

# Endogenous and Silencing-Associated Small RNAs in Plants <sup>W</sup>

Cesar Llave, Kristin D. Kasschau, Maggie A. Rector, and James C. Carrington<sup>1</sup>

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

**A large set of endogenous small RNAs of predominantly 21 to 24 nucleotides was identified in *Arabidopsis*. These small RNAs resembled micro-RNAs from animals and were similar in size to small interfering RNAs that accumulated during RNA silencing triggered by multiple types of inducers. Among the 125 sequences identified, the vast majority (90%) arose from intergenic regions, although small RNAs corresponding to predicted protein-coding genes, transposon-like sequences, and a structural RNA gene also were identified. Evidence consistent with the derivation of small RNAs of both polarities, and from highly base-paired precursors, was obtained through the identification and analysis of clusters of small RNA loci. The accumulation of specific small RNAs was regulated developmentally. We propose that *Arabidopsis* small RNAs participate in a wide range of post-transcriptional and epigenetic events.**

## INTRODUCTION

Eukaryotes have several types of tiny regulatory RNAs (Ruvkun, 2001; Waterhouse et al., 2001; Grosshans and Slack, 2002). Small temporal RNA (stRNA) and small interfering RNA (siRNA) are involved in the post-transcriptional control of gene expression. The stRNAs, such as those encoded by the *Caenorhabditis elegans* *lin-4* and *let-7* genes, bind to target mRNAs to repress translation (Lee et al., 1993; Moss et al., 1997; Olsen and Ambros, 1999; Reinhart et al., 2000; Slack et al., 2000; Grosshans and Slack, 2002). The siRNAs are associated with RNA silencing (or RNA interference [RNAi]), serving as guides for sequence-specific nucleolytic activity of the silencing-associated RISC complex (Hamilton and Baulcombe, 1999; Hammond et al., 2000, 2001; Zamore et al., 2000). The stRNA and siRNA are 21 to 25 nucleotides in length and arise from precursor molecules with extensive base-paired structure (Carthew, 2001; Ruvkun, 2001).

The functions of RNA silencing and stRNA-mediated repression of translation are partly understood. RNA silencing in plants is an antiviral response triggered by most or all viruses (Carrington et al., 2001; Vance and Vaucheret, 2001; Voinnet, 2001; Waterhouse et al., 2001). Based on analyses of virus-encoded silencing suppressors, RNA silencing limits the extent of virus multiplication and the ability of vi-

ruses to move long distances (Anandalakshmi et al., 1998; Brigneti et al., 1998; Kasschau and Carrington, 1998, 2001; Voinnet et al., 1999). siRNAs of plants also were suggested to direct the sequence-specific methylation of nuclear DNA and intercellular signaling to trigger silencing in distal tissues (Dalmay et al., 2000a; Sijen and Kooter, 2000; Jones et al., 2001; Matzke et al., 2001; Mette et al., 2001; Waterhouse et al., 2001).

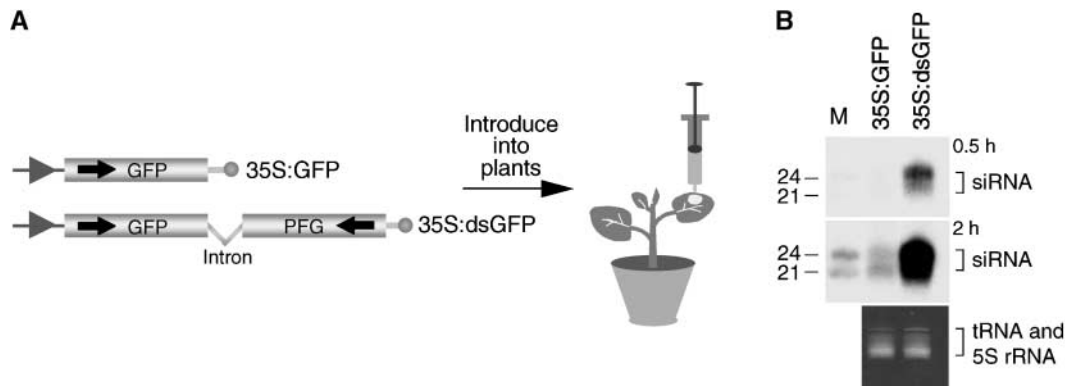
In *C. elegans*, RNAi serves to suppress the activity of several classes of transposons (Ketting et al., 1999; Tabara et al., 1999). Also, genetic defects in *C. elegans* genes (e.g., *ego1* and *dcr1* mutants) required for RNAi have specific developmental defects, suggesting that RNAi components are necessary for normal growth and development (Smardon et al., 2000; Grishok et al., 2001; Knight and Bass, 2001). stRNA-mediated repression of translation is required for the regulation of development in *C. elegans* (Ambros, 2000; Banerjee and Slack, 2002; Grosshans and Slack, 2002). The *lin-4* and *let-7* stRNAs, for example, bind to complementary sequences within the 3' nontranslated regions of sets of target mRNAs that are necessary for the temporal control of larval development (Lee et al., 1993; Wightman et al., 1993; Moss et al., 1997; Reinhart et al., 2000; Slack et al., 2000). Mutations that affect stRNA genes or stRNA target genes result in heterochronic defects (Ambros and Horvitz, 1984; Grishok et al., 2001).

Both stRNAs and siRNAs arise by cleavage of precursor molecules containing exclusively or extensive duplex secondary structure by an RNaseIII-like enzyme (Bernstein et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001), which in *Drosophila* was termed Dicer. Dicer is a multidomain protein with RNA helicase-like

<sup>1</sup> To whom correspondence should be addressed. E-mail carrington@orst.edu; fax 541-737-3045.

<sup>W</sup> Online version contains Web-only data.

Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.003210.



**Figure 1.** RNA Silencing and siRNAs in an Agrobacterium Infiltration System.

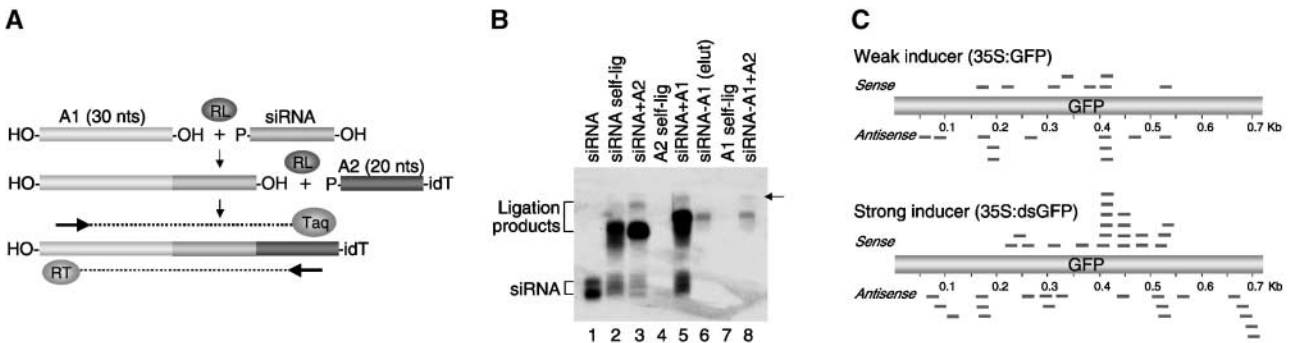
**(A)** Constructs used in the transient assay. The 35S:GFP construct contains a functional copy of the GFP coding region, and the 35S:dsGFP construct contains both sense and antisense GFP sequences separated by an intron. The constructs were introduced into Agrobacterium and injected into *N. benthamiana*.

**(B)** Silencing-associated siRNAs from Agrobacterium-inoculated tissue. Small RNA from injected tissues was analyzed with a radiolabeled GFP-specific probe. A series of two exposures (0.5 and 2 h) is shown. RNA standards of 21 and 24 nucleotides containing 5' phosphate and 3' hydroxyl ends were used (M). Equivalent loading of samples was shown by staining the gel with ethidium bromide before transfer (bottom gel).

and double-stranded (ds) RNA binding domains in addition to RNaseIII-like regions (Bernstein et al., 2001). Using fully duplexed RNA as a substrate, Dicer-catalyzed reaction products are predominantly dsRNA molecules containing 21 to 22 nucleotides in each strand and two nonpaired nucleotides at each 3' end (Bass, 2000; Bernstein et al., 2001; Elbashir et al., 2001). Unlike single-stranded ribonucleases, RNaseIII-like enzymes such as Dicer yield products with 5' monophosphate and 3' hydroxyl termini (Bass, 2000).

The relevance of small RNAs may be much broader than

was suspected previously. Recent analyses of *C. elegans*, *Drosophila*, and human revealed dozens of small RNAs termed micro-RNAs (miRNAs) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Most or all of the miRNAs arise from structured precursor RNAs from intergenic regions (IGRs). Several lines of evidence indicate that Dicer is required for the maturation of sRNA and possibly miRNA in general (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Ruvkun, 2001; Grosshans and Slack, 2002). Although the functional significance of most miRNAs has yet to be determined, at



**Figure 2.** Cloning and Analysis of siRNAs from *N. benthamiana* Tissue.

**(A)** 5' (A1) and 3' (A2) adapters with end modifications to ensure unidirectional ligation were ligated sequentially to gel-purified siRNAs using RNA ligase (RL), and the products were amplified by reverse transcriptase-mediated PCR. nts, nucleotides.

**(B)** RNA gel blot of intermediate and final siRNA-adaptor ligation products using a GFP probe. The A1-siRNA ligation product was gel eluted (elut) before ligation with A2. The arrow shows the final A1-siRNA-A2 product.

**(C)** Distribution of cloned siRNA sequences in the GFP sense and antisense sequences using 35S:GFP (weak inducer) and 35S:dsGFP (strong inducer).

least some are likely to function in a manner similar to that of stRNAs.

In this study, a large set of small RNAs was identified in Arabidopsis. The majority of these RNAs arose from intergenic sequences, although several were identified from coding sequences and introns of conventional genes and transposable elements. The Arabidopsis small RNA populations may be more diverse structurally and functionally than the animal small RNA populations.

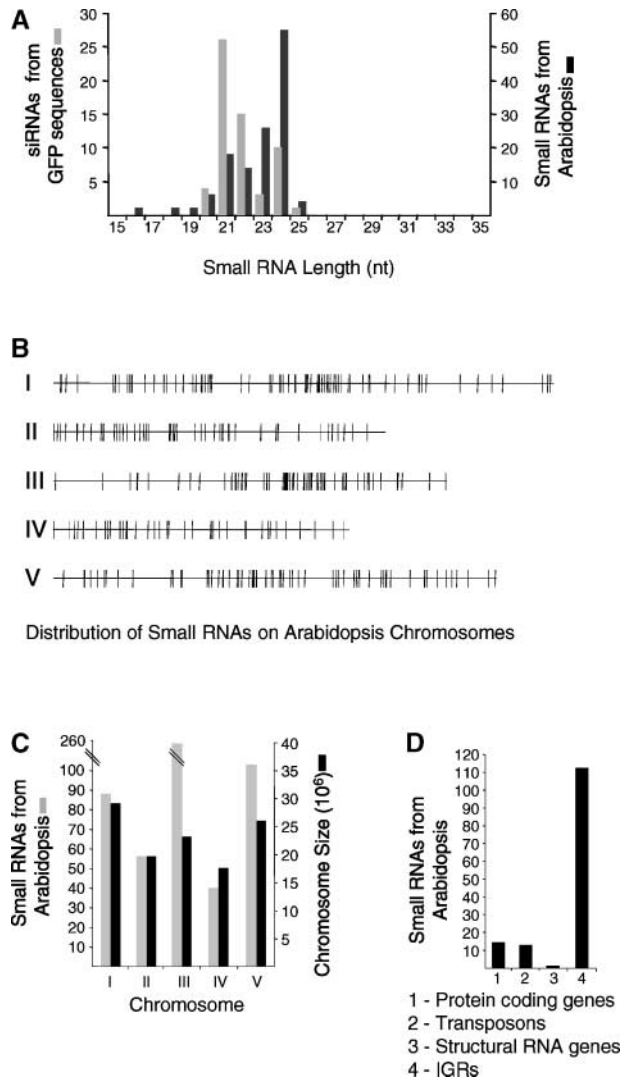
**RESULTS**

**Amplification and Analysis of siRNAs from Plants Using a Model RNA-Silencing System**

The primary goal of this study was to explore naturally occurring, endogenous small RNAs of plants. However, siRNAs induced in a transient RNA-silencing system were analyzed initially for two purposes. First, siRNAs formed using defined RNA-silencing inducers provided a relatively simple model to adapt and refine methods for analyzing small RNAs in plants. Second, determination of the chemical nature of siRNAs in plants was necessary for comparison with endogenous small RNAs. *Nicotiana benthamiana* leaves were infiltrated with *Agrobacterium tumefaciens* cultures containing expression cassettes for transcripts encoding green fluorescent protein (35S:GFP) or transcripts corresponding to intron-spliced, inverted-repeat forms of the GFP sequence (35S:dsGFP) (Figure 1A).

The 35S:GFP and 35S:dsGFP constructs are weak and strong inducers, respectively, of RNA silencing in the infiltration zone (Johansen and Carrington, 2001). The GFP-related siRNAs from the infiltration zone were analyzed by blot assay using a radiolabeled GFP-specific probe. The siRNAs produced using either inducer migrated as a heterogeneous population with two predominant bands of ~21 to 22 and 24 to 25 nucleotides in length (Figure 1B). The quantity of siRNA induced by the 35S:dsGFP construct was ~20-fold greater than that induced by the 35S:GFP construct. Additionally, the ratio of the two major bands differed between siRNAs triggered by the two inducers, with the slower migrating band accumulating to relatively high levels using the 35S:dsGFP inducer (Figure 1B).

The siRNAs were gel purified and ligated to RNA oligonucleotide adapters using T4 RNA ligase (Figure 2A). Adapter 1 (A1) contained hydroxyl groups at both 5' and 3' ends and could ligate only to the 5' end of siRNAs. Adapter 2 (A2) contained 5' monophosphate and 3' inverted deoxythymidine groups and could ligate only to the 3' end of siRNA molecules. Intermediate and final ligation products using siRNAs were analyzed by RNA gel blot assay using a radiolabeled GFP-specific probe that detected both sense and antisense sequences (Figure 2B). siRNA self-ligation control



**Figure 3.** Size of Small RNAs and Distribution of Loci with Complete Identity to Small RNAs in the Arabidopsis Genome.

(A) Histogram of sizes of siRNAs produced in *N. benthamiana* (gray bars) and small RNAs in Arabidopsis inflorescence tissue (black bars). nt, nucleotides.

(B) The 125 small RNAs isolated from Arabidopsis hit 539 genome loci. A vertical line marks each locus. Several loci had a high density of small RNA sequences, and these may appear as a single line.

(C) Number of small RNA loci (gray bars) on each of the five chromosomes (black bars).

(D) The Arabidopsis small RNA sequences corresponded to sequences from protein-coding genes, transposons and retroelements, structural RNA genes, and IGRs.

**Table 1.** Small RNAs in Arabidopsis

RNAs	Coding Regions		Intron Regions		Structural RNA <sup>a</sup>	IGR <sup>b</sup>
	Nontransposons	Transposons	Nontransposons	Transposons		
Single-locus small RNAs <sup>c</sup>	4	4	0	0	0	72
Multiple-locus small RNAs <sup>d</sup>	5	4	9	1	1	41
Total small RNAs (single + multiple)	9	8	9	1	1	113
Small RNA loci in the genome <sup>e</sup>	11	15	9	2	137	365

<sup>a</sup>One small RNA corresponds to the 5S rRNA repeat unit.

<sup>b</sup>Contains all small RNAs with sequence identity to IGRs, including pseudogenes.

<sup>c</sup>Identified as small RNAs with complete sequence identity to only one locus in the Arabidopsis genome.

<sup>d</sup>Identified as small RNAs with complete sequence identity to multiple loci in the Arabidopsis genome.

<sup>e</sup>Total number of loci with complete sequence identity to small RNAs.

reactions yielded products with electrophoretic mobilities consistent with the formation of siRNA dimers and larger concatemers (Figure 2B, lane 2). Ligation of siRNAs with the A1 adapter yielded major products with mobilities consistent with the joining of one siRNA and one adapter molecule (Figure 2B, lanes 3 and 5) and minor products with slower electrophoretic mobilities. The major A1-siRNA (using both inducers) ligation product was gel purified (Figure 2B, lane 6) and ligated to adapter A2.

The A1-siRNA-A2 ligation products were reverse transcribed, amplified by PCR, and cloned. The inserts in 72 recombinant plasmids containing GFP sequences (based on hybridization) were sequenced, revealing 59 unique siRNAs (Figure 2C and supplemental material). Nineteen sequences were from siRNA derived from the 35S:GFP inducer, whereas 40 sequences were from 35S:dsGFP-derived siRNA (Figure 2C). The siRNAs ranged from 20 to 25 nucleotides in length, with the most common lengths being 21 (44%) and 22 (25%) nucleotides (Figure 3A). The distribution of siRNA sizes using the two inducers generally was similar, although the 24-nucleotide size class was recovered more often using the 35S:GFP inducer (26%) compared with the 35S:dsGFP inducer (13%). The relative proportion of 24-nucleotide siRNA in the sequenced populations was not reflected in the blot assay, in which siRNAs that comigrated with the 24-nucleotide standard were overrepresented using 35S:dsGFP (Figure 1B).

The unidirectional adapter ligation strategy allowed unambiguous determination of siRNA polarity. The ratio of sense-to-antisense siRNA was 8:11 using the 35S:GFP inducer and 20:20 using the 35S:dsGFP inducer (Figure 2C). Some clustering of siRNAs along the inducer sequences was detected. This was particularly evident for siRNAs of sense polarity induced by 35S:dsGFP (Figure 2C). Whether the apparent nonrandom distribution was caused by an amplification or sampling problem, or by a biologically relevant process such as preferential cleavage zones in siRNA precursors or siRNA stability, was not investigated. None of the siRNAs contained nontemplated 5' or 3' nucleotides. Al-

though little or no bias was detected in the nucleotide composition of siRNAs, 5' guanosine and 3' thymidine were underrepresented among the total cloned sequences.

### Small RNA Populations in Arabidopsis

Using the same methods that were used with the siRNA analysis, naturally occurring small RNAs from inflorescence tissue of Arabidopsis ecotype Columbia-0 were extracted, ligated to adapters, amplified, cloned, and sequenced. The RNA samples used for ligation to adapters were eluted from denaturing polyacrylamide gels in the zone corresponding to 15 to 35 nucleotides (based on RNA oligonucleotide standards). Naturally occurring small RNAs in the absence of an exogenous inducer of silencing could reflect endogenous RNA silencing activity, the activity of stRNA-like genes similar to *C. elegans lin-4* and *let-7*, or other small RNA-mediated processes. The availability of the complete genome sequence of Arabidopsis enabled confirmation of the identity of each small RNA sequence.

A total of 131 inserts were analyzed, yielding 125 unique Arabidopsis sequences (Table 1 and supplemental material). These ranged in length from 16 to 25 nucleotides, although the vast majority contained 21 to 24 nucleotides (Figure 3A). The overall size distribution of Arabidopsis small RNAs was similar to that of siRNA generated in the GFP- and dsGFP-silencing system, although the relative proportion of RNAs containing 21, 22, 23, and 24 nucleotides differed between the two sets (Figure 3A). Four of the small RNAs (15, 17, 33, and 125) contained single internal mismatches compared to the published Arabidopsis genome sequence (see supplemental data). Several lines of evidence argue against these sequences being formed as a result of the random breakdown of cellular RNAs.

The size distribution of the Arabidopsis small RNAs was coincident, over a narrow range, with that of the siRNAs (Figure 3A), even though the RNA substrates for adapter ligation were eluted from gels in the zone corresponding to

15 to 35 nucleotides. Also, the terminal structures of RNA fragments formed as a result of single-strand-specific ribonucleases typically contain 5' hydroxyl and 3' phosphate groups, which would prevent ligation to adapters. These data support the hypotheses that Arabidopsis contains large populations of small RNAs resembling siRNAs and that the small RNAs originate through processing by an RNaseIII-like mechanism.

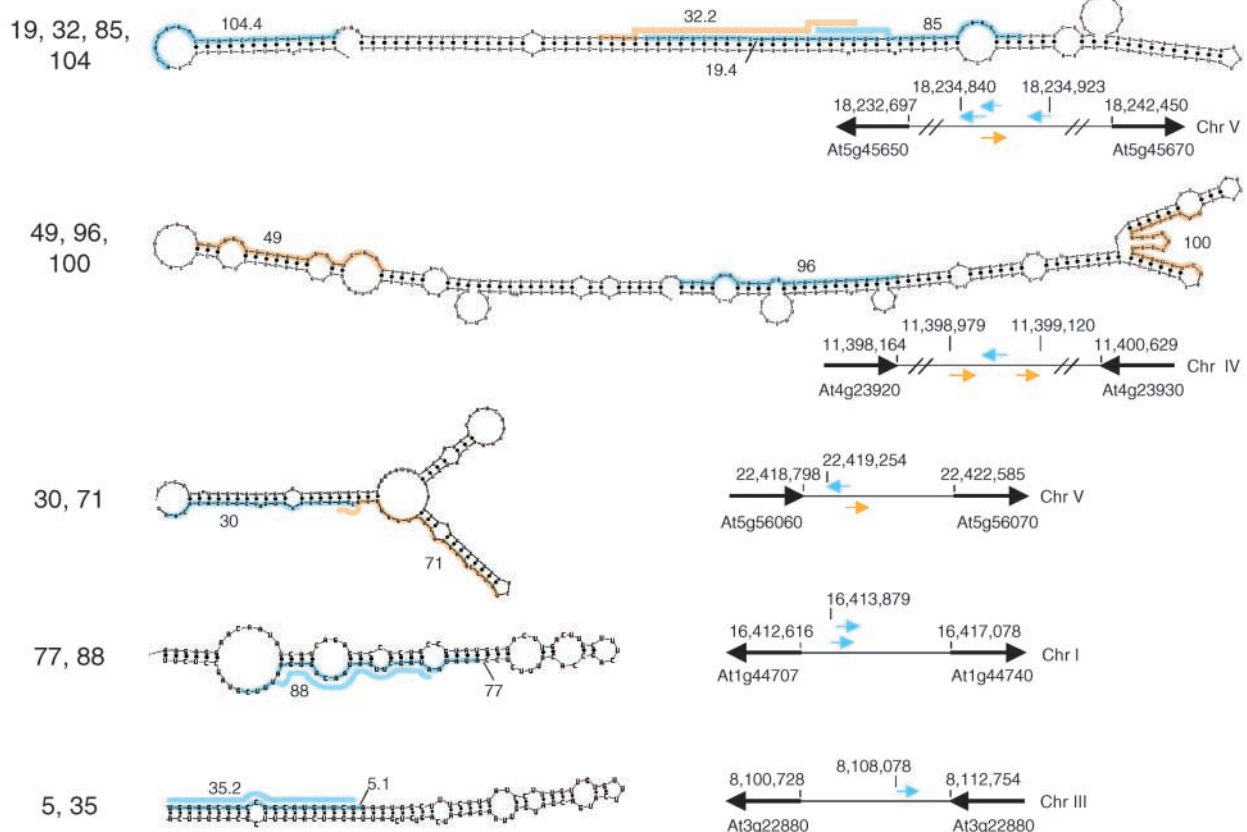
Although 125 unique sequences were identified, 539 genomic loci with sequence identity to small RNAs were mapped to IGRs, protein-coding genes, transposon-like sequences, and a structural RNA gene (Table 1, Figures 3B to 3D). Eighty small RNAs corresponded to unique loci, whereas 45 sequences corresponded to multiple loci (Table 1). Sequences from small RNAs were identified in each chromosome (Figure 3B), with the numbers of small RNA loci on chromosomes I, II, and IV being approximately proportional to chromosome size (Figure 3C). The numbers of

loci on chromosomes III and V, on the other hand, were disproportionately high as a result of small RNAs 75 and 123, which were present in sequences that were repeated 64 and 137 times, respectively.

**Small RNAs from Arabidopsis IGRs**

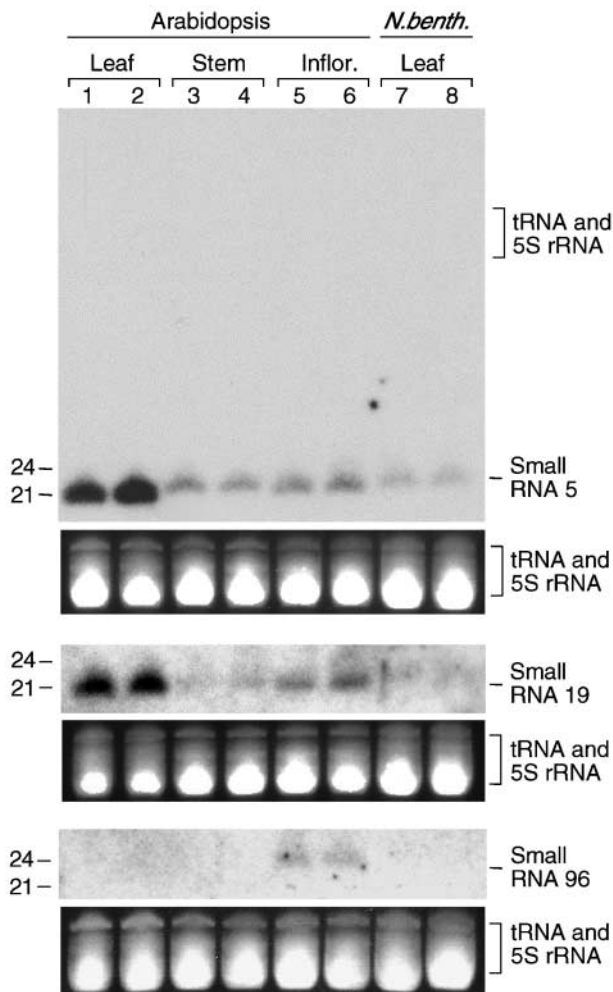
A total of 113 (90%) of the Arabidopsis small RNAs corresponded to at least one IGR sequence (Table 1, Figure 3D). Seventy-two of these derived from unique IGR positions, whereas 41 corresponded to multiple sites (IGR and/or predicted protein-coding genes). There was no correlation between the polarity of small RNAs and the polarity of known or predicted genes flanking these sites (see below), supporting the hypothesis that these RNAs originate through independent transcription events. The proportion of small RNAs mapping to IGR sequence was significantly higher than the

**Small RNAs**



**Figure 4.** Predicted Secondary Structures of RNA from Loci Containing Clusters of IGR-Derived Small RNA Sequences.

Small RNA sequences (highlighted) occurred in both sense (blue) and antisense (orange) polarities. The orientation of the known or predicted genes (accession numbers are indicated) flanking each small RNA locus is shown at right by the large arrows (black), along with genome coordinates (nucleotides) as annotated in GenBank. The orientation of the small RNAs in relation to the chromosome is represented by the orange and blue arrows. Chr, chromosome.



**Figure 5.** IGR-Derived Small RNAs in Arabidopsis Tissues.

Small RNA gel blots of duplicate samples isolated from three tissue types of Arabidopsis were probed with radiolabeled antisense oligonucleotides corresponding to small RNAs 19, 96, and 5. The zone in the small RNA 5 blot above the small RNA position is presented to show that putative precursor RNAs were not detected. The position of the tRNA/5S RNA zone in the blot was determined by staining with ethidium bromide. RNA from *N. benthamiana* leaf tissue also was analyzed. The RNA standards (21 and 24 nucleotides) and sample loading controls are described in Figure 1.

proportion of the Arabidopsis genome assigned as IGR (~50% [Arabidopsis Genome Initiative, 2000]), indicating that IGRs are relatively rich in small RNA-yielding sequences.

Several small RNAs originated from each of five loci containing clusters of two to four small RNA sequences spaced irregularly within an IGR interval (Figure 4). In four of these clusters, the small RNA sequences overlapped partially. Interestingly, three of the clusters contained small RNA sequences of both sense and antisense polarities (Figure 4).

The sequences surrounding and including each of the clustered small RNAs were predicted to fold into relatively stable base-paired structures containing up to 323 nucleotides (Figure 4). The predicted IGR structures generally were larger and more complex than those predicted to form in miRNA precursors in animals (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001).

The accumulation patterns of several IGR-derived small RNAs from the clustered loci were analyzed in replicated, normalized samples from Arabidopsis leaf, stem, and inflorescence tissue by blot hybridization using radiolabeled oligonucleotide probes. Discrete RNA bands that comigrated with either the 21- or 24-nucleotide standards were detected consistently using probes for small RNAs 19, 96, and 5 (Figure 5). In each case, the electrophoretic mobility corresponded precisely with the size determined through sequence analysis. As shown in the blot for small RNA 5 (Figure 5), no slower migrating species corresponding to putative precursors were detected using any of the labeled probes. Each small RNA exhibited tissue specificity, with small RNAs 5 and 19 accumulating predominantly in leaves and small RNA 96 accumulating in inflorescence tissue. In addition, a small RNA band was detected in *N. benthamiana* leaf extracts that were probed for small RNA 5 (Figure 5).

#### Small RNAs from Arabidopsis Coding Regions and Introns

Twenty-seven small RNAs (22%) corresponded to at least one genomic sequence of either sense or antisense polarity within a known or predicted protein-coding gene (Table 1, Figure 3D). Eight of these sequences corresponded to unique loci, and 19 were present at multiple sites (IGR or other protein-coding genes). Seventeen small RNAs mapped to sense or antisense sequences within predicted coding sequences, and 10 small RNAs mapped to sense or antisense intron sequences.

Nine small RNAs corresponded to genes resembling transposases or other proteins commonly found in transposable elements (Table 2). These included both long terminal repeat (LTR) and non-LTR retrotransposons (class I) and MULE and CACTA-like transposons (class II). The percentage (33%) of small RNAs mapping to transposon-like genes relative to the percentage corresponding to nontransposon protein-coding genes is substantially higher than that predicted based on the number of class I and II transposons in the Arabidopsis genome (Arabidopsis Genome Initiative, 2000; Le et al., 2000).

Small RNAs containing sense or antisense sequences of 16 nontransposon protein-coding genes were identified (Table 3). However, ESTs or cDNA clones have been identified for only four of these predicted genes. For two genes, which were predicted to encode a Ser carboxypeptidase-like protein and a protein of unknown function, multiple small RNAs were identified.

Two of the protein-coding genes were predicted to yield

**Table 2.** Transposon-like Genes Corresponding to Small RNAs

Annotation	Accession Number <sup>a</sup>	Chromosome (Small RNA Start Position) <sup>b</sup>	Small RNA <sup>c</sup>	Small RNA Orientation	Position in Gene <sup>d</sup>
<b>Class I</b>					
LTR retrotransposons	At3g44570	III (16168450)	4	Sense	CR
	At3g50490	III (18748214)	67.18	Sense	CR-NCR
	At4g03860	IV (1806346)	82	Antisense	CR
	At1g35970	I (13402547)	109.4	Antisense	Intron
	At2g10650	II (4118437)	109.33	Antisense	Intron
Non-LTR retrotransposons	At2g31080	II (13173209)	64	Antisense	CR
	At1g26860	I (9305299)	105.1	Antisense	CR
<b>Class II</b>					
MULEs	At2g25370	II (10756003)	42	Antisense	CR
CACTA-like elements	At1g41910	I (15208566)	6.1	Antisense	CR
	At2g10000	II (3735856)	6.2	Antisense	CR
	At1g36080	I (13484620)	53.1	Sense	CR
	At1g23930	I (8461386)	53.2	Sense	CR
	At5g37390	V (14547631)	53.3	Sense	CR
	At5g33220	V (12243390)	53.5	Sense	CR
	At4g09370	IV (4908254)	53.6	Sense	CR
	At4g28960	IV (13245060)	53.7	Sense	CR
	At3g47330	III (17445738)	53.8	Sense	CR

<sup>a</sup> Determined by BLAST search with the small RNA sequence.

<sup>b</sup> First nucleotide position in the chromosome corresponding to the small RNA sequence.

<sup>c</sup> Unique code for each small RNA or small RNA locus.

<sup>d</sup> Location of the corresponding small RNA sequence in the gene coding sequence. CR, coding region; NCR, noncoding region.

mRNA or pre-mRNA with unusual secondary structures. The small RNA 2 sequence was identified at two positions, one corresponding to a sequence annotated as IGR and the other corresponding to an antisense segment in the 3' proximal coding sequence of a predicted gene of unknown function (Figure 6, Table 3). These two small RNA loci were adjacent to one another on chromosome V. However, the sequences surrounding the two small RNA sites contained extensive complementarity, which would yield an exceptionally stable stem structure in an mRNA precursor (Figure 6A). Small RNAs corresponding to sense and antisense small RNA 2 were detected by blot assay, and both accumulated predominantly in inflorescence tissue (Figure 6B). This region of chromosome V also contained a cluster of two other overlapping small RNA sequences (30 and 71).

At a Ser carboxypeptidase-like gene, small RNA sequences 90 and 101 were identified at two sites, separated by 41 nucleotides, within the predicted coding sequence. Complementary sequences to both small RNA 90 and 101 also were identified in a predicted intron for this gene. Secondary structure prediction indicated that a significant portion of the putative pre-mRNA would form an extensive base-paired stem composed of both coding and intron sequences (data not shown). These data indicate that at least some of the small RNAs from protein-coding genes are associated with sequences capable of forming extensively base-paired structures.

Several additional small RNAs mapped to both IGR and

protein-coding genes. Small RNA 39 was identified at four positions in the Arabidopsis genome, including a chromosome III IGR sequence with the potential to form a stable stem structure (Figure 7A). This RNA also corresponded to an antisense segment from the central region of three closely related genes (Figure 7B), at least two of which are expressed. These genes belong to the *Scarecrow-like* (also known as GRAS) family of plant-specific transcription factors, >30 of which are predicted in Arabidopsis (Figure 7C). These factors are involved in a number of signaling and developmental processes, including the control of radial cell divisions in roots (*SHR* and *SCR1*), hormonal signaling by gibberellin (*GAI1* and *RGA1*), and light signaling (*PAT1*) (Di-Laurenzio et al., 1996; Silverstone et al., 1998; Bolle et al., 2000; Helariutta et al., 2000). Small RNA 39 was detected exclusively in inflorescence tissue (Figure 7D). Developmentally regulated small RNAs with complementarity to mRNA targets may enable a number of mechanisms for post-transcriptional regulation.

### Small RNAs from Arabidopsis Structural RNA

Two small RNAs (75 and 123) were clustered in or around the repeated 5S rRNA gene. The 5S rRNA gene is repeated within large head-to-tail arrays on three chromosomes (III, IV, and V), the typical repeat unit (~500 nucleotides) being composed of transcribed sequence (120 nucleotides) and

flanking spacer sequence (Cloix et al., 2002). The small RNA 123 sequence was identified in the sense orientation within the 5S transcribed region in 116 of the chromosome III repeats and 21 of the chromosome V repeats. This sequence overlapped box A of the internal promoter. The small RNA 75 sequence was in the opposite polarity relative to small RNA 123 and mapped to the spacer sequence on the 5' side of the transcribed sequence in 64 of the repeated units. All of the small RNA 75 sequences were in the centromere-proximal 5S rRNA array on chromosome III. This array was shown to be inactive in Arabidopsis (Cloix et al., 2002).

## DISCUSSION

### Endogenous Small RNAs in Arabidopsis

A diverse population of small RNAs in Arabidopsis was identified in this study. Although these were isolated from plants that lacked any exogenous inducers of RNA silencing, the Arabidopsis small RNAs resembled siRNAs triggered by two types of RNA-silencing inducers in *N. benthamiana*. The endogenous small RNAs of Arabidopsis and siRNAs of *N. benthamiana* were clearly related based on similarity of size

(predominantly 21 to 24 nucleotides in length) and inferred terminal structure, which strongly suggest that both types of RNAs are formed by an RNaseIII-like mechanism. Importantly, the Arabidopsis small RNAs also resemble miRNAs of animals, which in *C. elegans* and *Drosophila* are likely formed through the activity of Dicer (Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001; Lee and Ambros, 2001). Arabidopsis contains nine predicted genes with at least one RNaseIII-like domain similar to that in Dicer (Z. Xie and J.C. Carrington, unpublished data), although the roles of these genes in small RNA biosynthesis have yet to be assigned.

Among the 125 endogenous small RNAs identified in this study, only two were identified in more than one clone, suggesting that the screen was far from saturating. Further adding to the complexity is the likelihood that the production of small RNAs is under developmental control. Small RNA 39, for example, accumulated to detectable levels only in inflorescence tissue. Like the mRNA transcriptome, the composition of small RNA populations is unique in different cell types and tissues.

### Small RNAs from IGR Sequences

The predominant class of Arabidopsis small RNA derives from IGR sequences. These small RNAs have similarities to

**Table 3.** Protein-Coding Genes Corresponding to Small RNAs

Annotation	Accession Number <sup>a</sup>	Chromosome (Small RNA Start Position) <sup>b</sup>	Small RNA <sup>c</sup>	Small RNA Orientation	Position in Gene <sup>d</sup>	ESTs <sup>e</sup>
Inorganic phosphate transporter	At1g20860	I (7255467)	23.1	Antisense	Intron	–
Cytochrome P450	At1g47630	I (16956099)	104.1	Sense	CR-intron	–
Hypothetical protein, no known similarities	At1g79970	I (29737207)	107.1	Antisense	Intron	+
		I (29737208)	14.1	Antisense	Intron	–
Weak cathepsin L similarity	At2g39160	II (16282272)	40	Antisense	CR	–
Zinc-finger/BTB domain-containing protein	At2g40440	II (16839165)	17.2 <sup>f</sup>	Antisense	Intron	–
F-box-containing protein	At3g21170	III (7420343)	97	Sense	CR	–
Scarecrow-like	At4g00150	IV (57411)	39.1	Antisense	CR	+
Scarecrow-like	At3g60630	III (22420503)	39.2	Antisense	CR	–
Scarecrow-like	At2g45160	II (18567669)	39.4	Antisense	CR	+
Hypothetical protein, similar to nuclear protein in rice	At4g18980	IV (9360949)	24.19	Sense	Intron	–
Hypothetical protein, no known similarities	At4g27930	IV (12867464)	55	Antisense	CR	–
Ser carboxypeptidase-like	At5g22960	V (7660496)	90.1	Sense	Intron	–
		V (7659853)	90.2	Antisense	CR	–
		V (7659933)	101.1	Sense	CR	–
		V (7660416)	101.2	Antisense	Intron	–
Hypothetical protein, no known similarities	At5g56070	V (22423005)	2.2	Antisense	CR	–
Hypothetical protein, no known similarities	At5g49590	V (19422956)	105.2	Antisense	Intron	–
Hypothetical protein, no known similarities	At4g34690	IV (15520101)	124	Antisense	CR	–
SKI2-like RNA helicase protein	At5g61140	V (24315538)	110.2	Sense	Intron	+

<sup>a</sup> Determined by BLAST search with the small RNA sequence.

<sup>b</sup> First nucleotide position in the chromosome corresponding to the small RNA sequence.

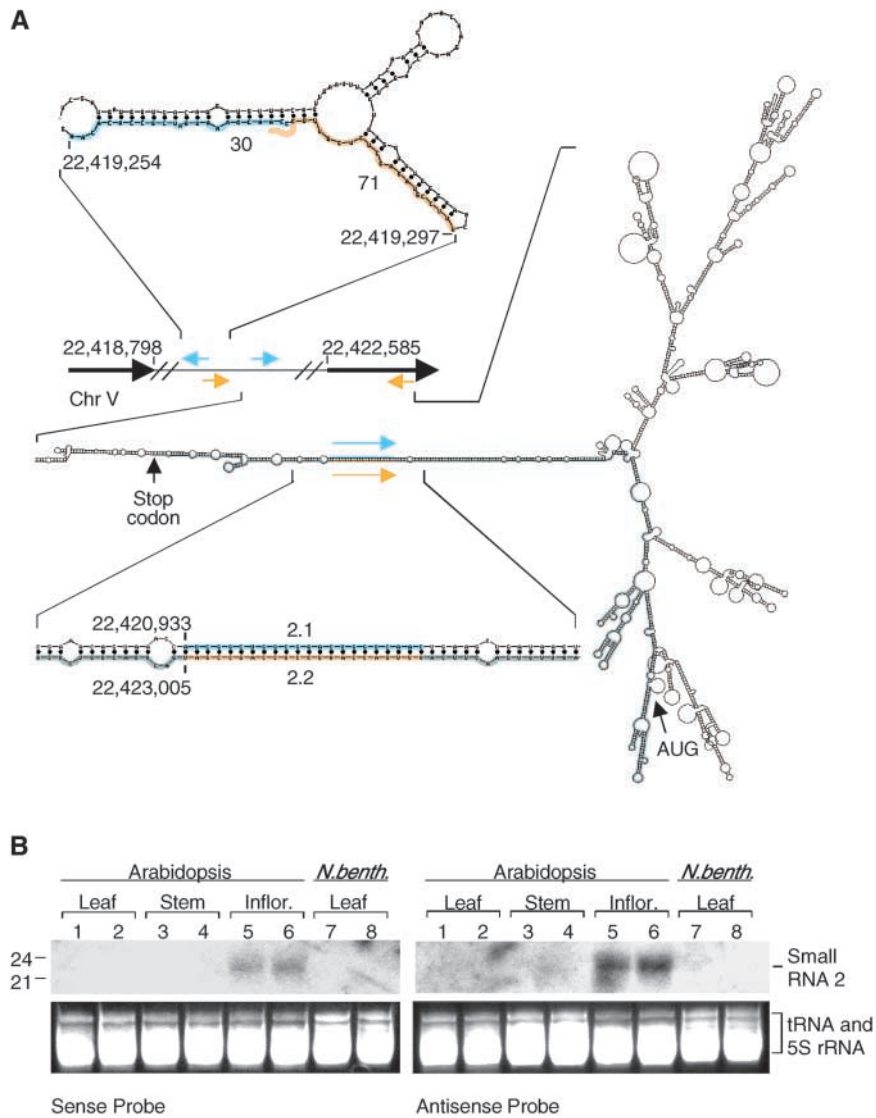
<sup>c</sup> Unique code for each small RNA or small RNA locus.

<sup>d</sup> Location of the corresponding small RNA sequence in the gene. CR, coding region.

<sup>e</sup> +, EST in GenBank; –, no EST in GenBank.

<sup>f</sup> Small RNA 17 contained a single mismatch relative to the published Arabidopsis genome sequence.





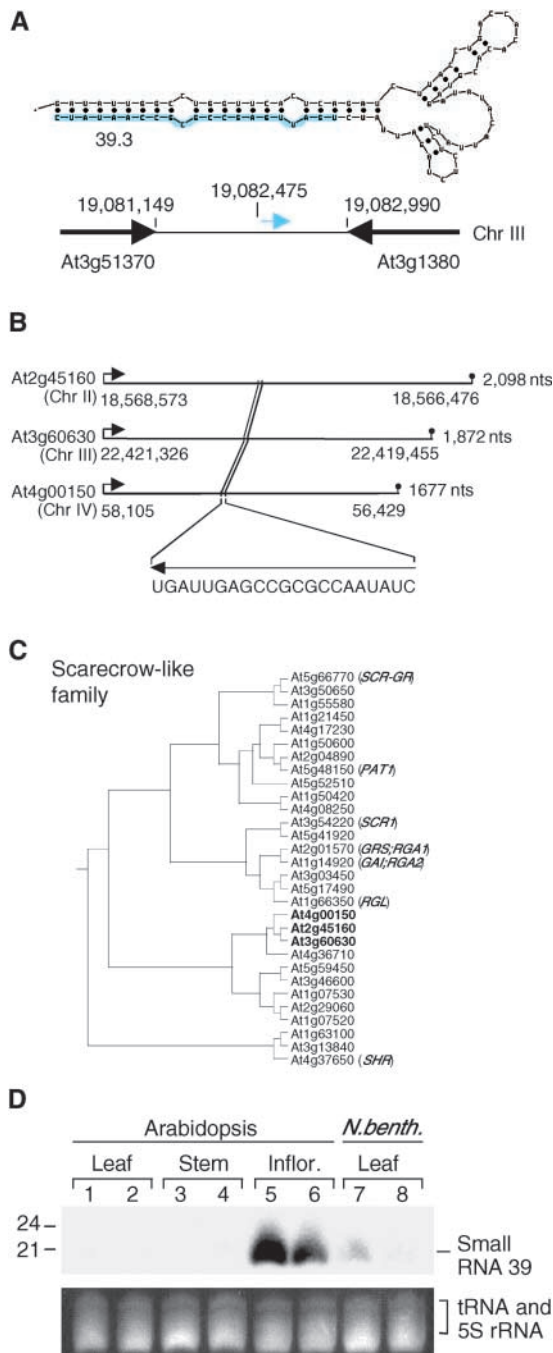
**Figure 6.** Small RNAs in and around Protein-Coding Sequences with Unusual Potential to Form Extensive Secondary Structure.

**(A)** Predicted secondary structures for RNA encompassing four loci within a 4000-nucleotide segment of chromosome V. Two loci in opposite orientations (2.1 and 2.2) correspond to small RNA 2, one within a predicted IGR (blue) and another in sequences near the 3' end of a protein-coding gene of unknown function (orange). The predicted open reading frame is indicated by gray shading. Chr, chromosome.

**(B)** Blot analysis of duplicate small RNA samples isolated from three Arabidopsis tissues. Duplicate sets of small RNA samples were analyzed using radiolabeled probes to detect sense or antisense versions of small RNA 2. The RNA standards (21 and 24 nucleotides) and sample loading controls are described in Figure 1.

the recently identified miRNAs of *C. elegans*, *Drosophila*, and human (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). The miRNAs are small (predominantly 21 to 24 nucleotides), arise by processing of precursor transcripts (~70 nucleotides) containing imperfectly paired stem structures in a Dicer-dependent manner, and originate primarily from intergenic sequences (Ruvkun,

2001; Grosshans and Slack, 2002). Both animal miRNAs and Arabidopsis small RNAs arise by transcription events that are independent of adjacent conventional genes. However, there are some notable differences that may reflect distinct mechanisms of biosynthesis. First, the Arabidopsis IGR small RNAs are predicted to arise from heterogeneous precursor RNAs. Some may contain relatively short and simple



**Figure 7.** Characterization of Small RNA 39.

**(A)** Predicted secondary structure of the IGR region corresponding to small RNA 39 (blue) and the upstream flanking sequence. The orientation (black arrows) and genome coordinates (nucleotides as annotated in GenBank) of predicted genes flanking the IGR sequence are shown. **(B)** Three genes in the *Scarecrow-like* gene family with sequences complementary (long arrow) to that of small RNA 39. For continuity with supplemental data online, the small RNA 39 sequence is written 5' to 3'. **(C)** Dendrogram of the *Scarecrow-like* gene family. Genes high-

lighted in boldface contained sequences complementary to that of small RNA 39. **(D)** Blot analysis of duplicate small RNA samples isolated from three tissue types of Arabidopsis using a radiolabeled oligonucleotide that was complementary to small RNA 39. The RNA standards (21 and 24 nucleotides) and sample loading controls are described in Figure 1. Chr, chromosome; nts, nucleotides.

stem-loop structures, such as those in animal miRNA precursors, whereas others are associated with extensive or complex stem-loop structures (Figure 4). Second, several small RNAs originate from irregularly spaced or overlapping loci within presumed precursor RNAs of variable length. Third, Arabidopsis small RNAs from clustered loci occur in both orientations. These distinct properties of Arabidopsis IGR small RNAs can be explained by a biosynthetic mechanism that includes transcription of IGR DNA to form a highly base-paired precursor, Dicer-like cleavage to yield limited small RNAs, and RNA-dependent RNA polymerase-mediated amplification of part of the precursor RNA. The initial and subsequent small RNAs would serve as primers in a PCR-like amplification mechanism, similar to that proposed for the amplification of target RNAs during RNAi in *C. elegans* and *Drosophila* (Lipardi et al., 2001; Sijen et al., 2001). The amplification products then would be cleaved efficiently by an RNaseIII-like mechanism to yield small RNAs of both polarities.

In fact, clustering of small RNAs in both orientations is precisely what happens during RNA silencing using either weak or strong inducers (Figure 2). Amplification of these IGR precursor transcripts to yield sequences in both orientations may be catalyzed by *SDE1/SGS2*, an RNA-dependent RNA polymerase-like protein that is required for RNA silencing triggered by transgenes and some viruses (Dalmay et al., 2000b; Mourrain et al., 2000). However, Arabidopsis contains at least six other genes with high similarity to *SDE1/SGS2* (L.K. Johansen and J.C. Carrington, unpublished data), one or more of which may have specialized functions in small RNA formation.

What are the functions of IGR-derived small RNAs? First, at least some small RNAs may be similar to *C. elegans* stRNAs. These would anneal imperfectly to mRNA targets to modulate translation. The imperfect nature of the base pairing between small RNAs and target mRNAs, however, precludes the computer-based identification of interacting partners using available data. Second, the IGR small RNAs might interact with target mRNAs, as do stRNAs, but have effects on processes other than translation. These interactions could affect RNA splicing, mRNA localization or RNA turnover, depending on how the small RNA interacted with the factors that normally control these events.

Third, IGR-derived small RNAs with perfect or nearly perfect complementarity to an mRNA might trigger site-specific

lighted in boldface contained sequences complementary to that of small RNA 39.

**(D)** Blot analysis of duplicate small RNA samples isolated from three tissue types of Arabidopsis using a radiolabeled oligonucleotide that was complementary to small RNA 39. The RNA standards (21 and 24 nucleotides) and sample loading controls are described in Figure 1. Chr, chromosome; nts, nucleotides.

cleavage of the mRNA. This could happen if the small RNA incorporates into a functional RISC-like complex with sequence-specific nucleolytic function, as occurs during RNA silencing (Hammond et al., 2000, 2001). The inflorescence-specific small RNA 39 is interesting in this respect, because it is perfectly complementary to mRNAs for three *Scarecrow* gene family members.

Fourth, some small RNAs may serve as guides for the modification of chromosomal DNA. RNA silencing is known to trigger cytosine methylation within homologous sequences in chromosomal DNA, and siRNAs have been proposed to function as guide sequences in this process (Jones et al., 1999; Sijen and Kooter, 2000; Matzke et al., 2001; Mette et al., 2001). Small RNA-directed methylation of chromosomal DNA would have implications for the epigenetic control of nuclear genomes (Wassenegger, 2000). The small RNAs also might guide the modification of functional or structural RNA molecules. Small nucleolar RNAs, for example, serve such a function during 2'-O-ribose methylation and pseudouridylation modification of rRNAs (Tollervey, 1996; Smith and Steitz, 1997).

### Small RNAs from Protein-Coding and Structural RNA Genes

A less abundant class of *Arabidopsis* small RNA contained sense or antisense sequences for coding regions and introns of conventional genes and transposon-like elements. In cases in which the small RNA sequences were unique (e.g., small RNAs 40 and 97), the small RNAs must have derived from the genes containing the homologous sequences. In cases of small RNAs corresponding to multiple genome loci, such as those containing identity with several of the transposon sequences or 5S rRNA, the origin was ambiguous.

Does the presence of this class of small RNA reflect the silencing of endogenous mRNA or structural RNA targets? In *C. elegans*, genetic analysis revealed that several factors required for RNAi also are required for transposon silencing (Ketting et al., 1999; Tabara et al., 1999). The highly repetitive nature of transposon sequences in the *Arabidopsis* genome (*Arabidopsis* Genome Initiative, 2000; Le et al., 2000), as well as the likelihood that some active transposons will integrate in a manner that yields dsRNA-forming sense and antisense transcripts, suggest that transposons are a probable target of RNA silencing in plants.

In addition to small RNAs from transposons, the small RNAs from protein-coding gene sequences that are predicted to form extensive duplex structures may reflect RNA-silencing activity. Two such small RNAs, one corresponding to a Ser carboxypeptidase-like gene and one corresponding to a gene of unknown function, were detected by blot assay or identified in the cloned library (Table 3, Figure 6). As with the artificial 35S:dsGFP RNA-silencing inducer, extensive duplex structures in mRNA or pre-mRNA may be tar-

geted efficiently for cleavage. These small RNAs then would function in the subsequent RNA silencing of homologous and closely related sequences.

### Endogenous Small RNAs and Signaling

The finding of so many small RNAs in plants and animals leads to the logical conclusion that a broad range of processes are affected by this class of RNA. Given the size of these species and the known capacity of intercellular junctions to transport nucleic acids in plants, it is reasonable to propose that small RNAs serve as local or systemic signaling molecules. Viral nucleoprotein complexes, proteins, and functional mRNAs all are known to move from cell to cell and long distances in plants (Wu et al., 2002).

Cell-to-cell transport of small RNAs through plasmodesmata may enable signaling for post-transcriptional or epigenetic control in cells near sites of small RNA production. Long-distance transport of small RNAs through the phloem could contribute to certain types of regulation that occur systemically (Kuhn et al., 1997; Voinnet and Baulcombe, 1997; Palauqui and Vaucheret, 1998; Ruiz-Medrano et al., 1999; Kim et al., 2001). The major challenges that lie ahead are the identification of targets for small RNAs and understanding the function of small RNA-target interactions.

## METHODS

### Plants, Plasmids, Strains, and *Agrobacterium tumefaciens* Injection

Nontransgenic *Nicotiana benthamiana* and *Arabidopsis thaliana* ecotype Columbia-0 were used. The 35S:green fluorescent protein (GFP) construct (pRTL2-smGFP) contained a functional copy of the soluble modified GFP coding sequence. The intron-spliced, inverted-repeat form of the GFP sequence (35S:dsGFP) construct (pRTL2-dsGFP) contained both sense and antisense soluble modified GFP sequences separated by a 120-nucleotide intron of the *RTM1* gene (Chisholm et al., 2000), as described previously (Johansen and Carrington, 2001). These plasmids contained the 35S promoter and terminator sequences. Expression cassettes were inserted into vector pSLJ75515, and the resulting constructs were introduced by triparental mating into *Agrobacterium tumefaciens* strain GV2260 (Johansen and Carrington, 2001).

*Agrobacterium* infiltration of leaves was performed as described (Llave et al., 2000). Bacteria containing either the 35S:GFP or 35S:dsGFP construct were resuspended in infiltration solution to 1 and 0.5 OD<sub>600</sub>, respectively. Infiltrated leaves were harvested after 6 to 7 days and used for RNA isolation.

### RNA Isolation and Blot Analysis

Total RNA was extracted using TRIZOL reagent (Bethesda Research Laboratories Life Technologies) as described (Johansen and

Carrington, 2001). Low molecular weight RNA was isolated subsequently by anion-exchange chromatography (RNA/DNA Midi Kit; Qiagen, Valencia, CA) according to the manufacturer's instructions. For amplification and cloning experiments, small RNAs were further gel purified as follows. RNA was resolved by electrophoresis on 15% polyacrylamide (30:0.8) gels containing 7 M urea in TBE buffer (45 mM Tris-borate, pH 8.0, and 1.0 mM EDTA). A gel slice containing RNAs of ~15 to 35 nucleotides (based on RNA oligonucleotide size standards) was excised and eluted in 0.3 M NaCl at 4°C for ~16 h. The eluted RNAs were precipitated using ethanol and resuspended in diethyl pyrocarbonate-treated deionized water.

Blot hybridization analysis was performed as described (Llave et al., 2000) with some modifications. Small RNA was resolved on denaturing 15% polyacrylamide (30:0.8) gels, electroblotted to Nytran SuperCharge membranes (Schleicher & Schuell) using a trans-blot semidry transfer cell (Bio-Rad) for 1 h at 400 mA, and UV cross-linked. Radiolabeled probes for the GFP coding sequence were made by random priming reactions in the presence of  $\alpha$ -<sup>32</sup>P-dATP. DNA oligonucleotides complementary to small RNA sequences were end labeled with  $\gamma$ -<sup>32</sup>P-ATP using T4 polynucleotide kinase (New England Biolabs, Beverly, MA). Blots used for the analysis of small RNA sequences were prehybridized and hybridized using PerfectHyb Plus buffer (Sigma). Hybridization of blots using random-primed GFP probes was performed at 38°C, whereas hybridization of blots using oligonucleotide probes was performed at 20°C below the calculated dissociation temperature ( $T_d$ ) [ $T_d$  (°C) = 4(G+C) + 2(A+T)]. Blots were washed at 50°C as described (Llave et al., 2000).

Densitometric analysis of at least four independent RNA gel blots exposed to x-ray film was used to assess the relative levels of small RNA accumulation. Synthetic oligoribonucleotides of 21 and 24 nucleotides, with 5' phosphate and 3' hydroxyl ends, were used as size standards during electrophoresis. Ethidium bromide staining and visualization of the 5S RNA/tRNA bands in small RNA gels were used to monitor the loading of RNA samples.

Amplification products from small RNA cloning experiments were resolved on 2% agarose gels in TBE buffer. Products from GFP-derived small RNAs were subjected to DNA gel blot analysis as described (Llave et al., 2000), except that hybridizations were performed at 40°C using a radiolabeled GFP probe.

#### Amplification and Cloning of Small Interfering RNAs from *N. benthamiana*

Small RNA was isolated from mature *N. benthamiana* leaves (8 to 10 g) that were infiltrated with *Agrobacterium* containing either 35S:GFP or 35S:dsGFP constructs (Johansen and Carrington, 2001). Gel-purified small interfering RNA (siRNA) molecules were ligated sequentially to 5' and 3' RNA oligonucleotide adapters using T4 RNA ligase (1 unit/ $\mu$ L) (New England Biolabs). 2'-ACE-protected RNA oligonucleotides (Dharmacon Research, Boulder, CO) were designed to contain either 5' or 3' end modifications that prevented self-ligation. The 5' RNA adapter 1 (5'-CACUGCGUUUGCUGGCUUUGGCUGAUG-GCG-3') possessed 5' and 3' hydroxyl groups. The 3' RNA adapter 2 (5'-CAGUAGUCGUCAAGAGUGGA-idT-3') possessed a 5' monophosphate and a 3' inverted deoxythymidine (idT). RNA oligonucleotides were 2' deprotected according to the instructions of the manufacturer and gel purified as described previously.

The siRNAs were first ligated to the 5' RNA adapter 1. The ligation products were gel eluted and ligated to the 3' RNA adapter 2 as described above. Individual ligation reactions of siRNA to each adapter

were performed as controls, and reaction products were monitored by electrophoresis on 15% polyacrylamide gels and RNA gel blots. The final ligation product then was used as a template in a reverse transcription reaction using the reverse primer (5'-ACTCTTGAC-GACTAC-3') and Superscript II reverse transcriptase (Bethesda Research Laboratories Life Technologies). This was followed by PCR using the reverse and forward (5'-CGTTTGCTGGCTTTG-3') primers and Taq DNA polymerase.

Amplification products were purified by phenol-chloroform extraction and ethanol precipitation and then ligated into the T/A site of a pGEM-T Easy vector (Promega) using T4 DNA ligase (Promega). Electroporation-competent *Escherichia coli* DH10B cells (Bethesda Research Laboratories Life Technologies) were transformed with the siRNA library. Positive colonies were identified by PCR using the M13F (5'-CCCAGTCACGACGTTGATAAACGACG-3') and M13R (5'-GCG-GATAACAATTTACACAGGAAACAGC-3') primers followed by DNA gel blot hybridization using a radiolabeled GFP probe. Plasmids were extracted from positive colonies, and the inserts were sequenced.

#### Amplification and Cloning of Small RNAs from Arabidopsis

Small RNA was isolated from inflorescence tissue of Arabidopsis Columbia-0 (3 to 4 g), gel purified, and ligated sequentially to RNA adapters 1 and 2 as described above. The final ligation product was reverse transcribed and amplified by PCR using a reverse transcriptase primer (5'-TTTTCTGCAGTCCACTCTTGACGAC-3') that contained a PstI restriction site and a forward primer (5'-AAACCATGGCACTGCGTTTGCTGGC-3') that contained a NcoI restriction site (sites are indicated in boldface). The PCR products were phenol extracted, precipitated with ethanol, and digested with PstI and NcoI. After additional phenol extraction and ethanol precipitation, the digested products were ligated to a PstI-NcoI-linearized pGEM-T Easy vector using T4 DNA ligase. The inserts from individual colonies were amplified by PCR using the M13F and M13R primers and sequenced.

Small RNAs were named with numerical suffixes (e.g., small RNA 39). In cases in which a small RNA sequence corresponded to multiple genome loci, the loci number is indicated after the small RNA number (e.g., small RNA 39.3). Detailed information about each small RNA is provided in the supplemental material.

#### Computer Analysis of Small RNA Sequences from Arabidopsis

The small RNA sequences were used to search the GenBank database using BLAST. Only those clones that contained matches to one or more Arabidopsis genome sequences were analyzed further.

The potential secondary structure of RNA surrounding small RNA sequences was predicted using *mfold* version 3.1 (Mathews et al., 1999) (<http://bioinfo.math.rpi.edu>). For IGR-derived small RNAs, secondary structure predictions were made using sense and antisense sequences containing 100 nucleotides (based on genomic sequence) on either side of the small RNA or clusters of small RNAs. For protein-coding gene-derived small RNAs, sense and antisense sequences of up to 3000 nucleotides flanking the small RNAs were analyzed. Phylogenetic trees were generated using Clustal W 1.8 (Thompson et al., 1994; Higgins et al., 1996).

Upon request, all novel material described in this article will be made available in a timely manner for noncommercial research purposes. No restrictions or conditions will be placed on the use of any

materials described in this article that would limit their use in non-commercial research.

#### Accession Numbers

GenBank accession numbers for the small RNAs are provided in the supplemental material. GenBank accession numbers for small RNAs cited in the text are as follows: small RNA 2 (AF501743), small RNA 4 (AF501745), small RNA 5 (AF501746), small RNA 6 (AF501747), small RNA 14 (AF501755), small RNA 17 (AF501758), small RNA 19 (AF501760), small RNA 23 (AF501764), small RNA 24 (AF501765), small RNA 30 (AF501771), small RNA 32 (AF501773), small RNA 35 (AF501776), small RNA 39 (AF501780), small RNA 40 (AF501781), small RNA 42 (AF501783), small RNA 49 (AF501790), small RNA 53 (AF501794), small RNA 55 (AF501796), small RNA 64 (AF501805), small RNA 67 (AF501808), small RNA 71 (AF501812), small RNA 77 (AF501818), small RNA 82 (AF501823), small RNA 85 (AF501826), small RNA 88 (AF501829), small RNA 90 (AF501831), small RNA 96 (AF501837), small RNA 97 (AF501838), small RNA 100 (AF501841), small RNA 101 (AF501842), small RNA 104 (AF501845), small RNA 105 (AF501846), small RNA 107 (AF501848), small RNA 109 (AF501850), small RNA 110 (AF501851), and small RNA 124 (AF501865).

#### ACKNOWLEDGMENTS

We thank Lisa Johansen, Stephen Chisholm, Zhixin Xie, and Robert Anderberg for valuable comments and advice during the course of this work. We also thank Ed Allen for constructing the *Scarecrow* gene family dendrogram. C.L. was the recipient of a postdoctoral fellowship from the Ministerio de Educacion y Cultura (Spain). This work was supported by Grants AI27832 and AI43288 from the National Institutes of Health and by Grant NRI 2002-35319-11560 from the U.S. Department of Agriculture.

Received March 19, 2002; accepted April 15, 2002.

#### REFERENCES

- Ambros, V.** (2000). Control of developmental timing in *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* **10**, 428–433.
- Ambros, V., and Horvitz, H.R.** (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226**, 409–416.
- Anandalakshmi, R., Pruss, G.J., Ge, X., Marathe, R., Smith, T.H., and Vance, V.B.** (1998). A viral suppressor of gene silencing in plants. *Proc. Natl. Acad. Sci. USA* **95**, 13079–13084.
- Arabidopsis Genome Initiative.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Banerjee, D., and Slack, F.** (2002). Control of developmental timing by small temporal RNAs: A paradigm for RNA-mediated regulation of gene expression. *Bioessays* **24**, 119–129.
- Bass, B.L.** (2000). Double-stranded RNA as a template for gene silencing. *Cell* **101**, 235–238.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J.** (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366.
- Bolle, C., Koncz, C., and Chua, N.H.** (2000). PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Genes Dev.* **14**, 1269–1278.
- Brigneti, G., Voinnet, O., Wan-Xiang, L., Ding, S.W., and Baulcombe, D.C.** (1998). Viral pathogenicity determinants are suppressors of transgene silencing. *EMBO J.* **17**, 6739–6746.
- Carrington, J.C., Kasschau, K.D., and Johansen, L.K.** (2001). Activation and suppression of RNA silencing by plant viruses. *Virology* **281**, 1–5.
- Carthew, R.W.** (2001). Gene silencing by double-stranded RNA. *Curr. Opin. Cell Biol.* **13**, 244–248.
- Chisholm, S.T., Mahajan, S.K., Whitham, S.A., Yamamoto, M.L., and Carrington, J.C.** (2000). Cloning of the Arabidopsis *RTM1* gene, which controls restriction of long-distance movement of tobacco etch virus. *Proc. Natl. Acad. Sci. USA* **97**, 489–494.
- Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.C., Picard, G., and Tourmente, S.** (2002). Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Res.* **12**, 132–144.
- Dalmay, T., Hamilton, A., Mueller, E., and Baulcombe, D.C.** (2000a). Potato virus X amplicons in Arabidopsis mediate genetic and epigenetic gene silencing. *Plant Cell* **12**, 369–379.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C.** (2000b). 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.
- DiLaurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N.** (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. *Cell* **86**, 423–433.
- Elbashir, S.M., Lendeckel, W., and Tuschl, T.** (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200.
- Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Bailly, D.L., Fire, A., Ruvkun, G., and Mello, C.C.** (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34.
- Grosshans, H., and Slack, F.J.** (2002). Micro-RNAs: Small is plentiful. *J. Cell Biol.* **156**, 17–21.
- Hamilton, A.J., and Baulcombe, D.C.** (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952.
- Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J.** (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296.
- Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J.** (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150.
- Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N.** (2000). The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell* **101**, 555–567.
- Higgins, D.G., Thompson, J.D., and Gibson, T.J.** (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402.

- Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D.** (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838.
- Johansen, L.K., and Carrington, J.C.** (2001). Silencing on the spot: Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system. *Plant Physiol.* **126**, 930–938.
- Jones, A.L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J., and Baulcombe, D.C.** (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291–2302.
- Jones, L., Ratcliff, F., and Baulcombe, D.C.** (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.* **11**, 747–757.
- Kasschau, K.D., and Carrington, J.C.** (1998). A counter-defensive strategy of plant viruses: Suppression of posttranscriptional gene silencing. *Cell* **95**, 461–470.
- Kasschau, K.D., and Carrington, J.C.** (2001). Long-distance movement and replication maintenance functions correlate with silencing suppression activity of potyviral HC-Pro. *Virology* **285**, 71–81.
- Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.** (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659.
- Ketting, R.F., Haverkamp, T.H., vanLuenen, H.G., and Plasterk, R.H.** (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of the Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141.
- Kim, M., Canio, W., Kessler, S., and Sinha, N.** (2001). Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**, 287–289.
- Knight, S.W., and Bass, B.L.** (2001). A role for the RNase III enzyme CDR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271.
- Kuhn, C., Franceschi, V.R., Schulz, A., Lemoine, R., and Frommer, W.B.** (1997). Macromolecular trafficking indicated by localization and turnover of sucrose transporters in enucleate sieve elements. *Science* **275**, 1298–1300.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T.** (2001). Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858.
- Lau, N.C., Lim, E.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.
- Le, Q.H., Wright, S., Yu, Z., and Bureau, T.** (2000). Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97**, 7376–7381.
- Lee, R.C., and Ambros, V.** (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864.
- 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.
- Lipardi, C., Wei, Q., and Paterson, B.M.** (2001). RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107**, 297–307.
- Llave, C., Kasschau, K.D., and Carrington, J.C.** (2000). Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proc. Natl. Acad. Sci. USA* **97**, 13401–13406.
- Mathews, D.H., Sabina, J., Zucker, M., and Turner, D.H.** (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940.
- Matzke, M., Matzke, A.J., and Kooter, J.M.** (2001). RNA: Guiding gene silencing. *Science* **293**, 1080–1083.
- Mette, M.F., Matzke, A.J., and Matzke, M.A.** (2001). Resistance of RNA-mediated TGS to HC-Pro, a viral suppressor of PTGS, suggests alternative pathways for dsRNA processing. *Curr. Biol.* **11**, 1119–1123.
- 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., et al.** (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542.
- Olsen, P.H., and Ambros, V.** (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680.
- Palauqui, J.C., and Vaucheret, H.** (1998). Transgenes are dispensable for the RNA degradation step of cosuppression. *Proc. Natl. Acad. Sci. USA* **95**, 9675–9680.
- 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.
- Ruiz-Medrano, R., Xoconostle-Cazares, B., and Lucas, W.J.** (1999). Phloem long-distance transport of CmNACP mRNA: Implications for supracellular regulation in plants. *Development* **126**, 4405–4419.
- Ruvkun, G.** (2001). Molecular biology: Glimpses of a tiny RNA world. *Science* **294**, 797–799.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H.A., and Fire, A.** (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 1–20.
- Sijen, T., and Kooter, J.M.** (2000). Post-transcriptional gene-silencing: RNAs on the attack or on the defense? *Bioessays* **22**, 520–531.
- Silverstone, A.L., Ciampaglio, C.N., and Sun, T.** (1998). The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* **10**, 155–169.
- Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G.** (2000). The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669.
- Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M.** (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178.
- Smith, C.M., and Steitz, J.A.** (1997). Sno storm in the nucleolus: New roles for myriad small RNPs. *Cell* **89**, 669–672.
- Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C.** (1999). The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J.** (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Tollervey, D.** (1996). Small nucleolar RNAs guide ribosomal RNA methylation. *Science* **273**, 1056–1057.

- Vance, V., and Vaucheret, H.** (2001). RNA silencing in plants: Defense and counterdefense. *Science* **292**, 2277–2280.
- Voinnet, O.** (2001). RNA silencing as a plant immune system against viruses. *Trends Genet.* **17**, 449–459.
- Voinnet, O., and Baulcombe, D.C.** (1997). Systemic signalling in gene silencing. *Nature* **389**, 553–555.
- Voinnet, O., Pinto, V.M., and Baulcombe, D.C.** (1999). Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proc. Natl. Acad. Sci. USA* **96**, 14147–14152.
- Wassenegger, M.** (2000). RNA-directed DNA methylation. *Plant Mol. Biol.* **43**, 203–220.
- Waterhouse, P.M., Wang, M.B., and Lough, T.** (2001). Gene silencing as an adaptive defence against viruses. *Nature* **411**, 834–842.
- 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.
- Wu, X., Weigel, D., and Wigge, P.A.** (2002). Signaling in plants by intercellular RNA and protein movement. *Genes Dev.* **16**, 151–158.
- Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P.** (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33.