

AN ABSTRACT OF THE DISSERTATION OF

Josh Cuperus for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on April 1, 2011.

Title: Processing and Functional Diversity of Plant MicroRNA.

Abstract approved:

James C. Carrington

RNA interference pathways can involve amplification of secondary siRNAs by RNA-dependent RNA polymerases. In plants, RDR6-dependent secondary siRNAs, including transacting siRNAs (tasiRNAs), arise from transcripts targeted by some microRNAs (miRNAs). In the case of TAS3 tasiRNA formation, ARGONAUTE7 (AGO7)-miR390 complexes interact with primary transcripts at two sites, resulting in recruitment of RNA-DEPENDENT RNA POLYMERASE6 for dsRNA biosynthesis. An extensive screen for Arabidopsis mutants with specific defects in TAS3 tasiRNA biogenesis or function was done. This yielded *ago7* mutants, *dcl4* mutants, and two mutants that accumulated low levels of miR390. A direct genome sequencing-based approach to both map and rapidly identify one of the latter mutant alleles was developed. This revealed a G-to-A point mutation (*mir390a-1*) that was calculated to stabilize a relatively nonpaired region near the base of the *MIR390a* foldback, resulting in misprocessing of the miR390/miR390* duplex and subsequent reduced TAS3 tasiRNA levels. Directed substitutions, as well as analysis of variation at paralogous miR390-generating loci (*MIR390a* and *MIR390b*), indicated that base-pair properties and nucleotide identity within a region 4-6 bases below the miR390/miR390* duplex region contributed to the efficiency and accuracy of precursor processing. *Arabidopsis thaliana* secondary siRNAs from mRNA as well as *TAS1/ TAS2*

and *TAS4* trans-acting siRNAs are shown to be triggered through initial targeting by a 22-nucleotide (nt) miRNA that associates with AGO1. In contrast to canonical 21-nt miRNAs, 22-nt miRNAs primarily arise from foldback precursors containing asymmetric bulges. Using artificial miRNA constructs, conversion of asymmetric foldbacks to symmetric foldbacks resulted in the production of 21-nt forms of miR173, miR472 and miR828. Both 21- and 22-nt forms associated with AGO1 and guided accurate slicer activity, but only 22-nt forms were competent to trigger RDR6-dependent siRNA production from target RNA. These data suggest that a region below the miR/miR* duplex contributes to both mature miRNA accumulation as well as precursor processing accuracy and that AGO1 functions differentially with 21- and 22-nt miRNAs to engage the RDR6-associated amplification apparatus.

©Copyright by Josh Cuperus

April 1, 2011

All Rights Reserved

Processing and Functional Diversity of Plant MicroRNA

by
Josh Cuperus

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 1, 2011
Commencement June 2011

Doctor of Philosophy dissertation of Josh Cuperus presented on April 1, 2011.

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Josh Cuperus, Author

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the following people: Jim Carrington for mentoring me during my thesis work; Kristin Kasschau, Alberto Carbonell, Hernan Garcia-Ruiz, Miya Howell, Atsushi Takeda, and Taiowa Montgomery for guidance; Chris Sullivan and Noah Fahlgren for computational guidance; fellow graduate students Elisabeth Chapman, Tai Montgomery, Noah Fahlgren, Sanjuro Jogdeo, Seong Jeon Yoo and Steen Hoyer for discussions; Russel Burke, Tiffany Townsend, Sarah Dvorak, Annie Bailey, Anna Scalley and Amy Shatswell for assistance with experiments and plant care; the CGRB office staff Rosa Hill, Gail Millimaki and Goretti Nguyen; Mark Dasenko and Matthew Peterson for assistance with Illumina sequencing; our collaborators from the labs of Detlef Weigel, Steven Jacobsen, and many others; fellow MCB students; and friends and family, especially for Marissa.

CONTRIBUTION OF AUTHORS

Chapter 1: Joshua T. Cuperus, Ignacio Rubio-Somoza, Detlef Weigel and James C. Carrington wrote the manuscript.

Chapter 2: Joshua T. Cuperus, Taiowa A. Montgomery, Russel T. Burke, Tiffany Townsend, Maxim Greenberg, and Keithanne Mockiatis performed research. Joshua T. Cuperus, Christopher M. Sullivan and Noah Fahlgren analyzed sequencing data. Joshua T. Cuperus and James C. Carrington wrote the manuscript.

Chapter 3: Joshua T. Cuperus, Taiowa A. Montgomery, Russel T. Burke and Tiffany Townsend performed research. Joshua T. Cuperus, Christopher M. Sullivan and Noah Fahlgren analyzed sequencing data. Joshua T. Cuperus and James C. Carrington wrote the manuscript.

Chapter 4: Joshua T. Cuperus, Alberto Carbonell and Russel T. Burke performed infiltration assays. Joshua T. Cuperus, Alberto Carbonell, Hernan Garcia-Ruiz and Sunny D. Gilbert were involved in preparing sequencing libraries. Joshua T. Cuperus, Christopher M. Sullivan and Noah Fahlgren analyzed sequencing data. Joshua T. Cuperus, James C. Carrington, Noah Fahlgren and Alberto Carbonell wrote the manuscript.

Chapter 5: Joshua T. Cuperus, James C. Carrington and Noah Fahlgren wrote the manuscript. Additional sections are not reproduced here.

TABLE OF CONTENTS

	<u>Page</u>
Introduction	1
EUKARYOTIC RNA SILENCING	2
PLANT ENDOGENOUS SILENCING PATHWAYS	3
Small interfering RNAs	3
MicroRNA	4
Conservation and diversification of microRNA processing	5
<i>MIRNA</i> foldback recognition features	5
Conserved size variation in mature miRNA	7
miRNA-triggered production of 21-nucleotide siRNA	8
MicroRNA-directed DNA methylation	10
Transcriptional regulation of miRNA and tasiRNA precursors	11
AGO-specific small RNA functionality	12
<i>TAS1</i> and <i>TAS3</i> -based forward genetic screen and identification of mutants in Arabidopsis by direct genome sequencing	19
SUMMARY	20
INTRODUCTION	20
RESULTS	23
Mutagenesis of transgenic Arabidopsis expressing <i>TAS1c</i> and <i>TAS3a</i> -based syn- tasiRNA	23
Map-assisted sequencing and identification of <i>mir390a-1</i>	26
Map assisted sequencing and identification of <i>nripd/e2-19</i>	28
DISCUSSION	29
MATERIALS AND METHODS	31
Plant Material	31
Mutagenesis of <i>35S:TAS3aPDS-1</i> and <i>35S:TAS1cPDS-2</i> -transformed plants	31
Transgene constructs	32
Conventional sequencing	32
RNA blot assays	32
Genomic sequencing	33
MASS (<u>M</u> apping and <u>A</u> ssembly with <u>S</u> hort <u>S</u> equences)	33
ACKNOWLEDGMENTS	34
<i>MIR390a</i> precursor processing-defective mutants in Arabidopsis	47
SUMMARY	48
INTRODUCTION	48
RESULTS	49

TABLE OF CONTENTS (Continued)

	<u>Page</u>
Defective Processing of the <i>mir390a-1</i> Foldback	49
Mutational Analysis of the <i>MIR390a</i> foldback.....	51
DISCUSSION.....	53
MATERIALS AND METHODS	54
Mutagenesis of <i>MIR390a</i>	54
RNA folding	55
Transgene constructs.....	55
Small RNA Sequencing from Transient Expression Assays.	55
ACKNOWLEDGMENTS.....	56
Unique Functionality of 22 nt miRNAs in Triggering RDR6-Dependent siRNA Biogenesis from Target Transcripts in Arabidopsis.....	64
SUMMARY	65
INTRODUCTION.....	65
RESULTS.....	67
Secondary siRNAs originate from some targeted mRNAs	67
Asymmetric <i>MIRNA</i> foldbacks yield 22-nucleotide miRNA.....	69
22-nucleotide forms of miR173, miR472, and miR828 trigger siRNAs.....	71
DISCUSSION.....	73
METHODS	77
Accession codes.....	77
Plant materials and growth conditions.....	77
Transgene constructs.....	77
Transient expression assays.....	77
RNA Blot assays.....	78
Small RNA sequencing.....	78
Small RNA analysis.....	79
Statistical analyses.....	79
5' RACE.....	79
Small RNA datasets.....	80
<i>MIRNA</i> secondary and tertiary structure prediction.....	80
ACKNOWLEDGMENTS.....	81
SUPPLEMENTAL DATA.....	93
General Conclusions	94
Forward genetics and identification of causative mutations	95

TABLE OF CONTENTS (Continued)

	<u>Page</u>
Processing of miRNA in plants	96
miRNA-trigger siRNA.....	97
Bibliography	100

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1. MicroRNA processing in plants.....	16
1.2. Functional cross-talk of the microRNA pathway in plants.....	18
2.1 Syn-tasiRNA strategy and mutant screen.....	35
2.2. <i>TAS1cPDS-2</i> mutant analysis.	36
2.3 Characterization of mutants with reduced <i>TAS1</i> tasiRNA	37
2.4 Characterization of <i>TAS3aPDS-1</i> class I, II and III mutants.	38
2.5. Arabidopsis <i>AGO7</i> mRNA and the position of <i>ago7</i> alleles.....	40
2.6. Identification of the causal mutation in 52b2 mutant.....	41
2.7. Col-0, <i>TAS3aPDS-1</i> , and <i>mir390a-1</i> rosettes.....	42
2.8. MASS identification of the causal mutation in the <i>m48</i> mutant.....	43
3.1. Foldback processing and <i>TAS3</i> tasiRNA-initiation activity of <i>mir390a-1</i> in transient assays.....	57
3.2. Small RNA read percentages mapped to the <i>ath-MIR390a</i> locus from transient overexpression assays.	58
3.3. Directed mutational analysis of the <i>MIR390a</i> foldback.....	61
4.1. Arabidopsis small RNA distribution.....	82
4.2. Small RNA from Arabidopsis annotated transcripts.....	83
4.3. <i>MIRNA</i> foldback asymmetry leads to formation of 22-nucleotide miRNAs	84
4.4. Production and activities of 21 and 22-nucleotide miR173 forms.....	85
4.5 Mapping of miR173 or miR173-related reads, and <i>TAS1c</i> transcript-derived reads, from the transient expression assay data shown in 3c.	86
4.6 miRNA accumulation in the Arabidopsis <i>dcl234</i> triple mutant.	87
4.7 Enrichment or depletion of miRNA in HA-AGO1 coimmunoprecipitates.	88
4.8. Role of the 3' end nucleotide in miR173.	90
4.9. Production and activities of 21 and 22-nucleotide miR472 and miR828 forms.	91
4.10. Predicted tertiary structures of <i>MIR390a</i> -derived amiR173-21 and amiR173 foldbacks using MC-Fold and MC-Sym.	92

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1. Position of <i>ago7</i> and <i>dcl4</i> mutations.....	44
2.2. Polymorphisms in the 1.52 Mb mapping interval in 52b2 mutant.....	45
2.3. Polymorphisms in the 3 Mb mapping interval in <i>m48</i> mutant.....	46
3.1. Small RNA reads mapping within a 29nt window centered on miR390 or miR390* from transient expression assay.	62
3.2. Primers used to design MIR390-based constructs.....	63

Introduction

Josh T. Cuperus, Ignacio Rubio-Somoza, Detlef Weigel and James C. Carrington

The Plant Cell
American Society of Plant Biologists
15501 Monona Drive
Rockville, Maryland 20855-2768
PMID: 21317375

Current Opinion in Plant Biology
<http://www.current-opinion.com/jpbl/about.htm?jcode=jpbl>
Volume 12 (5), 1 October 2009, Pages 622-627

EUKARYOTIC RNA SILENCING

Small RNAs direct transcriptional, posttranscriptional, and DNA methylation-based silencing activities that shape the transcriptomes and proteomes of eukaryotic organisms. These small RNAs include microRNA (miRNA), several classes of endogenous small interfering RNA (siRNA), Piwi-associated RNA (piRNA), and other classes that are differentiated by their distinct biogenesis pathways (Carthew and Sontheimer, 2009; Malone and Hannon, 2009). A unifying theme of all eukaryotic small RNAs is that they are loaded into effector complexes comprised of an ARGONAUTE (AGO) family member protein. AGO family members, guided by small RNA, make sequence-specific contact with target nucleic acids (primarily RNA), leading to silencing activities. AGO-guide-target complexes inhibit transcripts either directly (posttranscriptionally) or indirectly (transcriptionally at target loci) (Carthew and Sontheimer, 2009). The functional diversity of small RNA is, in part, due to association with different AGO effectors.

The majority of small RNA arise from processing of long double-stranded RNA by a member of the Dicer family. The size of small RNA derived from double-stranded RNA is dependent on association with different Dicer proteins. Dicer family members contain a PAZ domain that binds the 2-nucleotide, single-stranded 3' ends of double-stranded RNA. PAZ binding positions the RNase III domains of Dicer proteins for cleavage, resulting in a small RNA duplex of a specific size (Sashital and Doudna, 2010). The source of double-stranded RNA is pathway dependent and can arise from naturally occurring overlapping sense/antisense transcripts, through self-complementarity of single-stranded RNA transcripts, or in some organisms through the activity of RNA-dependent RNA polymerases (RdRp). RdRPs are widely distributed throughout eukaryotes, but have been lost in Vertebrates and flies.

PLANT ENDOGENOUS SILENCING PATHWAYS

In plants, endogenous small RNA pathways have diversified, leading to many functionally distinct classes, which involve members of the DICER-LIKE (DCL) family (4 members in Arabidopsis), AGO family (10 members in Arabidopsis), and sometimes RNA-DEPENDENT RNA POLYMERASE (RDR) family (6 members in Arabidopsis). The expansion of small RNA-associated proteins has also led to the specialization of endogenous small RNA pathways. These diverse pathways include miRNA, siRNA and a specialized class of siRNAs, trans-acting siRNAs.

Small interfering RNAs

Small interfering RNAs (siRNAs) are produced in three main sizes, 21, 22, and 24 nucleotides. Twenty-four nucleotide siRNAs are broadly spread throughout the genomes of plants, and are enriched in centromeric and pericentromeric regions, or other regions that are enriched in transposons and retrotransposons that are highly repressed by DNA methylation and heterochromatic structure (Lu et al., 2006; Rajagopalan et al., 2006; Kasschau et al., 2007; Lister et al., 2008). The 24-nucleotide siRNAs from these regions guide DNA methylation at target loci, a mechanism only known to occur in plants (Zhang et al., 2006; Henderson and Jacobsen, 2007; Zilberman et al., 2007). The exact mechanism is still poorly understood, however the plant-specific nuclear RNA polymerase Pol IV produces single-stranded RNA transcripts, which are in turn amplified into double-stranded RNA by RDR2 and diced into 24-nucleotide duplexes by DCL3 (Law and Jacobsen, 2010). These 24-nucleotide siRNAs associate with an AGO4-family member (AGO4, AGO6 or AGO9), and target transcripts produced by another plant specific polymerase, Pol V, at repetitive loci (Qi et al., 2006; Mi et al., 2008; Havecker et al., 2010) (Wu et al., 2010). It is thought that Pol V transcripts act as scaffolds that recruit AGO4-siRNA complexes, which recruit DNA methylation machinery, including the DNA methylase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) that marks DNA with repressive cytosine methylation, to target loci (Wierzbicki et al., 2008; Wierzbicki et al., 2009). This RNA-directed DNA methylation (RdDM) system acts as a

mechanism to detect and silence mobile elements in plant genomes. Both 21- and 22-nucleotide siRNAs are not as widely spread through plant genomes, and primarily act post-transcriptionally to silence RNA in trans. Highly repetitive regions of the genome, including mobile elements can be targeted by 21, 22 and 24-nucleotide siRNA mechanisms.

MicroRNA

miRNAs arise from transcripts containing self-complementary foldback structures that are initially processed to form 21-22nt miRNA/miRNA* duplexes. In animals, primary transcripts with miRNA foldbacks (pri-miRNA) are processed first by the Microprocessor complex, which contains the RNase III-type protein Drosha and its cofactor Pasha (also known as DGCR8 in humans) (Carthew and Sontheimer, 2009). Pasha/DGCR8 binds to pri-miRNAs and directs Drosha cleavage at ~11bp away from the ssRNA-dsRNA junction at the base of the stem (Denli et al., 2004; Han et al., 2004a; Zeng and Cullen, 2005; Zeng et al., 2005). Following export of the miRNA precursor (pre-miRNA) from the nucleus, Dicer, with partners that include the dsRNA-binding domain protein Loquacious, cleaves near the terminal loop, releasing ~22nt miRNA duplexes (Chendrimada et al., 2005; Forstemann et al., 2005; Haase et al., 2005; Jiang et al., 2005; Saito et al., 2005). Drosha-mediated cleavage, therefore, generates one end of the miRNA/miRNA* duplex, while Dicer “measures” ~22nt from Drosha cleavage sites and catalyzes the second set of cuts.

Plants orchestrate both pri-miRNA and pre-miRNA processing with the same (or very similar) complex, which includes the RNase-III like enzyme DICER-LIKE1 (DCL1) as the catalytic component (Golden et al., 2002; Park et al., 2002; Reinhart et al., 2002; Kurihara et al., 2006). DCL1 interacts with the double-stranded RNA binding protein HYPONASTIC LEAVES1 (HYL1), the zinc-finger protein SERRATE (SE) and the RNA-binding protein DAWDLE (DDL), all of which are generally necessary for efficient and accurate miRNA biogenesis (Han et al., 2004b; Hiraguri et al., 2005; Kurihara et al., 2006; Yang et al., 2006; Dong et al., 2008; Yu et al., 2008; Szarzynska et al., 2009). While HYL1 appears to be

uniquely associated with miRNA processing, SE has been implicated in both miRNA processing and mRNA splicing (Vazquez et al., 2004a; Laubinger et al., 2008). The specific roles of HYL1 and SE as cofactors during miRNA processing are poorly understood, and the diversity of plant miRNA foldback sizes and structures suggests that they may function with less rigid specificity than the non-catalytic cofactors during animal miRNA processing (Axtell, 2008). After DCL1-mediated processing, miRNA duplexes are methylated by the 3'-methyltransferase HUA ENHANCER1 (HEN1), and at least some miRNA may be transported to the cytoplasm by the exportin-5-like protein HASTY (Chen et al., 2002; Bollman et al., 2003; Yu et al., 2005).

Conservation and diversification of microRNA processing

The transition of a young *MIRNA* locus from a nascent duplication product to one yielding a transcript that is recognized by the canonical *MIRNA* foldback processing machinery (involving DCL1, HYL1, SE; (Voinnet, 2009)) presumably involves neutral mutations. It is clear that plant *MIRNA* foldbacks can qualify over a range of structures, as foldback sequence lengths in many species vary between less than 100 to over 900 nucleotides in length. This is in striking contrast to animal species foldback lengths, which are highly uniform and predominantly less than 100 nucleotides (Figure 1.1A). Several recent studies of *P. patens* and *A. thaliana* *MIRNA* have shed light on plant miRNA processing mechanisms. In this section, the "criteria" that qualify a transcript for miRNA processing in plants are discussed.

***MIRNA* foldback recognition features**

Several recent studies revealed two structural features that confer optimized foldback processing characteristics. One feature was identified in mutants with defects in *ath-MIR390a*, *ath-MIR172a*, *ath-MIR171a*, *ath-MIR167a*, *ath-MIR164c* and *ath-MIR398a*, with substitutions affecting the region 1 to 8 bp below the miRNA/miRNA* duplex and resulting in reduced or inaccurate miRNA accumulation (Cuperus et al., 2010a; Mateos et al., 2010; Song et al.,

2010; Werner et al., 2010). For comparison, single nucleotide mutations in the loop-proximal region, in many cases, were considerably less deleterious, or had no effect, on miRNA accumulation (Mateos et al.; Song et al., 2010; Werner et al.). Mutations in the region below the miRNA/miRNA* duplex that maintained wild type structure were generally neutral, while mutations that opened or closed predicted bulges reduced the accumulation or altered the processing of the mature miRNA (Cuperus et al., 2010a; Mateos et al., 2010; Song et al., 2010; Werner et al., 2010). In plant *MIRNA* foldbacks, this region is characterized by relatively weak base-pairing with high number of 1-3 nucleotide bulges (Figure 1.1B). In the case of *MIR390a* and *MIR390b*, which specify the identical miR390 sequence, natural variation in the strength of base-pairing in this region may account for differences in miR390 expression from each locus (Cuperus et al., 2010a). Additionally, the *ath-mir390a-1* point mutation, which reduces miR390 accumulation and processing accuracy, is predicted to lead to enhanced base-pairing in the region just below the miR390/miR390* region (Cuperus et al., 2010a). Weak or flexible base-pairing within the loop-distal region of the foldback may be necessary for efficient miRNA processing in plants, although this may not be the case in animals (Figure 1.1B).

Genetic analyses suggested that positioning of the initial DCL1 processing event was dependent on a stem length equivalent to approximately 15 bp below the miRNA/miRNA* duplex (Mateos et al., 2010; Song et al., 2010; Werner et al., 2010). Foldbacks from *A. thaliana* (Mateos et al., 2010; Song et al., 2010; Werner et al., 2010), and several additional plant species (Figure 1.1C), tend to have large loops, or terminate, at ~15 bp below the miRNA/miRNA* region, and this characteristic appears important for optimal processing. *Drosophila melanogaster* *MIRNA* foldbacks are strongly characterized by loss of base-pairing at ~11 bp below the miRNA/miRNA* region (Figure 1.1C), and this confers recognition by the RNA-binding protein DGCR8, which then positions the initial cleavage events by Drosha (Sashital and Doudna, 2010). Interaction between HYL1 and DCL1 is required for accurate

processing of pri- and pre-miRNA (Dong et al., 2008). HYL1 may assist in positioning DCL1 for the initial processing step, although direct evidence for how this occurs is lacking. Although loop-distal processing occurs first during miRNA maturation of most foldbacks in plants and animals, alternative primary processing events have been characterized (Okamura et al., 2007; Ruby et al., 2007; Addo-Quaye et al., 2009; Bologna et al., 2009; Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010). The *MIR159/MIR319* family of foldbacks is processed first by DCL1 on the loop-proximal side of the stem, and this is a deeply conserved property of the family (Addo-Quaye et al., 2009; Bologna et al., 2009). In fact, two small RNA are processed from *MIR159/MIR319* family members, a loop-proximal small RNA and mature miR159/miR319, which are separated by 20-30 bp (Bologna et al., 2009) (Figure 1.1D). Both mature miR319 and the loop-proximal small RNA are sensitive to *hyl1* mutations (Bologna et al., 2009), suggesting that they depend on similar processing recognition mechanisms.

Conserved size variation in mature miRNA

Cis-features of the foldback convey information to accurately position the first DCL1 processing events. The positioning of the second processing step to release the miRNA/miRNA* duplex from the foldback precursor is determined by the "molecular ruler" property of DCL1 (Sashital and Doudna, 2010). The PAZ domain binds to the 2-nucleotide 3' overhang that forms on the precursor after the initial processing events. The RNA duplex lays along the Dicer surface. The RNase III domains are positioned at a distance equivalent to ~21 bp away from the position of the bound end. Most plant foldbacks yield a dominant mature miRNA species, although most or all produce some positional and size variants of the miRNA and miRNA* sequences. These likely reflect misprocessing events at the first and/or second DCL1-catalyzed step. The majority of miRNA are 21 nucleotides long, but several families in *A. thaliana* and rice accumulate abundant 22-nucleotide miRNA (Johnson et al., 2009; Chen et al., 2010; Cuperus et al., 2010b). Accumulation of these miRNA is dependent on DCL1, and not DCL2, which generates 22-nucleotide siRNA from long dsRNA (Park et al., 2002; Gascioli

et al., 2005; Xie et al., 2005b; Bouché et al., 2006; Deleris et al., 2006; Lu et al., 2006; Montgomery et al., 2008b; Cuperus et al., 2010b). In the majority of cases, foldbacks that generate 22-nucleotide miRNA contain a single asymmetric, nonpaired nucleotide on the miRNA side of the duplex. Indeed, removal of the asymmetric bulged nucleotide leads to formation of 21-nucleotide miRNA, and introduction of an asymmetric bulged nucleotide leads to formation of 22-nucleotide forms (Chen et al., 2010; Cuperus et al., 2010b). Assuming that the asymmetric bulge does not extend the length of the helix, this explains most of the longer, 22-nucleotide miRNAs. The *MIR163* family is an extreme example of miRNA size variation. In *A. thaliana*, three bulged nucleotides result in a 24-nucleotide mature miRNA, while the *A. lyrata* locus only has two bulges and yields a 23-nucleotide miRNA (Griffiths-Jones et al., 2008; Fahlgren et al., 2010). Functional consequences of miRNA size variation are discussed below.

miRNA-triggered production of 21-nucleotide siRNA

Trans-acting siRNAs (tasiRNAs) are RDR6- and DCL4-dependent products of a refined RNAi pathway, and they function as repressors on specific, co-evolved target mRNA (Allen and Howell, 2010). tasiRNA accumulate among arrays of siRNAs formed from non-coding RNA transcripts from *TAS* loci. The siRNA arrays form in phase, or in register, with the site of precise miRNA-guided cleavage of the primary *TAS* transcript (Allen et al., 2005; Axtell et al., 2006; Rajagopalan et al., 2006; Montgomery et al., 2008a). Phasing of *TAS1/TAS2*, *TAS3* and *TAS4* tasiRNAs is set by cleavage guided by miR173-AGO1, miR390-AGO7 and miR828-AGO1 complexes, respectively (Allen et al., 2005; Axtell et al., 2006; Rajagopalan et al., 2006; Montgomery et al., 2008a; Montgomery et al., 2008b). Thus, miR173, miR390 and miR828 function as activators, rather than repressors, of an siRNA pathway. The involvement of these miRNAs in tasiRNA biogenesis is a specialized function, as most plant miRNAs, at least those associated with AGO1, cannot effectively substitute in these pathways (Montgomery et al., 2008a; Montgomery et al., 2008b). A unique property of miR390 is its association with AGO7,

which has extremely high specificity for miR390 (Montgomery et al., 2008a). miR390 is excluded from AGO1 due to the presence of a 5' adenosine, as AGO1 has a strong preference for miRNAs with a 5' uridine (Mi et al., 2008; Montgomery et al., 2008a). miR390-AGO7 complexes associate with *TAS3* transcripts at two sites - only one of which is sliced (Axtell et al., 2006; Howell et al., 2007). Interaction of miR390-AGO7 at the non-cleaved site is postulated to mark the transcript for recruitment of RDR6 (Montgomery et al., 2008a). Both *TAS3* and miR390 are deeply conserved in plants, although in *P. patens*, both miR390 target sites are cleaved (Arazi et al., 2005; Axtell et al., 2006). In contrast to *TAS3*, *TAS1/TAS2* and *TAS4* transcripts are targeted and sliced at only one site (Allen et al., 2005; Rajagopalan et al., 2006). While substitution of the miR173 target site of *TAS1* with other AGO1-associating miRNA targets sites leads to cleavage, no *TAS1* tasiRNAs are generated (Montgomery et al., 2008b). The only requirement for siRNA generation in the case of *TAS1* is the presence of a miR173 target site. Addition of a miR173 target site to an unrelated mRNA-generating loci results in siRNA formation (Montgomery et al., 2008b). Interestingly, miR173 and miR828 are both 22 nucleotides long, rather than the more common 21-nucleotide length. Twenty-one-nucleotide versions of miR173 and miR828 guide cleavage of target RNA, but do not trigger production of siRNA (Chen et al., 2010; Cuperus et al., 2010b).

In addition to tasiRNA from evolved non-coding loci, some miRNA-targeted mRNA yield secondary siRNA (Axtell et al., 2006; Ronemus et al., 2006; Howell et al., 2007; Chen et al., 2010; Cuperus et al., 2010b; Zheng et al., 2010). Some of these mRNA produce phased RDR6/DCL4-dependent, 21-nucleotide siRNAs like *TAS* transcripts (Axtell et al., 2006; Chen et al., 2007; Howell et al., 2007; Chen et al., 2010; Cuperus et al., 2010a). In some cases, targeting of transcripts by miRNA at multiple sites, or in combination with tasiRNA, may trigger secondary siRNA formation (Axtell et al., 2006). In other cases, generation of secondary siRNA is triggered by a single miRNA-targeting event, and removal or substitution of these target sites may result in a loss of siRNAs (Allen et al., 2005; Axtell et al., 2006; Ronemus et al., 2006; Howell et al., 2007; Vaucheret, 2009). Routing of miRNA-targeted transcripts into

secondary siRNA-generating pathways appears to be regulated, as only ~20% of miRNA or tasiRNA-targeted transcripts yield secondary siRNA (Cuperus et al., 2010b). As with *TAS1/TAS2* and *TAS4* tasiRNAs, siRNA from these protein-coding transcripts is frequently associated with 22-nucleotide triggers in both *Arabidopsis* and rice (Johnson et al., 2009; Chen et al., 2010; Cuperus et al., 2010b). In rice, miR2118 mediates recruitment of 21-nucleotide secondary siRNA-generating machinery, but miR2775-targeted transcripts generate 24-nucleotide siRNA (Johnson et al., 2009). The function of 24-nucleotide secondary siRNA is not clear, but similar 24-nucleotide phased siRNA were also found in *Brachypodium distachyon* (International Brachypodium Initiative, 2010). Most 22-nucleotide miRNA associate with AGO1, suggesting that 21- and 22-nucleotide miRNA-AGO1 complexes are functionally distinct (Chen et al., 2010; Cuperus et al., 2010b). The mechanisms underlying this distinction are not yet known, although it was postulated miRNA size affects an AGO1 functional state that mediates recruitment, either directly or indirectly, of RDR6 (Chen et al., 2010; Cuperus et al., 2010b).

MicroRNA-directed DNA methylation

A subset of miRNA variants preferentially associate with AGO proteins involved in RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2010). In *A. thaliana*, AGO4, AGO6 and AGO9 typically associate with 24-nucleotide siRNA containing a 5' adenosine, and use these siRNA as guides to direct the methylation of target loci DNA (Qi et al., 2006; Mi et al., 2008; Havecker et al., 2010). Twenty-four-nucleotide siRNA are generated by DCL3 from dsRNA produced through the activities of RDR2- and DNA-dependent RNA polymerase IV (Pol IV) (Law and Jacobsen, 2010). AGO4-siRNA complexes may interact with targets through interaction with non-coding transcripts generated by a second plant-specific nuclear RNA polymerase, PolV (Wierzbicki et al., 2009). In rice, DCL3a also processes multiple *MIRNA* foldbacks, yielding 24-nucleotide, miRNA-like siRNA (Wu et al., 2010). While biogenesis of these small RNAs are DCL3-dependent, some loci require initial processing by DCL1a, and

lead to production of both 21- and 24-nucleotide small RNAs. In contrast to 21-nucleotide miRNA variants, the 24-nucleotide miRNA-like siRNA preferentially associate with rice AGO4a and AGO4b, and like 24-nucleotide siRNA, can guide methylation of target DNA (Wu et al., 2010). Similarly, 23-27-nucleotide DCL3-dependent small RNAs that preferentially associate with AGO4 and direct methylation, and that originate from miRNA foldbacks, were identified in *A. thaliana* (Chellappan et al., 2010). Studies in *P. patens* also link miRNA to transcriptional gene silencing through DNA methylation, as mutants with increased levels of miRNA-target RNA duplexes exhibit elevated DNA methylation levels at the target locus (Khraiweh et al., 2010). It is clear, therefore, that miRNA and miRNA-AGO complexes have evolved functions that are distinct from canonical mRNA repression, and that miRNAs can function through cross-talk with both posttranscriptional and transcriptional silencing pathways (Figure 1.2).

Transcriptional regulation of miRNA and tasiRNA precursors

Most *Arabidopsis* *MIRNA* and *TAS* genes are transcribed independently, and expression can be under strict spatial, temporal, developmental, and environmental control. Transcriptional regulation contributes, at least partially, to small RNA expression patterns. Megraw et al. (Megraw et al., 2006) identified promoter motifs associated with hormonal control, flower and shoot meristem development, light response and stress responses (e.g. AUXIN RESPONSE FACTORS, MYC2, LFY, GATA factors) to be overrepresented in a set of 52, mostly conserved, *MIRNA* loci. Most families with these motifs had members with diverse combinations of motifs, suggesting transcriptional regulatory diversity (Megraw et al., 2006). Differentiation of *MIRNA* family member expression is clearly supported by promoter-reporter fusion and primary transcript analyses. Promoters of the four *MIR167* genes (a-d) in *Arabidopsis* drive expression in distinct floral organ domains that correspond to regional specialization of their targets, *ARF6* and *ARF8*. Thus, while all four *MIR167* family members function coordinately to regulate both targets, miR167c-ARF8 and miR167b-ARF8 nodes function uniquely in anther filaments and ovules, respectively (Wu et al., 2006).

Node specialization is also exemplified by expression of *MIR164* paralogs (a-c), their targets *CUC1* and *CUC2*, and the related *CUC3* gene, a non-target. All six genes are involved in shaping leaves and forming lateral buds. *MIR164a*, which contributes to a miR164-CUC2 node, and *MIR164c* are active in boundary regions between lateral organ primordia and meristems, whereas the *MIR164b* promoter directs expression in abaxial epidermal cells (Sieber et al., 2007; Raman et al., 2008). In maize, analysis of transcripts from laser-dissected vegetative meristems and leaf primordia indicate unique, although partially overlapping, expression patterns for seven of the *MIR165/MIR166* genes (Nogueira et al., 2009). Distinct functionality among some of the six members of the *MIR159/MIR319* family is explained partly by miRNA sequence diversity and partly by expression level/location diversity (Allen et al., 2007; Palatnik, 2007). Additionally, *MIR157a*, *MIR160b*, *MIR167d* and *TAS3a* transcripts oscillate diurnally, whereas at least some of the target transcripts encoding SPL (miR157) and ARF (miR160, miR167, *TAS3* tasiRNA) transcription factors do not (Hazen et al., 2009; Sire et al., 2009).

Additional node identity and functionality may also depend on target protein diversification. Initial evidence for biochemical specialization is seen with miR160-ARF nodes, where ARF10, ARF16, and ARF17 each possess a repression domain, but only ARF10 and ARF16 contain a protein dimerization domain. This distinction raises the possibility of miR160-mediated regulatory effects on other ARF proteins, or other partners in ARF10 or ARF16-containing complexes, encoded by transcripts that lack a miR160 target site (Guilfoyle and Hagen, 2007).

AGO-specific small RNA functionality

The idea that AGO proteins have diversified and specialized through evolution has strong support from analysis of the 10 family members from Arabidopsis. Seven AGO proteins (AGO1, AGO2, AGO4, AGO5, AGO6, AGO7/ZIP, AGO10/ZLL/PNH) have been shown to possess distinct biological functions and/or distinct biochemical properties (Vaucheret, 2008). Several AGO proteins are known to act through diverse mechanisms, including translational

repression, RNA cleavage, or RNA-directed DNA methylation. miRNA and tasiRNA function through, or physically associate with, AGO1, AGO2, AGO4, AGO5, AGO7 and AGO10, although the significance of miRNA/tasiRNA association with AGO2, AGO4, and AGO5 is currently not known (Vaucheret, 2008). Biochemical data indicates that AGO1 is loaded with most miRNA and tasiRNA (Baumberger and Baulcombe, 2005; Qi et al., 2005; Mi et al., 2008), while solid genetic but limited biochemical data indicate functionality of AGO10 with a subset miRNA or in a more limited set of biological contexts (Lynn et al., 1999; Liu et al., 2008; Tucker et al., 2008).

A basis for selective sorting of small RNA among AGO proteins, and thus a mechanism for functional specialization of small RNA activity, was recently identified through immunoprecipitation and RNA-seq analysis of AGO-small RNA complexes. AGO1 associates with most miRNA and tasiRNA by selective recognition of a 5' uridine (5'U) (Mi et al., 2008; Montgomery et al., 2008a; Takeda et al., 2008). This explains why the vast majority of miRNAs (82.5% of all, 80% of conserved) and tasiRNAs possess a 5'U. There are a few notable exceptions, such miR172, miR169 and miR395 isoforms, which possess either 5'A or 5'C but still associate with AGO1 (Mi et al., 2008). The composition of small RNAs associated with AGO10 has yet to be published, but it will be interesting to learn if the spectrum of associated miRNAs and tasiRNAs differs between AGO1 and AGO10. AGO2 and AGO5 have clear preferences for small RNA with 5'A and 5'C, respectively (Mi et al., 2008; Montgomery et al., 2008a; Takeda et al., 2008). AGO4, which acts with 24-nucleotide siRNAs to specify de novo RNA-directed DNA methylation, has an apparent preference for 5'A (Qi et al., 2006; Mi et al., 2008). In contrast, AGO7 has high specificity for a specific small RNA, miR390, but through determinants other than the 5' nucleotide (Montgomery et al., 2008a).

Arabidopsis AGO activities are also controlled at the transcriptional level, and at least two AGO proteins (AGO7 and AGO10) may contribute to cell-nonautonomous functions. AGO1 and AGO10 have overlapping activities during shoot apical meristem (SAM) development (Lynn et al., 1999; Tucker et al., 2008). AGO10 is expressed in vascular precursor cells below

the SAM and in adaxial regions of cotyledons and leaves, whereas AGO1 is expressed more broadly (Lynn et al., 1999; Tucker et al., 2008). This implies that an AGO10-dependent signal is transported to the meristem. AGO7 expression is also highly restricted to adaxial cells of young leaves, and may be a limiting factor in the expression of *TAS3* tasiRNA (Montgomery et al., 2008a; Chitwood et al., 2009). AGO7 functions specifically with miR390 in AGO7-miR390 complexes to initiate tasiRNA formation from *TAS3* primary transcripts, which are also expressed in adaxial leaf cells (Garcia et al., 2006; Montgomery et al., 2008a; Chitwood et al., 2009). However, *TAS3* tasiRNA regulate target transcripts, which include *ARF3/ETTIN* mRNA, in a gradient across leaf primordia, suggesting that the tasiRNA or tasiRNA precursor can move cell-to-cell over short distances (Chitwood et al., 2009). This is consistent with the evidence that DCL4 products (21-nucleotide siRNA) or dsRNA precursors from transgenes can move short distances and suppress target transcripts (Dunoyer et al., 2005). Interestingly, modeling of a polarized maize leaf suggests that morphogen-like movement of *TAS3* tasiRNA could have the effect of sharpening boundaries of target gene expression (Levine et al., 2007).

Additionally, the biochemical mode of activity of individual small RNA-AGO1/AGO7/AGO10 complexes may vary significantly, as has been characterized in animals (Ghildiyal and Zamore, 2009). Some may act primarily through target slicing and others primarily through translation inhibition, but it is likely that the majority repress targets through both mechanisms (Llave et al., 2002; Rhoades et al., 2002; Chen, 2004; Gandikota et al., 2007; Brodersen et al., 2008; Dugas and Bartel, 2008). Furthermore, the extent to which both mechanisms act may change through development, as shown for miR156 (Gandikota et al., 2007). Thus, small RNA-target node identity or functionality may be modulated at the transcriptional, AGO sorting, AGO activity, and intercellular transport levels.

Plants use multiple endogenous small RNA pathways to regulate gene expression at numerous levels. These pathways modulate target expression transcriptionally and post-transcriptionally, and may include amplification steps that produce secondary siRNA that reinforce or mobilize target silencing. Although many of the mechanisms that contribute to the

diversification and specialization of small RNAs in plants are understood, more work is needed to identify unrecognized mechanisms and to clarify known mechanisms. This thesis focuses on contributing answers to the following questions: What mechanisms and factors are involved in secondary siRNA production? How do endogenous small RNA pathways interact with each other? What cis features govern processing of microRNA from *MIRNA* foldbacks? And can differences in processing efficiency act to fine-tune expression?

Figure 1.1. MicroRNA processing in plants.

(A) Box plots of the predicted length of *MIRNA* foldbacks in plants and animals.

(B) Entropy of *MIRNA* foldbacks from 5 distinct regions. A foldback diagram (right) marks the location of the five regions. I: 5' arm, loop-distal. II: miRNA. III: loop, 5' and 3' arms, loop-proximal. IV: miRNA*. V: 3' arm, loop-distal. Entropy was calculated per base using the program RNAfold (Hofacker, 2003). The length of regions I and V were either 17 and 19 nucleotides, respectively, for plants or 11 and 13 nucleotides, respectively, for *C. elegans* based on predicted differences in plant and animal miRNA processing from the foldback base.

(C) Distribution of large loops (≥ 3 nucleotides) or ends of helices for foldbacks from 24 deeply conserved *MIRNA* families in five plant species. Both individual species and combined species data are shown (mean relative level +/- Standard Error of Means). The 17/15-nucleotide region shows an enrichment of loops or helix ends in all plants shown, while *D. melanogaster* large loops or helix ends are enriched at 13/11 nucleotides. The combined plant bar graph is overlaid with the cumulative number of predicted intact foldbacks remaining beyond each position. The combined plant foldback end distribution is overlaid on the *D. melanogaster* plot for comparison.

(D) Illustration of cis elements involved in animal and plant miRNA foldback processing. Bold numbers represent initial and secondary processing sites leading to mature small RNA.

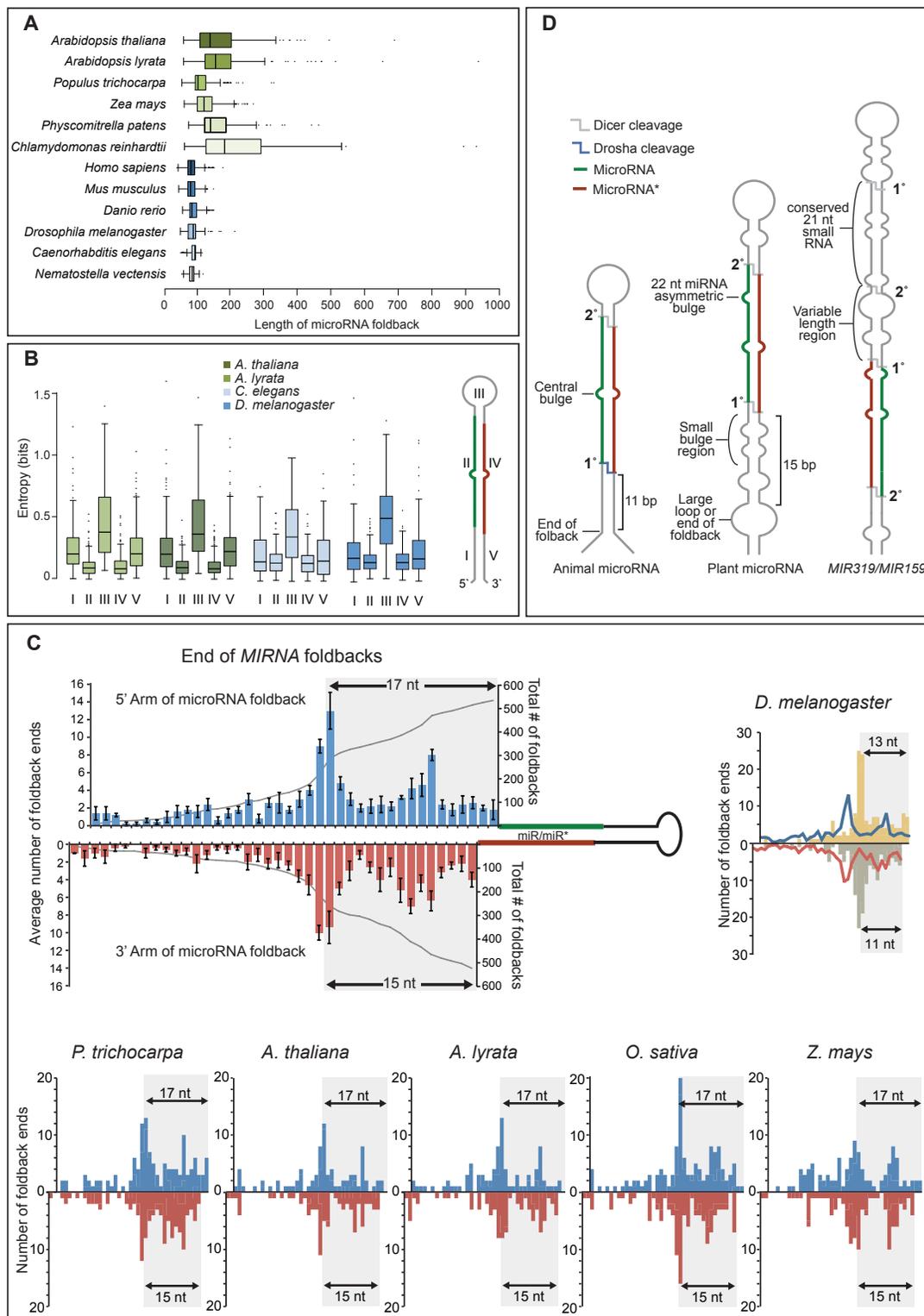


Figure 1.1. MicroRNA processing in plants.

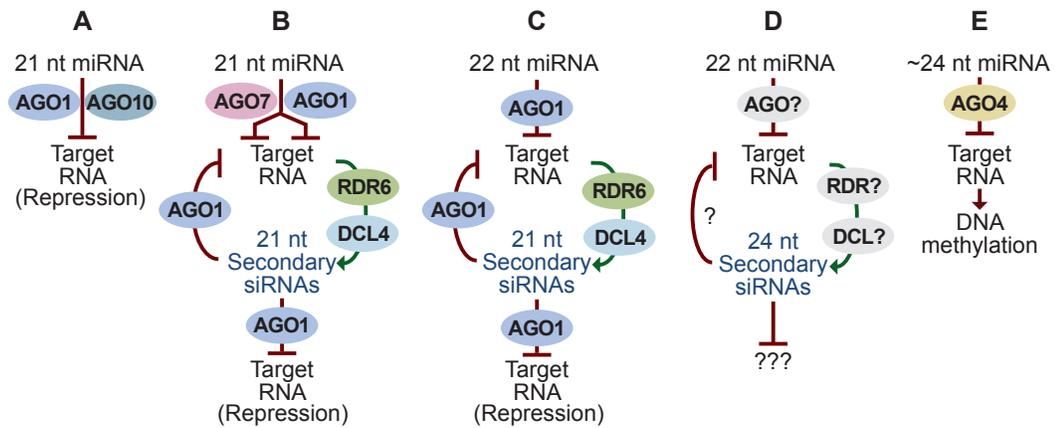


Figure 1.2. Functional cross-talk of the microRNA pathway in plants.

Plant miRNA functions, including repression of targets, triggering of siRNAs from targets, and triggering of RNA-directed DNA methylation. **(A)**, **(B)** and **(E)** have been described in monocots, eudicots and bryophytes, **(C)** in monocots and eudicots and **(D)** only in monocots.

TAS1 and TAS3-based forward genetic screen and identification of mutants in Arabidopsis by direct genome sequencing

Josh T. Cuperus, Taiowa A. Montgomery, Noah Fahlgren, Russel T. Burke, Christopher M. Sullivan, Maxim V.C. Greenberg, Steven E. Jacobsen and James C. Carrington

PNAS
700 11th Street, NW, Suite 450
Washington, DC 20001
Volume 107 (1), pages 466-471

Epigenetics
1806 Rio Grande Street
Austin, TX 78702
Volume 6, Issue 3, pages 334-344

SUMMARY

Trans-acting siRNA (tasiRNA) formation is initiated by miRNA-guided cleavage of primary *TAS* transcripts. In Arabidopsis, the miR390-ARGONAUTE7 complex is specialized for initiation cleavage and non-cleavage interaction with *TAS3* primary transcripts, resulting in recruitment of RNA-DEPENDENT RNA POLYMERASE6 for double-stranded RNA (dsRNA) biosynthesis, while miR173-AGO1 or miR828-AGO1 complexes interact with and cleave *TAS1/TAS2* and *TAS4* primary transcripts respectively and also recruit the dsRNA biosynthesis machinery. Screens for *TAS1* and *TAS3* tasiRNA-deficient mutants were done, yielding two independent lines with the *mir390a-1* allele that accumulated low levels of mature miR390. Using one of these miR390-deficient mutants, a direct genome sequencing-based approach to both map (using a bulk segregant population) and rapidly identify the mutant allele was developed. This revealed a G-to-A point mutation that was calculated to stabilize a relatively non-paired region near the base of the *MIR390a* foldback. This high-throughput technique was also used to identify a mutant deficient in RNA-directed DNA methylation, *nrrpd/e2-19*.

INTRODUCTION

Small RNAs, including microRNA (miRNA), several classes of endogenous small interfering RNA (siRNA), and Piwi-associated RNA (piRNA), direct silencing activities that shape transcriptomes and proteomes of eukaryotic organisms. miRNAs arise from transcripts containing self-complementary foldback structures that are initially processed to form 21-22nt miRNA/miRNA* duplexes. After excision from a foldback, most miRNA guide strands associates with ARGONAUTE 1 (AGO1), and target RNA in a sequence specific manner. In the majority of cases this interaction leads to cleavage and repression of target RNA (Allen and Howell, 2010).

The trans-acting siRNA (tasiRNA) class represents a specialized type of amplification-dependent siRNA (Peragine et al., 2004; Vazquez et al., 2004b). Primary tasiRNA-generating

transcripts are first processed by miRNA-guided cleavage (Allen et al., 2005; Yoshikawa et al., 2005). Either the 3' (*TAS1*, *TAS2* and *TAS4* families) or 5' (*TAS3* family) cleavage product is stabilized and converted to dsRNA by RNA-DEPENDENT RNA POLYMERASE6 (RDR6) (Peragine et al., 2004; Vazquez et al., 2004b; Allen et al., 2005; Rajagopalan et al., 2006). Phased, 21-nt siRNAs are generated in register with the miRNA-guided cleavage site through sequential processing by DCL4. Routing of *TAS3* precursor RNA requires two miRNA-guided events, both of which involve AGO7-miR390 complexes (Axtell et al., 2006; Montgomery et al., 2008a). Interaction of AGO7-miR390 at a 3' proximal target site results in primary transcript cleavage, and sets the register for phased siRNA generation. The 3' cleavage function of AGO7-miR390 is generic, as any of several heterologous miRNA working through AGO1 can substitute for AGO7-miR390 (Montgomery et al., 2008a). A second miR390 target site at a 5'-proximal position in the processed precursor interacts with AGO7-miR390 in a non-cleavage mode (Axtell et al., 2006; Howell et al., 2007).

RNA-directed DNA methylation (RdDM) in plants is a mechanism involving 24-nucleotide siRNAs that targets in cis, resulting in DNA methylation at target DNA loci. RdDM has been well studied and involves the concordant action of DNA-dependent RNA polymerases (Pol IV and Pol V), RNA-DEPENDENT RNA POLYMERASE2 (RDR2), three ARGONAUTES (AGO4, AGO6, AGO9), DICER-LIKE3 (DCL3), multiple accessory proteins (such as DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) and INVOLVED IN DE NOVO 2 (IDN2)), and the DNA methylation machinery including DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), SUPPRESSOR OF VARIATION 3-9 HOMOLOGUE 9 (SUVH9) and SUVH2. Because of the complexity of RdDM, the potential to identify new factors by forward genetic approaches still exists. Previously, mutagenesis approaches have successfully recovered many Arabidopsis mutant alleles deficient in RdDM (Zilberman et al., 2003; Kanno et al., 2005; Smith et al., 2007; Kanno et al., 2008).

Forward genetics has fundamentally influenced our knowledge of biological processes in many organisms, and has led to many discoveries, including the functional significance of

miRNAs (Lee et al., 1993; Wightman et al., 1993). Fundamentally, chemical- or radiation-induced mutations are easiest to administer and are mostly random throughout a genome. In plants, ethyl methane sulfonate (EMS) is useful for causing single-nucleotide mutations. This method allows for saturation of thousands of mutations per genome without extreme damage to the genome, and can be administered on Arabidopsis seeds, making it ideal for forward genetic screens. The high number of mutations per plant allows for screening of relatively few plants to find those with phenotypes of interest. The EMS-induced point mutations may lead to a variety of effects on gene function, including complete or partial loss of function, constitutive function or double mutants. This allows for identification of genes unlikely to be found in other forward genetics that cause severe alterations to the genome (Page and Grossniklaus, 2002).

One major challenge of point mutation forward genetics (also known as markerless forward genetics) is to identify a causative mutation in a background of thousands of other mutations. This is accomplished by determining the approximate position of the mutation of interest (mapping interval), then by sequencing the mapped region and identifying candidate causal mutations. The process takes upwards of several months depending on the method. Until recently, a map position was identified using PCR and the gain or loss of a restriction site based on the polymorphic site (Konieczny and Ausubel, 1993), or by using array hybridization methods (Hazen et al., 2005). After identification of a mapping interval, extensive Sanger sequencing was then necessary to identify mutations within the mapping interval. Recently, new technologies including microarray technology and high-throughput sequencing have increased the depth of known single-nucleotide polymorphisms (SNPs) between well-studied Arabidopsis ecotypes (Clark et al., 2007; Lister et al., 2009). Additionally, high-throughput sequencing has allowed for high-resolution re-sequencing of genomes and of mapping strains (Lister et al., 2009).

The advent of high-throughput sequencing (HTS) methods has enabled direct approaches to analyze plant genomes. Here we identified multiple *TAS1* and *TAS3* deficient mutants, including novel *TAS3* deficient mutants. We also applied high-throughput sequencing

to enable rapid mapping of *A. thaliana* mutations, normally taking months of work. Similar to traditional mapping, a bulk segregant population of two highly similar ecotypes was used, however using high-throughput sequencing, identification of a mapping interval and identification of candidate mutations within the mapping interval, were combined into a single computational pipeline. In this work we used high-throughput sequencing to identify two causative mutations in two independent forward genetic screens.

RESULTS

Mutagenesis of transgenic Arabidopsis expressing *TAS1c* and *TAS3a*-based syn-tasiRNA

TAS3a-based and *TAS1c*-based synthetic (syn)-tasiRNAs with complementarity to the *PDS* mRNA provide a visual readout for tasiRNA activity in transgenic Arabidopsis plants (Montgomery et al., 2008a; Montgomery et al., 2008b). The 35S promoter-driven syn-tasiRNA construct *35S:TAS3aPDS-1* yields tandem syn-tasiRNAs from the 5' D7[+] and 5' D8[+] positions in place of siRNA2141 and siRNA2142 (also known as tasi-ARFs) from the *TAS3a* transcript (Allen et al., 2005; Montgomery et al., 2008a) (Figure 2.1A). In wild-type (Col-0) plants expressing *35S:TAS3aPDS-1*, photobleaching emanates from the midrib and major veins, with the phenotype most prominent when viewed from the adaxial side of leaves (Figure 2.1B, (Montgomery et al., 2008a)). The 35S promoter-driven syn-tasiRNA construct *35S:TAS1cPDS-2* yields syn-tasiRNAs from the 3' D3[+] and 3' D4[+] positions in place of siR255 and siR850 from the *TAS1c* transcript (Allen et al., 2005; Montgomery et al., 2008b)(Figure 2.2A). In Col-0 plants expressing *35S:TAS1cPDS-2* photobleaching is widespread and leads to reduced biomass and fertility. Syn-tasiRNA accumulation and photobleaching are suppressed in plants containing loss-of-function *rdr6-15*, *dcl4-2* and for *35S:TAS3aPDS-1*-transformed, *zip-1* (AGO7-defective) mutations (Montgomery et al., 2008a) (Montgomery et al., 2008b).

The visual photobleaching readout was used in a genetic screen for mutants with loss of *TAS1* tasiRNAs. Due to severity of the *35S:TAS1cPDS-2* transgenic Arabidopsis phenotype (Figure 2.2E and (Montgomery et al., 2008b)), propagating mutants was difficult, however, initial screening of mutants lacking photobleaching was aided by the inability of non-mutants expressing *35S:TAS1cPDS* that had active syn-tasiRNA to survive on soil. However, not all plants that lacked photobleaching were mutants, as the extreme and variable phenotype yielded a large number of plants thought to have silenced the *TAS1cPDS-2* transgene or a reduction in photobleaching due to the variability of the phenotype (Matzke et al., 2000). Although specific loss of *TAS1*, *TAS2* and *TAS4* tasiRNAs was not predicted to have an obvious phenotype, mutants could be grouped based on other criteria. A large variety of mutants were recovered (Figure 2.2B), including many categorized based on their small RNA profiles including: specific loss or reduction of syn-tasiRNA, loss or reduction of endogenous tasiRNA, no loss of small RNA, and loss or reduction of all small RNA (Figure 2.2 C, D and E). The most interesting category of expected mutants were those having reduced *TAS1*, *TAS2* and *TAS4* tasiRNAs while maintaining *TAS3* tasiRNAs. Two mutants had reduced levels of *TAS1* tasiRNA with relatively normal levels of *TAS3* tasiRNA and a reduction in miR173, however other miRNA were not tested (Figure 2.3). These mutants had developmental defects, including slow growth and sterility, and therefore, further analysis of these mutants could not be completed (Figure 2.3).

The visual photobleaching readout was also used in a genetic screen for mutants with defects specific to the *TAS3* tasiRNA pathway. Besides loss of photobleaching, mutants with *TAS3*-specific defects were predicted to have 1) low or no syn-tasiRNA and endogenous tasi-ARF (siRNA2142), 2) normal levels of *TAS1* tasiRNA (siR255), 3) normal levels of miRNA, such as miR171, that do not function in the *TAS3* pathway, and 4) an accelerated vegetative phase change (AVPC) phenotype, which is tightly associated with loss of *TAS3* tasiRNA (Poethig, 2003; Adenot et al., 2006; Fahlgren et al., 2006; Garcia et al., 2006; Montgomery et al., 2008a). *TAS3* pathway-specific mutants were not expected to have severe developmental

defects, as would be expected for general loss-of-miRNA function mutants (e.g. (Clarke et al., 1999; Jacobsen et al., 1999)). The AVCP phenotype is characterized by downward-curved rosette leaves, giving the appearance of a narrow leaf phenotype, and early development of abaxial trichomes (Poethig, 2003). Transgenic seeds expressing *35S:TAS3aPDS-1* were mutagenized, and seedlings from the M2 generation (self-pollinated) of 250 pools were screened for loss of photobleaching. Three hundred fifty-five individuals with a reduced-photobleaching phenotype were recovered (Figure 2.1C). As exemplified by mutant 104a5, 216 mutants had an AVCP phenotype (Figure 2.1B, C). Ninety-five mutants had relatively strong or severe vegetative defects, and 44 had no visible abnormalities. Most of the severe phenotype plants (Class I) were dwarfed, had serrated leaves, and resembled known miRNA-debilitated mutants, such as those with *hen1* or strong hypomorphic *ago1* alleles (Figure 2.4A). Among 7 Class I mutants analyzed, each had reduced levels of miR171 and siR255, indicating that they were generally deficient in miRNA accumulation or activity (Figure 2.4A). Class I mutants were not analyzed further.

The levels of siR255, a *TAS1* tasiRNA, were measured in each of the 216 plants with AVCP phenotypes. Approximately 64% (138) or 25% (55) of these mutants lacked *TAS1* siR255 altogether, or produced siR255-related small RNA that migrated during electrophoresis as a 22-nucleotide RNA, respectively, indicating that they possessed general (*TAS3* and *TAS1*) tasiRNA defects (Figure 2.1C, 2.4B). These 193 plants comprised Class II mutants. Complementation analyses of a random sampling of 15 mutants, including 90b5 (Figure 2.4B), revealed that the loss-of-siR255 subgroup within Class II mutant was dominated by *rdm6* (10) and *sgs3* (5) mutants. In contrast, 14/14 mutants tested from the size-shifted subclass possessed *dcl4* mutations. Loss of DCL4 is known to result in 22-nucleotide size-shifted tasiRNA, due to the activity DCL2 (Dunoyer et al., 2005; Gascioli et al., 2005; Xie et al., 2005a).

Only 12% (26) of plants possessed an AVCP phenotype, normal levels of 21-nucleotide siR255, and normal levels of miR171 (Figure 2.1C). Nearly all of these, which were

designated as Class III mutants, possessed low or undetectable levels of *TAS3* siR2142. Complementation analysis revealed 22 independent *ago7* mutants, 14 of which were subjected to *ago7* allele sequencing. Most of the *ago7* alleles contained substitutions affecting the PIWI domain, while single mutants with mid domain or N-terminal domain substitutions were identified (Table 2.1 and Figure 2.5). A *TAS3*-specific *dcl4* mutant (70b1), in which siR2142-related small RNA, but not siR255, was shifted to 22-nucleotide, was recovered (Figure 2.4C, E). The 70b1 *dcl4* allele contained a non-conserved Gly-to-Arg substitution affecting a region between the PAZ domain and first RNaseIII domain (Table 2.1). Two recessive mutants, 52b2 and 87a3, could not be assigned to any of the complementation groups tested through crosses to *zip1*, *rdr6-15*, *sgs3-11*, and *dcl4-2*. These mutants had identical, moderate AVPC phenotypes (Figure 2.4C). In quantitative blot assays, the 52b2 mutant accumulated significantly reduced levels siR2142 (44.9% compared to Col-0; $p < 0.0028$), but normal levels of miR171 and *TAS1* siR255 (Figure 2.4C,D). Interestingly, both 52b2 and 87a3 had low levels of miR390 (Figure 2.4C), with quantitative blot assays revealing a significant difference ($p < 0.0001$) between 52b2 and Col-0 plants (Figure 2.4D).

Map-assisted sequencing and identification of *mir390a-1*

Because the Class III 52b2 and 87a3 mutants remained unassigned after the initial complementation analysis, they were investigated further using a genome-wide mutation-identification strategy. In principle, direct genome sequencing of a mutant genome using high-throughput sequencing (HTS) technology could identify sites of mutation. However, each EMS-mutagenized genome can possess hundreds or thousands of changes in addition to the mutation causing the phenotype of interest. We developed a strategy to apply direct sequencing to a bulk segregant population of genomes, and applied it to identification of the causative 52b2 mutation. A segregating F2 population from a cross between 52b2 (Col-0 background) and the polymorphic accession Ler was prepared under syn-tasiRNA transgene selection (hygromycin resistance), and 93 homozygous plants with both AVCP and low

photobleaching phenotypes were identified. Samples of DNA from the 93 plants were pooled and subjected to paired-end sequencing using an Illumina 1G system, which provided 221 million 36-base reads.

A multistep pipeline, Mapping and Assembly with Short Sequences (MASS, Figure 2.6A) was devised to map and sequence Col-0-enriched sequences, which were expected to cosegregate with the 52b2 mutation. Approximately 143,000 single nucleotide polymorphisms (SNPs) were identified by Clark et al. (Clark et al., 2007), and these were used to identify and quantify Col-0- and Ler-specific reads from repeat-filtered sequences. The ratios of summed Col-0 SNPs/summed Ler SNPs were calculated in 100,000 base windows (20,000 base scroll) across the Arabidopsis genome. A major peak of enriched Col-0 SNPs was identified on the long arm of chromosome II (Figure 2.6B). In addition, several minor peaks of Col-0-enriched SNPs were identified around pericentromeric regions. The basis for these minor peaks was not determined conclusively, although it is speculated they reflect miscalled SNPs that do not exist in Ler.

A 1.52 Mb region encompassing the major Col-0-enriched peak was assembled with the program Mapping and Assembly with Quality (MAQ, (Brodersen et al., 2008)) using all high-quality sequencing reads. The CNS2SNP function was used to identify polymorphisms between the assembled sequences and the Col-0 reference sequence (Brodersen et al., 2008). Polymorphisms were considered for further analysis if called as homozygous by CNS2SNP, had a PHRED-like score greater than or equal to 43, were supported by five or more reads but less than or equal to 50 reads (to avoid using reads derived from repeats), and had read coverage from both strands to avoid strand-biased sequencing errors. The MASS pipeline is available for download at <http://jcclab.science.oregonstate.edu/MASS>.

Five G-to-A or C-to-T changes, which were consistent with EMS-induced mutation, were identified in the interval (Table 2.1). An A-to-G difference at genome coordinate ChrII:16766679 was detected, but this was due to a bona fide difference between the reference and initially mutagenized genome. Four of the G-to-A mutations were sequenced

using the Sanger method and confirmed as post-EMS specific in the 52b2 mutant. Mutations were identified in AT2G40950 (bZIP17) and AT2G41470 (embryo-specific protein related) loci, but neither could readily explain the *TAS3* pathway-specific phenotypes. Two mutations were located in an intergenic region and a helitron-type transposable element. However, one mutation affected the sequence at the base of the foldback from MIR390a (Figure 2.6C,D). The phenotypically identical mutant 87a3 was sequenced manually across the loci corresponding to four of the 52b2 EMS-associated mutation sites, revealing exactly the same mir390 mutation but reference-identical sequences at each of the other positions. Thus, 52b2 and 87a3 were independent mutants containing the same mir390a allele (*mir390a-1*, Figure 2.7). The *mir390a-1* mutant Arabidopsis had distinct loss of photobleaching, accelerated vegetative phase change and leaf elongation (Figure 2.7). A genomic fragment containing wild-type MIR390a was introduced into 52b2 mutant plants. This transgene restored photobleaching to the 52b2 mutant line, and partially suppressed the leaf curling phenotype (Figure 2.6E), confirming that the *mir390a-1* mutation caused the defects in 52b2.

Map assisted sequencing and identification of *nripd/e2-19*

In plants, RDR2-, Pol IV-, and DCL3-dependent siRNAs are involved in de novo methylation of DNA, resulting in transcriptional repression of target genes (Chapman and Carrington, 2007). The locus *FLOWERING WAGENINGEN (FWA)* has been used extensively to study RdDM. In wild type plants, this locus is repressed by RdDM, leading to normal flowering time, while de-repression of this locus results in late flowering time (Cao and Jacobsen, 2002; Chan et al., 2004). Several forward genetics screens have been used to identify factors involved in methylation, however separating de novo and maintenance methylation has proven difficult. Using an EMS mutagenesis method, followed by transformation of mutagenized plants with a *FWA* transgene, mutants with reduced de novo methylation were screened based on the reduced flowering time phenotype as a result of an unmethylated *FWA* transgene. Multiple mutants had reduced methylation and reduced flowering time. One mutant, *m48*, was

prepared for bulk segregant and MASS analysis by collecting DNA from an F2 population of 38 Col-0, Landsberg hybrids with reduced flowering time. The DNA from the 38 recombinants was pooled and we performed paired-end sequencing using Illumina GA II technology, resulting in 142 million reads and 21.4x coverage of the Arabidopsis Col-0 genome.

Using MASS, a 3 MB region enriched in Col-0 SNPs was identified, with the highest peak between 8.25 and 8.8 MB on chromosome 3 (Figure 2.8A). Using the MAQ alignment tool, we mapped reads within this region, and recovered 11 mutants consistent with EMS mutagenesis (Table 2.3). Among these mutations, was one in *NRPD/E2*, a gene previously identified to be involved in RdDM (Herr et al., 2005; Kanno et al., 2005). A cross of *m48* to the null mutant *nripd/e2-2* confirmed this mutation as the causal mutation (Figure 2.8B) (Onodera et al., 2005). The *m48* mutation in *NRPD/E2* leads to a missense mutation changing the positively charged amino acid Arginine to the neutral amino acid Glutamine (Figure 2.8C).

DISCUSSION

Multiple Arabidopsis mutants with tasiRNA biosynthetic or activity defects were recovered from *TAS1*-based syn-tasiRNA genetic screen. Unfortunately silencing or variability of *35S:TAS1cPDS-2* and the lack of obvious developmental phenotypes, mutants were difficult to identify. Two mutants had reduced *TAS1c* tasiRNAs as well as reduced miR173 levels. Because the Arabidopsis mutant *mir173-1* and the miR173 mimic containing Col-0 are indistinguishable from wild-type plants, but has reduced miR173 levels (Todesco et al.; Gascioli et al., 2005), these mutants with obvious growth phenotypes seem unlikely to be mutations in *MIR173* and more likely a weak allele in a microRNA biogenesis factor that effects *TAS1* more than *TAS3*, or a novel factor.

Dozens of Arabidopsis mutants with tasiRNA biosynthetic or activity defects were recovered from a genetic screen for loss of *TAS3*-based syn-tasiRNA function. While the vast majority of mutants possessed general (*TAS1* and *TAS3*) tasiRNA or miRNA defects, three mutant classes contained *TAS3*-specific phenotypes. The *ago7* mutants extended the known

allele series (Hunter et al., 2003; Vazquez et al., 2004b), although none of the mutants characterized here possess unanticipated phenotypes. The *dcl4-7* allele resulted in a *TAS3* tasiRNA-specific defect leading to 22-nucleotide rather than 21-nucleotide RNA, which was unexpected. It is possible that the variant DCL4 protein does not recognize the *TAS3* syn-tasiRNA dsRNA precursor, and the 22-nucleotide siRNA are formed through DCL2, which is known to yield 22-nucleotide siRNA from dsRNA (Dunoyer et al., 2005; Gascioli et al., 2005; Xie et al., 2005a; Deleris et al., 2006; Mlotshwa et al., 2008). This mutant could be particularly interesting in studies to understand the basis for recruitment of DCL proteins to distinct precursors.

Finally, the direct sequencing-based approach that identified the *mir390a-1* mutation and *nrpd/e2-19*, as well as other sequencing-based methods (Sarin et al., 2008), should be broadly applicable to identification of other marker-less (e.g. EMS-induced) mutations. The major benefit of this approach, obviously, is the simultaneous mapping and sequencing at a genome-wide level. The ability to score all known polymorphisms in individuals from the mapping population affords tremendous marker density, and the MASS pipeline provides a straightforward route to identification of a small number of candidate genes within a relatively small interval of 1-3 Mb. Of course, the method is dependent on genetic resolution and sequencing depth, which in this study was approximately 12X and 21X genome-wide coverage for *mir390a-1* and *nrpd/e2-19* respectively. Although the *mir390a-1* and *nrpd/e2-19* mutations were sequenced directly from a mapping population, sequence-based mutant identification does not necessarily require a mapping population, although identification of causal mutations is confounded by inconsequential EMS changes (several hundred to thousands throughout the genome) following mutagenesis. However, sequencing of two independent mutants that are known to possess defects in the same gene, followed by identification of convergent mutant loci, could provide an alternative to the sequencing of bulked, segregating populations.

MATERIALS AND METHODS

Plant Material

References for *rdr6-15*, *dcl4-2*, *sgs3-11*, *hen1-1*, *hyl1-2*, *se-2*, *hst-15*, *dcl1-7*, *nrdp/e2-2* and *zip-1* alleles were previously listed (Allen et al., 2005), *Arabidopsis* ecotype Col-0 plants were transformed by the floral dip method using *Agrobacterium tumefaciens* GV3101 (Clough and Bent, 1998). Transgenic plants were grown on MS medium with kanamycin (50 ug/ml) or hygromycin (50 ug/ml) or both for 10 days, transferred to soil and grown in normal greenhouse conditions with supplemental light with a 16hr light/ 8 hr dark cycle.

Mutagenesis of *35S:TAS3aPDS-1* and *35S:TAS1cPDS-2*-transformed plants

Mutagenesis was done using 125,000 homozygous *35S:TAS3aPDS-1*-transformed F2 individuals and 50,000 homozygous *35S:TAS1cPDS-2*-transformed F2 individuals by soaking seeds in 0.2% EMS for 16 hours. M1 plants were pooled into groups of 500 plants, and seed was collected. 2500 M2 seed from each pool were planted directly on soil and initially screened based on loss of photobleaching. Plants with low or no photobleaching were separated based on their developmental phenotypes. Leaves 7 and 8 were collected for RNA isolated with TRIzol. In the case of *35S:TAS3aPDS-1*-transformed mutants, five-ug total RNA was used for Low Molecular Weight Denaturing 7M Urea gels, and probed for tasiRNA255, separating class II and class III mutants based on tasiRNA255 levels. Outcrossing was performed to identify plants harboring mutations in known factors involved in *TAS3* tasiRNA biogenesis. Each mutant was crossed with *dcl4-2*, *rdr6-15*, *sgs3-11*, and *zip-1*. Non-complemented crosses were categorized, and some were further characterized, by replanting M3 selfed plants and re-confirming phenotype and small RNA profile.

Transgene constructs

Transgene sequences were PCR-amplified from genomic DNA. Syn-tasiRNA, and miRNA constructs were introduced by site-overlap extension as previously described (Ho et al., 1989) and introduced in pMDC32, pGWB2 or pGWB1 (Curtis and Grossniklaus, 2003; Nakagawa et al., 2007). *MIR390a* constructs were designed similarly to *MIR319a* based artificial microRNAs (Schwab et al., 2006). These were introduced into pENTR/D-TOPO (Invitrogen) and subsequently recombined with LR clonase (Invitrogen) into pMDC32 (Supplemental Experimental Procedures (Curtis and Grossniklaus, 2003)). The *FWA* transgene was transformed into Arabidopsis using Agrobacterium harboring a pCAMBIA3300 vector with an engineered version of *FWA*.

Conventional sequencing

Sequencing using the Sanger method was done using the NCBI primer design tool available through <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>. Primers were designed to encompass the mutation within a 500 bp window. Two independent PCR products were gel purified and directly sequenced along with a non-mutagenized parental control.

RNA blot assays

RNA was isolated using TRIzol reagent (Invitrogen). Two chloroform extractions were done and RNA was precipitated in an equal volume of isopropanol for 20 min. RNA blot assays were done as described (Montgomery et al., 2008b). Briefly 5 ug, 10 ug or 20 ug of total normalized RNA was resolved by denaturing polyacrylamide-gel electrophoresis. RNA was transferred to positively charged nitrocellulose membrane. DNA or LNA probes were end-labeled using [³²]ATP and Optikinase (USB). Probes were hybridized to RNA on membranes at 38-50° C. Quantification of small RNA blot hybridization intensities was done using an Instant Imager (Packard Bioscience) and normalized relative to the parental *TAS3aPDS-1*-transformed control or appropriate transient assay control.

Genomic sequencing

A bulk segregant population (F2) from a cross between the 52b2 and *m48* mutants and the polymorphic parent Ler were generated. Homozygous mutant plants (92 individuals for 52b2 or 38 for *m48*), were inferred based on phenotype, and DNA was isolated. Sequence analysis was done with an Illumina Genome Analyzer I for 52b2 and an Illumina Genome Analyzer II for *m48*. Seven lanes of a paired-end flow cell were used in both cases. The Illumina (www.illumina.com) genomic DNA sample prep protocol was followed with modifications. After the PCR amplification step, additional gel purification was done to remove adapter-adapter product. Sequencing and base calling were done according to the manufacturers recommendations.

MASS (Mapping and Assembly with Short Sequences)

Reads from the bulk segregant population were mapped to the Arabidopsis (Col-0, TAIR8) genome using Cache Assisted Hash Search using XOR logic (CASHX; (Fahlgren et al., 2009)), resulting in ~12x average coverage for perfect-match reads (1.6 GB). Using 143,508 available SNPs (Clark et al., 2007), a database of 71 bp sequences centered on each SNP (Col-0 vs. Ler) was created. When 71mers overlapped, they were joined into one larger database entry. Illumina 1G reads were aligned to entries in the database using CASHX. Reads that hit Col-0 or Ler SNPs were summed in 100,000 bp windows, using a 20,000 bp scroll, and ratios were calculated. These ratios were plotted using R and visualized (Figure 2.4B). Illumina reads that aligned with up to two mismatches to ChrII:15800000-17320000 were parsed using Short Oligonucleotide Analysis Package (SOAP, (Li et al., 2008b)). Using the MAQ program easyrun (Li et al., 2008a), 967,616 sequences (with their Illumina-based quality scores) that mapped with 2 mismatches or less to the 1.5 MB interval were assembled. The MASS package contains scripts to run CASHX, SOAP and MAQ, and is available for download (<http://jcclab.science.oregonstate.edu/MASS>). In addition to the MASS mapping and

alignment tools, the MASS package contains the entire pipeline used to identify the *mir390a-1* mutation. It includes programs for creating plots of SNP enrichment, alignment with MAQ and filtering of SNPs. MASS is designed to take any indicated read length and create an appropriate database of sequences centered on a SNP nucleotide, forcing each read to align across the SNP site. The MASS pipeline filters the SNP data set (cns.snp) from the MAQ output. Using Illumina quality scores, data are filtered based on the following criteria: consensus base is a true base; a phred-like quality score of 43; a minimum read depth of 5; a maximum read depth of 50; and no second best base call. The phred-like quality score is based on Illumina quality scores. In part, these filtering values are based on ~12x coverage; quality scores and read depth may be adjusted based on coverage, read length and quality of reads.

ACKNOWLEDGMENTS

We thank Sarah Dvorak for assistance with the genetics screen and technical assistance. We also thank Detlef Weigel, Korbinian Schneeberger, and Richard Clark for helpful discussion, genomic sequencing test data sets, and SNP data. This work was supported by grants from the National Science Foundation (MCB-0618433), National Institutes of Health (AI43288) and United States Department of Agriculture–National Research Initiative (2006-35301-17420).

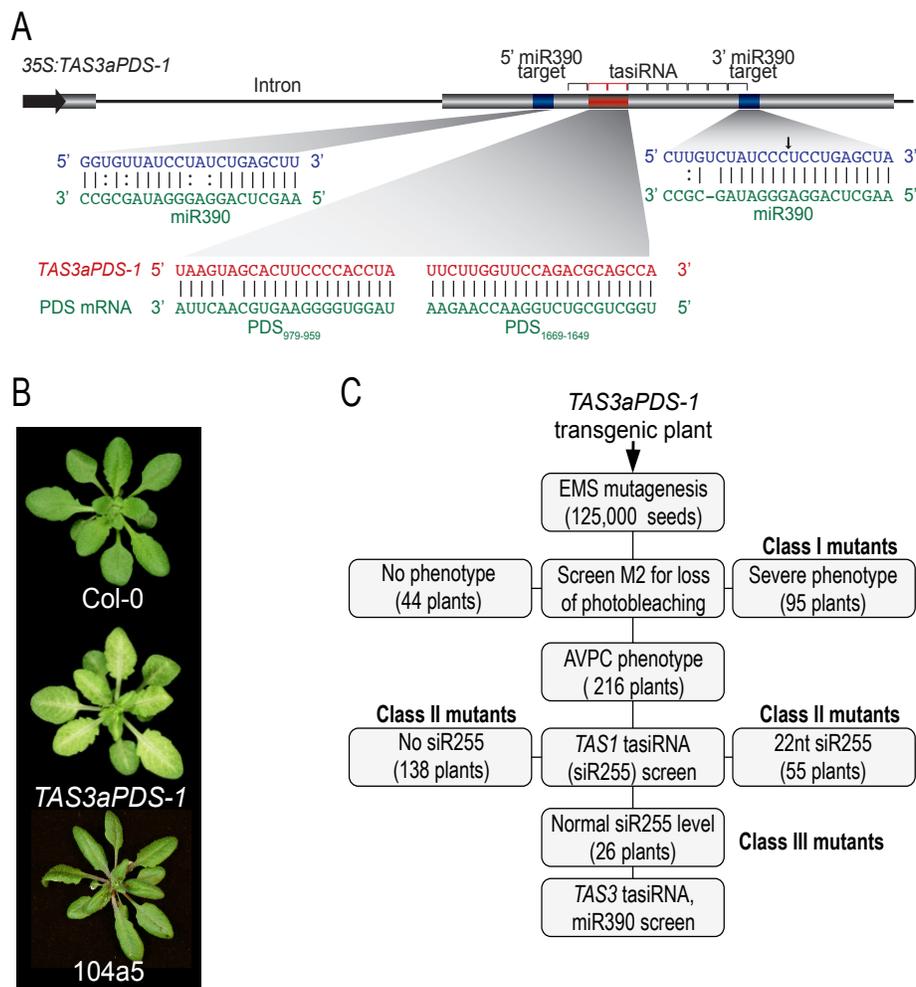


Figure 2.1 Syn-tasiRNA strategy and mutant screen

(A) Organization of syn-tasiRNA construct *TAS3aPDS-1*. The miR390/miR390 target and syn-tasiRNA/PDS target sequences are shown in the expanded diagrams. (B) Photobleached phenotype of *35S:TAS3aPDS-1*-transformed Col-0 and a Class II mutant (104a5) recovered from the screen are shown next to non-transformed Col-0 (wt). (C) Flowchart of the screen for *TAS3* tasiRNA-defective mutants using *35S:TAS3aPDS-1*-transformed plants.

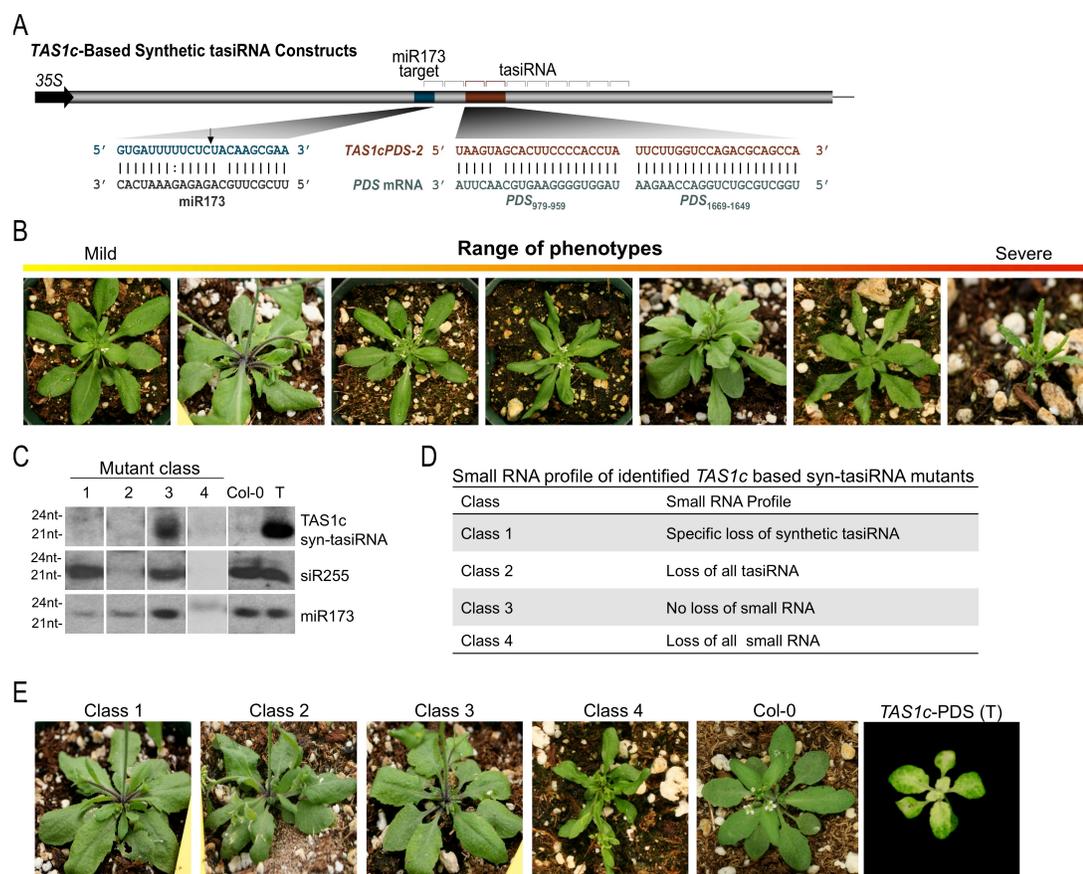


Figure 2.2. *TAS1cPDS-2* mutant analysis.

(A) Organization of syn-tasiRNA construct *TAS1cPDS-2*. The miR173 target and syn-tasiRNA/PDS target sequences are shown in the expanded diagrams. (B) Phenotypes of *TAS1cPDS-2* containing mutants. (C and D) Four mutant classes were identified based on small RNA blot assays. (E) Images of plants represented in small RNA blots in (C).

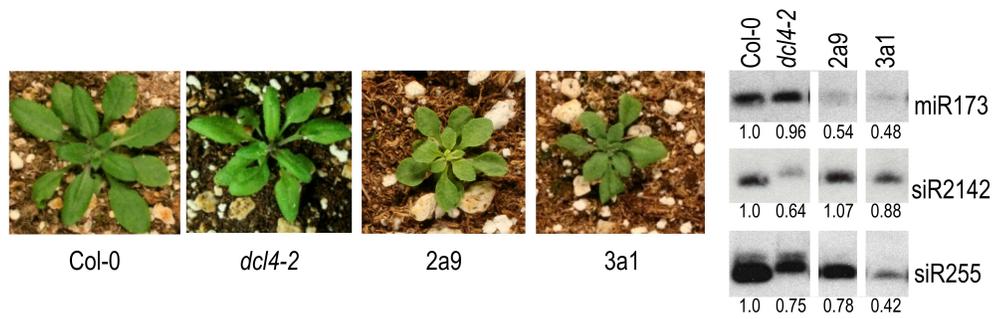


Figure 2.3 Characterization of mutants with reduced *TAS1* tasiRNA

(A) Representative images and select small RNA blot profiles from Col-0 plants, reference mutant (*dcl4-2*), and two mutants with reductions in *TAS1c* siR255 and miR173.

Figure 2.4 Characterization of *TAS3aPDS-1* class I, II and III mutants.

(A) Representative images and select small RNA blot profiles from parental *35S:TAS3aPDS-1* transformed Col-0 (T) plants, reference mutants (*hen1-5* and *ago1-25*), and Class I mutants. Small RNA data using each radiolabeled probe in each panel were from the same blot (B and C) Representative images and select small RNA blot profiles from parental *35S:TAS3aPDS-1* transformed Col-0 (T) plants, reference mutants (*rdr6-15* and *zip1*), and Class II and Class III mutants. Small RNA data using each radiolabeled probe in each panel were from the same blot. All small RNA, except for *TAS3* siR2142 in the *70b1* mutant, co-migrated with the 21-nucleotide standard. (D and E) Mean (n=3) relative level +/- Std. Dev. of *TAS3* siR2142, miR390, miR171, and *TAS1* siR255 (*TAS3aPDS-1* = 1.0).

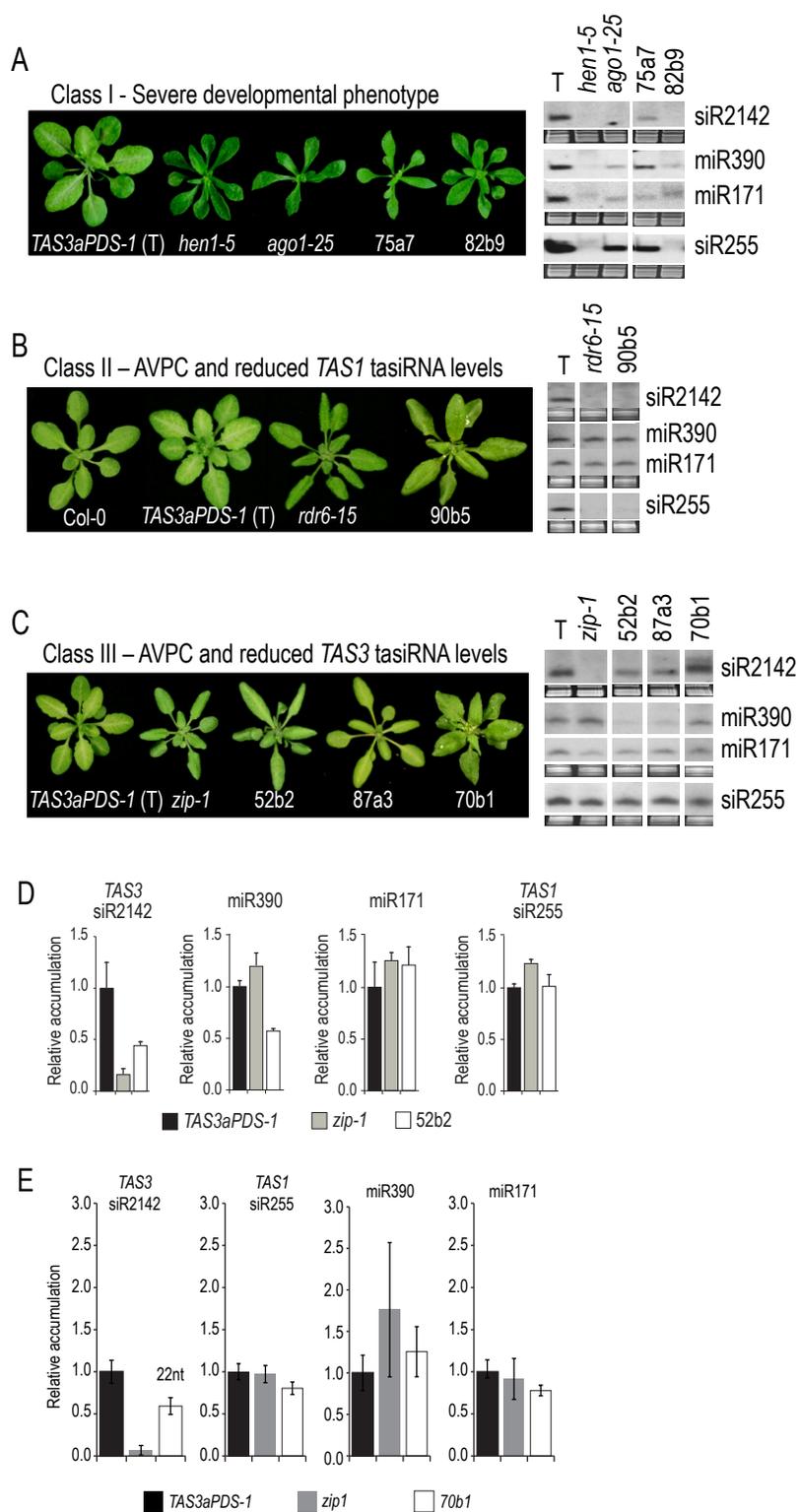


Figure 2.4 Characterization of Class I, II and III mutants

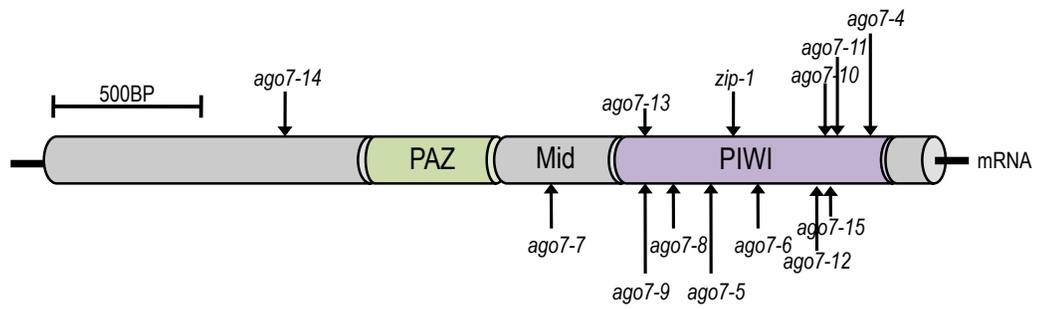


Figure 2.5. Arabidopsis AGO7 mRNA and the position of *ago7* alleles.

The mutations in *ago7-11*, *ago7-12* and *ago7-14* were nonsense, all other mutations were missense mutations. The majority of mutations resided in the catalytic PIWI domain of AGO7.

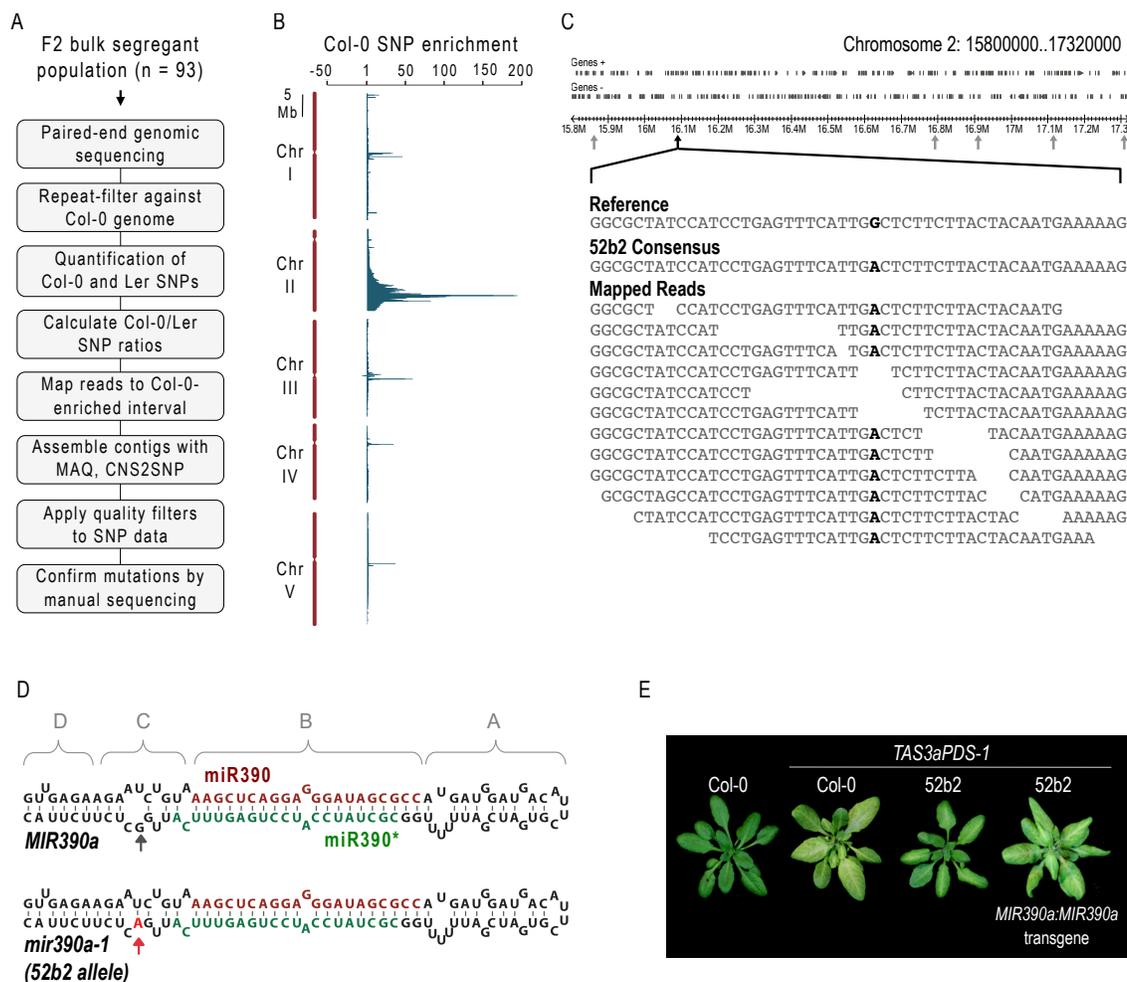


Figure 2.6. Identification of the causal mutation in 52b2 mutant

(A) Flowchart of sequence-based mapping and mutation identification using a bulk segregant population. (B) Scrolling window plot of ratios (Col-0/Ler) of total SNPs detected in the bulk segregant sequence dataset. (C) A 1.52 Mb interval spanning the major Col-0-enriched region of chromosome 2 is illustrated. Each nucleotide position that deviates from the reference genome position is indicated by an arrow. The complete or partial sequences of mapped reads from a 50 base segment (chromosome 2 16069100-16069149) from the *MIR390a* locus is shown in the expanded portion. (D) Foldback sequence and predicted structure from wild-type *MIR390a* and mutant 52b2 *mir390a-1* alleles. The position corresponding to the mutation is indicated by arrows. For comparative purposes, four foldback domains were assigned as indicated by the brackets. (E) Restoration of photobleaching phenotype in 52b2 mutant plants by transformation with a wild-type *MIR390a* transgene.

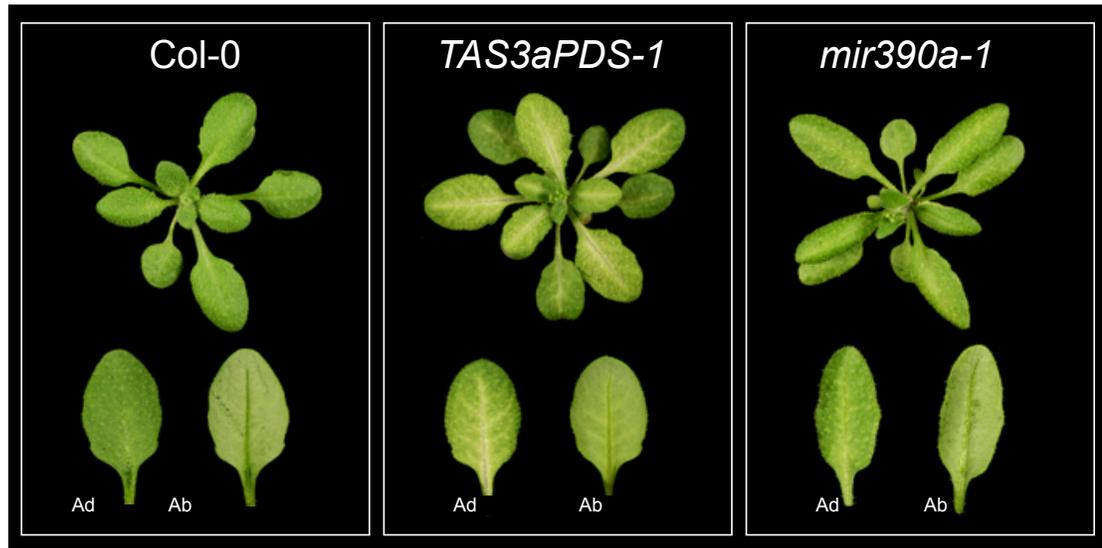


Figure 2.7. Col-0, *TAS3aPDS-1*, and *mir390a-1* rosettes.

The loss of photobleaching, accelerated vegetative phase change and leaf elongation phenotypes are evident in the *mir390a-1* mutant. The seventh mature leaf is shown separately.

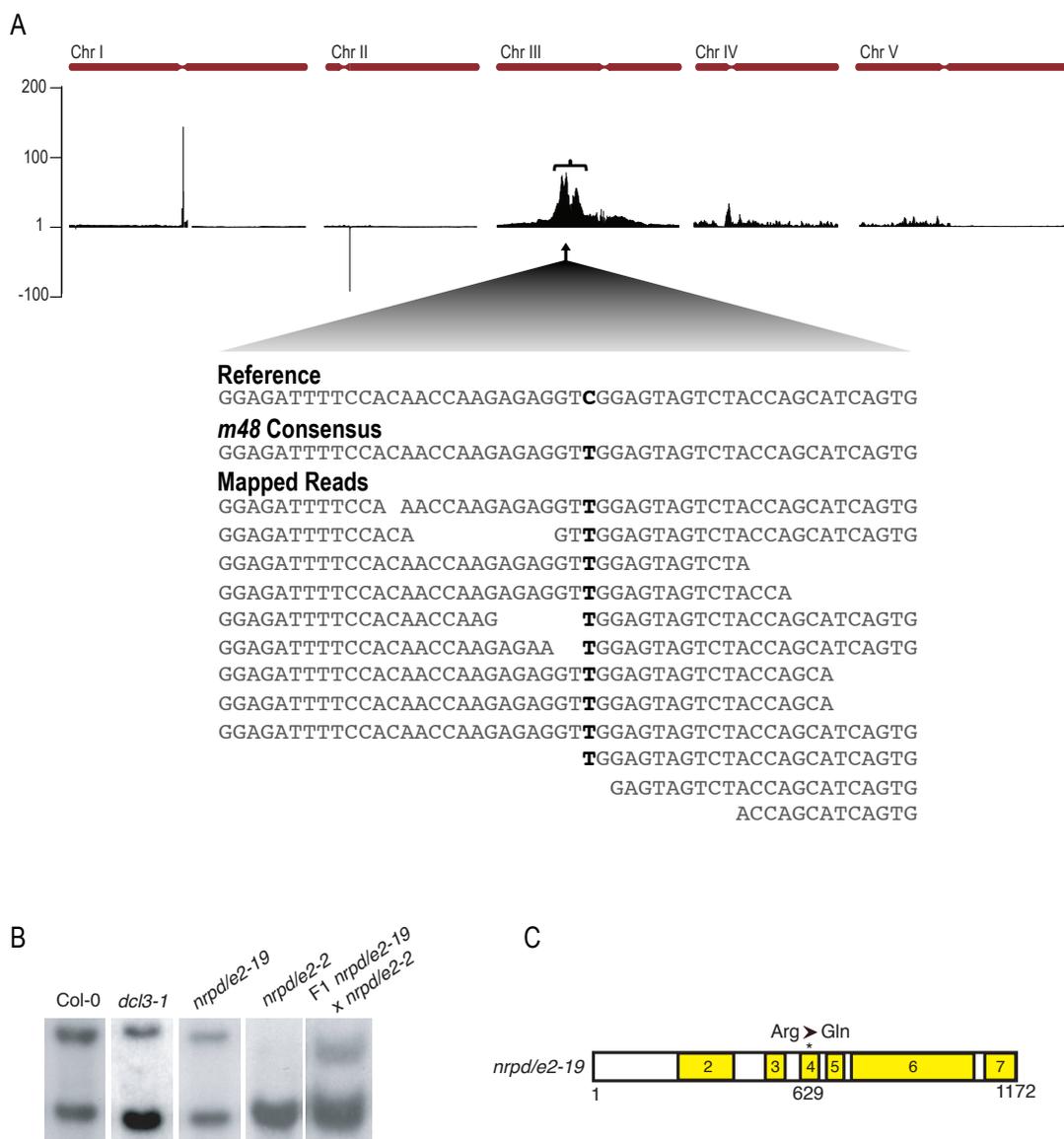


Figure 2.8. MASS identification of the causal mutation in the *m48* mutant.

(A) Scrolling window plot of ratios (Col-0/Ler) of total SNPs detected in the bulk segregant sequence dataset. The bracketed region represents the 3 MB Col-0 enriched region selected for mapping EMS mutations. The complete or partial sequences of mapped reads from a 50 base segment (chromosome 3 8570233-8570283) from the *NRPD/E2* locus is shown in the expanded portion. (B) Methylation was detected at the *MEA-ISR* locus (upper band methylated, lower band unmethylated) by southern blot analysis. *m48* has a weaker phenotype than the *nrpd/e2-2* null allele. A complementation cross with reduced methylation at *MEA-ISR* confirmed the *nrpd/e2-19* mutation was the causal mutation. (C) The cytosine to thymine mutation in *nrpd/e2* results in a missense mutation denoted by an asterisk. Numbered domains in yellow refer to conserved Rbp2 regions across the protein.

Table 2.1. Position of *ago7* and *dcl4* mutations

Allele	Genomic coordinate	AGI gene coordinate	Transcript coordinate	Reference base	Mutant base	Reference amino acid	Mutant amino acid	Independent alleles	Domain
<i>ago7-4</i>	1:26105371	3460	2831	G	A	R	K	2	Piwi
<i>ago7-5</i>	1:26105954	2877	2248	G	A	A	T	4	Piwi
<i>ago7-6</i>	1:26105896	2935	2306	C	T	S	F	1	Piwi
<i>ago7-7</i>	1:26106454	2377	1748	G	A	G	E	1	Mid
<i>ago7-8</i>	1:26106199	2632	2003	C	T	S	L	1	Piwi
<i>ago7-9</i>	1:26106703	2128	-	G	A	Splice junction		1	Piwi
<i>ago7-10</i>	1:26105750	3081	2452	C	T	L	F	1	Piwi
<i>ago7-11</i>	1:26105735	3096	2467	C	T	Q	Stop	1	Piwi
<i>ago7-12</i>	1:26105882	2949	2320	C	T	Q	Stop	1	Piwi
<i>ago7-13</i>	1:26106197	2634	2005	G	A	E	K	1	Piwi
<i>ago7-14</i>	1:26107506	1325	778	C	T	Q	stop	1	Near N-term
<i>ago7-15</i>	1:26105772	3059	2429	G	A	S	N	1	Piwi
<i>dcl4-10</i>	5:6862527	6742	3382	G	A	G	R	1	Between PAZ & RNase III
<i>dcl4-11</i>	5:6863775	5492	2546	G	A	W	Stop	1	Between dsRBD & PAZ
<i>dcl4-12</i>	5:6862527	5429	2483	G	A	G	D	1	Between dsRBD & PAZ

Table 2.2. Polymorphisms in the 1.52 Mb mapping interval in 52b2 mutant

Locus	Feature	Reference base	52b2	87a3	MAQ quality score	Read depth	Both strands
2:15841833	Intergenic	G	A	G	75	16	Y
2:16069126	MIR390a	G	A	A	54	9	Y
2:16766679	Intergenic	A ¹	G	G	51	8	Y
2:16883030	Transposon	C	T	C	45	6	Y
2:17096007	AT2G40950	G	A	G	51	8	Y
2:17302163	AT2G41470	G	A	G	63	12	Y

¹Reference base refers to TAIR8 sequence.

Table 2.3. Polymorphisms in the 3 Mb mapping interval in *m48* mutant

Chromosome	Chromosomal position	Reference base	<i>m48</i>	MAQ quality score	Read depth	Both strands	Feature
3	8183304	C	T	39	4	Y	Intergenic
3	8570260	C	T	63	12	Y	AT3G23780: NRPD/E2
3	8901305	C	T	33	3	Y	AT3G24480: Lucine Rich Repeat
3	9575990	C	T	51	8	Y	AT3G26170: Cytochrome p450
3	9872818	C	T	39	4	N	intergenic
3	9876968	C	T	36	3	N	MIR169k
3	10155521	C	T	96	23	Y	AT3G27430: 20S proteasome
3	10279989	C	T	69	14	Y	AT3G27730: DNA helicase
3	10627624	C	T	69	14	Y	AT3G28380: P-glycoprotein17
3	10972694	C	T	39	4	Y	AT3G28380: P-glycoprotein17
3	10995403	G	A	108	27	Y	AT3G28980

¹Reference base refers to TAIR8 sequence.

***MIR390a* precursor processing-defective mutants in Arabidopsis**

Josh T. Cuperus, Taiowa A. Montgomery, Noah Fahlgren, Russel T. Burke, Christopher M. Sullivan, and James C. Carrington

PNAS

700 11th Street, NW, Suite 450

Washington, DC 20001

Volume 107 (1), pages 466-471

SUMMARY

Processing of mature microRNA from *MIRNA* precursors is a poorly understood mechanism in plants. In general, *MIRNA* in plants are significantly more heterogeneous in length and structure than other eukaryotes. Here, the region near the single nucleotide mutation in the stem of *MIR390a*, leading to a significant reduction in mature miR390 levels, was investigated further. This mutation led to misprocessing of the miR390/miR390* duplex and subsequent reduction in *TAS3* tasiRNA levels. Transient expression assays in *Nicotiana benthamiana*, small RNA amplicon high-throughput sequencing, directed substitutions, and analysis of natural variation at miR390-generating loci (*MIR390a* and *MIR390b*) indicated that both base-pair properties and nucleotide identity within a region 4-6 bases below the miR390/miR390* duplex region are important for processing and that processing can vary dramatically amongst endogenous *MIRNA* loci.

INTRODUCTION

MicroRNAs (miRNAs) arise from transcripts containing self-complementary foldback structures that are initially processed to form 21-22nt miRNA/miRNA* duplexes and in conjunction with an ARGONAUTE (AGO)-containing complex usually repress mRNA in a sequence specific manner. In animals, primary transcripts with miRNA foldbacks (pri-miRNA) are processed first by the Microprocessor complex, which contains the RNase III-type protein Drosha and its cofactor Pasha (also known as DGCR8 in humans), then by Dicer, with partners that include the dsRNA-binding domain protein Loquacious (Carthew and Sontheimer, 2009). Plants orchestrate both pri-miRNA and pre-miRNA processing with the same (or very similar) complex, which includes the RNase-III like enzyme DICER-LIKE1 (DCL1) as the catalytic component (Golden et al., 2002; Park et al., 2002; Reinhart et al., 2002; Kurihara et al., 2006). DCL1 interacts with the dsRNA binding protein HYPONASTIC LEAVES1 (HYL1) and the zinc-finger protein SERRATE (SE), both of which promote efficient and accurate miRNA biogenesis (Hiraguri et al., 2005; Kurihara et al., 2006; Yang et al., 2006;

Dong et al., 2008). In contrast to animals, few distinguishable cis characteristics have been identified as promoting accurate processing of mature miRNA from *MIRNA* transcripts.

A mutation in the base of the stem of the *MIR390a* foldback was identified by a forward genetic screen to have reduced levels of mature miR390, and reduced levels of *TAS3* tasiRNA. Interestingly, this mutation increased the predicted base-pairing at the 94th nucleotide (from the 5' end of the predicted foldback), changing a guanine at position 94 (G94) to an adenine (A94), modulating a predicted weak U-G interaction to a stronger base-paired U-A. The drastic reduction of miR390 levels in *mir390a-1* was unexpected due to the redundancy of *MIR390* loci (*MIR390a* and *MIR390b* in Arabidopsis), and the predicted overlap in expression patterns of the two *MIR390* loci (Montgomery et al., 2008a). Here, the defects in processing of the recovered *mir390a-1* mutation, as well as other nucleotide changes, are further analyzed using a combination of *Nicotiana benthamiana* transient expression assays, high-throughput sequencing, and predicted secondary structure analysis of *MIR390* foldbacks, showing that the *mir390a-1* mutation, as well as *ath-MIR390b*, have reduced processivity of mature miR390a and therefore reduced *TAS3* tasiRNA production.

RESULTS

Defective Processing of the *mir390a-1* Foldback

Given that the miR390 sequence is not affected in the Arabidopsis *mir390a-1* mutant but reduces accumulation of mature miR390, *TAS3* tasiRNA, and has an accelerated phase change phenotype, it seems reasonable to suggest that the mutation in the *mir390a-1* foldback limits processing efficiency or accuracy. Interestingly, *MIR390a* and *MIR390b* loci specify the identical miR390 sequence, but the foldbacks differ in sequence and predicted base-pair structure at the base of the stem. The C region (loop distal region, see Figure 2.6) from the *MIR390b* foldback contains more predicted base-paired positions at and adjacent to C112, which occupies the spatially equivalent position as G94 in *MIR390a*. To analyze the effects of the *mir390-1* A94 mutation, as well as the differences in the C region between

MIR390a and *MIR390b* foldbacks, on miR390 biogenesis and *TAS3* tasiRNA formation, a transient expression assay based on *Agrobacterium*-mediated delivery of constructs to *Nicotiana benthamiana* leaves was used (Llave et al., 2002). All constructs were expressed using a constitutive 35S promoter.

35S:MIR390a, *35S:mir390a-1*, and *35S:MIR390b* were expressed individually to analyze miR390 biogenesis and accumulation, or coexpressed with *35S:TAS3aPDS-2* (syn-tasiRNA) and *35S:HA-AGO7* to test for *TAS3* tasiRNA initiation activity. When expressed individually or with the other *TAS3* tasiRNA components, *35S:mir390-1* yielded miR390 at 28.3% ($p < 3.02 \times 10^{-5}$) or 28.6% ($p < 0.002$) the levels measured from *35S:MIR390a* (Figure 3.1A), respectively, which was consistent with the low levels of *TAS3* tasiRNA detected in the 52b2 mutant plants (Figure 2.2C,D). Interestingly, *35S:MIR390b* also yielded low levels of miR390 when expressed individually (17.0%, $p < 9.11 \times 10^{-6}$) or with *TAS3* tasiRNA components (19.8%, $p < 0.0019$) (Figure 3.1A). Additionally, the functional amounts of miR390 as reflected in the amounts of *TAS3*-based syn-tasiRNA were significantly lower using *35S:mir390-1* (21.9%, $p < 0.0072$) and *35S:MIR390b* (33.5%, $p < 0.013$), compared to using *35S:MIR390a* (Figure 3.1A). These data suggest that processing of the *mir390a-1* foldback, as well as the naturally occurring *MIR390b* foldback, occurs inefficiently.

To analyze processing accuracy of the *MIR390a*, *mir390a-1* and *MIR390b* foldbacks, small RNA libraries from triplicate samples were subjected to high-throughput sequencing analysis after transient expression in *N. benthamiana*. Reads were first normalized based on library size and spike-in standards (Fahlgren et al., 2009). Reads from within 29-nt windows, centered around the middle of the annotated miR390 or miR390* sequences, were then analyzed for size, 5' position and 3' position by calculation of Shannon's entropy (H) (Shannon, 1948; Schneider, 1997), providing a set of measures of small RNA uniformity or processing accuracy at both ends of each sequence (Figure 3.1C). *MIR390a* yielded predominantly 21-nucleotide, canonical miR390 with highly uniform 5' and 3' ends, and moderately heterogenous 20-21-nucleotide miR390* sequences with uniform 5' ends but 3'

ends from two major positions (Figure 3.1B, Figure 3.2). In contrast, *35S:mir390a-1* yielded 5', 3'- and size-heterogeneous miR390 and miR390* sequences, with only $45.6\% \pm 22.6\%$ of miR390-related sequences containing accurately processed 5' and 3' ends (Figure 3.1, Figure 3.2). This was reflected in high *H* values for each miR390 and miR390* parameter (Figure 3.1C). *MIR390b* yielded sequences with intermediate processing accuracy. Both ends of miR390, and the 5' end of miR390*, exhibited more heterogeneity than the comparable ends of sequences from *MIR390a* (Figure 3.1B,C). Combined with the syn-tasiRNA biogenesis data (Figure 3.1A), these experiments indicate that the *mir390a-1* mutation affects both processing accuracy and efficiency, resulting in low levels of functional miR390. They also indicate that *MIR390b* possesses the properties of a low-efficiency mutant allele with a moderate processing heterogeneity.

Mutational Analysis of the *MIR390a* foldback

The G-to-A substitution in the *mir390a-1* mutant could conceivably debilitate processing due to a change in foldback base-pairing, loss of a base determinant, or both. Computational analysis of predicted foldback variants suggested that the *mir390a-1* structure possessed a higher probability of base-pairing between U12 and A94, compared to the probability of pairing between U12 and G94 in the wild-type foldback (Figure 3.3A). This was reflected in a lower calculated entropy at both positions in the *mir390a-1* foldback (Figure 3.3A, (Shannon, 1948; Schneider, 1997)). The *MIR390b* predicted foldback, with even more extensive base-pairing, yielded lower calculated positional entropies at nearly all bases in region C (Figure 3.3A, (Shannon, 1948; Schneider, 1997)). Seven *35S:MIR390a* mutants with substitutions at either position 94 and/or position 12 were constructed. Including *mir390a-1*, the series resulted in foldbacks containing all possible single-base substitutions at both positions, and two combinations of dual-base substitutions (Figure 3.3A). Additionally, the sequences comprising *MIR390a* region C were substituted for the approximate equivalent sequences from *MIR390b*.

Predicted foldback structures, positional entropies and miR390 biogenesis levels in a transient assay were determined (Figure 3.3A,B,C).

Each substitution at position 94 (*mir390a-1*, *mir390a-94U* and *mir390a-94C*) resulted in significantly ($p < 0.003$) lower miR390 levels compared to wild-type *MIR390a*, although the *mir390a-94U* and *mir390a-94C* defects were only modest (Figure 3.3B,C). *mir390a-1*, but not *mir390a-94U* and *mir390a-94C*, was predicted to result in increased base-pairing with U12 (Figure 3.3A). These data generally reinforce a role for G94, either as a single base-determinant or as a high-entropy, weak base-pair partner with U12, in foldback recognition or processing. Among the position 12 substitutions, *mir390-12C* was significantly ($p < 8.8 \cdot 10^{-9}$) debilitated for miR390 biogenesis and was predicted to form a low-entropy basepair with G94. *mir390a-12A* retained both A12 and G94 in a predicted non-paired configuration and yielded wild-type levels of miR390 (Figure 3.3A,B,C). These position 12 mutants lend support to the idea that a non-paired or weakly paired G94 contributes to miR390 biogenesis. In contrast, *mir390a-12G* was predicted to adopt a fold involving low-entropy, highly base-paired 12G and G94 positions, but led to wild-type levels of miR390 (Figure 3.3A,B,C). However, the *mir390a-12G* local stem structure was predicted to include novel, high-entropy asymmetric bulges that differed from the comparable positions from *MIR390a* (Figure 3.3A).

Among the double mutants, *mir390a-12C94A* contained the A94 mutation from *mir390a-1* and a basepair-disrupting change at position 12 (Figure 3.3A). This mutant was highly debilitated for miR390 biogenesis, indicating that the *mir390a-1* defect (A94) was not due solely to the increased base-pair configuration between positions 12 and 94 (Figure 3.3B,C). Interestingly, the *mir390a-12G94U* mutant foldback, which contained the G and U positions from wild-type *MIR390a* reversed, yielded nearly wild-type levels of miR390 (Figure 3.3B,C).

Finally, substitution of the base of the *MIR390a* stem with that from *MIR390b* led to significant ($p < 6.73 \cdot 10^{-6}$) debilitation of miR390 formation (Figure 3.3B,C). The mutant foldback region C was predicted to contain the same low-entropy, highly base-paired configuration as predicted for *MIR390b* foldback (Figure 3.3A).

DISCUSSION

The *mir390a-1* mutant allele was novel, as it possessed foldback defects that have not been identified previously. The position of the G-to-A mutation strongly suggests that a determinant for processing efficiency and accuracy resides at or around the proximal end of the foldback. To date, there are few known "rules" that define determinants for recognition and processing of plant foldbacks by DCL1-HYL1-SE complexes. The relatively high diversity of sizes and secondary structures of plant *MIRNA* foldbacks (Axtell, 2008) means that such determinants are not particularly obvious by sequence or structure comparisons. Based on elegant in vitro processing assays with *MIR167b* foldbacks, DCL1 is sufficient to catalyze ATP-dependent pri- and pre-miRNA transcript processing, although only a minority of such products possess accurate 5' and 3' ends (Dong et al., 2008). The dsRNA binding motifs of DCL1 may provide a basal function for foldback recognition. However, inclusion of both SE and HYL1 in these reactions increases the rate and accuracy of processing (Dong et al., 2008). This may indicate that SE and HYL1 function as accessory factors that position DCL1 accurately on substrates through interaction with one or more structural features.

We propose that the inaccurate and inefficient processing of the *mir390a-1* foldback is due to loss of interaction with key factors promoting miRNA biogenesis. In particular, it is attractive to consider G94 in a flexible, high-entropy context as a recognition determinant for HYL1 and/or SE. Both SE and HYL1 promote miR390 accumulation in vivo (Montgomery et al., 2008b; Chitwood et al., 2009). Importantly, the effects of the *mir390a-1* mutation on foldback processing in transient assays are very similar to the effects of *MIR167b* foldback processing in the absence of SE and HYL1 in vitro (Dong et al., 2008). It seems unlikely, however, that foldback position G94 is the sole determinant for such interactions, as there is high sequence and structural diversity at this position among foldbacks from conserved *MIRNA* families. By analogy with the Drosha-Pasha/DGCR8 complex interacting with the base of animal foldbacks (reviewed in (Kim et al., 2009)), features defining the junction between the

base of the stem and the non-paired region outside of the stem may also interact with the DCL1-HYL1-SE complex for positioning of the first set of cuts at the proximal end of the miRNA/miRNA* duplex.

Although *MIR390a* and *MIR390b* specify an identical miR390 sequence, the two Arabidopsis loci are not equivalent. *MIR390b* is insufficient to cover the requirements for miR390 in the absence of full *MIR390a* function, as the mutant screen revealed. miR390 functions at two sites within *TAS3* transcripts, and clearly *MIR390a* is necessary for adequate miRNA levels. The transient assays indicated that the *MIR390b* foldback is inherently inefficient for miR390 biogenesis, and the mutational analysis and genetic swap experiments showed that this was due, at least in part, to the stem structure below the miR390/miR390* region. The predicted structural variation between the two loci, whereby the *MIR390a* locus possesses a predicted high-entropy G94 position and the *MIR390b* locus possesses a low-entropy base-paired region, is conserved throughout the *Brassicaceae* (data not shown). However, the functional significance of the *MIR390b* locus remains unclear.

MATERIALS AND METHODS

Mutagenesis of *MIR390a*

Six oligonucleotides were used to create *MIR390a* substitution constructs. Briefly, 5' and 3' fragments that partially overlapped the *MIR390a* foldbacks were amplified with KOD polymerase (Novagen) using a plasmid containing 35S:*MIR390a* as template. The *MIR390a* genomic primers flanked the miR390 sequence by 250 nucleotides on each side:

MIR390a F [caccTATAGGGGGGAAAAAAGGTAG]

MIR390a R [GAGACTAAAGATGAGATCTA]

MIR390b F 5' [caccTTCCAAAATATGTAATATGGGGA]

MIR390b R 5' [CTAACAAACTGCTTAGATGTGTGAA].

Sequences CACC on forward primers are not genomic, but were added for cloning into pENTR/D-TOPO (Invitrogen). Forward primers were combined with RI primers, while reverse flanking primers were combined with F2 primers (Table 3.2). Fragments were gel purified, then mixed with loop containing fragments, or overlapping oligos containing loop sequence, in a second round of PCR. Round 2 PCR fragments were gel purified.

RNA folding

Computational analysis of foldback sequences shown in Figure 5A was done using RNAfold (rna.tbi.univie.ac.at, (Zuker, 2003)) with the following options: -p -T 22 -d2. Shannon's entropy (Shannon, 1948; Schneider, 1997) at each position was calculated using the program RNAdist.pl by summing entropy values for all base-pairing probabilities. Color-coding of entropy values was done using the program relplot.pl. All programs are part of the Vienna package (Bonhoeffer et al., 1993).

Transgene constructs

Transgene sequences were PCR-amplified from genomic DNA. *MIR390a* constructs were designed similarly to *MIR319a* based artificial microRNAs (Schwab et al., 2006). These were introduced into pENTR/D-TOPO (Invitrogen) and subsequently recombined with LR clonase (Invitrogen) into pMDC32 (Supplemental Experimental Procedures (Curtis and Grossniklaus, 2003)).

Small RNA Sequencing from Transient Expression Assays.

Small RNA amplicons were prepared in triplicate as described (Fahlgren et al., 2009). Four synthetic oligoribonucleotides (Std2, Std3, and Std6 [see (Fahlgren et al., 2009)] and Std11 [pUGUCCGACACGAUGCAGAUCC]) were added to 40 µg total RNA per sample before amplicon preparation at four concentrations (Std11, 0.0001 pmol; Std6, 0.001 pmol; Std3, 0.01 pmol; Std2, 0.1 pmol). In addition, samples were barcoded using four variants of the standard

5' adaptor (5'GUUCAGAGUUCUACAGUCCGACGAUAAC3' [barcode A],
5'GUUCAGAGUUCUACAGUCCGACGAUCCC3' [barcode C], 5'
GUUCAGAGUUCUACAGUCCGACGAUGGC3' [barcode G], and
5'GUUCAGAGUUCUACAGUCCGACGAUUUC3' [barcode U], barcoded sequence
underlined) and multiplexed. Sequencing by synthesis was done with an Illumina Genome
Analyzer I (GAI). Multiplexed amplicons (four samples, 2.5 pmol total) were added per lane.
Reads were computationally parsed based on detection of the 5' barcode (AAC, CCC, GGC,
or TTC) and the first six nucleotides of the 3' adapter (CTGTAG). Read proportions were
based on total reads (18–24-nucleotides) that matched perfectly within a 29-base window
surrounding the annotated miR390 or miR390* sequences from MIR390a and MIR390b.
Control samples to measure the low levels of endogenous miR390-related sequences in *N.*
benthamiana leaves were prepared after transient expression of 35S:GUS (Table 3.1).
Sequence and size information content from miR390-related sequences recovered after the
transient assays was analyzed using Shannon's entropy formula (Shannon, 1948).
Calculations were done independently for 5' end, 3' end and size

ACKNOWLEDGMENTS

We thank Tiffany Townsend and Amy Shatswell for plant growth help. We thank Mark
Desanko for Illumina sequencing sample prep. This work was supported by grants from the
National Science Foundation (MCB-0618433), National Institutes of Health (AI43288) and
United States Department of Agriculture–National Research Initiative (2006-35301-17420).

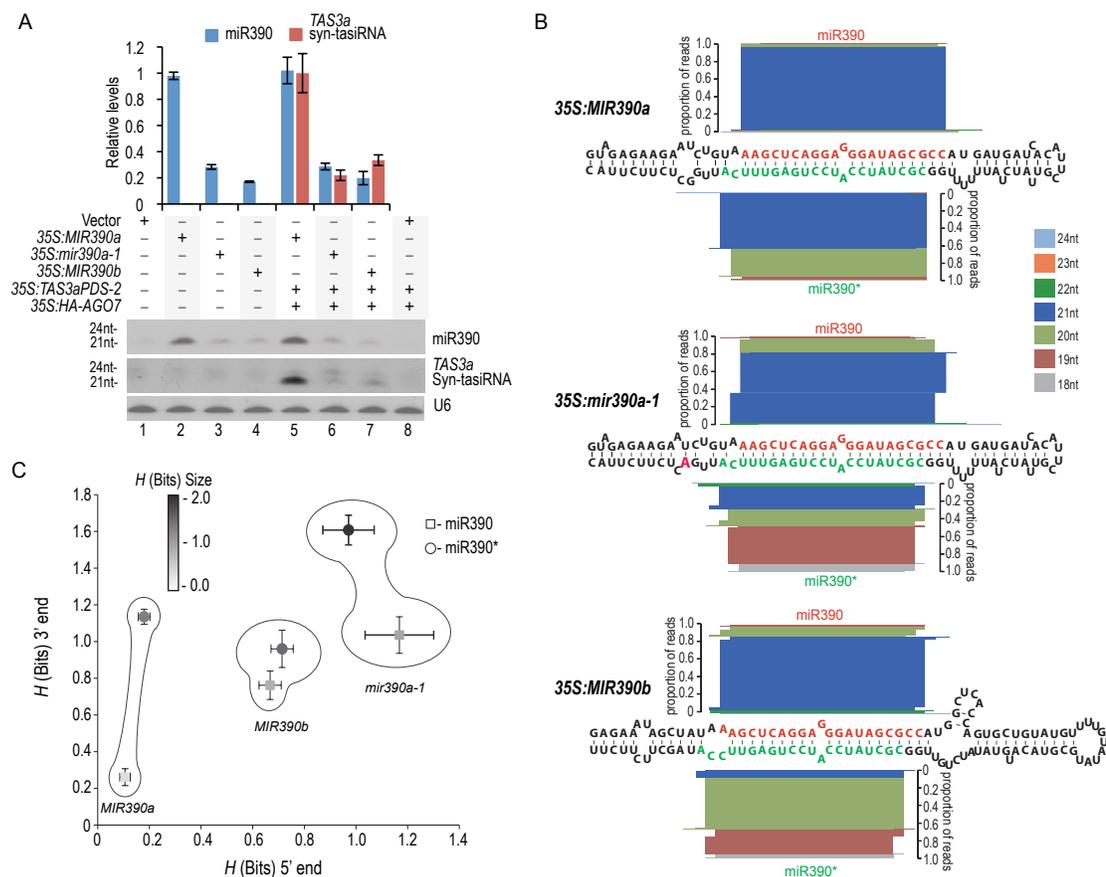


Figure 3.1. Foldback processing and *TAS3* tasiRNA-initiation activity of *mir390a-1* in transient assays

(A) Accumulation of miR390 and *TAS3* syn-tasiRNA in *N. benthamiana* transient assays. One of three independent replicates is shown. Mean ($n=3$) relative miR390 (blue) and *TAS3a* (red) syn-tasiRNA levels \pm Std. Dev. (lane 2 and lane 5 = 1.0) are shown in the graph. Syn-tasiRNA levels were measured only in assays containing 35S:*TAS3aPDS-2* (lanes 5-8). U6 RNA is shown as a loading control. (B) Analysis of miR390 and miR390* sequences after transient expression of *MIR390a*, *mir390a-1* and *MIR390b*. Proportions of reads containing specific sequences are plotted as stacked bars based on size (color coded), 5' position and 3' position, with end positions aligned to the respective sequences shown in the foldbacks. miR390 and related sequences are plotted upwards, and miR390* sequences are plotted downwards. (C) Shannon's Entropy (H) for 5' end (x-axis), 3' end (y-axis) and size (grayscale) of small RNA populations shown in panel B. High H values reflect high information content.

Figure 3.2. Small RNA read percentages mapped to the *ath-MIR390a* locus from transient overexpression assays.

Percentage of small RNA reads mapping within a 29-nt window centered on miR390 or miR390* from transient expression assay (A–D) Percentage of small RNA reads mapping to miR390 or miR390* from a transient expression assay. Offset refers to the 5' position of reads, where 0 is the 5' position of miR390 or miR390*. A negative offset refers to positions 5' upstream of miR390 or miR390*, whereas positive offset refers to positions 3' downstream of miR390 or miR390*. Reads were mapped against appropriate, infiltrated MIR390 foldback, or, in the case of 35S:GUS, reads were mapped to both MIR390a and MIR390b.

A

35S:MIR390a miR390

Offset	Small RNA Size Class																							
	18			19			20			21			22			23			24					
	Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
-4	-	0.158	-	-	-	-	-	-	-	-	-	-	-	-	0.228	-	-	-	-	-	-	-		
-3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.047		
-1	-	-	-	-	-	-	-	-	-	-	0.093	0.315	0.456	0.233	-	-	0.047	-	-	-	-	-		
0	0.093	-	-	-	-	-	2.33	2.681	4.556	96.32	96.06	94.31	-	-	0.456	-	-	-	-	-	-	-		
1	-	-	-	-	-	-	0.093	0.158	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
2	-	-	-	0.14	-	-	-	-	-	-	-	-	-	0.047	-	-	-	-	-	-	-	-		
3	-	0.158	-	-	-	-	-	-	-	-	-	0.158	-	-	0.158	-	-	-	-	-	-	0.28		
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.28	0.047		

35S:MIR390a miR390*

Offset	Small RNA Size Class																							
	18			19			20			21			22			23			24					
	Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.109	-	-	-	-	-	-	0.218		
-3	-	-	-	-	0.109	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.218		
-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
0	0.352	0.327	0.656	3.612	3.162	0.984	26.87	36.1	33.77	66.26	57.36	62.95	0.264	-	-	-	0.109	-	-	-	-			
1	0.264	0.436	-	1.145	1.418	0.328	0.705	0.218	0.328	0.088	-	-	-	-	-	-	-	-	-	-	-			
2	0.088	0.218	0.984	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.176			
3	0.088	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
4	-	-	-	-	-	-	-	-	-	-	-	-	0.088	-	-	-	-	-	-	-	-			

B

35S:mir390a-1 miR390

Offset	Small RNA Size Class																							
	18			19			20			21			22			23			24					
	Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
-4	-	-	-	-	-	-	0.484	-	0.423	1.937	0.196	2.114	-	-	-	-	-	-	-	-	1	-		
-3	-	-	-	-	-	0.423	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
-2	-	-	-	-	0.196	-	-	-	-	-	-	-	-	0.242	-	-	-	-	-	-	-	-		
-1	-	-	-	-	-	-	-	-	-	57.38	21.22	24.74	-	0.098	0.423	-	-	-	-	-	-	-		
0	-	0.196	-	-	0.423	17.68	13.56	13.53	20.1	63.26	53.49	-	-	-	-	-	-	-	-	-	-			
1	-	-	-	-	0.196	-	0.295	1.48	-	0.098	-	-	-	-	-	-	-	-	-	-	-			
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.196	0.211		
3	-	0.196	-	-	-	-	-	-	-	-	-	-	-	0.634	-	-	-	-	-	-	0.969	0.196		
4	-	-	-	-	-	-	-	-	-	-	-	0.423	-	-	-	-	-	-	-	-	0.969	0.098		

35S:mir390a-1 miR390*

Offset	Small RNA Size Class																							
	18			19			20			21			22			23			24					
	Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
-4	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	1	-	-	-	-		
-3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1		
-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
-1	-	-	-	-	-	-	-	-	0.153	-	-	-	-	-	-	-	-	-	-	-	-	1		
0	-	0.321	-	0.726	0.963	1.529	9.144	14.77	14.83	7.257	31.78	29.82	-	-	-	-	-	-	-	-	-	-		
1	12.05	7.223	5.046	59.51	32.74	32.87	3.483	5.618	7.034	4.499	2.729	4.893	2.032	2.889	3.364	-	-	-	-	-	-	-		
2	0.29	0.161	0.153	-	-	-	-	0.161	0.153	-	-	-	-	-	-	-	-	-	-	-	-	-		
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
4	0.726	-	-	-	-	-	0.145	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

Figure 3.2. Small RNA read percentages mapped to the *ath-MIR390a* locus from transient overexpression assays.

C

			Small RNA Size Class																							
			18			19			20			21			22			23			24					
			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
Offset	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3					
-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
-3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
-1	-	-	-	-	-	-	-	-	-	1.6	1.613	1.14	7.2	2.581	-	4.8	0.323	-	-	-	-	-				
0	-	-	-	0.285	-	0.968	5.128	5.6	8.387	86.61	76.8	75.81	0.57	-	-	-	-	-	-	-	-	0.8				
1	-	-	-	0.285	-	1.935	2.849	3.2	5.161	3.134	-	1.935	-	-	-	-	-	-	-	-	-	-				
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
3	-	-	-	-	-	-	-	-	-	-	-	-	0.968	-	-	-	-	-	-	-	-	-				
4	-	-	-	-	-	-	-	-	0.323	-	-	-	-	-	-	-	-	-	-	-	-	-				

			Small RNA Size Class																							
			18			19			20			21			22			23			24					
			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate			Replicate					
Offset	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3					
-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
-3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.253	-	-	-	-	-	-	-	-				
-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
-1	-	-	-	-	-	0.345	-	-	-	-	-	-	1.382	0.253	-	-	-	-	-	-	-	-				
0	1.034	-	0.253	7.241	9.677	5.808	66.55	43.78	62.63	6.552	7.834	8.586	-	-	-	-	-	-	-	-	-					
1	-	5.991	2.525	16.9	27.19	16.92	0.345	1.843	1.768	-	-	-	-	-	-	-	-	-	-	-	-					
2	1.034	1.843	0.758	-	-	-	-	0.461	0.253	-	-	-	-	-	-	-	-	-	-	-	-					
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					

Figure 3.2. Small RNA read percentages mapped to the *ath-MIR390a* locus from transient overexpression assays (Continued).

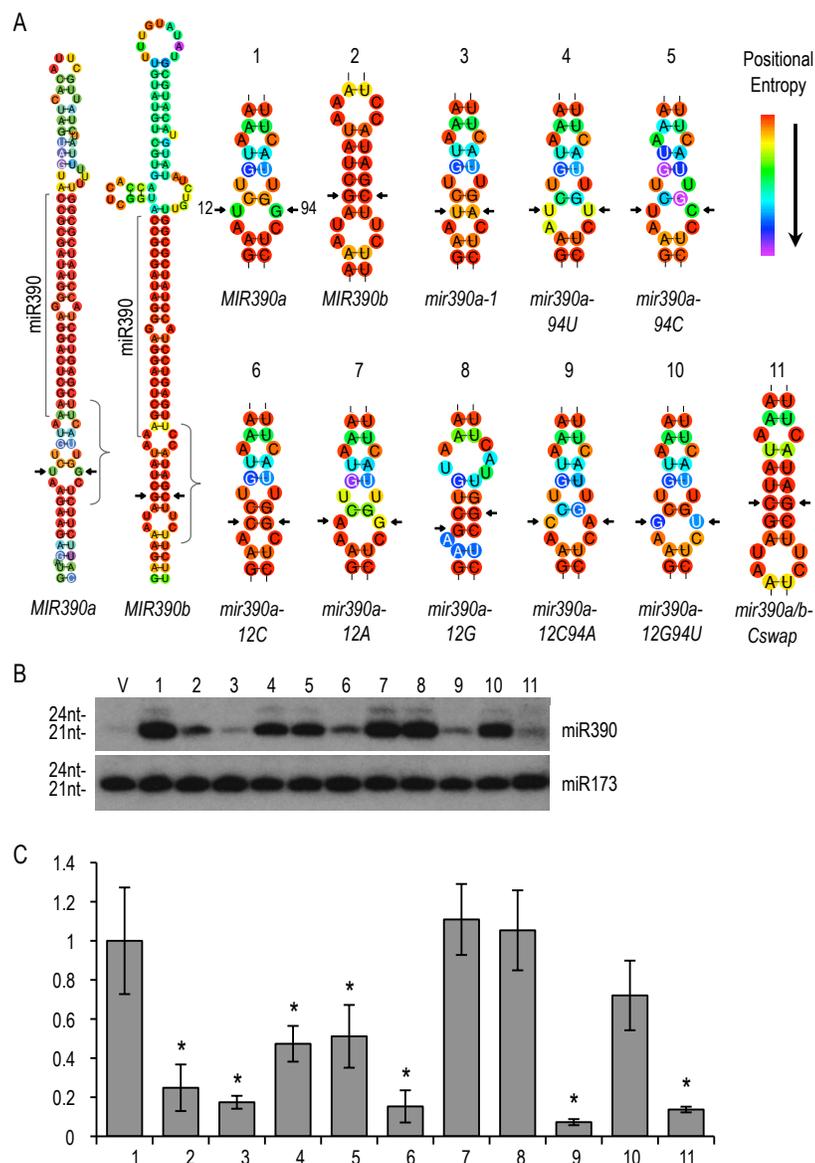


Figure 3.3. Directed mutational analysis of the *MIR390a* foldback

(A) Predicted basepair structure of *MIR390a* and *MIR390b* foldbacks (left). The enlarged region corresponds to the bracketed region for 11 mutant or variant foldbacks (right). Positions 12 (U in *MIR390a*) and 94 (G in *MIR390a*) are indicated by arrows. Positional entropy values range from 0 (red) to 1.6 (purple) for all *MIR390a*-based foldbacks and 1.8 (purple) for *MIR390b*. Folding and Shannon's entropy values reflect the probability of variant base-pair states, and were calculated using RNAfold (Zuker, 2003). (B) Blot assays for miR390 derived from expression of *35S:MIR390a* (lane 1), *35S:MIR390b* (lane 2), *35S:mir390a-1* (lane 3) and each directed mutant construct (lanes 4-11) are shown, along with a negative control sample expressing empty vector (V). All experimental constructs were co-expressed with *35S:MIR173* as an internal control. One of six independent replicates is shown. (C) Mean relative miR390 levels +/- Sts. Dev. (*35S:MIR390a* = 1.0)

Table 3.1. Small RNA reads mapping within a 29nt window centered on miR390 or miR390* from transient expression assay.

Sample	Replicates		
<i>35S:MIR390a</i> miR390	2146	634	439
<i>35S:MIR390a</i> miR390*	1135	917	305
<i>35S:mir390a-1</i> miR390	413	1018	473
<i>35S:mir390a-1</i> miR390*	689	623	654
<i>35S:MIR390b</i> miR390	351	125	310
<i>35S:MIR390b</i> miR390*	290	217	396
<i>35S:GUS</i> miR390	14	46	36
<i>35S:GUS</i> miR390*	11	30	28

Table 3.2. Primers used to design MIR390-based constructs

Primer description	Sequence*
MIR390a F	caccTATAGGGGGGAAAAAAGGTAG
MIR390a R	GAGACTAAAGATGAGATCTA
MIR390b F	caccTTCCAAAATATGTAATATGGGGA
MIR390b R	CTAACAACTGCTTAGATGTGTGAA
MIR390a R primer I	GGCGCTATCCCTCCTGAGCTTTACAGATTCTTCTCTACTTTG
MIR390a F primer II	GGCGCTATCCATCCTGAGTTTCATTGGCTCTTCTACTACAATG
MIR390a F primer III	AAGCTCAGGAGGGATAGCGCCATGATGATCACATTTCGTTATC
MIR390a R primer IV	TGAAACTCAGGATGGATAGCGCCAAAAATAGATAACGAATGTGATC
MIR390a R primer I 12C	GGCGCTATCCCTCCTGAGCTTTACAG G TTCTTCTCTACTTTG
MIR390a R primer I 12G	GGCGCTATCCCTCCTGAGCTTTACAG C TTCTTCTCTACTTTG
MIR390a R primer I 12A	GGCGCTATCCCTCCTGAGCTTTACAG T TTCTTCTCTACTTTG
MIR390a F primer II 96A	GGCGCTATCCATCCTGAGTTTCATTG A CTCTTCTTACTACAATG
MIR390a F primer II 96C	GGCGCTATCCATCCTGAGTTTCATTG C CTCTTCTTACTACAATG
MIR390a F primer II 96U	GGCGCTATCCATCCTGAGTTTCATTG T CTCTTCTTACTACAATG
390aSWAP R primer I:	GGCGCTATCCCTCCTGAGCTTT ATAGCTAT TTTCTCTACTTTG
390aSWAP F primer II:	GGCGCTATCCATCCTGAGTTTC ATAGCTTC TTTCTTACTACAATG
71mer_amiRNA	AAGCTCAGGAGGGATAGCGCCATGATGATCACATTTCGTTATCTATTTTTTGGCG CTATCCATCCTGAGTTTCA
71mer_amiRNA_rc	TGAAACTCAGGATGGATAGCGCCAAAAATAGATAACGAATGTGATCATCATGG CGCTATCCCTCCTGAGCTT
MIR390b F3	AGAATCTGTAAAGCTCAGGAGGGATAGCGCCATGGCTCACCAGTGCTGTATG
MIR390b R4	AGAGCCTTTGGAACCTCAGGATGGATAGCGCCAACAGATATACATGTA

* Red nucleotides represent nucleotide substitutions in MIR390a constructs

**Unique Functionality of 22 nt miRNAs in Triggering RDR6-Dependent
siRNA Biogenesis from Target Transcripts in Arabidopsis**

Josh T. Cuperus, Alberto Carbonell, Noah Fahlgren, Hernan Garcia-Ruiz, Russell T. Burke,
Atsushi Takeda, Christopher M. Sullivan, Sunny D. Gilbert, Taiowa A. Montgomery and
James C. Carrington

Nature Structural & Molecular Biology, Nature Publishing Group
The Macmillan Building,
4 Crinan Street,
London, N1 9XW, UK

SUMMARY

RNA interference pathways may involve amplification of secondary siRNAs by RNA-dependent RNA polymerases. In plants, RDR6-dependent secondary siRNAs arise from transcripts targeted by some microRNA (miRNA). Here, *Arabidopsis thaliana* secondary siRNA from mRNA, and trans-acting siRNA, are shown to be triggered through initial targeting by 22-nucleotide miRNA that associate with AGO1. In contrast to canonical 21-nucleotide miRNA, 22-nucleotide miRNA primarily arise from foldback precursors containing asymmetric bulges. Using artificial miRNA constructs, conversion of asymmetric foldbacks to symmetric foldbacks resulted in production of 21-nucleotide forms of miR173, miR472 and miR828. Both 21 and 22-nucleotide forms associated with AGO1 and guided accurate slicer activity, but only 22-nucleotide miRNA were competent to trigger RDR6-dependent siRNA from target RNA. These data suggest that AGO1 functions differentially with 21 and 22-nucleotide miRNA to engage the RDR6-associated amplification apparatus.

INTRODUCTION

During RNA interference (RNAi), double stranded RNA (dsRNA) is processed by Dicer, a dsRNA-specific RNase III class ribonuclease, into small ~20-30-nucleotide RNA duplexes. Typically, one strand of the duplex preferentially associates with an Argonaute protein to form an effector complex (RNA-induced silencing complex, or RISC), that targets and silences transcripts based on sequence complementarity (Kim et al., 2009). In plants, fungi, nematodes, and some other organisms, this process can be amplified through production of secondary small interfering RNAs (siRNAs) after transcription by RNA-dependent RNA polymerase (RdRp) on the primary target RNA (Ghildiyal and Zamore, 2009). As exemplified in *Caenorhabditis elegans*, primary siRNA, secondary siRNA and other small RNA may serve as guides in association with distinct Argonaute proteins (Wassenegger and Krczal, 2006; Guang et al., 2008; Claycomb et al., 2009; Gu et al., 2009). Additionally, dsRNA precursors or siRNA have the potential to act non-cell-autonomously in plants and some animals, leading to

spread of silencing signals, subsequent amplification, and transgenerational silencing in some cases (Carthew and Sontheimer, 2009; Ghildiyal and Zamore, 2009).

In amplification-competent organisms, the selective recruitment and activity of RdRp on primary targets is a key step that governs whether or not secondary siRNA and amplification occurs. The partially redundant RdRps RRF-1 and EGO-1 from *C. elegans*, QDE-1 from *Neurospora crassa*, Rdp1 from *Schizosaccharomyces pombe*, and RNA-DEPENDENT RNA POLYMERASE6 (RDR6) from *Arabidopsis thaliana* transcribe RNA in a primer-independent manner (Makeyev and Bamford, 2002; Meister and Tuschl, 2004; Motamedi et al., 2004; Pak and Fire, 2007; Sijen et al., 2007; Correa et al.). *A. thaliana* may encode six RdRps, at least three of which generate long dsRNA that serve as DICER-LIKE (DCL) substrates in several distinct pathways (Voinnet, 2008; Ghildiyal and Zamore, 2009). RDR6, which frequently functions in combination with DCL4, participates in posttranscriptional silencing of exogenous targets (transgene transcripts, some viral RNA), as well as several types of endogenous transcripts (Mallory et al., 2008; Voinnet, 2008). Indiscriminant entry of transcripts into the RDR6-dependent siRNA amplification pathway does not occur in plants. Amplification of secondary siRNA occurs, however, from a minority of small RNA-targeted transcripts, and the genetic requirements for biogenesis of these siRNA resemble those of trans-acting siRNA (tasiRNA) (Peragine et al., 2004; Vazquez et al., 2004b; Axtell et al., 2006; Rajagopalan et al., 2006; Ronemus et al., 2006; Howell et al., 2007). Tightly controlled entry of transcripts into amplification pathways makes intuitive sense, as feed-forward amplification of silencing signals could conceivably result in runaway suppression of off-target transcripts from related gene family members. In *Arabidopsis*, the XRN family of exonucleases functions to degrade sliced transcript fragments (Chiba and Green, 2009), and sequencing analysis using an *xrn4* mutant suggests that XRN4 is antagonistic to secondary siRNA biogenesis (Gregory et al., 2008). How secondary siRNA-producing transcripts are routed differentially from most other target transcripts is understood poorly.

Plant tasiRNA represent a case in which an RdRp-dependent, secondary siRNA-producing pathway has evolved to produce specialized small RNA that function to repress co-evolved targets. The four known families of tasiRNA (*TAS1-TAS4*) in Arabidopsis are particularly informative as experimental systems to understand formation of secondary siRNA. tasiRNAs form from primary *TAS* transcripts that are initially targeted and sliced by AGO1-miR173 (*TAS1*, *TAS2*), AGO1-miR828 (*TAS4*) or AGO7-miR390 (*TAS3*) complexes (Allen et al., 2005; Baumberger and Baulcombe, 2005; Yoshikawa et al., 2005; Rajagopalan et al., 2006; Montgomery et al., 2008a). *TAS1*, *TAS2* and *TAS4* tasiRNA-generating transcripts are cleaved at a 5'-proximal site. The 3' RNA product is then transcribed by RDR6, leading to dsRNA that is sequentially processed by DCL4 to yield phased tasiRNA in register with the cleavage site (Allen et al., 2005; Dunoyer et al., 2005; Gascioli et al., 2005; Xie et al., 2005a; Yoshikawa et al., 2005; Rajagopalan et al., 2006; Montgomery et al., 2008b).

Here, the basis for selective entry of miRNA-targeted transcripts through the RDR6-dependent siRNA-generating pathway was investigated. Combining genome-wide analyses of 21-nucleotide siRNA and miRNA-targeting patterns with directed experiments, miRNA length was found to be a key determinant in triggering amplification in the context of AGO1-miRNA complexes. In addition, determinants within *MIRNA* foldbacks that govern miRNA size were identified.

RESULTS

Secondary siRNAs originate from some targeted mRNAs

Six *A. thaliana* small RNA libraries were generated from aerial plant tissue and sequenced in a multiplexed format, yielding an average of 658,999 reads/library that mapped to at least one genomic position. Among 20-24-nucleotide reads, 53.9% were 21-nucleotides, the vast majority of which corresponded to miRNA (Figure 4.1). A significant proportion of 20-24-nucleotide reads (74.4%) mapped to 979 annotated transcripts, 55 of which are known to be

targeted by one or more miRNA or tasiRNA (Figure 4.2Aa, Supplementary Table 1). Although targeted transcripts represented only 5.6% of the 979 small RNA-generating transcripts, reads from the targeted set represented 27.0% of reads from the entire transcript set. Among reads mapping to the targeted transcripts, 85.4% were 21-nucleotide in length, whereas the majority (52.9%) of reads derived from non-targeted transcripts were 24-nucleotides in length (Figure 4.2A). Twenty-one-nucleotide siRNA originating from the targeted transcripts were shown previously to be largely RDR6- and DCL4-dependent (Howell et al., 2007). Thus, among Arabidopsis siRNA originating from annotated transcripts, those from targeted transcripts are more likely to yield siRNA of 21-nucleotides in length.

A total of 280 miRNA-target or tasiRNA-target pairs were collated from published data (Supplementary Table 2). Given that multigene families with many members targeted by a specific miRNA or tasiRNA family tend to bias representation on this list, "non-redundant" sets of miRNA-target and tasiRNA-target pairs were generated. One hundred non-redundant set iterations (50 pairs in each) contained only one paralog/target family/small RNA, with the targeted family member chosen randomly for each family in each iteration. Most targeted transcripts from the non-redundant list yielded no 21-nucleotide siRNA (Figure 4.2C, Supplementary Table 2). However, on average 33.8% of targets yielded at least a few siRNAs, with several targets yielding highly abundant 21-nucleotide siRNA (Figure 4.2C). Levels of these abundant siRNAs were decreased in the *dcl2-1 dcl3-1 dcl4-2* (termed *dcl234*) triple mutant (Figure 4.2C).

In exploring the differences between small RNA-target pairs that led to secondary siRNA production and those that did not, it was noticed that secondary siRNA-generating transcripts were frequently targeted by miRNA or tasiRNA of 22-nucleotides in length (Figure 4.2D). Targets that interacted with miRNA or tasiRNA containing less than 22-nucleotides yielded significantly ($p=0.034$, permutation test) less siRNA on average. Additionally, 47% of singly targeted transcripts and 43% of multiply targeted transcripts presented in Figure 4.2A were

targeted by a small RNA of predominantly 22-nucleotides in length. The biogenesis and activity of 22-nucleotide miRNAs were therefore subjected to further analysis.

Asymmetric *MIRNA* foldbacks yield 22-nucleotide miRNA

The vast majority of Arabidopsis miRNA reads were either 21 or 22-nucleotides in length, with 22-nucleotide families representing 22.5% of all annotated *MIRNA* families (Figure 4.3A). The overall abundance of 22-nucleotide miRNA reads, however, was relatively low (2.3%). Both the percentage of 22-nucleotide-generating *MIRNA* families, and the numbers of 22-nucleotide miRNA reads, were comparably low in rice (Figure 4.3A; Supplementary Table 3,4). Size was also a distinct characteristic for most *MIRNA* families in both Arabidopsis and rice, with relatively few families expressing a mixed size distribution. Comparing the proportions of 21 or 22-nucleotide miRNA from non-redundant loci in Arabidopsis and rice revealed a distinct bimodal distribution; over 90% of all *MIRNA* loci generated miRNA that were >80% either 21-nucleotide or 22-nucleotide in length (Figure 4.3B; Supplementary Table 3,4). To understand the basis for 21 vs. 22-nucleotide miRNA formation, foldback base-pairing patterns of Arabidopsis *MIRNA* were examined. Strikingly, those Arabidopsis foldbacks that contained asymmetry in the form of a non-paired nucleotide within the miRNA sequence were significantly enriched for 22-nucleotide miRNA reads ($p < 2.6 \times 10^{-9}$, Wilcoxon rank sum test) (Figure 4.3C). In nearly all cases of predominantly 22-nucleotide miRNA, including miR173, miR393 and miR472, an asymmetric non-paired base was detected; there were a few exceptions, such as miR828, which arose from a foldback containing only symmetric mispairs within the miRNA/miRNA* segment (Figure 4.3D, Supplementary Table 3).

To determine if the asymmetric positions within miRNA sequences of foldbacks led to formation of the 22-nucleotide size miRNA, artificial miRNA (amiRNA) were expressed using foldbacks containing either asymmetric (wild-type configuration) or symmetric miRNA/miRNA* segments. These were produced using the *MIR390a* foldback, in which the miR390/miR390* sequences were replaced by miR173/miR173* sequences (Montgomery et al., 2008a). The

symmetric foldback was engineered by the addition of a U residue to base-pair with the normally mispaired A residue within the *MIR173* foldback (Figure 4.4A). The authentic *MIR173* foldback, and the asymmetric and symmetric amiRNA foldbacks, were coexpressed with the tasiRNA-generating construct, *35S:TAS1c*, in transient assays using *Nicotiana benthamiana* (Allen et al., 2005), and both miRNA and tasiRNA accumulation (see below) were measured in blot assays and by sequence analysis. Neither miR173 nor *TAS1c* tasiRNA are conserved in *Nicotiana* species, so accumulation of these small RNAs is entirely dependent on the transient expression assay. The asymmetric foldback, from either the authentic (*35S:MIR173*) or artificial (*35S:amiR173*) constructs, yielded predominantly 22-nucleotide-size miR173 in blot assays (Figure 4.4B). In contrast, the symmetric foldback (*35S:amiR173-21*) yielded a 21-nucleotide-size miR173 (Figure 4.4B). Sequence analysis of small RNAs from these assays confirmed that the asymmetric foldbacks yielded predominantly 22-nucleotide miR173, although a substantial amount of off-size products of 19 and 20-nucleotides were sequenced (Figure 4.4C, Figure 4.5, Supplementary Table 5). The symmetric foldback, however, yielded miR173 that was 91% 21-nucleotide in length (Figure 4.4C, Figure 4.5, Supplementary Table 5). The 21-nucleotide form lacked the 3'C residue present in authentic miR173.

Several previous reports collectively showed that miR173 formation requires DCL1, but not DCL2 (Park et al., 2002; Xie et al., 2005b; Bouché et al., 2006; Montgomery et al., 2008b). DCL2, however, was shown to produce 22-nucleotide siRNA from endogenous Arabidopsis dsRNA (Gascioli et al., 2005; Xie et al., 2005a). To determine if DCL2, or siRNA-generating DCL3 and DCL4, are necessary for 22-nucleotide miRNA, miRNA reads sequenced from Col-0 and the *dcl234* triple mutant were compared. Accumulation and family representation levels of most miRNAs in the *dcl234* triple mutant, regardless of size, were similar to those in wild-type Col-0 plants (Figure 4.3A, Figure 4.6A,B). miRNA reads from Col-0 and *dcl234* mutant plants for most families, except for two (miR822 and miR839, which are known to be DCL4-dependent (Rajagopalan et al., 2006)), were highly correlated (Pearson product-moment correlation coefficient [r] = 0.95) (Figure 4.6B). A high correlation (r = 0.94) was shown when

only 22-nucleotide miRNA reads were compared (Figure 4.6B). Therefore, all known 22-nucleotide miRNA in Arabidopsis are likely DCL1-dependent.

22-nucleotide forms of miR173, miR472, and miR828 trigger siRNAs

As shown previously (Allen et al., 2005; Montgomery et al., 2008b) and in Figure 4.4B, coexpression of *35S:MIR173* with *35S:TAS1c* triggered tasiRNA biogenesis. Authentic-size miR173 generated from *35S:amiR173* also triggered *TAS1c* tasiRNA formation (Figure 4.4B, lane 5). The tasiRNA pools triggered by both 22-nucleotide miR173 sources were highly phased, with the register set by the miR173 cleavage site (Figure 4.4C, Figure 4.5, Supplementary Table 6). The 21-nucleotide form produced from the amiR173-21 foldback, however, possessed very low tasiRNA trigger activity, and the few *TAS1c*-derived siRNA that were produced exhibited no phasing pattern (Figure 4.4B,C; Figure 4.5; Supplementary Table 6).

The lack of tasiRNA-trigger activity of miR173-21 could be due to one or more reasons, including failure to associate with AGO1, failure to interact with and guide cleavage of the *TAS1c* primary transcript, or loss of post-cleavage functions necessary to recruit RDR6 to the precursor transcript. Mi et al. showed that, among AGO1, AGO2, AGO4 and AGO5, miR173 associated most commonly with AGO1. To confirm that both 21 and 22-nucleotide forms of miR173 associated with AGO1, *TAS1c* biogenesis assays in *N. benthamiana* were done with co-expression of HA-tagged AGO1, followed by co-immunoprecipitation analysis. Both 22 and 21-nucleotide forms of miR173 co-immunoprecipitated with HA-AGO1 (Figure 4.4D, lanes 4, 6 and 8). *TAS1c* tasiR255, produced in the presence of 22-nucleotide miR173, also co-immunoprecipitated with HA-AGO1 (Figure 4.4D, lanes 4 and 6). As HA-AGO1-nonassociated controls, accumulation *TAS1c* 3'D2(-) tasiRNA, which contains an AGO1-nonpreferred 5'A, and U6 RNA were analyzed. No or very low levels of *TAS1c* 3'D2(-) tasiRNA and U6 RNA were detected in the immunoprecipitated samples (Figure 4.4D), indicating that HA-AGO1 was selective. To assess more broadly whether or not AGO1 associates preferentially with 21 or

22-nucleotide miRNA, small RNAs from total RNA (input fraction) and immunoprecipitated HA-AGO1 complexes (IP fraction) were identified by high-throughput sequencing (two replicates each). Of the miRNAs that were enriched at least 2-fold in the IP fraction, miRNA families that were predominantly 20-24-nucleotides were recovered, suggesting that AGO1 does not include or exclude miRNA based on length (Figure 4.7). All predominantly 22-nucleotide miRNA families that met read thresholds for inclusion in the enrichment calculation were enriched in the IP fraction. Additionally, for two predominantly 21-nucleotide miRNAs (miR397a and miR167a,b,d) that had a >30% subpopulation of 22-nucleotide variants (Figure 4.3B), both the 21 and 22-nucleotide variants were enriched more than 2-fold in the IP fraction (Figure 4.7).

To assess guide function of 22 and 21-nucleotide miR173 forms, 5' RACE assays were done for each co-expression assay using primers to detect the 3' product of miR173-guided cleavage. Products were detected in all three co-expression assays, and sequencing confirmed that cleavage guided by both 22 and 21-nucleotide miR173 forms occurred at the canonical target site (Figure 4.4E,F). Furthermore, both 22 and 21-nucleotide miR173 forms destabilized *TAS1c* full-length transcripts to very low or undetectable levels (Figure 4.8).

Because the 21-nucleotide form derived from the *35S:amiR173-21* foldback contained a 3'A rather than a 3'C residue, as is present in 22-nucleotide authentic miR173, the effect of 3' nucleotide identity on both 21 and 22-nucleotide forms of miR173 was examined. First, a construct (*35S:amiR173-21-3'C*) that generated a symmetric foldback yielding a 21-nucleotide form containing a 3'C was tested (Figure 4.8A). This 21-nucleotide-size miR173 failed to trigger tasiRNA accumulation despite guiding accurate cleavage of the *TAS1c* transcript (Figure 4.8B,C lane 2). Second, to assess the possibility of a preference for a specific 3' nucleotide in a 22-nucleotide context, 22-nucleotide miR173 variants containing 3' G, A or U were tested. Each of the variants accumulated to within 36% of the level of 22-nucleotide miR173 containing a natural 3'C and guided accurate cleavage of the *TAS1c* transcript (Figure

4.8B,C). Interestingly, although each of the 22-nucleotide 3'G, 3'A and 3'U variants triggered tasiRNA accumulation, the levels of tasiRNA generated were relatively low (Figure 4.8B).

To determine if the 22-nucleotide requirement to trigger tasiRNA is specific to miR173, 22 and 21-nucleotide artificial miRNAs for both miR472 and miR828 were generated using the *MIR390a*-based foldback, and tested using the *TAS1c* tasiRNA transient assay system. Both asymmetric and symmetric foldbacks were engineered following the same strategy as for the miR173 variants (Figure 4.9A) and co-expressed with modified *TAS1c* constructs (*35S:TAS1c-472* and *35S:TAS1c-828*) in which the normal miR173 target site was substituted with sites recognized by miR472 or miR828. Based on mobility of miR472 and miR828 forms in blot assays, 22 and 21-nucleotide variants were generated as predicted from asymmetric and symmetric foldbacks, respectively, although the asymmetric miR828 foldback yielded a mixture of both size classes (Figure 4.9B). Importantly, tasiRNA were detected only in those samples containing 22-nucleotide variants of miR472 and miR828 (Figure 4.9B). Like the miR173 size variants, both the 22 and 21-nucleotide forms of miR472 and miR828 guided accurate cleavage of the engineered tasiRNA transcripts (Figure 4.9B,C) and co-immunoprecipitated with HA-AGO1 (Figure 4.9D). These data clearly show that 22-nucleotide forms of three Arabidopsis miRNAs possess unique functionality to direct tasiRNA or secondary siRNA biogenesis.

DISCUSSION

This study associated a unique class of Arabidopsis miRNA - those that are 22-nucleotides in length - with triggering formation of secondary, RDR6/DCL4-dependent 21-nucleotide siRNAs from primary miRNA targets. While both 22-nucleotide and 21-nucleotide forms of miR173, miR472 and miR828 associated with AGO1 and guided accurate cleavage of target transcripts, advancement of target RNA fragments through the 21-nucleotide siRNA biogenesis pathway was a unique property of the 22-nucleotide miRNA-target RNA pairs. Previous studies provided hints that 22-nucleotide miRNAs might be associated with

secondary siRNA. Arabidopsis *MIR168b* was shown to generate both 21 and 22-nucleotide forms of miR168 (Vaucheret, 2009), which targets the transcript encoding AGO1 and leads to secondary siRNAs in a homeostatic regulatory loop (Mallory and Vaucheret, 2009; Vaucheret, 2009). In rice, 22-nucleotide miRNA target sites were identified in several abundant siRNA-generating transcripts (Johnson et al., 2009), suggesting that the 22-nucleotide property may be conserved in angiosperms.

The fact that most plant miRNAs, particularly those from the most highly expressed families, are 21-nucleotide in length may explain why secondary siRNA from most targeted transcripts are absent or in low abundance. While the tasiRNA represent highly refined examples of functional, discrete secondary siRNAs from a non-coding RNA, it is interesting to consider potential functions for the cases in which abundant siRNAs originate from protein-coding transcripts. In at least some examples, secondary siRNA might specifically target transcripts from related family members. For gene families that are evolving or expanding rapidly, secondary targeting may suppress dosage effects. Evidence for the functionality of secondary siRNA from large multigene family transcripts has been presented (Chen et al., 2007; Howell et al., 2007; Addo-Quaye et al., 2008).

It is proposed that 22-nucleotide miRNA-AGO1 complexes, but not 21-nucleotide miRNA-AGO1 complexes, mark transcripts for RDR6-dependent siRNA formation. This could involve AGO1 adopting one of two states - an amplification-trigger state and a non-trigger state. Both states would be competent to interact with and slice targets containing a suitable target site, but only the trigger state would recruit RDR6 to the de-capped (sliced) target. Previous results from experiments with a non-cleavable target of miR173 suggested that secondary siRNA formation depends on a functional cleavage site (Montgomery et al., 2008b). Several studies have also shown that all known tasiRNA, and many other RDR6-dependent siRNA in Arabidopsis (Allen et al., 2005; Axtell et al., 2006; Rajagopalan et al., 2006; Howell et al., 2007), originate preferentially adjacent to target cleavage sites. Recruitment of RDR6, therefore, may depend on target cleavage.

Conformational changes of well-characterized Argonaute proteins are known to occur upon interaction with target RNA and release of the guide 3' end from the PAZ domain (Wang et al., 2008; Wang et al., 2009). How a 22-nucleotide RNA would effect an AGO1 conformation distinct from a 21-nucleotide RNA is not immediately obvious. The proposed trigger state, however, could involve a direct interaction with RDR6 or associated factors, such as SGS3 (Peragine et al., 2004; Yoshikawa et al., 2005; Kumakura et al., 2009). To date, there are no data suggesting direct interaction between AGO1 and RDR6 or known RDR6-associated factors. Alternatively, conformationally distinct AGO1-small RNA complexes may interact differentially with GW domain proteins, which interact directly with AGO proteins as scaffolds to mediate interactions with other factors to suppress translation and promote de-adenylation (Eulalio et al., 2009). Arabidopsis AGO1-specific GW proteins are not yet characterized, although AGO4 likely interacts with GW domains in an RNA PolIV subunit (NRPDE1) and the transcription elongation factor SPT5L (El-Shami et al., 2007; Bies-Etheve et al., 2009; He et al., 2009).

While it is clear that AGO1 exhibits a preference for small RNAs with a 5'U, which docks within a MID domain binding pocket, there are no clear structural or functional data supporting a 3' nucleotide preference for interaction with the PAZ domain. Among 22-nucleotide miRNAs in Arabidopsis, a 3'C is most common, although this implies neither a binding preference nor particular functionality. Within a 22-nucleotide context, a 3'C promoted siRNA-trigger function of miR173 most effectively. However, there was no strict requirement for any specific 3' nucleotide, and the 22-nucleotide siRNA-triggers miR828 and miR393 both contain 3'A. The studies presented here do not allow quantitative comparison of the affinity of AGO1 for 21 vs. 22-nucleotide miRNAs, or for 22-nucleotide miRNAs with different 3' nucleotides.

The finding that 22-nucleotide miRNA trigger RDR6-dependent secondary siRNAs leads to the question of whether or not other siRNA amplification pathways have guide RNA-specific activities to recruit RDR1 or RDR2. Signals for amplification mediated by RDR1 and RDR6 during antiviral RNAi are not known, although results presented here suggest 22-nucleotide

siRNA, working through AGO1 or other AGO proteins, might be plausible triggers. As DCL2 is known to both promote a full silencing response against several viruses (Blevins et al., 2006; Bouché et al., 2006; Deleris et al., 2006; Fusaro et al., 2006; Diaz-Pendon et al., 2007; Garcia-Ruiz et al., 2010) and to generate 22-nucleotide siRNAs (Gascioli et al., 2005; Xie et al., 2005a), this idea may have support. DCL2 was also shown to promote DCL4-dependent posttranscriptional silencing of several transgenes (Mlotshwa et al., 2008). Furthermore, overproduction of DCL2-dependent 22-nucleotide siRNA may account for the severe phenotypes of *dcl134* and *dcl14* mutant Arabidopsis plants, in which DCL2 might generate 22-nucleotide siRNA from dsRNA that would normally be processed to 21-nucleotide siRNA by DCL4 (Bouché et al., 2006).

Finally, the basis for production of either a 21-nucleotide or a 22-nucleotide miRNA depends on the nature of the foldback in plants. Most Arabidopsis miRNAs are formed by DCL1-mediated processing first at the loop-distal end of the foldback, then at the loop proximal end (Mateos et al.; Song et al.; Werner et al.). DCL1, like other Dicers, functions as a molecular ruler. For the loop-proximal cuts, DCL1 "measures" the length of helical RNA from the PAZ domain-bound end to the positions juxtaposed to the active centers (Qin et al.; MacRae et al., 2006; MacRae et al., 2007). For foldbacks that are symmetrically paired within the miRNA/miRNA* region, it is inferred that DCL1 measures a length equivalent to 21 base-pairs in most cases. However, it is also inferred that length of an A-form helix containing a single asymmetric bulge is frequently equivalent to that of a symmetrically paired helix. Three-dimensional modeling of 21-nucleotide- and 22-nucleotide-generating *MIR173* foldbacks using MC-Fold and MC-Sym (Parisien and Major, 2008) supports this idea. The most likely models show that an unpaired nucleotide does not substantially increase the length of the miRNA/miRNA* duplex region (Figure 4.10). This explains why, when the asymmetric nucleotide occurs within the miRNA sequence, the resulting miRNA contains a non-canonical 22-nucleotide length. Given the unique functionality of 22-nucleotide miRNA to trigger siRNA amplification, it seems reasonable to propose that foldback structures with asymmetric

positions, and that would result in 22-nucleotide miRNA, should be under evolutionary constraints.

METHODS

Accession codes.

High-throughput sequencing datasets for HA-AGO1-associated small RNA and input controls were deposited in Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under the series accession GSE22252. Datasets for Col-0 and *dcl2-1 dcl3-1 dcl4-2* small RNA libraries were published previously⁴⁷ and are available under the series accession GSE20197. Twenty-eight *O. sativa* Nipponbare libraries, as described in Supplementary Information, were also used.

Plant materials and growth conditions.

N. benthamiana and *A. thaliana* plants, including the previously described *dcl2-1 dcl3-1 dcl4-2* triple mutant (Garcia-Ruiz et al., 2010), were grown in normal greenhouse conditions with supplemental light on a 16hr light/ 8 hr dark cycle.

Transgene constructs.

35S:TAS1c and *35S:MIR173a* derived constructs were generated in pMDC32 and described previously (Allen et al., 2005; Montgomery et al., 2008a; Montgomery et al., 2008b). *MIR390a*-derived artificial miRNA constructs were designed by ligating overlapping oligonucleotides into a pMDC32-derived vector containing ~200bp upstream and downstream of the *MIR390a* foldback.

Transient expression assays.

Transient expression assays in *N. benthamiana* leaves were carried out as previously described (Llave et al., 2002) with *Agrobacterium tumefaciens* GV3101. Bacterial cultures were resuspended at OD₆₀₀=1.0. When multiple constructs were co-expressed, equal amounts

of each culture were used. Plants were grown in the greenhouse after infiltration, and zones of infiltration were harvested 48 hours post-infiltration for RNA isolation.

RNA Blot assays.

Total RNA from *Arabidopsis* and *N. benthamiana* was isolated using TRIzol reagent (Invitrogen). Two chloroform extractions were done and RNA was precipitated in an equal volume of isopropanol for 20 min. Triplicate samples from pools of *N. benthamiana* infiltrated leaves were analyzed. RNA blot assays were done as described (Montgomery et al., 2008b). Briefly 5 µg, 10 µg or 20 µg of total normalized RNA was resolved by denaturing polyacrylamide-gel electrophoresis for small RNA analysis, or 5 µg of total RNA was resolved by denaturing 1.5% agarose-gel electrophoresis. RNA was transferred to positively charged nitrocellulose membrane. DNA or LNA probes were end-labeled using [³²]ATP and Optikinase (USB). Probes were hybridized to RNA on membranes in Sigma Perfect-Hyb at 68°C (HMW blots) or 38-42°C (small RNA blots). An Instant Imager (Packard Bioscience) was used to measure blot hybridization signals.

Small RNA sequencing.

Small RNA amplicons from transient assays in *N. benthamiana* were prepared as described (Fahlgren et al., 2009; Cuperus et al., 2010a). Four synthetic oligoribonucleotides were added in different amounts (Std11, 0.0001 pmol; Std6, 0.001 pmol; Std3, 0.01 pmol; Std2, 0.1 pmol) to 90 µg total RNA per sample before amplicon preparation. Sequencing-by-synthesis was done with five pmol of each amplicon using an Illumina Genome Analyzer II (GAII, www.illumina.com). Reads were parsed using the first six nucleotides of the 3' adapter (CTGTAG). Reads were normalized per million as described (Fahlgren et al., 2009). Read proportions were based on total reads (18-24-nucleotides) that matched perfectly within a 29 base window surrounding the miR173 or miR173* sequences from samples containing 35S:*MIR173a*, 35S:*amiR173* and 35S:*amiR173-21*. Radar plots to display phasing from

35S:*TAS1c* were generated as previously described (Axtell et al., 2006; Montgomery et al., 2008a). Small RNA amplicons were generated and analyzed from HA-AGO1 IP samples as described (Montgomery et al., 2008a).

Small RNA analysis.

Small RNA from *A. thaliana* were mapped to the genome and annotated transcriptome (TAIR9; <http://www.arabidopsis.org>), including annotated *MIRNA* precursors (miRBase v14; <http://www.miRBase.org>), using the CASHX pipeline⁵⁹. Those transcripts to which four or greater small RNA reads/million (RPM) were mapped in at least four of the six libraries were defined as small RNA-generating transcripts. For analysis of small RNA-generating transcripts, annotated structural RNA, pseudogenes or transposable elements were removed.

Statistical analyses.

All statistical analyses were done using R v2.9.2. Permutation tests were done using the R “twotPermutation” function, with 1 million simulations, from the DAAG package (Maindonald and Braun, 2007). A Wilcoxon rank sum test was used to compare empirical logits for the ratio of 22 to 21-nucleotide miRNA reads. The rank sum test was done using the R “wilcox.test” function from the stats package.

5' RACE.

Modified RNA ligase-mediated 5' RACE was done as described (Montgomery et al., 2008b) with the Generacer kit (Invitrogen) using the following gene specific primers:

TAS1c_707_5'_RACE [GATGATGCTTCTTCGCTACACCTCGGAG] and

TAS1c_573_5'_RACE [AGCAACTGTTCTTTAGACGACTTGAAAATCTCAT]. 5' RACE

products were gel purified using Invitrogen PureLink gel extraction kit, cloned in TOPO TA (Invitrogen), introduced into *E. coli*, screened for inserts, and sequenced.

Small RNA datasets.

Small RNA data sets used in this paper are available from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>). Data for *Arabidopsis* wild type Col-0 and *dcl234* were previously described (Garcia-Ruiz et al., 2010) and are available from GEO (GSE20197). Samples used in this work were GSM506656, GSM506657, GSM506658, GSM506662, GSM506663 and GSM506664 for Col-0; and GSM506680, GSM506681, GSM506686, GSM506687 and GSM506688 for *dcl234*. Data from rice are available from GEO series GSE11014 (GSM278532, GSM278533, GSM278534, GSM278535, GSM278571 and GSM278572) (Zhu et al., 2008), GSE16350 (GSM409313, GSM409314, GSM409315, GSM409316, GSM409317, GSM409318, GSM409319, GSM409320, GSM409321, GSM409322, GSM409323 and GSM40924) (Johnson et al., 2009), GSE16248 (GSM407071 and GSM407072) (Johnson et al., 2009), GSE12317 (GSM309691, GSM309692 and GSM309693) (Zhou et al., 2009), GSE19602 (GSM489087) (He et al., 2010), and GSE13152 (GSM329296, GSM329297, GSM329298 and GSM329299) (Heisel et al., 2008).

MIRNA secondary and tertiary structure prediction.

MIRNA secondary structures were generated using MC-Fold, a knowledge-based folding algorithm that utilizes data from the Protein Data Bank (PDB) to take into account both canonical Watson-Crick and non-canonical RNA base-pairs commonly found in solved RNA structures (Parisien and Major, 2008).

MIRNA tertiary structures were predicted using MC-Sym (Parisien and Major, 2008). MC-Sym structures were constrained using the MC-Fold structure. Initially, 1000 tertiary models are returned. These initial models were sorted based on all-atoms force-fields and the ten models with lowest free energy were kept. Because MC-Sym constructs tertiary models using fragments from the PDB, the phosphate backbone has stretches of discontinuity. To correct for discontinuity, the models were first relieved until the atomic force's root-mean-square (RMS) was < 100 Kcal/mol/Å, then refined until the RMS was < 5 Kcal/mol/Å (Parisien and

Major, 2008). Tertiary structures were viewed using open-source PyMOL v1.2 (www.pymol.org).

ACKNOWLEDGMENTS

A.C. was supported by a postdoctoral fellowship from the Ministerio de Ciencia e Innovacion (BMC-2008-0188). This work was supported by grants from the National Science Foundation (MCB-0618433 and MCB-0956526), National Institutes of Health (AI43288) and Monsanto Corporation.

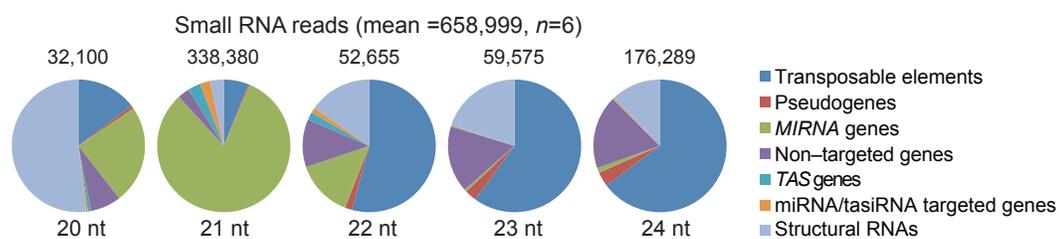


Figure 4.1. Arabidopsis small RNA distribution

Distribution of variously sized small RNA to the indicated features across the Arabidopsis genome. Data are from the means from six libraries.

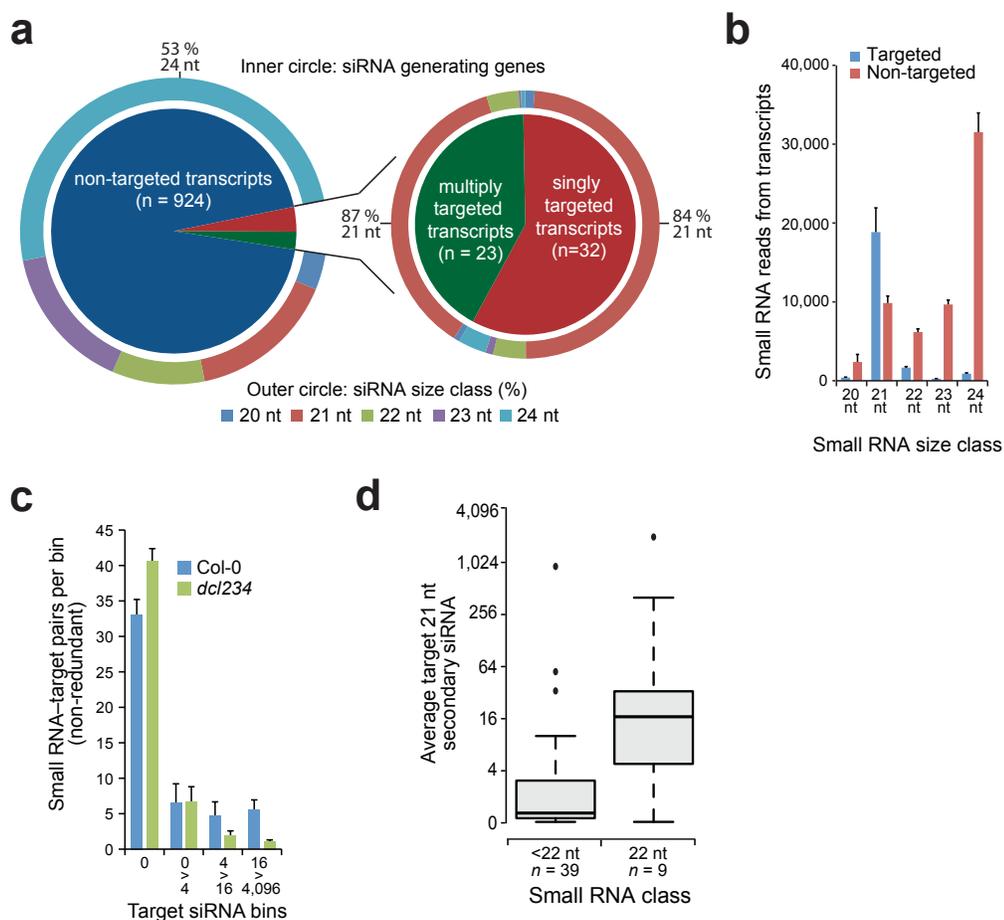


Figure 4.2. Small RNA from Arabidopsis annotated transcripts.

(a) Proportion of small RNA-generating transcripts that are targeted (at single or multiple sites) by miRNA or tasiRNA, or non-targeted. Outer rings show proportion of small RNA size. (b) Mean 20-24-nucleotide siRNA levels from targeted or non-targeted transcripts. (c) Non-redundant small RNA-target transcript pairs yielding four levels (bins) of 21-nucleotide siRNA in Col-0 and *dcl2-1 dcl3-1 dcl4-2 (dcl234)* mutant plants. Data are from averages of six (Col-0) and five (*dcl234*) replicates. (d) Box plots showing the mean numbers of 21-nucleotide siRNA originating from non-redundant transcripts targeted by small RNA that are 22-nucleotide or less than 22-nucleotide in length.

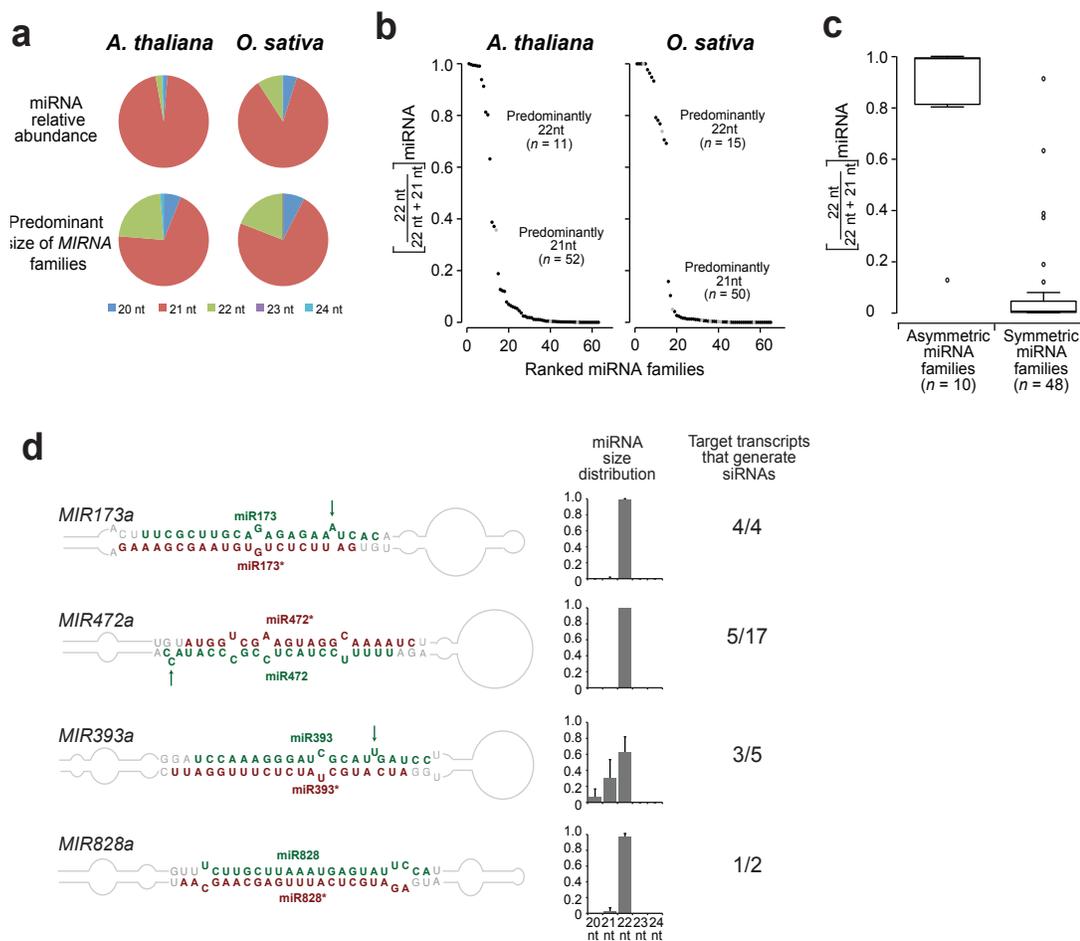


Figure 4.3. *MIRNA* foldback asymmetry leads to formation of 22-nucleotide miRNAs.

(a) Mean proportions of distinct miRNA size classes in read datasets (top), and of predominant size class for *MIRNA* families (bottom), from Arabidopsis and rice. The rice miRNA were from a filtered subset that passed basic criteria for bona fide miRNA⁵⁶ (see Supplemental Table 4). (b) Rank order showing proportion of 22-nucleotide size class, from averages of sequencing datasets, corresponding to non-redundant *MIRNA* loci in Arabidopsis and rice. Multigene *MIRNA* families with loci encoding the identical miRNA, but that have both symmetric and asymmetric foldbacks, are color coded grey. (c) Proportion of 22-nucleotide miRNA from non-redundant *MIRNA* loci with base-pair asymmetry or symmetry within the miRNA/miRNA* segment of the foldback. (d) Examples of 22-nucleotide miRNA-generating *MIRNA* foldbacks, average miRNA size distribution, and proportion of target transcripts that yield 21-nucleotide siRNA (at least four reads from 4/6 replicate libraries). Green arrows indicate the predicted asymmetric position within the foldbacks.

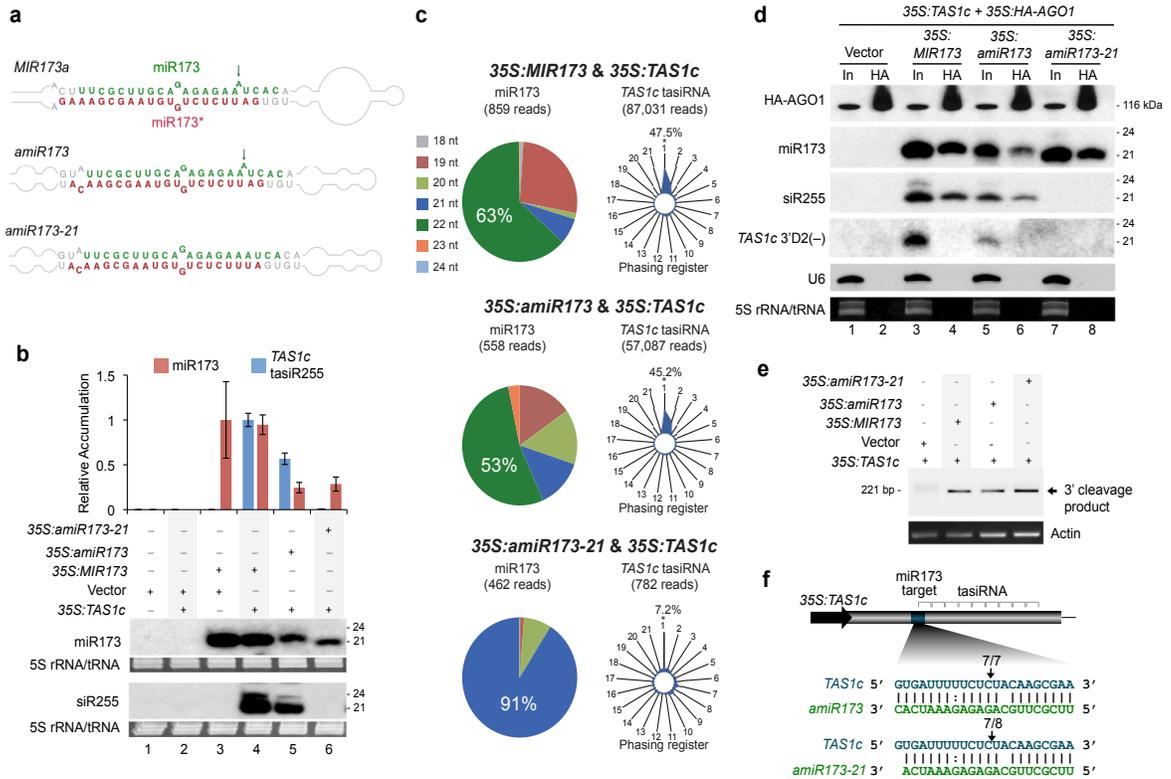
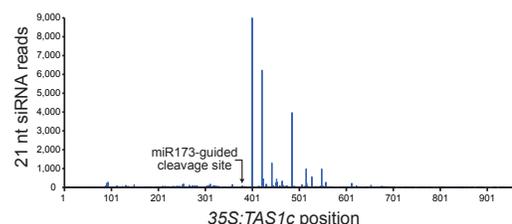
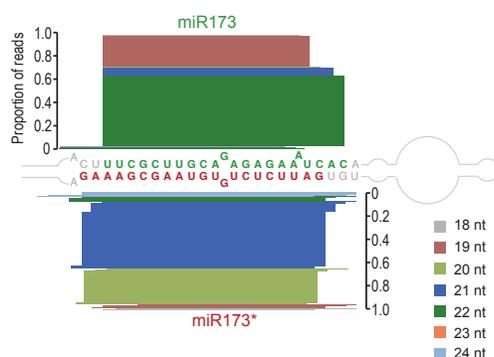


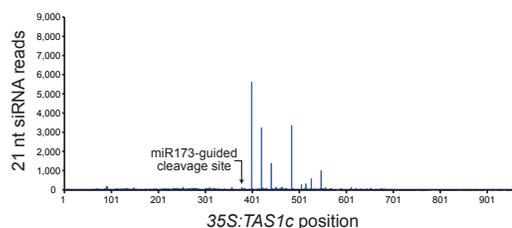
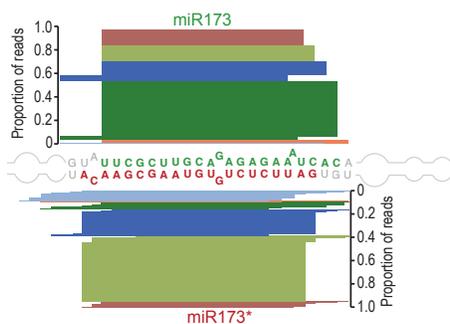
Figure 4.4. Production and activities of 21 and 22-nucleotide miR173 forms.

(a) Foldbacks of from wild-type *MIR173a*, *amiR173* and *amiR173-21*. Artificial miRNAs were engineered within the *MIR390a* foldback. miRNA guide and miRNA* strands are represented with green and red, respectively. The arrows indicate the predicted asymmetric position in *MIR173* and *amiR173* foldbacks. (b) Accumulation of miR173 and *TAS1c* tasiRNA (siR255) in *N. benthamiana* transient assays. Constructs were coexpressed as indicated above the blot panels. Mean ($n=3$) relative miR173 (red) and siR255 (blue) levels \pm SD (lane 2 and lane 3 = 1.0 for miR173 and tasiRNA255 respectively) were plotted (top). One of three biological replicates of the blot data, and EtBr-stained rRNA as loading controls, are shown (bottom). (c) Analysis of miR173 (from *35S:MIR173*, *35S:amiR173* and *35S:amiR173-21*) and *TAS1c*-derived siRNA sequences by high-throughput sequencing after transient assays in *N. benthamiana*. Pie charts display the percentage of 18-24-nucleotide reads. Radar plots display percentages of 21-nucleotide reads corresponding to each of the 21 registers from *TAS1c* transcripts, with position 1 designated as immediately after the miR173-guided cleavage site. (d) Analysis of co-immunoprecipitation of 21-nucleotide and 22-nucleotide miR173 variants with HA-AGO1. Protein and RNA blot assays using input (in) and IP (HA) fractions from *N. benthamiana* following coexpression of *35S:HA-AGO1* and *35S:TAS1c* with *35S:MIR173*, *35S:amiR173* and *35S:amiR173-21*. The *TAS1c* 3'D2(-) panel shows an HA-AGO1-nonassociated tasiRNA generated from the *TAS1c* transcript as an IP control. U6 RNA and EtBr-stained rRNA were included as input loading and HA-AGO1-non-associated controls. (e) EtBr-stained 5' RACE products corresponding to the 3' cleavage product from miR173-guided cleavage. *N. benthamiana* actin RT-PCR products are shown as a control. (f) Proportion of cloned 5' RACE products corresponding to cleavage within *TAS1c* transcripts at the canonical miR173-guided site in assays with *amiR173* and *amiR173-21*.

a 35S:MIR173 & 35S:TAS1c-derived reads



b 35S:amiR173 & 35S:TAS1c-derived reads



c 35S:amiR173-21 & 35S:TAS1c-derived reads

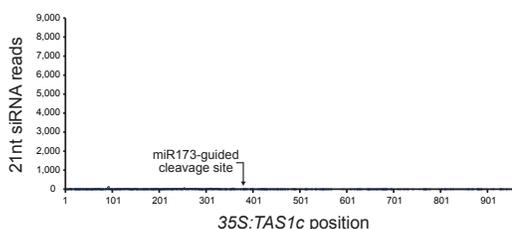
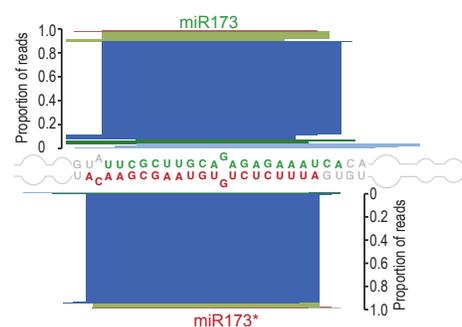


Figure 4.5 Mapping of miR173 or miR173-related reads, and TAS1c transcript-derived reads, from the transient expression assay data shown in Figure 3c.

(a,b,c) Analysis of miR173, miR173* and TAS1c reads after transient co-expression of 35S:MIR173a, 35S:amiR173 or 35S:amiR173-21 with 35S:TAS1c. Proportions of reads are plotted as stacked bars based on size (color coded), 5' position and 3' position, with end positions aligned to the respective sequences shown in the foldbacks. miR173 and related sequences are plotted upwards, and miR173* sequences are plotted downwards. Twenty-one-nucleotide TAS1c siRNA reads are plotted in the bar graphs along the transcript length, with the miR173-guided cleavage site indicated. The bar graphs show consolidated "sense" and "antisense" reads, in which the siRNAs from the two strands were normalized by 2-nucleotides to account for the known offset in duplexes resulting from DCL4-mediated cleavage.

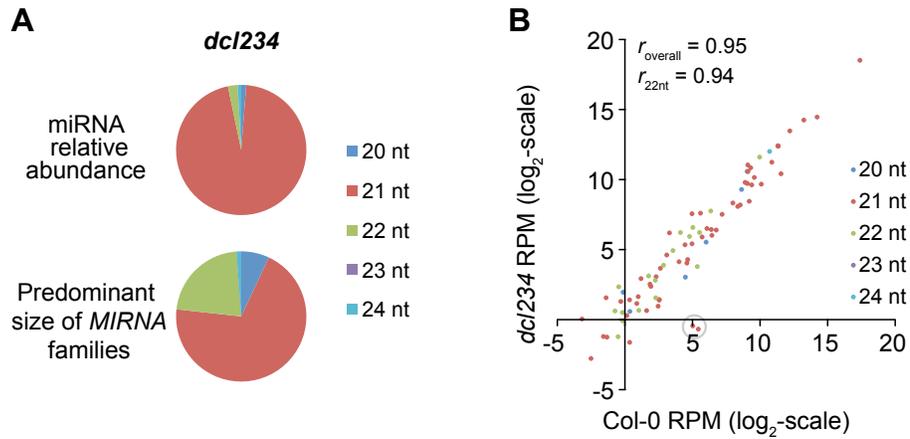


Figure 4.6 miRNA accumulation in the Arabidopsis *dcl234* triple mutant.

(a) Mean proportions of distinct miRNA size classes in read datasets (top), and of predominant size class for *MIRNA* families (bottom), from Arabidopsis *dcl234* triple mutants. (b) Scatter plot comparing average miRNA reads in wild-type Arabidopsis (Col-0) and *dcl234* triple mutants (six and five replicates, respectively). Reads were normalized per million (RPM). The Pearson product-moment correlation coefficients (r) for the overall and 22-nucleotide sets of miRNA families are shown. DCL4-dependent miR822 and miR839 are indicated (grey circle) (Rajagopalan et al., 2006).

Figure 4.7 Enrichment or depletion of miRNA in HA-AGO1 coimmunoprecipitates.

Mature miRNA from total RNA (input fraction) and immunoprecipitated HA-AGO1 complexes (IP fraction) were identified by high-throughput sequencing (two replicates each). Enrichment or depletion of miRNA in the IP fraction was determined by plotting the average ratio of miRNA reads/million (RPM) in the IP samples divided by the miRNA reads in the input sample. Only miRNA with an average of at least 5 RPM in either the IP or input samples were plotted. The predominant size class for most groups was plotted. For miR397a and miR167a,b,d, two size classes are plotted, as 22-nucleotide variants represented more than 30% of the 21 and 22-nucleotide reads for these families (percentages of 21 and 22-nucleotide reads are indicated as in Figure 2b).

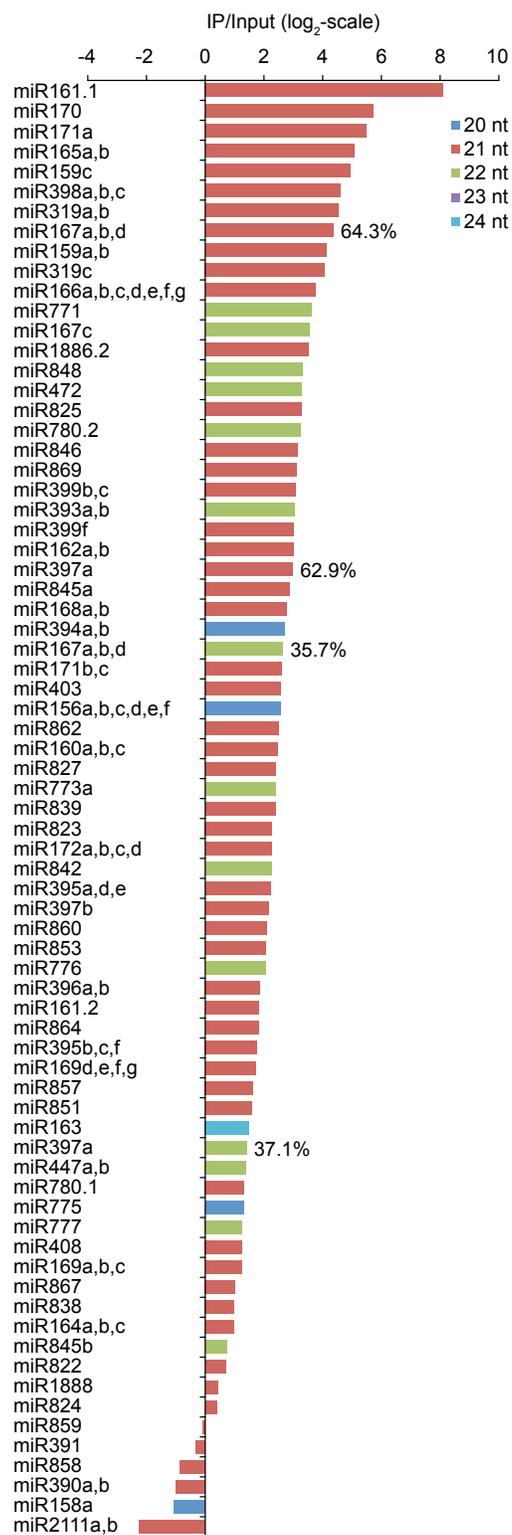


Figure 4.7 Enrichment or depletion of miRNA in HA-AGO1 coimmunoprecipitates.

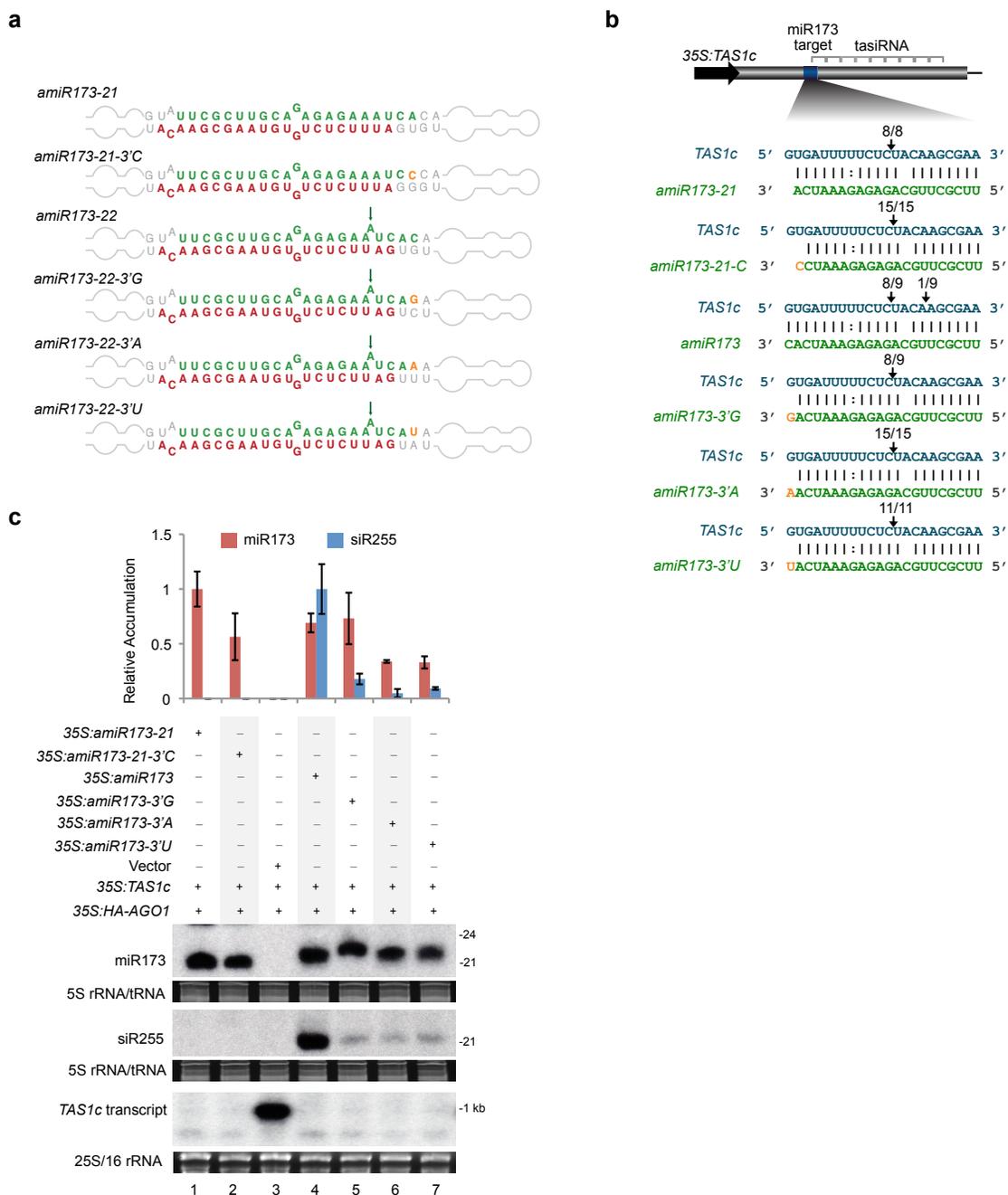


Figure 4.8. Role of the 3' end nucleotide in miR173.

(a) Foldbacks yielding 21-nucleotide or 22-nucleotide miR173 variants with distinct 3' end nucleotides. Mutagenized positions are shown in orange. (b) 5'RACE assays showing proportion of products mapping to the canonical cleavage site for each miR173 variant tested. (c) Accumulation of miR173 variants and *TAS1c* tasiRNA (siR255) in *N. benthamiana* co-expression assays with the constructs indicated above the blot panels. Mean (n=3) relative miR173 variant (red) and siR255 (blue) levels +/- SD (lane 1 and lane 4 = 1.0 for miR173 variant and siR255, respectively). Based on RNA migration, amiR173-22-3'G may be 23-nucleotides in size rather than 22-nucleotides.

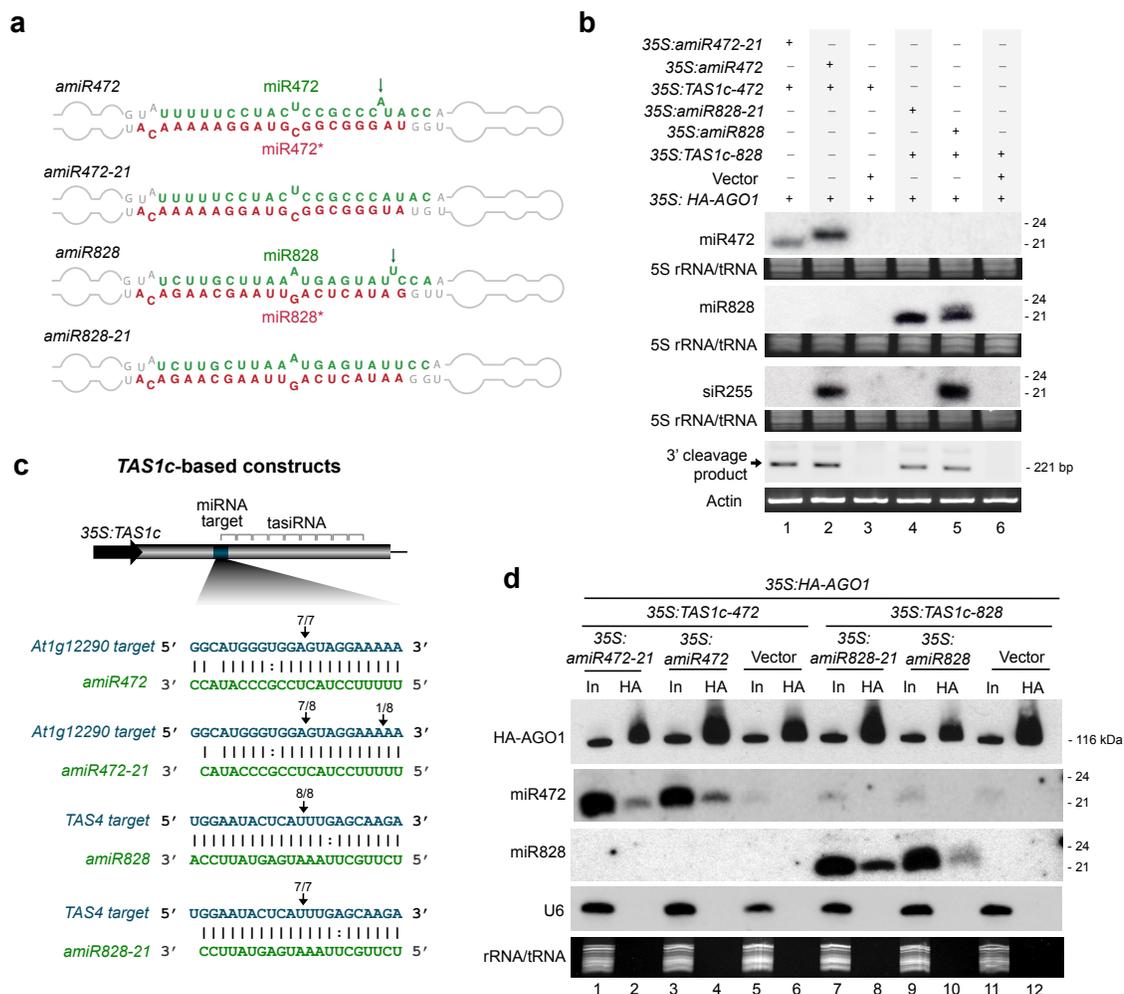


Figure 4.9. Production and activities of 21 and 22-nucleotide miR472 and miR828 forms. (a) Foldbacks of *amiR828*, *amiR828-21*, *amiR472*, and *amiR472-21*. (b) Accumulation of miR472, miR828 and modified *TAS1c* tasiRNA (siR255), and 5'RACE to detect miRNA-guided cleavage products of the modified *TAS1c* transcripts, in *N. benthamiana* transient assays. (c) Proportion of cloned 5' RACE products corresponding to cleavage within modified *TAS1c* transcripts at the canonical miR472- or miR828-guided sites in assays with the designated artificial miRNAs. The target site sequences are actual sites from *At1g12290* and *TAS4* transcripts, which are recognized by miR472 and miR828, respectively. (d) Analysis of co-immunoprecipitation of 21-nucleotide and 22-nucleotide amiR472 and amiR828 variants with HA-AGO1. Protein and RNA assays for input (in) and IP (HA) fractions from *N. benthamiana* expressing the amiR472 and amiR828 variants were done using blots containing samples from both sets of experiments. U6 RNA and EtBr-stained rRNA were included as input loading and HA-AGO1-non-associated controls.

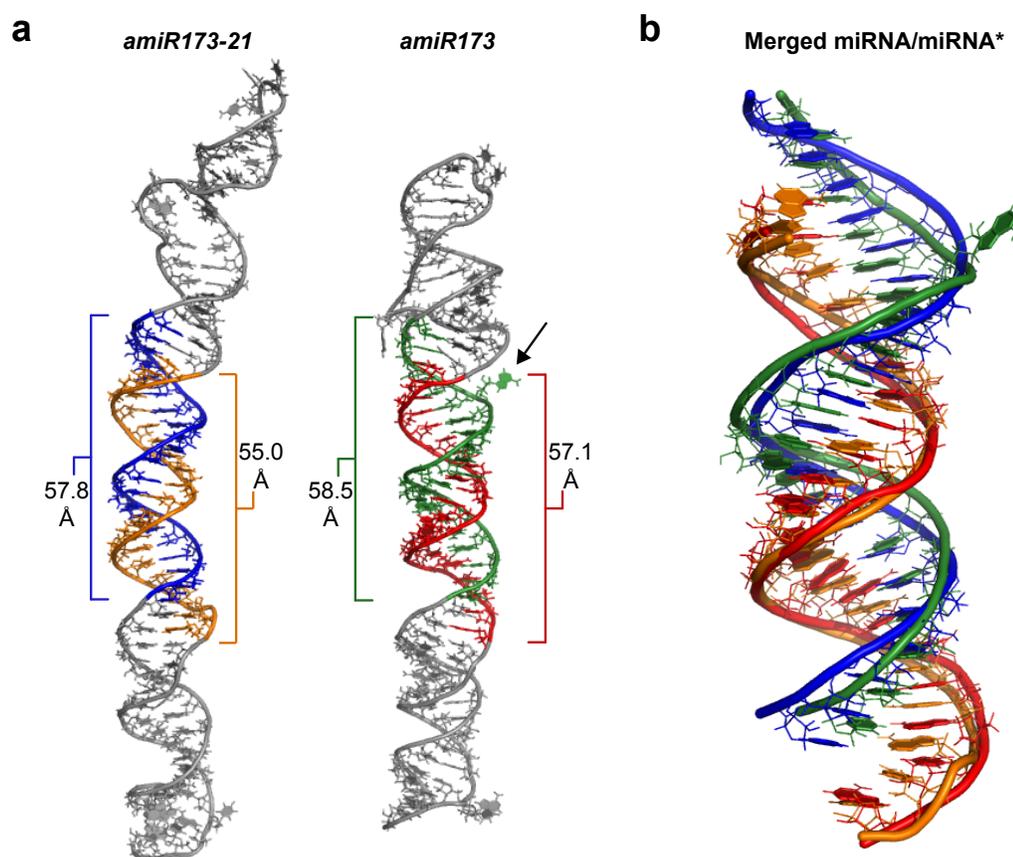


Figure 4.10. Predicted tertiary structures of MIR390a-derived amiR173-21 and amiR173 foldbacks using MC-Fold and MC-Sym.

(a) Structures of foldbacks were predicted using MC-Fold and MC-Sym7. The amiR173-21 guide and star strands are highlighted blue and orange, respectively. The amiR173 guide and star strands are highlighted green and red, respectively. The position of the nucleotide of the asymmetric bulge is indicated with an arrow. The lengths (in angstroms) of the guide and star strands are shown. (b) Overlay of miRNA/miRNA* duplexes of amiR173-21 and amiR173 show similarity in calculated duplex lengths.

SUPPLEMENTAL DATA

The following materials are available in the online version of this article

(<http://www.nature.com/nsmb/index.html>).

Supplementary Table 4.1. Small RNA reads from Arabidopsis transcripts.

Supplementary Table 4.2. Transcripts targeted by miRNA or tasiRNA.

Supplementary Table 4.3. Reads from Arabidopsis *MIRNA* loci.

Supplementary Table 4.4. Reads from rice filtered *MIRNA* loci.

Supplementary Table 4.5. Small RNA reads mapping to *MIR173* precursors from *Nicotiana benthamiana* leaves coexpressing *35S:TAS1c* together with *35S:MIR173*, *35S:amiR173* or *35S:amiR173-21*.

Supplementary Table 4.6. Small RNA reads mapping to *TAS1c* from *Nicotiana benthamiana* leaves coexpressing *35S:TAS1c* together with *35S:MIR173*, *35S:amiR173* or *35S:amiR173-21*.

General Conclusions

Josh T. Cuperus, Noah Fahlgren and James C. Carrington

The Plant Cell
American Society of Plant Biologists
15501 Monona Drive
Rockville, Maryland 20855-2768
PMID: 21317375

This thesis work describes a novel high-throughput sequencing-based method to identify putative causal mutations from a mapping population in *Arabidopsis* (Chapter 2), characterization of a mutation in the *MIR390a* foldback resulting in a miR390 processing defective mutant (Chapter 3), and identification of a 22-nucleotide-dependent miRNA-directed siRNA-producing mechanism in plants (Chapter 4). This work adds significantly to our understanding of small RNA biology and will be useful in identifying causal mutations using high-throughput sequencing.

Forward genetics and identification of causative mutations

Forward genetics has and will continue to be an instrumental tool in plant science. The advent of a high-throughput method to quickly, and simultaneously, map and identify causal mutations will drastically decrease the time associated with identifying causal mutations. Although forward genetic screens are useful, not all experimental designs yield useful screens. The inability of the *TAS1cPDS-2*-based mutagenesis screen to identify mutants demonstrates the importance of developing mutagenesis screens with less severe, consistent visual phenotypes. The *TAS1cPDS-2* parental line was unable to grow on soil and had widespread variable photobleaching, resulting in a high-proportion of plants that escaped photobleaching by transgene silencing or other means, which occasionally grew on soil. The problem of identifying mutants was compounded because putative mutants were not expected to have developmental phenotypes, making it difficult to separate silenced or reduced photobleaching plants from mutants. In contrast, the *TAS3aPDS-1*-based mutagenesis screen had both distinct photobleaching in leaf vasculature and a distinct accelerated vegetative phase-change (AVPC) developmental phenotype, which resulted in recovery of both novel and expected mutants. More severe phenotypes, like those in the putative *TAS1cPDS-2* screen mutants, can have complications due to infertility, therefore making propagation of the mutants difficult.

The pipeline Mapping and Assembly with Short Sequences (MASS) was successfully used in two independent mutant screens, using different numbers of individuals (92 for *mir390a-1* and 38 for *nripd/e-19*), and different depths of sequencing coverage (12X for *mir390a-1* and 21X for *nripd/e-19*) demonstrating the flexibility of the tool for both variables. The pipeline is publicly available, and has the potential to be used in alternative configurations including other *Arabidopsis* ecotypes, or other organisms. In theory, a mapping population may not be necessary for this type of analysis. Alternatively, enrichment of mutations near the causal mutation, relative to background mutagenesis-derived single nucleotide changes amongst mutagenized individuals, could be identified from bulk segregant populations of mutants homozygous for the causal mutation crossed to the parental strain instead of a mapping strain. As sequencing costs continue to decrease, direct sequencing of mutants could replace traditional complementation crossing techniques in *Arabidopsis*. This may be especially true in screens for components in the RdDM pathway, where more than 20 factors are known to influence DNA methylation at target loci.

Processing of miRNA in plants

The recovered mutation, *mir390a-1*, appeared to have only a mild effect on the structure or free energy of the *MIR390a* foldback, however the Guanine-to-Adenine mutation leads to a drastic reduction and inaccurate processing of mature miR390. The change from a weak, or unpaired U-G, to a more stable U-A base-pair, without other significant structural changes predicted, lead to the deficiency of processing from *mir390a-1*. Recently, other mutagenesis-based studies have discovered that mutations in a similar region, below the miR/miR* of multiple miRNAs, result in similar reductions in mature miRNA (Mateos et al., 2010; Song et al., 2010; Werner et al., 2010). While the base of *MIRNA* share little nucleotide homology, the majority of these mutations were also predicted to alter the predicted base-pairing. Interestingly, a second cis element required for proper miRNA processing, 15 base-pairs below the miR/miR* (See Figure 1.1C), is approximately one helical turn from the cis region

identified in this work (~11 nucleotides), suggesting it may act as a region of contact with an RNA binding protein such as HYL1 or DDL. The heterogeneity of, and the redundancy between, miRNA families makes the identification of additional cis elements difficult. In future work it will be critical to sequence the RNA bound by known miRNA processing factors like DCL1, HYL1, DDL and SE to identify the regions of *MIRNA* foldbacks that these factors associate with.

Does miRNA-processing heterogeneity, or differential processing activities during development, represent a biologically relevant regulatory mechanism? Differential processing of miRNA is well established in animals, but few examples have been demonstrated in plants (Newman and Hammond). Maize *zma-MIR166a* transcripts accumulate in the tip of the shoot apical meristem, but no mature miR166 was detected by in situ hybridization (Nogueira et al., 2009). *A. thaliana ath-MIR172b* transcripts were detected in wild type inflorescence tissues, but not in seedlings, even though transcripts were detected in both tissues in *dcl1* mutant plants (Laubinger et al., 2010). These examples suggest that processing efficiency can be modulated in a tissue- or precursor-specific manner, as was observed for some animal *MIRNA* (Michlewski et al., 2008). Efficiency could be based on the relative affinity of DCL1, SE, HYL1 and DDL for each pri-miRNA, but processing rates would be dependent on the availability of these proteins in each cell. Sequestration of one or more processing factors (e.g. titration of HYL1 by short interspersed elements [SINE] RNAs (Pouch-Pelissier et al., 2008)) would disfavor processing of pri-miRNA with low affinity for these proteins. Further work is needed to understand the prevalence of differential or allele-specific processing and how this would be regulated.

miRNA-trigger siRNA

Some miRNA possess specialized features that trigger the production of secondary siRNA from protein-coding transcripts. What is the biological role of these siRNAs? Secondary siRNA might reinforce silencing of transcripts from the miRNA target locus, or conceivably expand

the range of target repression to related family members that interact with the siRNAs. For example, targeting of members of the pentatricopeptide repeat (*PPR*) family by several miRNA and tasiRNA in *A. thaliana* triggers the production of secondary siRNA with the theoretical potential to target other *PPR* transcripts in trans, although compelling evidence for the activity of the secondary siRNA is lacking (Axtell et al., 2006; Chen et al., 2007; Howell et al., 2007; Addo-Quaye et al., 2008). Another fascinating possibility is that secondary siRNA might function as non-cell autonomous signals to form regulatory gradients, or to silence transcripts or loci at distal sites (Chitwood et al., 2009; de Felippes et al., 2010; Dunoyer et al., 2010; Molnar et al., 2010). Furthermore, DCL2-dependent siRNAs are important for repression virus accumulation at long distances in Arabidopsis (Garcia-Ruiz et al., 2010). Long distance travel and targeting of RNA by 22-nucleotide small RNA could act as an efficient method to not just degrade target RNA, but create a local population of 21-nucleotide small RNAs.

Interestingly, the distinction of a single nucleotide (21 nucleotides versus 22 nucleotides) in many plant miRNA is enough to alter function, and when 22-nucleotides lead to secondary siRNA production of target RNA instead of repression, the main consequence of 21-nucleotide miRNA targeting. Here, we showed that the majority of 22-nucleotide miRNA-generating loci have predicted asymmetrically bulged nucleotides. We, and others, demonstrated that 21-nucleotide miRNA can be converted to 22-nucleotide miRNA by the addition of a single nucleotide in the miRNA region of a precursor transcript (Cuperus et al., 2010b; Molnar et al., 2010). This suggests that a single nucleotide insertion or deletion event could alter the size, and potentially the function of a miRNA. While it may be evolutionarily advantageous for some miRNA to trigger siRNA production and further repress target RNA, this kind of silencing could be detrimental if it spread uncontrolled, and therefore may be well regulated or generally selected against. As more 22-nucleotide generating loci in additional plant species are identified, and the functional roles are assessed, more studies involving the evolution and regulation of 22-nucleotide miRNA will be possible. Additionally, the introduction of 22-

nucleotide artificial miRNAs that trigger secondary siRNAs may also be useful in future studies and biotechnology.

The ability of AGO1-22-nucleotide small RNA complexes, but not AGO1-21-nucleotide small RNA complexes to trigger the formation of RDR6-dependent siRNAs is novel characteristic of AGO1 as no other plant AGO-complexes have been shown to have two distinct functions. Structural analysis of different AGO1 complexes coupled with in vitro assays and forward genetics will help identify the important residues or conformational states of AGO1 that recruit either RNA degradation components or amplification components to target RNAs.

Finally, additional diversified functions of miRNAs will likely emerge from studies of diverse plants species. In depth analysis of rice and Arabidopsis species have already revealed distinct flavors of miRNAs, effector pathways and cross-talk between silencing pathways. Recently, 22-nucleotide miRNA-triggered 24-nucleotide siRNAs have been identified in monocots. The mechanism that leads to the differential formation of 24- instead of 21-nucleotide secondary siRNAs in rice has yet to be identified and undoubtedly will be a future direction of research. As additional plant lineages are explored, novel innovations involving miRNAs will likely become apparent. The rapidly growing collection of plant genomes, and advances in high-throughput small RNA sequencing has enhanced our understanding of the evolution of *MIRNA* genes and their functions. High-throughput sequencing will continue to be instrumental in our understanding of small RNA pathways and facilitate research in additional plants. It is becoming evident that understanding the mechanisms of small RNA biogenesis is fundamental to our understanding of developmental processes, genome stability, and response to both biotic and abiotic stresses. The majority of the work described here focuses on miRNA processing and diversification in Arabidopsis, however, these finding will likely have implications outside of Arabidopsis as well.

Bibliography

- Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J.** (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr Biol* **18**, 758-762.
- Addo-Quaye, C., Snyder, J.A., Park, Y.B., Li, Y.F., Sunkar, R., and Axtell, M.J.** (2009). Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA* **15**, 2112-2121.
- Adenot, X., Elmayan, T., Laressergues, D., Boutet, S., Bouche, N., Gascioli, V., and Vaucheret, H.** (2006). DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16**, 927-932.
- Allen, E., and Howell, M.D.** (2010). miRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Semin Cell Dev Biol* **21**, 798-804.
- 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.
- Allen, R.S., Li, J., Stahle, M.I., Dubroue, A., Gubler, F., and Millar, A.A.** (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. *Proc Natl Acad Sci U S A* **104**, 16371-16376.
- 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.** (2008). Evolution of microRNAs and their targets: are all microRNAs biologically relevant? *Biochim Biophys Acta* **1779**, 725-734.
- Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P.** (2006). A Two-Hit Trigger for siRNA Biogenesis in Plants. *Cell* **127**, 565-577.
- Baumberger, N., and Baulcombe, D.C.** (2005). *Arabidopsis* ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci USA* **102**, 11928-11933.
- Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T.** (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep* **10**, 649-654.
- Blevins, T., Rajeswaran, R., Shivaprasad, P.V., Beknazariants, D., Si-Ammour, A., Park, H.S., Vazquez, F., Robertson, D., Meins, F., Jr., Hohn, T., and Pooggin, M.M.**

- (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res* **34**, 6233-6246.
- Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S.** (2003). HASTY, the *Arabidopsis* ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* **130**, 1493-1504.
- Bologna, N.G., Mateos, J.L., Bresso, E.G., and Palatnik, J.F.** (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J* **28**, 3646-3656.
- Bonhoeffer, S., McCaskill, J.S., Stadler, P.F., and Schuster, P.** (1993). RNA multi-structure landscapes. A study based on temperature dependent partition functions. *Eur Biophys J* **22**, 13-24.
- Bouché, N., Laressergues, D., Gascioli, 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.
- Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O.** (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185-1190.
- Cao, X., and Jacobsen, S.E.** (2002). Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**, 1138-1144.
- Carthew, R.W., and Sontheimer, E.J.** (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-655.
- Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E.** (2004). RNA silencing genes control de novo DNA methylation. *Science* **303**, 1336.
- Chapman, E.J., and Carrington, J.C.** (2007). Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* **8**, 884-896.
- Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W., and Jin, H.** (2010). siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Res* **38**, 6883-6894.
- Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J.** (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584-589.
- Chen, H.M., Li, Y.H., and Wu, S.H.** (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*. *Proc Natl Acad Sci USA* **104**, 3318-3323.

- Chen, H.M., Chen, L.T., Patel, K., Li, Y.H., Baulcombe, D.C., and Wu, S.H.** (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A* **107**, 15269-15274.
- Chen, X.** (2004). A 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.
- Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R.** (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-744.
- Chiba, Y., and Green, P.J.** (2009). mRNA degradation machinery in plants. *J. Plant Biol.* **52**, 114-124.
- Chitwood, D.H., Nogueira, F.T., Howell, M.D., Montgomery, T.A., Carrington, J.C., and Timmermans, M.C.** (2009). Pattern formation via small RNA mobility. *Genes Dev* **23**, 549-554.
- Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., Wolfe, S.A., and Giraldez, A.J.** (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694-1698.
- Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H., Frazer, K.A., Huson, D.H., Scholkopf, B., Nordborg, M., Ratsch, G., Ecker, J.R., and Weigel, D.** (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338-342.
- Clarke, J.H., Tack, D., Findlay, K., Van Montagu, M., and Van Lijsebettens, M.** (1999). The SERRATE locus controls the formation of the early juvenile leaves and phase length in *Arabidopsis*. *Plant J* **20**, 493-501.
- Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F., Conte, D., Jr., and Mello, C.C.** (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**, 123-134.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743.
- Correa, R.L., Steiner, F.A., Berezikov, E., and Ketting, R.F.** (2010). MicroRNA-directed siRNA biogenesis in *Caenorhabditis elegans*. *PLoS Genet* **6**, e1000903.

- Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M., and Carrington, J.C.** (2010a). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. *Proc Natl Acad Sci U S A* **107**, 466-471.
- Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C.** (2010b). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* **17**, 997-1003.
- Curtis, M.D., and Grossniklaus, U.** (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**, 462-469.
- de Felippes, F.F., Ott, F., and Weigel, D.** (2010). Comparative analysis of non-autonomous effects of tasiRNAs and miRNAs in Arabidopsis thaliana. *Nucleic Acids Res.*
- 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.
- Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J.** (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-235.
- Diaz-Pendon, J.A., Li, F., Li, W.X., and Ding, S.W.** (2007). Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**, 2053-2063.
- Dong, Z., Han, M.H., and Fedoroff, N.** (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A* **105**, 9970-9975.
- Dugas, D.V., and Bartel, B.** (2008). Sucrose induction of Arabidopsis miR398 represses two Cu/Zn superoxide dismutases. *Plant Mol Biol* **67**, 403-417.
- Dunoyer, P., Himber, C., and Voinnet, O.** (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.
- Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J.C., and Voinnet, O.** (2010). Small RNA duplexes function as mobile silencing signals between plant cells. *Science* **328**, 912-916.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T.** (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**, 2539-2544.

- Eulalio, A., Tritchler, F., and Izaurralde, E.** (2009). The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *RNA* **15**, 1433-1442.
- 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.
- Fahlgren, N., Sullivan, C.M., Kasschau, K.D., Chapman, E.J., Cumbie, J.S., Montgomery, T.A., Gilbert, S.D., Dasenko, M., Backman, T.W., Givan, S.A., and Carrington, J.C.** (2009). Computational and analytical framework for small RNA profiling by high-throughput sequencing. *Rna* **15**, 992-1002.
- Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C.** (2010). MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* **22**, 1074-1089.
- Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D.** (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* **3**, e236.
- Fusaro, A.F., Matthew, L., Smith, N.A., Curtin, S.J., Dedic-Hagan, J., Ellacott, G.A., Watson, J.M., Wang, M.B., Brosnan, C., Carroll, B.J., and Waterhouse, P.M.** (2006). RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep* **7**, 1168-1175.
- Gandikota, M., Birkenbihl, R.P., Hohmann, S., Cardon, G.H., Saedler, H., and Huijser, P.** (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J* **49**, 683-693.
- 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.
- Garcia-Ruiz, H., Takeda, A., Chapman, E.J., Sullivan, C.M., Fahlgren, N., Brempelis, K.J., and Carrington, J.C.** (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *Plant Cell* **22**, 481-496.
- Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H.** (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* **15**, 1494-1500.

- Ghildiyal, M., and Zamore, P.D.** (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94-108.
- Golden, T.A., Schauer, S.E., Lang, J.D., Pien, S., Mushegian, A.R., Grossniklaus, U., Meinke, D.W., and Ray, A.** (2002). Short integuments1/suspensor1/carpel factory, a Dicer homolog, is a maternal effect gene required for embryo development in Arabidopsis. *Plant Physiol.* **130**, 808-822.
- Gregory, B.D., O'Malley, R.C., Lister, R., Urich, M.A., Tonti-Filippini, J., Chen, H., Millar, A.H., and Ecker, J.R.** (2008). A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev Cell* **14**, 854-866.
- Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J.** (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-158.
- Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., Chen, C.C., Chaves, D.A., Duan, S., Kasschau, K.D., Fahlgren, N., Yates, J.R., 3rd, Mitani, S., Carrington, J.C., and Mello, C.C.** (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* **36**, 231-244.
- Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S.** (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**, 537-541.
- Guilfoyle, T.J., and Hagen, G.** (2007). Auxin response factors. *Curr Opin Plant Biol* **10**, 453-460.
- Haase, A.D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gagnol, A., and Filipowicz, W.** (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* **6**, 961-967.
- Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N.** (2004a). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-3027.
- Han, M.H., Goud, S., Song, L., and Fedoroff, N.** (2004b). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci USA* **101**, 1093-1098.
- Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H., and Baulcombe, D.C.** (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**, 321-334.
- Hazen, S.P., Naef, F., Quisel, T., Gendron, J.M., Chen, H., Ecker, J.R., Borevitz, J.O., and Kay, S.A.** (2009). Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol* **10**, R17.

- Hazen, S.P., Borevitz, J.O., Harmon, F.G., Pruneda-Paz, J.L., Schultz, T.F., Yanovsky, M.J., Liljegren, S.J., Ecker, J.R., and Kay, S.A.** (2005). Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol* **138**, 990-997.
- He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., and Deng, X.W.** (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**, 17-33.
- He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K.** (2009). An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**, 498-508.
- Heisel, S.E., Zhang, Y., Allen, E., Guo, L., Reynolds, T.L., Yang, X., Kovalic, D., and Roberts, J.K.** (2008). Characterization of unique small RNA populations from rice grain. *PLoS One* **3**, e2871.
- Henderson, I.R., and Jacobsen, S.E.** (2007). Epigenetic inheritance in plants. *Nature* **447**, 418-424.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C.** (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118-120.
- Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., Koiwa, H., Seki, M., Shinozaki, K., and Fukuhara, T.** (2005). Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Mol Biol* **57**, 173-188.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R.** (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.
- Hofacker, I.L.** (2003). Vienna RNA secondary structure server. *Nucleic Acids Research* **31**, 3429-3431.
- Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C.** (2007). Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**, 926-942.
- Hunter, C., Sun, H., and Poethig, R.S.** (2003). The *Arabidopsis* heterochronic gene *ZIPPY* is an ARGONAUTE family member. *Curr Biol* **13**, 1734-1739.
- International Brachypodium Initiative.** (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763-768.

- 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.
- Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q.** (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev* **19**, 1674-1679.
- Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.L., Walbot, V., Sundaresan, V., Vance, V., and Bowman, L.H.** (2009). Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res* **19**, 1429-1440.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J.** (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**, 761-765.
- Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M., and Matzke, A.J.** (2008). A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet* **40**, 670-675.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C.** (2007). Genome-Wide Profiling and Analysis of *Arabidopsis* siRNAs. *PLoS Biol* **5**, e57.
- Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R., and Frank, W.** (2010). Transcriptional control of gene expression by microRNAs. *Cell* **140**, 111-122.
- Kim, V.N., Han, J., and Siomi, M.C.** (2009). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**, 126-139.
- Konieczny, A., and Ausubel, F.M.** (1993). A procedure for mapping *Arabidopsis* mutations using a co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403-410.
- Kumakura, N., Takeda, A., Fujioka, Y., Motose, H., Takano, R., and Watanabe, Y.** (2009). SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. *FEBS Lett* **583**, 1261-1266.
- Kurihara, Y., Takashi, Y., and Watanabe, Y.** (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* **12**, 206-212.
- Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G., and Weigel, D.** (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **105**, 8795-8800.

- Laubinger, S., Zeller, G., Henz, S.R., Buechel, S., Sachsenberg, T., Wang, J.W., Ratsch, G., and Weigel, D.** (2010). Inaugural Article: Global effects of the small RNA biogenesis machinery on the *Arabidopsis thaliana* transcriptome. *Proc Natl Acad Sci U S A* **107**, 17466-17473.
- Law, J.A., and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**, 204-220.
- 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.
- Levine, E., Zhang, Z., Kuhlman, T., and Hwa, T.** (2007). Quantitative characteristics of gene regulation by small RNA. *PLoS Biol* **5**, e229.
- Li, H., Ruan, J., and Durbin, R.** (2008a). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851-1858.
- Li, R., Li, Y., Kristiansen, K., and Wang, J.** (2008b). SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713-714.
- Lister, R., Gregory, B.D., and Ecker, J.R.** (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol* **12**, 107-118.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R.** (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523-536.
- Liu, Q., Yao, X., Pi, L., Wang, H., Cui, X., and Huang, H.** (2008). The ARGONAUTE10 gene modulates shoot apical meristem maintenance and leaf polarity establishment by repressing miR165/166 in *Arabidopsis*. *Plant J.*
- 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., Kulkarni, K., Souret, F.F., Muthuvallappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C.** (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res* **16**, 1276-1288.
- Lynn, K., Fernandez, A., Aida, M., Sedbrook, J., Tasaka, M., Masson, P., and Barton, M.K.** (1999). The PINHEAD/ZWILLE gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the ARGONAUTE1 gene. *Development* **126**, 469-481.

- MacRae, I.J., Zhou, K., and Doudna, J.A.** (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* **14**, 934-940.
- MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A.** (2006). Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195-198.
- Maindonald, J.H., and Braun, J.** (2007). Data analysis and graphics using R : an example-based approach. (Cambridge ; New York: Cambridge University Press).
- Makeyev, E.V., and Bamford, D.H.** (2002). Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol Cell* **10**, 1417-1427.
- Mallory, A.C., and Vaucheret, H.** (2009). ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO Rep* **10**, 521-526.
- Mallory, A.C., Elmayer, T., and Vaucheret, H.** (2008). MicroRNA maturation and action--the expanding roles of ARGONAUTES. *Curr Opin Plant Biol* **11**, 560-566.
- Malone, C.D., and Hannon, G.J.** (2009). Small RNAs as guardians of the genome. *Cell* **136**, 656-668.
- Mateos, J.L., Bologna, N.G., Chorostecki, U., and Palatnik, J.F.** (2010). Identification of MicroRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Curr Biol* **20**, 49-54.
- Matzke, M.A., Mette, M.F., and Matzke, A.J.** (2000). Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* **43**, 401-415.
- Megraw, M., Baev, V., Rusinov, V., Jensen, S.T., Kalantidis, K., and Hatzigeorgiou, A.G.** (2006). MicroRNA promoter element discovery in Arabidopsis. *RNA* **12**, 1612-1619.
- Meister, G., and Tuschl, T.** (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-349.
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y.** (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116-127.
- Michlewski, G., Guil, S., Semple, C.A., and Caceres, J.F.** (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* **32**, 383-393.

- Mlotshwa, S., Pruss, G.J., Peragine, A., Endres, M.W., Li, J., Chen, X., Poethig, R.S., Bowman, L.H., and Vance, V.** (2008). DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis. *PLoS One* **3**, e1755.
- Molnar, A., Melnyk, C.W., Bassett, A., Hardcastle, T.J., Dunn, R., and Baulcombe, D.C.** (2010). Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* **328**, 872-875.
- Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C.** (2008a). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133**, 128-141.
- Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H., and Carrington, J.C.** (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc Natl Acad Sci U S A* **105**, 20055-20062.
- Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D.** (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789-802.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T.** (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34-41.
- Newman, M.A., and Hammond, S.M.** Emerging paradigms of regulated microRNA processing. *Genes Dev* **24**, 1086-1092.
- Nogueira, F.T., Chitwood, D.H., Madi, S., Ohtsu, K., Schnable, P.S., Scanlon, M.J., and Timmermans, M.C.** (2009). Regulation of small RNA accumulation in the maize shoot apex. *PLoS Genet* **5**, e1000320.
- Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C.** (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89-100.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S.** (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613-622.
- Page, D.R., and Grossniklaus, U.** (2002). The art and design of genetic screens: Arabidopsis thaliana. *Nat Rev Genet* **3**, 124-136.
- Pak, J., and Fire, A.** (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241-244.

- Palatnik, J.F., Wollman, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J.C., Weigel, D.** (2007). Sequence and Expression Differences Underlie Functional Specialization of *Arabidopsis* MicroRNAs miR159 and miR319. *Dev Cell* **13**, 115-125.
- Parisien, M., and Major, F.** (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55.
- Park, W., Li, J., Song, R., Messing, J., and Chen, X.** (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* **12**, 1484-1495.
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S.** (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev* **18**, 2368-2379.
- Poethig, R.S.** (2003). Phase change and the regulation of developmental timing in plants. *Science* **301**, 334-336.
- Pouch-Pelissier, M.N., Pelissier, T., Elmayan, T., Vaucheret, H., Boko, D., Jantsch, M.F., and Deragon, J.M.** (2008). SINE RNA induces severe developmental defects in *Arabidopsis thaliana* and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet* **4**, e1000096.
- Qi, Y., Denli, A.M., and Hannon, G.J.** (2005). Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* **19**, 421-428.
- Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J.** (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008-1012.
- Qin, H., Chen, F., Huan, X., Machida, S., Song, J., and Yuan, Y.A.** Structure of the *Arabidopsis thaliana* DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA* **16**, 474-481.
- Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P.** (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* **20**, 3407-3425.
- Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P., and Theres, K.** (2008). Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in *Arabidopsis thaliana*. *Plant J* **55**, 65-76.
- 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.
- Ronemus, M., Vaughn, M.W., and Martienssen, R.A.** (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., Jan, C.H., and Bartel, D.P.** (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*.
- Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C.** (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol* **3**, e235.
- Sarin, S., Prabhu, S., O'Meara, M.M., Pe'er, I., and Hobert, O.** (2008). Caenorhabditis elegans mutant allele identification by whole-genome sequencing. *Nat Methods* **5**, 865-867.
- Sashital, D.G., and Doudna, J.A.** (2010). Structural insights into RNA interference. *Curr Opin Struct Biol* **20**, 90-97.
- Schneider, T.D.** (1997). Information content of individual genetic sequences. *J Theor Biol* **189**, 427-441.
- Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D.** (2006). Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* **18**, 1121-1133.
- Shannon, C.E.** (1948). A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423 and 623-656.
- Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J.L., and Meyerowitz, E.M.** (2007). Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness. *Development* **134**, 1051-1060.
- Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H.** (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**, 244-247.
- Sire, C., Moreno, A.B., Garcia-Chapa, M., Lopez-Moya, J.J., and San Segundo, B.** (2009). Diurnal oscillation in the accumulation of Arabidopsis microRNAs, miR167, miR168, miR171 and miR398. *FEBS Lett* **583**, 1039-1044.
- Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C.** (2007). An SNF2 Protein Associated with Nuclear RNA Silencing and the Spread of a Silencing Signal between Cells in Arabidopsis. *Plant Cell* **19**, 1507-1521.

- Song, L., Axtell, M.J., and Fedoroff, N.V.** (2010). RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr Biol* **20**, 37-41.
- Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.R., Mueller-Roeber, B., Jarmolowski, A., and Szweykowska-Kulinska, Z.** (2009). Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. *Nucleic Acids Res* **37**, 3083-3093.
- Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y.** (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol* **49**, 493-500.
- Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D.** A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. *PLoS Genet* **6**, e1001031.
- Tucker, M.R., Hinze, A., Tucker, E.J., Takada, S., Jurgens, G., and Laux, T.** (2008). Vascular signalling mediated by ZWILLE potentiates WUSCHEL function during shoot meristem stem cell development in the Arabidopsis embryo. *Development* **135**, 2839-2843.
- Vaucheret, H.** (2008). Plant ARGONAUTES. *Trends Plant Sci* **13**, 350-358.
- Vaucheret, H.** (2009). AGO1 homeostasis involves differential production of 21-nt and 22-nt miR168 species by MIR168a and MIR168b. *PLoS One* **4**, e6442.
- Vazquez, F., Gascioli, 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., Gascioli, 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.
- Voinnet, O.** (2008). Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci* **13**, 317-328.
- Voinnet, O.** (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**, 669-687.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J.** (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921-926.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J.** (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754-761.

- Wassenegger, M., and Krczal, G.** (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci* **11**, 142-151.
- Werner, S., Wollmann, H., Schneeberger, K., and Weigel, D.** (2010). Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr Biol* **20**, 42-48.
- Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S.** (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635-648.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S.** (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**, 630-634.
- 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, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y.** (2010). DNA methylation mediated by a microRNA pathway. *Mol Cell* **38**, 465-475.
- Wu, M.F., Tian, Q., and Reed, J.W.** (2006). *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133**, 4211-4218.
- Xie, Z., Allen, E., Wilken, A., and Carrington, J.C.** (2005a). 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.
- Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C.** (2005b). Expression of *Arabidopsis* MIRNA genes. *Plant Physiol* **138**, 2145-2154.
- Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C.** (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc Natl Acad Sci U S A* **107**, 15163-15168.
- Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H.** (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J* **47**, 841-850.
- Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S.** (2005). A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes Dev* **19**, 2164-2175.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X.** (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935.

- Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C., and Chen, X.** (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* **105**, 10073-10078.
- Zeng, Y., and Cullen, B.R.** (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**, 27595-27603.
- Zeng, Y., Yi, R., and Cullen, B.R.** (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo J* **24**, 138-148.
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R.** (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**, 1189-1201.
- Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.S., and Gregory, B.D.** (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet* **6**.
- Zhou, X., Sunkar, R., Jin, H., Zhu, J.K., and Zhang, W.** (2009). Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. *Genome Res* **19**, 70-78.
- Zhu, Q.H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., and Helliwell, C.** (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res* **18**, 1456-1465.
- 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.
- Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S.** (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**, 61-69.
- Zuker, M.** (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415.