* phenotype (Hunter et al. 2006). The *RDR6-SGS3-AGO7* pathway also

genetically interacts with the *ASYMMETRIC LEAVES1* (*AS1*)-*AS2* pathway to specify proper leaf morphogenesis (Garcia et al. 2006; Xu et al. 2006). Thus, it appears that the most obvious morphological phenotypes associated with hypomorphic mutants in the *TAS* pathway overlap very well with those of the *TAS3* tasiRNA target *ARF3*. The role played by the other *TAS* loci in the biology of *A. thaliana* remains an open question.

Which of the four plant Dicers processes the tasiRNA dsRNA precursor? Plants homozygous for the presumed null alleles *dcl4-1* and *dcl4-2* show decreased accumulation of *TAS1*, *TAS2* and *TAS3* tasiRNAs, and misprocessing of *TAS1* and *TAS2* tasiRNAs (Gasciolli et al. 2005; Xie et al. 2005b; Yoshikawa et al. 2005). Furthermore, just as DCL1 cooperates with HYL1 to process miRNAs, DCL4 must partner with a dsRNA binding protein, known as DRB4, with which it interacts *in vitro* (Hiraguri et al. 2005). DRB4 shares sequence similarity with HYL1, and *drb4-1* hypomorphs (with a T-DNA insertion between the transcription start site and the first ATG that eliminates mRNA accumulation) display defects in leaf patterning similar to *ago7/zip*, *rdr6*, *sgs3*, and *tas3* mutants (Adenot et al. 2006). Curiously, *drb4-1* mutants show reduced accumulation of *TAS3* and *TAS1* tasiRNAs in leaves and only a slight reduction in *TAS2* tasiRNAs, while in inflorescence tissues the levels of *TAS1* and *TAS2* tasiRNAs are much more affected than *TAS3* tasiRNAs. Together, these results suggest that DCL4 is the major Dicer responsible for tasiRNA biogenesis, and that DRB4 is likely the primary interaction partner of DCL4, though other DRBs may functionally compensate for DRB4 deficiency (Adenot et al. 2006) (Fig. I.3).

While *TAS* loci produce abundant small RNAs, very few small RNAs emanate from loci encoded by other miRNA cleavage targets (Lu et al. 2005; Rajagopalan et al. 2006; Ronemus et al. 2006). Why are *TAS* transcripts funneled through the RDR6-SGS3-DCL4 pathway, but the canonical protein-coding targets of miRNAs are not, following miRNA cleavage? The finding that dual miR390 complementary sites flank four non-coding loci in the moss *Physcomitrella patens* prompted the discovery that *AtTAS3* also contains a 5′ miR390 site (albeit with unusually weak complementarity), that this site is conserved within angiosperms, and that this site binds miR390 *in vitro* but does not stimulate cleavage (Axtell et al. 2006). In addition, siRNAs in phase with this terminus are abundant in *A. thaliana* (Axtell et al. 2006; Rajagopalan et al. 2006), and genes with

two target sites for miRNAs or tasiRNAs showed an enhanced propensity to spawn small RNAs from the region bounded by the dual sites (Axtell et al. 2006). Thus, it appears that two small RNA complementary sites, often but not necessarily triggering two cleavage events, predisposes a cleavage product towards entry into the siRNA-generating pathways as an RdRP substrate. Mechanistically, this explains both tasiRNA precursor processing and the ability of abundant transcripts from exogenous sequences such as transgenes or viruses to stimulate an siRNA cascade. Such highly-expressed transcripts would be more likely to incur random nonspecific cleavages (Axtell et al. 2006). The model is supported by the fact that DCL4, which catalyzes processing of tasiRNA precursors, is also the Dicer primarily responsible for production of viral siRNAs (Bouche et al. 2006; Deleris et al. 2006).

Natural antisense siRNAs

The natural antisense RNAs (nat-siRNAs) provide a pathway by which siRNAs can mediate PTGS in response to fluctuating environmental conditions to regulate one of a pair of *cis*-antisense, overlapping genes (Borsani et al. 2005; Katiyar-Agarwal et al. 2006). Both biotic and abiotic stresses seem capable of inducing nat-siRNA production. For example, under high salt stress, *SRO5* messages are upregulated and presumably pair with transcripts from the overlapping *P5CDH* gene, generating a substrate for the production of a 24-nt DCL2/RDR6/SGS3/NRPD1a-dependent siRNA. This siRNA then specifies cleavage of *P5CDH* messages, establishing a terminus for RdRP synthesis of the antisense strand, and promoting amplification of *P5CDH* silencing via DCL1-catalyzed production of secondary siRNAs (Borsani et al. 2005). In the only other known example of nat-siRNA mediated gene regulation, the inducer is a bacterial pathogen, the *Ps pathovar tomato* carrying the avirulence (*avr*) gene *avrRpt2* (*Pst* (*avrRpt2*)). Upon introduction of *Pst* (*avrRpt2*), the *ATGB2* gene is induced and a 22-nt nat-siRNA accumulates specifically from the region antisense to the overlapping *PPRL* gene, via a DCL1-HEN1-RDR6-SGS3-NRPD1a pathway. Presumably, this siRNA downregulates *PPRL* as part of a disease resistance response because *PPRL* is a negative regulator of the endogenous host resistance signaling pathway (Katiyar-Agarwal et al. 2006).

siRNAs and chromatin silencing

Epigenetic silencing constitutes an important regulatory mechanism in eukaryotes, maintaining genomic stability by transcriptionally silencing elements of the genome whose untimed expression (and, in the case of transposons, recombination) could interfere with normal developmental processes. Chromatin exists in two states, a compact and transcriptionally inert, repeat-rich form, and a transcriptionally active gene-rich form. That RNA can direct chromatin-level modifications was first noted when genomic integrants of RNA viroids were seen to undergo *de novo* methylation at cytosine nucleotides after viroid replication (Wassenegger et al. 1994). This cytosine methylation could occur in all three sequence contexts: CpG, symmetric or CpNpG (where N is A, T, C, or G), and asymmetric or CpHpH (where H is A, T or C) (Pelissier et al. 1999). The phenomenon was called RNA-dependent DNA methylation (RdDM). Transcriptional gene silencing (TGS) is often accompanied by epigenetic modifications to the DNA or histones, and accordingly it was found that transcriptional silencing could result from methylation activity directed at promoter sequences and was associated with small RNAs, suggesting a central role for siRNAs in TGS and RdDM (Mette et al. 2000).

In *Arabidopsis*, METHYLASE1 (MET1) is responsible for CG maintenance methylation, while *de novo* methylation at non-CpG sites is catalyzed by the *DOMAINS REARRANGED METHYLASE1* methyltransferase and CpNpG maintenance methylation is carried out by *CHROMOMETHYLASE3* (*CMT3*). A significant proportion of CMT3 activity depends on the histone H3 lysine-9 methyltransferase SUVH4/KRYTONITE. The DRM proteins catalyze *de novo* cytosine methylation in all sequence contexts (Cao and Jacobsen 2002; Cao et al. 2003). Non-CpG methylation is much more prevalent in plants than in animals, and is particularly linked to transposon loci via the action of *CMT3* (Lindroth et al. 2001; Tompa et al. 2002). Methylation in promoters is quite rare compared to open reading frames (ORFs) (5% and 33% respectively in genome-wide studies) and promoter methylation correlates with low and tissue-specific expression, whereas ORF body methylation correlates with high expression (Zhang et al. 2006b; Zilberman et al. 2007).

The connection between small RNAs and epigenetics was strengthened when SDE4 and AGO4 were shown to be necessary for siRNA accumulation and DNA

cytosine methylation at the *AtSN1* retrotransposon locus (Hamilton et al. 2002; Zilberman et al. 2003). In addition, *ago4* mutants showed a concomitant reduction of siRNAs and histone H3K9 methylation at this locus (Zilberman et al. 2003). Some AGO4-guided chromatin modifications depend to some extent on intact "slicer" activity of the protein (Qi et al. 2006), suggesting a role for siRNA guided cleavage. More conclusive evidence for the link between small RNAs and chromatin modifications came with the observation that *AtSN1* small RNAs, H3K9 methylation, and DNA methylation were all reduced in *rdr2-1* and *dcl3-1* mutants (Xie et al. 2004). Other studies have linked siRNA production to genome-level methylation (Lippman et al. 2004; Lu et al. 2005; Zhang et al. 2006b; Zilberman et al. 2007). Another ARGONAUTE protein, AGO6, appears to function redundantly with AGO4 in siRNA accumulation and DNA methylation in all sequence contexts at endogenous repetitive loci such as *SIMPLEHAT* and *AtREP2* (Zheng et al. 2007). However, whether the sequence-specificity conferred by siRNAs involves RNA-RNA or DNA-RNA hybrid intermediates remains a mystery, though the prevailing model in *S. pombe* proposes that siRNAs mediate the recognition of nascent transcripts (Buhler et al. 2006).

The production and action of some siRNAs also appears to be associated with a plant-specific RNA polymerase, called RNA Polymerase IV (RNAPIV) (Fig. I.4). Two forms of RNAPIV holoenzymes exist in plants, differing in the identity of their largest subunit (Nuclear RNA Polymerase D1) and distinguished as RNAPIVa (NRPD1a NRPD2a) and RNAPIVb (NRPD1b NRPD2a). *NRPD1a* corresponds to the same locus as *SDE4* (Herr et al. 2005), and the NRPD2b gene is a pseudogene (Pontier et al. 2005). In *nrpd1a*, *nrpd1b* and *nrpd2a* mutants siRNA accumulation is compromised at *AtSN1*, 5S ribosomal DNA (rDNA) repeats and the siRNA-generating locus 1003, all of which are hypomethylated in all sequence contexts (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). It is thought that RNAPIV involvement in epigenetic events is engaged at facultative heterochromatin (comprising chromatin that cycles between decondensed and condensed forms, such as transposon and 5S sequences) rather than constitutive heterochromatin (because methylation of the constitutively heterochromatic centromeric signature 180-bp repeat array is largely unaffected in *nrpd* mutants) (Onodera et al. 2005). Notably, methylation is most noticeably lost in the CpNpN

context, the only context for which there is no maintenance methyltransferase (Onodera et al. 2005).

RNAP IVa and RNAP IVb cannot act redundantly to compensate for deficiencies in the other, implying that *NRPD1a* and *NRPD1b* homologs are probably functionally distinct. Consistent with this hypothesis, siRNA generation at some loci is RNAP IVb-independent while others require both RNAP IVa and RNAP IVb (Pontier et al. 2005). Furthermore, AGO4 co-localizes with NRPD1b, RDR2, DCL3 and siRNAs in nuclear Cajal bodies, sites of ribonucleoprotein maturation, while RNAP IVa is external to the nucleolus and localizes to repeat loci (Li et al. 2006; Pontes et al. 2006). NRPD1a is thought to act upstream of siRNA production, and may be involved in transcribing loci at which chromatin silencing machinery has already deposited histone and DNA methylation and other epigenetic marks (Kanno et al. 2005; Pontier et al. 2005) (Fig. I.4). This is consistent with the cytological data described above. Pontes *et al.* (2006) suggest that RNAP IVa is the polymerase responsible for generating transcripts using DNA-associated, methylation-induced aberrant RNA templates synthesized at heterochromatic loci by RNAPI, RNAPII, or RNAPIII. These transcripts translocate by an unknown mechanism to the nucleolus for RDR2-DCL3 siRNA production, and the siRNAs then associate with AGO4 and RNAP IVb to guide DNA methylation and heterochromatic modifications at repetitive loci (Pontes et al. 2006). NRPD1b may be responsible for amplification of the silencing signal, potentially by using siRNAs from primary transcripts as primers for secondary siRNA production. Other epigenetic components tied to the heterochromatic siRNA pathway include the SNF2-like chromatin remodeling protein DRD1, which plays a role in RNA-induced non-CpG methylation (Kanno et al. 2004; Chan et al. 2006).

What is the nature of the small RNA signals which recruit epigenetic silencing machinery to relelvant loci? Cloning and sequencing of small RNA libraries indicate that the overwhelming majority of small RNAs in plants corresponds to intergenic or repetitive loci, particularly in the pericentromeres (Xie et al. 2004; Lu et al. 2006; Rajagopalan et al. 2006; Kasschau et al. 2007). As described above, current data implicates at least some of these siRNAs in the suppression of endogenous transposable elements and heterochromatin assembly at transcriptionally silent loci. At the sequence

level, these small RNAs are distinguished from the miRNAs and tasiRNAs by their longer length (~22–24 nt, with a peak at 24 nt) and their preference for a 5′ adenine nucleotide (Hamilton et al. 2002; Xie et al. 2004; Rajagopalan et al. 2006). The genetic requirements for the formation of such siRNAs appears to vary from locus to locus, but the siRNAs are proposed to direct transcriptional gene silencing, because they are consistently associated with heterochromatinization, including methylation of lysines in histones, and DNA cytosine methylation, as described above. DCL3, RDR2 and AGO4 are all associated with siRNAs at heterochromatic or transposon loci, but AGO4 and PolIVb exhibit more locus-specific effects in terms of methylation and siRNA accumulation (Zilberman et al. 2003; Xie et al. 2004; Pontier et al. 2005; Pontes et al. 2006). At least a subset of endogenous small RNAs are likely to nucleate epigenetic events at heterochromatin-fated loci, but the mechanisms and sequence of events remains to be clarified.

miRNA-directed chromatin methylation

In plants, post-transcriptional gene silencing (PTGS) is effected by miRNAs and some siRNAs, while transcriptional gene silencing (TGS) appears to be mediated almost entirely through siRNAs. However, evidence suggests that an intact cleavage-competent miR165/166 complementary site in the *PHABULOSA* and *PHAVOLUTA* genes are necessary for DNA cytosine methylation, which occurs in *cis* downstream of the exon-spanning complementary site at the 3′ end of the PHB gene (Bao et al. 2004). It is possible that miRNA-directed cleavage of the nascent, spliced *PHB* transcript directs DNA methylation of the template allele via local production of secondary siRNAs, however it is unclear if these methylation markers affect PHB transcription. A couple of small RNAs have been observed downstream and upstream of the complementary site in the PHB gene, though the few small RNAs detected at PHB all lie upstream of the complementary site (Rajagopalan et al. 2006; Kasschau et al. 2007).

Genetic redundancy in small RNA pathways

The number of predicted or empirically validated Dicer proteins in any one genome ranges widely, and in organisms such as plants with more extensive siRNA

activities the Dicer protein repertoire appears to have expanded. For example, among the fully sequenced plant genomes, there are four DICER-LIKE (DCL) proteins in *Arabidopsis thaliana*, five predicted homologs in *Populus trichocarpa* (black cottonwood), and six predicted homologs in *Oryza sativa* spp *japonica* (rice) (Margis et al. 2006). However, in metazoans the Dicer family is typically much smaller; there are two Dicer proteins in *Drosophila melanogaster* (fruit fly), and one each in *Homo sapiens* and *Caenorhabditis elegans* (worm).

While members of the DCL family in *Arabidopsis* are apparently specialized for certain cleavage functions within the different small RNA pathways, some redundancy of roles has been retained. DCL1 is primarily responsible for miRNA biogenesis, DCL2 for natural antisense siRNA (nat-siRNA) biogenesis, DCL3 for the biogenesis of the ~24-nt heterochromatic siRNAs, and DCL4 for the ~21 nt trans-acting siRNAs (tasiRNAs). Viral siRNAs are processed primarily by DCL4 and DCL2 (Bouche et al. 2006; Deleris et al. 2006). Unsurprisingly, DCL4 is also the endonuclease responsible for siRNA formation at transgene loci, again likely acting downstream of RDR6. Despite these specialized functions, a large degree of functional overlap exists among the four DCL proteins in *A. thaliana* (Gasciolli et al. 2005). In double and triple mutant backgrounds, the catalytic deficiencies of some hypomorphic DCL variants can be partially overcome by wild-type DCL family members (Gasciolli et al. 2005). For example, each *TAS* locus seems to depend to a varying extent on DCL4 and DRB4 for processing of their respective transcripts (Gasciolli et al. 2005), a phenomenon that points to DCL redundancy and mirrors that of varying dependencies for different miRNAs on DCL1 and AGO1 (Vaucheret et al. 2004; Vaucheret et al. 2006). It appears that *DCL1* can compensate for hypomorphic *dcl4* in the tasiRNA pathway in the production of ~21-nt siRNAs, because in *dcl2 dcl3 dcl4* plants, *TAS1*, *TAS2*, and *TAS3* tasiRNAs accumulate to apparently normal levels (Bouche et al. 2006). In addition, though DCL1 processes most miRNA precursors, at least some recently evolved miRNA hairpins are processed only by DCL4 (Rajagopalan et al. 2006). The preservation of functional redundancy of DCL proteins in *Arabidopsis* highlights the importance of the small RNA pathways in the plant's lifecycle and may confer adaptive advantages that are more important for survival under conditions of environmental stress which unevenly compromise the activities of

DCL proteins. Perhaps the levels of DCL and DRB proteins differ among cell types, and consequently the accumulation of small RNAs in each is subject to the relative activity of various combinations of DCL and DRB proteins.

Interestingly, DCL2 may act antagonistically to inhibit miRNA and siRNA production by DCL1, according to genetic data (Gasciolli et al. 2005; Bouche et al. 2006). In *dcl2* mutants, miRNAs overaccumulate. This antagonistic function is more pronounced when both *DCL1* and *DCL4* are compromised, because while *dcl1 dcl4* and *dcl1 dcl3 dcl4* mutants (hypomorphic for *DCL1*) display severe developmental phenotypes, quadruple mutants with compromised *DCL2* are viable (Bouche et al. 2006).

Unlike mutations affecting proteins involved in miRNA processing or function, severe developmental defects are not immediately incurred by plants with compromised siRNA biogenetic machinery. For example, *dcl3-1* mutants (presumably null) have a wild-type vegetative phenotype even after five generations (Xie et al. 2004), and *dcl4-1* mutants (also presumably null), compromised in tasiRNA production, display only mild morphological phenotypes (Gasciolli et al. 2005). After three generations, stochastic phenotypes become visible in *dcl2 dcl3* and *dcl3 dcl4* mutants but not in *dcl2 dcl4* or *dcl2 dcl3 dcl4* mutants (Gasciolli et al. 2005). This absence of strong deleterious outcomes is mirrored by *cmt3* and *drm1 drm2* homozygotes, neither of which display abnormalities even after five generations of inbreeding (Lindroth et al. 2001; Cao and Jacobsen 2002). The generational and stochastic nature of severe phenotypes is consistent with genetic redundancy combined with a progressive loss of epigenetic modifications at pernicious loci that become reactivated and lead to further genomic instability. Indeed, phenotypes in some of these mutants may be unmasked as a result of hypersusceptibility to viral infections or debilitated responses to other environmental stresses, as has been observed in *rdr6* and *dcl2* mutants (Dalmay et al. 2000; Xie et al. 2004).

PTGS and pathogen defense

Before endogenous siRNAs were discovered in plants, small RNAs were known to derive from exogenous elements such as viral agents or highly-expressed transgenes introduced into plant genomes. Thus PTGS was long considered to be a *cis*-acting RNA

immune system that was activated in plants by the invasion of foreign genetic elements (Vaucheret 2006).

Both viral-induced gene silencing (VIGS) and transgene-induced gene silencing (TIGS) pathways share components with other PTGS mechanisms. The silencing of endogenous genes is mediated at low frequency by homologous sense transgenes (S-PTGS), and at higher frequency by inverted-repeat transgenes (IR-PTGS) (Beclin et al. 2002). IR transgene transcripts can presumably fold back on themselves to form stable dsRNA for the production of small RNAs, while S-PTGS requires additional factors to convert primary transcripts into substrates for small RNA generation, namely RDR6 and SGS3, proteins which are also involved in tasiRNA biogenesis.

VIGS is mediated by DCL4 and DCL2 (Bouche et al. 2006) which process tasiRNA precursors and nat-siRNAs respectively. PTGS pathways constitute a battleground for the arms races that characterize most pathogen-host co-evolution; evidently, some fraction of adaptation in viral and plant genomes is concentrated on disabling or buttressing, respectively, the protective roles of silencing pathways. Differential susceptibility likely stems in part from the biochemical armor that viruses use to evade detection; many plant viruses express proteins that interfere with or otherwise suppress different steps and components along the PTGS pathways (as reviewed in (Vance and Vaucheret 2001; Dunoyer and Voinnet 2005)) and the miRNA pathway (Mallory et al. 2002; Kasschau et al. 2003; Chapman et al. 2004; Dunoyer et al. 2004; Lakatos et al. 2006). The PTGS pathways comprise a plant defense system that viruses may have evolved to counteract in order to ensure efficient and systemic infection. Rationalizing viral interference with miRNA pathway components is more difficult, though it may result from the overlap in machinery between the pathways, or because the upregulation of some miRNA targets establishes conditions conducive to viral proliferation, or as a response to the antiviral action of as-yet undiscovered miRNAs which may direct targeting of viral sequences or activate other defense mechanisms in the plant by downregulating repressors of these mechanisms (Vaucheret 2006).

Endogenous silencing RNAs in metazoan species

Distinct varieties of small RNAs presumed to act in silencing have been discovered in yeast, flies, and mammals. In the fission yeast *Saccharomyces pombe*, siRNAs derived from centromeric repeat sequences (Reinhart and Bartel 2002) are associated with a complex called RITS (RNA-induced initiation of transcriptional gene silencing) and target heterochromatin assembly to these domains via TGS and PTGS pathways (Verdel et al. 2004; Buhler et al. 2006). The Slicer activity of Ago1 in *S. pombe* is necessary for RITS localization to heterochromatic loci and spreading of H3K9 methylation, suggesting that transcript cleavage accompanies or precedes transcriptional silencing (Irvine et al. 2006).

Characterized metazoan silencing RNAs appear to associate with a subfamily of Argonaute proteins, known as Piwi proteins, which possess both PAZ and PIWI domains. Some of these small RNAs seem to function in maintaining genome stability by silencing foreign DNA such as transposable elements. In *D. melanogaster*, 24-29 nucleotide repeat-associated siRNAs (rasiRNAs) matching transposable elements and satellite DNA have been isolated from testes and developing embryos (Aravin et al. 2003) which partition into predominantly antisense (Vagin et al. 2006) and predominantly sense populations by associating with different Piwi silencing proteins (Brennecke et al. 2007; Gunawardane et al. 2007). In mammals, the 26–31-nt Piwi-interacting RNAs (piRNAs) accumulate strand-specifically from predominantly non-genic and non-repetitive genomic clusters in the male germline during spermatogenesis (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006), although a subset appears to derive earlier than the rest, from developmentally regulated transposon loci, suggesting a role in transposon suppression (Aravin et al. 2007). The piRNA class has also been identified in zebrafish, with some notable differences; they occur in ovaries as well as testes, they derive from both strands, and they are more biased than mammalian piRNAs to originate from repetitive loci (Houwing et al. 2007). In these respects, zebrafish piRNAs resemble *Drosophila* rasiRNAs (a subclass of fly piRNAs) more than mammalian piRNAs. The biogenesis of mammalian piRNAs is unclear, but in zebrafish they may be generated without the involvement of Dicer (Houwing et al. 2007). *Drosophila* DmHEN1, the homolog of HEN1, methylates the 2′ hydroxyl at the 3′ end of rasiRNAs (Saito et al. 2007), a process

that occurs after Ago2-RISC assembly, RISC loading of siRNA duplex, and passenger strand cleavage (Horwich et al. 2007; Saito et al. 2007). The stark similarities between rasiRNA activity in *Drosophila* and the activity of some piRNAs in vertebrates suggest that they may together represent the extant versions of an ancient RNA-based germline defense system (O'Donnell and Boeke 2007).

Summary of Thesis

The initial characterizations of small RNAs in *A. thaliana* suggested that most plant miRNAs could be grouped into multigenic families, and regulated key developmental processes within regulatory circuits conserved, in many cases, throughout the angiosperms, and in some cases predating the origins of vaculature and seeds in the plant kingdom. 97 confidently annotated miRNA genes had been identified in *A. thaliana*, comprising 26 families, 21 of which had recognizeable orthologs in the two other sequenced plant genomes, *Oryza sativa* (rice) and *Populus trichocarpa* (black cottonwood). Furthermore, while the list of animal miRNAs was still expanding, it appeared that the vast majority of plant miRNAs had been discovered, a hypothesis that could not be dismissed by my own early small RNA cloning efforts, which was limited in scale and yielded candidates with only very weak evidentiary support. The confident identification of non-conserved miRNAs remained beyond the reach of the traditional approaches of cloning and Sanger sequencing, and bioinformatic prediction, leaving open the question of the extent of non-conserved miRNAs in *Arabidopsis* that might exist in addition to the five highly-expressed and previously identified lineage-specific miRNAs.

However, with the advent of deep sequencing technologies, it has become possible to more carefully probe small RNA libraries for previously unreported miRNAs and siRNAs, and to develop more sensitive criteria for identifying and evaluating apparently non-conserved miRNA candidates. Employing such an approach, this work dispels the notion that the activity of plant miRNAs is predominantly circumscribed in ancient developmental programs, suggesting new possibilities for their function in lineage-specific pathways of pathogen defense, chromatin and epigenetic modifications, and transcriptional control. In Chapter 1, I describe the discovery and existence of at least 37 additional non-conserved miRNAs in *Arabidopsis* and at least one additional

conserved family, as well as some new insights gleaned from high-throughput sequencing about miRNA biogenesis, activity, and the possible evolutionary mechanisms underlying miRNA gene emergence. These findings uncovered a previously unknown layer of regulatory control effected by miRNAs with apparently lineage-specific roles, whose expression in wild-type plants under standard growth conditions is relatively low. One could therefore postulate that the relative ease of evolving a miRNA locus has been used to great advantage during the processes of natural selection, affording a means by which the space of possibly beneficial post-transcriptional regulatory interactions can be explored, with minimal fitness consequences for the plant. Thus, this work expands the scope of miRNA biology in plants.

In Chapter 2, I elaborate on the various populations of siRNAs in *A. thaliana* as sampled by deep sequencing, and the regions of the genome that give rise to them, including at least one newly discovered *trans*-acting siRNA locus and several thousands of other small RNA-spawning loci in the plant genome. This data prompted revisitation of the notion that repetitive and mobile element sequences are disproportionately high contributors to the endogenous small silencing RNA pool, and demonstrated that in fact, while such loci were slightly overrepresented by small RNA species, the vast majority of small RNAs described mapped to regions that have yet to be annotated. Further biological characterization of these small RNAs is necessary to determine their functional roles in gene silencing.

All *Arabidopsis thaliana* small RNAs described in this work can be browsed on the genome assembly at the online database of the Arabidopsis Small RNA Project at http://asrp.cgrb.oregonstate.edu/cgi-bin/gbrowse/thaliana-v5

Table I.1. Distinguishing features of *Arabidopsis* miRNAs and siRNAs

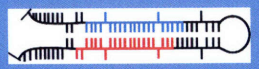
		miRNAs	siRNAs
Length		• ~21 nt	• 21-24 nt
Precursor 2° structure			
Processing/silencing machinery	Second strand synthesis (RNA DEPENDENT RNA POLYMERASES)		• RDR2, RDR6
	Processing (DICER-LIKE proteins)	• DCL1	• DCL2, DCL3, DCL4
	Silencing (ARGONAUTE proteins)	• AGO1	• AGO4, AGO7
Duplexes per precursor		• typically 1	• typically >1
Regulatory mechanism		• PTGS	• PTGS • TGS
Regulatory mode of action		• target in *trans*	• target in *cis* or in *trans*

FIGURE LEGENDS

Figure I.1. Prevailing model for miRNA biogenesis and target recognition in *Arabidopsis*.

MIRNA loci in plants are transcribed by RNA Pol II generating transcripts that adopt a stemloop secondary structure (1). This stemloop undergoes processing by DCL-HYL1 or DCL4-DRB4 and 3′ methylation by HEN1 (2). The liberated duplex may be unwound in the nucleus or in the cytoplasm; regardless, nuclear exit is mediated by HASTY or other export proteins (3). The miRNA strand, which is often less stably paired at its 5′ end, becomes selectively incorporated into and stabilized in the AGO1-containing RISC complex while the miRNA* strand is excluded. The miRNA-programmed RISC recognizes and binds complementary mRNAs at the miRNA target site (4). AGO1 catalyzes cleavage of the target message guided by miRNA-mRNA pairing, leaving a 5′ monophosphate. Cleaved fragments are released (5). AGO1-RISC with bound miRNA is free to recognize another complementary message and catalyze its cleavage.

Figure I.2. A cloning protocol that enriches for DCL products.

Total RNA is isolated and 18–26-nt species are gel-purified using labeled size markers (1). A 3′ pre-adenylated oligo with a blocked 3′ end is directionally ligated to small RNA species in the absence of ATP and ligated small RNAs are gel-purified as band-shifted species (2). Ligation of the 5′ adaptor occurs in the presence of ATP and doubly-ligated species are gel-purified (3). Reverse transcription is followed by a primary PCR amplification generating a small RNA cDNA library (4). For dideoxy sequencing, PCR products are restriction digested with BanI at a site engineered into the adaptor sequences, concatamerized, and cloned. Individual clones with inserts are sequenced (5). For 454 pyrophosphate-based sequencing, a secondary large-scale PCR step is followed by native gel purification to retain small cDNAs in double-stranded form for submission to the company (6). This protocol is based on that published in Lau et al. 2001.

Figure I.3. Prevailing model for tasiRNA biogenesis and target recognition in *Arabidopsis*.

RNA Pol II synthesizes primary transcripts from one of the five *TAS* loci which are dispersed in the genome (1). The primary transcript is targeted for cleavage at a site 5′ of the tasiRNA-spawning region by miR173 (*TAS1a*, *TAS1b*, *TAS1c*, *TAS2*) or at a site 3′ of the tasiRNA-spawning region by miR390 (*TAS3*) (2). This establishes transcripts with a precisely defined terminus, presumably serving as a start or stop signal for RDR6-mediated synthesis of the antisense strand, in which SGS3 is also thought to be involved (3). Upon formation of double-stranded RNA, DCL4-DRB4 succesively cleaves siRNA duplexes in 21 nucleotide (nt) registers, leaving a 2-nt 3′ overhang. These duplexes are methylated by HEN1 (4). RISC complexes (presumably containing AGO1 or AGO7) are loaded with one of the siRNA strands from a duplex, and are then competent to recognize and bind target messages (5). Target messages are cleaved at sites complementary to the RISC-bound tasiRNA, leaving a 5′ phosphate on the 3′ cleavage product (6).

Figure I.4. A model for heterochromatic siRNA biogenesis in *Arabidopsis*.

Heterochromatin-fated loci are transcribed, but it is unclear which polymerase gives rise to primary transcripts (1). If these transcripts lack inherent self-pairing character or do not co-localize with convergent transcripts, then they become substrates for RDR2-dependent synthesis of the complementary strand, generating double-stranded RNA (dsRNA) (2). The dsRNA precursor is presumably processed by DCL3, and HEN1 methylates at least some heterochromatic siRNAs (4). PolIVa (NRPD1a-NRPD2a) is thought to be upstream of siRNA production (3). PolIVb (NRPD1b-NRPD2a) and the chromatin remodeling protein DRD1 are thought to be downstream of siRNA production (5). AGO4, DRM1 and DRM2 target DNA cytosine methylation marks at the locus of origin or loci highly homologous to siRNAs, in a process that is associated with siRNA production (6).

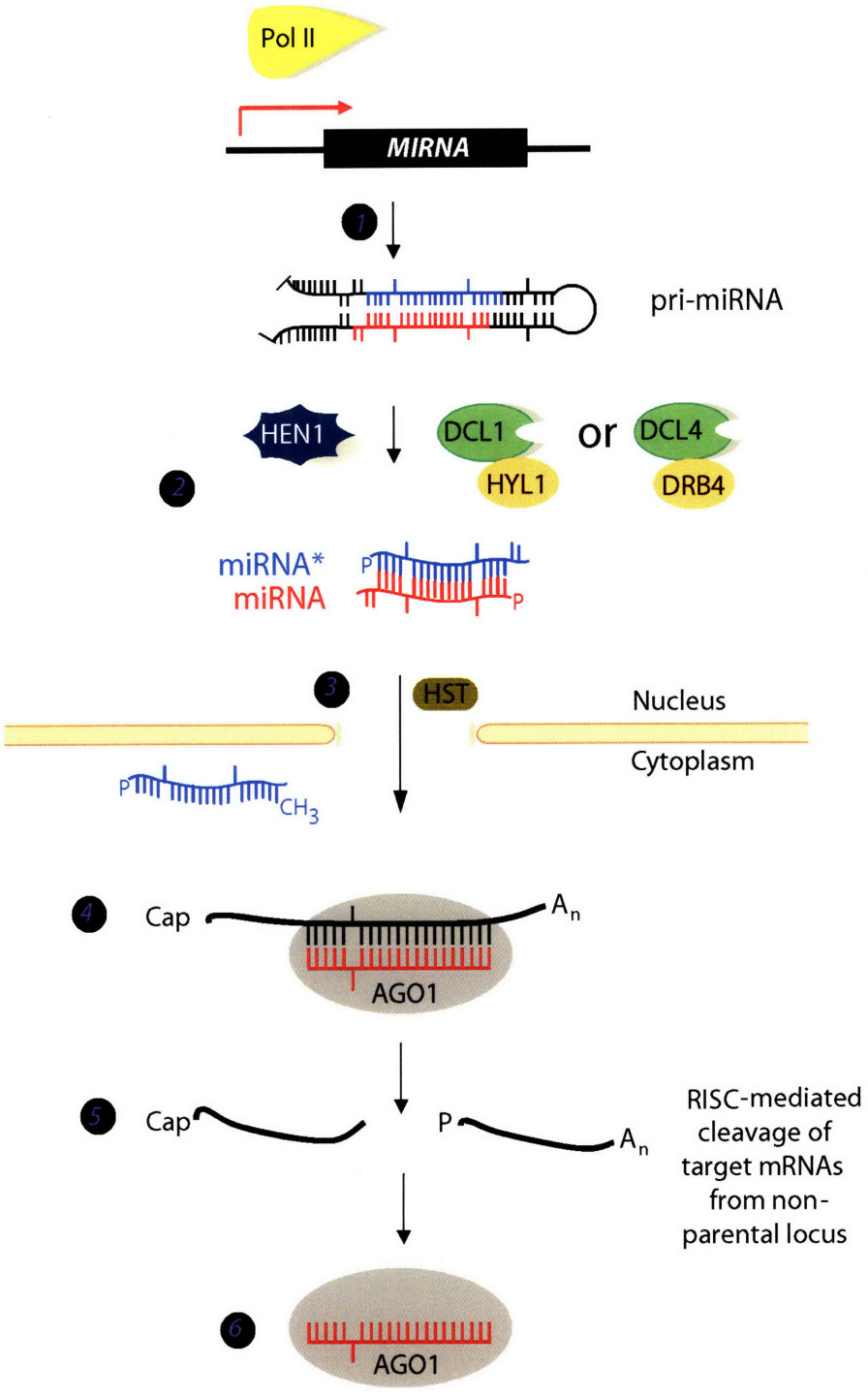


Figure I.1. Prevailing model for miRNA biogenesis and target recognition in *Arabidopsis.*

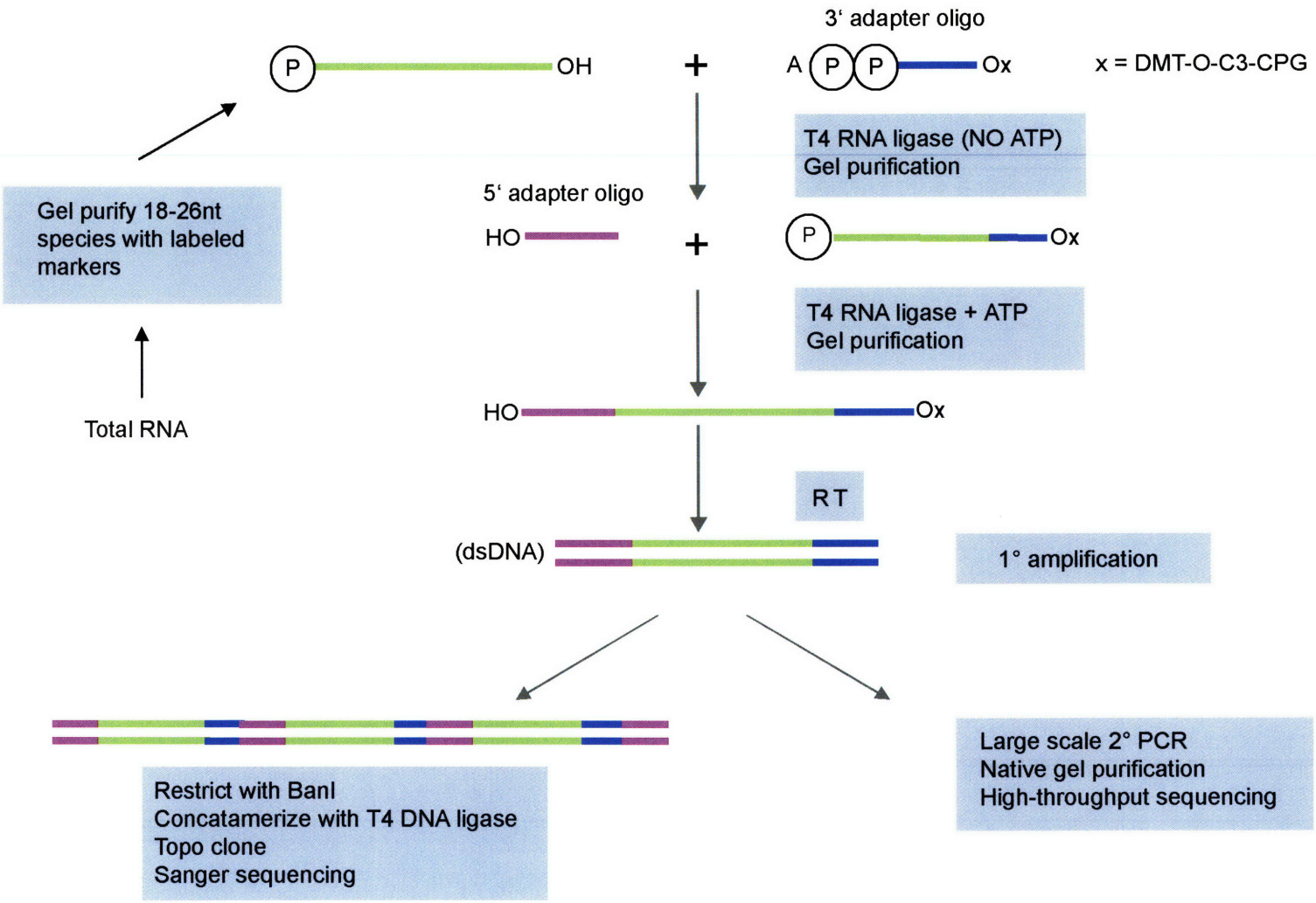


Figure I.2. A cloning protocol that enriches for DCL products.

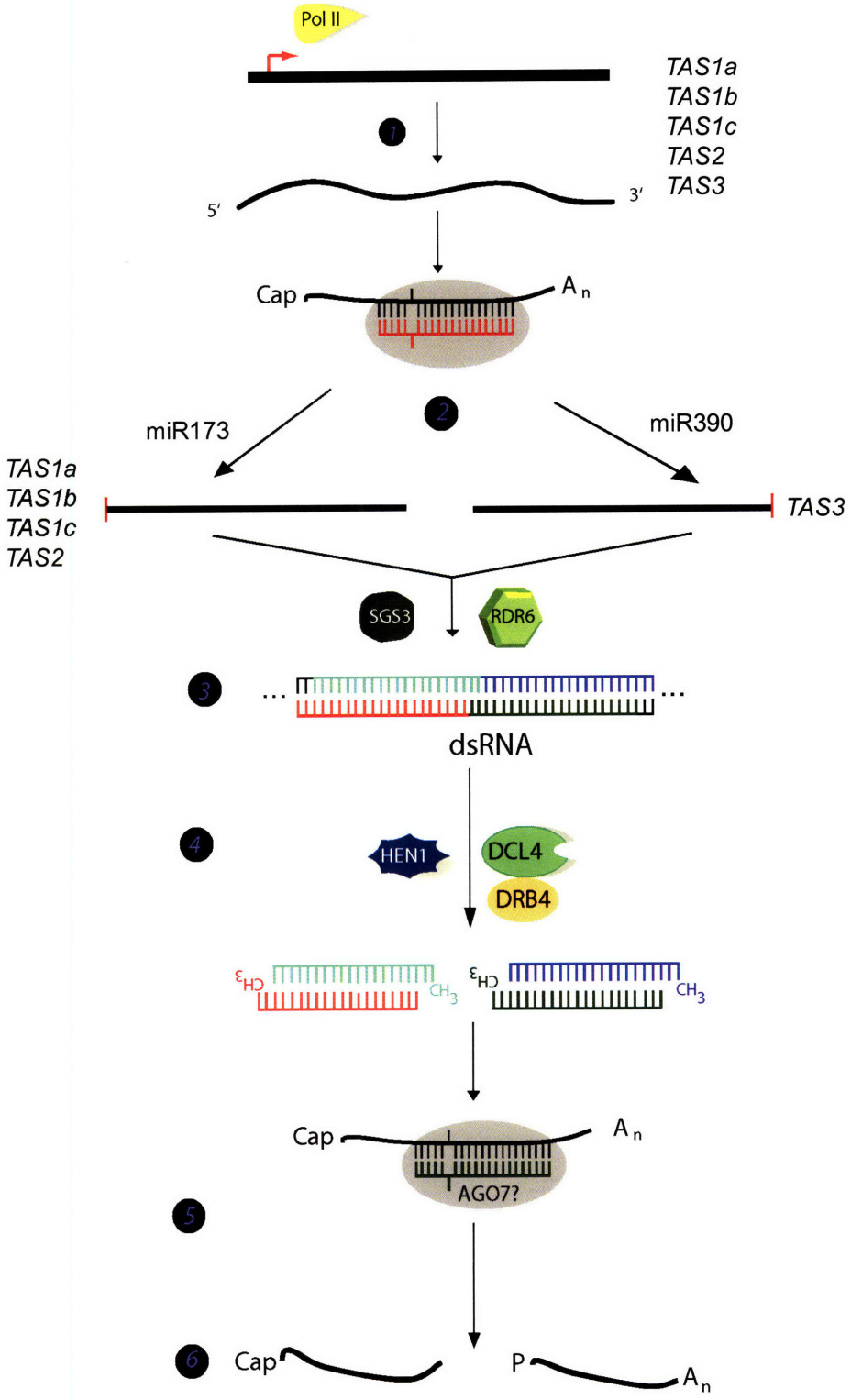


Figure I.3. Prevailing model for tasiRNA biogenesis and target recognition in *Arabidopsis*.

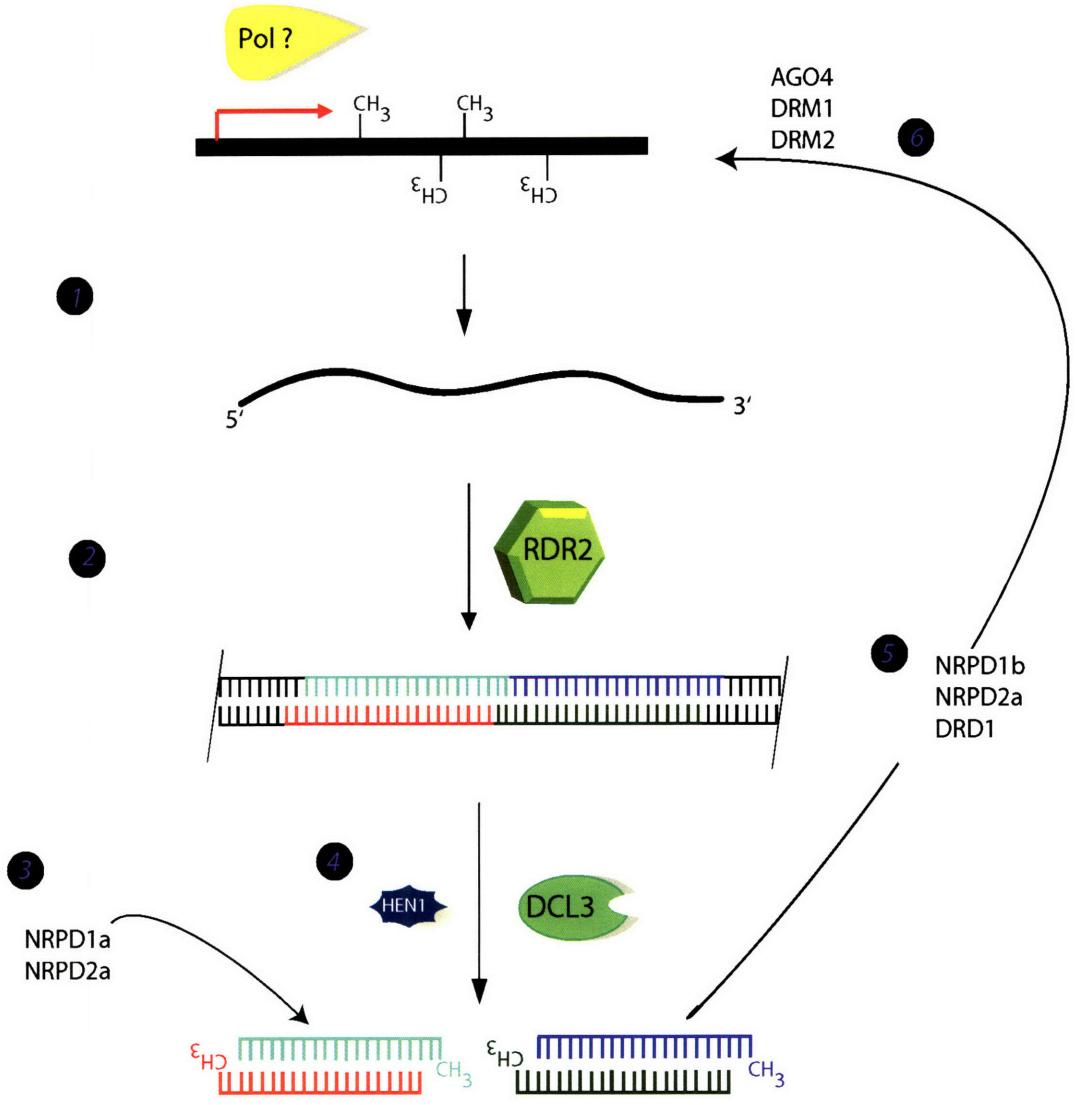


Figure I.4. A model for heterochromatic siRNA biogenesis in *Arabidopsis.*

REFERENCES

Achard, P., A. Herr, D.C. Baulcombe, and N.P. Harberd. 2004. Modulation of floral development by a gibberellin-regulated microRNA. *Development* **131**: 3357-3365.

Adenot, X., T. Elmayan, D. Lauressergues, S. Boutet, N. Bouche, V. Gasciolli, and H. Vaucheret. 2006. DRB4-dependent *TAS3 trans*-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16**: 927-932.

Allen, E., Z. Xie, A.M. Gustafson, and J.C. Carrington. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121**: 207-221.

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

Ambros, V. 1989. A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*. *Cell* **57**: 49-57.

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

Ambros, V., B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, and T. Tuschl. 2003a. A uniform system for microRNA annotation. *Rna* **9**: 277-279.

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

Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M.J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203-207.

Aravin, A.A., M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, T. Gaasterland, J. Meyer, and T. Tuschl. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* **5**: 337-350.

Aravin, A.A., R. Sachidanandam, A. Girard, K. Fejes-Toth, and G.J. Hannon. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744-747.

Arazi, T., M. Talmor-Neiman, R. Stav, M. Riese, P. Huijser, and D.C. Baulcombe. 2005. Cloning and characterization of micro-RNAs from moss. *Plant J* **43**: 837-848.

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

Axtell, M., C. Jan, R. Rajagopalan, and D.P. Bartel. 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**: 565-577.

Axtell, M.J. and D.P. Bartel. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell.*

Axtell, M.J., J.A. Snyder, and D.P. Bartel. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell.*

Bagga, S., J. Bracht, S. Hunter, K. Massirer, J. Holtz, R. Eachus, and A.E. Pasquinelli. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**: 553-563.

Baker, C.C., P. Sieber, F. Wellmer, and E.M. Meyerowitz. 2005. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. *Curr Biol* **15**: 303-315.

Bao, N., K.W. Lye, and M.K. Barton. 2004. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev Cell* **7**: 653-662.

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

Bartel, D.P. and C.Z. Chen. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* **5**: 396-400.

Baskerville, S. and D.P. Bartel. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**: 241-247.

Baumberger, N. and D.C. Baulcombe. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**: 11928-11933.

Beclin, C., S. Boutet, P. Waterhouse, and H. Vaucheret. 2002. A branched pathway for transgene-induced RNA silencing in plants. *Curr Biol* **12**: 684-688.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366.

Bohmert, K., I. Camus, C. Bellini, D. Bouchez, M. Caboche, and C. Benning. 1998. AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *Embo J* **17**: 170-180.

Bohnsack, M.T., K. Czaplinski, and D. Gorlich. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**: 185-191.

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

Borevitz, J.O., Y. Xia, J. Blount, R.A. Dixon, and C. Lamb. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**: 2383-2394.

Borsani, O., J. Zhu, P.E. Verslues, R. Sunkar, and J.K. Zhu. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**: 1279-1291.

Bouche, N., D. Lauressergues, V. Gasciolli, and H. Vaucheret. 2006. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *Embo J.*

Boutet, S., F. Vazquez, J. Liu, C. Beclin, M. Fagard, A. Gratias, J.B. Morel, P. Crete, X. Chen, and H. Vaucheret. 2003. *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr Biol* **13**: 843-848.

Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089-1103.

Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**: 25-36.

Brennecke, J., A. Stark, R.B. Russell, and S.M. Cohen. 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**: e85.

Brenner, S., M. Johnson, J. Bridgham, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S.R. Williams, K. Moon, T. Burcham, M. Pallas, R.B. DuBridge, J. Kirchner, K. Fearon, J. Mao, and K. Corcoran. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**: 630-634.

Buhler, M., A. Verdel, and D. Moazed. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**: 873-886.

Calado, A., N. Treichel, E.C. Muller, A. Otto, and U. Kutay. 2002. Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *Embo J* **21**: 6216-6224.

Cao, X., W. Aufsatz, D. Zilberman, M.F. Mette, M.S. Huang, M. Matzke, and S.E. Jacobsen. 2003. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* **13**: 2212-2217.

Cao, X. and S.E. Jacobsen. 2002. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**: 1138-1144.

Catalanotto, C., G. Azzalin, G. Macino, and C. Cogoni. 2002. Involvement of small RNAs and role of the *qde* genes in the gene silencing pathway in *Neurospora*. *Genes Dev* **16**: 790-795.

Chan, S.W., I.R. Henderson, X. Zhang, G. Shah, J.S. Chien, and S.E. Jacobsen. 2006. RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. *PLoS Genet* **2**: e83.

Chapman, E.J., A.I. Prokhnevsky, K. Gopinath, V.V. Dolja, and J.C. Carrington. 2004. Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes Dev* **18**: 1179-1186.

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

Chen, X. 2005. MicroRNA biogenesis and function in plants. *FEBS Lett* **579**: 5923-5931.

Chen, X., J. Liu, Y. Cheng, and D. Jia. 2002. HEN1 functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. *Development* **129**: 1085-1094.

Chendrimada, T.P., K.J. Finn, X. Ji, D. Baillat, R.I. Gregory, S.A. Liebhaber, A.E. Pasquinelli, and R. Shiekhattar. 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447**: 823-828.

Dalmay, T., A. Hamilton, S. Rudd, S. Angell, and D.C. Baulcombe. 2000. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for

posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**: 543-553.

Deleris, A., J. Gallego-Bartolome, J. Bao, K.D. Kasschau, J.C. Carrington, and O. Voinnet. 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**: 68-71.

Denli, A.M., B.B. Tops, R.H. Plasterk, R.F. Ketting, and G.J. Hannon. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-235.

Dharmasiri, N., S. Dharmasiri, and M. Estelle. 2005a. The F-box protein TIR1 is an auxin receptor. *Nature* **435**: 441-445.

Dharmasiri, N., S. Dharmasiri, D. Weijers, E. Lechner, M. Yamada, L. Hobbie, J.S. Ehrismann, G. Jurgens, and M. Estelle. 2005b. Plant development is regulated by a family of auxin receptor F box proteins. *Dev Cell* **9**: 109-119.

Doench, J.G., C.P. Petersen, and P.A. Sharp. 2003. siRNAs can function as miRNAs. *Genes Dev* **17**: 438-442.

Doench, J.G. and P.A. Sharp. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**: 504-511.

Dunoyer, P., C.H. Lecellier, E.A. Parizotto, C. Himber, and O. Voinnet. 2004. Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. *Plant Cell* **16**: 1235-1250.

Dunoyer, P. and O. Voinnet. 2005. The complex interplay between plant viruses and host RNA-silencing pathways. *Curr Opin Plant Biol* **8**: 415-423.

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

Elmayan, T., S. Balzergue, F. Beon, V. Bourdon, J. Daubremet, Y. Guenet, P. Mourrain, J.C. Palauqui, S. Vernhettes, T. Vialle, K. Wostrikoff, and H. Vaucheret. 1998. *Arabidopsis* mutants impaired in cosuppression. *Plant Cell* **10**: 1747-1758.

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

Fahlgren, N., M.D. Howell, K.D. Kasschau, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, T.F. Law, S.R. Grant, J.L. Dangl, and J.C. Carrington. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. *PLoS ONE* **2**: e219.

Fahlgren, N., T.A. Montgomery, M.D. Howell, E. Allen, S.K. Dvorak, A.L. Alexander, and J.C. Carrington. 2006. Regulation of AUXIN RESPONSE FACTOR3 by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr Biol* **16**: 939-944.

Fang, Y. and D.L. Spector. 2007. Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr Biol* **17**: 818-823.

Filipowicz, W. 2005. RNAi: the nuts and bolts of the RISC machine. *Cell* **122**: 17-20.

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

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

Forstemann, K., Y. Tomari, T. Du, V.V. Vagin, A.M. Denli, D.P. Bratu, C. Klattenhoff, W.E. Theurkauf, and P.D. Zamore. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* **3**: e236.

Fujii, H., T.J. Chiou, S.I. Lin, K. Aung, and J.K. Zhu. 2005. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr Biol* **15**: 2038-2043.

Garcia, D., S.A. Collier, M.E. Byrne, and R.A. Martienssen. 2006. Specification of leaf polarity in *Arabidopsis* via the *trans*-acting siRNA pathway. *Curr Biol* **16**: 933-938.

Gasciolli, V., A.C. Mallory, D.P. Bartel, and H. Vaucheret. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr Biol*.

Giraldez, A.J., R.M. Cinalli, M.E. Glasner, A.J. Enright, J.M. Thomson, S. Baskerville, S.M. Hammond, D.P. Bartel, and A.F. Schier. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833-838.

Girard, A., R. Sachidanandam, G.J. Hannon, and M.A. Carmell. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199-202.

Gregory, R.I., K.P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235-240.

Griffiths-Jones, S., R.J. Grocock, S. van Dongen, A. Bateman, and A.J. Enright. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140-144.

Grigg, S.P., C. Canales, A. Hay, and M. Tsiantis. 2005. SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. *Nature* **437**: 1022-1026.

Grimson, A., K.K. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, and D.P. Bartel. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91-105.

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

Gunawardane, L.S., K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M.C. Siomi. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**: 1587-1590.

Gwizdek, C., B. Ossareh-Nazari, A.M. Brownawell, A. Doglio, E. Bertrand, I.G. Macara, and C. Dargemont. 2003. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J Biol Chem* **278**: 5505-5508.

Hamilton, A., O. Voinnet, L. Chappell, and D. Baulcombe. 2002. Two classes of short interfering RNA in RNA silencing. *Embo J* **21**: 4671-4679.

Hamilton, A.J. and D.C. Baulcombe. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950-952.

Han, J., Y. Lee, K.H. Yeom, J.W. Nam, I. Heo, J.K. Rhee, S.Y. Sohn, Y. Cho, B.T. Zhang, and V.N. Kim. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**: 887-901.

Han, M.H., S. Goud, L. Song, and N. Fedoroff. 2004. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* **101**: 1093-1098.

Henderson, I.R., X. Zhang, C. Lu, L. Johnson, B.C. Meyers, P.J. Green, and S.E. Jacobsen. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**: 721-725.

Hernandez-Pinzon, I., N.E. Yelina, F. Schwach, D.J. Studholme, D. Baulcombe, and T. Dalmay. 2007. SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of *trans*-acting endogenous siRNA. *Plant J* **50**: 140-148.

Herr, A.J., M.B. Jensen, T. Dalmay, and D.C. Baulcombe. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118-120.

Hipfner, D.R., K. Weigmann, and S.M. Cohen. 2002. The *bantam* gene regulates Drosophila growth. *Genetics* **161**: 1527-1537.

Hiraguri, A., R. Itoh, N. Kondo, Y. Nomura, D. Aizawa, Y. Murai, H. Koiwa, M. Seki, K. Shinozaki, and T. Fukuhara. 2005. Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Mol Biol* **57**: 173-188.

Horwich, M.D., C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P.D. Zamore. 2007. The *Drosophila* RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Curr Biol.*

Houwing, S., L.M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D.V. Filippov, H. Blaser, E. Raz, C.B. Moens, R.H. Plasterk, G.J. Hannon, B.W. Draper, and R.F. Ketting. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**: 69-82.

Hunter, C., M.R. Willmann, G. Wu, M. Yoshikawa, M. de la Luz Gutierrez-Nava, and S.R. Poethig. 2006. *Trans*-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* **133**: 2973-2981.

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

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

Irvine, D.V., M. Zaratiegui, N.H. Tolia, D.B. Goto, D.H. Chitwood, M.W. Vaughn, L. Joshua-Tor, and R.A. Martienssen. 2006. Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**: 1134-1137.

Jackson, R.J. and N. Standart. 2007. How do microRNAs regulate gene expression? *Sci STKE* **2007**: re1.

Jacobsen, S.E., M.P. Running, and E.M. Meyerowitz. 1999. Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* **126**: 5231-5243.

Johnston, R.J. and O. Hobert. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845-849.

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

Jones-Rhoades, M.W., D.P. Bartel, and B. Bartel. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol.*

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

Kanno, T., B. Huettel, M.F. Mette, W. Aufsatz, E. Jaligot, L. Daxinger, D.P. Kreil, M. Matzke, and A.J. Matzke. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**: 761-765.

Kanno, T., M.F. Mette, D.P. Kreil, W. Aufsatz, M. Matzke, and A.J. Matzke. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol* **14**: 801-805.

Kasschau, K.D., N. Fahlgren, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, and J.C. Carrington. 2007. Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol* **5**: e57.

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

Katiyar-Agarwal, S., R. Morgan, D. Dahlbeck, O. Borsani, A. Villegas, Jr., J.K. Zhu, B.J. Staskawicz, and H. Jin. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* **103**: 18002-18007.

Kepinski, S. and O. Leyser. 2005. The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* **435**: 446-451.

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

Kidner, C.A. and R.A. Martienssen. 2004. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* **428**: 81-84.

Kim, J., J.H. Jung, J.L. Reyes, Y.S. Kim, S.Y. Kim, K.S. Chung, J.A. Kim, M. Lee, Y. Lee, V. Narry Kim, N.H. Chua, and C.M. Park. 2005. microRNA-directed cleavage of *ATHB15* mRNA regulates vascular development in *Arabidopsis* inflorescence stems. *Plant J* **42**: 84-94.

Kim, V.N. and J.W. Nam. 2006. Genomics of microRNA. *Trends Genet* **22**: 165-173.

Kim, Y.K. and V.N. Kim. 2007. Processing of intronic microRNAs. *Embo J* **26**: 775-783.

Krek, A., D. Grun, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, and N. Rajewsky. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**: 495-500.

Kurihara, Y., Y. Takashi, and Y. Watanabe. 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.

Kurihara, Y. and Y. Watanabe. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**: 12753-12758.

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

Lakatos, L., T. Csorba, V. Pantaleo, E.J. Chapman, J.C. Carrington, Y.P. Liu, V.V. Dolja, L.F. Calvino, J.J. Lopez-Moya, and J. Burgyan. 2006. Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *Embo J* **25**: 2768-2780.

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

Lau, N.C., A.G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D.P. Bartel, and R.E. Kingston. 2006. Characterization of the piRNA complex from rat testes. *Science* **313**: 363-367.

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

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

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-854.

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

Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, and V.N. Kim. 2004. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**: 4051-4060.

Lewis, B.P., C.B. Burge, and D.P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.

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

Li, C.F., O. Pontes, M. El-Shami, I.R. Henderson, Y.V. Bernatavichute, S.W. Chan, T. Lagrange, C.S. Pikaard, and S.E. Jacobsen. 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**: 93-106.

Li, J., Z. Yang, B. Yu, J. Liu, and X. Chen. 2005. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr Biol* **15**: 1501-1507.

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

Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769-773.

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

Lindbo, J.A. and W.G. Dougherty. 2004. Plant pathology and RNAi: A brief history. *Annu Rev Phytopathol.*

Lindroth, A.M., X. Cao, J.P. Jackson, D. Zilberman, C.M. McCallum, S. Henikoff, and S.E. Jacobsen. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**: 2077-2080.

Lingel, A., B. Simon, E. Izaurralde, and M. Sattler. 2003. Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* **426**: 465-469.

Lingel, A., B. Simon, E. Izaurralde, and M. Sattler. 2004. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**: 576-577.

Lippman, Z., A.V. Gendrel, M. Black, M.W. Vaughn, N. Dedhia, W.R. McCombie, K. Lavine, V. Mittal, B. May, K.D. Kasschau, J.C. Carrington, R.W. Doerge, V. Colot, and R. Martienssen. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471-476.

Liu, J., M.A. Carmell, F.V. Rivas, C.G. Marsden, J.M. Thomson, J.J. Song, S.M. Hammond, L. Joshua-Tor, and G.J. Hannon. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437-1441.

Liu, J., F.V. Rivas, J. Wohlschlegel, J.R. Yates, 3rd, R. Parker, and G.J. Hannon. 2005a. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* **7**: 1261-1266.

Liu, J., M.A. Valencia-Sanchez, G.J. Hannon, and R. Parker. 2005b. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**: 719-723.

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

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

Lobbes, D., G. Rallapalli, D.D. Schmidt, C. Martin, and J. Clarke. 2006. SERRATE: a new player on the plant microRNA scene. *EMBO Rep* **7**: 1052-1058.

Lu, C. and N. Fedoroff. 2000. A mutation in the *Arabidopsis HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* **12**: 2351-2366.

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

Lu, C., S.S. Tej, S. Luo, C.D. Haudenschild, B.C. Meyers, and P.J. Green. 2005. Elucidation of the small RNA component of the transcriptome. *Science* **309**: 1567-1569.

Lund, E., S. Guttinger, A. Calado, J.E. Dahlberg, and U. Kutay. 2004. Nuclear export of microRNA precursors. *Science* **303**: 95-98.

Ma, J.B., K. Ye, and D.J. Patel. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**: 318-322.

Maher, C., L. Stein, and D. Ware. 2006. Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res* **16**: 510-519.

Mallory, A.C., D.P. Bartel, and B. Bartel. 2005. MicroRNA-directed regulation of Arabidopsis *AUXIN RESPONSE FACTOR17* is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* **17**: 1360-1375.

Mallory, A.C., D.V. Dugas, D.P. Bartel, and B. Bartel. 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, A.C., B.J. Reinhart, D. Bartel, V.B. Vance, and L.H. Bowman. 2002. A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. *Proc Natl Acad Sci U S A* **99**: 15228-15233.

Mallory, A.C., B.J. Reinhart, M.W. Jones-Rhoades, G. Tang, P.D. Zamore, M.K. Barton, and D.P. Bartel. 2004b. MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5' region. *Embo J* **23**: 3356-3364.

Margis, R., A.F. Fusaro, N.A. Smith, S.J. Curtin, J.M. Watson, E.J. Finnegan, and P.M. Waterhouse. 2006. The evolution and diversification of Dicers in plants. *FEBS Lett* **580**: 2442-2450.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

Matranga, C., Y. Tomari, C. Shin, D.P. Bartel, and P.D. Zamore. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**: 607-620.

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

Mette, M.F., W. Aufsatz, J. van der Winden, M.A. Matzke, and A.J. Matzke. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *Embo J* **19**: 5194-5201.

Millar, A.A. and F. Gubler. 2005. The *Arabidopsis GAMYB*-like genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* **17**: 705-721.

Mourrain, P., C. Beclin, T. Elmayan, F. Feuerbach, C. Godon, J.B. Morel, D. Jouette, A.M. Lacombe, S. Nikic, N. Picault, K. Remoue, M. Sanial, T.A. Vo, and H. Vaucheret. 2000. *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**: 533-542.

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

O'Donnell, K.A. and J.D. Boeke. 2007. Mighty Piwis defend the germline against genome intruders. *Cell* **129**: 37-44.

Onodera, Y., J.R. Haag, T. Ream, P.C. Nunes, O. Pontes, and C.S. Pikaard. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613-622.

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

Palatnik, J.F., H. Wollmann, C. Schommer, R. Schwab, J. Boisbouvier, R. Rodriguez, N. Warthmann, E. Allen, T. Dezulian, D. Huson, J.C. Carrington, and D. Weigel. 2007. Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Dev Cell* **13**: 115-125.

Papp, I., M.F. Mette, W. Aufsatz, L. Daxinger, S.E. Schauer, A. Ray, J. van der Winden, M. Matzke, and A.J. Matzke. 2003. Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol* **132**: 1382-1390.

Parizotto, E.A., P. Dunoyer, N. Rahm, C. Himber, and O. Voinnet. 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, M.Y., G. Wu, A. Gonzalez-Sulser, H. Vaucheret, and R.S. Poethig. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci U S A* **102**: 3691-3696.

Park, W., J. Li, R. Song, J. Messing, and X. Chen. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* **12**: 1484-1495.

Pasquinelli, A.E., B.J. Reinhart, F. Slack, M.Q. Martindale, M.I. Kuroda, B. Maller, D.C. Hayward, E.E. Ball, B. Degnan, P. Muller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86-89.

Pelissier, T., S. Thalmeir, D. Kempe, H.L. Sanger, and M. Wassenegger. 1999. Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. *Nucleic Acids Res* **27**: 1625-1634.

Peragine, A., M. Yoshikawa, G. Wu, H.L. Albrecht, and R.S. Poethig. 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.

Pillai, R.S., S.N. Bhattacharyya, C.G. Artus, T. Zoller, N. Cougot, E. Basyuk, E. Bertrand, and W. Filipowicz. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**: 1573-1576.

Pillai, R.S., S.N. Bhattacharyya, and W. Filipowicz. 2007. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol* **17**: 118-126.

Pontes, O., C.F. Li, P.C. Nunes, J. Haag, T. Ream, A. Vitins, S.E. Jacobsen, and C.S. Pikaard. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**: 79-92.

Pontier, D., G. Yahubyan, D. Vega, A. Bulski, J. Saez-Vasquez, M.A. Hakimi, S. Lerbs-Mache, V. Colot, and T. Lagrange. 2005. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev* **19**: 2030-2040.

Qi, Y., A.M. Denli, and G.J. Hannon. 2005. Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* **19**: 421-428.

Qi, Y., X. He, X.J. Wang, O. Kohany, J. Jurka, and G.J. Hannon. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008-1012.

Quatrano, R.S., S.F. McDaniel, A. Khandelwal, P.F. Perroud, and D.J. Cove. 2007. *Physcomitrella patens*: mosses enter the genomic age. *Curr Opin Plant Biol* **10**: 182-189.

Rajagopalan, R., H. Vaucheret, J. Trejo, and D.P. Bartel. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* **20**: 3407-3425.

Ray, A., J.D. Lang, T. Golden, and S. Ray. 1996. *SHORT INTEGUMENT* (*SIN1*), a gene required for ovule development in Arabidopsis, also controls flowering time. *Development* **122**: 2631-2638.

Rehwinkel, J., I. Behm-Ansmant, D. Gatfield, and E. Izaurralde. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *Rna* **11**: 1640-1647.

Reinhart, B.J. and D.P. Bartel. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906.

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

Reyes, J.L. and N.H. Chua. 2007. ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J* **49**: 592-606.

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

Rodriguez, A., S. Griffiths-Jones, J.L. Ashurst, and A. Bradley. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**: 1902-1910.

Romano, N. and G. Macino. 1992. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol Microbiol* **6**: 3343-3353.

Ronemus, M., M.W. Vaughn, and R.A. Martienssen. 2006. MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-Dependent RNA Polymerase in *Arabidopsis*. *Plant Cell* **18**: 1559-1574.

Ruby, J.G., C. Jan, C. Player, M.J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D.P. Bartel. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**: 1193-1207.

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

Saito, K., A. Ishizuka, H. Siomi, and M.C. Siomi. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol* **3**: e235.

Saito, K., Y. Sakaguchi, T. Suzuki, H. Siomi, and M.C. Siomi. 2007. Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**: 1603-1608.

Schwab, R., S. Ossowski, M. Riester, N. Warthmann, and D. Weigel. 2006. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell*.

Schwab, R., J.F. Palatnik, M. Riester, C. Schommer, M. Schmid, and D. Weigel. 2005. Specific effects of microRNAs on the plant transcriptome. *Dev Cell* **8**: 517-527.

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

Sen, G.L. and H.M. Blau. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* **7**: 633-636.

Shibata, S., M. Sasaki, T. Miki, A. Shimamoto, Y. Furuichi, J. Katahira, and Y. Yoneda. 2006. Exportin-5 orthologues are functionally divergent among species. *Nucleic Acids Res* **34**: 4711-4721.

Song, J.J., J. Liu, N.H. Tolia, J. Schneiderman, S.K. Smith, R.A. Martienssen, G.J. Hannon, and L. Joshua-Tor. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**: 1026-1032.

Song, J.J., S.K. Smith, G.J. Hannon, and L. Joshua-Tor. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434-1437.

Tang, G., B.J. Reinhart, D.P. Bartel, and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes Dev* **17**: 49-63.

Telfer, A. and R.S. Poethig. 1998. *HASTY*: a gene that regulates the timing of shoot maturation in *Arabidopsis thaliana*. *Development* **125**: 1889-1898.

Tomari, Y., C. Matranga, B. Haley, N. Martinez, and P.D. Zamore. 2004. A protein sensor for siRNA asymmetry. *Science* **306**: 1377-1380.

Tompa, R., C.M. McCallum, J. Delrow, J.G. Henikoff, B. van Steensel, and S. Henikoff. 2002. Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr Biol* **12**: 65-68.

Vagin, V.V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P.D. Zamore. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320-324.

Valoczi, A., E. Varallyay, S. Kauppinen, J. Burgyan, and Z. Havelda. 2006. Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. *Plant J* **47**: 140-151.

Vance, V. and H. Vaucheret. 2001. RNA silencing in plants--defense and counterdefense. *Science* **292**: 2277-2280.

Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* **20**: 759-771.

Vaucheret, H., A.C. Mallory, and D.P. Bartel. 2006. AGO1 homeostasis entails coexpression of *MIR168* and *AGO1* and preferential stabilization of miR168 by AGO1. *Mol Cell* **22**: 129-136.

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

Vazquez, F., H. Vaucheret, R. Rajagopalan, C. Lepers, V. Gasciolli, A.C. Mallory, J.L. Hilbert, D.P. Bartel, and P. Crete. 2004b. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell* **16**: 69-79.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S.I. Grewal, and D. Moazed. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672-676.

Wassenegger, M., S. Heimes, L. Riedel, and H.L. Sanger. 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**: 567-576.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855-862.

Wikstrom, N., V. Savolainen, and M.W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc Biol Sci* **268**: 2211-2220.

Williams, L., S.P. Grigg, M. Xie, S. Christensen, and J.C. Fletcher. 2005. Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. *Development* **132**: 3657-3668.

Wu, G. and R.S. Poethig. 2006. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target *SPL3*. *Development* **133**: 3539-3547.

Wu, M.F., Q. Tian, and J.W. Reed. 2006. *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133**: 4211-4218.

Xie, Z., E. Allen, N. Fahlgren, A. Calamar, S.A. Givan, and J.C. Carrington. 2005a. Expression of *Arabidopsis MIRNA* genes. *Plant Physiol* **138**: 2145-2154.

Xie, Z., E. Allen, A. Wilken, and J.C. Carrington. 2005b. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **102**: 12984-12989.

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

Xie, Z., K.D. Kasschau, and J.C. Carrington. 2003. Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr Biol* **13**: 784-789.

Xu, L., L. Yang, L. Pi, Q. Liu, Q. Ling, H. Wang, R.S. Poethig, and H. Huang. 2006. Genetic interaction between the AS1-AS2 and RDR6-SGS3-AGO7 pathways for leaf morphogenesis. *Plant Cell Physiol* **47**: 853-863.

Yan, K.S., S. Yan, A. Farooq, A. Han, L. Zeng, and M.M. Zhou. 2003. Structure and conserved RNA binding of the PAZ domain. *Nature* **426**: 468-474.

Yang, L., Z. Liu, F. Lu, A. Dong, and H. Huang. 2006a. SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J* **47**: 841-850.

Yang, Z., Y.W. Ebright, B. Yu, and X. Chen. 2006b. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**: 667-675.

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

Yi, R., Y. Qin, I.G. Macara, and B.R. Cullen. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**: 3011-3016.

Yoshikawa, M., A. Peragine, M.Y. Park, and R.S. Poethig. 2005. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev* **19**: 2164-2175.

Yu, B., E.J. Chapman, Z. Yang, J.C. Carrington, and X. Chen. 2006. Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in *Arabidopsis*. *FEBS Lett* **580**: 3117-3120.

Yu, B., Z. Yang, J. Li, S. Minakhina, M. Yang, R.W. Padgett, R. Steward, and X. Chen. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**: 932-935.

Zamore, P.D., T. Tuschl, P.A. Sharp, and D.P. Bartel. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zeng, Y., R. Yi, and B.R. Cullen. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* **100**: 9779-9784.

Zhang, B., X. Pan, C.H. Cannon, G.P. Cobb, and T.A. Anderson. 2006a. Conservation and divergence of plant microRNA genes. *Plant J* **46**: 243-259.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S.W. Chan, H. Chen, I.R. Henderson, P. Shinn, M. Pellegrini, S.E. Jacobsen, and J.R. Ecker. 2006b. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *arabidopsis*. *Cell* **126**: 1189-1201.

Zheng, X., J. Zhu, A. Kapoor, and J.K. Zhu. 2007. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *Embo J* **26**: 1691-1701.

Zhong, R. and Z.H. Ye. 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., X. Cao, and S.E. Jacobsen. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716-719.

Zilberman, D., M. Gehring, R.K. Tran, T. Ballinger, and S. Henikoff. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61-69.

Chapter 1:

A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*

A substantial part of this chapter was originally published in
"A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*"
Ramya Rajagopalan, Hervé Vaucheret, Jerry Trejo and David P. Bartel
Genes & Development 20:3407, 2006.

SUMMARY

In order to discover novel miRNAs in *Arabidopsis thaliana*, we employed a high-throughput pyrosequencing approach to deeply sample the sequences represented in wild-type small RNA libraries. We obtained 887,000 reads corresponding to *Arabidopsis thaliana* small RNAs, representing 340,000 unique sequences, a substantially greater diversity than previously obtained in any species. The deeper coverage afforded by high-throughput small RNA sequencing provided an opportunity to propose additional criteria that can be used to confidently classify miRNAs. We used these criteria to identify 38 previously uncharacterized miRNAs, and dozens of other plausible candidates. The newly identified miRNAs had several characteristics in common that differentiate them from highly conserved miRNAs and imply a recent evolutionary origin along the *Arabidopsis* lineage. They were classifiable into single-gene families, tending to derive from unique loci in the genome, and were typically expressed at lower levels in wild-type tissues compared to the highly conserved miRNAs. Several of these newly discovered miRNAs also retained high homology to their predicted target genes, suggesting that they had evolved from duplications of these genes. One of these patterns suggests a novel sequential duplication mechanism by which miRNA gene evolution can proceed in plants.

Other notable features emerged from closer study of some of the newly identified miRNAs. The *MIR838* locus mapped within an intron of *DICER-LIKE 1* (*DCL1*), suggesting a second homeostatic auto-regulatory mechanism for DCL1 expression. Two miRNAs, miR822 and miR839, depended on DCL4 rather than DCL1 for their accumulation, indicating a second pathway for miRNA biogenesis in plants. The *MIR824* locus produces a miRNA that specifies the downregulation of messages encoding a methyltransferase involved in epigenetic chromatin modifications, *CHROMOMETHYLASE3*, indicating that miRNA-based control also feeds back on transcriptional gene silencing machinery. More generally, our results revealed the existence of a layer of miRNA-based control beyond that found previously, which is evolutionarily much more fluid, employing many newly emergent and diverse miRNAs, each expressed in specialized tissues or at low levels under standard growth conditions.

INTRODUCTION

Small silencing RNAs direct transcriptional and post-transcriptional gene-silencing activities that shape eukaryotic transcriptomes and protein output (Chen 2005; Jones-Rhoades et al. 2006; Mallory and Vaucheret 2006). In plants, these small regulatory RNAs are comprised of microRNAs (miRNAs) and several classes of endogenous small interfering RNAs (siRNAs), which can be differentiated by their distinct modes of biogenesis and the types of genomic loci from which they derive.

The miRNAs derive from primary transcripts that form characteristic stem-loop structures (Carrington and Ambros 2003; Ambros 2004; Bartel 2004; Jones-Rhoades et al. 2006). For characterized *Arabidopsis* miRNAs, this miRNA stem-loop precursor is processed by a Dicer-like RNAseIII-type ribonuclease (DCL1) to generate the miRNA:miRNA* duplex, with 2-nt 3′ overhangs (Park et al. 2002; Reinhart et al. 2002). The miRNA* species derives from the opposing arm of the hairpin and pairs imperfectly to the miRNA (Reinhart et al. 2002). The miRNA strand preferentially incorporates into a silencing complex that has at its core the ARGONAUTE1 (AGO1) protein (Vaucheret et al. 2004; Baumberger and Baulcombe 2005; Qi et al. 2005).

Plant miRNAs have imperfect but extensive complementarity to their mRNA targets, enabling these targets to be predicted with confidence, particularly when the miRNA:target pairing is conserved in multiple species (Rhoades et al. 2002; Jones-Rhoades and Bartel 2004). Plant miRNAs typically direct cleavage of their targets (Llave et al. 2002; Tang et al. 2003). The conserved targets of plant miRNAs are predominantly messages of transcription factors, and the importance of miRNA-mediated regulation of many of these targets for proper embryonic, vegetative, and/or floral development is well established (Chen 2005; Jones-Rhoades et al. 2006; Mallory and Vaucheret 2006). Conserved miRNA targets also include messages for other developmental factors, such as F-box proteins, DCL1 and AGO1, and messages for non-developmental factors, such as stress-response proteins (Chen 2005; Jones-Rhoades et al. 2006; Mallory and Vaucheret 2006).

The identification of previously uncharacterized miRNAs has proven possible using both experimental approaches (Park et al. 2002; Reinhart et al. 2002) and computational prediction (Jones-Rhoades and Bartel 2004), but the latter can miss many

authentic miRNAs that are non-conserved, and expression of the small RNA must in any case be confirmed. We therefore employed a high-throughput protocol to generate a dataset of more than 340,000 unique small RNAs, matching nuclear, plastid, and/or mitochondrial genomes. Within this set of sequences we uncovered 38 novel miRNAs which are unrelated in sequence to previously identified miRNA families, as well as several candidate miRNA loci, suggesting that a majority of plant miRNAs have remained unnoticed due to their low expression in wild-type plants and that many more remain to be discovered.

RESULTS

A diverse set of endogenous small RNAs

To apply high-throughput sequencing to novel miRNA discovery, we adapted our small RNA purification and sequencing protocol, designed to identify RNAs with the size and covalent structure (5′ phosphate and 3′ OH) of DCL products (Lau et al. 2001). *Arabidopsis* small RNAs were sequenced from libraries made from whole seedlings, rosette leaves, whole flowers and siliques. These four runs yielded over 1,500,000 reads. Those with recognizable flanking adaptor sequences and with lengths between 16 and 28 nucleotides were compared to *Arabidopsis* nuclear, chloroplast and mitochondrial genomes. Including another 4239 reads obtained by using conventional methods, 887,266 reads perfectly matched at least one locus and were analyzed further (188,954 from seedling, 186,899 from rosette, 205,649 from flower, and 305,764 from siliques). The 887,266 reads represented 340,114 unique, although sometimes partially overlapping sequences, of which approximately 1/3 matched previously identified or newly discovered miRNAs (described below) (Table 1.1).

The most abundant reads corresponded to conserved, previously identified miRNAs

As expected, known miRNAs were the sequences most redundantly retrieved from the pool, boasting the highest read frequency of all small RNA classes and 15% of the total (Table 1.1). All of the miRNA families known to be conserved to poplar and rice (20 families) or just poplar (one additional family; (Jones-Rhoades et al. 2006) were

represented among our reads, with frequencies as high as 36,093 (miR167). Even the stress-inducible miRNAs miR395 and miR399, previously undetectable in plants grown under normal conditions (Jones-Rhoades and Bartel 2004; Fujii et al. 2005), were represented (13 and 580 reads, respectively), suggesting that some other miRNAs induced in specific conditions might also be represented by multiple reads in our dataset. For a few of the miRNAs, including miR319a/b, the sequenced species differed slightly from the annotated species, suggesting refinements of the annotated species (Supplementary Database 1.1).

Most previously identified conserved miRNA families have multiple, paralogous loci, which combined total 92 loci (Jones-Rhoades et al. 2006). In some cases, members of the same family have slightly different sequences, which can sometimes target distinct sets of messages (Schwab et al. 2005). In other cases, paralogous loci give rise to identical mature miRNAs, raising the question of which paralogs are expressed. Mapping the miRNA* species, rare side products, or degradation fragments unique to a single paralog enabled us to confirm the expression of all but 13 of the 92 loci (Supplementary Database 1.1), including 19 whose transcription had not been previously confirmed, either by cloning or by mapping the 5′ end of primary transcripts (Xie et al. 2005a). The exceptions were for loci for which no reads could be uniquely mapped.

For most miRNAs, variants of the most abundant read were isolated with 5′ or 3′ heterogeneity, evidenced by missing or extra nucleotides at each terminus (Supplementary Database 1.1). Occasional slippage of DCL1 processing presumably gives rise to the extra bases, whereas missing nucleotides could result from slippage or end degradation. In contrast to metazoan miRNAs (Lau et al. 2001), we found that 5′ heterogeneity was common in *Arabidopsis* miRNA pools, and only slightly less prevalent than 3′ heterogeneity, with no correlation between the extent of heterogeneity and the arm of the foldback that produced the miRNA (Supplementary Database 1.1). Plant miRNAs might tolerate more extensive 5′ heterogeneity because seed pairing represents a smaller portion of their targeting specificity (Mallory et al. 2004), whereas animal miRNAs truncated or extended by a single nucleotide at their 5′ end would no longer recognize many normal targets and would instead recognize many other messages (Lim et al. 2005).

The miRNA* species had ~9% as many reads as the mature miRNAs, which was higher than the 1% observed in worms (Ruby et al. 2006). This percentage varied widely. For two of the 21 conserved families (miR395 and miR397), represented by 13 and 361 reads, respectively, no miRNA* species were observed (Supplementary Database 1.1). At the other extreme, miR403* was observed more frequently than mature miR403 (1643 and 66 reads, respectively). Mature miR403 directs cleavage of *AGO2* mRNA (Allen et al. 2005) and is more easily detected by RNA blotting than is miR403* (H.V., unpublished data). We infer that sequencing abundance does not always correlate with *in vivo* abundance, which in any event might not always predict the functional strand.

Several characteristics of *Arabidopsis* miRNAs and their foldbacks emerged from analysis of reads corresponding to miRNA loci that had previously been confidently identified. First, relatively few unique non-overlapping reads mapped to authentic miRNA foldbacks (Supplementary Database 1.1). Although DCL1 processing on some foldbacks appeared a little sloppy, it appeared at least globally very precise, in that most reads centered on the miRNA/miRNA*, even for stems that were quite extensive. Second, the miRNA* sequence was observed for most loci. Third, less than 1% of reads mapped to the strand antisense to that giving rise to miRNA and miRNA* (Supplementary Database 1.1).

Arabidopsis has many miRNAs lacking close orthologs in other sequenced plants

In addition to the 21 conserved miRNA families, another 5 apparently nonconserved miRNA genes (miR158, miR161, miR163, miR173, and miR447) have been confidently identified in *Arabidopsis* (Jones-Rhoades et al. 2006). Each was represented among our reads, with read frequency ranging from 29 for miR447 to 10,573 for miR161 (Supplementary Database 1.1). How many additional nonconserved miRNAs might exist in *Arabidopsis*? The multitude of other endogenous small RNAs, some of which derive from regions with fortuitous potential to fold into miRNA-like hairpins, has complicated miRNA identification in plants, leading to the suggestion that biogenetic requirements be confirmed using mutant backgrounds prior to annotation (Jones-Rhoades et al. 2006). High-throughput sequencing offered an alternative approach for distinguishing miRNAs from other small RNAs. Candidates from loci with a substantial

number of reads deriving from the antisense strand can be excluded because such antisense reads suggest origin from a perfect dsRNA rather than a hairpin. For remaining candidates meeting the conventional hairpin-pairing criteria, a sequenced miRNA* species, especially one with 2-nt 3′ overhangs, provides strong evidence that the candidate originates from a DCL-processed stemloop. As a result, demonstrating that the candidate accumulates in prescribed mutant backgrounds becomes less important, which is particularly helpful for miRNAs difficult to detect on blots.

Using these criteria, we identified 38 additional *Arabidopsis* miRNA families, thereby increasing by 2.5 fold the known diversity of miRNAs in *Arabidopsis* (Fig. 1.1A-B, Table 1.2, Supplementary Database 1.2; this table and Supplementary Database also include a 39th miRNA, miR391, which was absent in miRBase version 7.1). To receive miRNA designation, a miRNA* species (or close variant if the miRNA was sequenced at least 3 times) must have been observed among the reads, with the exception of miR823, which was validated using RNA blotting and the conventional set of mutants (below). The most abundant read on the foldback was deemed the miRNA, although in cases where read density was roughly equivalent for the most abundant reads from each arm of the foldback, both were together annotated to represent the new miRNA locus (using the 5′ and 3′ designations adopted in similar cases for metazoan miRNAs). Many other miRNAs might exist in *Arabidopsis*; another 40 candidates mapped to plausible hairpins but lacked reads representing the miRNA* species (Table 1.3). Four particularly compelling candidates, each sequenced more than 25 times (Table 1.2, candidates A-D), were carried forward in subsequent analysis, anticipating that they will eventually be validated.

A search in plant expressed sequence tag (EST) data sets, and the *Oryza sativa* (rice) and *Populus trichocarpa* (poplar) genomes, revealed potential orthologs for only one of the newly identified miRNAs, miR828, which had recognizable orthologs in poplar (representing the eurosids I lineage) and leafy spurge (representing the eurosids II lineage), each with one substitution in the mature miRNA (Fig. 1.5). For all other newly identified miRNAs, potential orthologs were either absent in sequenced genomes or found only after relaxing the homology criterion to allow three point substitutions. However, most, if not all, of these candidates appeared to be false-positives, because at

this stringency an equivalent number of hits were found that satisfied the homology and pairing criteria but mapped to the non-homologous arms of predicted hairpins. We concluded that most of the newly identified miRNAs do not have identifiable orthologs in the sequence databases and henceforth refer to them all as "nonconserved," while recognizing that a few might have divergent orthologs difficult to identify with confidence, and that many might have orthologs in unsequenced species more closely related to *Arabidopsis thaliana*. Mirroring the search for orthologs, we found no convincing *Arabidopsis* paralogs of the newly identified miRNAs.

Although screening was performed on 20- to 24-nt reads, without preference for a particular length or 5′ nucleotide, 74% of the newly identified miRNA loci encoded a 21-nt miRNA, and 87% encoded a miRNA beginning with a U (Table 1.2). Thus, these characteristics of the conserved miRNAs (Reinhart et al. 2002) were shared by the newly identified miRNAs. Some intriguing tissue specificities were also evident (Table 1.2). For example, miR771 and miR839 were sequenced primarily from flowers, miR391 and miR825 appeared preferentially in rosette leaves, miR822 and miR842 were preferentially sequenced in seedlings, and miR828 was sequenced most often from siliques. For some with the most striking specificities, we speculate that expression might be at a high level within just a few specialized cells within that organ.

DCL4 processes some *Arabidopsis* miRNAs

Most of the newly identified miRNAs were infrequently recovered by deep sequencing, with median read frequencies of only 13, compared to 731 for the conserved families, suggesting that in plants nonconserved miRNAs are generally expressed at low levels or primarily in specific cells or growth conditions. When RNA from plants grown under normal lab conditions was blotted and probed for the ten most abundant newly identified miRNAs, only eight could be detected using either DNA or LNA probes. Accumulation of six of these eight miRNAs displayed the classical biogenetic profile of DCL1-dependency, with insensitivity to defects in any of the other DCL enzymes or RDR proteins (Fig. 1.1C).

In contrast to the previously characterized *Arabidopsis* miRNAs, accumulation of two of the eight miRNAs detectable on RNA blots depended on DCL4, not DCL1 (Fig.

1.2A). One, miR822, was previously classified as an siRNA (ASRP1729) because it accumulates in *dcl1* plants (Allen et al. 2004; Xie et al. 2004). Both miR822 and miR839 were insensitive to defects in RDR2 and RDR6, as expected for RNAs that derive from hairpins rather than dsRNA. Further supporting a hairpin precursor structure was the pattern of reads from these loci (Fig. 1-2B). Over 99% of reads arose from one strand, with only two of the 1892 *MIR822* reads and one of the 332 *MIR839* reads deriving from the antisense strand—a pattern inconsistent with a perfect dsRNA intermediate. Furthermore, the major species from each arm of the predicted foldbacks paired to each other, with 2-nt 3′ overhangs observed for the miR822:miR822* duplex. Although the cleavage precision did not appear to match that of DCL1, this preferred processing from a localized region of an RNA hairpin stem satisfied the defining feature of miRNAs. We concluded that transcripts from a few miRNA loci are processed by DCL4 rather than by DCL1. The dependency on DCL4 for miR822 and miR839 accumulation appeared even higher than that for tasiRNA accumulation; in the absence of DCL4, tasiRNA precursors are processed into 22-nt and 24-nt species by DCL2 and DCL3, respectively (Gasciolli et al. 2005; Xie et al. 2005b), whereas miR822 and miR839 species are not detectable in either *dcl4-1* or *dcl4-2* plants (Fig. 1.2A, data not shown).

Predicted targets of newly identified miRNAs

Conserved miRNA targets can be predicted with very high confidence, whereas in single-genome analyses only the more extensively paired interactions can be predicted with reasonable confidence (Jones-Rhoades and Bartel 2004). To better predict nonconserved interactions, scoring rubrics have been developed that preferentially penalize mismatches to the 5′ and central regions of the miRNA (Allen et al. 2005; Schwab et al. 2005), which are more disruptive than those to the 3′ region of the miRNA (Mallory et al. 2004). When applying the rubric of Allen et al. (2005) in a single-genome search to predict targets of 22 unrelated miRNAs, scoring cutoffs that captured 86% of the experimentally confirmed targets of these miRNAs gave a ratio of authentic to false-positive predictions of 6.9:1, estimated by summing the number of targets predicted for the miRNAs and comparing to the average predicted for ten shuffled cohorts. Using these score cutoffs, we applied the rubric to predict targets of the newly discovered miRNAs,

achieving a lower, although still significant, estimated signal:noise ratio of 3.0:1 (Table 1.2, Supplementary Database 1.3).

One explanation for the apparently lower specificity was that for six miRNAs, the miRNA and miRNA* species were difficult to distinguish from each other, and thus both were included in the target-prediction analysis, recognizing that one of the strands might contribute only false-positive predictions. Similarly, two register-shifted sequences of roughly equally abundance from the miR829 foldback were included. Another explanation might be that some of the newly identified miRNA families have fewer targets with extensive complementarity than do the previously identified families. Indeed, some might not have any biological targets, a subset of which might be "young" DCL1/DCL4 substrates whose processing will soon be lost in the course of neutral evolutionary drift unless a beneficial targeting interaction emerges first. Nonetheless, the prediction of three times as many targets as expected by chance suggested that many of the newly identified miRNAs downregulate genes. Targets for three of the more abundant miRNAs were validated by 5′ RACE (Fig. 1.6). These were *CMT3*, a miR823 target, which encodes a CpNpG DNA cytosine methyltransferase, *AGL16*, a miR824 target, which encodes a MADS-box transcription factor, and *MYB113*, a miR828 target, which encodes a MYB transcription factor.

Predicted targets of the newly-identified miRNAs included transcription factors in the MYB and AP2 families, which each have paralogs known to be targeted by previously identified miRNA families (Supplementary Database 1.3). In addition, members of transcription-factor families not previously known to be regulated by *Arabidopsis* miRNAs, such as MADS-box, ERF (ethylene response factor), WHIRLY and Dof (DNA-binding with one finger) proteins, were among the predicted targets. F-box-containing proteins added to the list of known miRNA targets implicated in protein degradation. A PPR gene distinct from those known to be targeted by miR161, miR400, or *TAS1* or *TAS2* trans-acting siRNAs was also among the predictions. Other predictions extended the biological processes thought to be regulated by miRNAs. For example, 9 jacalin lectins, predicted miR842 and miR846 targets, bind complex carbohydrates and are thought to be involved in initiating pathogen defense responses (Geshi and Brandt 1998). Histone variants, and epigenetic silencing machinery such as CMT3 and a bromo-

adjacent homology (BAH) domain-containing protein, were predicted targets, suggesting that *Arabidopsis* miRNAs regulate transcriptional silencing pathway components in addition to targeting miRNA biogenetic and effector proteins like DCL1 and AGO1.

Evolutionary origins of miRNA genes

Some miRNAs might have arisen from duplication of their target loci, and if so, those that were recently derived might exhibit similarity to their targets that extends beyond the mature miRNA sequence, as observed previously for miR161 and miR163 (Allen et al. 2004). Six of the newly identified miRNA loci displayed extended sequence similarity with their predicted target genes, diagnostic of common origins. Both arms of the *MIR822* gene were previously observed to have an extended alignment to several DC1 domain-containing genes (Allen et al. 2005). The same pattern was seen for *MIR841* and *MIR826* and their predicted targets (Fig. 1.3A-B).

A different pattern was observed for *MIR842* and *MIR846*, suggesting an alternative pathway for miRNA gene emergence. As illustrated for *MIR846*, these genes appeared to derive from two regions of their predicted targets, rather than one (Fig. 1.3C). The simplest explanation for the dual alignment to their targets, with the miRNA arm of the hairpin aligning to one region of the target and the miRNA* arm aligning to the other region, was that a duplication within the targets preceded the duplications that gave rise to the miRNA locus.

Another interesting miRNA–target configuration involved *MIR840*, which was expressed from the opposite strand of its predicted target gene, *AtWhirly3*. This is an arrangement first observed for an Epstein-Barr Virus miRNA and its target (Pfeffer et al. 2004), but one that had not been seen in plants. *AtWhirly3* encodes a homolog of potato p24, a known transcriptional regulator of plant-defense and disease-resistance genes. In the sense orientation, the miRNA was found within the annotated 3′ UTR of a PPR mRNA, At2g02750. Although both strands encode a hairpin, our reads did not include any small RNA sequences from the *AtWhirly3* strand. Either the presumptive miRNA or its star sequence could target the *AtWhirly3* 3′ UTR for cleavage. This implies a mechanism by which the expression of one member of a convergent gene pair influences

the output of the other—a miRNA counterpart to that observed previously for a convergent gene pair that generates nat-siRNAs (Borsani et al. 2005).

Of the 44 genes for the miRNAs and candidates listed in Table 1.2, 35 were in regions between annotated genes, as is typical of plant miRNA genes (Reinhart et al. 2002), whereas nine overlapped protein-coding genes. One was miR840, described above. Of the remaining eight, miR837, miR838, miR848, miR852 and CandidateD overlapped introns, in the same orientation as the protein-coding host gene—an arrangement that bypasses the need to acquire an independent promoter (Baskerville and Bartel 2005). Mature miR837 also had a second match in the genome, located within the same intron that contains the miR837 stem-loop, but in the antisense orientation, suggesting that miR837 might target the pre-mRNA of its host gene, an oligopeptide transporter. miR841 derived from the strand antisense to the intron of At4g13570, a gene closely related to one of its predicted targets but itself not predicted because our search was limited to spliced messages. miR777 and miR834 were localized to the 5′ UTR and 3′ UTR of genes, respectively, with their foldbacks potentially extending into annotated protein-coding regions.

A homeostatic self-regulatory mechanism for *DCL1*

miR838 derived from a hairpin within intron 14 of the *DCL1* mRNA (Fig. 1.4A). The foldback potential of this intron was previously noted, and RACE mapping of the *DCL1* transcript revealed a 4.0 Kb fragment whose 3′ end terminates at the exon 14/15 junction, and a population of ~2.5 Kb fragments, some of which have 5′ ends falling within intron 14 (Xie et al. 2003). Because small RNAs were not detected, the fragments were attributed to aberrant splicing at intron 14 (Xie et al. 2003).

We propose that the presence of this intronic miRNA enables a self-regulatory mechanism that helps maintain *DCL1* homeostasis (Fig. 1.4B). When nuclear DCL1 protein levels are high, the miRNA biogenesis machinery (including DCL1 and HYL1) could compete more efficiently than the splicing machinery for the *DCL1* precursor transcript. If DCL1 began to process the miRNA hairpin before the intron 14 splice sites were defined and juxtaposed during spliceosome formation, then *DCL1* expression would shift towards a pool of truncated, non-functional *DCL1* transcripts, thereby providing a

regulatory feedback mechanism that supplements miR162-directed cleavage (Fig. 1.4C). 5′ RACE confirmed that a population of fragments had 5′ ends terminating at the ends of miR838 (Fig. 1.4A). The low abundance of the miRNA can be explained by the idea that four linkages must be cut to generate the miRNA:miRNA* duplex, whereas just a single cut bisects the mRNA. Perhaps very little of the duplex is fully excised, and as a result the miRNA never accumulates to sufficient levels to direct efficient target cleavage. We suggest that the processing of other intronic miRNAs might also influence the expression of their host genes—speculation bolstered by the presence of a conserved miRNA-like hairpin in the mammalian *DGCR8* gene, whose protein product functions in pri-miRNA processing (Pedersen et al. 2006).

DISCUSSION

A Diverse Set of Newly Emergent miRNAs

Many miRNA candidates have been proposed over the last few years, some of which have been published and annotated in miRBase as authentic *Arabidopsis* miRNAs. Our large dataset provided an opportunity to evaluate these candidates and the methods used to identify them. Beyond the 97 confidently identified genes, none of the other current *Arabidopsis* miRNA annotations (miRBase 7.1) were supported by our data from wild-type plants grown under standard conditions; some of these proposed hairpins matched reads but in a pattern suggestive of endogenous siRNAs (Supplementary Database 1.1). Furthermore, none of the mature miRNA and candidate sequences of Table 1.2 matched recently proposed computational candidates, although for seven of 592 recent miRNA predictions (Lindow and Krogh 2005) there was some overlap, which ranged between 7 and 19 nucleotides (Supplementary Database 2). Apart from homologs of known miRNAs, it appears that the only plant miRNAs to have been identified computationally and subsequently confirmed experimentally were those initially reported by Jones-Rhoades and Bartel (2004), at a time when computational searches that required evolutionary conservation could still be productive because some highly conserved miRNAs remained to be found. miR771, miR772, miR775, miR777, miR779, and one of our candidates (Candidate I) corresponded to miRNA hairpins reported while our

manuscript was in review (Lu et al. 2006). For *MIR772*, the species we annotated as the miRNA appears to be the miRNA*; for *MIR779* the species we sequenced more frequently and annotated as the miRNA derived from a different portion of the hairpin than did miR779.1. Five of our newly identified miRNAs were in a set of 86 candidates previously suggested by analysis of MPSS signatures (Lu et al. 2005) and whose sequences were provided by B. Meyers (personal communication).

Of the 38 newly identified miRNAs, only one, miR828, was clearly conserved in other sequenced genomes.The inferred emergence of the new miRNAs after the divergence of the eurosids I (represented by *Arabidopsis*) and II (represented by poplar), about 90 million years ago (Wikstrom et al. 2001) significantly changes our view of miRNAs in plants. Previously, the proportion of known miRNAs that were conserved among eudicots (*Arabidopsis* and poplar) was quite striking—92 of the 97 known genes, 21 of the 26 known families. With respect to the number of microRNA molecules in wild-type plants, this domination by conserved miRNAs still holds, in that over 87% of the miRNA molecules we sequenced were conserved throughout sequenced flowering plants. However, with respect to the diversity of plant miRNAs, the picture has dramatically broadened to encompass twice as many nonconserved miRNA families as conserved. In addition to the previously known set of highly conserved miRNAs, each typically expressed from multiple genes at high levels, we now know of a much more evolutionarily flexible set of miRNAs, each expressed from single genes at low levels or in very specialized tissues in plants grown under standard conditions. A plot of the cumulative distribution of sequencing frequency illustrates the relationship between conservation and expression that delineates these two sets of miRNAs (Fig. 1.1D). All but three of the 14 families sequenced at a frequency of greater than 1/1000 were conserved, whereas only 11 of the 51 families sequenced at a frequency less than 1/1000 appeared to be conserved.

The identification and characterization of these additional miRNAs also expanded our view of plant miRNA biogenesis. At least two miRNAs, miR822 and miR839, depended on DCL4 rather than DCL1 for their accumulation. Just as DCL1, which is primarily responsible for miRNA biogenesis, can generate some siRNAs (Borsani et al. 2005; Bouche et al. 2006; Henderson et al. 2006), DCL4, which is primarily responsible

for siRNA biogenesis, can generate some miRNAs. The imprecise cleavage of *MIR161*, which yields miR161 5′ termini ranging over 16 nucleotides, and the dual, apparently sequential processing of the *MIR163* hairpin, which yields miR163.1 and miR163.2 (Kurihara and Watanabe 2004), both illustrate that DCL1 processing of some apparently young miRNA hairpins can be quite heterogeneous (Supplementary Database 1.1). DCL4-catalyzed cleavage appears even less precise, with a signature yielding numerous minor products often in phase with the miRNA:miRNA* duplex, suggestive of sequential processing after liberation of the miRNA:miRNA* (Fig. 1.2B). DCL4 can also process perfect hairpins to generate transgene siRNAs (Dunoyer et al. 2005), which are presumably far less defined. To the extent that these transgene hairpins might resemble evolutionary precursors of some miRNAs (Allen et al. 2004), we suggest an adaptive switch from DCL4- to DCL1-mediated processing during the course of miRNA gene emergence and evolution, which is driven by selective pressure for enhanced processing precision as the hairpin acquires substitutions and elevated expression, increasing both the probability and consequences of off-target repression. We suspect that the accumulation of some of the miRNAs that accumulate to levels insufficient to detect by RNA blot might also be DCL4 dependent. One attractive candidate would be *MIR841*, which appears to have emerged recently (Fig. 1.3) and for which register-shifted variants were isolated (Supplementary Database 1.2).

The nonconserved plant miRNAs presumably emerge and dissipate in short evolutionary timescales. Such rapid emergence of new genes is likely facilitated by the small size and simple architecture of miRNA genes. It could be further facilitated by mechanisms in which they can derive from their future targets (Fig. 1.3; (Allen et al. 2004), although it is unclear whether such mechanisms are relevant for most newly emergent miRNAs or just a minority of them. High-throughput sequencing of small RNAs from species closely related to *Arabidopsis* would help define the lifespan of these transient miRNA genes as well as the types of processes that they are particularly prone to control. We suspect that these processes will include those under strong positive selection, such as those involved in pathogen response and reproductive isolation.

With the discovery of this diverse, evolutionarily fluid set of miRNAs sequenced at low frequency, the question arises as to how many more miRNAs remain to be reliably

identified in *Arabidopsis*. Extrapolating from the sequencing frequencies of the conserved miRNAs, there is little reason to suspect that many more conserved families remain to be discovered (Fig. 1.1D). Indeed, the curve for the conserved miRNAs was already beginning to plateau with the identification of the first 13 plant miRNA families (Reinhart et al. 2002). The forecast is quite different for the nonconserved families, for which the curve shows no sign of a plateau, particularly when considering the 40 plausible candidates that appeared to derive from miRNA-like hairpins but did not meet our criteria for confident annotation because their miRNA* species had not been sequenced (Fig. 1.1D, gray symbols; Supplemental Table 1.1). Based on the large number of genomic segments with predicted potential to give rise to miRNA-like hairpins, it has long been easy to speculate that many non-abundant, nonconserved miRNAs might exist in a given plant or animal. For *Arabidopsis*, such speculation now has experimental support.

MATERIALS AND METHODS

Libraries and Sequencing

Wild-type *Arabidopsis* (Columbia accession) plants were grown under standard greenhouse conditions, except seedlings, which were grown as in Reinhart et al. (2002). Total RNA was extracted (Mallory et al. 2001) from whole seedlings, flowers, rosette leaves, and siliques, harvested six days, four weeks, six weeks, and two months after planting, respectively. Small RNA cDNA libraries were prepared for standard sequencing as in Lau et al. (2001) and for bead-in-well pyrophosphate sequencing as in Axtell et al. (2006). Pyrosequencing was performed at 454 Life Sciences (Branford, CT, USA).

Initial processing of reads

cDNA sequences were extracted from raw reads, excluding reads lacking perfect matches to the most proximal 11 nt of both adapter sequences. Unique sequences were mapped to the TAIR/NCBI genome Version 6.0 (November 2005), chloroplast and mitochondrial genomes.

miRNA identification

20- to 24-nt sequences with more than one read, 16 or fewer hits to the genome, and no matches to annotated non-coding RNA were folded using RNAfold with 330 nt of upsteam and downstream flanking sequence. For efficiency, candidate reads were clustered and only the most abundant in a set of overlapping hits was considered. Structures were evaluated using mirCheck, a script that assesses the quality of a foldback based on a battery of parameters that capture known miRNA hairpins (Jones-Rhoades and Bartel 2004). Hairpins that passed this initial filter were then manually screened. Designation as a miRNA required a) a foldback in which the duplex region that included 25 nt centered on the most frequently sequenced read had less than 8 unpaired nt (summing unpaired nt on both arms of the stem) and no more than 3 consecutive unpaired nt, of which no more than 2 were asymmetrically bulged, b) a sequenced miRNA* species (paired to the miRNA within the duplex with 2-nt 3′ overhangs) or for candidates with 3 or more reads, a slight variant of the miRNA*, and c) a sense:antisense read ratio > 0.90. In practice, all but 6 foldbacks that passed manual inspection and were named as miRNA loci had a ratio > 99% (Supplementary Database 1.2). *MIR824*, which has more than 330 nt between the miRNA and the miRNA*, was found in a separate analysis of genomic regions with abundant 21-nt reads.

Target site prediction for miRNAs

Patscan was used to search for near matches (up to six mismatches, or four mismatches and one bulged nucleotide) in TAIR Version 6.0 *Arabidopsis* cDNA database (www.arabidopsis.org) to each miRNA, and target sites were scored as described (Allen et al. 2005). To assess performance, we applied this algorithm to a control set of diverse *Arabidopsis* miRNAs, choosing the most frequently sequenced miRNA variant to represent each known miRNA family for which mRNA targets have been experimentally validated by 5′ RACE, as listed in Jones-Roades et al. (2006). We also generated ten different shuffled cohorts of these 22 miRNAs, preserving dinucleotide composition. Signal:noise ratios were calculated by comparing the total predictions for authentic miRNAs (signal) and the average for shuffled cohorts (noise). Analogously selected

cohorts were also used to estimate specificity of target prediction for the newly identified miRNAs.

RNA gel blot analysis of small RNA expression

RNA gel blots for miRNAs were performed using mutants and protocols as described previously (Vaucheret et al. 2004; Vazquez et al. 2004). Blots were probed with ^{32}P end-labeled DNA (miR822, miR823, miR391(5′), miR771, miR824, miR775) or LNA (miR391(3′), miR839, miR846) oligonucleotides, each complementary to the entire length of the miRNA.

5′ RACE

5′ RACE was performed as described in Jones-Rhoades and Bartel (2004), except that RNA samples were obtained from whole siliques or seedlings, gene-specific primers (Supplemental Text) were designed to be 70 to 400 bases from the predicted cleavage site, and the first gene-specific amplification for *DCL1* was done with the GeneRacer 5′ outer primer and was followed by two nested amplifications done with the GeneRacer 5′ nested primer. The following 5′ RACE gene-specific primers were used to map cleavage sites:

Transcript	Primer	Gene-specific primer sequence
At1g69770 (*CMT3*)	Outer Nested	CGTCCCACCAAAGACGACCAAAAGG CGCTCCCTTTCTCTTGGGGACCTGAC
At3g57230 (*AGL16*)	Outer Nested	CGCTTATGCAGAGGCAAGCTGACAC CGGTTGGCTGAGCTGAAGATGGACA
At1g01040 *(DCL1)*	Outer 1st Nested 2nd Nested	GGGTAACCTCTGAGCACCTCGGATAAGT CCGAGGAGAAAGAAGGTTCTTGC GCAATACGAAACACCACGTCC

Accession Numbers

All genome-matched small RNA sequences generated in this study are accessible at http://www.ncbi.nlm.nih.gov/geo/ as Platform GPL3968, Samples GSM118372, GSM118373, GSM118374, and GSM118375, and Series GSE5228.

Table 1.1. Summary statistics of *Arabidopsis* small RNAs matching miRNA loci

Locus Class	Unique sequences		Reads		Mean frequency[a]	Mean hits[b]
Silencing RNAs						
miRBase annotated miRNA hairpin[c]	960	(0.3%)	138416	(15.6%)	144	1.6
Newly identified miRNA hairpin[c]	361	(0.11%)	7002	(0.8%)	19.4	1.0

[a] Average number of reads per unique sequence

[b] Average number of hits to the nuclear genome

[c] Includes all sequences and reads that mapped to predicted miRNA hairpin precursors

Table 1.2. Newly identified miRNAs in *Arabidopsis*

miRNA	Sequence	Len[a]	Se[b]	R[b]	F[b]	Si[b]	Total	miRNA*[c]	Predicted targets[d]	Proteins of targeted messages[e]
miR391(5')[f]	UUCGCAGGAGAGAUAGCGCCA	21	1	259	12	108	380	yes		
miR391(3')	ACGGUAUCUCUCCUACGUAGC	21	13	215	33	53	314	yes	*At1g72000*	beta-fructofuranosidase
miR771[g]	UGAGCCUCUGUGGUAGCCCUCA	22	0	2	109	26	137	yes		
miR772(5')[g]	UGUAUGUAUGGUCGAAGUAGG	21	7	1	1	11	20	yes*	*At2g28010*	aspartyl protease
miR775[g]	UUCGAUGUCUAGCAGUGCCA	20	97	104	29	155	385	yes**	*At1g53290*	galactosyltransferase Avr9 elicitor
miR777[g]	UACGCAUUGAGUUUCGUUGCUU	22	6	2	0	2	10	yes		
miR779.2[g]	UGAUUGGAAAUUUCGUUGACU	21	17	2	0	1	20	yes		
miR822	UGCGGGAAGCAUUUGCACAUG	21	922	26	81	34	1063	yes	*At2g13900, At5g02330*	4 DC1-domain proteins (4)
miR823	UGGGUGGUGAUCAUAUAAGAU	21	305	14	78	296	693	no	*At1g69770*	CMT3 (2)
miR824	UAGACCAUUUGUGAGAAGGGA	21	254	33	46	90	423	yes	*At3g57230*	AGL16 MADS-box protein
miR825	UUCUCAAGAAGGUGCAUGAAC	21	13	134	1	57	205	yes		
miR826	UAGUCCGGUUUUGGAUACGUG	21	0	0	16	23	39	yes	*At4g03060*	AOP2
miR827	UUAGAUGACCAUCAACAAACU	21	0	8	2	14	24	yes	*At1g02860*	2 SPX C3HC4 RING zinc finger (2)
miR828	UCUUGCUUAAAUGAGUAUUCCA	22	1	0	0	14	15	yes	*At1g66370, At5g52600*	MYB113, MYB82 (3)
miR829.1	CAAAUUAAAGCUUCAAGGUAG	21	11	0	1	0	12	yes	*At5g18560*	AP2-domain ethylene response factor
miR829.2	AGCUCUGAUACCAAAUGAUGGAAU	24	10	1	4	0	15	no		
miR830	UAACUAUUUUGAGAAGAAGUG	21	0	1	5	3	9	yes		
miR831	UGAUCUCUUCGUACUCUUCUUG	22	1	0	0	5	6	yes	*At3g12190*	unknown protein
miR832(5')	UGCUGGGAUCGGGAAUCGAAA	21	0	0	0	6	6	yes	*At2g46960*	CYP709B1 cytochrome P450
miR832(3')	UUGAUUCCCAAUCCAAGCAAG	21	0	0	0	3	3	yes	*At4g30840*	WD-40 protein
miR833(5')	UGUUUGUUGUACUCGGUCUAGU	22	1	2	0	2	5	yes		
miR833(3')	UAGACCGAUGUCAACAAACAAG	22	1	0	2	4	7	yes		
miR834	UGGUAGCAGUAGCGGUGGUAA	21	0	0	2	1	3	yes	*At4g00930*	COP1 interacting protein (5)
miR835(5')	UUCUUGCAUAUGUUCUUUAUC	21	0	0	2	0	2	yes	*At1g49560*	MYB trancription factor (4)
miR835(3')	UGGAGAAGAUACGCAAGAAAG	21	1	0	0	1	2	yes	*At5g46170*	F-box family protein (2)
miR836	UCCUGUGUUUCCUUUGAUGCGUGG	24	0	0	0	2	2	yes		
miR837(5')	AUCAGUUUCUUGUUCGUUUCA	21	1	1	0	0	2	yes	*At1g01160, At4g00850*	2 GIF transcription factors (8)
miR837(3')	AAACGAACAAAAAACUGAUGG	21	1	0	0	2	3	yes		
miR838	UUUUCUUCUACUUCUUGCACA	21	0	1	0	1	2	yes	*At2g45720*	Armadillo/beta-catenin protein (6)
miR839	UACCAACCUUUCAUCGUUCCC	21	39	7	184	4	234	yes*		
miR840	ACACUGAAGGACCUAAACUAAC	22	4	22	3	17	46	yes*	*At2g02740*	WHIRLY3
miR841	UACGAGCCACUUGAAACUGAA	21	2	13	0	9	24	yes*	*At2g38810, At4g13570*	2 H2A.F/Z (3)
miR842	UCAUGGUCAGAUCCGUCAUCC	21	14	0	0	0	14	yes*	*At1g60130, At5g38550*	5 jacalin lectins (7)
miR843	UUUAGGUCGAGCUUCAUUGGA	21	1	4	0	2	7	yes*	*At3g13830*	2 F-box proteins (3)
miR844(5')	UGGUAAGAUUGCUUAUAAGCU	21	4	0	0	1	5	yes*		
miR844(3')	UUAUAAGCCAUCUUACUAGUU	21	3	2	0	3	8	yes*	*At3g46540*	epsin N-terminal homology protein
miR845	CGGCUCUGAUACCAAUUGAUG	21	142	4	153	104	403	yes**		
miR846	UUGAAUUGAAGUGCUUGAAUU	21	26	51	0	21	98	yes**	*At2g25980, At5g49850*	5 jacalin lectins (5)
miR847	UCACUCCUCUUCUUCUUGAUG	21	45	2	2	5	54	yes**	*At1g53720*	cyclophilin-RNA interacting protein (6)
miR848	UGACAUGGGACUGCCUAAGCUA	22	16	22	2	11	51	yes**		
miR849	UAACUAAACAUUGGUGUAGUA	21	12	0	0	4	16	yes**		
miR850	UAAGAUCCGGACUACAACAAAG	22	1	5	0	9	15	yes**		
miR851(5')	UCUCGGUUCGCGAUCCACAAG	21	0	1	9	2	12	yes**		
miR851(3')	UGGGUGGCAAACAAAGACGAC	21	0	0	5	4	9	yes**		
miR852	AAGAUAAGCGCCUUAGUUCUG	21	2	2	0	0	4	yes**		
miR853	UCCCCUCUUUAGCUUGGAGAAG	22	2	0	0	1	3	yes**		
CandidateA	UAAUCCUACCAAUAACUUCAGC	22	0	1	52	0	53	no	*At5g41610*	ATCHX18 cation/H+ exchanger (4)
CandidateB	UAGUAACAGAAUUUGGUGUUA	21	0	0	0	40	40	no	*At1g51700*	22 Dof zinc-finger txn factors (22)
CandidateC	UGAGAUGAAAUCUUUGAUUGG	21	8	0	17	6	31	no	*At2g30690*	unknown protein
CandidateD	UUCGUUGUCUGUUCGACCUUG	21	4	0	16	6	26	no	*At3g08500*	6 MYB transcription factors (7)

[a] Length of mature miRNA sequence

[b] Number of reads from seedlings (Se), rosette leaves (R), flowers (F), siliques (Si)

[c] The sequencing of a miRNA* species is denoted "yes" if the perfect match to the miRNA* (with the 2-nt 3′ overhangs typical of DCL products) was sequenced and was the most abundant read from that arm of the hairpin, "yes*" if the perfect miRNA* was sequenced but was not the most abundant read from that arm of the hairpin, "yes**" if only a close heterogeneous variant of the perfect miRNA* was sequenced, and "no" if no star species was recovered.

[d] AGI codes are given for genes with top-scoring target sites for the miRNA. A complete list of predicted mRNA targets of newly identified miRNAs, along with the associated target-site alignments and scores, is given in Supplementary Database 4.

[e] Protein products of predicted targets in the best score class. The total number of predicted targets falling within the cutoff is given in parentheses if more than one target was predicted.

[f] Reported as a miRNA in Xie et al. (2005a) but was not annotated in miRBase 7

[g] Locus was reported as a miRNA in Lu et al. (2006)

Table 1.3. MicroRNA candidates

Name	ID	Chr[a]	Start	End	Strand	Sequence	Read frequency
CandidateE	AtsRR303453	1	6318325	6318345	-	UACAUUGACCUCCAAGAUCUC	2
CandidateF	AtsRR118557	1	15902750	15902770	-	AGAACUUACCAAUAAUGAUGC	2
CandidateG	AtsRR028649	1	16668107	16668127	-	UAGUGGCUCUCAUUAGGUUAA	2
CandidateH	AtsRR266832	1	22153400	22153420	+	UCUCUCUGUUGUGAAGUCAAA	5
CandidateI	AtsRR187572	1	22799364	22799385	+	UCUAAGUCUUCUAUUGAUGUUC	4
CandidateJ	AtsRR188214	1	23358770	23358790	-	UAGUGGAAGCAGCAACGAGAA	7
CandidateK	AtsRR292789	1	24451488	24451508	+	UGAUUGCCAUCCGACUGUUUC	2
CandidateL	AtsRR030064	1	28060612	28060632	-	UUAACAAUUUCAAGCAAAGAA	5
CandidateM	AtsRR220218	1	11786328	11786350	+	UGGAAGAUGCUUUGGGAUUUAUU	9
CandidateN	AtsRR191288	1	22376144	22376164	+	GUCAUGGGGUAUGAUCGAAUG	6
CandidateO	AtsRR134954	1	2438059	2438079	+	AGGAGGAAGCUCAGAGGUUUG	2
CandidateP	AtsRR095156	2	10725185	10725205	+	UCCAAUAGGUCGAGCAUGUGC	7
CandidateQ	AtsRR098604	2	19694030	19694050	+	UGUUUUGGAUCUUAGAUACAC	2
CandidateR	AtsRR118697	3	5914438	5914458	-	UAAGAAGUGUUACGAAUUGGG	3
CandidateS	AtsRR169958	3	6217565	6217585	+	UGGAUUGGUCAAGGGAAGCGU	3
CandidateT	AtsRR240821	3	6488293	6488313	+	UCAUGUCGUAAUAGUAGUCAC	4
CandidateU	AtsRR271492	4	2185873	2185896	+	UCGCUUGUUGAAUUUAGUCUCGAA	3
CandidateV	AtsRR127497	4	7846826	7846846	+	UGCGAUUGAGAGCAACAAGAC	11
CandidateW	AtsRR306644	4	7891689	7891709	-	UCAAUUGUGCUUUGAAUAAUA	2
CandidateX	AtsRR028199	4	8110327	8110347	-	UUGCAUAUCUCUCGAGUUUGG	4
CandidateY	AtsRR109068	4	11375398	11375418	+	UUGAACAUGGUUUAUUAGGAA	3
CandidateZ	AtsRR054490	4	13295977	13295998	+	UAUAUAACAAUUAGAGGUUGUA	2
CandidateAA	AtsRR135803	4	4800330	4800350	+	UAAAACAGACAAGGAUGUGUG	2
CandidateAB	AtsRR183192	5	113404	113424	+	UAAGUGUACGAAUCAAACAGU	5
CandidateAC	AtsRR102633	5	7168881	7168901	+	UAAGUUAAGAUUUGUGAAGAA	3
CandidateAD	AtsRR183848	5	7583135	7583154	-	UCAGAGUAUCAGCCAUGUGA	4
CandidateAE	AtsRR009294	5	9098879	9098899	+	UCAAUAGAUUGGACUAUGUAU	5
CandidateAF	AtsRR274472	5	12987531	12987551	-	UGUAAUGCCCCUGACCUUGAC	3
CandidateAG	AtsRR114346	5	15908842	15908862	-	UCUGGUGUUGAGAUAGUUGAC	22
CandidateAH	AtsRR049156	5	17020830	17020850	-	UCUGGUUAGAACUUGAAUAAC	2
CandidateAI	AtsRR002060	5	20572201	20572221	-	UUGGUGUUAUGUGUAGUCUUC	2
CandidateAJ	AtsRR195937	5	21178835	21178855	+	ACAAAGUUUUAUACUGACAAU	10
CandidateAK	AtsRR263368	5	21385916	21385935	+	UCGAACUCAUCUUUGGUGUC	2
CandidateAL	AtsRR260180	5	23340235	23340257	+	ACAAAUUCUCGUUGACUAAGCCA	3
CandidateAM	AtsRR332741	5	11060603	11060623	-	UGAGCUUGACCUUCAGGUGGA	2
CandidateAN	AtsRR274472	5	11092222	11092242	+	UGUAAUGCCCCUGACCUUGAC	3

Listed are candidates that mapped to miRNA-like foldbacks but that were sequenced less than 25 times and for which the miRNA* sequence was not recovered.

[a] Chromosome; coordinates are for the TAIR v 6.0 assembly of the *Arabidopsis thaliana* genome.

FIGURE LEGENDS

Figure 1.1. Newly identified *Arabidopsis* miRNAs.
(*A*) Predicted secondary structures of miRNA hairpins highlighting the miRNA (red) and miRNA* species (blue). (*B*) The miRNA hairpins of panel A, shown in bracket notation with a tally of reads mapping to the hairpin and nucleotides colored as in *A*. (*C*) RNA blots demonstrating that accumulation of six detectable miRNAs depended on DCL1, not DCL2, DCL3, DCL4, RDR2 or RDR6. As a loading control, blots were stripped and re-probed for U6. (*D*) Sequencing frequencies of *Arabidopsis* miRNA families. Shown are cumulative plots for all *Arabidopsis* miRNA families (red squares) conserved families (violet diamonds) and all families plus the 40 sequenced candidates (gray triangles). Fourteen families, 11 of which were conserved, were sequenced at a frequency of greater than one per 1000 (dashed line).

Figure 1.2. DCL4-dependent miRNAs in *Arabidopsis*.
(*A*) RNA blots demonstrating that the accumulation of two detectable miRNAs depended on DCL4 and not on DCL1, DCL2, DCL3, RDR2 or RDR6. As a loading control, blots were stripped and re-probed for U6. Similar results were obtained with *dcl4-1* and *dcl4-2* alleles (data not shown). (*B*) Predicted secondary structures of the miRNA hairpins, with lines denoting the sequences mapping to the miRNA (top) and miRNA* (bottom) arm of each hairpin. The thickness and color of the lines correspond to the number of total reads representing each small RNA species, as indicated in the key. The two reads corresponding to the antisense of miR822 and the single read mapping to the antisense of miR839 are not depicted.

Figure 1.3. Extended homology between miRNA genes and their predicted target genes, suggestive of common origin. (*A*) *MIR841*, which illustrates a pattern of extended homology (orange shading) resembling that observed previously for *MIR161* and *MIR163* and their respective targets (Allen et al. 2005). Segments corresponding to the mature miRNA (red) and miRNA* (blue) are indicated. The diagram (*top*) depicts the target gene in right-left polarity, and the alignment (*bottom*) depicts the target gene and miRNA* segment as their reverse complements (rc). Numbers indicate positions in the protein-

coding gene (At2g38810), counting from its first annotated nucleotide. Nucleotides are shaded to indicate those shared by all (orange) or most (yellow) aligned sequences. (*B*) *MIR826*, for which extended homology suggested an evolutionary pathway whereby a later duplication creating the miRNA* arm was nested within an earlier duplication. The genomic proximity of the miRNA and target gene is shown in the upper diagram. For the middle and bottom diagrams, drawing conventions are as in *A*. (*C*) *MIR846*, for which extended homology suggested tandem duplication within an ancestral gene whereby the duplicated regions independently gave rise to the miRNA and miRNA* segments.

Figure 1.4. An intronic hairpin positioned so as to mediate DCL1 auto-regulation. (*A*) Intron 14 of the *DCL1* primary transcript, and the predicted hairpin structure of miR838. Arrows indicate 5′ ends of RACE-mapped fragments. (*B*) Alternative fates of the *DCL1* primary transcript, which appears to undergo either splicing to generate full-length *DCL1* mRNA or processing by DCL1 itself to generate transcript fragments severed within intron 14. (*C*) A schematic of *DCL1* posttranscriptional auto-regulation. When DCL1 protein levels are high, it could compete with splicing machinery for access to intron 14, thereby supplementing miR162-mediated regulation to maintain the proper level of *DCL1* mRNA.

Figure 1.5. Predicted orthologs of miR828 in other eudicots. The EST accession (*E. esula*) or chromosomes (*P. trichocarpa*) are noted. The predicted miRNA is rendered in red.

Figure 1.6. Mapping of miRNA-directed cleavage using 5′ RACE.
The arrows indicate positions at which cleavage products were detected. The fraction of clones mapping to that site are indicated.

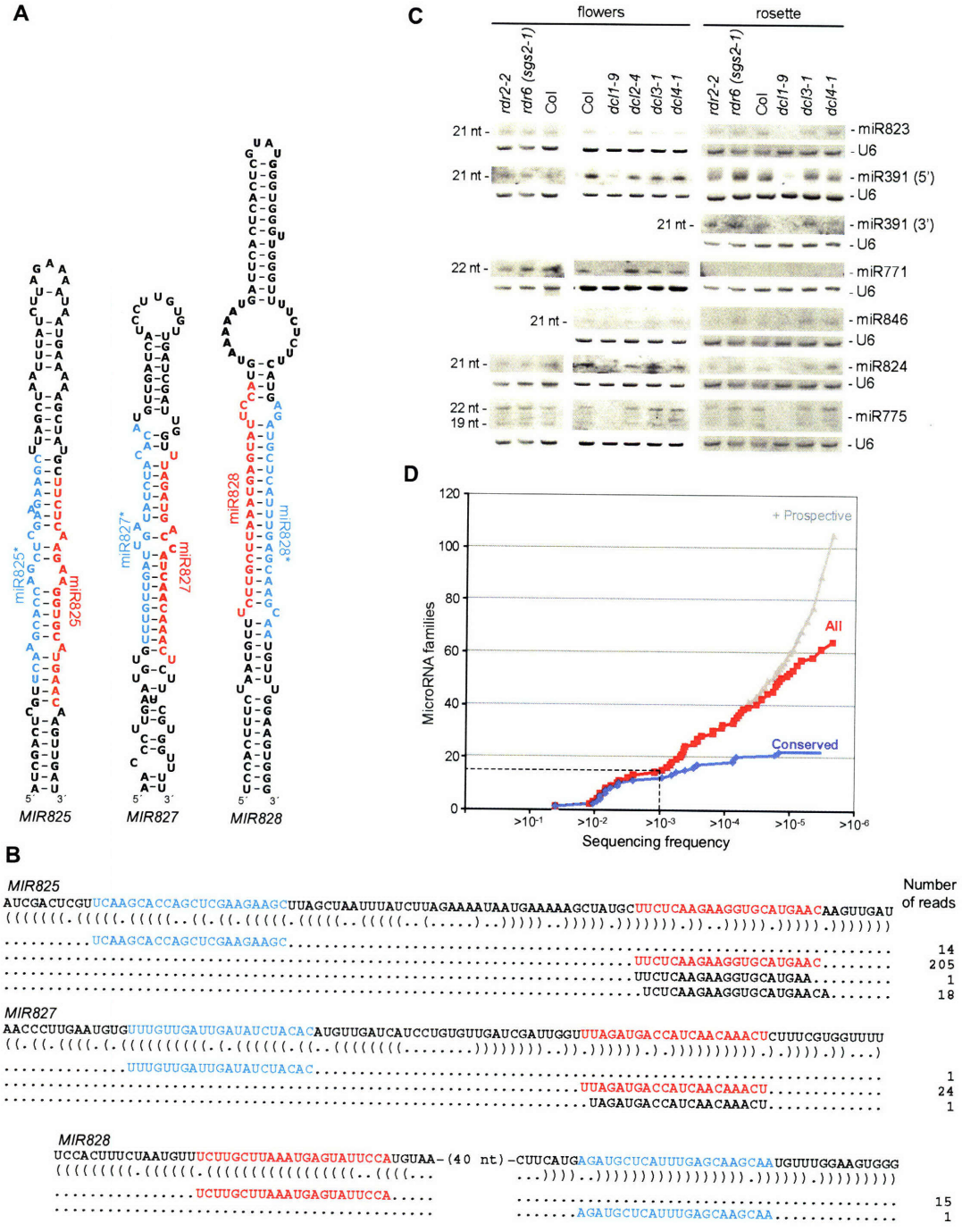


Figure 1.1. Newly identified *Arabidopsis* miRNAs.

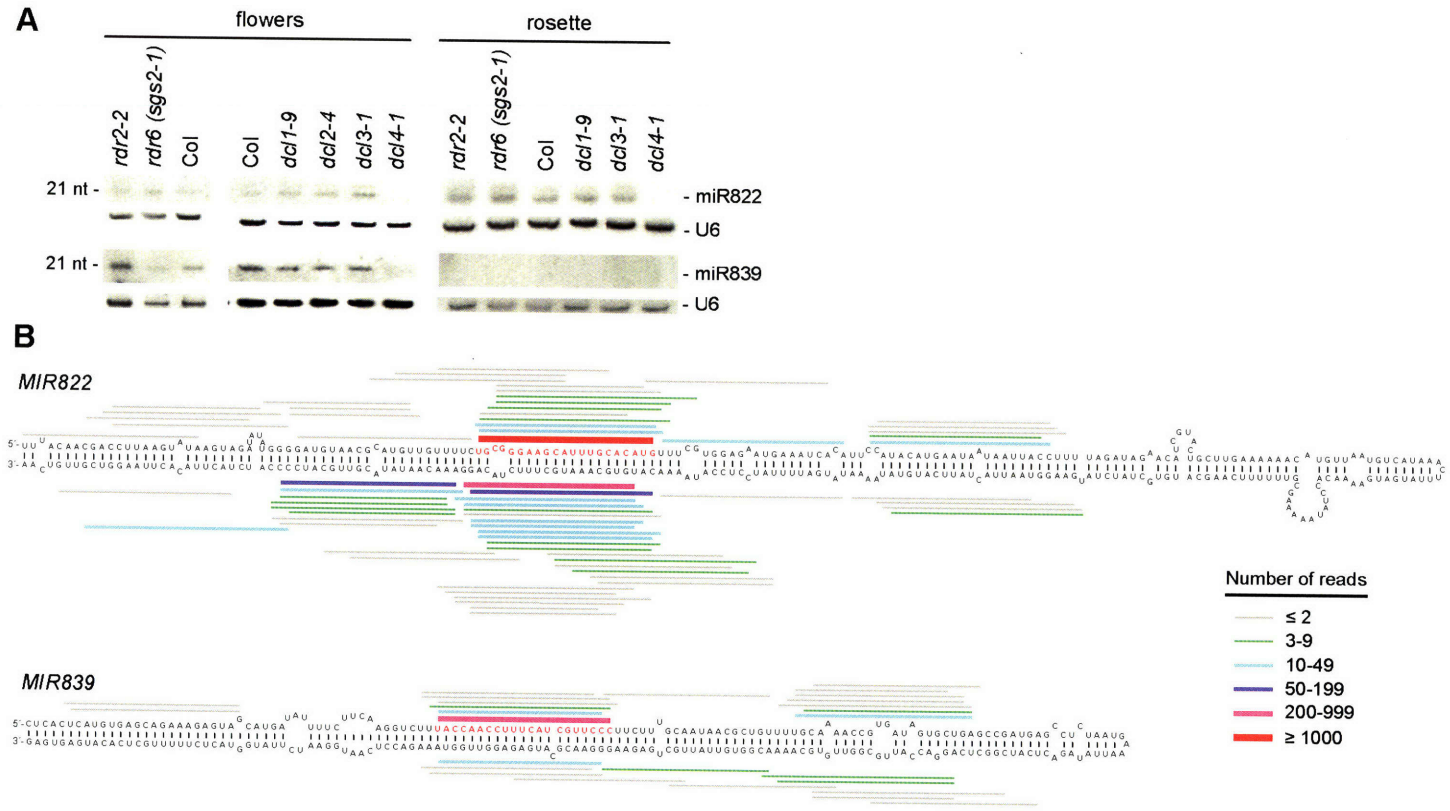


Figure 1.2. DCL4-dependent miRNAs in *Arabidopsis*.

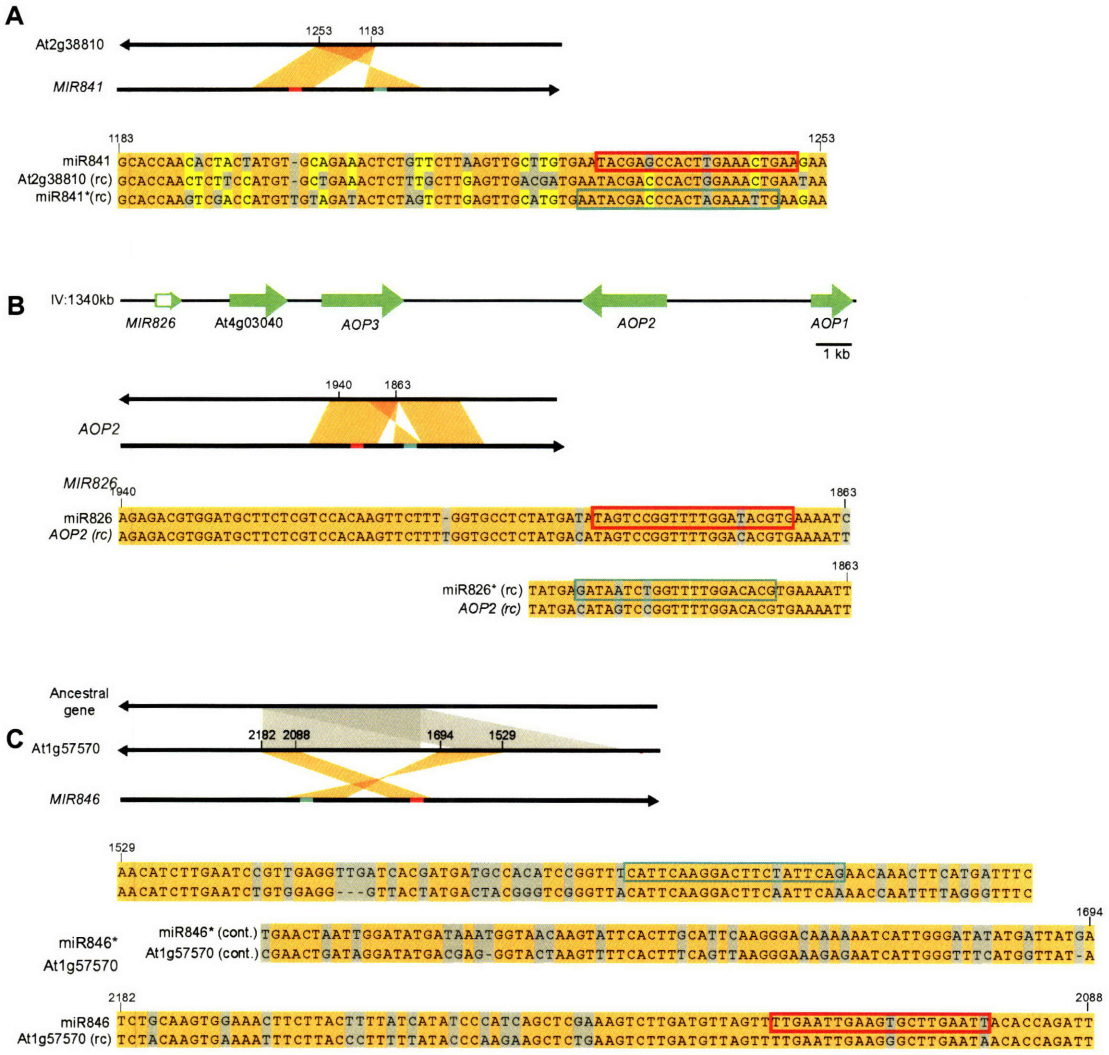


Figure 1.3. Extended homology between miRNA genes and their predicted target genes, suggestive of common origin.

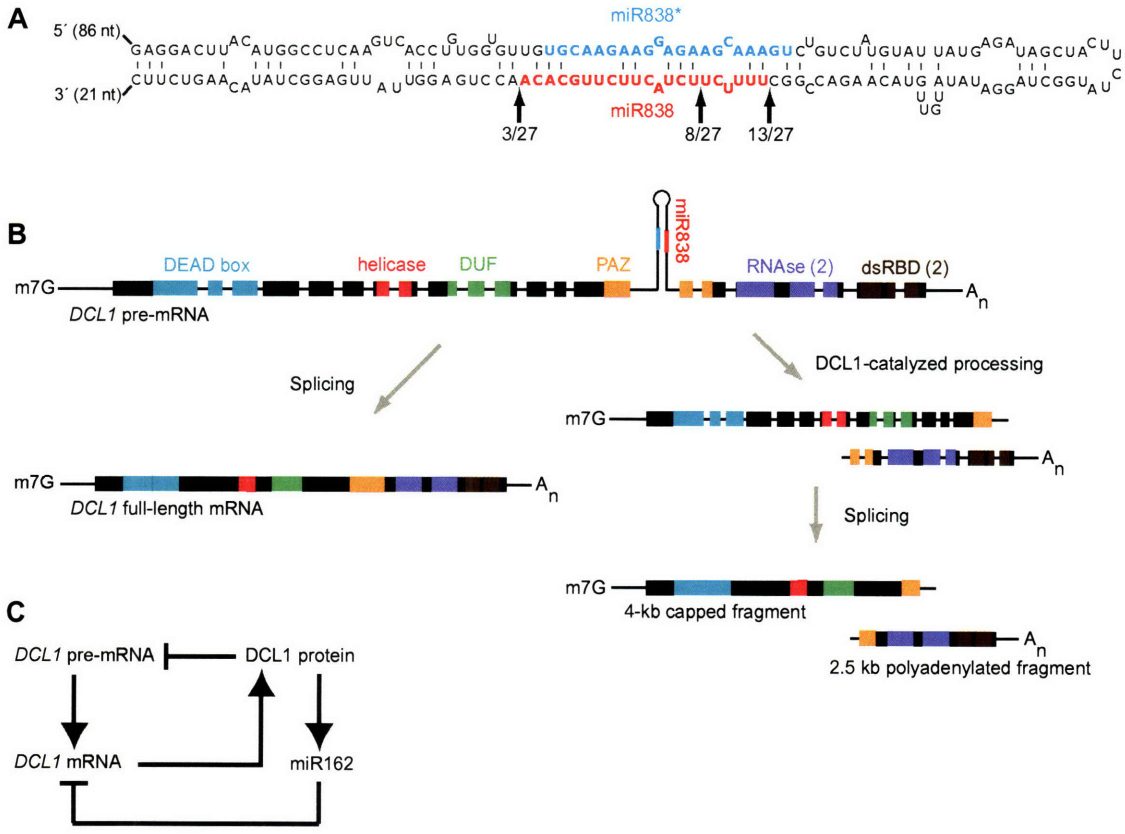


Figure 1.4. An intronic hairpin positioned so as to mediate DCL1 autoregulation.

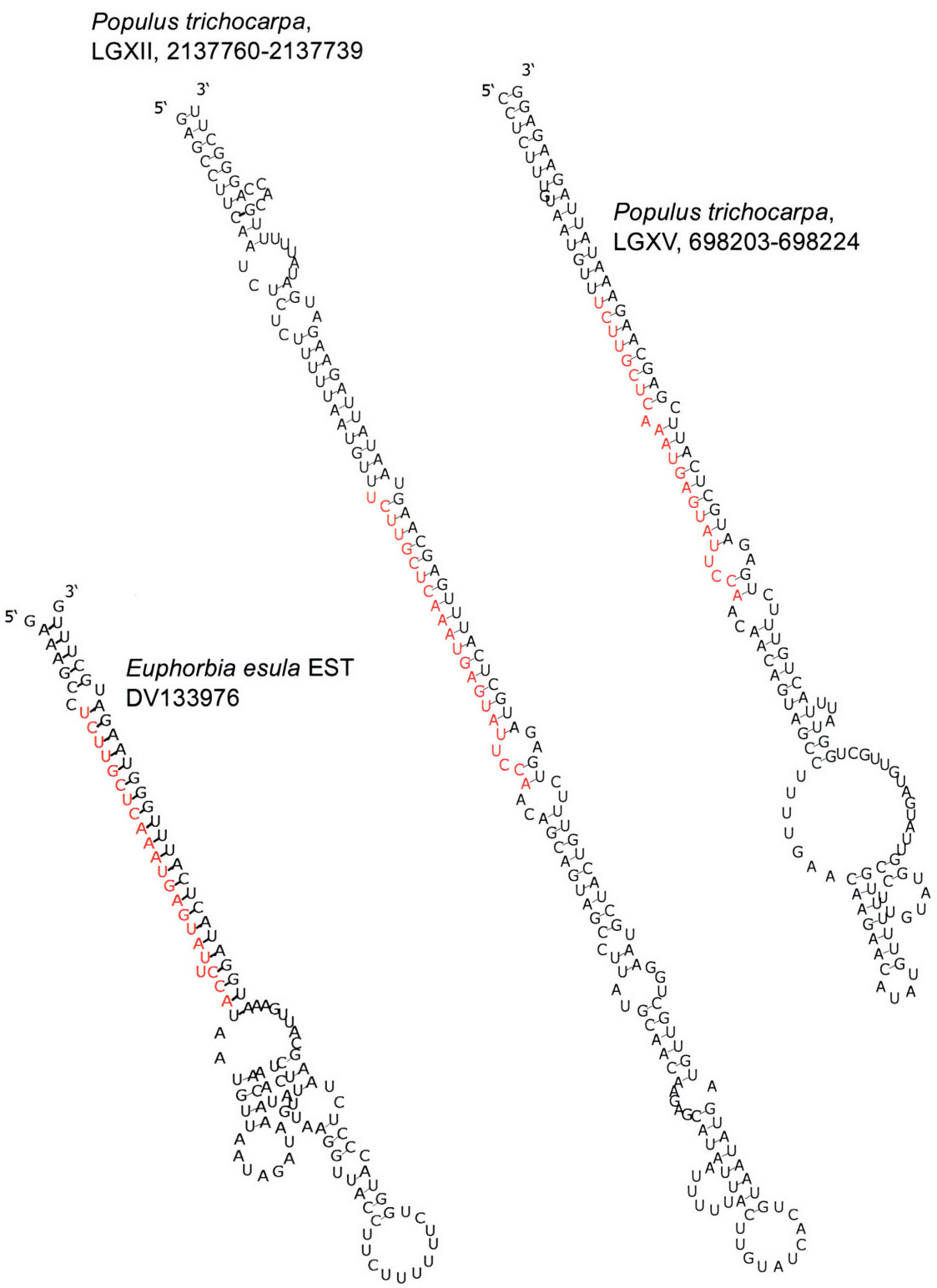


Figure 1.5. Predicted orthologs of miR828 in other eudicots.

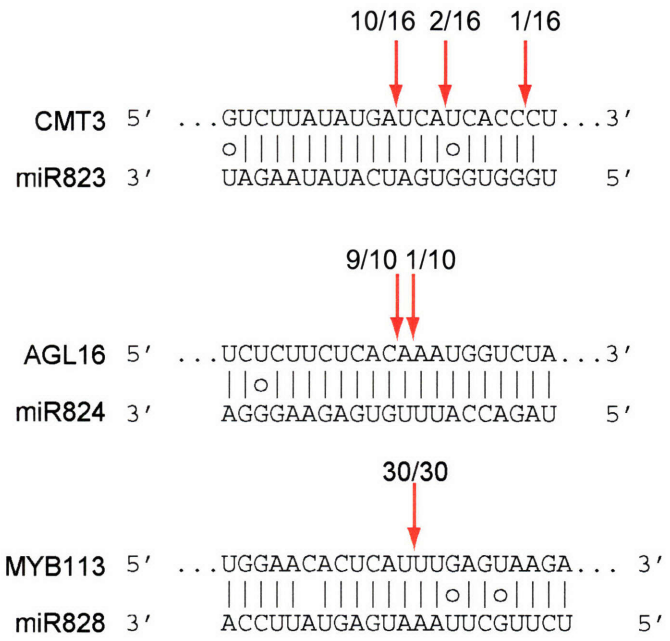


Figure 1.6. Mapping of miRNA-directed cleavage by 5' RACE.

SUPPLEMENTARY DATABASES

Supplementary Databases 1.1–1.3 can be accessed at the following WWW links:

Supplementary Database 1.1. Small RNA abundance profiles at previously annotated miRNA loci.

http://web.wi.mit.edu/bartel/pub/Supplemental%20Material/Rajagopalan2006/SD1/_SUPPL-DATABASE_1.htm

Supplementary Database 1.2. Small RNA abundance profiles at newly identified miRNA loci.

http://web.wi.mit.edu/bartel/pub/Supplemental%20Material/Rajagopalan2006/SD2/_SUPPL-DATABASE_2.htm

Supplementary Database 1.3. Predicted target genes of newly identified miRNAs. Within each entry, the first line specifies the AGI code and the coordinates of the target site within the cDNA sequence (given as [Start, End] from the 5′ end of the cDNA). The score of the best-scoring 20mer of the miRNA and its corresponding target site follows. The second line of each entry gives the gene annotation, and the final lines of each entry display the alignment of the miRNA to the predicted target site. All annotations and cDNA sequences are from the TAIR v6.0 build of the *A. thaliana* genome.

http://www.genesdev.org/content/vol20/issue24/images/data/3407/DC1/SupplDatabase4.txt

REFERENCES

Allen, E., Z. Xie, A.M. Gustafson, and J.C. Carrington. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121: 207-221.

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

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

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

Baskerville, S. and D.P. Bartel. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11: 241-247.

Baumberger, N. and D.C. Baulcombe. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* 102: 11928-11933.

Borsani, O., J. Zhu, P.E. Verslues, R. Sunkar, and J.K. Zhu. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123: 1279-1291.

Bouche, N., D. Lauressergues, V. Gasciolli, and H. Vaucheret. 2006. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *Embo J.*

Carrington, J.C. and V. Ambros. 2003. Role of microRNAs in plant and animal development. *Science* 301: 336-338.

Chen, X. 2005. MicroRNA biogenesis and function in plants. *FEBS Lett* 579: 5923-5931.

Dunoyer, P., C. Himber, and O. Voinnet. 2005. DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat Genet* 37: 1356-1360.

Fujii, H., T.J. Chiou, S.I. Lin, K. Aung, and J.K. Zhu. 2005. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr Biol* 15: 2038-2043.

Gasciolli, V., A.C. Mallory, D.P. Bartel, and H. Vaucheret. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr Biol.*

Geshi, N. and A. Brandt. 1998. Two jasmonate-inducible myrosinase-binding proteins from *Brassica napus* L. seedlings with homology to jacalin. *Planta* 204: 295-304.

Henderson, I.R., X. Zhang, C. Lu, L. Johnson, B.C. Meyers, P.J. Green, and S.E. Jacobsen. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* 38: 721-725.

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

Jones-Rhoades, M.W., D.P. Bartel, and B. Bartel. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol.*

Kurihara, Y. and Y. Watanabe. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* 101: 12753-12758.

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

Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433: 769-773.

Lindow, M. and A. Krogh. 2005. Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics* 6: 119.

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

Lu, C., K. Kulkarni, F.F. Souret, R. Muthuvalliappan, S.S. Tej, R.S. Poethig, I.R. Henderson, S.E. Jacobsen, W. Wang, P.J. Green, and B.C. Meyers. 2006. MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.*

Lu, C., S.S. Tej, S. Luo, C.D. Haudenschild, B.C. Meyers, and P.J. Green. 2005. Elucidation of the small RNA component of the transcriptome. *Science* 309: 1567-1569.

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

Mallory, A.C., B.J. Reinhart, M.W. Jones-Rhoades, G. Tang, P.D. Zamore, M.K. Barton, and D.P. Bartel. 2004. MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5' region. *Embo J* 23: 3356-3364.

Mallory, A.C. and H. Vaucheret. 2006. Functions of microRNAs and related small RNAs in plants. *Nat Genet* 38 Suppl 1: S31-36.

Park, W., J. Li, R. Song, J. Messing, and X. Chen. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 12: 1484-1495.

Pedersen, J.S., G. Bejerano, A. Siepel, K. Rosenbloom, K. Lindblad-Toh, E.S. Lander, J. Kent, W. Miller, and D. Haussler. 2006. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* 2: e33.

Pfeffer, S., M. Zavolan, F.A. Grasser, M. Chien, J.J. Russo, J. Ju, B. John, A.J. Enright, D. Marks, C. Sander, and T. Tuschl. 2004. Identification of virus-encoded microRNAs. *Science* 304: 734-736.

Qi, Y., A.M. Denli, and G.J. Hannon. 2005. Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* 19: 421-428.

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

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

Ruby, J.G., C. Jan, C. Player, M. Axtell, W. Lee, C. Nusbaum, H. Ge, and D.P. Bartel. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *Caenorhabditis elegans*. *Cell*.

Schwab, R., J.F. Palatnik, M. Riester, C. Schommer, M. Schmid, and D. Weigel. 2005. Specific effects of microRNAs on the plant transcriptome. *Dev Cell* 8: 517-527.

Tang, G., B.J. Reinhart, D.P. Bartel, and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes Dev* 17: 49-63.

Vaucheret, H., F. Vazquez, P. Crete, and D.P. Bartel. 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., H. Vaucheret, R. Rajagopalan, C. Lepers, V. Gasciolli, A.C. Mallory, J.L. Hilbert, D.P. Bartel, and P. Crete. 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell* 16: 69-79.

Wikstrom, N., V. Savolainen, and M.W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc Biol Sci* 268: 2211-2220.

Xie, Z., E. Allen, N. Fahlgren, A. Calamar, S.A. Givan, and J.C. Carrington. 2005a. Expression of *Arabidopsis* MIRNA genes. *Plant Physiol* 138: 2145-2154.

Xie, Z., E. Allen, A. Wilken, and J.C. Carrington. 2005b. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 102: 12984-12989.

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

Xie, Z., K.D. Kasschau, and J.C. Carrington. 2003. Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr Biol* 13: 784-789.

Chapter 2:

Small silencing RNAs in *Arabidopsis thaliana*

A substantial part of this chapter originally appeared in
"A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*"
Ramya Rajagopalan, Hervé Vaucheret, Jerry Trejo and David P. Bartel
Genes & Development 20:3407, 2006.

SUMMARY

To better understand the diversity of small silencing RNAs expressed in plants, we employed high-throughput pyrosequencing to obtain 887,000 reads corresponding to *Arabidopsis thaliana* small RNAs. They represented 340,000 unique sequences, a substantially greater diversity than previously obtained in any species. Most of the small RNAs had the properties of heterochromatic small interfering RNAs (siRNAs) associated with DNA silencing, in that they were preferentially 24 nt long and mapped to intergenic regions. The density of non-miRNA small RNAs was greatest in the proximal and distal pericentromeric regions. Contrary to other published reports, these candidate siRNAs displayed only a slightly preferential propensity to match repetitive elements. Many derived from small RNA "hotspots" in the genome, which give rise to an abundance of small RNA products over short intergenic distances. We also identified a novel and apparently lineage-specific *trans*-acting siRNA gene (*TAS4*), which may be implicated in regulating anthocyanin production and leaf senescence. A newly identified miRNA, miR824, cleaves the *TAS4* precursor transcript, which presumably sets the phase for siRNA production at this locus.

INTRODUCTION

In plants, small regulatory RNAs are categorized as belonging to the microRNAs (miRNAs) or to one of several classes of endogenous small interfering RNAs (siRNAs). MicroRNAs and siRNAs can be distinguished from each other by their unique biogenetic dependencies and by the types of genomic loci from which they derive.

Plant miRNAs are expressed initially as primary transcripts with enough self-complementarity to enable their folding into characteristic stem-loop secondary structures (Carrington and Ambros 2003; Ambros 2004; Bartel 2004; Jones-Rhoades et al. 2006). This miRNA stem-loop precursor is processed by one of two RNAseIII-type ribonucleases known as DICER-LIKE (DCL) proteins (DCL1 or DCL4) to generate the miRNA:miRNA* imperfectly paired duplex, with 2-nt 3′ overhangs (see Chapter I and (Park et al. 2002; Reinhart et al. 2002). The miRNA strand is preferentially selected for incorporation into the RNA-Induced Silencing Complex (RISC) where it specifies downregulation of complementary mRNA messages by ARGONAUTE1-catalyzed cleavage (Vaucheret et al. 2004; Baumberger and Baulcombe 2005; Qi et al. 2005).

Endogenous siRNAs derive from long double-stranded RNA (dsRNA) formed as a product of an RNA-dependent RNA polymerase (RdRP), convergent transcription, or transcription of repeats. They typically perform auto-silencing, in that they target DNA or transcripts corresponding to (or related to) the loci from which they derive. Exceptions are the ~21-nt *trans*-acting siRNAs (tasiRNAs), which derive from non-protein-coding genes known as *TRANS-ACTING siRNA* (*TAS*) genes and posttranscriptionally down-regulate protein-coding transcripts from unrelated loci, in a fashion reminiscent of miRNA-directed repression (Peragine et al. 2004; Vazquez et al. 2004; Allen et al. 2005; Yoshikawa et al. 2005). A segment of the *TAS* transcript is converted to dsRNA by RDR6, which is then successively cleaved by DCL4 into 21-nt siRNAs, which are then loaded into an AGO1- or AGO7-containing silencing complex where they direct cleavage of the mRNA targets (Peragine et al. 2004; Vazquez et al. 2004; Allen et al. 2005; Gasciolli et al. 2005; Xie et al. 2005; Yoshikawa et al. 2005; Adenot et al. 2006; Fahlgren et al. 2006; Hunter et al. 2006). One hallmark of tasiRNAs is that they are processed in phase from predominantly one register, which greatly decreases the diversity of siRNAs

that accumulate to appreciable levels from a particular TAS locus and ensures production of those with intended targets (Vazquez et al. 2004; Allen et al. 2005). To define the proper phasing register, each of the known *TAS* transcripts is cleaved by a miRNA-programmed silencing complex (Allen et al. 2005; Yoshikawa et al. 2005).

Another type of endogenous siRNA directing PTGS in plants is natural antisense siRNA (nat-siRNA). In the founding example of nat-siRNA-directed silencing, high salt levels induce the expression of *SRO5*, one of a pair of convergently transcribed genes, such that in the presence of transcripts from the other gene, *P5CDH*, a DCL2/RDR6/SGS3/NRPD1a-dependent 24-nt siRNA is produced that directs cleavage of *P5CDH* transcripts (Borsani et al. 2005). This creates a terminus for RdRP production of dsRNA and subsequent processing into secondary siRNAs by DCL1, which can also target *P5CDH* messages (Borsani et al. 2005).

A third type of endogenous siRNA found in plants is heterochromatic siRNA. The concerted activity of plant-specific DNA-dependent RNA polymerases, PolIVa and PolIVb, correlates with the accumulation of 24-nt heterochromatic siRNAs via RDR2-mediated dsRNA formation and DCL3-mediated processing (Xie et al. 2004; Chan et al. 2005; Vaucheret 2005). A fraction of these siRNAs associate with AGO4 to form a silencing complex thought to direct sequence-specific methylation events at the DNA and/or chromatin level, which in turn can lead to heterochromatin formation and maintenance at loci from which the siRNAs arise, such as retroelements and the 5S rDNA arrays (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). Other siRNAs depend on PolIVa-RDR2-DCL3 but not PolIVb or AGO4, and are not associated with DNA methylation and heterochromatin (Zilberman et al. 2003; Xie et al. 2004; Pontier et al. 2005; Pontes et al. 2006). Their function remains unknown.

Conventional cloning and sequencing of small RNAs from *Arabidopsis* has suggested that plants have a remarkable diversity of endogenous small RNAs (Llave et al. 2002; Park et al. 2002; Reinhart et al. 2002; Sunkar and Zhu 2004; Xie et al. 2004). Recently, massively parallel signature sequencing (MPSS) was employed to obtain a set of 77,434 unique 17-nt signatures of endogenous small RNAs from wild-type *Arabidopsis* (Lu et al. 2005). An appealing alternative that combines the full-length small RNA information of conventional sequencing with the high-throughput character of

MPSS is a pyrophosphate-based high-throughput sequencing technique (Margulies et al. 2005). This technique has recently been applied on a pilot scale to the sequencing of *Arabidopsis* small RNAs, generating between 13,000 to 45,000 unique sequences that match the genome (Henderson et al. 2006; Lu et al. 2006; Qi et al. 2006). In order to more broadly characterize the genomic distribution of loci that produce small RNAs in *Arabidopsis*, we used high-throughput pyrosequencing to obtain more than 340,000 unique small RNA sequences that matched the nuclear, plastid or mitochondrial genomes. Analysis of this dataset provided insights into the evolution, genomics, expression, biogenesis and function of small silencing RNAs in *Arabidopsis*.

RESULTS

Deep sequencing reveals a high degree of small RNA sequence diversity in *Arabidopsis*

Initally, we obtained 4239 reads by conventional Sanger sequencing of libraries made using a cloning protocol that enriches for RNAs with the characteristic termini (5′ phosphate and 3′ OH) and length of DCL products (Lau et al. 2001). We next adapted this small RNA purification and sequencing protocol to take advantage of high-throughput pyrophosphate sequencing. *Arabidopsis* small RNAs were sequenced from libraries made from total RNA isolated from whole seedlings, rosette leaves, whole flowers and siliques of wild-type plants. When pyrosequencing data generated from these four libraries was pooled, the resulting dataset contained over 1,500,000 reads. We filtered out sequences without recognizable flanking adaptor sequences and outside the size range of 16-28 nucleotides. The remaining sequences were aligned to the *Arabidopsis* nuclear, chloroplast and mitochondrial genomes. Including the 4239 pilot reads, 887,266 reads perfectly matched at least one locus and were analyzed further (188,954 from seedling, 186,899 from rosette, 205,649 from flower, and 305,764 from siliques).

These 887,266 reads represented 340,114 unique, although sometimes partially overlapping sequences (Table 2.1). About 65% (221,676) of the unique sequences were only sequenced once. The distribution of lengths and 5′ nucleotides for the set of

singletons and for the set of sequences with multiple reads were virtually identical (data not shown), suggesting that the two sets represented similar classes of small RNAs. Comparison to a dataset of 77,434 unique 17-nt MPSS signatures representing small RNAs that match the *Arabidopsis* genome (Lu et al. 2005) found only 13,596 unique signatures that matched the first 17 nucleotides of at least one of our unique reads. Together, the preponderance of singletons in our library and the modest overlap with the MPSS dataset indicated that deep sequencing approaches were still far from saturating the small RNA pools in *Arabidopsis*. Although many small RNAs expressed in *Arabidopsis* remained unidentified, our dataset represented a substantial increase in the known diversity of small RNAs matching the genome.

A newly identified tasiRNA gene in *Arabidopsis*

In a search for tasiRNA loci, we implemented a clustering algorithm that scanned the genome for phased clusters of ~21-nt reads. This procedure found all five of the previously identified *Arabidopsis* tasiRNA genes (*TAS1a*, *TAS1b*, *TAS1c*, *TAS2*, and *TAS3*; Supplementary Database 1) and discovered an additional locus, *TAS4* (Fig 2.1A), mapping between At3g25800 and At3g25790 (a MYB transcription factor).

Because miRNA-directed cleavage sets the phase for and stimulates production of tasiRNAs (Allen et al. 2005; Gasciolli et al. 2005; Yoshikawa et al. 2005; Axtell et al. 2006), we searched *TAS4* for miRNA complementary sites upstream and downstream of the region that generated small RNAs. It identified a single miR828 complementary site. Cleavage at this site, validated by 5′ RACE, defined a 5′ terminus that matched that of the most proximal siRNAs arising from this locus and was in perfect register with the other predominant siRNAs (Fig 2.1A). The EST mapping to this region (AU226008) corresponded to the opposite strand of the inferred primary transcript and presumably represented the RDR6-polymerized strand. Although poplar ESTs with miR828 complementary sites were found, conservation of *AtTAS4* to poplar was unclear.

We also predicted three targets for *TAS4*-siR81(-), one of the dominant *TAS4* siRNAs. The predicted targets, *PAP1*/*MYB75*/At1g56650, *PAP2*/*MYB90*/At1g66390, and *MYB113*/At1g66370 (Fig 2.1B), encoded three MYB transcription factors, which were distinct from the MYB genes targeted by miR159, and those with complementarity to

miR835 and CandidateD (see Chapter I). PAP1 and PAP2 regulate expression of anthocyanin/flavenoid and phenylpropanoid biosynthetic genes, and might also be involved in regulating leaf senescence (Borevitz et al. 2000; Pourtau et al. 2006). Intriguingly, miR828 was also predicted to down-regulate *MYB113* at an independent target site (Fig 2.1B), which suggested a close functional evolutionary relationship among these MYB target genes, miR828 and the *TAS4* cluster. Using 5′ RACE, we identified mRNA cleavage fragments diagnostic of miR828-directed cleavage of *MYB113*, and tasiRNA-directed cleavage of *PAP2*, thereby experimentally confirming these predicted targets and demonstrating that the *TAS4* locus was indeed *trans*-acting.

When considered as a group, the 10,469 reads from all six *TAS* genes were predominantly 21 nt and tended to begin with a uridine, as also observed for *Arabidopsis* miRNAs (Fig 2.2). MicroRNA and synthetic siRNA duplexes assemble into the silencing complex asymmetrically, such that the strand that pairs with less stability at its 5′ terminus is incorporated as the guide strand, while the other strand is degraded (Khvorova et al. 2003; Schwarz et al. 2003). Analysis of the initial 12 siRNA reads from *TAS1a* suggests that tasiRNAs also obey the asymmetry guidelines (Vazquez et al. 2004). The acquisition of many additional tasiRNA reads enabled us to revisit this issue. For each of the six *TAS* genes, each possible duplex represented by a 21mer read (including those out of register with the dominant phasing register) was considered and evaluated for which strand of the duplex yielded more reads. For 57% of the duplexes with energetically distinct terminal basepairing, the strand with more reads was the one that appeared to be least stably paired at its 5′ terminus. Confounding this analysis, however, was the preference for a U at tasiRNA 5′ termini. If assembly or stability of the silencing complex simply preferred a U at the 5′ terminus of the guide strand, without regard for the differential pairing stabilities at the duplex ends, then there would more frequently be an A:U pair at the 5′ terminus of the guide strand than at the 5′ terminus of the passenger strand, thereby generating artifactual adherence to the pairing asymmetry guidelines. Indeed, the weak adherence vanished when repeating our analysis considering only the duplexes with an A or U at the 5′ termini of both the guide and passenger strands. Apparently, the pairing asymmetry guidelines do not apply for tasiRNAs.

Small RNAs deriving from rRNA, tRNA and organellar loci

A little over 15% of the reads (7.5% of the unique sequences) in the dataset represented the sense strand of 25S, 18S, 5.8S or 5S ribosomal RNAs from the nuclear, mitochondrial, and/or chloroplast genomes (Table 2.1). This was a substantial but unsurprising proportion given that between 35-60% of the transcriptional output of active eukaryotic cells comprises ribosomal RNAs (Santoro 2005). In addition, 2.8% of the dataset corresponded to the sense strand of tRNAs in the nuclear, chloroplast, or mitochondrial genomes, and 0.04% matched snoRNA sequence (Table 2.1).

Functional small RNAs tend to have characteristic size and initial nucleotide distributions (Reinhart et al. 2002; Tang et al. 2003; Xie et al. 2004; Gasciolli et al. 2005). The small RNAs corresponding to rRNAs were distributed in size, and they tended to have a guanine or cytosine 5′ nucleotide more frequently than an adenine or a uridine (Fig 2.2). None of the snoRNA-derived small RNAs was sequenced more than 24 times, and the negligible number of reads matching to snoRNAs rendered insignificant the slight bias we saw towards 17-19mer snoRNA products. However, the tRNA-derived small RNAs did display a bias in sequence characteristics in that they were highly enriched overall for sequences beginning with a 5′ guanine. In addition there were twice as many 16mers as 19mers, the next most abundant size class. Of the 71 unique sequences with a 5′ guanine nucleotide in the 16mer class, 9 were sequenced 100 times or more, contributing the bulk (91%) of the 16mer 5′ guanine peak. 8 of these 9 sequences corresponded to the extreme 5′ end of nuclear Val, Gly, Ala or Phe tRNAs and we suspect that they represent the 5′ leader which is cleaved from pre-tRNAs by RNAse P. The ninth sequence was an internal sequence common to all four tRNA-Thr loci in the genome. Nevertheless, the relatively uniform distribution of sizes and initial nucleotide composition of reads matching the rRNAs, tRNAs and snoRNAs (Fig 2.2I-N) suggested that most if not all of these small RNAs were degradation products; they were not analyzed further.

After subtracting the organellar rRNA and tRNA matches, there were still 3,474 unique sequences (represented by 13,351 reads) that did not hit the nuclear genome but mapped to either the plastid or mitochondrial genomes, or to both (Table 2.1). The only striking pattern in the set of sequences derived from mitochondria and chloroplast was an

abundance of 27nt sequences beginning with a cytosine residue and a peak at 22nt for sequences beginning with an adenine nucleotide, each owing to a single highly abundant small RNA species. Although we expect artifacts of non-enzymatic degradation and RNA turnover products to be randomly produced and therefore maintain a low probability of sequence redundancy within the library, some organellar small RNAs not corresponding to rRNA or tRNA were sequenced very frequently. These may have been generated during cellular recycling of highly abundant mRNAs. Generally, however, as observed for the structural RNAs, the distribution of sizes and 5′ nucleotides of these reads was roughly uniform across all other lengths, suggesting that these were also degradation fragments of mitochondrial and plastid mRNAs (Fig 2.2M).

A significant proportion of both organellar genomes is duplicated in the nuclear genome; perhaps most notable is a massive insertion in the genetically-defined centromere of Chromosome 2, comprising ~75% of the mitochondrial genome (Lin et al. 1999). After filtering out small RNAs matching rRNA or tRNA, we found that a total of 2092 sequences represented by 3308 reads mapped to either or both of the organellar genomes in addition to the nuclear genome. The majority (79%) of these reads matched mitochondrial sequence. Because of the centromeric context of the mitochondrial genome insertion, and because some nuclear paralogs of organelle genes might be transcriptionally silenced with the participation of heterochromatic siRNAs, we included any organellar, non-rRNA, non-tRNA sequences that also matched the nuclear genome in the set of candidate silencing RNAs.

Other endogenous siRNAs in *Arabidopsis* mapped predominantly to intergenic regions

After removing the RNAs that corresponded to the sense strand of ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) (Table 2.1) and those matching previously annotated and newly identified miRNAs and tasiRNAs, we considered the remaining RNAs that matched the nuclear genome. These included a majority (63%) of the reads and a large majority (90%) of the unique sequences (Table 2.1). Of the reads that did not match non-coding RNA transcripts, only 10% corresponded solely to annotated mRNAs or introns, which

represented substantial depletion when considering that 49% of the sequenced genome is annotated as mRNA and intron. Of those that hit annotated mRNAs or introns, about 46% were exclusively in the antisense orientation to a protein-coding gene. The length and 5′-nucleotide profiles of small RNAs mapping exclusively to the sense strand of genes closely resembled that of small RNAs mapping exclusively to the antisense strand of genes (Fig 2.2G-H), suggesting that a majority of the sense as well as antisense reads might be siRNAs. We considered them, together with the other small RNAs that did not match non-coding RNA transcripts, as endogenous siRNA candidates.

The candidate siRNAs included 20,720 reads that mapped to the antisense of rRNAs, or to ribosomal DNA-like (rDNA-like) repeats but not to the mature rRNA sequences (Table 2.1). These are likely to include *bona fide* siRNAs acting by targeting the rDNA arrays for chromatin or histone modifications (Xie et al. 2004; Pontier et al. 2005; Li et al. 2006; Pontes et al. 2006). Because the fraction of the genome comprised by the rDNA arrays as well as their copy number was unknown, and because much of the rDNA sequence was missing from the current assembly, it was difficult to determine if this represented an enrichment.

The candidate siRNAs were mostly 24mers, the size of siRNAs associated with PolIV, heterochromatin formation and DNA methylation (Chan et al. 2004; Xie et al. 2004; Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). As expected based on initial sequencing efforts (Tang et al. 2003), the 24mers were enriched for a 5′-terminal adenosine (Fig 2.2E-F). As for the tasiRNAs, a tendency to adhere to the pairing asymmetry guidelines for silencing complex assembly was observed only in a naïve analysis that did not consider the presumably independent preference for a particular nucleotide at the 5′ termini of the siRNA reads. When correcting for the preference for an A at the 5′ terminus of the candidate siRNAs, the strand with less stable pairing at its 5′ terminus was sequenced no more frequently than expected by chance.

Small RNAs matching annotated protein-coding genes

Some protein-coding genes had a particularly high propensity for spawning small RNAs (Supplementary Table 2.1). Eleven of the 20 genes most frequently hit, when read counts were normalized by the number of genomic hits and assigned only when unambiguously

sense or antisense to genes, were convergently transcribed with a neighboring gene. These included an antisense gene pair, At2g16580/At2g16575, ranked 16th and 19th. Both genes have ORFs with unknown functions, with one ORF falling largely within the intron of the other. Convergent, overlapping transcription presumably generated dsRNA from which the small RNAs were derived. For the nine remaining genes in a convergent context, the 3′ termini were either uncharacterized or had non-overlapping annotation.

RdRPs provide another mechanism for generating dsRNA. A search for reads that matched a cDNA database but failed to match the genome found 32 reads that spanned mRNA splice junctions in the antisense orientation (Supplementary Table 2.2). Such reads provided evidence for siRNAs generated by RdRP acting on a spliced mRNA template. The two cDNAs with the most non-overlapping antisense hits to splice junctions encoded a TIR-NBS-LRR disease-resistance protein (At5g38850) and a basic helix-loop-helix protein (At3g23690). Both were among the top 20 genes hit by small RNAs (Supplementary Table 2.1). Neither had hits to introns, indicating that for these two genes the RdRP activity acted primarily on spliced templates. However, many of the genes frequently hit by small RNAs had hits to introns. Moreover, many more small RNAs matched the sense strand of mRNA splice junctions (641) than matched the antisense (Supplementary Table 2.2). The 20-fold difference between sense and antisense reads to splice junctions, compared to the nearly even numbers of sense and antisense reads matching protein-coding genes more generally, suggests that if RdRPs play a major role in producing siRNAs from protein-coding regions then the templates are usually unspliced transcripts.

Candidate siRNAs derived preferentially from pericentromeric regions, with a slight preference for repeats

When candidate siRNA reads were mapped to each chromosome, plotting sequencing abundance normalized by the number of genome matches (Fig 2.3), the largest peak included mostly 22mers and corresponded to a cluster of Gypsy and MuDr elements on the long arm of Chromosome 3. The next two largest peaks corresponded to two rDNA arrays with neighboring repetitive elements on Chromosome 3 and

Chromosome 2. Most of the remaining small RNAs were 24mers that mapped to numerous loci dispersed throughout the genome (Fig 2.3).

Though the *Arabidopsis* genome is relatively compact, repetitive loci are abundant. They are most dense at and near the centromeres, and their density gradually tapers off in the 2–3 Mb on both sides of each centromere as protein-coding density increases (Fig 2.3) Along the remainder of the chromosomes, repeats are present but occur at much lower density. The siRNA density did not peak at the same regions as the repeat density peaked and was instead greatest in the proximal and distal pericentromeric regions, characterized by an intermediate density of both repeats and annotated protein-coding genes (Fig 2.3). A look at a diagnostic centromeric repeat class, the ~180 bp repeat satellite arrays (Copenhaver et al. 1999; Nagaki et al. 2003), illustrated this result. Only 2386 reads (1055 unique sequences) matched ~180-bp centromeric satellite repeats annotated by RepeatMasker. This was 0.43% of our reads, whereas the annotated ~180-bp repeat represented 0.39% of the current genome assembly. Because many ~180-bp repeats are missing from the assembly, this slight apparent enrichment was undoubtedly an overestimate; siRNAs deriving from unassembled repeats would artifactually add to the perceived density at any assembled repeats that they match. The unremarkable correspondence between candidate siRNAs and known heterochromatin was also illustrated at the heterochromatic knobs, which were rich in candidate siRNAs, but not more enriched than were the pericentromeric regions that surrounded them (Fig 2.3).

The observation that siRNAs were often associated with repeats, but were not highest where repeats were most dense, raised the question of whether siRNAs derived preferentially from repeat loci. Transposons, retroelements, and low-complexity sequences identified by RepeatMasker comprised ~15% of the current genome assembly. Of the 558,481 candidate siRNA reads, 188,502 (34%) hit these regions annotated by RepeatMasker—a modest, two-fold enrichment over the 15% that would have been expected if the siRNAs derived uniformly from repetitive and non-repetitive regions throughout the genome. The two-fold enrichment was largely attributed to the depletion of both repeats and siRNA matches within annotated protein-coding genes. Of the 51% of the genome that fell between annotated protein-coding genes, ~30% corresponded to repeats annotated by RepeatMasker. Of the 491,180 siRNAs mapping between annotated

protein-coding genes 188,502 (38%) hit regions annotated by RepeatMasker, indicating only a 1.2-fold preference for repeat regions within intergenic regions. The modest preference for repeat regions decreased further when excluding the 20,720 reads deriving from rDNA repeats. When considering only those RNAs associated with AGO4 (Qi et al., 2006), this slight enrichment increased, but not by much. Local (<100 kb) inverted-repeat regions did not appear to be over-represented among genomic hits of intergenic siRNA candidates that fell outside of repetitive elements identified by RepeatMasker. Of the repetitive DNA detected by RepeatMasker, 80% was either class I (retrotransposon-derived) or class II (DNA transposon-derived). About 95% of the repeat-associated siRNAs corresponded to these two classes, in the proportion expected based on the contribution of these two classes to the genome. Representation of some of the more well-characterized transposable element families is listed (Table 2.4).

Small RNA Hotspots Corresponded to Unannotated Genomic Regions

Some intergenic loci had a high propensity to give rise to candidate siRNAs. To supplement the low-resolution analysis (Fig 2.3), we performed a higher-resolution search for such siRNA hotspots and then surveyed the annotations corresponding to the top 20, which ranged in length from 0.5 –50 kb. Although most were in the vicinity of mobile elements or low-complexity sequence, only one hotspot had a transposon at the densest region of siRNAs. Two of the top 20 were very near centromeres, and half were in pericentromeric regions, within 4 Mb of the centromeres. One hotspot, ranked twelfth, corresponded to the 5S rDNA array on chromosome 2. Although the topmost ranked hotspot corresponded to a predicted but unlikely ORF, the other highly ranked hotspots were typically lacking in annotated features within the region producing the majority of small RNAs and represented uncharacterized intergenic regions (Fig 2.4). A preference for being in local (<100 kb) inverted repeats was not found among the top 20 hotspots, but three lower-ranking loci (ranking 27, 36, and 37) were found in an inverted context. Nine of the top 20 hotspots were in a convergent context with regard to flanking annotated genes. This was higher than might have been expected if convergent, non-convergent and divergent contexts were randomly distributed with respect to siRNA-

generating loci. However, without mapped transcripts for these convergent flanking genes, the mechanism for siRNA production remains to be elucidated.

DISCUSSION

Endogenous siRNAs in *Arabidopsis*

Perhaps the most surprising property of the candidate siRNAs was their underwhelming tendency to derive from repeat loci. Of course, RepeatMasker is limited to the identification of repetitive DNA with detectable homology to known repetitive element families, and cannot recognize genomic regions corresponding to novel transposable elements. Although some candidate siRNAs that did not match annotated repeats had multiple genomic matches, suggesting that they might derive from uncharacterized repeats (Table 2.1), most had only one hit. Moreover, unknown repeats would substantially increase the 1.2-fold enrichment found in intergenic regions only in the unlikely event that the repeats not yet identified were a far richer source of siRNAs than were known repeats.

Part of the reason that repeats generally were not a more rich source of siRNAs was that the regions within and immediately flanking the centromeres, which are mostly annotated repeats, were somewhat depleted in siRNAs when compared to the more distal pericentromeric regions, which had only an intermediate density of repeats (Fig 2.3). This observation differed from the report that siRNA density closely mirrors repeat density (Lu et al. 2005). We attribute this apparent contradiction to our normalization of read counts based on the number of times the sequence hit the genome assembly. That is, if a sequence with two reads hit the genome 200 times, we assigned one hundredth of a count to each locus, rather than two counts to each locus. Our approach attempted to reflect both the fact that a given molecule cannot arise simultaneously from more than one locus, and recent results showing that heterochromatic siRNAs act preferentially at the locus of origin (Buhler et al. 2006), while at the same time leaving ambiguous which repeat locus gave rise to a particular siRNA molecule. Our finding that siRNAs were less abundant at the centromeres, compared with the pericentromeric regions, was reminiscent of the heterochromatic siRNAs of *Schizosaccharomyces pombe*, which map to the

heterochromatic outer repeats of the centromeres but not to centromere cores (Reinhart and Bartel 2002). We suggest that heterochromatic siRNAs might function primarily near the boundaries of heterochromatin and euchromatin and play less of a role within large stretches of heterochromatin at the centromeres.

Pairing asymmetry, known to influence incorporation of miRNAs and synthetic siRNAs into silencing complexes (Khvorova et al. 2003; Schwarz et al. 2003), had no detectable correlation with accumulation of tasiRNAs and other *Arabidopsis* siRNA candidates. The same was found when we analyzed (data not shown) a set of *Arabidopsis* transgene siRNAs previously reported to follow the guidelines (Khvorova et al. 2003). Therefore, for no known cases in animals or plants do endogenously expressed siRNAs preferentially follow the asymmetry guidelines. One explanation might be that most siRNAs in the cell are in the duplex configuration and have not been loaded into the silencing complex. However, no preference was observed when repeating the analysis with a recently reported set of Ago4-associated siRNAs. Therefore, we favor the notion that for most if not all classes of endogenous siRNAs, pairing asymmetry plays little or no role in deciding which strand of the duplex serves as the guide strand. MicroRNAs are a different story; in every plant or animal species examined, miRNA accumulation tends to follow the asymmetry guidelines, even after accounting for their propensity to begin with a U (data not shown). In mammals, synthetic siRNAs also obey the asymmetry guidelines, presumably because vertebrate cells recognize and utilize a synthetic siRNA duplex as if it were an endogenous miRNA duplex. Perhaps more important than pairing asymmetry for endogenous siRNAs is the identity of the 5′ residue. For plant 24mer siRNAs, a 5′ terminal A may favor incorporation or stabilization within the silencing complex, whereas for tasiRNAs, a 5′ U may do the same. The identity of the 5′ nucleotide might also influence the incorporation or stability of miRNAs, which would help explain the strong preferences for U over A and C over G observed at their 5′ termini in all species.

MATERIALS AND METHODS

Libraries and Sequencing

Wild-type *Arabidopsis* (Columbia accession) plants were grown under standard greenhouse conditions, except seedlings, which were grown as in Reinhart et al. (2002). Total RNA was extracted (Mallory et al. 2001) from whole seedlings, flowers, rosette leaves, and siliques, harvested six days, four weeks, six weeks, and two months after planting, respectively. Small RNA cDNA libraries were prepared for standard sequencing as in Lau et al. (2001) and for bead-in-well pyrophosphate sequencing as in Axtell et al. (2006). Pyrosequencing was performed at 454 Life Sciences (Branford, CT, USA).

Initial processing of reads

cDNA sequences were extracted from raw reads, excluding reads lacking perfect matches to the most proximal 11 nt of both adapter sequences. Unique sequences were mapped to the TAIR/NCBI genome Version 6.0 (November 2005), chloroplast and mitochondrial genomes, and rRNA and tRNA sequences from TAIR (www.arabidopsis.org), the 5S rRNA Database (http://biobases.ibch.poznan.pl/5SData/), the published chromosomal rRNA sequences (Unfried et al. 1989, (Unfried and Gruendler 1990), the *Arabidopsis* tRNA database (http://lowelab.ucsc.edu/GtRNAdb/Athal/), the *Arabidopsis* snoRNA database (http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/arabidopsis, Brown et al. 2001), miRBase 7.1, and the TIGR v2 *Arabidopsis* Repeat Database. Repetitive regions of the genome were mapped by RepeatMasker (http://www.repeatmasker.org) using AtRepbase (Jurka et al. 2005).

Phased siRNA discovery

For each unique small RNA sequence (excluding those matching miRNAs, other noncoding RNAs, or protein-coding genes) a 500-nt window, anchored at one end by that sequence, was evaluated for phased small RNAs. If three or more unique 20- to 23-nt sequences with non-overlapping hits existed in the window, each was evaluated for phasing with any of the others in the window, allowing ± 2 nt of divergence from perfect 21-nt phasing. Phased sequences were extracted from the window and the process was

repeated for any remaining 20- to 23-nt sequences until two or fewer unique non-overlapping hits were left. Each potential phase was then evaluated according to five parameters: a) a count of all unique sequences in phase b) a count of the reads in the window c) a hits-normalized score, whereby the sum of the read frequencies of all phased sequences was divided by the sum of their genomic hits d) a normalized 21mer score, which divided the sum of the 21mer reads in phase by the sum of the reads in the window and e) a phasing score, which divided the sum of the reads in phase by the sum of the reads in the window. Cutoffs for each score were empirically adjusted to find values that captured all known tasiRNA clusters but restricted the number of false positives.

Target site prediction for TAS4 siRNAs

Patscan was used to search for near matches (up to six mismatches, or four mismatches and one bulged nucleotide) in TAIR Version 6.0 *Arabidopsis* cDNA database (www.arabidopsis.org) to each tasiRNA, and target sites were scored as described (Allen et al. 2005).

Hotspot identification

For candidate siRNAs, non-overlapping 500-bp windows were ranked by scoring small-RNA density as a sum of abundances of all sequences in the window, normalized by the sum of the total number of times each sequence hit the genome. Top-ranking windows were then used as seeds for extension in both directions until a 500-bp window lacking any siRNA hits was encountered.

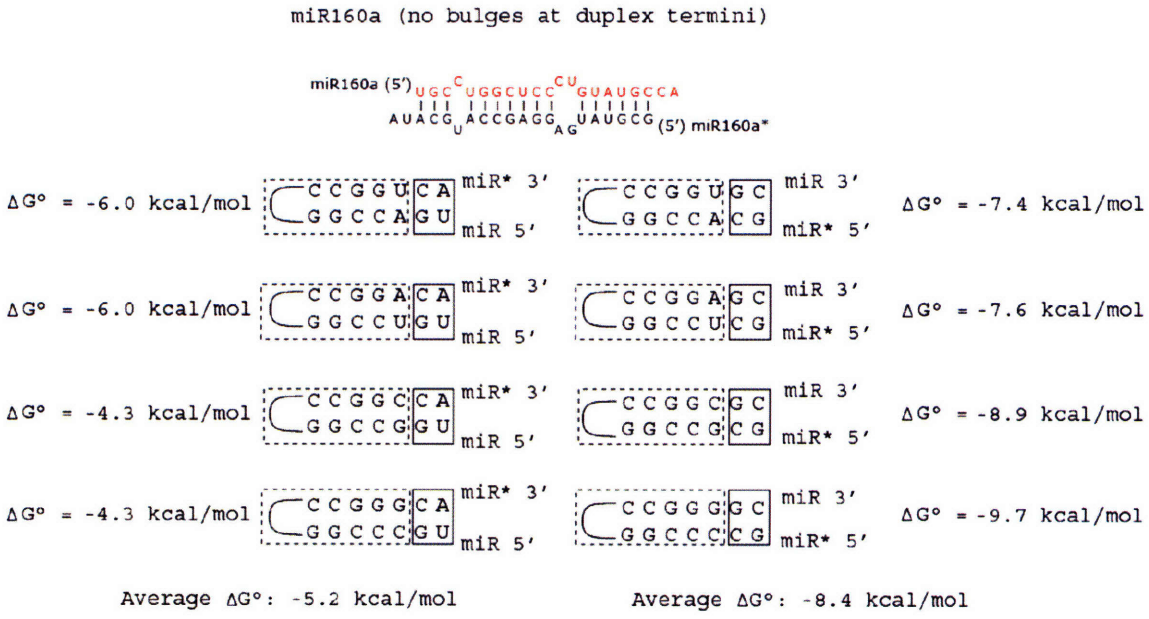
Analysis of siRNA and miRNA pairing asymmetry

For analysis of tasiRNA and candidate siRNA pairing asymmetry, conceptual duplexes with 2-nt 3′ overhangs were constructed by determining the reverse complement of the sequenced strand. The terminal three pairs (two nearest neighbors) on each end of the duplex were analyzed, comparing sums of the two $-\Delta G°_{37}$ nearest-neighbor parameters for RNA duplex stability (Xia et al. 1998).

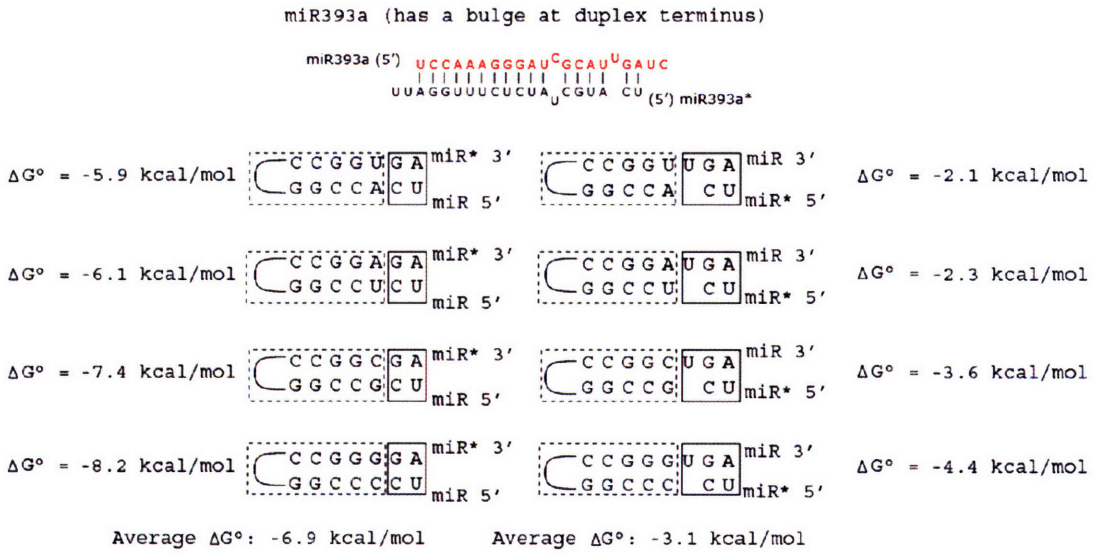
More complex algorithms were also implemented and yielded similar conclusions.

Evaluating the terminal four base pairs (three nearest neighbors) or terminal two base pairs (one nearest neighbor) gave similar results as the algorithm described above (which considered two nearest neighbors). Reasoning that the terminal basepair stability would contribute more than successively interior basepairs, we designed variants of the algorithm that gave increasing weight to the more terminal pairs. However, after controlling for the preference of the siRNA 5' nucleotide, no influence of pairing asymmetry was detected.

To assess the asymmetry of pairing at the ends of miRNA-miRNA* duplexes, we considered the pairing of all previously annotated miRNA hairpins whose mature miRNA sequence was represented among our sequencing reads. Analysis of these termini was more complex than that of siRNAs because the miRNA-miRNA* duplexes contained mismatches and bulges in addition to Watson-Crick pairs. Therefore, we appended the termini to model hairpins, and then used RNAfold (Hofacker et al. 1994) to evaluate the relative pairing stability at each terminus of the "hairpin" using four variants of a linker sequence, comparing the average free energies as exemplified below (the dashed box indicates the linker sequence, and the solid box indicates the terminus of the duplex with the specified ends):



For duplexes with a bulged nucleotide on either strand, the bulged nucleotide was included, allowing RNAfold to decide how best to minimize its energetic contribution, as shown for miR158b below:



The terminus of the duplex with higher average predicted ΔG° was deemed to have less stable pairing; for miR160a, the terminus specified by the 5′ end of the miRNA was destabilized, while for miR393a, the terminus specified by the 3′ end of the miRNA was destabilized.

When the miRNA loci were scored, 82% were found to observe the asymmetry guidelines. To control for the preference for a U or an A at the 5' end of miRNAs, the analysis was repeated for only those hairpins for which both the miRNA and the miRNA* began with an A or a U. The bias was still observed for these miRNAs; 67%, of the duplexes still obeyed the asymmetry guidelines. This percentage was greater than the 50% expected by chance if duplex asymmetry had no role in strand selection, suggesting that the asymmetry rules for incorporation play a role in determining which strand of the miRNA:miRNA* duplex becomes loaded in the RISC complex in plants.

5′ RACE

5′ RACE was performed as described in Jones-Rhoades and Bartel (2004), except that RNA samples were obtained from whole siliques or seedlings, gene-specific primers

(Supplemental Text) were designed to be 70 to 400 bases from the predicted cleavage site, and the first gene-specific amplifications for *PAP2*, and *MYB113* were done with the GeneRacer 5′ outer primer and were followed by two nested amplifications done with the GeneRacer 5′ nested primer.

The following 5′ RACE gene-specific primers were used to map cleavage sites:

Transcript	Primer	Gene-specific primer sequence
At1g66370 (*MYB113*)	Outer 1st Nested 2nd Nested	GTCTCTTCATCAAACCGAGCCCAAA CCGTAGCTTCTGGACCCAACACATC GGAGAAGGATCGAGGCCGAGGCTTTA
At1g66390 (*PAP2*)	Outer 1st Nested 2nd Nested	CAAACGCCAAAGTGGCCCCATGT CAGCTGTCGTCGCTTCAGGAACAATCG CATCAGCTTCTTGGTTTTCCCCCAGT
TAS4	Outer Nested	GAAGGATCGAGGTCGAGGCACCA CGTCCTTCACCACGGCAATTTCATGC

Accession Numbers

All genome-matched small RNA sequences generated in this study are accessible at http://www.ncbi.nlm.nih.gov/geo/ as Platform GPL3968, Samples GSM118372, GSM118373, GSM118374, and GSM118375, and Series GSE5228.

All genomic loci matching the small RNAs in this study can be accessed at: http://www.genesdev.org/cgi/content/full/20/24/3407/DC1/SupplTable5.zip or browsed online at the *Arabidopsis* Small RNA Project database: http://asrp.cgrb.oregonstate.edu/cgi-bin/gbrowse/thaliana-v5

ACKNOWLEDGMENTS

We thank M. Axtell for many helpful discussions, G. Bell for computational advice, A. Hall and J. Jurek (Preuss Lab) for pilot analysis of centromeric candidate siRNAs, and M. Jones-Rhoades and G. Ruby for scripts. Supported by the Prix Louis D. award from the Institut de France and a grant from the N.I.H. to D.B.

Table 2.1. Summary statistics of small RNAs sequenced from *Arabidopsis*

Locus Class	Unique sequences		Reads		Mean frequency[a]	Mean hits[b]
Silencing RNAs						
miRBase annotated miRNA hairpin[c]	960	(0.3%)	138416	(15.6%)	144	1.6
Newly identified miRNA hairpin[c]	361	(0.11%)	7002	(0.8%)	19.4	1.0
Trans-acting siRNA locus	1366	(0.4%)	10358	(1.2%)	7.6	1.0
Newly identified tasiRNA	34	(0.01%)	111	(0.01%)	3.3	1.0
Candidate silencing RNAs						
rDNA	5318	(1.6%)	20720	(2.3%)	3.9	24.8
Protein-coding genes						
Sense	18626	(5.5%)	31475	(3.5%)	1.7	1.2
Antisense	13196	(3.9%)	23441	(2.6%)	1.8	1.2
Sense and antisense	728	(0.2%)	1191	(0.1%)	1.6	1.5
Annotated repeat and mobile elements	111345	(32.7%)	188502	(21.2%)	1.7	16.5
Other nuclear genomic	156407	(46.0%)	293152	(33.0%)	1.9	1.7
Non-protein coding RNAs[d]						
snoRNA	199	(0.06%)	368	(0.04%)	1.8	1.3
snRNA	237	(0.1%)	390	(0.04%)	1.7	5.8
tRNA	2224	(0.7%)	24606	(2.8%)	11.1	8.5
rRNA	25639	(7.5%)	134183	(15.1%)	5.2	3.9
Organellar small RNAs						
Mitochondrial/chloroplast[e]	3474	(1.0%)	13351	(1.5%)	3.8	n/a
TOTAL	340114		887266			

(n/a) Not applicable
[a] Average number of reads per unique sequence
[b] Average number of hits to the nuclear genome
[c] Includes all sequences and reads that mapped to predicted miRNA hairpin precursors
[d] Matching mature sense strand
[e] Does not include rRNA or tRNA

Table 2.2. Small RNAs deriving from mobile elements

Mobile element family	% of genome	Reads
DNA/MuDR	2.41	44407.2
LTR/Gypsy	5.12	38045.4
RC/Helitron	1.67	30799.2
LINE/L1	0.97	17058.5
SINE	0.05	13234.9
LTR/Copia	1.41	12990.0
DNA/hAT	0.38	8708.3
DNA	0.33	4997.5
DNA/En-Spm	0.82	3889.3
DNA/Harbinger	0.14	2903.7
DNA/Pogo	0.11	979.2
DNA/Mariner	0.04	168.0
DNA/hAT-Ac	0.01	57.0
DNA/Tc1	0.02	26.4

Listed are the major classes of transposable elements and the hits-normalized sequence abundance of reads corresponding to each.

FIGURE LEGENDS

Figure 2.1. The *TAS4* locus gives rise to tasiRNAs predicted to down-regulate *MYB* transcripts. (*A*) The number of reads with a 5′ terminus at each position is plotted. Bars above the axis represent sense reads; those below represent antisense reads. The miR828 complementary site is marked by a red arrow and shown below the graph, together with the fraction of 5′ RACE clones supporting the indicated cleavage site. *TAS4*-siR81(-), the siRNA predicted to target MYB genes, is indicated (blue bar), as is the spacing separating the phased species at each interval; spacing for the species not represented by reads is indicated in grey. (*B*) *TAS4*-siR81(-) and miR828 complementary sites in three *MYB* genes. Cleavage confirmed by 5′ RACE is indicated (arrows), along with the fraction of clones mapping to the cleavage site. The remaining 14 *PAP2* clones mapped more than 20 nt from the cleavage site.

Figure 2.2. Size distribution and 5′-nt identity of *Arabidopsis* small RNAs classes.

Figure 2.3. Normalized abundance of candidate siRNAs in 0.1 Mb windows spanning the nuclear genome. Colored bars above the axis represent matches to the plus strand; colored bars below the axis represent those to the minus strand, with the colors indicating the proportion of 21mers (red), 24mers (light and dark blue), 24mers with a 5′ A (light blue), and other lengths (yellow). Below the siRNA profiles are histograms plotting the fraction of nucleotides falling within annotated protein-coding genes (black; scale, 0 – 100%) and the fraction falling within repetitive elements annotated by RepeatMasker (gray; scale, 0 – 100%). Centromeres are indicated by solid grey bars, heterochromatic knobs by hashed grey bars.

Figure 2.4. Two high-ranking siRNA hotspots, with neighboring genomic annotations.
(A) Normalized abundance of siRNAs in the highest ranking hotspot.
(B) Normalized abundance of siRNAs in the third highest ranking hotspot. Colors and bars are as in Figure 6.

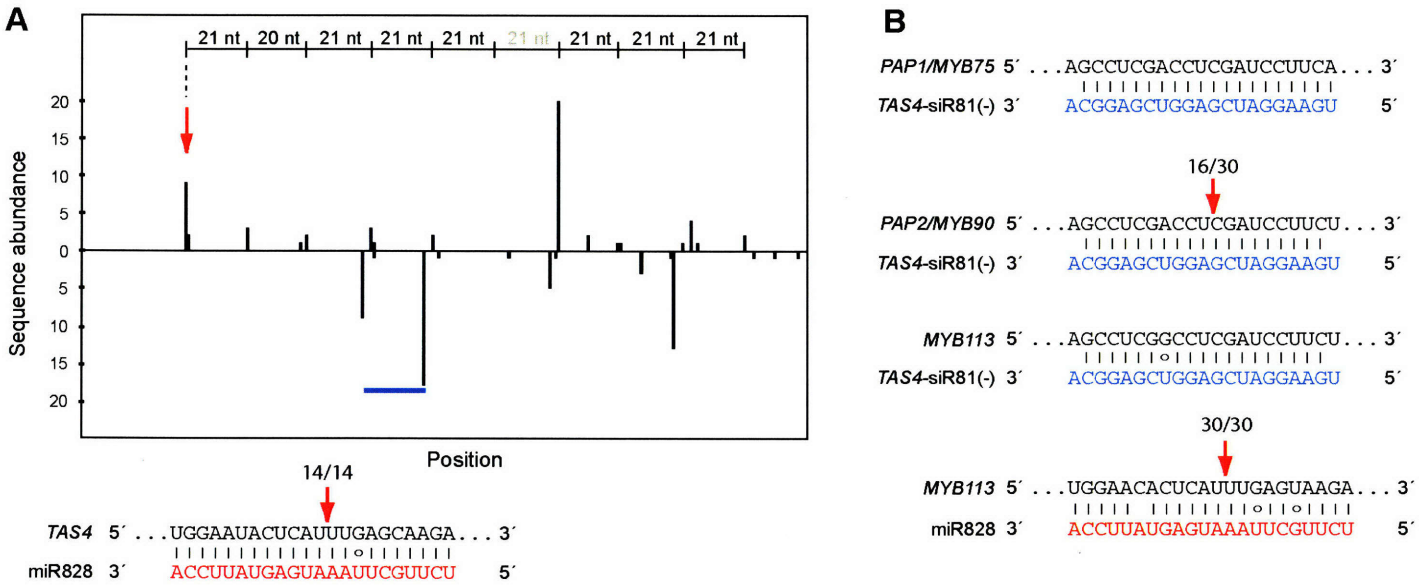


Figure 2.1. The *TAS4* locus gives rise to tasiRNAs predicted to down-regulate MYB transcripts.

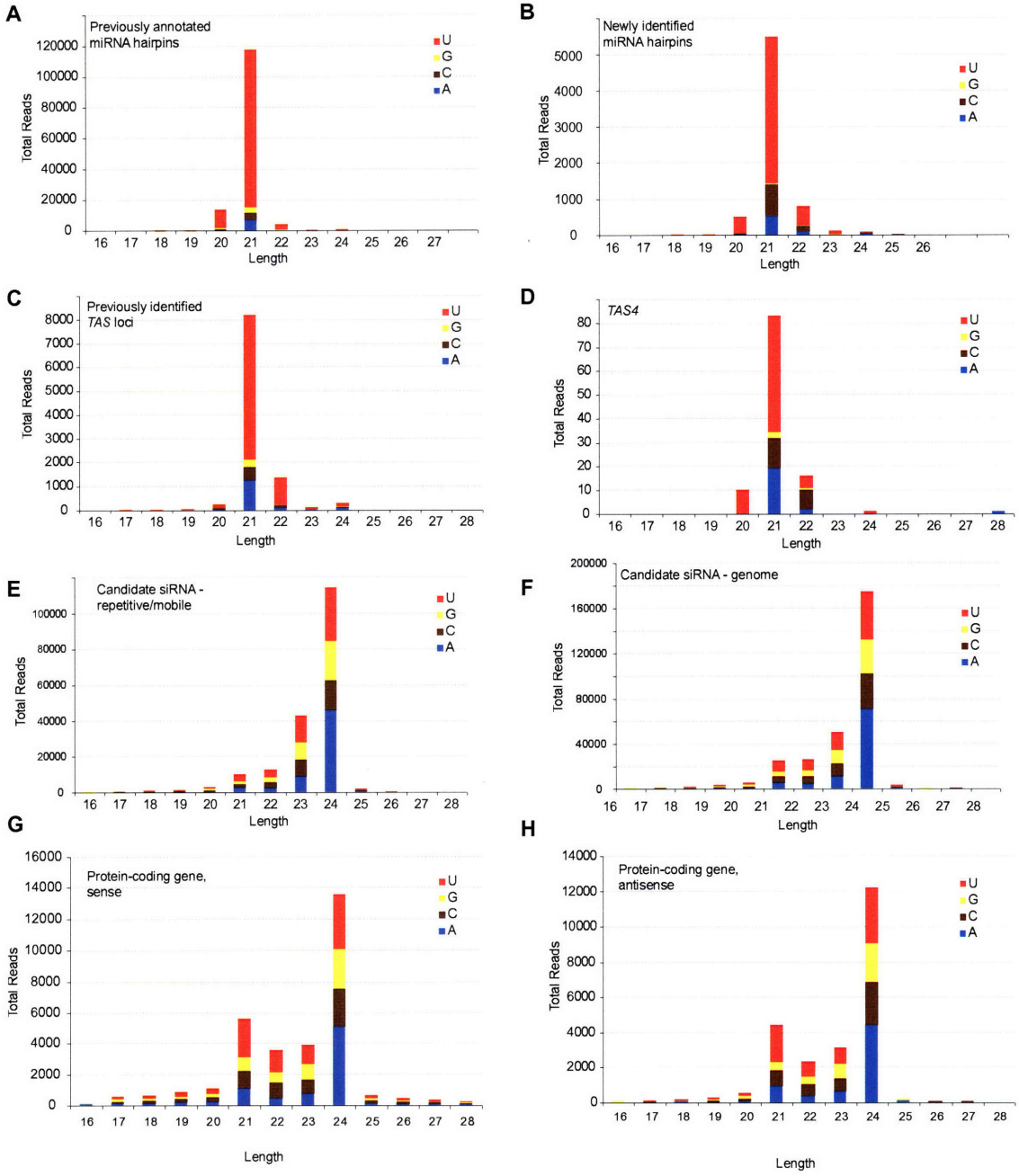


Figure 2.2. Size distribution and 5′-nt identity of *Arabidopsis* small RNA classes.

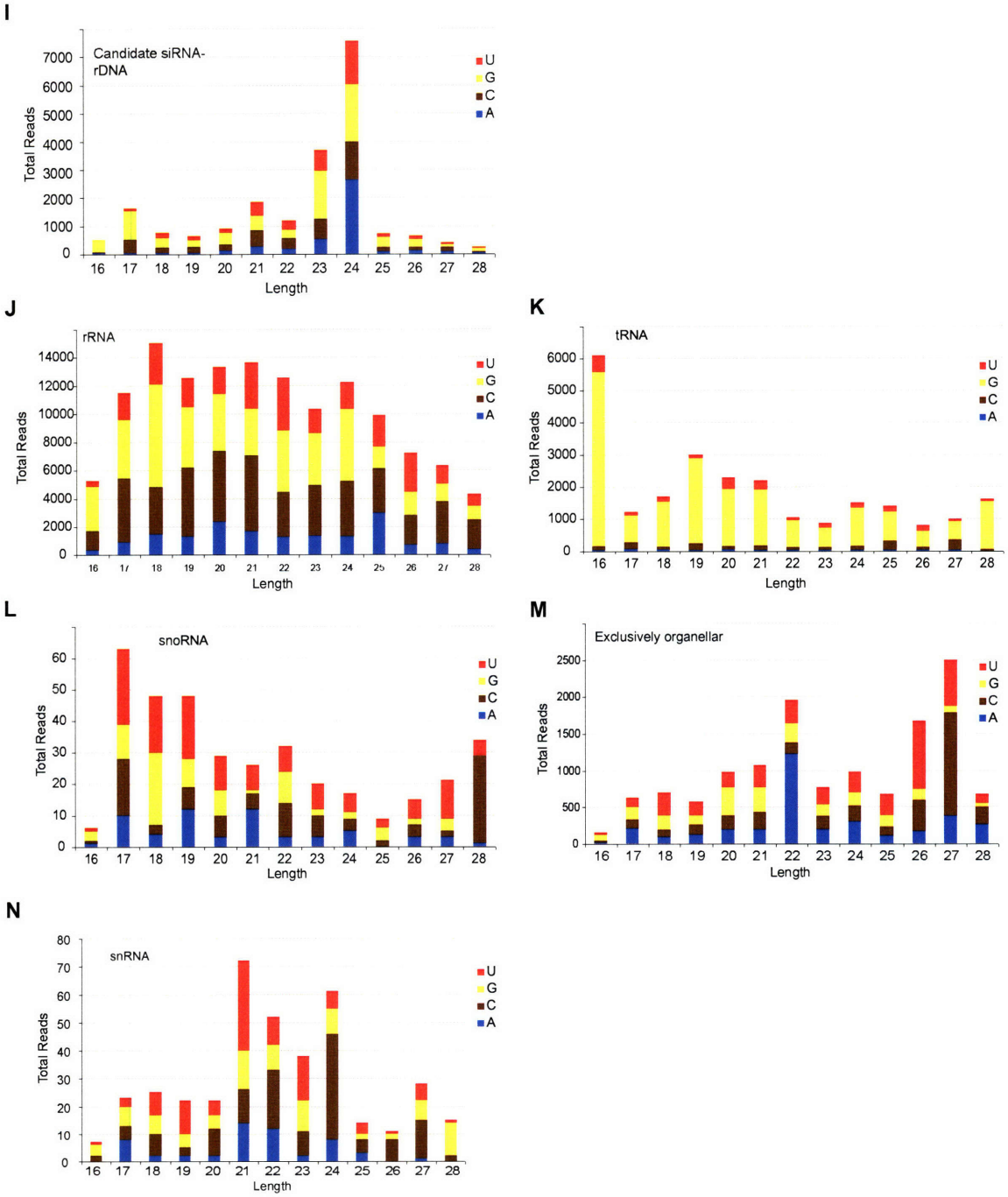


Figure 2.2 cont'd. Size distribution and 5′-nt identity of Arabidopsis small RNA classes.

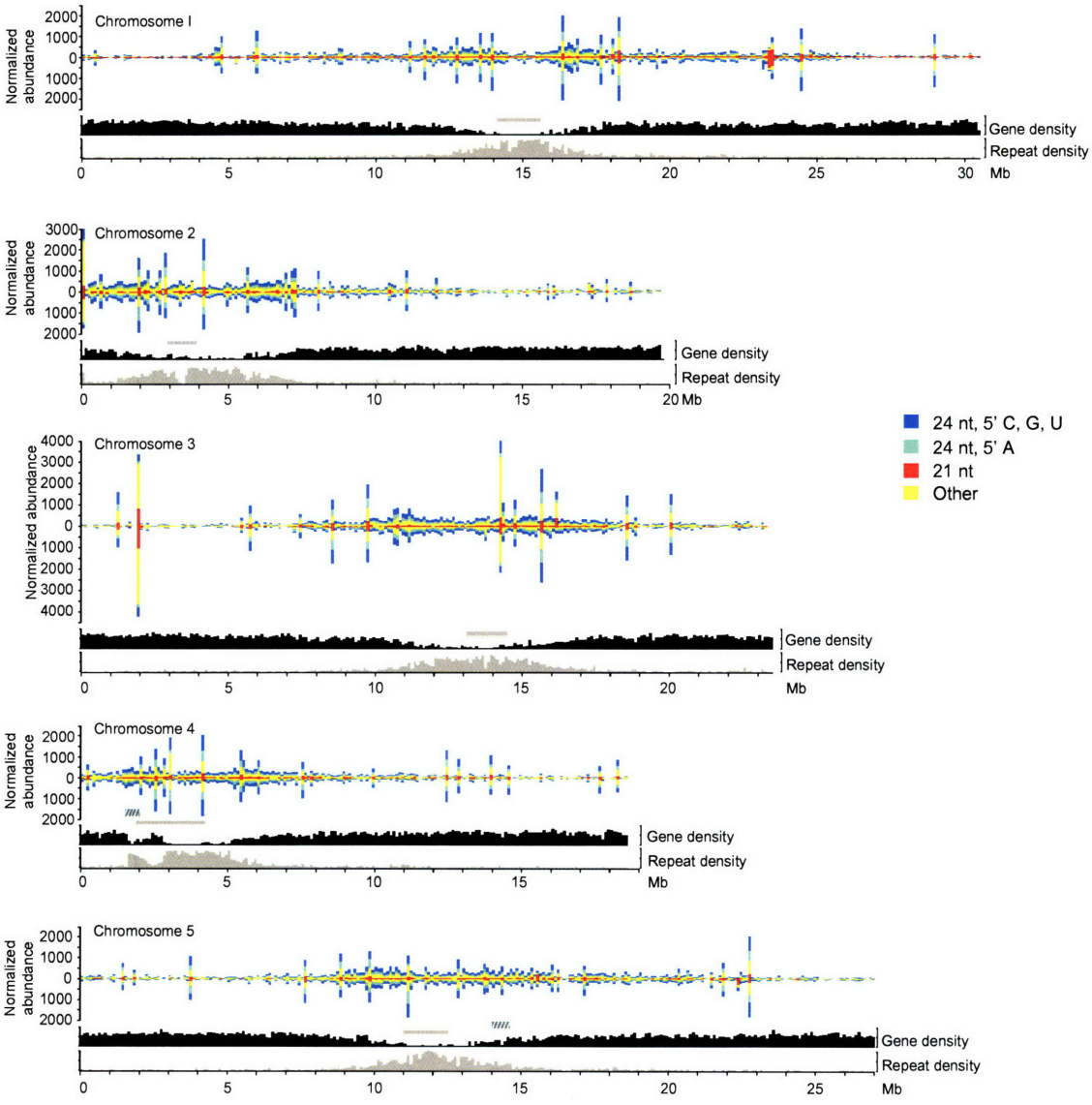


Figure 2.3. Normalized abundance of candidate siRNAs in 0.1 Mb windows spanning the nuclear genome.

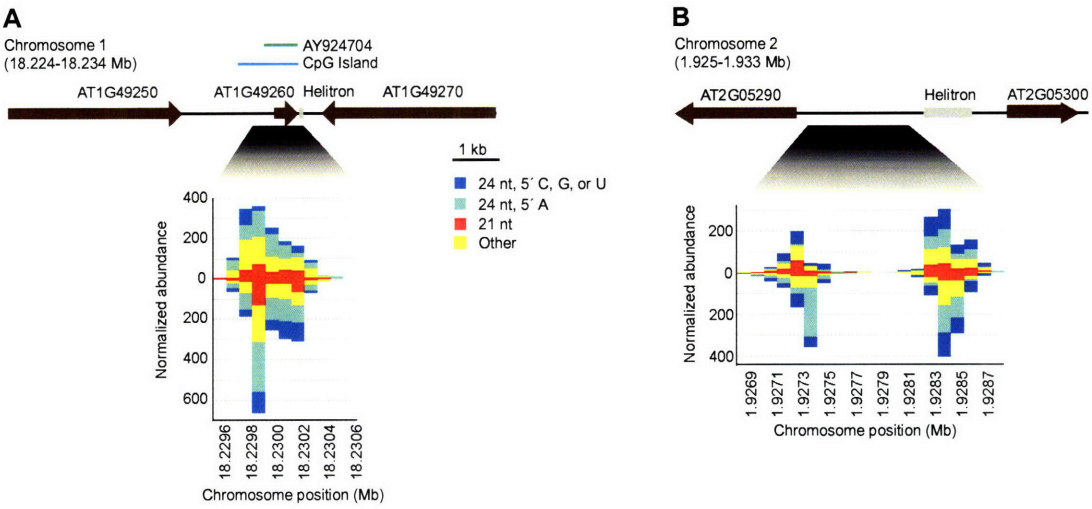


Figure 2.4. Two high-ranking siRNA hotspots, with neighboring genomic annotations.

SUPPLEMENTARY TABLES AND DATABASE

Supplementary Tables 1-2 and Supplementary Database 1 can be accessed at the following WWW links:

Supplementary Table 2.1. Protein-coding genes with a high number of small RNA hits.

http://www.genesdev.org/cgi/content/full/20/24/3407/DC1/Supp Table 2.2.xls

Supplementary Table 2.2. Sequences that did not match the genome but matched the cDNA database.

http://www.genesdev.org/cgi/content/full/20/24/3407/DC1/Supp Table 2.3.xls

Supplementary Database 2.1. Small RNAs corresponding to *Arabidopsis TAS* loci.

http://web.wi.mit.edu/bartel/pub/Supplemental%20Material/Rajagopalan2006/SD1/_SUPPL-DATABASE_3.htm

REFERENCES

Adenot, X., T. Elmayan, D. Lauressergues, S. Boutet, N. Bouche, V. Gasciolli, and H. Vaucheret. 2006. DRB4-dependent *TAS3 trans*-acting siRNAs control leaf morphology through AGO7. *Curr Biol* 16: 927-932.

Allen, E., Z. Xie, A.M. Gustafson, and J.C. Carrington. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121: 207-221.

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

Axtell, M., C. Jan, R. Rajagopalan, and D.P. Bartel. 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* 127: 565-577.

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

Baumberger, N. and D.C. Baulcombe. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* 102: 11928-11933.

Borevitz, J.O., Y. Xia, J. Blount, R.A. Dixon, and C. Lamb. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 12: 2383-2394.

Borsani, O., J. Zhu, P.E. Verslues, R. Sunkar, and J.K. Zhu. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123: 1279-1291.

Buhler, M., A. Verdel, and D. Moazed. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125: 873-886.

Carrington, J.C. and V. Ambros. 2003. Role of microRNAs in plant and animal development. *Science* 301: 336-338.

Chan, S.W., I.R. Henderson, and S.E. Jacobsen. 2005. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 6: 351-360.

Chan, S.W., D. Zilberman, Z. Xie, L.K. Johansen, J.C. Carrington, and S.E. Jacobsen. 2004. RNA silencing genes control *de novo* DNA methylation. *Science* 303: 1336.

Copenhaver, G.P., K. Nickel, T. Kuromori, M.I. Benito, S. Kaul, X. Lin, M. Bevan, G. Murphy, B. Harris, L.D. Parnell, W.R. McCombie, R.A. Martienssen, M. Marra, and D. Preuss. 1999. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286: 2468-2474.

Fahlgren, N., T.A. Montgomery, M.D. Howell, E. Allen, S.K. Dvorak, A.L. Alexander, and J.C. Carrington. 2006. Regulation of *AUXIN RESPONSE FACTOR3* by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr Biol* 16: 939-944.

Gasciolli, V., A.C. Mallory, D.P. Bartel, and H. Vaucheret. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr Biol*.

Henderson, I.R., X. Zhang, C. Lu, L. Johnson, B.C. Meyers, P.J. Green, and S.E. Jacobsen. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* 38: 721-725.

Herr, A.J., M.B. Jensen, T. Dalmay, and D.C. Baulcombe. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* 308: 118-120.

Hofacker, I.L., W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. 1994. Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte fur Chemie* 125: 167-188.

Hunter, C., M.R. Willmann, G. Wu, M. Yoshikawa, M. de la Luz Gutierrez-Nava, and S.R. Poethig. 2006. *Trans*-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* 133: 2973-2981.

Jones-Rhoades, M.W., D.P. Bartel, and B. Bartel. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol.*

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

Kanno, T., B. Huettel, M.F. Mette, W. Aufsatz, E. Jaligot, L. Daxinger, D.P. Kreil, M. Matzke, and A.J. Matzke. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* 37: 761-765.

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

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

Li, C.F., O. Pontes, M. El-Shami, I.R. Henderson, Y.V. Bernatavichute, S.W. Chan, T. Lagrange, C.S. Pikaard, and S.E. Jacobsen. 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* 126: 93-106.

Lin, X., S. Kaul, S. Rounsley, T.P. Shea, M.I. Benito, C.D. Town, C.Y. Fujii, T. Mason, C.L. Bowman, M. Barnstead, T.V. Feldblyum, C.R. Buell, K.A. Ketchum, J. Lee, C.M. Ronning, H.L. Koo, K.S. Moffat, L.A. Cronin, M. Shen, G. Pai, S. Van Aken, L. Umayam, L.J. Tallon, J.E. Gill, M.D. Adams, A.J. Carrera, T.H. Creasy, H.M. Goodman, C.R. Somerville, G.P. Copenhaver, D. Preuss, W.C. Nierman, O. White, J.A. Eisen, S.L. Salzberg, C.M. Fraser, and J.C. Venter. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761-768.

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

Lu, C., K. Kulkarni, F.F. Souret, R. Muthuvalliappan, S.S. Tej, R.S. Poethig, I.R. Henderson, S.E. Jacobsen, W. Wang, P.J. Green, and B.C. Meyers. 2006. MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.*

Lu, C., S.S. Tej, S. Luo, C.D. Haudenschild, B.C. Meyers, and P.J. Green. 2005. Elucidation of the small RNA component of the transcriptome. *Science* 309: 1567-1569.

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

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Nagaki, K., P.B. Talbert, C.X. Zhong, R.K. Dawe, S. Henikoff, and J. Jiang. 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis* thaliana centromeres. *Genetics* 163: 1221-1225.

Onodera, Y., J.R. Haag, T. Ream, P.C. Nunes, O. Pontes, and C.S. Pikaard. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120: 613-622.

Park, W., J. Li, R. Song, J. Messing, and X. Chen. 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., M. Yoshikawa, G. Wu, H.L. Albrecht, and R.S. Poethig. 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.

Pontes, O., C.F. Li, P.C. Nunes, J. Haag, T. Ream, A. Vitins, S.E. Jacobsen, and C.S. Pikaard. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126: 79-92.

Pontier, D., G. Yahubyan, D. Vega, A. Bulski, J. Saez-Vasquez, M.A. Hakimi, S. Lerbs-Mache, V. Colot, and T. Lagrange. 2005. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev* 19: 2030-2040.

Pourtau, N., R. Jennings, E. Pelzer, J. Pallas, and A. Wingler. 2006. Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in *Arabidopsis*. *Planta* 224: 556-568.

Qi, Y., A.M. Denli, and G.J. Hannon. 2005. Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* 19: 421-428.

Qi, Y., X. He, X.J. Wang, O. Kohany, J. Jurka, and G.J. Hannon. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*.

Reinhart, B.J. and D.P. Bartel. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* 297: 1831.

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

Santoro, R. 2005. The silence of the ribosomal RNA genes. *Cell Mol Life Sci.*

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

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

Tang, G., B.J. Reinhart, D.P. Bartel, and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes Dev* 17: 49-63.

Unfried, I. and P. Gruendler. 1990. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Res* 18: 4011.

Vaucheret, H. 2005. MicroRNA-dependent *trans*-acting siRNA production. *Sci STKE* 2005: pe43.

Vaucheret, H., F. Vazquez, P. Crete, and D.P. Bartel. 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., H. Vaucheret, R. Rajagopalan, C. Lepers, V. Gasciolli, A.C. Mallory, J.L. Hilbert, D.P. Bartel, and P. Crete. 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell* 16: 69-79.

Xia, T., J. SantaLucia, Jr., M.E. Burkard, R. Kierzek, S.J. Schroeder, X. Jiao, C. Cox, and D.H. Turner. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37: 14719-14735.

Xie, Z., E. Allen, A. Wilken, and J.C. Carrington. 2005. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 102: 12984-12989.

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

Yoshikawa, M., A. Peragine, M.Y. Park, and R.S. Poethig. 2005. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev* 19: 2164-2175.

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

Chapter 3:

Discussion and future directions

The biological relevance of recently evolved miRNAs

Three characteristics distinguish the microRNAs described in this work. First, they appear to have emerged relatively recently in the *Arabidopsis* genome. This is supported by several lines of evidence. By far the most compelling is the lack of convincing orthologs in other sequenced plant genomes for all but one of the newly identified miRNAs. In addition, at least five of these *MIR* loci are similar in sequence to predicted target genes, consistent with a model for their evolution by gene duplication processes so recent as to preclude sufficient intervening time for sequence divergence of the hairpin. Second, these miRNAs are all derived from single loci in the genome, and not large multigenic families. Third, many of these miRNAs do not seem to have canonical cleavage targets, at least using algorithmic prediction to search for target sites that mimic features of known target sites. Lastly, they are apparently non-abundant in wild-type tissues. We may have sampled a basal level of leaky expression for some miRNAs that are only induced after exposure to a stress or other environmental stimulus, while others might be cell-type specific and only weakly represented in the whole-organ samples from which our libraries are derived. Nevertheless, the inferred expression patterns of this set of miRNAs are quite different from the generally broad expression of the conserved miRNAs (Axtell and Bartel 2005).

The apparent lineage specificity of the miRNAs described in this work may be intimately tied to their low expression. The trend that emerges from this work is that abundance (as assayed by sequencing) does correlate with conservation to other plant species (Chapter 1 and Figure 1.1D). The low abundance of newly described miRNAs in wild-type tissues, as well as many more non-conserved than conserved miRNAs without targets satisfying the canonical pairing requirements for plant miRNA-mediated repression, leaves unresolved the question of the functional roles, if any, for these miRNAs. Certainly they do not immediately suggest a critical role for these miRNAs in plant development. Nevertheless, at least some of the miRNAs appear to be implicated in lineage-specific functions which deserve more *in vivo* study, if only because they might shed some insight on speciation or the various ways in which plants have diversified, adapted and specialized for a particular set of growth conditions or an ecological niche.

Some of these miRNAs may lend themselves to *in vivo* study of their biological consequences, particularly the more abundant ones which can be detected easily by RNA blotting. Generating miR-resistant transgenic plants and miRNA-overexpression plants might reveal their functional roles. Further description of the less abundant miRNAs will be more challenging but mutational phenotypes or induction of some these miRNAs may be unmasked under a serendipitous combination of growth conditions and mutant screens not specifically aimed at elucidating small RNA-mediated silencing functions.

miR823-directed repression of *CMT3*

miR823, which negatively regulates *CMT3*, is most abundant in tissue isolated from young seedlings and siliques, and the 3' cleavage products of *CMT3* messages were detectable in siliques (Chapter 1). The high relative abundance of miR823, compounded with the accumulation of miR823-directed *CMT3* cleavage products in siliques, suggests that this interaction has been selectively stabilized and that miR823 has acquired the necessary promoter elements and other regulatory controls to drive its expression in a tissue-specific manner. CMT3 methylates asymmetric cytosines in DNA at transposable and repetitive elements (Lindroth et al. 2001; Tompa et al. 2002). Since silique tissue is the site of embryogenesis, containing fertilized seeds, presumably this would result in a slight reactivation of transposable elements and other repetitive sequences. It is unclear how the plant might benefit from mobile element reactivation in the embryonic tissue fated to give rise to the next generation. One speculation holds that transposable elements are important evolutionary architects, shaping genomes by enhancing or inhibiting expression of neighboring genes in a manner that is favored by natural selection. Perhaps some minimal amount of transposon hopping is tolerated in developing or mature siliques in order to facilitate genomic evolution. Massive genomic instability may be avoided by the intact methylation activities of DMT1 and DMT2, a hypothesis that agrees well with data confirming the viability and apparently normal developmental phenotypes of *cmt3* null plants for up to 5 generations (Lindroth et al. 2001). Alternatively, CMT3 may negatively impinge on imprinting in the endosperm (the only plant tissue known to be a site for imprinting activity) during embryogenesis, and dampening its expression could be necessary for proper imprinting. Endosperm imprinting takes advantage of the double

fertilization process of plants whereby a fertilized diploid egg cell is nourished by a triploid endosperm consisting of a diploid female cell fertilized independently by another male gamete. Reprogramming of the endosperm is unnecessary since it does not contribute to the next generation (Kinoshita et al. 2004). However, CMT3 does not seem to be involved in epigenetic events at the *FWA* locus, which is reactivated by the 5-methylcytosine DNA glycosylase protein *DEMETER* and is unaffected by mutation of *CMT3* (Kinoshita et al. 2004).

miR824-directed downregulation of *AGL16*

According to the small RNA sequencing data described in this work, miR824 is present at high levels in seedlings (Table 1.2) and weakly expressed in rosette leaves. Its target gene *AGL16*, encoding a MADS-box transcription factor, is highly expressed in wild-type rosette leaves (especially in guard cells and trichomes), and moderately expressed in roots and stems (Alvarez-Buylla et al. 2000), exhibiting an inverse relationship with miR824 sequencing abundance in these tissues consistent with miR824-mediated targeting of *AGL16*. This miRNA-target pair may be involved in fine-tuning nitrogen assimilation, because *AGL16* is upregulated under conditions of nitrogen deprivation (Bohnsack et al. 2004), suggesting that miR824 expression may be dampened under nitrogen stress or that *AGL16* escapes miRNA-mediated regulation by some other mechanism. The target genes whose transcription is modulated by AGL16 are unknown, and since MADS-box genes can encode transcriptional activators or repressors, several possible regulatory circuits present themselves which merit further study. Under conditions of nitrogen starvation, AGL16 protein may activate or repress genes involved in nitrogen uptake and metabolic pathways, either enhancing or attenuating nitrogen assimilation. In either case, miR824 may act to maintain nitrogen assimilation at homeostatic levels under standard growth conditions when ample but not excess nitrogen is available in the environment. Analysis of the relative expression profiles of both miR824 and *AGL16* under nitrogen stress, as well as the identification of genes whose promoters are bound by AGL16, could illuminate one mechanism by which regulated absorption and metabolism of this macronutrient is effected.

Discovery of novel *MIR* genes

How many more miRNAs might exist in *Arabidopsis*? Our data suggests that species-specific miRNAs may far outnumber the core set of conserved miRNAs in plants. The discovery of such miRNAs will be facilitated by improvements in deep sequencing technologies. As more and more high-throughput sequencing projects reveal additional miRNA candidates that are less and less abundant, the challenge of annotating *bona fide* miRNAs will only increase. Since experimental cloning is limited to uncovering miRNAs expressed under the sampled conditions, and since we know that miRNAs are often responsive to environmental shifts, sequencing from a variety of mutant lines, growth and stress conditions, and precisely defined cell lineages and developmental stages will allow identification of miRNAs that are more highly expressed under a limited range of conditions, and aid attempts at cataloguing the diversity of miRNAs that plants employ for regulatory activities. Some potentially informative results could be derived, for example, by sampling specifically from apical and basal meristematic tissue, roots, and embryonic tissues, and plants grown in unseasonal temperature and light cycles. Also, plants grown on soils depleted for one or more macronutrients may harbor additional stress-induced miRNAs; miR395 and miR399 may act within a larger network of miRNAs which are responsive to the environment and whose concerted action can shift the gene expression profile away from those whose expression is not needed, or towards those whose expression is required, under certain conditions.

With new genomes comes the possibility of tracing evolutionary histories of what seem now to be species-specific miRNAs in *Arabidopsis*. Sequencing of species more closely related to *Arabidopsis* could aid such efforts. In addition, lineage-restricted miRNAs have been sequenced from several other species, including *Oryza sativa* (Sunkar et al. 2005; Nobuta et al. 2007), *Populus trichocarpa* (Lu et al. 2005), *Physcomitrella patens* (moss) (Arazi et al. 2005; Talmor-Neiman et al. 2006; Axtell et al. 2007), *Triticum aestivum* (wheat) (Yao et al. 2007), *Solanum lycopersicon* (tomato) (Pilcher et al. 2007) and even a unicellular photosynthetic eukaryote, *Chlamydymonas reinhardtii* (green alga) (Molnar et al. 2007; Zhao et al. 2007). The biogenetic requirements of species-specific miRNAs merit further characterization to confirm their designation, and

establishing their presence or absence in related species will more accurately define the mechanisms of miRNA evolution in land plants.

Models for *MIR* gene evolution

The 38 non-conserved miRNAs discussed in this work point to a hitherto uncharacterized diversity of evolving miRNA loci in *Arabidopsis*. These 38 miRNAs are represented by single loci in the genome, consistent with the idea that their relatively recent emergence in the *Arabidopsis* lineage has not afforded sufficient time for these miRNAs to expand and diversify in the way that their deeply conserved counterparts have. The simple architecture of miRNA genes, and the perhaps minimal fitness consequences incurred by the plant due to their initially low expression, may make emergent miRNAs easy substrates for the evolution of new regulatory circuits which may prove advantageous under certain growth conditions.

One mechanism for miRNA gene evolution appears to be duplication of future target genes (Allen et al. 2004). If the most direct mechanism for miRNA gene emergence is by inverted duplication of a future target gene, then miRNA gene loci should initially display a high degree of similarity to these future target genes. However, the historical era during which the *MIR* locus materializes in the genome, the point during its evolutionary trajectory at which it is discovered, and the nature of the selective forces (if any) acting on the gene, will impact the relative degree of sequence similarity that might be observable at a given timepoint. Furthermore, different *MIR* genes appear to be evolving at different rates in the genome (Maher et al. 2006). Any duplications of future target genes giving rise to ancient, highly conserved miRNAs likely happened in the common ancestor of land plants, while duplications that gave birth to miRNAs conserved only to angiosperms are likely to have occurred before the monocots and dicots split from each other. In either case, hundreds of millions of years of evolution have occurred, which is likely to blur the signal. Duplications giving rise to non-conserved miRNAs are likely to have occurred much more recently, perhaps along the eurosids I lineage or later; however, if these duplications did take place, then sequence divergence of the *MIR* loci and the target genes appears to have proceeded rapidly for many loci described in this study. More sensitive algorithms for detecting homology at a significant level may be

capable of mining such miRNA/target pairs for signals of inverted duplication evolution that escape detection using our algorithm.

At least five miRNAs described in this work retain significant homology to predicted target genes, a genomic footprint suggestive of recent evolution; evidently, insufficient time has elapsed for significant sequence divergence to have erased this resemblance. Several fates are potentiated by this process. One region of the hairpin could retain sufficient sequence complementarity to be competent to specify target cleavage of the parental gene, and mutation in this region might then be selected against, while the rest of the hairpin could diverge from the parental gene to the extent that compensatory mutations in both arms of the hairpin maintained sufficient basepairing and other necessary sequence determinants to allow proper recognition and excision of the nascent miRNA by the small RNA silencing machinery. Over evolutionary time, the acquisition of promoter elements that enhance expression might be selected for, allowing spatially-defined accumulation of the miRNA in those tissues where its action is required or advantageous. Together these processes would stabilize *MIR* genes in the genome; acquisition of biologically relevant functions would help protect them from elimination by random mutation. Such selective forces are likely to act at loci encoding conserved miRNAs, and relatively abundant non-conserved miRNAs like miR823 and miR824 which are expressed to an appreciable degree. However, it is unclear if these miRNA genes evolved from ancestral target gene duplications.

An alternative post-duplication fate might be mutational drift and rapid divergence of the entire hairpin, either eliminating the potential for stemloop formation (which would presumably attenuate DCL-mediated miRNA production), or maintaining a hairpin-like structure that is nevertheless distinct in sequence from the parental gene. This could provide a means for inverted duplication of one gene to give rise to a regulatory circuit involving miRNA targeting of another related or sufficiently homologous gene. The liberation of a single duplex in excess of all others might correlate with stabilization of the miRNA within AGO1-RISC even in the absence of an evolved functional target interaction, if no deleterious consequences are invoked (such as competitive inhibition of AGO1-RISC that reduces the activity of functionally relevant miRNAs).

Of course, inverted duplications of non-genic loci may occur as frequently by chance as duplications of genic loci, giving rise to regions that express miRNA-like hairpins. A subset of the newly-identified miRNAs, perhaps enriched in those without predictable targets and/or those lacking resemblance to predicted target genes, may have emerged by such duplications. While some are likely to be non-functional in terms of targeting or regulatory activities, and may be purged from the genome, others may benefit from advantageous sequence divergence and be co-opted into useful regulatory interactions.

Regardless of their origins, miRNAs in transition might comprise a reservoir of incipient regulatory molecules which can serve as easy substrates for natural selection. Most of the miRNAs described in this work would fit in this category. *MIR* loci that are recruited into regulatory circuits and result in fitness advantages for the plant would be expected to be retained, while those with neutral or deleterious effects would be subject to mutational forces and eventually dissipate. Given that only about 17 nt of complementarity is required for efficient cleavage, accumulated mutations might also by chance occasionally result in an enhanced interaction with a target gene related to the parental gene.

As new genome sequences emerge, it will be informative to determine if orthologs of these newly identified miRNAs exist in near relatives of *A. thaliana*. For example, recent cloning of small RNAs from the basal eudicot *Eschscholzia californica* (California poppy) revealed the presence of an ortholog of *MIR845* (Barakat et al. 2007). Similar data from other sister clades can help in mapping the evolutionary trajectories of individual miRNA families across many clades, which will allow us to refine our models for the evolution of miRNAs using the genomic footprints of more recently emerged miRNAs in the genome.

Trans-acting siRNA loci in the *Arabidopsis* genome

What might the role of *TAS4* tasiRNAs be, given that the vegetative phase change phenotypes typical of tasiRNA biogenesis mutants appears to be almost entirely attributable to the disruption of *TAS3* tasiRNA-mediated repression of *ARF3* and *ARF4*? Recent work suggests a novel phenotype for *dcl4-2* and *drb4-1* mutant plants that may

indicate a role for *TAS4* tasiRNAs in regulating leaf senescence. When aged to12-weeks old, the rosettes of wild-type plants are completely green, whereas *dcl4-2* and *drb4-1* leaves display various shades of red or purple pigmentation which is most concentrated at leaf tips and edges. The red-leaf phenotype is almost 100% penetrant in both mutants. The light absorbance of leaf extracts indicates that the pigment molecule is anthocyanin in both cases (Nakazawa et al. 2007). This imbalance in anthycyanin pigmentation is phenocopied in *pap1-D* activation-tagged mutants, displaying a purple coloration throughout the plant body at all stages of development that is enhanced under stressful growth conditions such as high-intensity light, drought, and pathogen infection, symptomatic of the underlying over-accumulation of anthocyanins and phenylpropanoid products (Borevitz et al. 2000). Leaf senescence in *A. thaliana* seems to be induced rather than repressed by glucose, and both *PAP1* and *PAP2* are upregulated after glucose treatment (Pourtau et al. 2006). Perhaps *TAS4*-siR81 downregulation of *PAP1* in plants delays leaf senescence until an appropriate point during the lifecycle of the plant. The study of leaf phenotypes associated with *TAS4* T-DNA insertion lines may be able to shed light on the question of whether *TAS4* forestalls premature leaf aging.

Do other *TAS* loci lie hidden in the *Arabidopsis* genome? The likelihood is high; a handful of loci displayed phased small RNA accumulation, though not nearly as extensively or abundantly as that exhibited by *TAS4*, with its ~10 registers. Higher-throughput sequencing may be able to unearth new tasiRNA clusters. A major unanswered question is why and how some tasiRNAs without predictable or empirically-verified targets are more stabilized in cells than others, to a level beyond that which can be rationalized by the asymmetry rules for RISC incorporation? In a similar vein, why does *TAS3* seem to give rise to such a relatively high abundance of apparently inactive siRNAs and such comparatively low amounts of the siRNAs that repress *ARF3* and *ARF4?* A third open question is how *TAS* loci evolve, given that they potentially encode silencing RNAs that can act coordinately but are also restricted in their expression due to the polycistronic arrangement of the small RNAs. These unresolved aspects require further investigation.

Candidate heterochromatic siRNAs in *Arabidopsis*

The scope of candidate siRNA production in *Arabidopsis* is vast, and we seem to be far from the point at which we might believe we have identified all siRNA-spawning loci. Indeed, the propensity or capacity for any locus to give rise to siRNAs may be dependent on and fluctuate with shifts in the spatial distribution of silencing machinery, the developmental stage, and environmental stimuli including biotic and abiotic reagents, as in the case of the nat-siRNAs. Several conundrums shroud in mystery the biogenesis and function of these small RNAs. Why do some loci thought to be intergenic or non-expressed spawn such an abundance of small RNAs? How exactly are silencing signals amplified by PolIV variants, histone modifying proteins, and DNA methyltransferases? What is the function of AGO4 in assembling some heterochromatic loci? Why do almost all silencing-associated proteins exhibit locus-specific effects? Does methylation of DNA and/or histones correlate with reduced transcription? Is PolIV capable of generating transcripts at loci that are inaccessible to PolII, in order to both maintain the condensed chromatin state and simultaneously prime siRNA biogenesis? What are the mechanisms by which some siRNAs direct chromatin modifications? These and many other questions are likely to keep small-RNA biologists occupied for quite some time.

Large-scale evolutionary consequences of small RNA-mediated gene regulation

The discovery of a large class of apparently lineage-specific miRNAs suggests that the birth of new and useful miRNA genes may also correlate with large-scale speciation events. For example, miR172-regulation of *GLOSSY15* in maize promotes juvenile-to-adult phase transitions in maize (Lauter et al. 2005). Persistence of juvenile traits into adulthood (a process known as neoteny) is exhibited by *Corngrass1* maize mutants, and hearkens to likely phenotypes of ancestral wild grasses like teosinte, the progenitor of maize. The *Cg1* mutation over-expresses a tandem miR156 locus which promotes juvenility by decreasing miR172 levels, implicating the adaptation of miR156-mediated regulatory circuits in the domestication of corn (Chuck et al. 2007). Eukaryotic transcriptomes are necessarily fluid from cell to cell and stage to stage, and small RNA regulatory control is one contributing aspect to this fluidity and the adaptive plasticity it

confers to plants in various environmental contexts. Therefore, if miRNA gene evolution in plants is quite widespread and relatively frequent in response to the challenges presented by the sessile lifecycle, the characterization of the miRNA complements of related but distinct species may help to expose mechanisms of speciation that are directly impacted by *MIR* gene emergence.

REFERENCES

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

Alvarez-Buylla, E.R., S.J. Liljegren, S. Pelaz, S.E. Gold, C. Burgeff, G.S. Ditta, F. Vergara-Silva, and M.F. Yanofsky. 2000. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* **24**: 457-466.

Arazi, T., M. Talmor-Neiman, R. Stav, M. Riese, P. Huijser, and D.C. Baulcombe. 2005. Cloning and characterization of micro-RNAs from moss. *Plant J* **43**: 837-848.

Axtell, M.J. and D.P. Bartel. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell*.

Axtell, M.J., J.A. Snyder, and D.P. Bartel. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell*.

Barakat, A., K. Wall, J. Leebens-Mack, Y.J. Wang, J.E. Carlson, and C.W. Depamphilis. 2007. Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants. *Plant J*.

Bohnsack, M.T., K. Czaplinski, and D. Gorlich. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**: 185-191.

Borevitz, J.O., Y. Xia, J. Blount, R.A. Dixon, and C. Lamb. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**: 2383-2394.

Chuck, G., A.M. Cigan, K. Saeteurn, and S. Hake. 2007. The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat Genet* **39**: 544-549.

Kinoshita, T., A. Miura, Y. Choi, Y. Kinoshita, X. Cao, S.E. Jacobsen, R.L. Fischer, and T. Kakutani. 2004. One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**: 521-523.

Lauter, N., A. Kampani, S. Carlson, M. Goebel, and S.P. Moose. 2005. microRNA172 down-regulates *glossy15* to promote vegetative phase change in maize. *Proc Natl Acad Sci U S A* **102**: 9412-9417.

Lindroth, A.M., X. Cao, J.P. Jackson, D. Zilberman, C.M. McCallum, S. Henikoff, and S.E. Jacobsen. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**: 2077-2080.

Lu, S., Y.H. Sun, R. Shi, C. Clark, L. Li, and V.L. Chiang. 2005. Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17**: 2186-2203.

Maher, C., L. Stein, and D. Ware. 2006. Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res* **16**: 510-519.

Molnar, A., F. Schwach, D.J. Studholme, E.C. Thuenemann, and D.C. Baulcombe. 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447**: 1126-1129.

Nakazawa, Y., A. Hiraguri, H. Moriyama, and T. Fukuhara. 2007. The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the *trans*-acting siRNA pathway. *Plant Mol Biol* **63**: 777-785.

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

Pilcher, R.L., S. Moxon, N. Pakseresht, V. Moulton, K. Manning, G. Seymour, and T. Dalmay. 2007. Identification of novel small RNAs in tomato (*Solanum lycopersicum*). *Planta* **226**: 709-717.

Pourtau, N., R. Jennings, E. Pelzer, J. Pallas, and A. Wingler. 2006. Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in *Arabidopsis*. *Planta* **224**: 556-568.

Sunkar, R., T. Girke, P.K. Jain, and J.K. Zhu. 2005. Cloning and characterization of microRNAs from rice. *Plant Cell* **17**: 1397-1411.

Talmor-Neiman, M., R. Stav, W. Frank, B. Voss, and T. Arazi. 2006. Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. *Plant J* **47**: 25-37.

Tompa, R., C.M. McCallum, J. Delrow, J.G. Henikoff, B. van Steensel, and S. Henikoff. 2002. Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr Biol* **12**: 65-68.

Yao, Y., G. Guo, Z. Ni, R. Sunkar, J. Du, J.K. Zhu, and Q. Sun. 2007. Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biol* **8**: R96.

Zhao, T., G. Li, S. Mi, S. Li, G.J. Hannon, X.J. Wang, and Y. Qi. 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev* **21**: 1190-1203.

Molecular Cell, Vol. 16, 69–79, October 8, 2004, Copyright ©2004 by Cell Press

Endogenous *trans*-Acting siRNAs Regulate the Accumulation of *Arabidopsis* mRNAs

Franck Vazquez,[1,2,5] Hervé Vaucheret,[2,3,5,*] Ramya Rajagopalan,[3,4] Christelle Lepers,[1] Virginie Gasciolli,[2] Allison C. Mallory,[3] Jean-Louis Hilbert,[1] David P. Bartel,[3,4] and Patrice Crété[1,6,*]

[1]Laboratoire de physiologie de la différenciation végétale
UPRES-EA3569, IFR118, EPT1016
Université des Sciences et Technologies de Lille
Bât. SN2, cité scientifique
59650 Villeneuve d'Ascq Cedex
France

[2]Laboratoire de Biologie Cellulaire
Institut Jean-Pierre Bourgin, INRA
78026 Versailles Cedex
France

[3]Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, Massachusetts 02142

[4]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

Here we describe a set of endogenous short interfering RNAs (siRNAs) in *Arabidopsis*, some of which direct the cleavage of endogenous mRNAs. These siRNAs correspond to both sense and antisense strands of a noncoding RNA (At2g27400) that apparently is converted to double-stranded RNA and then processed in 21 nt increments. These siRNAs differ from previously described regulatory small RNAs in two respects. First, they require components of the cosuppression pathway (RDR6 and SGS3) and also components of the microRNA (miRNA) pathway (AGO1, DCL1, HEN1, and HYL1) but not components needed for heterochromatic siRNAs (DCL3 and RDR2), another class of endogenous plant siRNAs. Second, these siRNAs repress the expression of genes that have little overall resemblance to the genes from which they originate, a characteristic previously reported only for miRNAs. The identification of this silencing pathway provides yet another dimension to posttranscriptional mRNA regulation in plants.

Introduction

Endogenous noncoding small RNAs (20 to 25 nt long) are important regulators of gene expression in both plants and animals (Carrington and Ambros, 2003; Bartel, 2004; He and Hannon, 2004; Mallory and Vaucheret, 2004). Two types of endogenous small RNA regulators have been reported, microRNAs (miRNAs) and short interfering RNAs (siRNAs). miRNAs are processed from transcripts that can form local hairpin precursor structures, whereas siRNAs are processed from long bimolecular RNA duplexes or extended hairpins (Bartel and Bartel, 2003). miRNAs have been cloned from various organisms and are typically evolutionarily conserved (Bartel, 2004). Many of the genes that are regulated by miRNAs are essential for proper development in plants (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004), and the importance of various miRNA-mediated regulatory relationships has been demonstrated in plants and animals (Lee et al., 1993; Wightman et al., 1993; Moss et al., 1997; Reinhart et al., 2000; Aukerman and Sakai, 2003; Boutla et al., 2003; Brennecke et al., 2003; Emery et al., 2003; Johnston and Hobert, 2003; Palatnik et al., 2003; Xu et al., 2003; Chen, 2004; Mallory et al., 2004a, 2004b; Vaucheret et al., 2004).

In plants, proper miRNA accumulation depends on the activity of the nuclear proteins DCL1, HEN1, and HYL1 (Park et al., 2002; Reinhart et al., 2002; Boutet et al., 2003; Kasschau et al., 2003; Han et al., 2004; Sunkar and Zhu, 2004; Vazquez et al., 2004). *dcl1* null alleles are embryo-lethal, indicating that DCL1 is required for plant viability, at least during reproduction and/or at early stages of development (Schauer et al., 2002). Partial loss-of-function *dcl1* mutants with point mutations in the RNA helicase domain (*dcl1-7*, *dcl1-8*) or truncation of the second dsRNA binding domain (*dcl1-9*) are viable (Schauer et al., 2002) but show reduced miRNA accumulation and striking developmental defects including sterility, confirming the crucial role of DCL1 and presumably miRNAs during plant reproduction (Park et al., 2002; Reinhart et al., 2002). *hen1* and *hyl1* null alleles exhibit reduced miRNA levels and developmental defects that overlap with those of partial loss-of-function *dcl1* mutants. However, in contrast to *dcl1* null alleles, *hen1* and *hyl1* null alleles are viable. AGO1 also participates in the miRNA pathway (Kidner and Martienssen, 2004; Vaucheret et al., 2004). AGO1 is the founding member of the ARGONAUTE protein family (Bohmert et al., 1998). Hypomorphic *ago1* alleles exhibiting developmental defects overlapping with those of *hen1* and *hyl1* accumulate miRNAs but show increased accumulation of the corresponding target mRNAs, an observation that when combined with analogous roles of Argonaute proteins in animals (Carmell et al., 2002) suggests that AGO1 is part of the miRNA-programmed RNA-induced silencing complex (RISC) (Vaucheret et al., 2004). Most miRNAs have reduced accumulation in *ago1* null alleles, also suggesting that AGO1 stabilizes miRNAs (Vaucheret et al., 2004). A role for AGO1 in controlling the cellular localization of miRNAs has also been proposed (Kidner and Martienssen, 2004).

In plants, two types of siRNA pathways have been described. The first type involves siRNAs that trigger changes in the chromatin state of elements from which they derive (Hamilton et al., 2002; Finnegan and Matzke,

*Correspondence: herve.vaucheret@versailles.inra.fr (H.V.), crete@luminy.univ-mrs.fr (P.C.)

[5]These authors contributed equally to this work.

[6]Present address: Laboratoire de Génétique et Biophysiques des Plantes, UMR 6191 CEA-CNRS-Université de la Méditerranée, Faculté des sciences Case 901, 163 Avenue de Luminy, 13288 Marseille cedex 9.

2003; Zilberman et al., 2003; Xie et al., 2004). Accumulation of these siRNAs requires DCL3, one of the four *Arabidopsis* DCL proteins (Xie et al., 2004). In contrast to miRNAs, four chromatin-associated siRNAs described by Xie et al. (2004) require RDR2, one of multiple RNA-dependent RNA polymerases (RDR) in plants. A second pathway involves siRNAs deriving from and defending against exogenous RNA sequences such as viruses or sense transgene transcripts (S-PTGS). The production of siRNAs associated with S-PTGS requires another RDR protein, RDR6 (also known as SDE1/SGS2), and a protein of unknown function, SGS3 (Dalmay et al., 2000; Mourrain et al., 2000; Boutet et al., 2003). Thus far, no role for regulating endogenous RNAs has been assigned to RDR6 and SGS3.

Here, we identify an endogenous siRNA pathway that requires AGO1, DCL1, HEN1, HYL1, SGS3, and RDR6. This pathway differs from the heterochromatic siRNA pathway, which requires DCL3 and RDR2 but not SGS3 and RDR6, but resembles the miRNA pathway, which requires AGO1, DCL1, HEN1, and HYL1. Like miRNAs, these siRNAs direct cleavage of endogenous complementary mRNAs that have little overall resemblance to the genes from which the siRNAs originate; however, unlike miRNAs, their production requires RDR6 and SGS3. The identification of endogenous mRNAs that are regulated at the post-transcriptional level by these siRNAs raises the possibility that numerous endogenous genes may be regulated by a similar pathway.

Results

Identification of Endogenous RNAs that Are Deregulated in rdr6 *and* sgs3 *Mutants*

RDR6 and SGS3 are required for posttranscriptional gene silencing mediated by sense-transgene (S-PTGS) or DNA viruses but not inverted repeat transgenes (IR-PTGS) or RNA viruses (Dalmay et al., 2000; Mourrain et al., 2000; Beclin et al., 2002; Muangsan et al., 2004). To determine if RDR6 and SGS3 could also play a role in posttranscriptional regulation of endogenous mRNAs, RNA steady-state levels in wild-type plants, *sgs2(rdr6)* and *sgs3* mutants were compared by cDNA-amplified fragment length polymorphism (cDNA-AFLP). Out of the 256 possible pairs, 80 primer pairs were tested, representing approximately 4000 polyadenylated transcripts. Three cDNA-AFLP products reproducibly showed elevated levels in *rdr6* and *sgs3* mutants, and we analyzed one of these products in more detail (Figure 1).

This product corresponds to two genomic DNA regions on BAC F12K2 separated by a sequence of 572 nucleotides (Figure 1). Sequence analysis revealed cannonical donor and acceptor splice-site sequences, both at expected positions within the genomic sequences, suggesting that the cDNA-AFLP product results from the splicing of a single intron. RT-qPCR confirmed a 3- to 4-fold increase in the level of the At2g27400 transcript in *rdr6* and *sgs3* mutants, first indicated by our cDNA-AFLP analysis (data not shown). 5′-rapid amplification of cDNA ends (RACE) and 3′-RACE followed by sequencing revealed that the full-length 926 nt sequence of the transcript is composed of two exons of 164 and 190 nucleotides separated by an intron of 572 nucleotides. This transcript is polyadenylated and capped and originates from a nonannotated genomic region located between the annotated genes At2g27410 and At2g27395. Therefore, we named this transcript At2g27400. Because this transcript contains only one ORF, 21 amino acids in length, starting at the first AUG at position 55, relative to the 5′ end of the transcript, it is likely that RNA, not protein, is the final product of this gene. Two ESTs corresponding to At2g27400 were found in the *Arabidopsis* database. Both ESTs start at position 52, relative to the 5′ end of the transcript determined by 5′-RACE, and end within the intron. One is not polyadenylated and ends at position 674. The other is polyadenylated and ends at position 711, either due to the presence of a polyadenylation signal at position 714 or to oligo-dT mispriming in an A-rich region that follows this position. The absence of full-length spliced or unspliced EST in the *Arabidopsis* database suggests that the At2g27400 gene is not strongly transcribed or that the primary RNA and its spliced product are immediately subjected to further processing.

A Cluster of Small RNAs Originates from the Intron of the At2g27400 Transcript through RDR6 and SGS3 Activity

Because the At2g27400 transcript is a noncoding RNA, we looked for small RNAs deriving from the sequence of At2g27400 in sequenced libraries of small RNA databases from plants (Park et al., 2002; Reinhart et al., 2002; Sunkar and Zhu, 2004; Xie et al., 2004; R.R. and D.P.B., unpublished data). The At2g27400 unspliced transcript shares perfect identity with twelve independently cloned small RNAs within the portion of the intron that is common to both the unspliced transcript and the truncated EST (Figure 1). Ten small RNAs correspond to the At2g27400 transcript whereas two small RNAs are complementary to At2g27400 (Table 1). We refer to these RNAs by their position relative to the 5′ end of the At2g27400 transcription start (Figure 1) and the strand from which they originate (+ or −). Using an antisense At2g27400 RNA probe we detected 21 nt small RNAs of sense polarity in wild-type plants but not in *sgs2(rdr6)* and *sgs3* mutants (Figure 2A), indicating that the accumulation of these small RNAs depends on RDR6 and SGS3 activities. Small RNAs of antisense polarity were not detected using a sense At2g27400 RNA probe (data not shown), consistent with the smaller fraction of cloned small RNAs representing antisense polarity versus sense polarity (Table 1).

Nine of the twelve independently cloned small RNAs occur in perfect 21 nt increments, eight on the sense strand, one on the antisense strand (Table 1; Figure 1). Two of the others, siR521(+) and siR523(+), are variants of siR522(+) and deviate from perfect phasing by only one nucleotide at their 5′ termini. Only one, siR466(−) was processed in a different register, having a 10 nt overlap with siR477(−). The two 21 nt strands of each siRNA duplex produced by Dicer are not equally eligible for assembly into the cleavage complex (RISC); the strand with lowest base-pairing stability at its 5′ end preferentially enters into RISC (Khvorova et al., 2003; Schwarz et al., 2003). All eleven At2g27400-derived small RNAs that are in phase obey the asymmetry rule

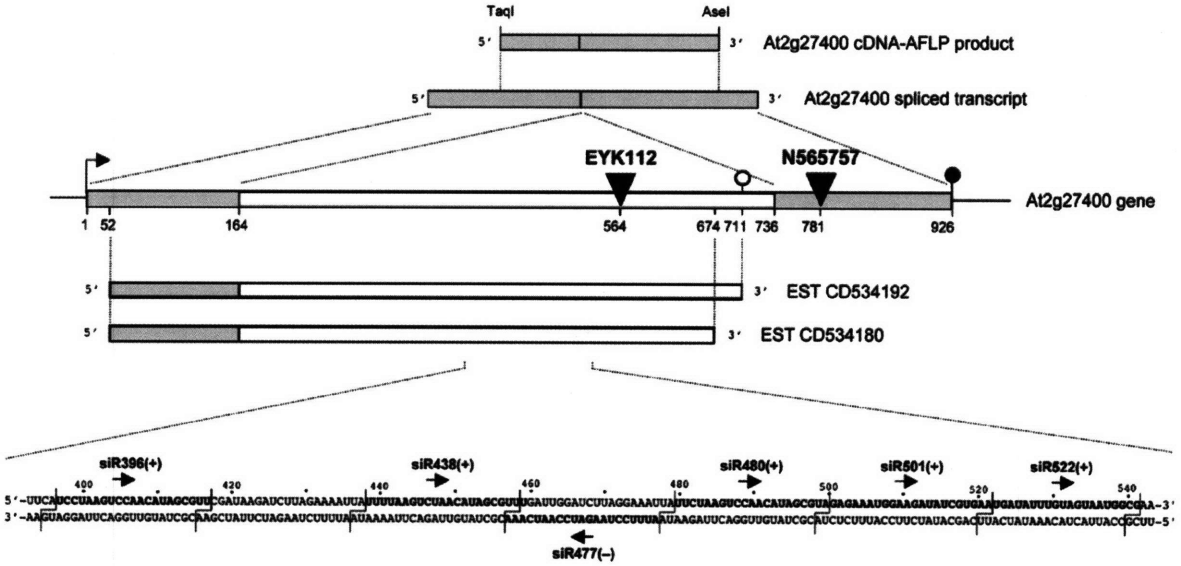


Figure 1. siRNAs Correspond to Both Sense and Antisense Strands of the At2g27400 Noncoding RNA

The diagrammatic representation of the At2g27400 gene is shown in the middle of the figure. The two exons are represented by gray boxes and the intron is represented in white. The cDNA-AFLP clone is a TaqI-AseI fragment internal to the spliced transcript. The 5′ and 3′ extensions of the transcript were obtained by 5′- and 3′-RACE. The arrow and filled circle indicate the position of transcription start and polyadenylation signal, respectively. The open circle indicates the position of the putative polyadenylation signal. Black triangles indicate the locations of the T-DNA insert within At2g27400 in the EYK112 and N565757 lines. The intron sequence corresponding to the cloned siRNAs is shown in the expanded region and the sequence of the cloned strand of the siRNA duplex is shown in bold. Numbers indicate positions on the full-length transcript, relative to the 5′ end.

that characterizes siRNAs and miRNAs in animals and plants, explaining the bias in polarity of the eleven cloned small RNAs and the difficulty in detecting siRNAs from the antisense strand on RNA blots (Figure 2A). This observation also implies that these small RNAs enter into a RISC-like complex (see below).

The requirement of RDR6 and SGS3 for small RNA accumulation (Figures 2A and 2B), the 21 nt phasing of small RNAs (Figure 1), and the presence of 21 nt clones from both the sense and antisense strands (Figure 1) suggest that these small RNAs likely derive from the successive Dicer-mediated cleavage of a long dsRNA formed by the RNA-dependent RNA polymerase activity of RDR6. Multiple small RNAs, ~21 nt in length, deriving from both strands of a long dsRNA precursor have the defining features of siRNAs (Elbashir et al., 2001) and thus we call these small RNAs the At2g27400 siRNAs.

Accumulation of At2g27400 siRNAs Requires AGO1, DCL1, HEN1, and HYL1 1 but not AGO7, DCL2, DCL3, RDR2, SDE3, and WEX

To determine the genetic requirements for At2g27400 siRNA production, their accumulation was assayed in silencing-related mutants (*ago1*, *ago7*, *dcl1*, *dcl2*, *dcl3*, *hen1*, *hyl1*, *rdr2*, *sde3*, and *wex*). At this stage of development, At2g27400 siRNAs were detected in wild-type

Table 1. Cloned siRNAs Deriving from the At2g27400 Transcript

At2g27400 siRNAs	Length (nt)	Phasing	# clones	alternative name
siR396(+)	21	Yes	2[a,b]	ASRP752[b], miR175c[d] miR389b.3[c], miR175d[d]
siR438(+)	21	Yes	1[c]	
siR466(−)	21	No	1[a]	
siR477(−)	21	Yes	1[a]	
siR480(+)	21	Yes	3[b,c,d]	ASRP255[b], miR389b.1[c], miR175a-2[d]
siR501(+)	21	Yes	1[a]	
siR521(+)	20	Yes[e]	1[b]	ASRP1745[b]
siR522(+)	21	Yes	1[b]	ASRP1852[b]
siR523(+)	22	Yes[e]	1[b]	ASRP1786[b]

(+) sense strand, (−) antisense strand
[a]this work
[b]Xie et al., 2004, http://cgrb.orst.edu/smallRNA/db/search_user_seq.html
[c]Sunkar and Zhu, 2004
[d]Park et al., 2002
[e]in phase on 3′ end, one nucleotide out of phase on 5′ end

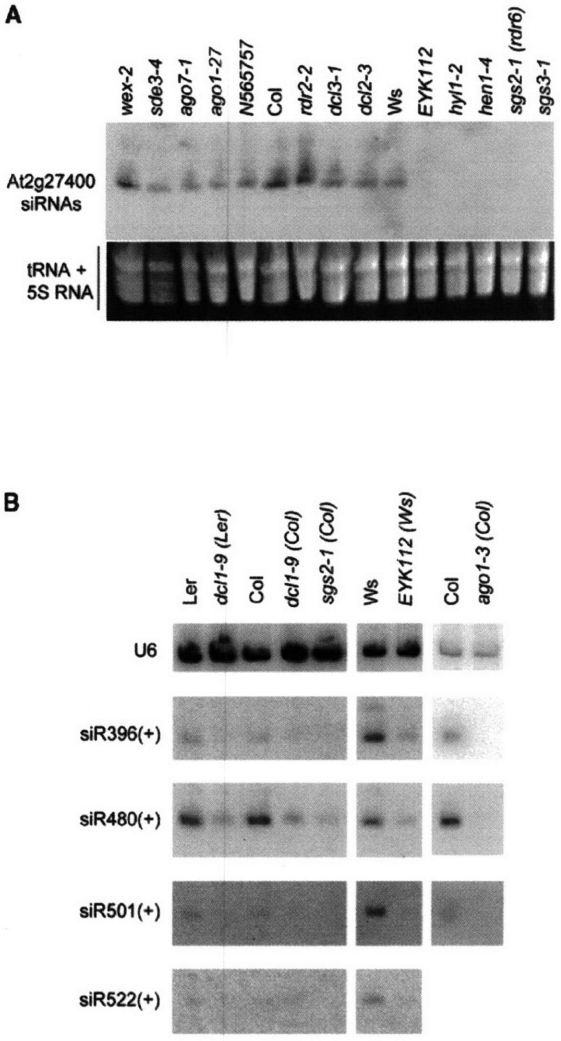


Figure 2. Accumulation of the At2g27400 siRNAs Requires AGO1, DCL1, HEN1, HYL1, RDR6, and SGS3

(A) RNA gel blot analysis of At2g27400 siRNA accumulation in wild-type plants (Col or Ws) and *wex-2*, *sde3-4*, *ago7-1*, *ago1-27*, *N565757*, *rdr2-2*, *dcl3-1*, *dcl2-3*, *EYK112*, *hyl1-2*, *hen1-4*, *sgs2-1 (rdr6)* and *sgs3-1* mutants (all in Col, except for EYK112 in Ws). 15 μg of total RNA was extracted from the aerial part of 4-6 leaf-old plants and hybridized to a full-length unspliced At2g27400 antisense RNA probe. EtBr staining of tRNA + 5S RNA is shown as a loading control.

(B) RNA gel blot analysis of siR396(+), siR480(+), siR501(+), and siR522(+) accumulation in wild-type plants (Col, Ler or Ws) and *dcl1-9*, *sgs2-1 (rdr6)*, *EYK112* and *ago1-3* mutants. 10 μg of total RNA was extracted from the aerial part of 4–6 leaf-old plants and successively hybridized to DNA oligo probes complementary to each siRNA. U6 is shown as a loading control.

plants, *ago7*, *dcl2*, *dcl3*, *rdr2*, *sde3*, and *wex* null mutants and in hypomorphic *ago1* mutants (Figure 2). In contrast, the accumulation of these siRNAs was reduced in the partial loss-of-function *dcl1* mutant and they were below detectable level in *ago1*, *hen1*, and *hyl1* null mutants (Figure 2), indicating that, like miRNAs, the accumulation of these siRNAs requires AGO1, DCL1, HEN1, and HYL1 but, unlike heterochromatic siRNAs, does not require DCL3 or RDR2.

siRNA Accumulation Is Decreased in At2g27400 Intron- but Not Exon2-Insertion Mutants

Among the siRNAs that correspond to At2g27400 on BAC F12K2, one, siR480(+), shows a perfect homology with two other intergenic regions located on BAC F2J10 and F17A14. Since the At2g27400 probe used in the previous experiment shares 80% nucleotide identity with F2J10 and 72% with F17A14, it is possible that some of the siRNAs detected with this probe could derive from one of these sequences. To determine if the siRNAs detected with the At2g27400 RNA probe derive from At2g27400, we analyzed small RNA accumulation in a mutant (*EYK112*) carrying a T-DNA inserted in the At2g27400 intron at position 568, relative to the 5′ end of the transcript determined by 5′-RACE, downstream of the location of the siRNAs, but within the portion of the intron that is common to both the unspliced transcript and the truncated EST (Figure 1). Using an antisense At2g27400 RNA probe, siRNAs were detected in wild-type plants but not in the *EYK112* mutant (Figure 2A), indicating that, at this stage of development, the siRNAs detected with this probe derive from BAC F12K2 (At2g27400) and not from BACs F2J10 or F17A14. Additional RNA blots were hybridized with DNA oligo probes complementary to individual siRNAs (Figure 2B). The accumulation of each of the siRNAs was strongly reduced in the *EYK112* mutant, although traces could still be detected. Since a chimeric transcript starting at the 5′ end of At2g27400 and ending in the T-DNA is present in *EYK112* homozygous plants (data not shown), it is possible that the residual amount of siRNAs is processed from this chimeric transcript or transcripts deriving from BACs F2J10 or F17A14. A second insertion mutant (*N565757*) was identified, carrying a T-DNA inserted in the At2g27400 exon2 at position 781, relative to the 5′ end of the transcript determined by 5′-RACE (Figure 1). Using an antisense At2g27400 RNA probe, siRNAs were detected at similar levels in wild-type plants and N565757 mutants (Figure 2A), suggesting that the T-DNA in N565757 is inserted beyond the 3′ end of the siRNA precursor transcript, and that siRNAs do not derive from the full-length unspliced transcript but rather from the spliced intron or a transcript corresponding to the truncated ESTs.

Targets of At2g27400 siRNAs Are Cleaved in Wild-Type Plants

In plants, miRNAs cleave endogenous mRNAs by binding to unique segments of near-perfect complementarity (Llave et al., 2002b, Kasschau et al., 2003; Tang et al., 2003). Ten out of eleven At2g27400-siRNAs obey the asymmetry rule that characterizes miRNAs and siRNAs (Khvorova et al., 2003; Schwarz et al., 2003), suggesting that they could enter a RISC-like complex to guide cleavage of complementary mRNAs. Indeed, two of the previously cloned siRNAs, siR438(+) and siR480(+), have been proposed to target At5g18040, At1g51670, At4g29760, and At4g29770 (Park et al., 2002; Sunkar and Zhu, 2004). siR396(+), which is nearly identical to siR438(+) and siR480(+), also has the potential to target the same genes. At5g18040 has only one mismatch with the three siRNAs, and At1g51670, At4g29760, and At4g29770 each have two mismatches with siR480(+)

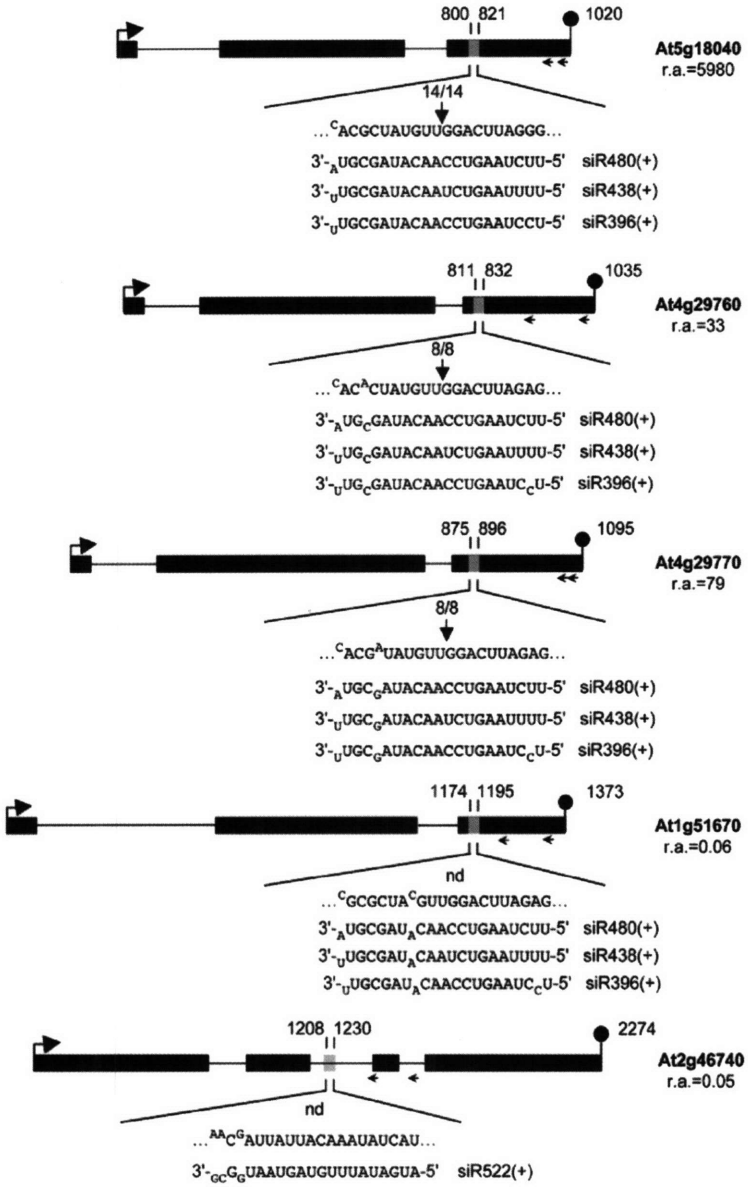


Figure 3. At2g27400 siRNAs Cleave Target mRNAs

Black boxes represent open reading frames. The arrows and filled circles represent transcription starts and polyadenylation signals, respectively. Gray boxes represent the siRNA complementary site, with the nucleotide positions relative to the transcription start indicated. The sequence of each complementary site and siRNA are shown in the expanded regions. The position of primers used for 5′-RACE is indicated by horizontal arrows. The frequency of 5′-RACE clones corresponding to each cleavage site (vertical arrows) is shown as a fraction. "nd" indicates that no cleavage product was detected. The relative abundance of mRNAs or pre-mRNAs (At2g46740) in wild-type Col plants versus *ACTIN2* is indicated (r.a.). The wild-type value for *ACTIN2* is arbitrarily fixed to 100,000.

and three mismatches with siR396(+) (Figure 3). These four genes encode related proteins of unknown function (data not shown). In a search for potential targets of the other siRNAs, we found that siR522(+) and its variants, siR521(+) and siR523(+), have three mismatches with the intron 2 of At2g46740, a gene encoding an FAD binding domain-containing protein.

To determine if At2g27400 siRNAs direct cleavage of complementary mRNAs, we performed 5′-RACE on these predicted targets, as previously done for miRNA targets (Llave et al., 2002b). Cleavage products were amplified for the three most abundant predicted targets (Figure 3). Sequencing of these 5′-RACE products mapped the 5′ ends to the position corresponding to the middle of the siRNA complementarity (Figure 3), indicating that At5g18040, At4g29760, and At4g29770 mRNAs are targeted for cleavage by siR396(+), siR438(+), and/or siR480(+). We did not detect cleavage products for At1g51670, which was predicted to also be targeted by siR396(+), siR438(+), and/or siR480(+), and for At2g46740, which also was predicted to be targeted by siR522(+). The absence of cleavage products from At2g46740 may be due to the location of the complementarity site, which is in an intron. Pre-messenger RNAs are often short-lived molecules. Consistent with this observation, we found that At2g46740 pre-messenger RNA is much less abundant than the other mRNA targets for which we detected 5′ cleavage products (Figure 3). The absence of At1g51670 cleavage products could be due to the low expression of At1g51670 in adult leaves and flowers that were used to make the 5′-RACE cDNA pool (Figure 3).

To further investigate if the five mRNAs are regulated by the At2g27400 derived siRNAs, we quantified the steady-state level of uncleaved target mRNAs in mutants impaired in At2g27400-derived siRNA accumula-

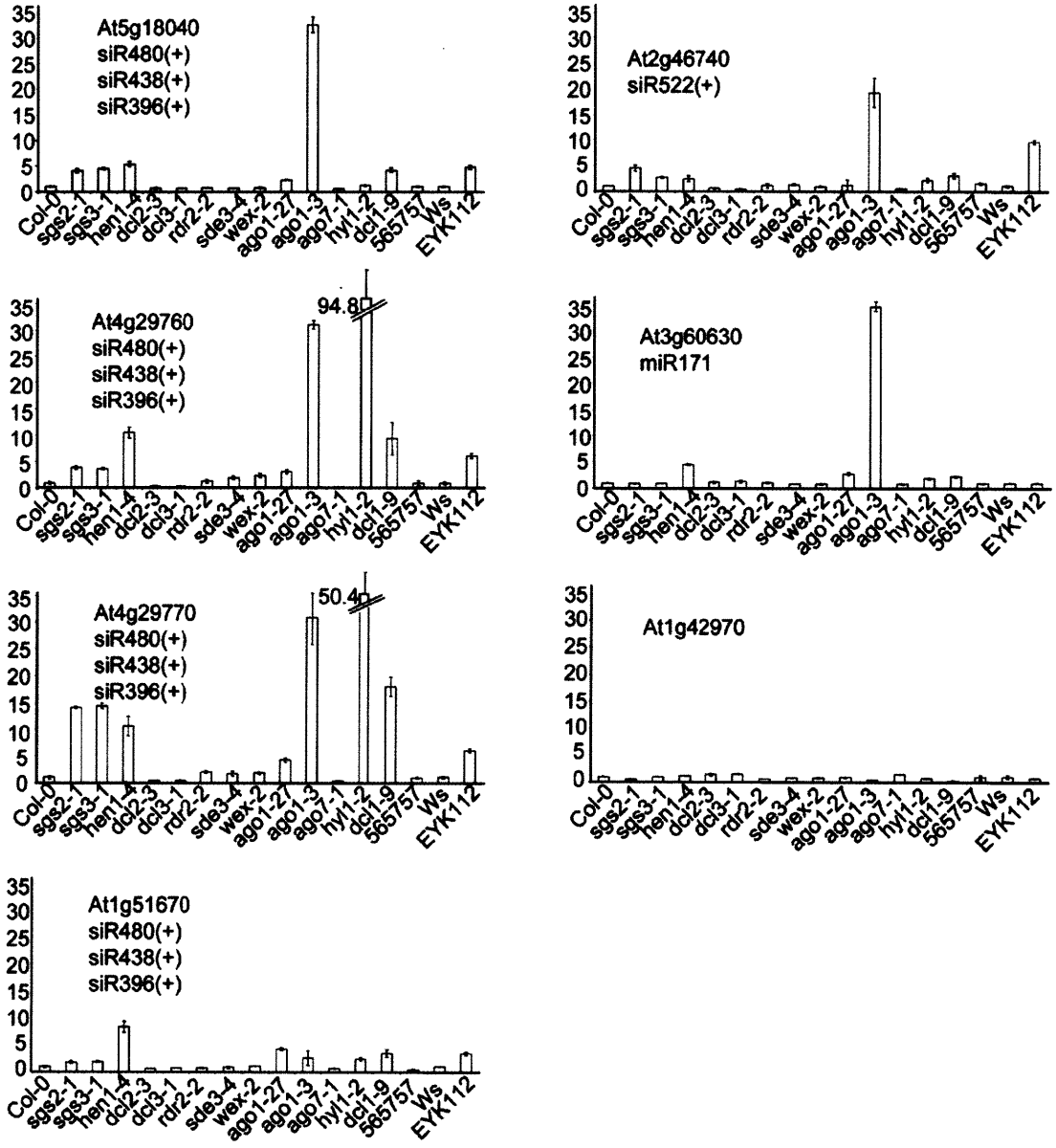


Figure 4. At2g27400 siRNAs Regulate the Expression of Target mRNAs

Steady-state levels of uncleaved target mRNAs was determined by RT-qPCR from total RNA extracted from the aerial part of 4-6 leaf-old plants using primers surrounding the siRNA complementary site. The At3g60630/*SCL6-III* miR171 target was used as a positive control and At1g42970/*GAPDH* as a negative, non-target, control. Quantifications were normalized to *ACTIN2*. The values in wild-type plants were arbitrarily fixed to 1. siRNA numbers and gene names corresponding to each pair are indicated at the left.

tion by RT-qPCR using primers surrounding the cleavage site. GAPDH (At1g42970) was used as a nontargeted mRNA control, while SCL6-III (At3g60630) was used as a miRNA target control. As expected for miRNA targets, uncleaved SCL6-III RNA levels increased in *dcl1*, *hen1*, *hyl1*, and *ago1* mutants but, importantly, not in *sgs2*(*rdr6*) and *sgs3* mutants (Figure 4). Uncleaved mRNA levels of At2g27400-derived siRNA targets were increased in *ago1, dcl1, hen1, and hyl1* mutants, as well as *rdr6, sgs3,* and the *EYK112* insertion mutant. Conversely, the accumulation of uncleaved mRNA was unchanged in *ago7*, *dcl2*, *dcl3*, *rdr2*, *sde3*, *wex*, and the *N565757* insertion mutant. These data are consistent with the change we observed in At2g27400-derived siRNAs accumulation (Figure 2) and indicate that At2g27400-derived siRNAs act to regulate endogenous mRNAs. As explained in the Discussion, we refer to this class of siRNAs as endogenous *trans*-acting siRNAs.

RDR6 and SGS3 Are Required for the At2g27400 siRNA Pathway but Have Limited or No Impact on the miRNA or Chromatin siRNA Pathways

The steady-state level of uncleaved SCL6-III mRNA that is a target of miR171 is unchanged in seedlings of *sgs2*(*rdr6*) and *sgs3* mutants (Figure 4), suggesting that SGS3 and RDR6 may be dispensable for miRNA-

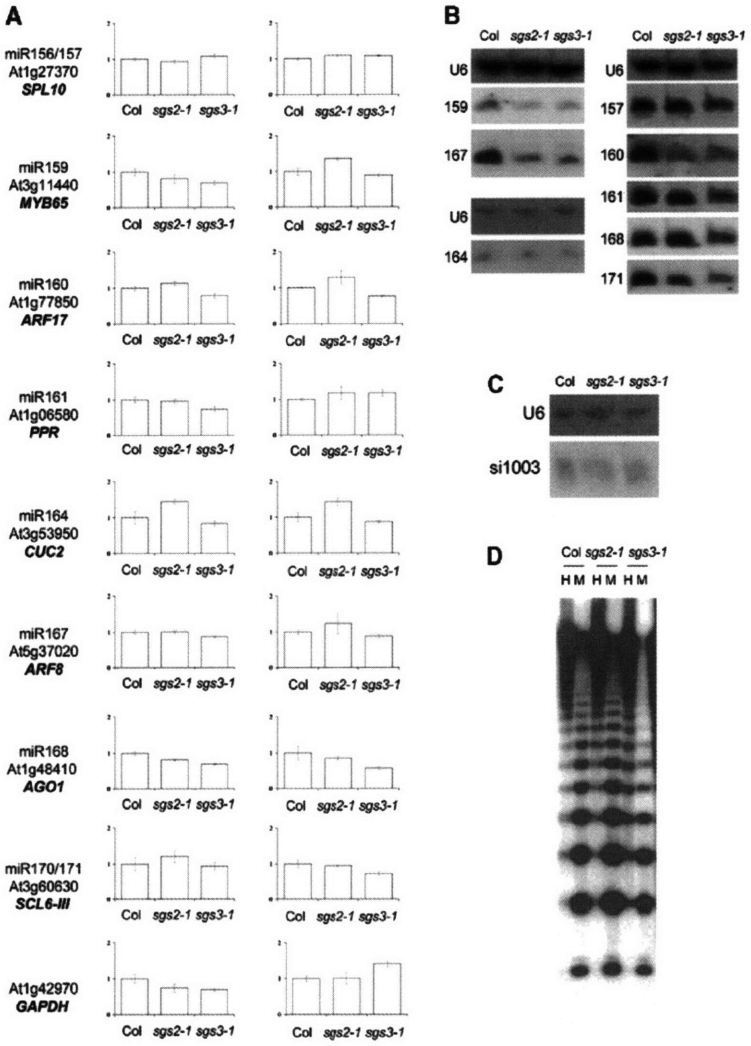


Figure 5. Analysis of miRNAs and Chromatin siRNAs in *rdr6* and *sgs3* Mutants

(A) *SPL10, MYB65, ARF17, PPR, CUC2, ARF8, AGO1*, and *SCL6-III* mRNA relative accumulation was determined in RNA extracted from inflorescences of wild-type plants (Col) and *sgs2*(*rdr6*) and *sgs3* mutants by RT-qPCR using primers surrounding the cleavage sites. RT-qPCR values determined for two independent cDNA synthesis are shown. Quantifications were normalized to *ACTIN2*. The wild-type Col value was arbitrarily fixed to 1. miRNA and target gene names are indicated on the left.

(B) RNA gel blots of 10 μg of RNA extracted from inflorescences of wild-type plants, *sgs2*(*rdr6*), and *sgs3* mutants were successively hybridized to specific miRNA probes, and finally hybridized to U6 as a loading control.

(C) RNA gel blot of 10 μg of RNA extracted from inflorescences of wild-type plants and *sgs2*(*rdr6*) and *sgs3* mutants was hybridized to an siRNA probe (ASRP1003) corresponding to 5S rDNA gene arrays and reprobed to U6 as a loading control.

(D) DNA gel blot analysis of DNA extracted from wild-type plants, *sgs2*(*rdr6*), *and sgs3* mutants. Genomic DNA was digested with the methylation-sensitive enzyme HpaII (H) or MspI (M) and hybridized to a 5S DNA probe.

directed cleavage. To test whether this applies at a different stage of development and for other miRNA targets, RT-qPCR was performed using RNAs extracted from inflorescences where most miRNA targets are abundantly expressed. No significant change in mRNA steady-state levels was observed between wild-type Col plants and *sgs2*(*rdr6*) or *sgs3* mutants (Figure 5A), indicating that RDR6 and SGS3 are indeed dispensable for miRNA-directed cleavage. Given this result, we expected miRNAs to accumulate in *sgs2*(*rdr6*) and *sgs3* mutants. Indeed, miRNAs accumulated in the mutants, although sometimes at a reduced level (Figure 5B), suggesting that this level is sufficient for proper miRNA-directed cleavage of the analyzed targets.

A previous report indicated that RDR6 and SGS3 are not required for the accumulation of 24 nt siRNAs derived from the retroposon AtSN1 or for DNA methylation of AtSN1 elements (Hamilton et al., 2002). To further analyze if mutations in RDR6 or SGS3 could impact the heterochromatic siRNA pathway, we looked at the accumulation of the 21 to 24 nt siRNAs deriving from tandem arrays of 5S rDNA genes, and at the methylation status of 5S DNA, which both require DCL3 and RDR2 (Xie et al., 2004). No change in 5S siRNA accumulation or 5S DNA methylation was observed in *sgs2*(*rdr6*) or *sgs3* mutants (Figures 5C and 5D), indicating that RDR6 and SGS3 are not required for proper functioning of the heterochromatic siRNA pathway and suggesting that the AGO1-DCL1-HEN1-HYL1-SGS3-RDR6-dependent siRNA pathway defined here is an endogenous posttranscriptional pathway.

Discussion

We identified a class of endogenous 21 nt regulatory siRNAs in plants that have similarities to but also differences from the previously characterized small silencing RNAs of plants. As is typical of plant miRNAs (Bartel, 2004), they are endogenous 21 nt regulatory RNAs, appear to derive from short double-stranded products of Dicer-like enzymes, appear to obey the asymmetry rules that determine which of the two strands of each short duplex enters RISC and accumulates, regulate endogenous target mRNAs in *trans* by directing cleavage in the middle of a single complementary site, target multiple members within a single gene family, and require AGO1,

DCL1, HEN1, and HYL1 for their biogenesis, accumulation, or function. In contrast to miRNAs, they derive from long double-stranded RNA, and their accumulation depends on RDR6 and SGS3. In these two respects, they resemble the siRNA mediators of S-PTGS, triggered by transgenes, and thus their characterization reveals the first endogenous regulatory role for RDR6 and SGS3. They differ from heterochromatic siRNAs, the other class of endogenous siRNAs that have been characterized in plants, in their posttranscriptional mode of silencing and the proteins required for their biogenesis, accumulation, or function. They differ from all siRNAs that have been described in plants in that they target genes that bear little resemblance to the genes from which the siRNAs derive, a function sometimes called "hetero-silencing" (Bartel, 2004), previously attributed only to miRNAs.

The siRNAs we describe here also differ from all endogenous siRNAs that have been described in animals, which like those previously found in plants, are known or suspected to mediate "auto-silencing," i.e., the silencing of the same locus (or very similar loci) from which they originate. For example, endogenous siRNAs perform important posttranscriptional regulation in *Drosophila*, where male fertility depends on the silencing of *Stellate* by endogenous siRNAs from the *Su(Ste)* repeats (Aravin et al., 2001). As a host gene for siRNAs, the *Su(Ste)* repeats differ from the source of the siRNAs we describe here in that the *Su(Ste)* repeats have extensive homology (greater than 1 kb of 90% identity) with their regulatory target, and their dsRNA product is generated by convergent transcription. Candidates for endogenous siRNAs that might mediate posttranscriptional silencing have also been reported in nematodes (Lau et al., 2001; Ambros et al., 2003; Lim et al., 2003). These are cloned ~21 nt RNAs that have perfect antisense complementarity to mRNAs of *C. elegans*. Although characterization of their biogenesis and potential functions has not been reported, it is reasonable to assume that these putative siRNAs ultimately derive from the same locus as their presumed targets, either through antisense transcription or an RdRp activity, because they generally match only a single locus in the genome.

Because of their unusual role in hetero-silencing, we refer to the new class of siRNAs identified here as endogenous "*trans*-acting siRNAs." Of course, all siRNAs are *trans*-acting in the sense that they are not part of the silenced molecule but are instead part of a ribonucleoprotein complex that recognizes the *cis*-acting siRNA complementary site. The newly identified siRNAs get their name because they are *trans*-acting in an additional sense—they direct hetero-silencing, repressing the expression of genes that bear little resemblence to the genes from which they derive. In this sense, other endogenous siRNAs that have been characterized are *cis*-acting, performing auto-silencing to repress the expression of genes that are the same or very similar to the loci from which they derive.

The endogenous *trans*-acting siRNAs analyzed in this paper derive from the At2g27400 gene, which likely produces RNA as its final product, in that it has no known or deduced function other than to serve as a host gene for the production of siRNAs. Among the small RNAs deriving from the sense strand, two nearly identical ones, siR480(+) and siR438(+), were previously reported as miRNAs (Table 1), annotated as miR175a-2/miR389b.1 and miR389b.3, respectively (Park et al., 2002; Sunkar and Zhu, 2004). Furthermore, siR396(+) was predicted to be a miRNA based on its similarity to siR480(+) (Park et al., 2002; Sunkar and Zhu, 2004). A predicted fold of the At2g27400 transcript does place siR396(+) in a hairpin with some resemblance to those of the known miRNAs (see Supplemental Figure S1 at http://www.molecell.org/cgi/content/full/16/1/69/DC1/), and in different suboptimal predicted folds, the segment that pairs to siR396(+) would presumably pair to siR438(+) and siR480(+), the two other clones that have been previously annotated as miRNAs. However, these potential stem loops differ subtly from those of the typical plant miRNAs. For the potential stem loops involving siR396(+), siR438(+), and siR480(+) pairing ends abruptly at both termini of the cloned RNA, whereas for plant miRNAs pairing typically extends, without interruption, beyond either the 5′ or 3′ terminus of the mature miRNA (Jones-Rhoades and Bartel, 2004). More convincing evidence comes when considering the siR396(+), siR438(+), and siR480(+) clones in the context of all the other clones that have now been found from the At2g27400 region. These three clones occur in 21 nt phasing with nearly all of the other At2g27400 siRNAs that have been found in wild-type plants (Table 1; Figure 1), including an siRNA from the antisense strand of the transcript, and like the other siRNAs, they do not accumulate in *rdr6* and *sgs3* mutants (Figure 2). They have the hallmarks of deriving from the processing of a long dsRNA, likely formed by the RNA-dependent RNA polymerase activity of RDR6, into multiple small RNA duplexes. siRNAs are produced by the Dicer-like processing of a long double-stranded precursor into multiple small RNA duplexes, whereas miRNAs are produced by the Dicer-like processing of each stem loop precursor into a single miRNA duplex (Bartel, 2004). Accordingly, we classified siR396(+), siR438(+), and siR480(+) as siRNAs, and attribute the location of siR396(+) within a predicted stem loop (Supplemental Figure S1) as a chance coincidence, perhaps not even an improbable one, given the observation that nearly 400,000 intergenic *Arabidopsis* 20-mers fall within miRNA-like stem loops (Jones-Rhoades and Bartel, 2004).

A mutant carrying a T-DNA inserted within the At2g27400 intron no longer accumulates the At2g27400 siRNAs (Figure 2). The coordinate reduction of siR396(+) with the other siRNAs further supports the hypothesis that all the cloned 21 nt RNAs derive from a common double-stranded RNA precursor. This insertion, which lies downstream of all the At2g27400 siRNAs (Figure 1), is not expected to influence the formation of stem loops (Supplemental Figure S1) but does truncate the 3′ sequence of the transcript (data not shown), perhaps compromising the RDR6-dependent conversion of the transcript into dsRNA. In contrast, siRNAs accumulate to wild-type levels in a mutant carrying a T-DNA inserted in exon2, suggesting that the elements needed for RDR6 recognition might reside at the 3′ end of either the spliced intron or a truncated unspliced transcript terminating at an alternative polyadenylation site located upstream of the 3′ end of the intron (Figure 1). The identification of an endogenous transcript that serves as a

template for RDR6-mediated production of dsRNA as well as the T-DNA lesion that might prevent RDR6 recognition could provide reagents for exploring the molecular basis of posttranscriptional gene silencing mediated by sense transgenes (S-PTGS/cosuppression). S-PTGS differs from IR-PTGS mediated by hairpin transgenes in that it requires the conversion of transgene-derived RNA into dsRNA (Beclin et al., 2002). It has been suggested that S-PTGS is triggered by "aberrant" RNAs transcribed from certain transgene loci (Dalmay et al., 2000; Mourrain et al., 2000); however, the features that define the transcripts as aberrant and therefore substrates for the production of dsRNA have been a mystery.

At2g27400 siR396(+), siR438(+), and/or siR480(+) regulate the expression of at least four genes encoding related proteins (Figure 4), all of which have siRNA complementary sites within their mature transcripts (Figure 3; Park et al., 2002, Sunkar and Zhu, 2004). In contrast, At2g27400 siR522(+) exhibits near-complementarity to only one target (At2g46740) and the siRNA complementary site is located within an intron. The overaccumulation of both At2g46740 spliced and unspliced transcripts in *rdr6* and *sgs3* mutants and in the *EYK112* insertion mutant (Figure 4; data not shown), which all have reduced At2g27400 siR522(+) accumulation (Figure 2), hints at the idea that siRNAs could guide cleavage within the introns of pre-mRNAs. Consistent with the intron-targeting hypothesis, the level of both At2g27400 unspliced and spliced transcripts was elevated in *rdr6* and *sgs3* mutants (data not shown), suggesting that antisense siRNAs derived from At2g27400 guide cleavage of the unspliced At2g27400 transcript, providing a possible feedback regulatory loop for the production of these siRNAs. However, this intriguing possibility of posttranscriptional regulation in the nucleus could not be corroborated by 5′-RACE experiments, perhaps because of the low abundance of the pre-mRNAs (Figure 3). Evidence for siRNA/miRNA targeting of intronic sequence within pre-mRNA has been reported only once in animals (Bosher et al., 1999), perhaps because efficiently spliced messages are not suitable targets for this type of regulation. Nuclear targeting of partially spliced *Arabidopsis DCL1* RNA by miR162 has been suggested (Xie et al., 2003), making our hypothesis of targeting of nuclear RNAs by siRNAs more plausible.

It is possible that additional *trans*-acting siRNAs exist in *Arabidopsis*, as suggested by other small RNA clones with clustered loci in the genome (Llave et al., 2002a; Reinhart et al., 2002). This raises the possibility that many endogenous mRNAs could be regulated posttranscriptionally by this class of siRNAs. The role of endogenous genes regulated by *trans*-acting siRNAs is yet unclear, whereas most of the genes regulated by miRNAs play important roles in plant development (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004). Indeed, *ago1*, *dcl1*, *hen1*, and *hyl1* mutants exhibit dramatic developmental defects (Bohmert et al., 1998; Jacobsen et al., 1999; Lu and Fedoroff, 2000; Chen et al., 2002), some of which have been confirmed to be due to the lack of miRNA regulation (Aukerman and Sakai 2003; Emery et al. 2003; Palatnik et al. 2003; Chen 2004; Mallory et al., 2004a, 2004b; Vaucheret et al., 2004). In contrast, *rdr6* and *sgs3* mutants exhibit a much milder phenotype, mostly characterized by downward curling of the leaf margin (Figure 6). This phenotype is very similar to that of *ago7* (*zippy*) mutants (Figure 6), which exhibit an early transition into the adult phase of vegetative development (Hunter et al., 2003); however, it is unlikely that misregulation of the At2g27400 siRNAs contribute to the phenotype of *ago7*, *rdr6*, and *sgs3* mutants. Indeed, the T-DNA insertion mutant EYK112 (Figure 1) that exhibits a strong reduction in the accumulation of these siRNAs (Figure 2) and increased accumulation of uncleaved target mRNAs (Figure 4) has no obvious phenotype (Figure 6). In addition, both siRNA accumulation and target mRNA accumulation are not affected by mutations in AGO7 (Figures 2 and 4), indicating that AGO7 is not required for this pathway, at least at this stage of development, and suggesting that the developmental defects of *ago7*, *rdr6*, and *sgs3* mutants could be due to the impairment of yet another siRNA pathway that remains to be discovered. Given the slight decrease in the accumulation of some miRNAs in *rdr6* and *sgs3* mutants (Figure 5), the developmental defects of these mutants could also be due to misregulation of miRNA targets not examined in our analysis.

Is the AGO1-DCL1-HEN1-HYL1-SGS3-RDR6-dependent pathway, which produces endogenous *trans*-acting siRNAs, an ancient pathway from which the miRNA

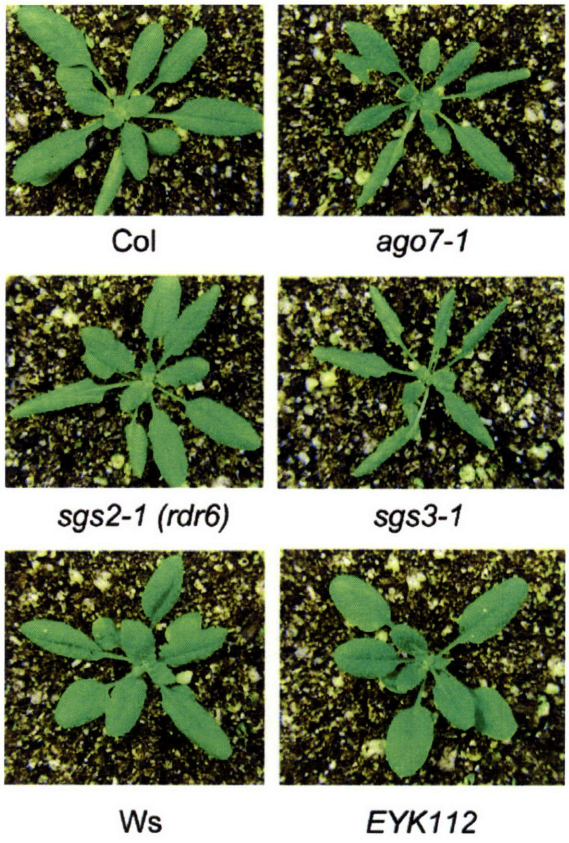


Figure 6. *rdr6* and *sgs3* Mutants Exhibit Developmental Defects Similar to *ago7*

Wild-type plants (Col or Ws), *sgs2-1(rdr6)*, *sgs3-1*, and *ago7-1(zip)* mutants (in Col) and the *EYK112* mutant (in Ws) were grown in soil for 25 days. *sgs2-1*, *sgs3-1*, and *ago7-1* mutants but not *EYK112* exhibit similar downward curling of the leaf margin.

pathway derived, or is it a more recent pathway? Whereas most miRNA/target pairs are conserved between *Arabidopsis* and rice (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004), we have not detected such conservation among the siRNA/target pairs described here. In addition, we do not find any matches to the At2g27400 siRNAs (0–2 substitutions) or At2g27400 transcript in EST databases of other plant species, including rice and *brassica*. This suggests that the At2g27400 siRNAs have emerged quite recently, and that the pathway as a whole might have recent origins. However, a few known miRNAs also appear to be of very recent origin (Jones-Rhoades and Bartel, 2004), and thus a more comprehensive list of AGO1-DCL1-HEN1-HYL1-SGS3-RDR6-dependent siRNAs will be needed to address the questions of which pathway might be more ancient and the degree to which the two pathways, which have such similar regulatory outcomes, might have overlapping origins.

Experimental Procedures

Plant Material

ago1-3, ago1-27, dcl1-9, dcl3-1, hen1-4, sgs2-1(rdr6) and *sgs3-1* mutants (all in the Col-0 ecotype) have been previously described (Bohmert et al., 1998; Elmayan et al., 1998; Jacobsen et al., 1999; Mourrain et al., 2000; Morel et al., 2002; Boutet et al., 2003; Vazquez et al., 2004; Xie et al., 2004). The *dcl1-9* mutant in Col was obtained by five backcrosses of the original *dcl1-9* mutant (Jacobsen et al., 1999) to Col. *ago7-1* (SALK_037458), *dcl2-3* (SALK_095069), *rdr2-2* (SALK_059661), *sde3-4* (SALK_092019), *wex-2* (SALK_003278) and the mutant carrying a T-DNA inserted in exon2 of the At2g27400 gene (SALK_065757) were identified in the collection of SALK mutants (all in the Col-0 ecotype) available from the *Arabidopsis* Stock Center (Alonso et al., 2003). The EYK112 mutant carrying a T-DNA inserted in the intron of the At2g27400 gene (FLAGDB_356B08) was identified in the INRA Versailles T-DNA collection of mutants (in the Ws ecotype) (Samson et al., 2002).

The molecular characterization of plant material, the methods for cDNA-AFLP profiling, real-time quantitative RT-PCR, RNA extraction and hybridization, modified 5′RLM-RACE for cloning of cleavage products, and 3′-RACE and 5′ RLM-RACE for cloning of full-length At2g27400 transcript are described in the Supplemental Data.

Acknowledgments

We thank the SALK Institute and INRA for T-DNA insertion mutants, Pascale Goupil, Sylvain Legrand, Marie-Christine Quillet, Serge Rambour, and Jacques Vasseur for fruitful discussions, Philippe Grappin for his cDNA-AFLP protocol, Tiphaine Lefevre, David Windels, and Dorothée Deprost for technical assistance, and Taline Elmayan for comments on the manuscript and help. This work was partly supported by a grant from the French minister of Research (ACI number 0220580) to P.C. and H.V., a PhD fellowship from the French Minister of Research to F.V., a grant of the European Commission (Riboreg program) to H.V., and NIH grants to A.C.M. and D.P.B.

Received: July 1, 2004
Revised: September 15, 2004
Accepted: September 16, 2004
Published online: September 30, 2004

References

Alonso, J.M., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301*, 653–657.

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

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr. Biol. *11*, 1017–1027.

Aukerman, M.J., and 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.

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

Bartel, B., and Bartel, D.P. (2003). MicroRNAs: at the root of plant development? Plant Physiol. *2*, 709–717.

Beclin, C., Boutet, S., Waterhouse, P., and Vaucheret, H. (2002). A branched pathway for transgene-induced RNA silencing in plants. Curr. Biol. *12*, 684–688.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. EMBO J. *17*, 170–180.

Bosher, J.M., Dufourcq, P., Sookhareea, S., and Labouesse, M. (1999). RNA interference can target pre-mRNA: Consequences for gene expression in a *Caenorhabditis elegans* operon. Genetics *153*, 1245–1256.

Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J.B., Crete, P., Chen, X., and Vaucheret, H. (2003). *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr. Biol. *13*, 843–848.

Boutla, A., Delidakis, C., and Tabler, M. (2003). Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in *Drosophila* and the identification of putative target genes. Nucleic Acids Res. *31*, 4973–4980.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell *113*, 25–36.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science *301*, 336–338.

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

Chen, X., Liu, J., Cheng, Y., and Jia, D. (2002). *HEN1* functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. Development *129*, 1085–1094.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543–553.

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

Elmayan, T., Balzergue, S., Beon, F., Bourdon, V., Daubremet, J., Guenet, Y., Mourrain, P., Palauqui, J.C., Vernhettes, S., Vialle, T., et al. (1998). Arabidopsis mutants impaired in cosuppression. Plant Cell *10*, 1747–1758.

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

Finnegan, E.J., and Matzke, M.A. (2003). The small RNA world. J. Cell Sci. *116*, 4689–4693.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. *21*, 4671–4679.

Han, M.H., Goud, S., Song, L., and Fedoroff, N. (2004). The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc. Natl. Acad. Sci. USA *101*, 1093–1098.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Hunter, C., Sun, H., and Poethig, R.S. (2003). The Arabidopsis heterochronic gene ZIPPY is an ARGONAUTE family member. Curr. Biol. *13*, 1734–1739.

Jacobsen, S.E., Running, M.P., and Meyerowitz, E.M. (1999). Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. Development *126*, 5231–5243.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature *426*, 845–849.

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.

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.

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

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

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

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

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

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

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

Lu, C., and Fedoroff, N. (2000). A mutation in the *Arabidopsis* HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell *12*, 2351–2366.

Mallory, A.C., and Vaucheret, H. (2004). MicroRNAs: something important between the genes. Curr. Opin. Plant Biol. *7*, 120–125.

Mallory, A.C., Dugas, D.V., Bartel, D.P., and 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, A.C., Reinhart, B.J., Jones-Rhoades, M., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. (2004b) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. EMBO J. *23*, 3356–3364.

Morel, J.B., Godon, C., Mourrain, P., Beclin, C., Boutet, S., Feuerbach, F., Proux, F., and Vaucheret, H. (2002). Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell *14*, 629–639.

Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. Cell *88*, 637–646.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., et al. (2000). *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell *101*, 533–542.

Muangsan, N., Beclin, C., Vaucheret, H., and Robertson, D. (2004). Geminivirus VIGS of endogenous genes requires SGS2/SDE1 and SGS3 and defines a new branch in the genetic pathway for silencing in plants. Plant J. *6*, 1004–1014.

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

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr. Biol. *12*, 1484–1495.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901–906.

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.

Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., and Lecharny, A. (2002). FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of Arabidopsis thaliana T-DNA transformants. Nucleic Acids Res. *30*, 94–97.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in *Arabidopsis* development. Trends Plant Sci. *7*, 487–491.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *2*, 199–208.

Sunkar, R., and Zhu, J-K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell online (www.plantcell.org/cgi/doi/10.1105/tpc.104.022830.)

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. *17*, 49–63.

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

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

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell *75*, 855–862.

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(5): e104 DOI:10.1371/journal.pbio.0020104.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. *13*, 790–795.

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

Cell

A Two-Hit Trigger for siRNA Biogenesis in Plants

Michael J. Axtell,[1,2,3] Calvin Jan,[1,2] Ramya Rajagopalan,[1,2] and David P. Bartel[1,2,*]
[1] Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[3] Present address: Department of Biology, The Huck Institutes of the Life Sciences, Pennsylvania State University, 101 LSB., University Park, PA 16802, USA.
*Contact: dbartel@wi.mit.edu
DOI 10.1016/j.cell.2006.09.032

SUMMARY

In *Arabidopsis*, microRNA-directed cleavage can define one end of RNAs that then generate phased siRNAs. However, most miRNA-targeted RNAs do not spawn siRNAs, suggesting the existence of additional determinants within those that do. We find that in moss, phased siRNAs arise from regions flanked by dual miR390 cleavage sites. *AtTAS3*, an siRNA locus important for development and conserved among higher plants, also has dual miR390 complementary sites. Both sites bind miR390 in vitro and are functionally required in *Arabidopsis*, but cleavage is undetectable at the 5′ site—demonstrating that noncleavable sites can be functional in plants. Phased siRNAs also emanate from the bounded regions of every *Arabidopsis* gene with two known microRNA/siRNA complementary sites, but only rarely from genes with single sites. Therefore, two "hits,"—often, but not always, two cleavage events—constitute a conserved trigger for siRNA biogenesis, a finding with implications for recognition and silencing of aberrant RNA.

INTRODUCTION

MicroRNAs and endogenous siRNAs modulate the transcriptomes of both animals and plants (Bartel, 2004; Zamore and Haley, 2005). MicroRNAs are ~21-nt RNAs that are processed from primary transcripts with characteristic stem-loop secondary structures. siRNAs are processed from long double-stranded RNA (dsRNA) formed by convergent transcription, RNA-dependent RNA polymerization, or extended hairpin structures. Both miRNAs and siRNAs are incorporated into silencing complexes where they pair to target transcripts and mediate association between the silencing complexes and targeted transcripts.

Most characterized interactions involving miRNAs or siRNAs result in the negative regulation of the target, either at the transcriptional level (heterochromatic siRNAs) or at the posttranscriptional level (siRNAs and miRNAs; Bartel, 2004). Plant miRNAs have extensive pairing to their targets (Rhoades et al., 2002), and as a result, posttranscriptional repression generally is caused by mRNA cleavage (Llave et al., 2002; Tang et al., 2003). An interesting exception to the tendency to mediate repression comes with the finding that miRNA-directed cleavage is needed for the production of *trans*-acting siRNAs (tasiRNAs; Allen et al., 2005; Yoshikawa et al., 2005), which differ from classical siRNAs in that they silence messages from loci that are unrelated to those from which the siRNAs derive (Peragine et al., 2004; Vazquez et al., 2004b). Each tasiRNA locus (known as a *TAS* gene) produces a non-protein-coding transcript, a portion of which is converted by the RNA-dependent RNA polymerase (RdRp) RDR6 into dsRNA that in turn is successively cleaved into mostly 21-nt siRNAs by DCL4, a Dicer-like enzyme (Peragine et al., 2004; Vazquez et al., 2004b; Allen et al., 2005; Gasciolli et al., 2005; Xie et al., 2005; Yoshikawa et al., 2005). All known *Arabidopsis TAS* (*AtTAS*) loci have miRNA complementary sites at which miRNA-directed cleavage appears to define one end of the dsRNA intermediate, thereby setting the register of phased tasiRNA production (Allen et al., 2005; Yoshikawa et al., 2005). miR173 directs cleavage upstream of the *AtTAS1a-c* and *AtTAS2* siRNAs, whereas miR390 directs cleavage downstream of the *AtTAS3* siRNAs. This role for miRNAs in setting a phasing register for tasiRNA biogenesis is important because siRNAs produced in most other registers would not have sufficient homology to direct cleavage of their mRNA targets. The coupling of miRNA-mediated cleavage to RdRp activity and siRNA production for the *AtTAS* loci raises the question of why such transitivity is very rare for the majority of non-*TAS* miRNA targets (Lu et al., 2005), which also undergo miRNA-directed cleavage. It seems likely that endogenous miRNA targets that enter the *TAS* pathway possess additional, as of yet unknown molecular features beyond a single miRNA complementary site.

In this study, we adapt high-throughput DNA sequencing to the discovery of endogenous small RNAs from the

moss *Physcomitrella patens* and *Arabidopsis*. Our data demonstrate that *P. patens* has four loci that give rise to phased siRNAs resembling tasiRNAs. These moss siRNAs are in phase with cleavage sites for miR390, the same miRNA important for tasiRNA phasing in flowering plants (Allen et al., 2005). We find that the four moss siRNA loci are each flanked by dual miR390 complementary sites that are both cleaved in vivo. *AtTAS3* also has a second, conserved miR390 complementary site that had not been originally recognized; as in *P. patens*, the region that generates phased siRNAs falls between the two miR390 complementary sites. The newly identified upstream miR390 complementary site of *AtTAS3* is not cleaved, even though it binds an miR390-associated silencing complex in vitro and is necessary for full *AtTAS3* function in vivo, which indicates that conserved miRNA complementary sites can function independently of target cleavage in plants. We also find that *Arabidopsis* genes with two or more small RNA complementary sites universally produce phased siRNAs from the regions that fall between the sites. Our results indicate that dual miRNA complementary sites are a trigger for siRNA biogenesis that has been conserved for the past 400 million years and provide potential insights into the recognition and silencing of aberrant RNAs.

RESULTS

Endogenous Small RNAs from Moss

Small RNA was prepared from three developmentally staged samples of the moss *P. patens* and sequenced using either the standard di-deoxy method (Lau et al., 2001) or a recently described pyrosequencing technology (Margulies et al., 2005) to yield a total of 561,102 small RNA reads, representing 214,996 unique sequences (see Table S1 in the Supplemental Data). To begin to categorize these sequences, we compared them to the 5.4 million traces available at the time of our analysis from the *P. patens* whole-genome shotgun (WGS) project. A total of 127,135 unique small RNA sequences, represented by 384,441 reads, had at least one perfect match to the WGS traces (Table S1). The remaining sequences that did not perfectly match any WGS trace were not analyzed further and were presumed to be sequences from genomic regions missed by the WGS project, sequencing errors, or unclassified contaminant sequences. Some of the genome-matched sequences were classified as miRNAs (M.J.A. and D.P.B., unpublished data). These accounted for 42% of the genome-matched reads, leaving 58% that did not appear to arise from loci with the characteristics of known miRNAs, suggesting that, similar to *Arabidopsis*, *P. patens* might express many endogenous siRNAs. Among these potential siRNAs, 18% (39,975 reads representing 19,974 sequences) matched more than 1,000 traces, which would correspond to more than 100 endogenous loci; these were designated as repetitive small RNAs (Table S1). Once a genome assembly is available, it will be interesting to evaluate whether repeat elements are more or less likely than other regions of the genome to give rise to moss small RNAs.

The Antiquity of tasiRNAs

The first characterized tasiRNA loci showed no sign of conservation beyond *Arabidopsis* (Peragine et al., 2004; Vazquez et al., 2004b). However, two related tasiRNAs from the *AtTAS3* locus, required for the proper timing of vegetative development and regulation of organ polarity (Adenot et al., 2006; Fahlgren et al., 2006; Garcia et al., 2006; Hunter et al., 2006), are conserved among diverse flowering plants (Allen et al., 2005; Williams et al., 2005). To explore the possibility that tasiRNAs might have emerged much earlier in plant evolution, we searched the nonrepetitive WGS moss traces for clustered small RNA hits in both sense and antisense orientation, wherein a significant fraction were phased in ~21-nt increments. Four candidate tasiRNA loci were found (*PpTAS1-4*), one of which was represented by 15,730 reads—4.1% of all of the nonrepetitive, genome-matched reads (Figure 1A). Indeed, a cDNA corresponding to this locus (*PpTAS1*) had previously been suspected of being a tasiRNA precursor based on the sequencing of three corresponding small RNAs (Arazi et al., 2005).

All known *Arabidopsis* tasiRNA loci have miRNA complementary sites that are thought to be important for setting the register of phased siRNA production (Allen et al., 2005; Yoshikawa et al., 2005). We found that all four moss *TAS* loci had complementary sites for miR390 positioned so as to set the phasing register of the dominant siRNA species (Figure 1). miR390 is conserved throughout land plants (Axtell and Bartel, 2005) and is the same miRNA that sets the phasing register for the *AtTAS3* siRNAs (Allen et al., 2005). *PpTAS2* had one siRNA in common with *PpTAS1* and another in common with *PpTAS3*, but otherwise the four *P. patens TAS* loci shared little sequence identity beyond their miR390 complementary sites (Figure S1). Moreover, no sequences resembling the *Arabidopsis* tasiRNAs were discerned in any of the four *PpTAS* loci, begging the question as to the evolutionary relationship between the *TAS* loci of moss and those of higher plants. Nonetheless, the presence of miR390 complementary sites in the *PpTAS* loci supported the idea that these four loci were indeed tasiRNA genes and suggested that miR390 has been setting the phasing register of tasiRNAs since the last common ancestor of moss and angiosperms.

Surprisingly, each of the *PpTAS* loci contained not one but two miR390 complementary sites, one upstream and another downstream of the siRNA-corresponding region (Figure 1). The phasing register of the siRNAs was consistent with cleavage at both complementary sites: 71.3% of the *PpTAS*-derived small RNAs began within one nucleotide of the residues predicted by cleavage at either the 5′ or 3′ site. We confirmed miR390-mediated cleavage at both sites of *PpTAS1* using 5′-RACE (Llave et al., 2002); 11 out of 17 sequenced cleavage products had 5′ residues corresponding precisely to those predicted by

cleavage at either the 5′ or 3′ miR390 complementary site (Figure 1A). These observations indicate that tasiRNA biogenesis from *PpTAS* loci is triggered by two miR390-directed cleavage events that together define the intervening cleavage product as a substrate for subsequent RdRp and Dicer activity.

Dual miR390 Complementary Sites Are a Conserved Feature of tasiRNA Precursors

Having found two miR390 complementary sites in each of the four moss *TAS* loci, we examined the *AtTAS3* locus to see if it might also have two sites. A second miR390 site was found, located upstream of the tasiRNA region (Figure 2A). Pyrosequencing of small RNAs from wild-type *Arabidopsis* inflorescences, leaves, seedlings, and siliques yielded 887,266 reads that matched the *Arabidopsis* genome (R.R. and D.P.B., unpublished data), of which 1,806 were *AtTAS3*-derived tasiRNAs. Many were in phase with the 3′ cleavage site, but a sizable proportion, particularly from the center and 5′ region of the locus, were not (Figure 2A). These alternatively phased siRNAs were largely in a register predicted by miR390-directed cleavage at the newly identified 5′ site. However, these alternatively phased siRNAs were almost equally consistent with the register that would be set by the most abundant *TAS3* siRNA, TAS3 5′D2(−) (Figure 2A). Because the newly identified 5′ miR390 complementary site in *AtTAS3* was unable to direct miR390-directed cleavage (see below), we favor the hypothesis put forward by Allen et al. (2005) that the alternatively phased *AtTAS3* tasiRNAs are phased by TAS3 5′D2(−)-mediated cleavage.

EST sequences representing *AtTAS3* homologs from diverse seed plants have two reported regions of nucleotide conservation: an ~42-nt region corresponding to the *AtTAS3* tasiRNAs that target *ARF3* and *ARF4* (tasi*ARF*s) and an ~21-nt region corresponding to the 3′ miR390 complementary site (Allen et al., 2005). We found a third conserved region of seed-plant *AtTAS3* homologs, which corresponded to the 5′ miR390 complementary site (Figure 2B), suggesting that dual targeting of tasiRNA precursors is an evolutionarily conserved function of miR390. For most of the *AtTAS3* homologs, cleavage at the 3′ miR390 complementary site would set the phasing register required for the accurate production of the two conserved tasi*ARF*s, which explains why the length of the region between tasi*ARF*s and the 3′ miR390 complementary site is relatively constant (Figure 2B; Allen et al., 2005). In contrast, the lengths of the regions between the newly identified 5′ miR390 complementary sites and the tasi*ARF*s were variable and generally out of phase with the tasi*ARF*s (Figure 2B). The exception was the sole gymnosperm *AtTAS3* homolog; for the loblolly pine (*Pinus taeda*) *TAS3* homolog, the 5′ miR390 complementary site was the one that was in the proper register for directing accurate production of the tasi*ARF*s, whereas the 3′ site was nine nucleotides out of phase.

Cleavage-Independent Function of a Conserved Plant miRNA Complementary Site

The 5′ miR390 complementary site of *AtTAS3* was unusual for a plant miRNA target in that it contained a mismatch and two G:U wobbles involving nucleotides 9–11 of miR390 (Figure 2A). Such mismatches in nucleotides surrounding the potential scissile phosphate inhibit miRNA-directed endonucleolytic cleavage in vitro and in vivo (Mallory et al., 2004; Schwab et al., 2005). Mismatches involving positions 9–11 of miR390 were a conserved feature of the 5′ sites of *TAS3* homologs and were in stark contrast to the pairing preferences for the 3′ sites of the same homologs (Figure 2C) and those observed for plant miRNA complementary sites in general (Mallory et al., 2004). 5′ RACE failed to detect cDNA ends terminating within the upstream site (data not shown), which suggested that the conserved, noncanonical 5′ miR390 complementary sites of flowering-plant *TAS3* genes might function independently of target cleavage.

We next tested, using wheat-germ extract (Tang et al., 2003), the biochemical properties of the miR390 complementary sites. A substrate containing the 5′ miR390 complementary site of a moss *TAS* gene (*PpTAS3*) was cleaved in vitro (Figure 3A). Similarly, the 5′ complementary site from the gymnosperm homolog (*PtTAS3*) was efficiently cleaved despite the presence of a mismatch at position 10 (Figure 3A). Cleavage at the *PtTAS3* 5′ site demonstrated that some cleavage targets could be missed when using target-prediction guidelines, such as those of Schwab et al. (2005), that forbid mismatches at position 10. In contrast, cleavage of the 5′ site of *AtTAS3*, which lacked Watson-Crick pairing at positions 9, 10, and 11, was not detected under conditions in which cleavage of the 3′ site was observed (Figures 3A and 3B). Mutations disrupting the 3′ site abolished cleavage, whereas adding additional miR390 to increase the amount of miR390-programed silencing specifically enhanced cleavage (Figure 3B). Repairing the mismatches to nucleotides 9–11 of miR390 resulted in efficient cleavage at the 5′ site, indicating that comparable cleavage of the wild-type 5′ site would have been detected had it occurred (Figure 3B).

Cleavage of the *PtTAS3* 5′ site provided an explanation for why the pine tasi*ARF*s were in phase with the 5′ miR390 complementary site rather than the 3′ site (Figure 2B); because the *PtTAS3* site can be cleaved, it could be the site that sets the phasing register for tasi*ARF* production in pine. As a corollary, the mismatches found in the 5′ site of *AtTAS3* and conserved among flowering plants appear to prevent cleavage that would set an inappropriate register for tasi*ARF* production from the *TAS3* genes of most flowering plants.

If the newly identified 5′ miR390 complementary site of *AtTAS3* evolved to interact with miR390 without being cleaved, we reasoned that it would efficiently bind the miR390-programmed silencing complex. To test this idea, we measured the ability of different sites to bind and act as competitive inhibitors of the endogenous silencing complex. Unlabeled RNA containing the *AtTAS3* 3′ site

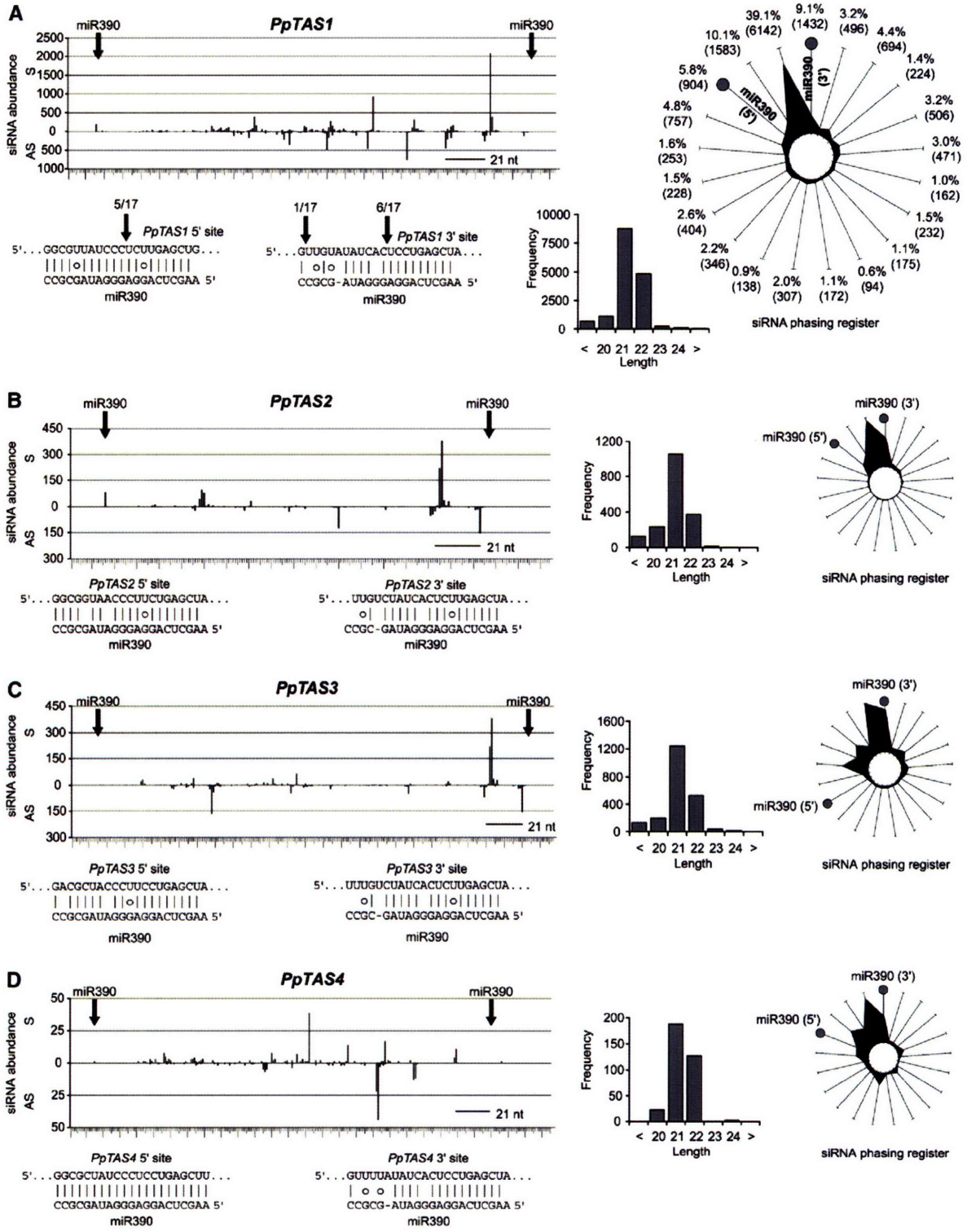


Figure 1. Phased *P. patens* siRNAs Mapped between miR390 Complementary Sites

(A) *PpTAS1*. The number of sequenced small RNAs with 5′ residues at each position along the *PpTAS1* locus is plotted for the sense (S) and antisense (AS) strands. The positions corresponding to the miR390 complementary sites are indicated by arrows, and pairing between the complementary sites and miR390 is shown below. The positions of 5′ ends mapped by RACE are indicated at the complementary sites by arrows, along with the fraction of sequenced clones mapping to the position. The length distribution and phasing of the small RNAs is plotted on the right. Each spoke of the radial

competed for cleavage of the identical labeled substrate with classical Michaelis-Menten behavior (K_m = 8.4 nM; k_{cat} = 0.006/s, estimating the concentration of endogenous miR390 silencing complex at 32 pM by quantitative northern). Adding *AtTAS3* 5′ site reduced the rate of cleavage even more efficiently, with the K_i of the 5′ site ~5-fold below the K_m of the 3′ site (Figure 3C). Disrupting complementarity to miR390 abolished the ability of both the 3′ and 5′ substrates to act as competitive inhibitors of miR390-mediated cleavage, thereby demonstrating the specificity of inhibition (Figure 3C). These observations, coupled with the conservation of central mismatches, indicated that the newly identified 5′ sites of flowering plants have evolved to bind the miR390 silencing complex while simultaneously avoiding cleavage. Indeed, analysis of flowering plant *AtTAS3* homologs showed that mismatches outside positions 8–11 were rare and almost never more disruptive than a G:U wobble (Figure 2C).

Both miR390 Complementary Sites Are Required for Full *AtTAS3* Function

We transformed *tas3-1*$^{-/-}$ plants with an *AtTAS3* genomic construct that can complement the *tas3-1* developmental defects (Adenot et al., 2006) and compared the frequency of complementation to that observed with variant constructs that had changes in one or both of the miRNA complementary sites (Figure 4). Of the primary transformant (T1) plants transformed with the wild-type *AtTAS3* construct (++), 37% were complemented for *AtTAS3* function, as indicated by lack of the elongated and curled leaves characteristic of the *tas3-1* line. Disrupting the 3′ complementary site (+Δ) lowered the frequency of complementation to 19.4%, demonstrating the importance of the 3′ site. Mutations of the 5′ complementary site that either disrupted binding (Δ+) or enhanced cleavage (R+) in vitro lowered the frequency of complementation to levels comparable to those of the 3′ site disruption. These results demonstrate the importance of binding without cleavage at this site, which was out of phase with the functional siRNAs. Simultaneous disruption of both sites (ΔΔ) lowered the apparent complementation frequency to 13.1%, a value indistinguishable from the background level of the assay. The lower complementation frequency of the ΔΔ construct compared with all constructs with single sites suggested some complementation by the single-site constructs, even the RΔ construct which contained a cleavable 5′ site and a disrupted 3′ site. We suspect that the relatively infrequent cases in which the single-site constructs complemented the *tas3-1* phenotype were due to miRNA-enhanced siRNA generation from the integrated transgene locus; transgenes often trigger siRNA production, especially when they include an miRNA complementary site (Parizotto et al., 2004), and the inventory of small RNAs produced from *TAS3* transgenes might occasionally include sufficient levels of tasi*ARF*s to enable wild-type development. Overall, the differential efficacies of the wild-type and single-site constructs demonstrate the importance of both miR390 complementary sites for *AtTAS3* function, and when considered together with our other results for the *AtTAS3* 5′ site, they indicate that an miRNA binding site that is not cleaved can nonetheless play an important, evolutionarily conserved role in plants.

A Conserved Trigger for siRNA Biogenesis

To test the hypothesis that dual small RNA complementary sites predispose the bounded region toward phased siRNA production, we examined *Arabidopsis* genes with multiple complementary sites for evidence of siRNA production. *ARF3* and *ARF4* both possess two sites complementary to the tasi*ARF*s TAS3 5′D7(+) and TAS3 5′D8(+) (Allen et al., 2005; Williams et al., 2005). For both *ARF3* and *ARF4*, our set of 887,266 *Arabidopsis* small RNA reads contained siRNAs from the region bounded by the tasi*ARF* complementary sites. These included 32 siRNA reads from the bounded region of *ARF4* and one read from the bounded region of *ARF3*. All 32 of the *ARF4* reads were in phase with each other (see Figure 5A and Figure S4), and the single *ARF3* read was in perfect phase with the cleavage sites (data not shown). For *ARF3*, no additional reads were observed outside of the bounded region, but for *ARF4*, an additional population of sense and antisense small RNAs arose from the region downstream of the 3′ complementary site. However, in contrast to the 21-nt, phased siRNAs from the region spanning the two sites, the siRNAs from the downstream cluster were not in phase with each other and were a mixture of 21-mers and 24-mers (Figure 5A). On the whole, the endogenous siRNAs from *ARF3* and *ARF4* supported the

graph represents 1 of the 21 possible phasing registers, with the total number of small RNAs mapping to that register plotted as distance from the center. The registers proceed clockwise from 5′ to 3′. The percentages and total number of sequenced siRNAs from each register are noted. The phasing registers of siRNAs from the antisense strand were corrected to account for the 2-nt, 3′ overhangs characteristic of Dicer-like cleavage. The specific registers predicted by 21-nt processing from the 5′ and 3′ cleavage sites are indicated with gray circles. Thus, phasing of the small RNA populations consistent with cleavage at one or the other site is indicated by abundant siRNAs that are in registers proximal to those predicted by the cleavage sites. For *PpTAS1*, 11,314 (72%) of the siRNAs were in phase with one of the complementary sites, in that they were in the same register or in immediately adjacent registers.

(B) *PpTAS2*, as in (A). The most populated register contained 709 siRNAs (39.1%); 1,648 (91%) of the siRNAs fell in the same or adjacent register(s) as a complementary site.

(C) *PpTAS3*, as in (A). The most populated register contained 524 siRNAs (24.3%); 1,097 (51%) of the siRNAs fell in the same or adjacent register(s) as a complementary site.

(D) *PpTAS4*, as in (A). The most populated register contained 80 siRNAs (23.3%); 221 (64%) of the siRNAs fell in the same or adjacent register(s) as a complementary site.

High-resolution graphs are also available (Figure S1).

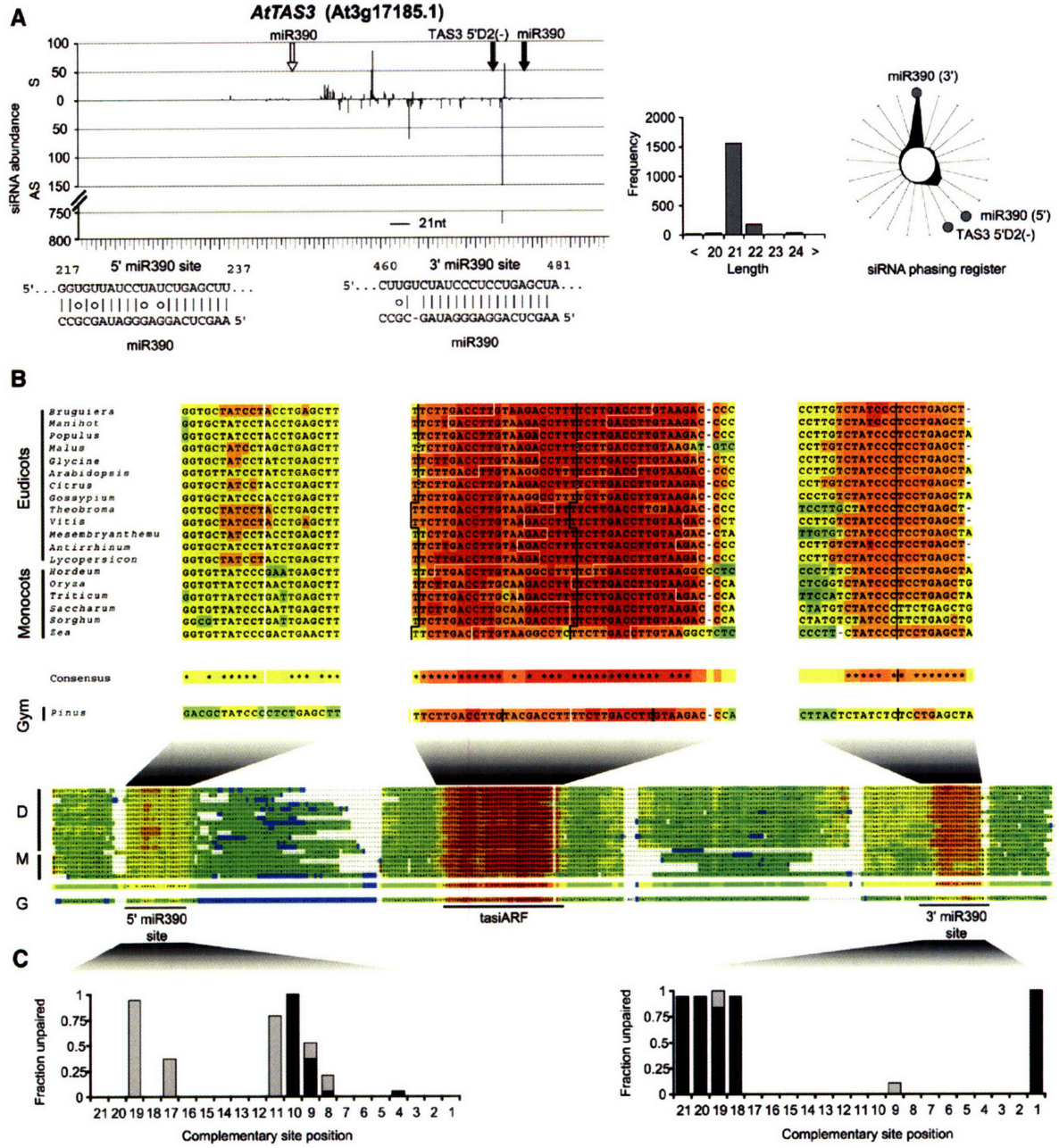


Figure 2. Dual miR390 Complementary Sites Were Conserved in Seed-Plant *TAS3* Loci

(A) *AtTAS3* siRNAs. The positions, length distribution, and phasing register of *Arabidopsis* small RNAs corresponding to *AtTAS3* are plotted as in Figure 1. A high-resolution graph is also available (Figure S1). The two miR390 complementary sites are indicated by arrows, with the open arrow indicating that no evidence of cleavage at the 5′ site was found. The site of secondary cleavage proposed to be directed by the highly abundant TAS3 5′D2(−) tasiRNA is also indicated (Allen et al., 2005). The most populated register contained 883 siRNAs (48.5%); 1,440 (79%) of the siRNAs fell in the same or adjacent register(s) as a complementary site.

(B) *TAS3* ESTs from diverse flowering plant genera contained dual miR390 complementary sites that flank the area of predicted tasiRNA production. Alignments were generated using ClustalW and color-coded based on the confidence of the local alignment using the CORE function of T-Coffee. tasi*ARF* refers to the regions homologous to *AtTAS3* 5′D7(+) and *AtTAS3* 5′D8(+). The regions corresponding to the 5′ miR390 complementary site, tasi*ARF*, and the 3′ miR390 complementary site are expanded. Black and white lines indicate the registers in phase with the 3′ and 5′ sites, respectively. EST details are given in Table S2. Gym, Gymnosperm.

(C) Analysis of miR390 complementary sites in EST homologs of *AtTAS3*. Complementary site positions were numbered starting with the residue corresponding to the 5′ nucleotide of miR390 and scored based on their pairing to *Arabidopsis* miR390a. Gray, G:U wobbles; black, other non-Watson-Crick pairs. This analysis was restricted to the flowering-plant *TAS3* homologs because the 5′ site of the *Pinus taeda* homolog is cleaved.

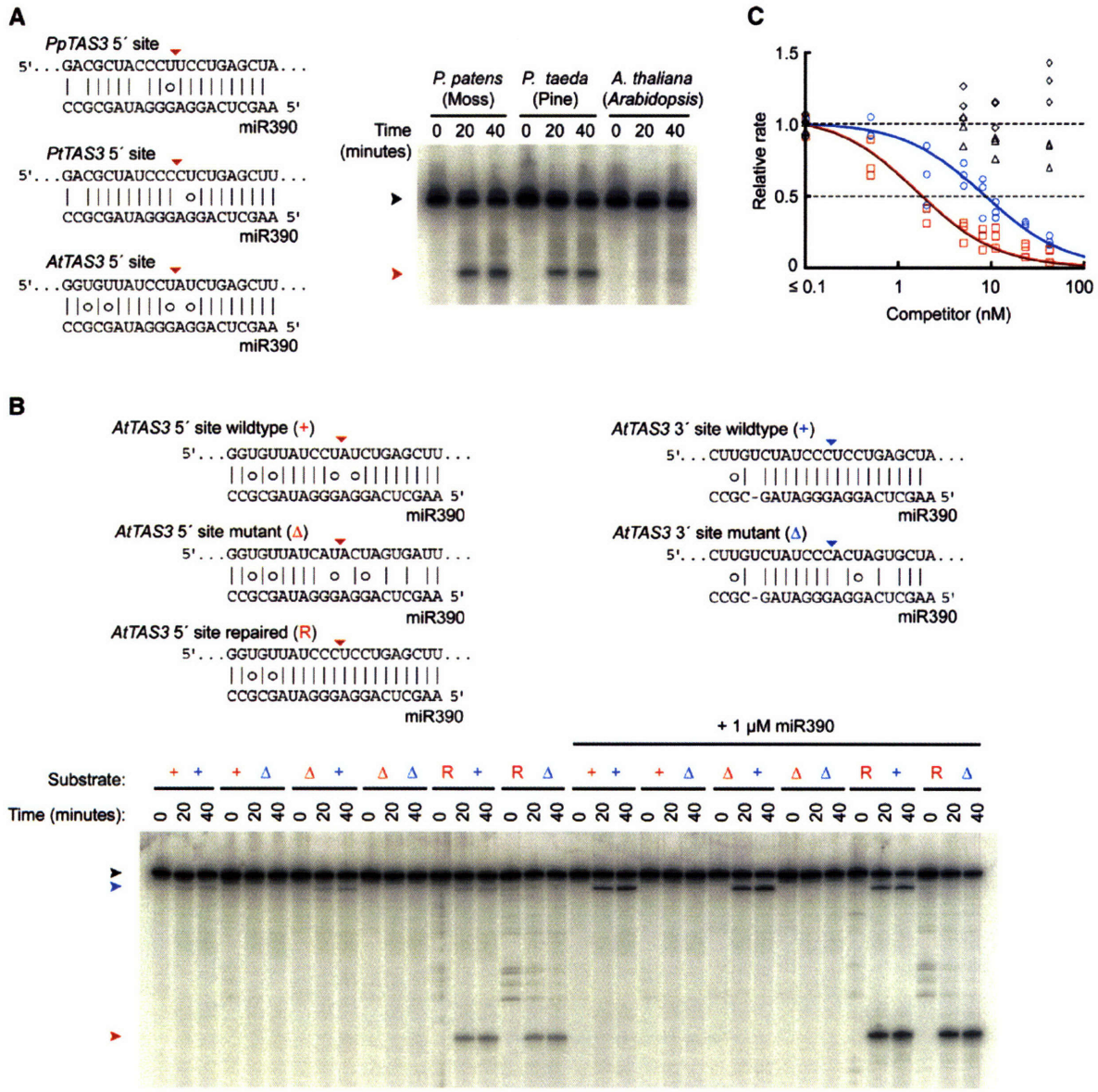


Figure 3. A Cleavage-Independent miRNA-Target Interaction in Flowering Plants

(A) miR390-directed cleavage of a target with a mismatch at position 10, but not one with a mismatch flanked by two G:U wobbles. The 5′ complementary site of *AtTAS3* was changed to the indicated sequences to mimic the sites found in *P. patens TAS3* or *P. taeda TAS3*. Cap-labeled RNAs containing only the 5′ complementary site were incubated with miR390-programmed wheat-germ lysate for the indicated times. Black and red arrowheads indicate the positions of uncut and cleaved substrate, respectively.

(B) The 5′ miR390 site of *AtTAS3* resisted miR390-mediated cleavage because of conserved mismatches at positions 9–11. Cap-labeled *AtTAS3* RNA encompassing both sites was incubated for the indicated times with wheat-germ lysate, with or without supplemental miR390. The sequences of 5′ and 3′ complementary site variants are shown paired to miR390. Red, 5′ site; blue, 3′ site; +, wild-type; Δ, disruptive mutation; R, repaired site. The positions of RNAs cleaved at the 3′ and 5′ site are indicated with blue and red arrowheads, respectively, while uncleaved substrate RNA is indicated by a black arrowhead.

(C) The cleavage-refractory 5′ miR390 complementary site of *AtTAS3* was a potent inhibitor of miR390-mediated cleavage. The relative rates of in vitro target cleavage using 1 nM of radiolabeled *AtTAS3*-derived RNA containing only the wild-type 3′ complementary site as substrate are plotted with varying concentrations of unlabeled RNA containing the wild-type 3′ site (blue circles), the wild-type 5′ site (red squares), the disrupted 3′ site (black triangles), or the disrupted 5′ site (black diamonds). The blue line shows the best fit to the data for the wild-type 3′ site and indicates a K_m for this site of 8.4 nM, whereas the red line shows the best fit to the data for the wild-type 5′ site and indicates a K_i for this site of 1.4 nM.

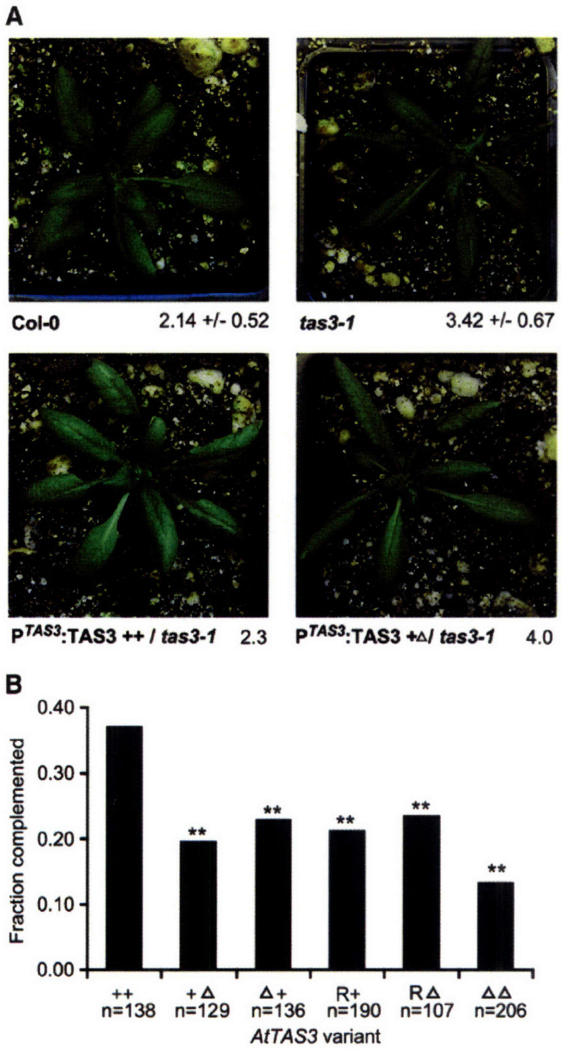


Figure 4. Dual miR390 Complementary Sites Are Required for Full *AtTAS3* Function

(A) Representative *tas3-1* transformants. The mean and standard deviation of the length to width ratio of the sixth leaf is reported for Col-0 (n = 40) and *tas3-1* (n = 41) control plants. Transformants with a ratio of less than 2.7 were scored as complementing (example at lower left), while those with ratios higher than 2.7 were scored as noncomplementing (example at lower right).

(B) The fraction of complemented *tas3-1* primary transformants after transformation with the indicated variants of *AtTAS3*. The number (n) of independent T1 plants examined for each variant is listed below. The sequences of 5′ and 3′ site variants are as in Figure 3. All statistically significant differences from the wild-type, as evaluated based on Chi-square goodness-of-fit tests, are indicated (**$p < 0.01$).

hypothesis that dual small RNA-mediated cleavage events predispose the intervening fragment toward recognition by an RdRp, leading to the production of phased siRNAs.

Several repetitive pentatricopeptide repeat (*PPR*) genes have been predicted or validated as targets of miR161, miR400, and *AtTAS1b*- and *AtTAS2*-derived tasiRNAs (Rhoades et al., 2002; Allen et al., 2004, 2005; Sunkar and Zhu, 2004; Vazquez et al., 2004a). We found 15 *PPR* genes that had at least two sites complementary to miR161.1, miR400, TAS1b 3′D4(−), or TAS2 3′D6(−). All 15 of these genes, including *At1g62670*, which was recently reported to give rise to secondary siRNAs (Ronemus et al., 2006), produced siRNAs from between small RNA complementary sites (Figure 5B and Figures S2 and S4). The *PPR*-associated siRNA populations were predominantly 21 nt in length, with a small proportion being 22 nt. Because of the highly repetitive nature of these target genes, many of the siRNAs could not be assigned to a single locus; however, even with this complication, the majority of the *PPR*-derived siRNAs were in a 21-nt phase at registers consistent with cleavage at known or predicted target sites, with 58% beginning within one nucleotide of the one predicted by cleavage at one of the sites (Figure 5B and Figures S2 and S4). This represents a substantial enrichment of phased siRNAs when compared with a random distribution of small RNAs falling into each of the possible 21 registers.

Analyses of secondary siRNAs deriving from miRNA targets with a single complementary site underscored the importance of dual complementary sites. Very little evidence of the production of siRNAs from single-site miRNA targets has been reported, with only the targets of miR168, miR393, and miR408 generating small RNAs that have been detected by sequencing of *Arabidopsis* small RNAs (Lu et al., 2005; Ronemus et al., 2006). The very low abundance of secondary siRNAs corresponding to *Arabidopsis* miRNA targets was also observed when we analyzed our large set of sequenced *Arabidopsis* small RNAs: as previously reported, significant numbers of secondary siRNAs were derived from *AGO1*, the only known target of miR168, as well as from the targets of miR393 (Figures S3 and S4). Sense and antisense small RNAs, predominantly 21 nt in length, arose only from the region downstream of the miRNA complementary sites, and they tended to be in phase with the end defined by miRNA-mediated cleavage (Figures S3 and S4). The phasing register, sizes, and downstream location of these secondary siRNAs were reminiscent of *AtTAS1a-c* and *AtTAS2* tasiRNAs, suggesting that they may have arisen through a common mechanism. Nonetheless, these examples were exceptions to the general observation that plant miRNA targets with single miRNA complementary sites were not efficient substrates for RdRp activity and subsequent production of secondary siRNAs. Even deeper sequencing of *Arabidopsis* small RNA populations may reveal that siRNA formation from single-site miRNA targets is a more widespread, albeit very low-efficiency, phenomenon. Taken together, our results strongly support a model in which dual miRNA complementary sites consistently predispose the bounded region of the target toward entry into a tasiRNA-like pathway, whereas a single miRNA complementary site triggers siRNA production less reliably and less efficiently.

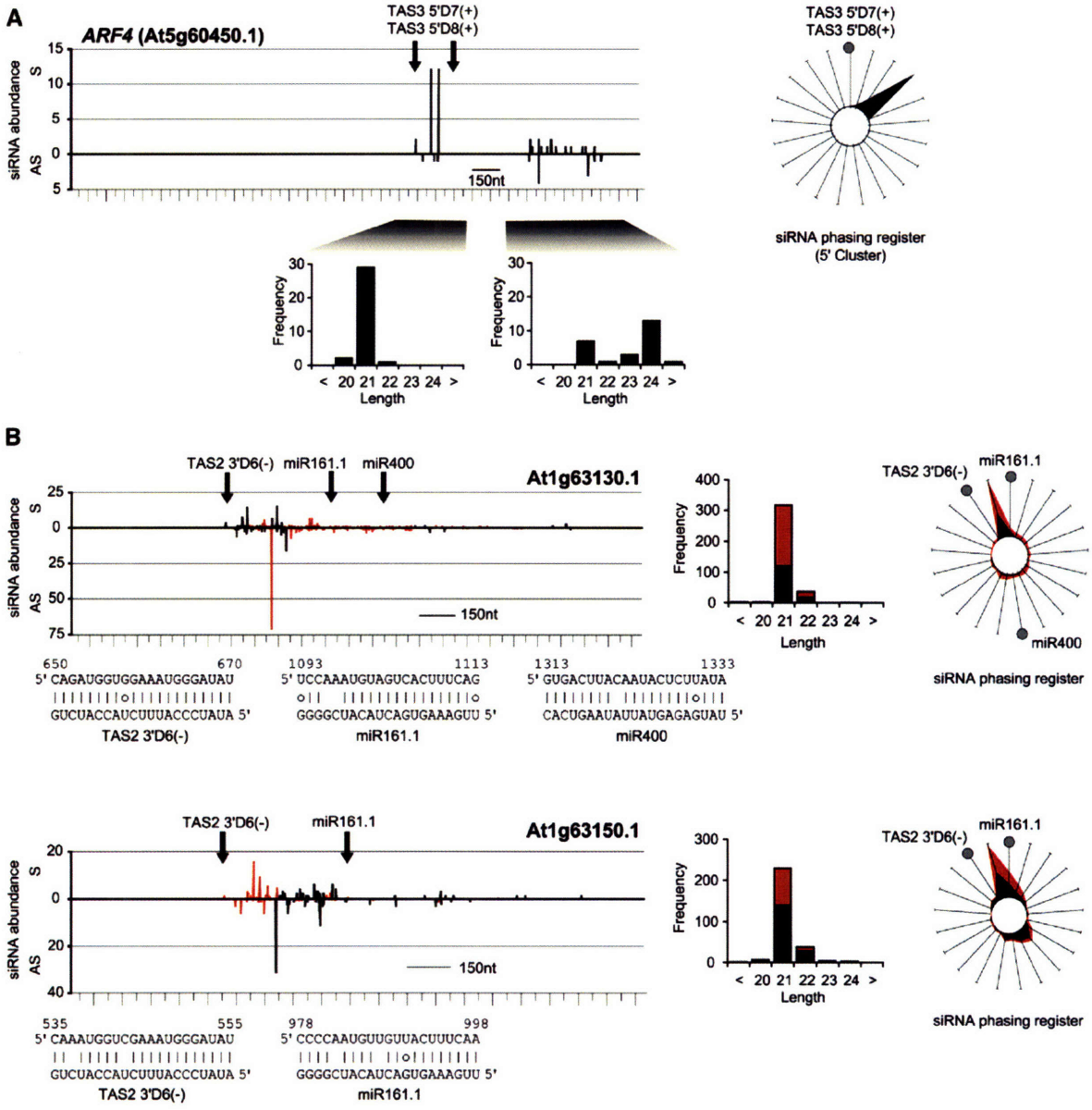


Figure 5. Multiple Small RNA Complementary Sites Correlate with siRNA Production from *Arabidopsis* Genes

(A) The positions, length distribution, and phasing register of *Arabidopsis* small RNAs corresponding to *ARF4* mRNA are plotted as in Figure 1. The most populated register contained 26 siRNAs (81.3%). The positions within the mRNA of the complementary sites to the tasi*ARF*s TAS3 5′D7(+) and TAS3 5′D8(+) RNAs are indicated by black arrows. The length distributions of two distinct clusters of *ARF4*-derived small RNAs are indicated separately.

(B) Small RNAs corresponding to the *PPR* mRNAs *At1g63130.1* and *At1g63150.1* are plotted as in (A). Complementary sites for miR161.1, miR400, and TAS2 3′D6(−) are indicated by arrows and displayed below. Small RNAs that matched only one site in the *Arabidopsis* genome are represented by black, whereas small RNAs with more than one match to the genome are plotted in red, after being normalized for the number of matching loci (e.g., a sequence with three reads that matches the genome twice would contribute 1.5 counts to each locus). The most populated register for *At1g63130.1* contained 96 (43.3%) unique and 78 (56.1%) ambiguous siRNAs; 118 (85%) of the unique and 149 (67%) of the ambiguous siRNAs fell in the same or adjacent register(s) as a complementary site. The most populated register for *At1g63150.1* contained 41 (39.9%) unique and 44 (24.6%) ambiguous siRNAs; 92 (51%) of the unique and 70 (68%) of the ambiguous siRNAs fell in the same or adjacent register(s) as a complementary site. The 13 other *PPR* genes with multiple small RNA complementary sites also gave rise to phased siRNAs (Figure S2).

DISCUSSION

One of the first reports on tasiRNAs noted their ~21-nt phasing, which was suggestive of successive DCL-catalyzed cleavage beginning from a defined point on a dsRNA substrate (Vazquez et al., 2004b). It was subsequently recognized that the single-stranded tasiRNA precursors are cleaved by miRNAs (Allen et al., 2005; Yoshikawa et al., 2005), and that this cleavage site defines a starting point for DCL4-catalyzed siRNA processing (Gasciolli et al., 2005; Xie et al., 2005). Yet the majority of *Arabidopsis* miRNA targets are cleaved without subsequently initiating the biogenesis of detectable amounts of siRNAs (Lu et al., 2005). What molecular features allow a plant cell to discriminate between the minority of cleavage products that are efficiently converted to siRNAs and the majority that are not? In the case of the siRNAs triggered by miR390, the expressed siRNAs emanate from a region flanked by miR390 complementary sites. The phasing registers of the *PpTAS* siRNA populations (Figure 1) and the observation of miR390-directed cleavage at both complementary sites in *PpTAS1* (Figure 1A) were consistent with successive DCL activity initiated from both ends of a dsRNA whose termini were defined by dual miR390-mediated cleavage events. The location of the pine tasi*ARF*s, which also appeared to be flanked by two cleavable sites (Figure 2B and Figure 3A), suggested that the same process produces siRNAs in gymnosperms. Furthermore, phased siRNAs were universally observed to emanate from regions of *Arabidopsis* genes bounded by small RNA complementary sites (Figure 5 and Figures S2 and S4). Small RNAs whose 5′ ends were within one nucleotide of the residues predicted by successive cleavage in precise 21-nt increments account for 71% of all small RNAs observed from *P. patens* and *Arabidopsis* genes with two or more complementary sites, indicating that such loci produced phased siRNAs. Precise 21-nt phasing appears to have degenerated as a result of occasional 22-nt cleavages, as indicated by the occurrence of both 21- and 22-nt small RNAs; otherwise, the fraction in phase would have been even higher. We conclude that one discriminating molecular feature of an RNA that triggers RdRp activity and subsequent entry into a phased siRNA pathway is the occurrence of dual small RNA-mediated cleavage events. Perhaps the resulting lack of any molecular signatures of normally processed mRNA (i.e., lack of 5′ cap, 3′ poly-A tail, and other mRNA-associated factors such as those deposited during splicing) direct such dual cleavage products to become RdRp substrates (Figure 6A).

AtTAS3 also contains dual miR390 complementary sites, the importance of which was highlighted by their conservation during seed-plant evolution and their requirement for efficient complementation of *tas3* plants (Figures 2–4). The observation that tasiRNAs from such a wide breadth of plant lineages derive from loci falling between miR390 complementary sites suggests an ancient and broadly conserved hallmark of siRNA biogenesis. Curiously, the 5′ miR390 complementary sites of flowering plant *TAS3* loci consistently contained mismatches to miR390 at positions critical for target cleavage (Figure 2C). These conserved mismatches at the center of the complementary site prevented miR390-mediated cleavage in vitro, yet they permitted efficient binding (Figure 3). Taken together, these results indicate that this site functions independent of target cleavage in initiating production of *AtTAS3* tasiRNAs (Figure 6B). We suggest that there might be additional instances in which miRNA binding without cleavage could have important functions in plants.

Dual miRNA complementary sites are an ancient trigger for siRNA biogenesis, but they are not the only trigger. *AtTAS1a-c* and *AtTAS2* precursors are cleaved at a single site upstream of the segment converted to siRNAs (Allen et al., 2005). Similarly, some other singly-cleaved *Arabidopsis* miRNA targets give rise to siRNAs from the area downstream of miRNA-mediated cleavage (Lu et al., 2005; Ronemus et al., 2006; Figures S3 and S4). We considered the possibility that these or other *Arabidopsis* miRNA targets might possess a second, previously unrecognized, miRNA complementary site. Even after drastically relaxing the stringency of target-site prediction to a level where past efforts have been unable to distinguish signal from noise, we found potential second sites in only 3 of 77 targets for which cleavage has been validated; none of the 3 were in *AtTAS1a-c*, *AtTAS2*, or the targets of miR168 and miR393 (data not shown). Thus, we conclude that there are very few, if any, unknown second sites among the currently known miRNA targets. What might replace the downstream miRNA complementary site and act as the second hit to trigger siRNA biogenesis from a few of these loci? Perhaps these transcripts have another mechanism for downstream cleavage involving another type of ribonuclease. Alternatively, downstream cleavage might be bypassed; in principle, RDR6 could be recruited to these transcripts by some other means; e.g., through a downstream binding element (Figure 6C). We note that in each of these cases (*AtTAS1a-c*, *AtTAS2*, and the targets of miR168 and miR393), the miRNA itself cannot be acting as a primer for RDR6-mediated RNA polymerization because all sequenced siRNAs arise exclusively from the 3′ region of the miRNA complementary site.

Production of siRNAs upstream of a single miRNA complementary site has been observed for a highly transcribed gene designed to be an miRNA "sensor" (Parizotto et al., 2004). Our hypothesis points to a possible mechanism for how such upstream siRNAs might be triggered. Perhaps a rare, upstream nonspecific cleavage, coupled with efficient miRNA-mediated processing, defines an initial RDR6 template in much the same way as dual miRNA-mediated cleavage does (Figure 6D). Because the miRNA sensor is highly expressed, initial siRNAs would likely encounter additional transcripts and direct their cleavage upstream of the miRNA complementary site, thereby generating more RDR6 substrates and initiating a cascade that results in the repression of the

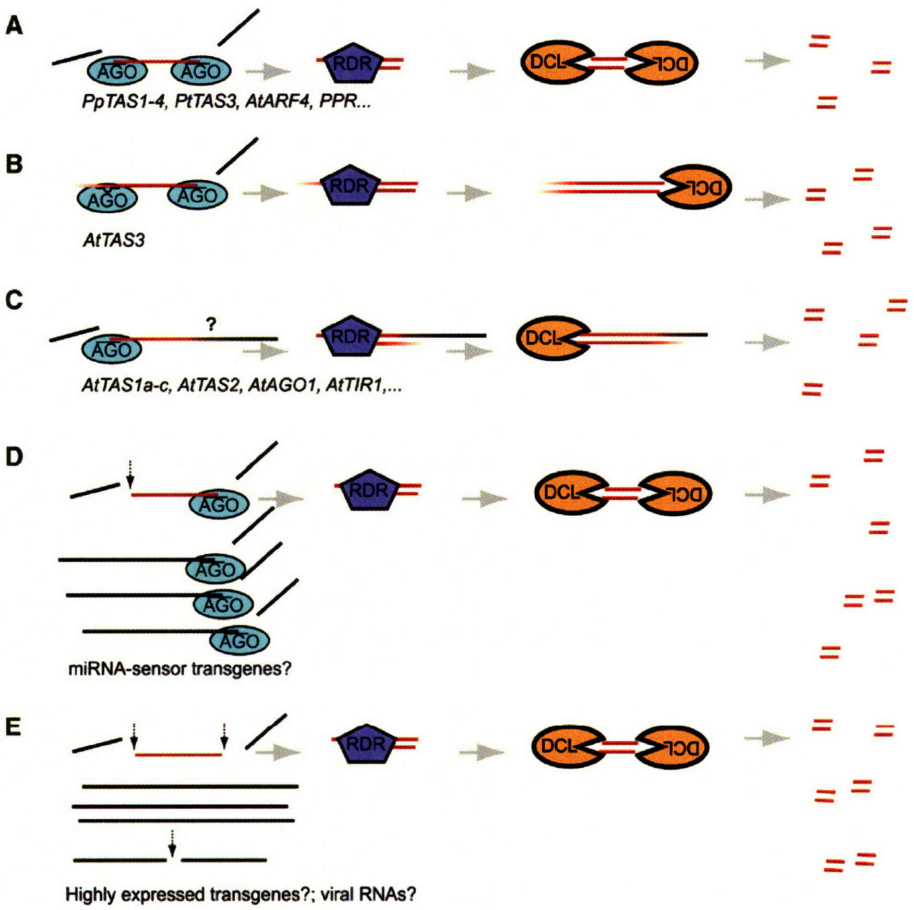


Figure 6. Different Embodiments of the Two-Hit Model for siRNA Biogenesis in Plants

(A) The internal product of dual miRNA- or siRNA-directed cleavage is recognized as a substrate for RdRp activity, which produces a dsRNA with two well-defined ends, as for *PpTAS1-4*, *PtTAS3*, *AtARF4*, and many *PPR* genes. Subsequent processing by a DCL enzyme produces populations of siRNAs in phase with one or the other end.

(B) The segment of an RNA flanked by miRNA complementary sites, only one of which is competent for AGO-catalyzed cleavage, defines an RdRp substrate, as in *AtTAS3*. Subsequent DCL processing of this dsRNA proceeds chiefly from the terminus defined by the miRNA-mediated cleavage.

(C) The segment of RNA defined on the 5′ by miRNA-directed cleavage and defined on the 3′ by an unknown element (question mark) that helps recruit RdRp activity gives rise to dsRNA, as in *AtTAS1a-c*, *AtTAS2*, and a limited number of other miRNA targets (Figure S3). Subsequent DCL processing of the dsRNA proceeds chiefly from the terminus defined by the miRNA-mediated cleavage.

(D) The segment of a very abundant RNA defined on the 5′ by a very rare, random cleavage event and defined on the 3′ by miRNA-directed cleavage becomes recognized as an RdRp substrate. This mechanism may trigger silencing of miRNA-sensor transgenes in plants.

(E) The segment of a very abundant RNA defined by two rare, random cleavage events becomes recognized as a substrate for RdRp activity. This mechanism may trigger transgene and virus silencing.

High-resolution graphs are also available (Figure S4).

sensor. Indeed, the probability of rare, nonspecific RNA cleavage occurring twice on the same mRNA molecule will increase as the abundance of that mRNA increases. For very highly expressed messages, such as those from viruses or transgenes, random nonspecific cleavage events might generate a few fragments that lack both a 5′ cap and a 3′ poly-A tail (Figure 6E). If, as our observations suggest, this type of RNA fragment is efficiently recognized by the *RDR6/DCL4* pathway, the resulting siRNAs would target other copies of the highly expressed transcript, facilitating the formation of more fragments without caps and tails and initiating a cascade that ultimately silences the virus or transgene. We note that this model postulates that the triggers for tasiRNA biogenesis and siRNA-mediated virus resistance are mechanistically similar—a postulate supported by the fact that a single Dicer-like protein, DCL4, is primarily responsible for both processes in wild-type *Arabidopsis* (Gasciolli et al., 2005; Xie et al., 2005; Yoshikawa et al., 2005; Bouche et al., 2006; Deleris et al., 2006). This two-hit model also

provides a molecular mechanism that rationalizes threshold models proposed previously to explain the observation that high initial expression levels of transgenes correlate with high frequencies of silencing initiation (Lindbo et al., 1993; Smith et al., 1994; Elmayan and Vaucheret, 1996). In our model, the plant cell detects over-abundant mRNA species because more abundant transcripts have a greater chance of both triggering and propagating the silencing cascade, i.e., more abundant species are more likely to include spontaneously generated fragments lacking both caps and tails, and more abundant species are more likely to be targeted by the resultant siRNA molecules, which in turn helps to generate more fragments lacking both caps and tails. The two-hit hypothesis thus outlines a mechanistic model for the initial identification of aberrant/dangerous RNA by the silencing machinery, and further suggests that the triggering of siRNA production can be directed to less abundant transcripts by the presence of dual miRNA complementary sites, which appears to have been occurring for the miR390-targeted *TAS* loci for the past 400 million years.

EXPERIMENTAL PROCEDURES

RNA Extractions

Total RNA from *P. patens* was isolated from three wild-type samples cultivated on minimal media agar overlaid with cellophane discs under standard conditions: Protonemata (7 day culture, 22°, 16 hr light), protonemata + young gametophores (14 day culture, 22°, 16 hr light), and gametophore + sporophytes (on media lacking NH_4-tartarate; 21 days at 22°, 16 hr light, followed by transfer to 15°, 8 hr light, with irrigation, for 39 days). Specimens were ground in 100 mM Tris-HCl (pH 9.0), 2% hexadecyltrimethylammonium bromide, 0.5% SDS, 2% polyvinyl pyrrolidone 40, 5 mM EDTA, and 10 mM β-mercaptoethanol. Samples were then phenol/chloroform extracted, ethanol precipitated, resuspended in 250 mM NaCl, and placed on ice for 20 min to precipitate carbohydrates. After centrifugation, the supernatant was ethanol precipitated to recover total RNA.

Small RNA Sequencing and Data Analysis

Construction of small RNA cDNA libraries was performed as described (Lau et al., 2001) and adapted for pyrosequencing (Supplemental Experimental Procedures). After phenol/chloroform extraction and native PAGE purification, 5 μg of purified PCR products for each library was delivered to 454 Life Sciences (Branford, CT, USA) for pyrosequencing. After discarding small RNAs that matched the *P. patens* chloroplast genome or the sense polarities of the nuclear 5S, 5.8S, 18S, or 26S rRNAs, the sequenced small RNAs were matched to the ~5.4 million *P. patens* WGS traces available at the time of analysis. Matching RNAs were classified as repetitive or nonrepetitive (Table S1, Supplemental Experimental Procedures). *PpTAS1-4* were found by using an algorithm that searched for clusters of nonrepetitive small RNAs from both the sense and antisense strand of WGS traces in which a significant fraction of the small RNAs were in phase with each other (in the same or adjacent registers, accounting for the 2-nt offset expected between the sense and antisense strands). Homologs of *AtTAS3* (Figure 2B, Table S2) were found by searching the est_others database for EST sequences that contained a sequence highly similar to 5′- TTCT TGACCTTGTAAGGCCTTTTCTTGACCTTGTAAGACCCC-3′ (representing the two tasi*ARF*s).

5′-RACE

Cleaved transcripts were detected using 5′-RACE (Llave et al., 2002).

Oligonucleotides

Oligos used for library preparation, RACE, mutagenesis, and transcription are listed in the Supplemental Experimental Procedures.

In Vitro Assays

Wheat-germ extract was prepared as previously described (Tang et al., 2003). Templates for in vitro transcription were made by PCR from constructs used for *tas3-1* complementation (pART27-*AtTAS3*; Adenot et al., 2006). All RNAs were gel-purified, and substrate RNA was radiolabeled by capping using guanylyl transferase (Ambion, Houston, TX) and α-^{32}P GTP. Cleavage reactions contained 50% wheat-germ extract (v/v), 5 mM DTT, 0.1 U/μl RNasin (Promega), 25 mM phosphocreatine, 1 mM ATP, 40 mM KOAc, and 0.03 μg/μl creatine kinase. Extract was incubated with buffer ± 1 μM phosphorylated miR390 for 20 min at 26°C, then added to 10,000 cpm of RNA (~5 fmol) per reaction. Reactions were stopped by the addition of 25 volumes of TRI reagent (Ambion, Houston, TX) at the indicated time points, followed by RNA extraction and PAGE analysis. For competition assays, extract was preincubated for 10 min at 26°C and added to labeled substrate (1 nM final) premixed with unlabeled competitor. Percent cleavage was calculated as the density of the cleaved band divided by the sum of the densities of the cleaved and full-length bands, and initial rates were calculated by regression. For 3′ wild-type competition, data were fit to the Michaelis-Menten equation, correcting for the fraction of total RNA that was radiolabeled: $V_{obs} = (1\text{ nM} / 1\text{ nM} + X\text{ nM}) \times (v_{max} \times [1 + X]\text{ nM}) / (K_m + [1 + X]\text{ nM})$, where X is the concentration of unlabeled 3′ wild-type competitor. For 5′ wild-type competition, data were fit for competitive inhibition: $V_{obs} = (v_{max} \times 1\text{ nM}) / (K_m \times (1 + X/K_i) + 1\text{ nM})$, where X is the concentration of the unlabeled 5′ wild-type competitor.

tas3-1 Complementation

Variants of pART27-*AtTAS3* (Figure 3B) were produced using QuikChange mutagenesis (Stratagene, La Jolla, CA) and transformed into *tas3-1* plants (Adenot et al., 2006). After 23–26 days of growth at ~22° in 16 hr light, 8 hr dark, the length-to-width ratio of the sixth rosette leaf was determined for each transformant. Nontransformed wild-type and *tas3-1* plants gave ratios of 2.14 ± 0.52 and 3.42 ± 0.67, respectively. Therefore, we classified transformants as "complementing" if this ratio was less than 2.7, recognizing that 17% of nontransformed *tas3-1* plants also scored as complemented, which represents the background of the assay.

Supplemental Data

The Supplemental Data for this article can be found online at http://www.cell.com/cgi/content/full/127/3/565/DC1/.

ACKNOWLEDGMENTS

We thank the Joint Genome Institute for the availability of *P. patens* WGS traces, Mitsuyasu Hasebe for the gift of *P. patens* spores and protocols, Herve Vaucheret for sharing the *tas3-1* homozygous line and pART27-*AtTAS3* construct prior to publication, Wendy Johnston for technical assistance, Guiliang Tang for advice on wheat-germ assays, and Matt Jones-Rhoades for computational assistance. This work was supported by a grant from the NIH, the Prix Louis D. from the Institut de France (D.P.B.), a Helen Hay Whitney Postdoctoral Fellowship (M.J.A.), and an NSF predoctoral fellowship (C.J.).

Received: March 21, 2006
Revised: July 7, 2006
Accepted: September 25, 2006
Published: November 2, 2006

REFERENCES

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouche, N., Gasciolli, V., and Vaucheret, H. (2006). DRB4-dependent *TAS3 trans*-acting siRNAs control leaf morphology through AGO7. Curr. Biol. *16*, 927–932.

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.

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.

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

Bouche, N., Lauressergues, D., Gasciolli, V., and Vaucheret, H. (2006). An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. EMBO J. *25*, 3347–3356.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C., and Voinnet, O. (2006). Hierarchical action and inhibition of plant Dicer-Like proteins in antiviral defense. Science *313*, 68–71.

Elmayan, T., and Vaucheret, H. (1996). Expression of single copies of a strongly expressed *35S* transgene can be silenced post-transcriptionally. Plant J. *9*, 787–797.

Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L., and Carrington, J.C. (2006). Regulation of *AUXIN RESPONSE FACTOR3* by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. Curr. Biol. *16*, 939–944.

Garcia, D., Collier, S.A., Byrne, M.E., and Martienssen, R.A. (2006). Specification of leaf polarity in *Arabidopsis* via the *trans*-acting siRNA Pathway. Curr. Biol. *16*, 933–938.

Gasciolli, 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.

Hunter, C., Willmann, M.R., Wu, G., Yoshikawa, M., de la Luz Gutierrez-Nava, M., and Poethig, R.S. (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. Development *133*, 2973–2981.

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

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

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). 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.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. (2004). MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5′ region. EMBO J. *23*, 3356–3364.

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.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and 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.

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.

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.

Ronemus, M., Vaughn, M.W., and Martienssen, R. (2006). MicroRNA-targeted and small interfering RNA–mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent RNA polymerase in *Arabidopsis*. Plant Cell *18*, 1559–1574.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev. Cell *8*, 517–527.

Smith, H.A., Swaney, S.L., Parks, T.D., Wernsman, E.A., and Dougherty, W.G. (1994). Transgenic plant virus resistance mediated by untranslatable sense RNAs: Expression, regulation, and fate of nonessential RNAs. Plant Cell *6*, 1441–1453.

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

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. *17*, 49–63.

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

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004b). Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Mol. Cell *16*, 69–79.

Williams, L., Carles, C.C., Osmont, K.S., and Fletcher, J.C. (2005). A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis ARF2, ARF3*, and *ARF4* genes. Proc. Natl. Acad. Sci. USA *102*, 9703–9708.

Xie, Z., Allen, E., Wilken, A., and Carrington, J.C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA *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.

Zamore, P.D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519–1524.

Accession Numbers

The consensus sequences of *PpTAS1-4* were deposited in Genbank (BK005825, BK005826, BK005827, and BK005828). 127,135 unique, genome-matched *P. patens* and 340,114 unique, genome-matched *Arabidopsis* small RNAs were deposited with the Gene Expression Omnibus (GSE5103 and GSE5228, respectively). The sequences of the *P. patens* and *Arabidopsis* small RNAs analyzed in this study are also available in Supplemental Databases S1 and S2, respectively.