

**FUNCTIONAL ANALYSIS OF microRNA TRIGGERS OF PHASED siRNA
BIOGENESIS IN PLANTS**

by

Qili Fei

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Plant and Soil Sciences

Summer 2016

© 2016 Qili Fei
All Rights Reserved

ProQuest Number: 10192168

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10192168

Published by ProQuest LLC (2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

**FUNCTIONAL ANALYSIS OF microRNA TRIGGERS OF PHASED siRNA
BIOGENESIS IN PLANTS**

by

Qili Fei

Approved: _____
D. Janine Sherrier, Ph.D.
Chair of the Department of Plant and Soil Sciences

Approved: _____
Mark Rieger, Ph.D.
Dean of the College of Agriculture & Natural Resources

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Blake C. Meyers, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

D. Janine Sherrier, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Nicole Donofrio, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Cheng Lu, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor Dr. Blake Meyers, who is a tremendous mentor during the years of my Ph.D. study. Dr. Meyers not only provided me continuous guidance and inspiration by his extraordinary wisdom in academics, but also influenced me by his courteous personality in getting along with colleagues and friends. Dr. Meyers is remarkably supportive to members in the lab. During my Ph.D. study, I have gained a lot of valuable experience by attending many scientific conferences to communicate with peers, and visiting the labs of our collaborators. The support from my advisor Dr. Meyers greatly helped me to develop my future career in research.

I also would like to thank my committee members, including Dr. Janine Sherrier, Dr. Nicole Donofrio and Dr. Cheng Lu, for their persistent assistance and valuable ideas input to my projects. They have devoted significant amounts of time and effort in attending my committee meetings in the past few years, especially during the time when big changes occurred after the Meyers lab moved to the Donald Danforth Plant Science Center in Missouri, and Dr. Cheng Lu moved to Iowa.

I am grateful to so many great colleagues during my Ph.D. study in the Meyers lab, including Dr. Tzoo-fen Lee, Dr. Pingchuan Li, Dr. Jixian Zhai, Dr. Arikrit Siwaret,

Dr. Kun Huang, Dr. Rui Xia, Dr. Chong Teng, Dr. Sandra Mathioni, Dr. Lanjuan Hu, Dr. Yu Zhang, Mayumi Nakano, Jing Xu, Reza Hammond, Saleh Tamim, Parth Patel, Suresh Pokhrel, Ayush Dusia, Deepti Ramachandrani, Feray Demirci, and Sherin Mathews. Special thanks are given to Dr. Jixian Zhai, Dr. Pingchuan Li and Dr. Tzuu-fen Lee for providing me plenty of help and guidance when I freshly joined the Meyers lab in early 2012. I also want to thank my classmates and friends from when we were in Delaware - Katie Bi, Yuanlong Tian, and Jia Ren, who shared lots of joy with me throughout my years of study. I thank many other old friends - the friendships that we cherish kept inspiring me while I was growing up.

Finally, I genuinely express my thanks to my parents and my brother. Their unfailing care and support not only encouraged me to pursue a Ph.D. degree, but also lighten up my entire life.

TABLE OF CONTENTS

LIST OF FIGURES	viii
ABSTRACT	x

Chapter

1	INTRODUCTION AND LITERATURE REVIEW	1
1.1	Overview of plant small RNAs	1
1.2	tasiRNA biogenesis, their functions, and their diversification in plants ...	4
1.3	miRNA triggers of phased, secondary siRNAs	11
1.4	Phased secondary siRNAs as a regulatory mechanism for protein-coding genes	16
1.5	Plant NB-LRRs as sources and targets of secondary siRNAs.....	22
2	SECONDARY siRNAs FROM MEDICAGO <i>NB-LRRs</i> MODULATED VIA miRNA-TARGET INTERACTIONS AND THEIR ABUNDANCES ...	27
2.1	Introduction	27
2.2	Results	31
2.2.1	Modulation of levels of miRNAs targeting <i>NB-LRRs</i> in Medicago	31
2.2.2	New targets of miRNAs, identified from miRNA overexpression in Medicago.....	39
2.2.3	Biogenesis of phasiRNAs determined by miRNA-target pairing and expression level of <i>PHAS</i> transcript.....	49
2.2.4	PhasiRNA production is determined by 3' pairing of a miRNA and its target	51
2.2.5	The role of target site flanking sequences in the function of miRNAs.....	55
2.3	Discussion.....	63
2.4	Conclusions and Future Directions	68
2.5	Materials and methods.....	70

3	STUDIES OF phasiRNA BIOGENESIS AND PLANT miRNA FUNCTIONS VIA CRISPR/Cas9.....	75
3.1	Introduction	75
3.2	Results	79
3.2.1	Genome editing of the passenger strand of <i>MIR160a</i> using a single guide RNA	79
3.2.2	Genome editing of the miRNA strand of <i>MIR160a</i> using a single guide RNA	87
3.2.3	Genome editing on <i>MIR160a</i> using double guide RNA	89
3.3	Discussion.....	93
3.4	Conclusions and Future Directions	95
3.5	Materials and Methods	97
4	SUMMARY AND FUTURE WORK	100
	REFERENCES	109
	Appendix	
	PUBLICATIONS RELATED TO THIS DISSERTATION	120

LIST OF FIGURES

Figure 1. Pathways for the biogenesis of phased, secondary siRNAs, modeled on Arabidopsis.....	10
Figure 2. Triggers and processing mechanisms of phased, secondary siRNAs.	15
Figure 3. miRNAs target nucleotides encoding conserved protein motifs of several gene families.....	21
Figure 4. miRNAs target conserved sequences in members of the NB-LRR gene families.	25
Figure 5. miRNA overexpression via hairy root transformation in Medicago.....	36
Figure 6. Schema of target mimic constructs for suppression of miRNAs targeting <i>NB-LRRs</i> in Medicago.....	38
Figure 7. Genome-wide analysis of phased small RNAs in Medicago.....	43
Figure 8. Fold change (FC) of summed phasiRNA abundances from each <i>PHAS</i> locus in miRNA overexpression lines.	45
Figure 9. PhasiRNA levels in transgenic tissues of four target mimic constructs.	46
Figure 10. Numbers of miRNA-target pairs within each penalty scores.	47
Figure 11. PARE signals increase by miRNA overexpression.	47
Figure 12. Differentially accumulating phasiRNAs in Medicago hairy roots overexpressing miRNAs.....	48
Figure 13. Factors that affect phasiRNA production.....	50
Figure 14. Pairing of 22-nt miRNAs with their targets, and its role in phasiRNA production.....	54
Figure 15. Ectopic expression of Medicago miRNAs in Arabidopsis.	59

Figure 16. PhasiRNAs are triggered by ectopic expression of Medicago miRNAs in Arabidopsis.....	60
Figure 17. Validated and potential targets in Medicago and Arabidopsis for mtr-miR2109.	61
Figure 18. Abundance levels of the 159 NB-LRR genes in Arabidopsis three-week old leaves.	62
Figure 19. Single guide RNA design for miRNA*.	82
Figure 20. Sanger sequencing results of individual T1 plants transformed with CRISPR/Cas9 targeting the miRNA* strand of miR160a.....	83
Figure 21. Sanger sequencing results of individual T2 plants transformed with CRISPR/Cas9 targeting the miRNA* strand of miR160a.....	84
Figure 22. T3 plants of line <i>160a*-13-16</i> and <i>160a*-9</i>	85
Figure 23. Small RNAs leaves of wildtype and <i>160a*-9</i> plants, mapped to target <i>ARF</i> genes to look for phasiRNA production.	86
Figure 24. Single guide RNA design for the miRNA strand of <i>MIR160</i>	88
Figure 25. CRISPR/Cas9-mediated genome editing via double guide RNAs.	90
Figure 26. Fragment deletion on <i>MIR160a</i> generated by CRISPR/Cas9 with double guide RNAs.	91
Figure 27. Mutants of <i>MIR160a</i> with a 47-bp fragment deletion display pleotropic phenotypes in plant development.	92
Figure 28. The integration of miRNAs and phasiRNAs in the “zig-zag-zig” model of the plant immune system.....	108

ABSTRACT

Small RNAs are a class of noncoding RNAs which are of great importance in gene expression regulatory networks. Different families of small RNAs are generated via distinct biogenesis pathways. One such family specific to plants is that of phased, secondary siRNAs (phasiRNAs); these require RDR6, DCL4, and (typically) a microRNA (miRNA) trigger for their biogenesis. Protein-encoding genes are an important source of phasiRNAs, and the model legume *Medicago truncatula* generates phasiRNAs from many *PHAS* loci.

We aimed to investigate their biogenesis and mechanism by which miRNAs trigger these molecules. We modulated miRNA abundances in transgenic tissues showing that the abundance of phasiRNAs correlates with the levels of both miRNA triggers and the target, precursor transcripts, and identified sets of phasiRNAs or *PHAS* loci that predominantly and substantially increase in response to miRNA overexpression. In the process of validating targets from miRNA overexpression tissues, we found that in the miRNA-mRNA target pairing, the 3' terminal nucleotide (the 22nd position), but not the 10th position, is important for phasiRNA production. Mutating the single 3' terminal nucleotide dramatically diminishes phasiRNA production. Ectopic expression of *Medicago NB-LRR*-targeting miRNAs in

Arabidopsis showed that only a few *NB-LRRs* are capable of phasiRNA production; our data indicate that this might be due to target inaccessibility determined by sequences flanking target sites. Our results suggest that target accessibility is an important component in miRNA-target interactions that could be utilized in target prediction, and the evolution of mRNA sequences flanking miRNA target sites may be impacted.

CRISPR/Cas9 has become a powerful technique in genome editing. In my study, CRISPR/Cas9 was employed to edit the passenger strand of *MIR160a* to convert pre-miR160 into an asymmetric structure in Arabidopsis, because evidence has been shown to support that the length of 22-nt miRNAs is important to trigger phasiRNA production. In the mutant with a single nucleotide insertion on miR160a*, we found that target transcripts of miR160, including *ARF10*, *ARF16* and *ARF17*, did not produce secondary siRNAs, suggesting that the asymmetric structure of miRNA might not be a determinant of phasiRNA production. Moreover, we tested the efficiency of fragment deletions in Arabidopsis *MIR160a* via the CRISPR/Cas9 vector with double guide RNAs, which would potentially generate the null mutant of miR160a. We found that CRISPR/Cas9 with a double guide RNA worked successfully by the floral dip method of transformation, reaching ~30% of fragment deletions (~50 bp) in the T1 generation. The *mir160a* mutants with a 47- or 48-bp fragment deletion showed severe pleiotropic developmental defects, such as serrated leaf, inward-curved and thin petal, short siliques, reduced fertility and arrested embryo development.

These results show that CRISPR/Cas9 is an efficient tool for functional studies of noncoding RNAs by fragment deletions in the plant genome.

Studies in the past a few years have shown that miRNAs, together with phasiRNAs are important regulators of plant NB-LRRs. In the Chapter 4 of this dissertation, we integrated small RNAs into the classic “zig-zag-zig model” of plant defenses, highlighting the roles of small RNAs in the modulation of host immunity.

Chapter 1

INTRODUCTION AND LITERATURE REVIEW

(This chapter has been published previously as Fei *et al.* (2013), modified to meet the formatting requirements of the dissertation.)

1.1 Overview of plant small RNAs

Plant small RNAs are in the size range of approximately 21 to 24 nucleotides; these short, processed transcripts play crucial roles in a variety of biological regulation processes, such as development, plant defense, and epigenetic modifications. Small RNAs in plants can be categorized into several major classes, including microRNAs (miRNAs), heterochromatic small, interfering RNAs (hc-siRNAs), phased secondary siRNAs (phasiRNAs), and natural antisense transcript siRNA (NAT-siRNAs). These categories are defined according to their origin and biogenesis (Axtell, 2013), with functions at both transcriptional and post-transcriptional levels. Common features of small RNAs are that members of the Dicer-like family (DCLs) are employed to cut longer RNAs into specific smaller lengths, and the resulting small RNAs are thereafter incorporated into Argonaute (AGO) family proteins to target complementary nucleotide sequences, functioning in a suppressive manner. In addition, recent data demonstrate that plant small RNAs are mobile, so that they can have effects over a long distance, including causing post-transcriptional silencing or epigenetic changes (Chitwood and Timmermans, 2010; Dunoyer *et al.*, 2010; Molnar *et al.*, 2010).

miRNAs are typically processed from a hairpin-like secondary structure of a non-coding mRNA (ncRNA), with this precursor mRNA generated by RNA polymerase II (Pol II). The RNase III enzyme DICER-LIKE 1 (DCL1) is responsible for biogenesis of the mature miRNA via processing of the mRNA precursor. DCL1 is one of four Dicer proteins encoded in a typical dicot genome (or one of five encoded in a typical monocot genome – see below). MicroRNAs function in a homology-dependent manner against target mRNAs to typically either (1) direct cleavage at highly specific sites, or (2) suppress translation; this mode of action depends largely on the miRNA complementarity with their targets (Valencia-Sanchez *et al.*, 2006; Voinnet, 2009). siRNAs are defined by their dependency on an RNA-dependent RNA polymerase (RdRP, or RDR) for their biogenesis. The activity of at least three of the six RDRs (RDR1/2/6) encoded in the Arabidopsis genome is believed to generate a double-stranded RNA (dsRNA) intermediary which is recognized and cleaved by a Dicer enzyme to generate different classes of siRNAs; so far little is known about the function of the triplicated paralogs RDR3/4/5 (Willmann *et al.*, 2011). Hc-siRNAs are ~24-nt in length, generated from DICER-LIKE 3 (DCL3) activity from intergenic or repetitive regions of genome via the activity of the plant-specific RNA polymerases Pol IV and possibly Pol V (Law and Jacobsen, 2010; Lee *et al.*, 2012; Matzke *et al.*, 2009). The function of hc-siRNAs is largely to maintain genome integrity, by maintenance of suppressive levels and types of DNA methylation on transposable elements. PhasiRNAs are described at length in this review and these are derived from an mRNA converted to dsRNA by RDR6, and processed by DICER-LIKE4 (DCL4),

exemplified by the category of *Arabidopsis* *trans*-acting siRNAs (tasiRNAs) (Vazquez *et al.*, 2004). In an exceptional case, phasiRNAs may also be 24-nt products of DICER-LIKE5 (DCL5, previously known as DCL3b) in grass reproductive tissues (Song *et al.*, 2012). The *trans*-acting name (tasiRNAs) of some phasiRNAs comes from their ability to function like miRNAs in a homology-dependent manner, directing AGO1-dependent slicing of mRNAs from genes other than that of their source mRNA (see below). NAT-siRNAs are a narrowly-described, unusual, and perhaps questionable category of small RNAs purportedly derived from two distinct, homologous, and interacting mRNAs (Borsani *et al.*, 2005). While hc-siRNAs play a crucial role in chromatin modifications, miRNAs, phasiRNAs and NAT-siRNAs function mainly at the post-transcriptional level by either cleavage or translational suppression of target transcripts, although a few instances have been described in which they can direct DNA methylation (Wu *et al.*, 2012; Wu *et al.*, 2010).

In the last few years, as a result of extensive genome sequencing in plants coupled with small RNA analysis, many new small RNAs have been described. Typically, with each new genome, a new cohort of miRNAs is described along with their mRNA targets. In parallel to these miRNAs studies, one of the most interesting findings of recent years in these new genomes has been the identification of a set of loci generating phased, secondary siRNAs, larger in number in most non-Brassica plant genomes than described for *Arabidopsis*. These secondary siRNAs are in many cases derived from a variety of protein-coding transcripts and in other cases from newly described long, non-coding mRNAs.

1.2 tasiRNA biogenesis, their functions, and their diversification in plants

Trans-acting siRNAs are a class of secondary siRNAs generated from non-coding *TAS* transcripts by miRNA triggers in a “phased” pattern (Allen *et al.*, 2005; Peragine *et al.*, 2004; Vazquez *et al.*, 2004; Yoshikawa *et al.*, 2005). The term “phased” indicates simply that the small RNAs are generated precisely, in a head-to-tail arrangement, starting from a specific nucleotide; this arrangement results from miRNA-triggered initiation followed by DCL4-catalyzed cleavage (Figure 1). The primary proteins that participate in tasiRNA biogenesis include RDR6, SUPPRESSOR OF GENE SILENCING 3 (SGS3), DCL4, AGO1, AGO7, and DOUBLE-STRANDED RNA BINDING FACTOR 4 (DRB4) (Adenot *et al.*, 2006; Fukudome *et al.*, 2011; Montgomery *et al.*, 2008a; Peragine *et al.*, 2004; Vazquez *et al.*, 2004; Xie *et al.*, 2005). While the roles of RDR6 and DCL4 are relatively clear, the role of SGS3 has not been well-described until recently. An *in vitro* analysis demonstrated that SGS3 can be recruited to AGO1-RISC (RNA-Induced Silencing Complex) complex by the 3' nucleotides of the 22-nt miR173 paired with the *TAS2* target RNA, and the function of SGS3 may be to stabilize the 3' target fragment resulting from miRNA-directed cleavage (Yoshikawa *et al.*, 2013). It's likely that there are other proteins involved in this process which have yet to be described, or which have minor roles, while yet other proteins may participate less directly via partially redundant roles; for example, DCL2 and DCL3 have partial redundancy with DCL4 in tasiRNA biogenesis (Gascioli *et al.*, 2005; Henderson *et al.*, 2006). Most importantly, there are two distinct pathways by which 21-nt tasiRNAs are produced, known as the “one hit” or

“two hit” pathways (Figure 1A). In the one-hit pathway, a single miRNA cleaves the mRNA target triggering the production of phasiRNAs in the fragment 3’ to (or “downstream” of) the target site (Figure 1B) (Allen *et al.*, 2005). We now know that this “one hit” miRNA trigger is typically 22-nt in length (Figure 2A) (Chen *et al.*, 2010; Cuperus *et al.*, 2010). In the two-hit model, a pair of 21-nt miRNA target sites is employed, of which cleavage occurs at only the 3’ target site, triggering the production of phasiRNAs in the fragment 5’ to (or “upstream” of) the target site (Figures 1, 2B) (Axtell *et al.*, 2006).

TasiRNAs, like miRNAs and other siRNAs, are usually incorporated into the RISC, leading to silencing of corresponding targets. TasiRNA functions have been well described from extensive work in *Arabidopsis*, which has a set of *TAS* genes that represent a core set of loci varying in their levels of conservation compared to other plants. Four families of *TAS* genes comprising eight loci have been identified in the *Arabidopsis* Col-0 genome, among which miR173 targets both *TAS1a/b/c* family and the *TAS2* locus, miR390 targets the *TAS3a/b/c* family, while miR828 triggers the production of *TAS4*-derived tasiRNAs (Allen *et al.*, 2005; Rajagopalan *et al.*, 2006; Yoshikawa *et al.*, 2005). *TAS3* is unique for several reasons: (1) it’s the only “two hit” locus in *Arabidopsis*, and (2) the 21-nt miR390 trigger is exclusively loaded to a specialized Argonaute, AGO7 (Axtell *et al.*, 2006; Montgomery *et al.*, 2008a). A subset of *TAS3a* derived tasiRNAs (tasi-ARFs) are involved in auxin responses, such as determining phase change or regulating root development, by altering transcript levels of Auxin Response Factor (ARF) members, including *ARF2*, *ARF3/ETT* and

ARF4 (Allen *et al.*, 2005; Fahlgren *et al.*, 2006; Hunter *et al.*, 2006; Marin *et al.*, 2010; Williams *et al.*, 2005). These “tasi-ARFs” form a concentration gradient from adaxial side to abaxial side of the leaf, suggesting they can move intercellularly as a regulator of ARF3-involved development (Chitwood *et al.*, 2009). The functions of the other Arabidopsis *TAS* genes are not well described: *TAS1* tasiRNAs target both pentatricopeptide repeat (PPR)-encoding transcripts as well as approximately five genes of unknown functions; *TAS2*-derived tasiRNAs target PPR-encoding transcripts as well (Allen *et al.*, 2005); *TAS4* tasiRNAs increase in the shoot under phosphate-deficient conditions, and perhaps participate in anthocyanin biosynthesis by targeting a group of MYB transcription factors (Hsieh *et al.*, 2009; Rajagopalan *et al.*, 2006). Another function of tasiRNAs is to mediate DNA methylation in *cis* at the *TAS* loci (Wu *et al.*, 2012), which is unusual given that these are 21-nt small RNAs. However, since this methylation does not obviously suppress the expression level of *TAS* genes, the functional importance of this observation is not yet clear.

TasiRNAs have been characterized in mosses, indicating that the utilization of tasiRNAs for gene regulatory functions is an ancient pathway in plants. In *Physcomitrella patens*, miR390, *TAS3a* and the resulting tasiR-ARFs have all been described (Axtell *et al.*, 2006; Talmor-Neiman *et al.*, 2006), as well as additional *TAS* loci, some of which are not conserved with Arabidopsis (Arif *et al.*, 2012). The *Physcomitrella TAS6* is a two-hit locus like *TAS3*, but is targeted by different conserved miRNAs, and has important roles in development, including bud formation (Cho *et al.*, 2012). *TAS3* is believed to be the most well-conserved *TAS* locus, as it has

been identified across a broad range of species, from *Physcomitrella* to monocots such as rice and maize (Heisel *et al.*, 2008; Williams *et al.*, 2005), and including gymnosperms such as pine (Axtell *et al.*, 2006). In *Arabidopsis*, only the 3' miRNA target site in the *TAS3* transcript is cleaved, as in other flowering plants, while both miR390 complementary sites in moss and pine showed cleavage (Figure 2B) (Axtell *et al.*, 2006). A shorter variant of *TAS3* than that found in *Arabidopsis* is also conserved in many eudicots and found alongside the canonical *TAS3* locus; this variant has cleavable target sites at both 5' and 3' positions and includes only a single tasiARF (Figure 2A)(Krasnikova *et al.*, 2009; Xia *et al.*, 2012). A recent study of phasiRNA trigger evolution (e.g. miRNAs) across a broad range of plant species demonstrated that after the appearance of the miR390 family ~450 million years ago, duplication, divergence, and neofunctionalization gave rise to at least seven families of miRNAs in two major superfamilies, the miR7122 and miR4376 superfamilies, while still maintaining miR390 as an important miRNA (Xia *et al.*, 2013). These three closely-related miRNA groups share a common origin and regulate distinct gene families (Xia *et al.*, 2013). Thus, land plants have widely exploited the regulatory functions of phasiRNA to regulate large gene families.

The characterization of tasi- or phasiRNAs in many plant genomes has utilized bioinformatics methods for genome-wide scans. Due to the precise 21-nt phasing of tasiRNAs (Allen *et al.*, 2005), genome-wide analysis with computational algorithms can identify potentially phased loci (Chen *et al.*, 2007; Howell *et al.*, 2007). These scans empirically define a specific “P-value” or “phasing score” as a threshold or cut-

off; in Arabidopsis, this identified the known tasiRNAs (see above), as well as several protein-coding genes, such as *PPR* transcripts, with 21-nt phased siRNAs (Chen *et al.*, 2007; Howell *et al.*, 2007). It was even shown that one tasiRNA, tasiR2140, plays a role in triggering tertiary tasiRNAs, as part of an expanded cascade of tasiRNA regulation (Chen *et al.*, 2007). Interestingly, *PPR* transcripts were shown to generate 21-nt secondary phased siRNAs, some of which were triggered by tasiRNAs, and some triggered by other miRNAs (Chen *et al.*, 2007; Howell *et al.*, 2007), but with the important observation that phased, secondary siRNAs are generated not only from non-coding *TAS* loci, but also from protein-coding transcripts. In more recent work, we have attempted to clarify the “tasi” versus “phasi” nomenclature (Zhai *et al.*, 2011); *trans*-acting function is often not experimentally confirmed coincident with the identification of phased siRNAs, and thus “phasiRNAs” are loci merely identified as phased, whereas “tasiRNAs” have been demonstrated to act in *trans* (Zhai *et al.*, 2011). In addition, the “*TAS*” name has been given only to non-coding transcripts with no function other than to give rise to secondary siRNAs. Recent work has described the *TAS6* loci (Arif *et al.*, 2012), and many *TAS*-like (*TASL*) loci (Xia *et al.*, 2013) as well as an unnamed *TAS*-like ncRNA locus (Zhai *et al.*, 2011), indicating that additional ncRNA-derived *TAS* loci will continue to be described and named, some of which may be lineage-specific. *TAS5* has also been described (Li *et al.*, 2012a), but we believe it is inappropriately named, as it appears to be an incorrectly annotated protein coding (NB-LRR) transcript. With the proliferation of sequenced plant genomes in

recent years, an integral part of genome annotation is to identify the full complement of phasiRNA-generating loci.

As a consequence of the relatively well-understood biogenesis pathway for tasiRNA and mechanism of their function, several labs have exploited this effective RNA silencing method for the study of gene function. Montgomery et al. used a synthetic *TAS3a* and *TAS1c* in Arabidopsis to produce artificial tasiRNAs targeting the *PDS* gene, resulting in photobleaching at the site of activity (Montgomery *et al.*, 2008a; Montgomery *et al.*, 2008b). In separate work, silencing of CHLORINA 42 (CH42) gene produced photobleaching, and was achieved by use of a modified *TAS1a* transcript (Felippes and Weigel, 2009). Finally, a *TAS1c* silencing system containing anywhere from a single *FAD2*-specific siRNA to a 210 bp fragment of *FAD2* could successfully phenocopy the *FAD2* loss-of-function mutant (de la Luz Gutierrez-Nava *et al.*, 2008). Presumably, in any of these systems or using other phased siRNA-producing transcripts, multiple artificial tasiRNAs could be developed to silencing several genes at once. Thus, artificial tasiRNAs are a powerful tool for gene functional analysis.

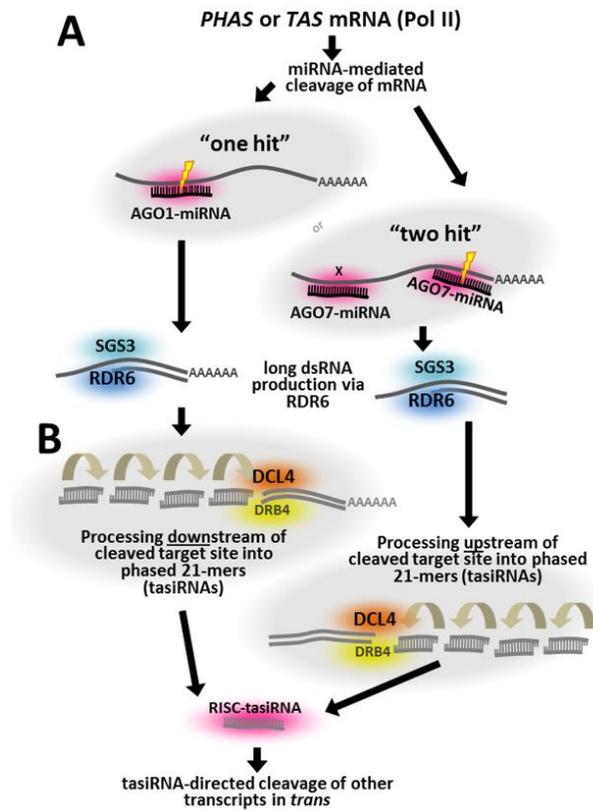


Figure 1. Pathways for the biogenesis of phased, secondary siRNAs, modeled on Arabidopsis.

A. As the first step in secondary siRNA biogenesis, mRNA targets are cleaved by a miRNA. In the “one-hit” model, exemplified in Arabidopsis by *TAS1*, *TAS2*, and *TAS4*, a 22-nt miRNA targets a single site. In the “two-hit” model, there are two target sites for a 21-nt miRNA, exemplified in Arabidopsis by *TAS3* transcripts cleaved by an AGO7-loaded miR390. Activity of the trigger miRNA recruits RDR6 and SGS3, resulting in production of a second strand of the target mRNA. B. The double-stranded RNA is successively processed by DCL4 and other components to generate 21-nt tasiRNAs; the direction of processing depends on the miRNA trigger mechanism. The secondary siRNAs are loaded onto an Argonaute protein and go on to function against other mRNAs.

1.3 miRNA triggers of phased, secondary siRNAs

As mentioned above, an intriguing early observation was that either one or two miRNA target sites can trigger tasiRNA biogenesis. The two-hit model provided the first mechanistic insights into the process for tasiRNA biogenesis, describing *TAS3* as the prototypical two-hit locus (Axtell *et al.*, 2006). Early experimental examination of the two *TAS3* target sites demonstrated that miR390 has an unusual association with AGO7 that is important for tasiRNA biogenesis, and that the 5' proximal miR390 target site must not be cleaved (Montgomery *et al.*, 2008a) – although outside of Arabidopsis, *TAS3* variants may be cleaved at the 5' position (Axtell *et al.*, 2006; Krasnikova *et al.*, 2009; Xia *et al.*, 2012). However, more recent data from other plant genomes describe two-hit loci for which the miRNA triggers are believed to be AGO1-associated, and for which the 5' proximal site may be cleaved. For example, in Medicago, in addition to *TAS3*, another “2₂₁” *TAS* locus was described (identified as such because it is a two-hit locus with two 21-nt miRNA target sites); like *TAS3*, the 5' proximal site is not cleaved and the 3' site is cleaved, but the triggers are miR172 and miR156, two well-conserved miRNAs that are AGO1-loaded in Arabidopsis (Figure 2B)(Zhai *et al.*, 2011). In both Medicago and apple, “2₂₂” loci have been described, with cleavage by two 22-nt miRNAs at both 5' and 3' proximal target sites resulting in bidirectional processing into phasiRNAs of the fragment between the target sites (Figure 2A)(Xia *et al.*, 2013; Zhai *et al.*, 2011). In yet another 2₂₂ variant, a cleavable 5' site and non-cleavable 3' site trigger phasiRNAs (Figure 2A)(Shivaprasad

et al., 2012; Xia *et al.*, 2013). More recent work in *Physcomitrella* has emphasized the diversity of two-hit loci, confirming via a newly described *TAS6* locus that 2₂₁ *PHAS* loci can be triggered by a pair of presumably AGO1-loaded 21-nt miRNAs that can be different from one another (Figure 2B)(Cho *et al.*, 2012). Thus, our current understanding is that the non-cleaving 5' proximal miRNA target site is apparently a unique feature of AGO7-loaded miR390 for some *TAS3* loci, with other two-hit loci utilizing cleaved 5' proximal sites via AGO1-loaded miRNAs.

Analysis of “one-hit” triggers of *PHAS* loci and experiments using these miRNAs have also producing intriguing findings. In 2010, a pair of articles described that a shared feature of one-hit loci is that the triggers are 22-nt and not 21-nt miRNAs (Chen *et al.*, 2010; Cuperus *et al.*, 2010). This led to the hypothesis that 22-nt miRNAs have special properties – the ability to trigger the production of phased siRNAs, confirmed via experiments employing a variety of constructs to generate miRNAs of specific lengths (Chen *et al.*, 2010; Cuperus *et al.*, 2010). In their experiments, canonical 21-nt miRNAs known to not trigger phased siRNAs, when produced as 22-nt variants, triggered the production of secondary siRNAs. Consistent with these results, tasiR2140, an unusual 22-nt tasiRNA triggers phasiRNA biogenesis from its target transcripts (Chen *et al.*, 2010). More recent work has demonstrated that alterations in the 3' nucleotide of the trigger miRNA can disrupt phasiRNA biogenesis, indicative of a role specifically for the small RNA length or target interactions (Zhang *et al.*, 2012b).

A recent publication indicates that the secondary structure of the miRNA duplex, rather than the 22-nt length, is the primary determinant of activity in triggering secondary siRNAs (Manavella *et al.*, 2012). Manavella and co-authors observed that not only 22-nt miRNAs but also 21-nt miRNAs with 22-nt miRNA* sequences can trigger secondary siRNA biogenesis. This led them to identify the characteristic shared by these miRNAs as an asymmetric duplex in the precursor, with the asymmetry resulting from a bulge or unpaired nucleotide. While AGO7-associated miRNAs (like miR390) are known to trigger secondary siRNAs (AGO2 may be similar), some miRNA triggers that they examined are typically loaded into AGO1 which is predominantly not associated with secondary siRNA production; this suggests that the same RISC components (AGO proteins) can either produce or not produce secondary siRNAs (Manavella *et al.*, 2012). To test the role of an asymmetric duplex, they created a synthetic version of miR173 with two asymmetric bulges that they demonstrated gave rise to 21-nt versions of both the miRNA and miRNA* - yet still triggered secondary siRNAs as effectively as a 22/21-nt asymmetric precursor (Manavella *et al.*, 2012). They thus inferred that RISC is reprogrammed upon interaction with an asymmetric duplex (i.e. a bulge caused by an unpaired base), and this reprogrammed RISC recruits proteins for secondary siRNA biogenesis. Manavella *et al.* also showed that the siRNAs were produced through via RDR6/SGS3/DCL4 – the cofactors likely recruited by the RISC.

While these data are quite convincing, other data suggest that duplex asymmetry cannot entirely explain the ability of some plant miRNAs to “trigger”

secondary siRNAs. Our group observed that in *hen1*, two targets of miR170 and miR171a, miRNAs produced from symmetric precursors, give rise to phasiRNAs (Zhai *et al.*, 2013). In wildtype, no phasiRNAs are produced from the targets, and these miRNAs are 21 nucleotides, but in the *hen1* mutant, 3' uridylation of the miRNAs after biogenesis gives rise to 22-nt variants. We inferred that in this case, it is the 22-nt length that confers the triggering activity for secondary siRNAs. Consistent with our observations, there are several reports of secondary siRNAs which for inexplicable reasons are consistently generated as 22-nt siRNAs – and themselves then trigger secondary siRNAs at targets in *trans*; this includes the Arabidopsis tasiR2140 (Chen *et al.*, 2007), as well as tasiRNAs from several *TASL* loci in multiple species (Xia *et al.*, 2013). Also inconsistent with the requirement of an asymmetric precursor, miR828 is produced from a symmetrical stem-loop precursor yet triggers phasiRNA production from *TAS4* and many *MYB* genes (Rajagopalan *et al.*, 2006; Xia *et al.*, 2012). Finally, mismatches between 3'-terminus of miRNA triggers and their *TAS* targets reduce the stability of the interaction between the cleavage fragment and RISC complex, inhibiting tasiRNA production (Zhang *et al.*, 2012b). This is likely because a mismatched 3' end would fail to recruit SGS3 and thus fail to stabilize the 3' mRNA fragment (Yoshikawa *et al.*, 2013). These results suggest the importance of miRNA-target interactions in generating tasiRNAs.

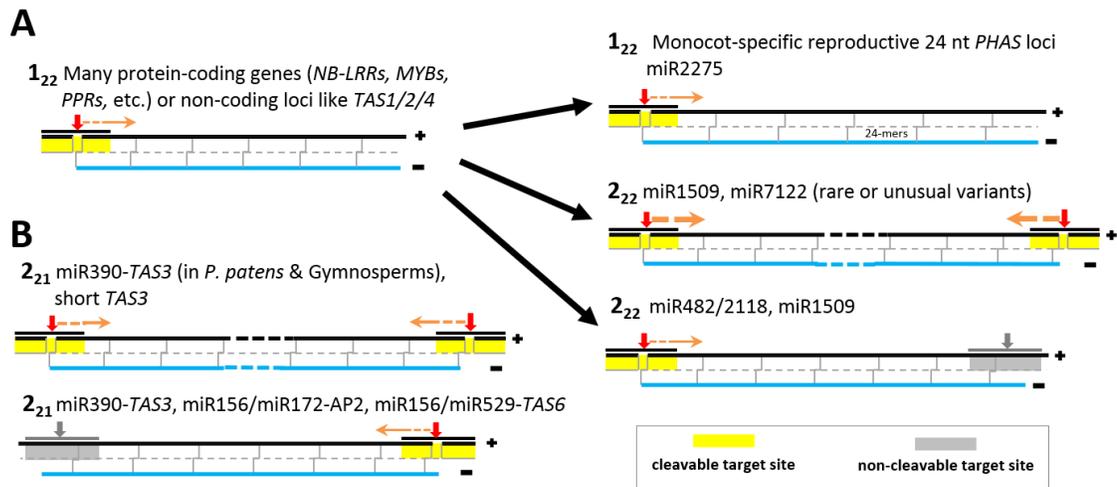


Figure 2. Triggers and processing mechanisms of phased, secondary siRNAs.

The primary mechanisms of processing for plant phased, secondary siRNAs are described along with prototypical loci and the miRNAs that trigger siRNA biogenesis at these loci. Red arrows indicate cleavage sites, orange arrows indicate the direction of precursor processing into phasiRNAs – which are indicated by grey lines in the double-stranded black/blue precursors. A. The “one-hit” pathway is typified by a single target site for a 22 nt miRNA that results in downstream processing of the target transcript into ~21 nt, phased siRNAs. This is denoted as a 1_{22} locus. There are at least three notable variations on the one-hit model, including: (i) the reproductive lncRNAs of monocots that are processed by DCL5 into 24 nt phased siRNAs, triggered by miR2275 and thus also 1_{22} loci, but with different biogenesis components; (ii) and (iii) are both 2_{22} loci, but the 3’ site can be either cleaved or not cleaved. B. The “two-hit” pathway is typified by two target sites of a 21-nt miRNA that results in processing upstream of the 3’ site. This is denoted as a 2_{21} locus, and the best characterized examples are *TAS3* and related loci, although a few other examples have been described. The 5’ site may be cleaved, which may result in processing from both directions, or the 5’ site may be non-cleaved, as originally described for the Arabidopsis *TAS3* locus. (Rui Xia designed this figure.)

1.4 Phased secondary siRNAs as a regulatory mechanism for protein-coding genes

MicroRNA-triggered secondary siRNAs are also generated from protein-coding loci in many plant genomes, first described in *Arabidopsis* (Howell *et al.*, 2007). A significant number of *PPR*, *NB-LRR* and *MYB* families were shown to generate phasiRNAs, from *Arabidopsis*, *Medicago*, apple and peach (Howell *et al.*, 2007; Xia *et al.*, 2012; Zhai *et al.*, 2011; Zhu *et al.*, 2012). The *PPR* family is one of the largest gene families in *Arabidopsis*, containing about 450 members in total, some of which were shown to be related with organelle RNA processes (Lurin *et al.*, 2004; O'Toole *et al.*, 2008). In *Arabidopsis*, a small number of miRNAs and tasiRNAs have been shown to collectively target ~40 PPRs, among which 28 are closely related (Howell *et al.*, 2007). A comparative analysis across plant species demonstrated conservation of the ability of this subgroup of PPRs to spawn secondary siRNAs, targeting a broader group of PPRs in many but not all plants (Xia *et al.*, 2013). The triggers of these PPR-derived secondary siRNAs are a superfamily of miRNAs, as well as unusual, 22-nt secondary siRNAs that function in *trans* (Chen *et al.*, 2007; Xia *et al.*, 2013). These small RNAs target variable sites within the *PPR* domains (Figure 3A). Since this regulatory network includes both miRNAs and tasiRNAs, it represents a highly redundant, interconnected set of PPR-targeting small RNAs. It was proposed that this regulation could be beneficial to the evolutionary expansion of *PPR* genes (Howell *et al.*, 2007). The superfamily of miRNAs that trigger secondary siRNAs from *PPRs* is also unusual because (1) it is derived from the prototypical phasiRNA

trigger, miR390, and (2) it gave rise to a different superfamily of miRNAs that target Ca²⁺ ATPases, some of which (but perhaps not all) also generate phasiRNAs (Wang *et al.*, 2011; Xia *et al.*, 2013).

Genes encoding *MYB* transcription factors are also rich sources of miRNA-triggered secondary siRNAs. MYBs are a family of DNA-binding proteins playing important roles in a variety of transcriptional regulation processes, such as cellular morphogenesis, meristem formation, cell cycle, and anthocyanin biosynthesis (Jin and Martin, 1999; Petroni and Tonelli, 2011). MYB transcription factors are encoded by one of the largest of gene families in many plant genomes (Feller *et al.*, 2011). Several MYBs have been identified to be responsible (among other activities) for anthocyanin biosynthesis both in fruit development of apple and other Rosaceae species, as well as in maize kernels (Feller *et al.*, 2011; Lin-Wang *et al.*, 2010; Takos *et al.*, 2006). In the case of apple, phasiRNAs are produced from a number of MYB-coding genes, as miR828 and miR858 target the conserved motifs of up to 81 *MYB* transcripts (Figure 3B)(Xia *et al.*, 2012). A comparative phylogenetic analysis revealed that those *MYB* genes containing target sites of both miR858 and miR828 are conserved across a broad range of plants with the miRNAs found only in eudicots thus far (Xia *et al.*, 2012); perhaps phasiRNA regulation of MYBs is an adaptation specific to the eudicots. The apple *MYB*-derived phasiRNAs are predicted to target a variety of genes with distinct functions, potentially expanding this miRNA-mediated regulatory network (Xia *et al.*, 2012). Similar to apple, peach also produces a large number of *MYB*-derived phasiRNAs (Zhu *et al.*, 2012). As with PPRs, MYBs are encoded by a large and

complex gene family in plant genomes, but the functional or evolutionary role of phasiRNA-transcriptional suppression of the family is unclear.

NB-LRR-encoding genes comprise one of the largest families found to be targeted by small RNAs. Compared to the PPR- and MYB-encoding gene families in other plant genomes, a much larger number of *NB-LRRs* were found to be *PHAS* loci in the *Medicago* genome (Zhai *et al.*, 2011). Many phasiRNAs target *NB-LRR* transcripts either in *cis* or in *trans* at other *NB-LRR* loci, representing a self-reinforcing regulatory network (Zhai *et al.*, 2011). As with the *PPR* family in *Arabidopsis* and other plants (Howell *et al.*, 2007; Xia *et al.*, 2013), *NB-LRRs* can be redundantly targeted by both miRNAs and secondary siRNAs (Zhai *et al.*, 2011). *NB-LRRs* regulation by secondary siRNAs has also been reported to exist widely in the Solanaceae (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Zhai *et al.*, 2011). Most recently, an examination of a wider variety of *NB-LRRs* in a wider variety of plant species demonstrated significant levels of secondary siRNAs in Norway spruce (a gymnosperm), *Amborella* (a basal angiosperm), cotton, poplar, grapevine, apple, and peach, indicating broad conservation and an ancient origin for the role of phasiRNAs in regulation of *NB-LRRs* (Kallman *et al.*, 2013). Possible reasons for phasiRNA regulation of *NB-LRR* transcripts are discussed in more detail below.

Many transcripts of protein-coding genes other than *PPRs*, *MYBs*, or *NB-LRRs* also generate phasiRNAs, but thus far, outside of these three large gene families, these protein-coding *PHAS* loci are solitary or very small families. For example, in soybean, the small RNA biogenesis machinery is itself subject to phasiRNA regulation,

evidenced by secondary siRNAs mapping to both *DCL2* and *SGS3* transcripts, and a number of other single- or low-copy genes are sources of phasiRNAs (Zhai *et al.*, 2011). Likewise in peach, phasiRNAs are produced from many single- or low-copy genes, including (among others) those encoding TIR/AFB, AUXIN RESPONSIVE FACTOR (ARF), and a Ca²⁺-ATPase (Xia *et al.*, 2013). The *PHAS* characteristic of many low copy protein-coding genes is conserved across species, for example, the tomato ortholog of the peach Ca²⁺-ATPase is also a *PHAS* locus (Wang *et al.*, 2011). Computational analysis of grape small RNAs identified nearly 50 phased loci in total, among which at least 20 are protein-coding genes and some of which are members of the *NB-LRR* family (Zhang *et al.*, 2012a). These results indicate phasiRNA-associated regulatory networks are utilized by many low-copy genes and gene families involved in diverse biological processes and pathways. It's possible that phasiRNAs perform regulatory functions such as tuning or heavily suppressing transcript levels that are equally important for single copy genes, although in such cases, these secondary siRNAs are presumably functioning in *cis* since there may be no *trans* targets.

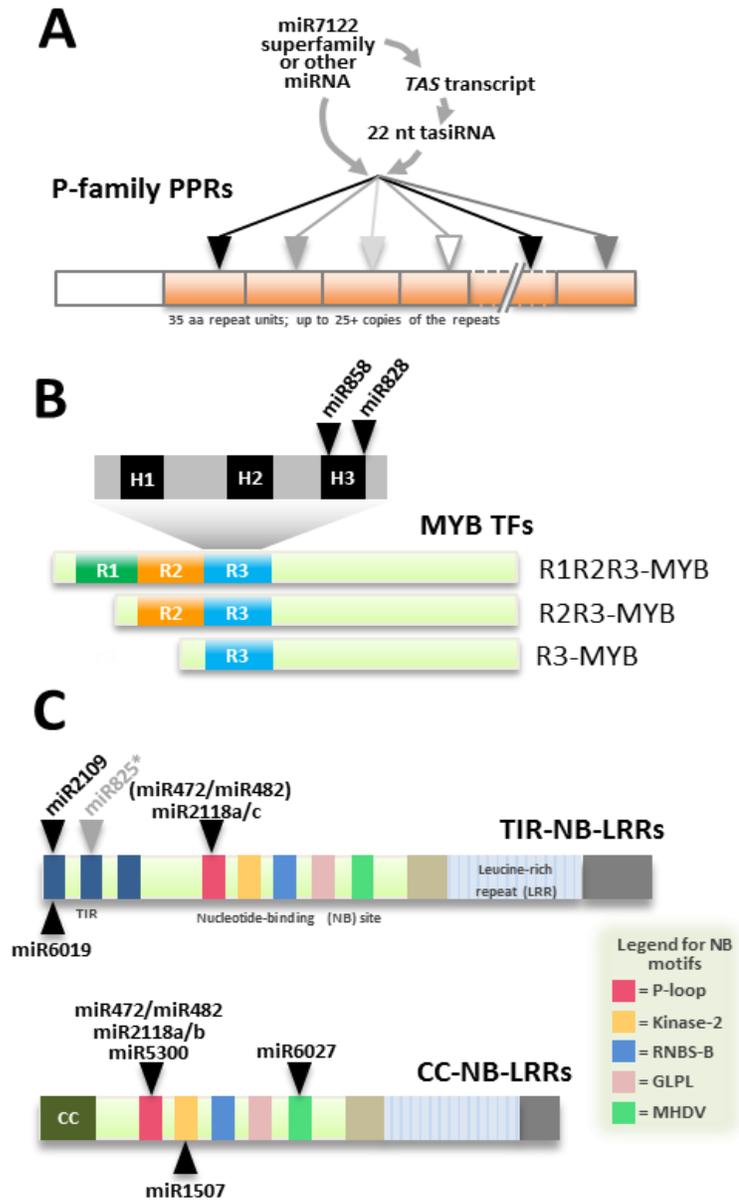


Figure 3. miRNAs target nucleotides encoding conserved protein motifs of several gene families.

A. PPR genes encoding the P subclass of PPRs are targeted by both miRNAs and tasiRNAs. Each grey box represents one degenerate repeat of ~35 amino acids. PPR proteins have a widely varying number of these repeat units (indicated by the broken repeat unit). Grey or outlined arrowheads indicate that miRNA or tasiRNA target sites may exist at varying levels, or may not exist at all in some repeats, due to the degeneracy of the repeat sequences. B. miRNAs target nucleotides encoding H3 motifs in the conserved R3 domains of MYB transcription factors in plants. C. Numerous miRNAs target nucleotides encoding conserved motifs of NB-LRRs in many plant species. The NB domain has five conserved motifs indicated by colored boxes; other conserved domains and motifs characterize these proteins, as indicated. Considering many plant species, multiple encoded motifs of NB-LRRs are targeted, including the TIR1, TIR2, P-loop, kinase-2, and MHDV motifs. miR472 and miR482 are nearly identical (see Figure 4A), and indicated parenthetically for TNLs, as CNLs are the preferential targets (with TNLs as less frequent targets). miR825* is indicated in grey, as it is observed to target an encoded TIR2 only in Arabidopsis.

1.5 Plant NB-LRRs as sources and targets of secondary siRNAs

Analysis of siRNAs matched to *NB-LRRs* in *Medicago* and several Solanaceous species identified many phased, secondary siRNAs (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Zhai *et al.*, 2011). In *Medicago*, transcripts encoding NB-LRRs are targeted by miRNAs at several conserved motifs, triggering phasiRNA production from these genes, to which we refer as “*phasi-NB-LRRs*” or *pNLs* (Zhai *et al.*, 2011). While more than 114 phasiRNA-producing *NB-LRRs* were identified, >60% of *Medicago* genomic *NB-LRRs* had significant levels of 21-nt small RNAs – suggesting that most members of this gene family are targeted by 22-nt miRNAs (Zhai *et al.*, 2011). Because phasiRNAs can also function both in *cis* and in *trans*, targeting other related transcripts, a limited number of miRNA triggers can dramatically amplify their suppressive functions through the production of secondary phasiRNAs, and these two kinds of small RNAs seem to have a joint effect in regulating the great majority of *NB-LRRs* in *Medicago*. Thus, miRNAs act as “master regulators” of the *NB-LRR* gene family via the production of phasiRNAs (Zhai *et al.*, 2011). Compared to *Medicago*, the numbers of *PHAS* loci are relatively smaller in other legume species, such as soybean (Zhai *et al.*, 2011). However, due to the synergistic effect by miRNAs and secondary phasiRNAs, a significant proportion of *NB-LRRs* could be targeted and down-regulated in legumes other than *Medicago*. In the Solanaceous species (tomato, potato, tobacco), numerous *pNLs* have been described, although our unpublished analysis suggests a lower proportion of genomic *NB-LRRs* in Solanaceous genomes

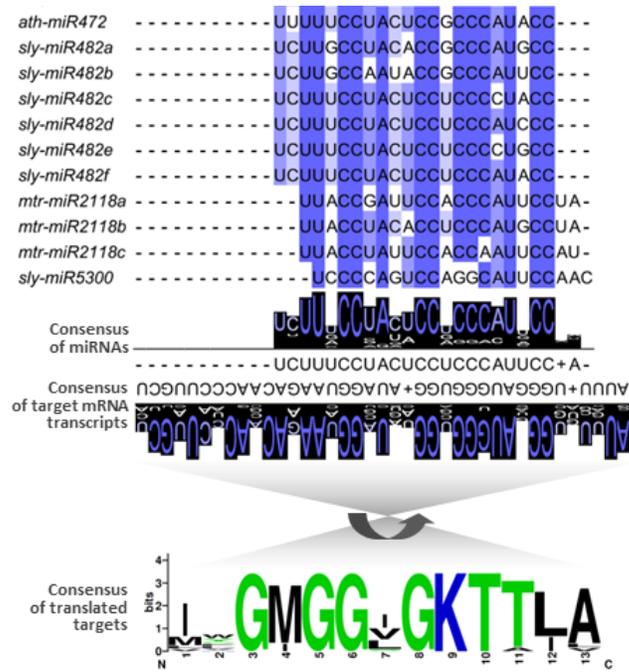
are *pNLs* than in *Medicago* (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Zhai *et al.*, 2011).

In the relationship between miRNAs and their *NB-LRR* targets, there is an unusual level of redundancy (Figure 3C). In *Medicago*, three families of 22-nt miRNA (miR1507, miR2109, and miR2118a/b/c) target the sequences encoding highly conserved protein motifs, such as TIR-1, P-loop, and Kinase-2, triggering phasiRNA production (Zhai *et al.*, 2011). These three unrelated families of miRNAs show no specialization for clades or subgroups within the *NB-LRRs*, implying that any one of these miRNAs is capable of targeting very diverse members of the *NB-LRR* family. Together with additional *NB-LRR*-targeting miRNAs from the Solanaceae, there are at least six miRNA families that target *NB-LRRs* (Figure 3C). Typically, plant miRNAs and their target families of genes show a one-to-one relationship, with a single miRNA that targets a single set of genes. For example, there are five copies of the miR172 family in *Arabidopsis*, which all target members of the *APETALA2* gene family (Aukerman and Sakai, 2003). Thus the case of *NB-LRRs*, targeted independently by six different miRNA families, is highly unusual. There are two additional levels of redundancy in *NB-LRR*-miRNA interactions worth considering: (1) the phasiRNAs generated via miRNA cleavage may function in *trans* to silence related targets. This *trans*-acting activity was confirmed in *Medicago* (Zhai *et al.*, 2011), and given the tremendous abundance of phasiRNAs produced from *NB-LRRs* in many species, this is likely a significant mechanism for silencing within the family. (2) An additional level of redundancy is represented by the diversity of miRNAs which target

nucleotides encoding conserved protein motifs. The most extreme case of this is the superfamily of miRNAs that target the encoded P-loop, a group which includes miR472, miR482, miR2089, miR2118, and miR5300. While some of this variation in naming is simply a historical artifact (i.e. miR472 and miR482 are nearly identical), there is substantial sequence variation in members of this superfamily such that the members wouldn't fit the definition of a single family (Figure 4A) (Meyers *et al.*, 2008). This superfamily could be known by its inclusion of both the miR482-type, more predominant in the Solanaceae (Shivaprasad *et al.*, 2012), or the miR2118-type, more predominant in the Fabaceae (Zhai *et al.*, 2011). The TIR-1 motif is similarly targeted by two unrelated miRNAs, miR2109 and miR6019, which target non-overlapping nucleotides that encode the motif (Figure 4B). It's possible that future miRNA annotation in more diverse species will identify even more divergent members of these families or superfamilies which may contribute further to the high level of redundancy of in the suppression of *NB-LRR* transcripts.

The functional relevance of endogenous *NB-LRR* silencing is unknown, yet the data undeniably demonstrate that it is widespread within the gene family, robust, and found in many very diverse angiosperms and as far back evolutionarily as the gymnosperms. Although the Poaceae apparently lack this *NB-LRR*-suppressive regulatory machinery (i.e. the miRNAs and therefore the phasiRNAs), and it's greatly reduced in the Brassicaceae, the phenomenon of *pNLs* is so prevalent that they must have an important function. Furthermore, understanding the mechanistic importance of phasiRNAs in *NB-LRR* regulation may provide insights into the analogous

miRNA/phasiRNA suppression of the PPRs and MYB TFs, or other protein-coding genes.



B

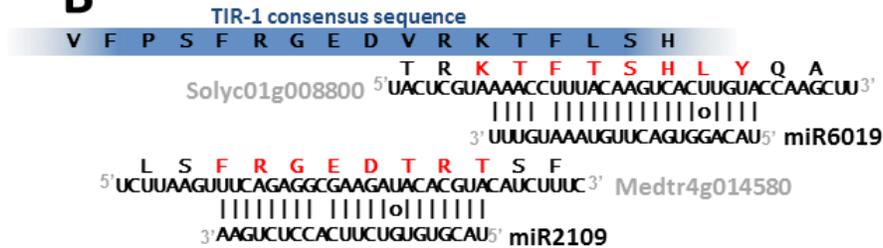


Figure 4. miRNAs target conserved sequences in members of the NB-LRR gene families.

A. The miR482/miR2118 superfamily of miRNAs is a relatively diverse group (above) that typically target nucleotides encoding the P-loop motif of NB-LRR proteins (below). Consensus sequences of either the miRNAs or their targets demonstrate a high degree of conservation (illustrated by WebLogo). In this figure, for illustrative purposes, we've randomly selected a diverse set of miRNA superfamily members (listed by their names in miRBase) and ~16

targets from the same source species. The miRNAs are shown 5' to 3' in the consensus, the mRNA targets are shown 3' to 5', and the translated protein motif is inverted relative to the target mRNAs (indicated by the curved arrow).

B. Redundancy in miRNA targeting at the encoded TIR-1 motif of NB-LRRs. miR6019 and miR2109 aligned to their TNL-encoding targets show they target the same encoded motif (blue bar at top), but at adjacent, non-overlapping sites. The consensus at the top is from Meyers et al. (1999). Red letters indicate the amino acids encoded by the target region. Example targets from tomato ("Soly...") and Medicago ("Medtr...") are indicated aligned to the miRNA sequences.

Chapter 2

SECONDARY siRNAs FROM MEDICAGO *NB-LRRs* MODULATED VIA miRNA-TARGET INTERACTIONS AND THEIR ABUNDANCES

(This chapter has been published previously as Fei *et al.* (2015), modified to meet the formatting requirements of the dissertation.)

2.1 Introduction

Small RNAs play important roles in gene silencing at both transcriptional and post-transcriptional levels. MicroRNAs (miRNAs), a special class of small RNAs processed by DICER-LIKE 1 (DCL1) from mRNA precursors with a stem-loop secondary structure, participate in many biological processes in plants, such as development, stress responses, disease resistance, etc. (Voinnet, 2009). In addition to the direct silencing of target genes, a subset of miRNAs are also able to trigger the production of phased, secondary siRNAs (phasiRNAs) from the cleavage products of target transcripts in plants, predominantly 21-nt in length (occasionally 22-nt). This process requires the participation of proteins that include RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), SUPPRESSOR OF GENE SILENCING 3 (SGS3), DICER-LIKE 4 (DCL4), and DOUBLE-STRANDED RNA BINDING PROTEIN 4 (DRB4) (Fei *et al.*, 2013). With just a few exceptions, including highly-conserved miR390, only 22-nt miRNAs are capable of triggering phasiRNA production (Chen *et al.*, 2010; Cuperus *et al.*, 2010). An investigation into the mechanism by which these

phasiRNA triggers function demonstrated that an asymmetric bulged structure of a miRNA/miRNA* duplex contributes to the reprogramming of RNA-induced silencing complex (RISC) endowing these miRNAs with phasiRNA ‘triggering’ activity (Manavella *et al.*, 2012).

Plant phasiRNAs are produced from both coding and non-coding RNAs (the loci that produce them are all called “*PHAS*” loci). *Trans*-acting siRNAs (tasiRNAs) are a special class of phasiRNAs generated from non-coding *TAS* genes. The function of *TAS3*-derived tasiRNAs (tasiARFs) is to modulate transcript levels of auxin response factors (ARFs), known in *Arabidopsis* and other plants; the functions of three other *Arabidopsis* tasiRNA loci, *TAS1/2/4*, in contrast, are poorly characterized. Previous studies have shown that tasiARFs play crucial roles in development in diverse tissues (Adenot *et al.*, 2006; Fahlgren *et al.*, 2006; Garcia *et al.*, 2006; Marin *et al.*, 2010). The functions of the regulatory pathway involving miR390/tasiARFs are also highly conserved across land plants, from mosses (*Physcomitrella patens*) to grasses (maize) (Cho *et al.*, 2012; Dotto *et al.*, 2014). In grasses, in addition to the non-coding *TAS3* gene, reproductive-specific non-coding RNAs producing both 21- and 24-nt phasiRNAs were found in large numbers, firstly in rice and recently in maize (Johnson *et al.*, 2009; Zhai *et al.*, 2015). These are triggered by miR2118 and miR2275, and their biogenesis requires distinct Dicer proteins, DCL4 and DICER-LIKE 5 (DCL5) (previously known as DCL3b) (Song *et al.*, 2012). Recent work shows that reproductive 21-nt phasiRNAs triggered by miR2118 are in rice associated with a specific AGO protein (MEIOSIS ARRESTED AT LEPTOTENE 1, or MEL1); the

mell mutant is sterile, suggesting a potentially important role of these phasiRNAs in reproductive development in monocots (Komiya *et al.*, 2014).

In addition to non-coding *PHAS* loci, plant genomes also contain a large number of protein-coding *PHAS* loci. Genome-wide experimental and computational analysis has identified phasiRNAs from many gene families, including *NB-LRRs*, pentatricopeptide repeat (*PPRs*), *MYBs*, Ca^{2+} -*ATPase*, and *TIR1/AFB* (Chen *et al.*, 2007; Howell *et al.*, 2007; Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Si-Ammour *et al.*, 2011; Wang *et al.*, 2011; Xia *et al.*, 2012; Zhai *et al.*, 2011). A recent study in soybean identified a total of 20 miRNAs that can trigger phasiRNA production from a broad range of gene families (Arikiti *et al.*, 2014). Among these, *NB-LRR*-encoding *PHAS* loci are conspicuous, because *NB-LRRs* comprise one of the largest gene families in plants, and they are redundantly targeted by diverse miRNA families in many species (Fei *et al.*, 2013). For example, in Medicago, the 22-nt miRNAs miR1507, miR2109 and miR2118a/b/c target highly conserved sites encoding protein motifs (Kinase-2, TIR-1, and the P-loop respectively) (Zhai *et al.*, 2011). In Solanaceous species like tobacco and tomato, miR6019 and miR482 target *NB-LRRs* and trigger phasiRNAs (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012). The phasiRNAs produced from *NB-LRRs* function in regulating plant immunity; for example, in Arabidopsis, miR472- and RDR6-mediated gene silencing helps modulate both PAMP- and effector-triggered immunity (Boccardo *et al.*, 2014), while in barley, miR9863 targets *Mla* alleles, a set of coiled-coil (CC) type *NB-LRRs* conferring

resistance to powdery mildew, modulating *Mla* allele transcript levels via synergistic action with the phasiRNAs it triggers (Liu *et al.*, 2014a).

Although miRNA-phasiRNA pathways have been demonstrated in a number of studies across different plant species, mechanistic insights into the regulation of phasiRNA accumulation are not well described. In addition, many plant genomes contain hundreds of *NB-LRRs* and their regulatory relationships or interactions with a much smaller set of miRNAs are poorly understood. To investigate these aspects of phasiRNAs, we chose *Medicago truncatula* (hereafter, ‘Medicago’), a species that contains a rich set of miRNA-phasiRNA-*NB-LRRs* interactions (Zhai *et al.*, 2011). We modulated the expression of miRNA triggers in Medicago, either by overexpression or silencing, and tested these miRNAs in Arabidopsis, providing insights into miRNA-target interactions and phasiRNA biogenesis.

2.2 Results

2.2.1 Modulation of levels of miRNAs targeting *NB-LRRs* in *Medicago*

miR1507, miR2109, and miR2118a/b/c target *NB-LRRs* transcripts at sites of encoded, conserved motifs (Fei *et al.*, 2013). Therefore, each of these miRNAs is competent to target a subset of the *NB-LRR* family. However, because these miRNAs target multiple *NB-LRRs*, and *NB-LRR* transcript levels are usually low, the specificity of miRNAs in targeting individual *NB-LRRs* is challenging to define systematically using standard techniques (namely PARE combined with target prediction). To more closely examine the spectrum of targets for these five *Medicago* miRNAs and the resulting impact on phasiRNA biogenesis, we modulated their expression by both overexpression and down-regulation (described below). For overexpression, we constructed binary vectors in which the *MIRNA* precursor expression is driven by the cauliflower mosaic virus (CaMV) 35S promoter (*35S::MIRNA*) (Figure 5A). These vectors were transformed to produce hairy roots from *Medicago* seedlings using *Agrobacterium rhizogenes* strain ARqual (Chabaud *et al.*, 2006). DsRed was used to select successfully transformed roots via fluorescence microscopy (Figure 5B). Small RNA libraries were constructed from the transgenic hairy root tissue, and sequenced, for two biological replicates in each experiment. The data were processed, aligned to the *Medicago* genome (version 3.5) (Young *et al.*, 2011), and normalized to RP10M (reads per 10 million reads, or RP10M) for data analysis.

To assess mature miRNA levels in the hairy roots with the overexpression constructs, using the set of Medicago miRNAs from miRBase (release 21) (Kozomara and Griffiths-Jones, 2013), we selected the 100 most-abundant miRNAs and plotted the miRNA abundance levels, using the average of the two replicates for each construct, compared to an empty vector control (Figure 5C). In each case, the miRNA of interest was increased in abundance by ~30 to ~170 fold. Surprisingly, we observed that two other miRNAs were also upregulated sharply with the overexpression of miR2118b. An examination of the sequences of those two miRNAs showed that one was the “star” or passenger strand (miR2118b*) generated from the miR2118b precursor. The other is named “miR5261” in miRBase (Figures 5C-G), derived from a stem-loop structure together an upstream flanking sequence. However, the stem-loop structure contains a large bulge in the center, inconsistent with typical plant miRNA precursor structures (Figure 5E) (Meyers *et al.*, 2008), suggesting that miR5261 is not a real miRNA. Mapping of this miRNA sequence, as well as PARE data, clearly showed that “miR5261” is a phasiRNA triggered by miR2118b produced from locus Medtr8g012200, which encodes an NB-LRR (Figure 5E-F). Apart from these two small RNAs, the hairy roots showed high specificity and an absence of apparent indirect effects of miRNA overexpression.

Our next experiments aimed to reduce levels of the same set of five mature miRNAs, using target mimic constructs and Medicago hairy roots (Todesco *et al.*, 2010; Yan *et al.*, 2012). Short tandem target mimic (STTM) sequences were designed with one construct for suppression of each miRNA family (miR1507, miR2109, or

miR2118) and a fourth targeting all three families at once (Figure 6A). As before, we used *DsRed* as a reporter gene for the selection of transformed hairy roots using fluorescence microscopy. Again, small RNA sequencing data were generated and analyzed, replicated as above. We found that the miRNAs of interest were down-regulated in Our next experiments aimed to reduce levels of the same set of five mature miRNAs, using target mimic constructs and *Medicago* hairy roots (Todesco et al., 2010, Yan et al., 2012). Short tandem target mimic (STTM) sequences were designed with one construct for suppression of each miRNA family (miR1507, miR2109, or miR2118) and a fourth targeting all three families at once (Figure 6A). As before, we used *DsRed* as a reporter gene for the selection of transformed hairy roots using fluorescence microscopy. Again, small RNA sequencing data were generated and analyzed, replicated as above. We found that the miRNAs of interest were down-regulated in these target mimic lines, although the abundances of miR2118a/b/c in the MIM2118 line showed a less substantial reduction (Figure 6B). In the target mimic line ‘MIM-all’ designed to down-regulate all five miRNAs, four of the five miRNAs were reduced by ~4 to ~9 fold (with the exception of miR1507, which showed only a ~1.5-fold decrease), demonstrating that the STTM approach can, by increasing the number of tandem repeats, simultaneously down-regulate multiple unrelated miRNAs by sequestering these miRNAs via the expressed target mimic transcripts containing tandem miRNA binding sites. In subsequent sections, we describe the global impact on targets observed from these hairy-root lines of miR2118a/b/c in the MIM2118 line showed a less substantial reduction (Figure 6B).

In the target mimic line ‘MIM-all’ designed to down-regulate all five miRNAs, four of the five miRNAs were reduced by ~4 to ~9 fold (with the exception of miR1507, which showed only a ~1.5-fold decrease), demonstrating that the STTM approach can, by increasing the number of tandem repeats, simultaneously down-regulate multiple unrelated miRNAs by sequestering these miRNAs via the expressed target mimic transcripts containing tandem miRNA binding sites. In subsequent sections, we describe the global impact on targets observed from these hairy-root lines.

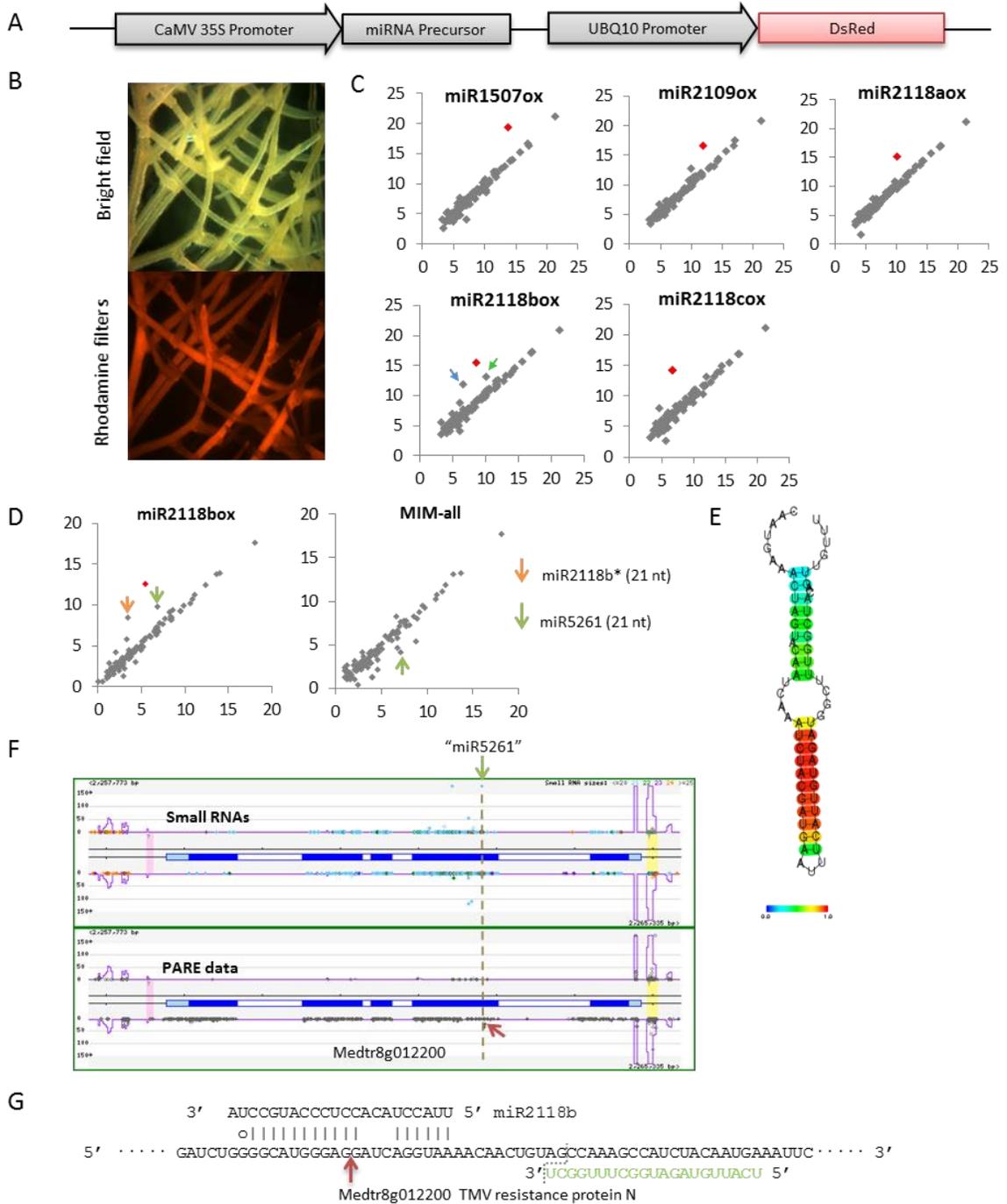


Figure 5. miRNA overexpression via hairy root transformation in *Medicago*.

A. Schema of the constructs used for miRNA overexpression experiments.

B. Transformed hairy roots were differentiated through fluorescent microscopy; only fluorescent roots were harvested. The images were taken with a dissecting scope; roots in the upper image were illuminated with white light, and the same roots using Rhodamine filters reveals fluorescence from transformation with the DsRed marker (lower image).

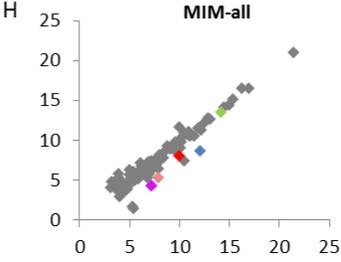
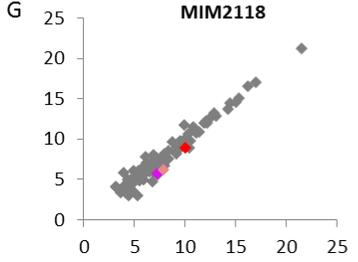
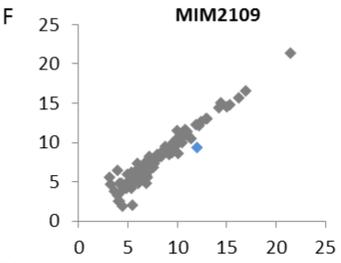
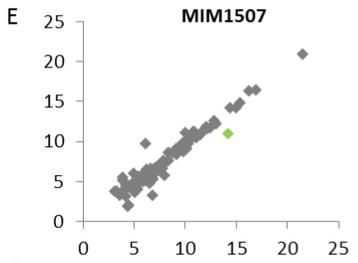
C. The abundance level (\log_2 value of reads per 10 million, $\log_2\text{RP10M}$) of each miRNA in miRNA overexpression lines; the X-axis is the level in an empty vector control, while the Y-axis is the level from an overexpression line. The same control data were used for each of the five plots. Each data point is the average of two biological replicates. The red color indicates the miRNA selected for overexpression in each experiment; grey dots indicate other miRNAs with abundances in the top 100. Blue and green arrowheads indicate two miRNAs that were associated with miR2118b overexpression.

D. Accumulation of miR5261 is increased when miR2118b is overexpressed (the same data as shown in panel C); miR5261 decreases in abundance when miR2118b is decreased in a target mimic line.

E. The secondary structure of the miRBase-annotated miR5261 precursor.

F. miR5261 maps to the coding region of gene Medtr8g012200. The red arrow indicates the cleavage site identified by PARE data.

G. The position that corresponds to miR5261, relative to the target site of miR2118b.



◆ miR1507
 ◆ miR2109
 ◆ miR2118a
 ◆ miR2118b
 ◆ miR2118c

Figure 6. Schema of target mimic constructs for suppression of miRNAs targeting *NB-LRRs* in *Medicago*.

The upper, colored schema shows the parts of the construct, with the promoters in light blue, the miRNA-aligned mimics in yellow, and the DsRed marker gene in red; in all constructs, the spacers in dark blue are 48 bp long. The bulge indicated in yellow signifies the anticipated bulge in the resulting target mRNA when paired with the miRNA; the alignments of those molecules are shown below each mimic, including the target site sequence and the mature miRNA sequence. The long dashes in the mature miRNA sequences indicate adjacent nucleotides opposite the three unpaired bases of the target mimic mRNA (shown in lowercase and red text). The construct design is essentially as described by Yan *et al.* (2012). The dot plots are as described for Figure 5; each plot shows the average miRNA levels for two biological replicates of the target-mimic constructs (Y-axis) compared to an empty vector control (X-axis). Values are the log₂ value of reads per 10 million (log₂RP10M). The same control data were used for each of the four target mimic experiments. Colored dots indicate specific miRNAs utilized in these experiments, with the legend shown at the bottom of the figure.

- A. Target mimics for miR1507 (construct called ‘MIM1507’).
- B. Target mimics for miR2109 (construct called ‘MIM2109’).
- C. Target mimics for miR2118a/b/c (construct called ‘MIM2118’).
- D. Target mimics for combined suppression of all miRNAs of interest in this study, including miR1507, miR2109, miR2118a/b/c (construct called ‘MIM-all’).
- E. miRNA levels in hairy roots expressing a target mimic construct against miR1507 (construct called ‘MIM1507’).
- F. miRNA levels in hairy roots expressing a target mimic construct against miR2109 (construct called ‘MIM2109’).
- G. miRNA levels in hairy roots expressing a target mimic construct against miR2118a/b/c (construct called ‘MIM2118’).
- H. miRNA levels in hairy roots expressing a target mimic construct against of all miRNAs utilized, including miR1507, miR2109, miR2118a/b/c (construct called ‘MIM-all’).

2.2.2 New targets of miRNAs, identified from miRNA overexpression in *Medicago*

In parallel to the production of the transgenic hairy root tissues, we prepared small RNA libraries from different tissues of wild type *Medicago*, including leaf, root, flower, seedling, and nodule, augmenting previously generated data for analysis of phased small RNAs (Zhai *et al.*, 2011). The new data from eight libraries comprised more than 80 million genome-matched reads, an eight-fold increase over our 2011 dataset; our aim was to saturate the set of known *PHAS* loci for *Medicago*. In total, we identified 220 *PHAS* loci in *Medicago*, almost doubling the set of known *Medicago* *PHAS* loci; of these, 134 were *NB-LRR* genes (Figure 7). We hypothesized that if miRNA triggers are a limiting factor, the abundance of phasiRNAs will increase when their miRNA triggers are overexpressed. To test this hypothesis, we calculated the abundance of all 21- and 22-mers at each *PHAS* locus in our miRNA overexpression lines (summing their abundances and comparing this to an empty vector control). Since some phasiRNAs map to multiple related genes, we ‘hits-normalized’ the data, dividing the abundance of each phasiRNA by its genome matches (aka ‘hits’); a small proportion of phasiRNAs matching ten or more genomic locations were excluded. Many *PHAS* loci demonstrated an increased abundance of phasiRNAs. Using the R package “OutlierD”, we next identified a total of 118 *PHAS* loci with significant up-regulation of phasiRNAs (Figure 8A). Among the 118 *PHAS* loci, 99 of them were *NB-LRRs*. We next assessed the opposite: whether a reduction in miRNA trigger levels

resulted in a decrease in phasiRNA accumulation, calculating in the target mimic lines the phasiRNA abundance at each *PHAS* loci. Compared to the control, in the STTM lines, phasiRNA abundances were decreased (Figure 9). In the “MIM-all”-expressing tissues, almost all verified target *PHAS* loci showed reduced phasiRNA abundances. In the individual STTM lines, phasiRNA abundances were most reduced at the same *PHAS* loci that were increased in the miRNA overexpression lines. We concluded that each miRNA targets a subset of *PHAS* loci, and alteration of the miRNA levels has a direct effect (correlated in the direction of change) on phasiRNA abundance at its targets.

The small RNA data clearly defined the extent of miRNA-target interactions for each overexpressed miRNA, showing that each miRNA has an almost mutually-exclusive scope of targets among the *PHAS* loci (Figure 8A). This indicates that most *NB-LRRs* are regulated by just a single miRNA. However, there were several *NB-LRRs* for which phasiRNA levels were increased by multiple miRNAs. For example, phasiRNAs generated from Medtr3g015550 (Figure 8B) in both miR1507 and miR2118b overexpression lines increased by ~3 and ~9 fold respectively, indicating interactions by both miR1507 and miR2118b at positions encoding Kinase-2 and P-loop motifs, respectively. Compared to miR2118b, additional mismatches exist in the complementary region between miR2118a and the target site, perhaps explaining why miR2118a fails to target this gene. In some cases, however, overexpression of several miRNAs had similar effects on phasiRNA abundances. For example, miR2118a and miR2118b both target Medtr3g033080 at the encoded P-loop and trigger robust

phasiRNAs when overexpressed (Figure 8C); the better pairing of miR2118b than miR2118a was reflected in the higher level of phasiRNAs induced upon overexpression. Finally, since a successful miRNA-target interaction is required for phasiRNA biogenesis, the overexpression data provide experimental support for 106 direct miRNA-target interactions (the number of induced *PHAS* loci from the five 22-nt miRNAs) with penalty scores as high as 8.5 (using the conventional TargetFinder scores), a substantially worse score than is typically considered allowable (Figure 10).

We prepared PARE libraries, which is a high-throughput method to identify the cleavage sites of mRNAs by sequencing the 5' ends of uncapped mRNAs (German *et al.*, 2008), from miRNA overexpression tissues of Medicago to validate miRNA-mRNA interactions. To calibrate the target scores used to analyze these PARE data, we utilized the “TargetFinder” scores mentioned above, which suggested that even a cutoff penalty score of 8.5 identifies valid interactions. In most cases, PARE read abundances at a predicted target site increased sharply in miRNA overexpression lines (compared to the empty vector control). For example, Medtr5g071850, targeted by miR1507, had a dramatic increase of the PARE signal from 65 to 311 RP20M (reads per 20 million reads) at the target site of miR1507 in the lines overexpressing this miRNA (Figure 11). These data were consistent with a quantitative or semi-quantitative measurement of the accumulation of mRNA cleavage products using PARE libraries. Combining miRNA targets acquired using the overexpression data with the PARE results, we were able to validate at least 162 targets of the five miRNAs. Because both increased mRNA cleavage and increased phasiRNAs were

measured in the presence of overexpressed miRNAs, we conclude that miRNA levels are limiting factors in the biogenesis of plant phasiRNAs.

We noticed that phasiRNAs are not uniformly abundant across a *PHAS* locus, i.e. some phasiRNAs accumulate to a high level, others at low levels. One explanation may be stabilization in AGO proteins, protecting them from degradation. Such differentially abundant phasiRNAs are present in wildtype tissues at all *PHAS* loci; in the miRNA overexpression tissues, the abundance of stabilized phasiRNAs would presumably increase, always higher than the background level of phasiRNAs at a *PHAS* locus. We performed an analysis to identify such putatively stabilized phasiRNAs using “edgeR” (Robinson *et al.*, 2010). This identified 2337 21- or 22-nt up-regulated small RNAs and 168 down-regulated small RNAs in the five miRNA overexpression lines ($p < 0.05$, FDR < 0.05) (Figure 12A-B). Of the 2337 small RNAs, ~57% (1331) mapped to the 220 *PHAS* loci (Figure 12C). The 5' terminal nucleotide of miRNAs sorting of miRNAs into specific AGO proteins (Mi *et al.*, 2008), so we analyzed this determines the nucleotide composition and found that the majority (~52%) of the 2337 have a 5' “U”, while among the 1331, the proportion of 5' “U” is slightly higher (56%) (Figure 12D). Finally, from our visual inspect of *PHAS* loci, we noticed that there were numerous 22-nt phasiRNAs, a class of variants for which the biogenesis is unknown. We examined the nucleotide frequencies 5' and 3' ends between 21- and 22-nt differentially expressed phasiRNAs (Figure 12E). We did not observe any significant difference at the 5' end nucleotide between 21- and 22-nt phasiRNAs. However, the proportion of 3' “U” ends is significantly higher for 22-nt

than 21-nt phasiRNAs. Knowing that data from Arabidopsis show that 5' "U" small RNAs are enriched in AGO1 (Mi *et al.*, 2008), while 5' "A" 21- to 22-nt small RNAs are enriched in AGO2, AGO4, AGO6, or AGO9 (Havecker *et al.*, 2010; McCue *et al.*, 2015; Mi *et al.*, 2008), it is likely that the differentially abundant phasiRNAs may be sorted to AGOs based on their 5' sequences. The function of the 3' variation in the 22-nt phasiRNAs remains to be determined, but could result from 3' tailing mediated by HESO1 or related nucleotidyl transferases (Ren *et al.*, 2012; Tu *et al.*, 2015; Wang *et al.*, 2015; Zhao *et al.*, 2012).

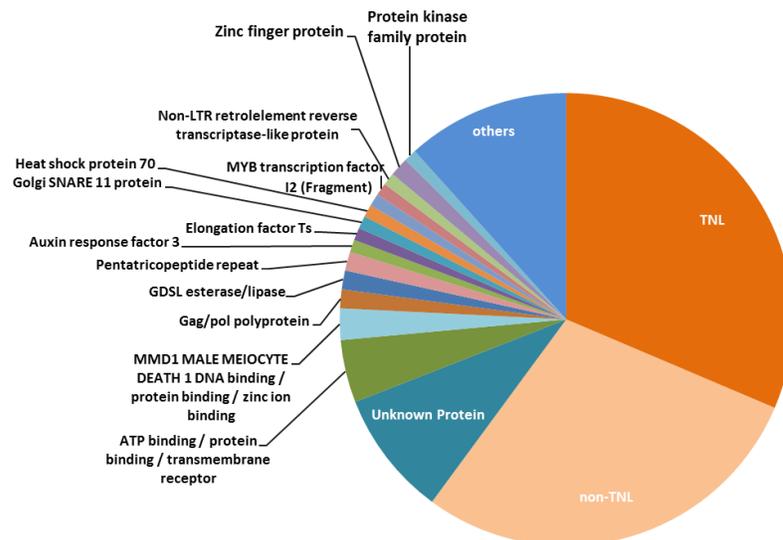


Figure 7. Genome-wide analysis of phased small RNAs in Medicago. The pie chart represents the proportion of the 220 loci identified as sources of phasiRNAs in *Medicago truncatula* that encode different families of proteins. Among these 220 loci, 134 encoded NB-LRRs (70 TIR-NB-LRRs, or TNLs; 64 non-TNLs, which have a varied N-terminal domain that is not a TIR-type). Loci predicted to generate phasiRNAs from noncoding RNAs (such as known *TAS* loci) were excluded from the figure.

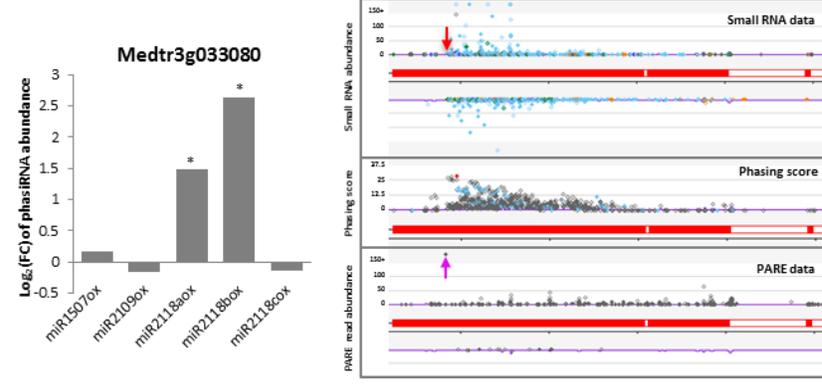
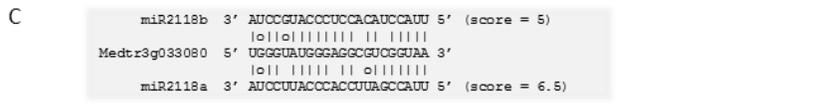
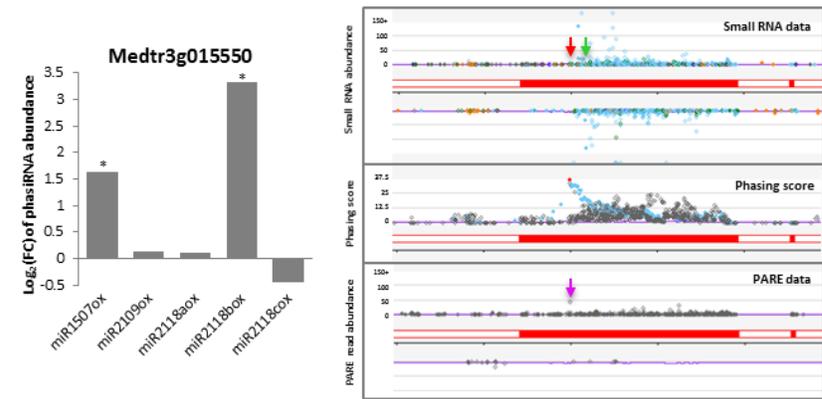
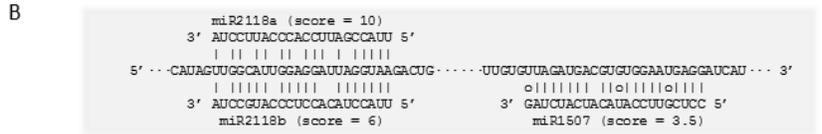
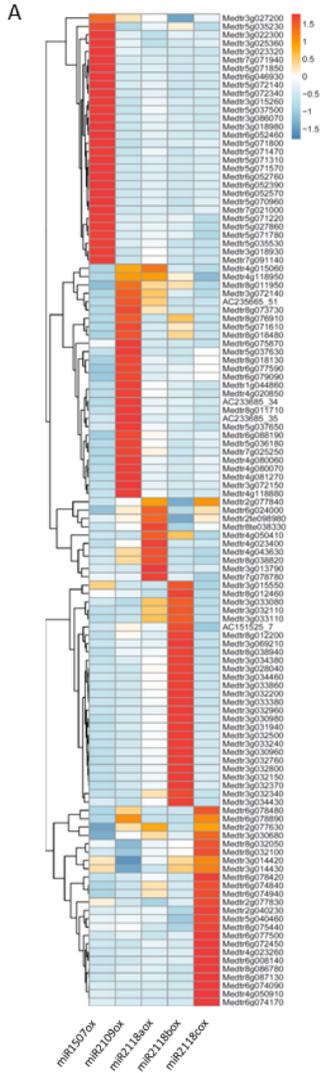


Figure 8. Fold change (FC) of summed phasiRNA abundances from each *PHAS* locus in miRNA overexpression lines.

A. Heat map representing the log₂ value fold change of phasiRNAs in each miRNA overexpression line compared to that of the empty vector control. In total, 118 *PHAS* loci were significantly upregulated by miRNA overexpression, of which 90 are *NB-LRRs*.

B. Medtr3g015550 is targeted by both miR2118b and miR1507 (but not miR2118a) at two different conserved motifs, as shown in the alignment at top, and supported by the increased phasiRNAs in the presence of elevated levels of miR2118b and miR1507 (histogram at lower left of the panel; asterisks indicate significance as measured by OutlierD in the R package). At the lower right of the panel, screenshots from our genome browser of this locus demonstrate the 21- and 22-nt phasiRNAs (light blue and green dots, respectively); red and green arrows indicate miR2118a/b and miR1507 target sites. The middle screenshot shows the calculation of the phasing score, in which each dot represents a "window" of ten cycles of 21 nt, with the score for the degree of phasing indicated on the Y axis (scores calculated approximately as described by Howell et al., 2007). Blue dots are the highest scoring windows and are positions in phase with the best score in this region (the single red dot in the image). Below this, the PARE data confirms cleavage at the miR2118 target site (purple arrow).

C. Medtr3g033080 is targeted by both miR2118a and miR2118b, at the same conserved motif. The three sections of this panel are as described for panel B.

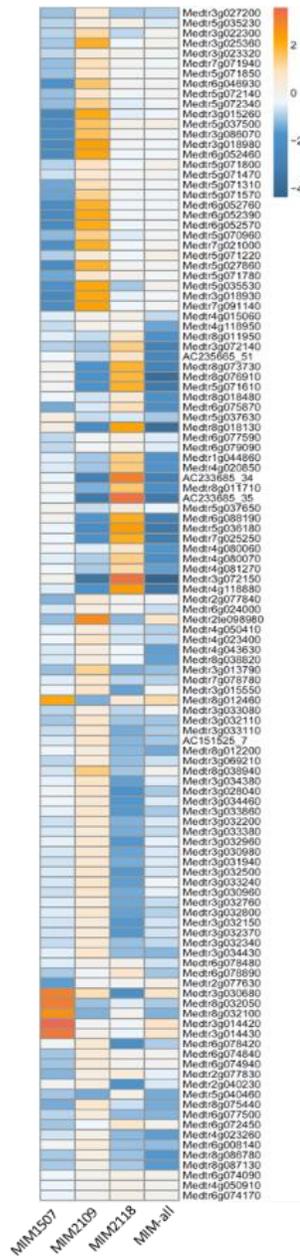


Figure 9. PhasiRNA levels in transgenic tissues of four target mimic constructs.

The rows are *PHAS* loci shown in the same order as Figure 8A, to make it easy to compare them. The data were rescaled from Figure 8A to reflect the reduction of phasiRNAs in the target mimic lines compared the miRNA overexpression lines.

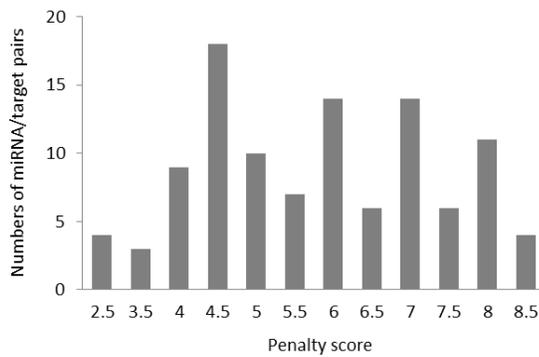


Figure 10. Numbers of miRNA-target pairs within each penalty scores. The histogram displays the count of miRNA-target interaction scores for the 106 *PHAS* loci that showed increased abundances in the miRNA overexpression tissues. There were no target sites among these loci with scores lower than 2.5. Scores were calculated using “TargetFinder”, as described in the methods section (n=106).

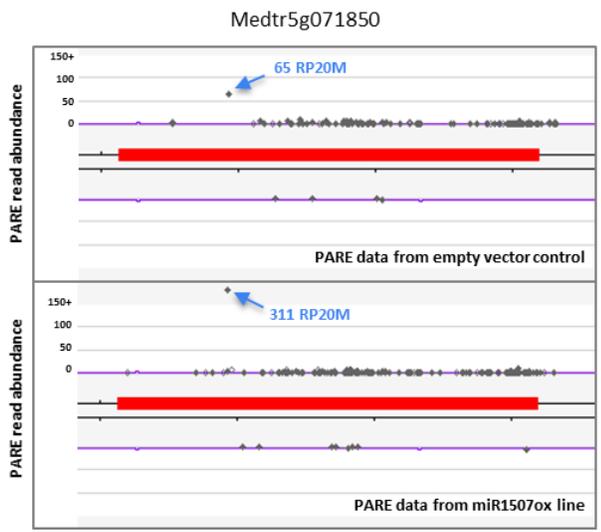


Figure 11. PARE signals increase by miRNA overexpression. One example of a miRNA target that shows increased PARE signal at the target site, perhaps due to increased levels of the miRNA leading to increased cleavage. Medtr5g071850, targeted by miR1507, had an increase of PARE signal from 65 to 311 RP20M at the target site of miR1507 in the lines overexpressing this miRNA. The data represent a single PARE library in each case.

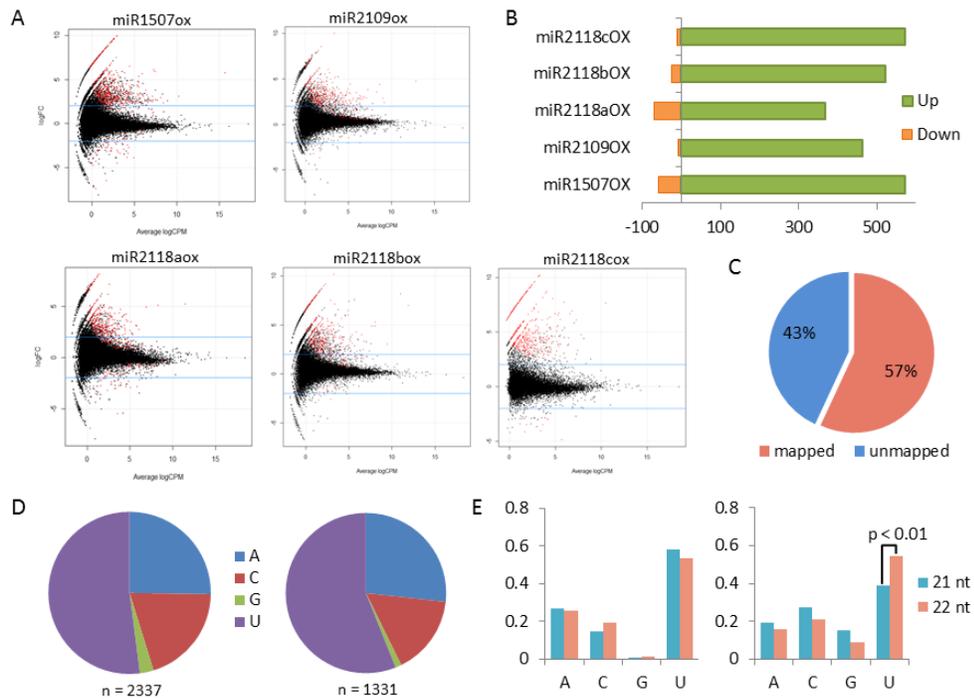


Figure 12. Differentially accumulating phasiRNAs in Medicago hairy roots overexpressing miRNAs.

A. MA plot of differentially accumulating small RNAs in each data set. Red dots indicate small RNAs with significantly different levels ($p < 0.05$, $FDR < 0.05$).

B. Numbers of significantly up- and down-regulated siRNAs ($p < 0.05$, $FDR < 0.05$). In total, 2337 siRNAs were upregulated, while 168 were down-regulated.

C. The proportion of differentially accumulating siRNAs that either map to *PHAS* loci (57%), or map elsewhere in the genome (43%).

D. The proportion of each nucleotide found at the 5' end of the set of differentially expressed small RNAs. At left, the 2337 upregulated small RNAs in the tissues overexpressing miRNAs; 52% have a 5' uracil. At right, the subset of the 2337 small RNAs that mapped to 220 *PHAS* loci (1331 in total) have a slightly larger proportion (56%) of 5' uracil.

E. The frequency of the 5' nucleotide of 21- and 22-nt small RNAs mapped to *PHAS* loci; at right, the frequency of the 3' nucleotide for the same set of small RNAs (Z test, $p < 0.01$).

2.2.3 Biogenesis of phasiRNAs determined by miRNA-target pairing and expression level of *PHAS* transcript

We next investigated other factors that determine phasiRNA accumulation in addition to miRNA levels. For example, knowing that base pairing between miRNAs and their targets is crucial in RNA silencing (Filipowicz *et al.*, 2008), we assessed base pairing between the miRNA trigger and *PHAS* target sequence. We calculated the penalty scores for miRNA-target interactions (Fahlgren and Carrington, 2010), using the *PHAS* loci up-regulated in our miRNA overexpression data, and we compared the phasiRNA abundance against the target penalty scores. We observed that phasiRNAs abundances were higher at loci with lower miRNA-target penalty scores (Figure 13A). Rules for the calculation of penalty scores were proposed based on the theory that different positions of miRNA/target pairing may have varied importance in successful targeting. A more simplistic calculation focusing on just the number of mismatches in the pairing also showed an inverse correlation to phasiRNA induction (Figure 13B). These results indicate that phasiRNA accumulation at *PHAS* loci is sensitive to the pairing with the miRNA trigger.

In addition to the role of miRNA triggers, we examined the role of the *PHAS* precursor transcripts in phasiRNA production. Using RNA-seq data from the empty vector control, we calculated the abundance of all 220 *PHAS* loci, and compared these to the overall phasiRNA abundance at each *PHAS* loci. The comparison demonstrated a good linear correlation between these two values, with a correlation coefficient of 0.447 (Figure 13C). The correlation coefficient was also calculated using only the

target genes of miR1507, miR2109, and miR2118a/b/c, showing a similarly strong correlation between *PHAS* loci expression level and phasiRNA abundance (Figure 13D). We concluded that more abundant phasiRNAs tend to generate from transcripts of *PHAS* loci that are expressed at higher levels.

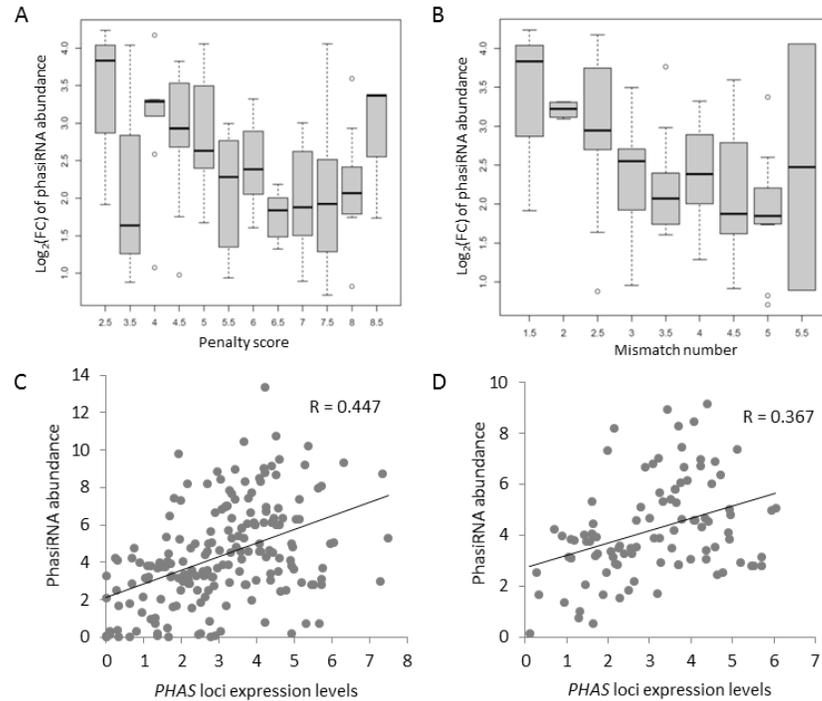


Figure 13. Factors that affect phasiRNA production.

As measured in the tissues overexpressing miRNAs, the change in phasiRNA abundance levels is related to both the penalty score (A) and the number of mismatches (B) of miRNA/target pairing ($n = 106$).

PhasiRNA abundance is correlated to the abundance of transcripts from which they are generated.

(C) The abundance of all 220 Medicago *PHAS* loci as measured by RNA-seq. Abundances were calculated as \log_2 values of RPKM (reads per kilobase per million reads) and plotted against the \log_2 RPM values of overall phasiRNA abundance.

(D) The subset of *PHAS* loci targeted by miR1507, miR2109, and miR2118a/b/c was also assessed. Abundances were calculated and plotted as in panel C.

2.2.4 PhasiRNA production is determined by 3' pairing of a miRNA and its target

Our miRNA overexpression experiments generated an extensive list of miRNAs and their cognate targets validated via production of phasiRNAs; we decided to use this list to assess the concordance with published empirical observations about miRNA-target interactions (Fahlgren and Carrington, 2010). We calculated the frequencies of nucleotide pairing between miRNAs and their targets, and we found that generally there was better pairing across the 3' half of the miRNA-target pair than the 5' half (Figure 14A). Similarly, the pairing of the 3'-terminal nucleotides of the miRNA includes fewer wobbles plus mismatches compared to 5'-terminal nucleotides (Figure 14A). Both of these observations were not expected under the canonical rules for miRNA-target interactions (Fahlgren and Carrington, 2010). We also observed that the 10th and 19th positions were perfectly paired in every case (Figure 14A). We focused in particular on the 19th to 22nd positions, since the 19th position was well-paired, and the addition of a 22nd nucleotide can make a miRNA competent to trigger phasiRNA production, perhaps exemplified by miR171a in a *hen1* background. We hypothesized that good target pairing across the 3' end of the miRNA might be important for phasiRNA production.

To test this hypothesis, we generated constructs for transient assays that allowed us to vary miRNA-target interactions and measure the impact on phasiRNA production. We used the Arabidopsis miR173-*TAS1c* system because it's not present in tobacco; the miRNA and mRNA target were co-expressed (as separate constructs)

by tobacco leaf infiltration of *Agrobacterium tumefaciens*, and RNA gel blotting was employed to measure a particularly abundant *TAS1c* phasiRNA, called 3'D3(+). A schematic diagram of this locus is shown in Figure 14B. We then assayed 3'D3(+) levels from constructs each containing mutations in one of the four 3' terminal nucleotides (Figure 14C); experiments were performed in triplicate, with consistent results. Mutations at the 19th, 20th, and 22nd positions dramatically reduced the level of 3'D3(+) (Figure 14C). In contrast, 3'D3(+) levels was only slightly affected by the single mutation of the 21st position. Double mutations at the 19th/20th or 21st/22nd positions abolished the production of phasiRNAs, as did a triple mutation (20th/21st/22nd positions; Figure 14C). Mutation of the 10th position only slightly affected the phasiRNA production from *TAS1c*. Our results are consistent with a previous study showing that the 3' *TAS2* cleavage product is subject to degradation when 3' terminal nucleotide pairing is reduced (Yoshikawa *et al.*, 2013). In order to cross-check these observations with a different phasiRNA-triggering miRNA, we replaced the miR173 target site of *TAS1c* with a natural target site from gene Medtr3g034460 that pairs well with miR2118b. We co-expressed the swapped *TAS1c* vector together with Medicago miR2118b and repeated the experiment, again in tobacco (Figure 14C). This showed again that phasiRNA 3'D3(+) was dramatically reduced when the 22nd position was mutated. However, phasiRNA production was only slightly reduced for single mutations at the 19th, 20th, or 21st positions, suggesting a difference with miR173. Double mutations at the 19th/20th substantially diminished 3'D3(+) levels, while mutated 21st/22nd positions had less of an impact with miR2118b

than in the miR173 experiment. Consistent with miR173/*TAS1c*, the impact of the single mutation at the 10th position was not discernable in 3'D3(+) levels. Our results indicate that the pairing between 3' end of 22-nt miRNAs and their targets, especially the 3' terminal nucleotide of the miRNA, is important for phasiRNA production. In contrast, the 10th position, conventionally known as a crucial position for target cleavage usually occurring between 10th and 11th position of a miRNA (Liu *et al.*, 2014b), is dispensable for triggering phasiRNA production.

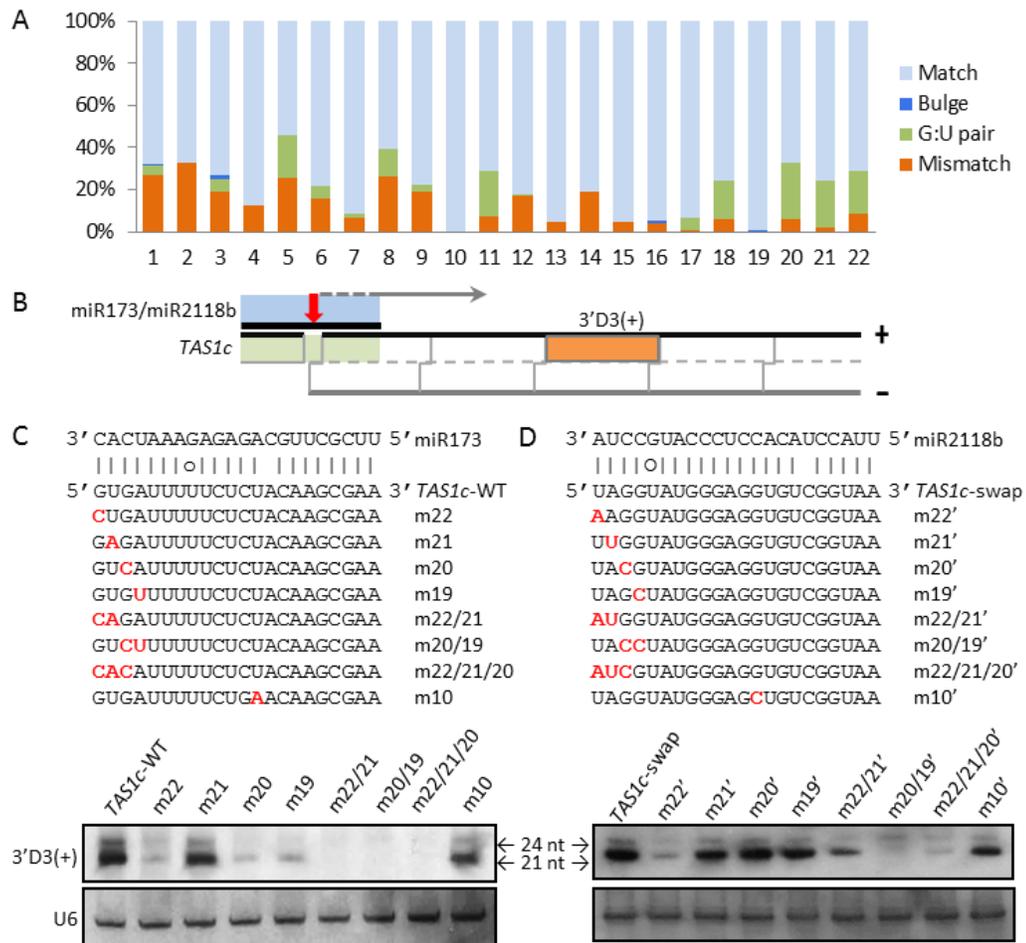


Figure 14. Pairing of 22-nt miRNAs with their targets, and its role in phasiRNA production.

A. Frequencies of nucleotide pairing between miRNAs and their targets that result in production of secondary siRNAs. Pairing was measured at each position from the 5' end (position #1) to the 3' end (position #22) of the miRNA. G:U pairing is considered a 'wobble' or partial mismatch. (n = 106)

B. A schematic diagram of the *TAS1c* transcript used in these experiments. The positions of the miR173 or miR2118 target sites are indicated, with the miRNA shown on top for illustration purposes, and the cleavage site marked with the red arrow. The location of phasiRNA 3'D3(+) is indicated in orange.

C. Above, the mature miR173 sequence aligned with the variants of *TAS1c* target site sequences that were utilized for transient assays; note that the miRNA is shown on top, in 3' to 5' orientation merely for illustrative purposes. Each line shows the sequence of the altered miR173 target sequence contained in the *TAS1c* variant, with the name of this construct at the right, indicating the position of the mutated base relative to the miRNA (i.e. m22 = a mutated base in the 22nd position of the miRNA). Below, an RNA gel blot showing (upper image) the levels of the 3'D3(+) phasiRNA for each *TAS1c* variant co-expressed with miR173 in leaves of *Nicotiana benthamiana*, via *Agrobacterium* infiltration; each lane contains 20 ug of total RNA combined from several leaves. The lower image shows the same blot hybridized with U6, as a loading control.

D. The sections are as in panel C, but the target site of miR173 in *TAS1c* was replaced by a target site for the Medicago miR2118b (from gene Medtr3g034460); nucleotides were mutated at the same positions as in panel B.

2.2.5 The role of target site flanking sequences in the function of miRNAs

NB-LRRs are present in the genomes of a wide range of species in land plants, but with variation reflecting their ancient age plus natural selection (often from pathogens) (Yue *et al.*, 2012). Our Medicago miRNAs of interest target regions of transcripts encoding motifs conserved across all plant *NB-LRRs* (Zhai *et al.*, 2011). In *Arabidopsis*, there are two known triggers of *phasi-NB-LRRs* (*pNLs*), including miR472 and miR825*, and few *pNL* loci (Chen *et al.*, 2010; Howell *et al.*, 2007). This relatively paucity of *pNLs* makes *Arabidopsis* a good ectopic system in which to study the activities of Medicago miR1507, miR2109, and miR2118. Using target prediction, these miRNAs all have the potential to target motifs found in *Arabidopsis NB-LRRs*, and the range of target scores is not obviously different from the same miRNA targets in Medicago. We made stable transgenic lines of *Arabidopsis* with the same constructs used for hairy root transformation in Medicago; small RNA libraries were prepared from leaf tissue of T1 (hemizygous) plants (two replicates, from independent T1 plants). These five miRNAs accumulated in *Arabidopsis* to robust levels (Figure 15A). We surveyed 159 *NB-LRRs* in *Arabidopsis* (Guo *et al.*, 2011) to identify whether phasiRNAs were produced, summing the abundance of 21- and 22-nt small RNAs from these genes. We identified only six *NB-LRRs* with increased phasiRNA levels across the five overexpression lines, in addition to three endogenous *pNLs* at which the small RNA levels were unchanged (Figure 15B, and two examples in Figure 16). Of these six, five were in the miR1507 overexpression lines, and one in both

miR2118a and miR2118c overexpression lines (Figure 15B). The miR1507 target genes included two genes with known roles in disease resistance, *RPS2* and *RPP8* (Bent *et al.*, 1994; Cooley *et al.*, 2000); transcripts from both were cleaved at the predicted miR1507 target site (Figure 15C). Given the large number of potential targets, the percentage of *pNLs* in the transgenic materials was unexpectedly low (< 4%).

We were curious why the Medicago miRNAs largely failed to trigger phasiRNA production from many *NB-LRRs* in Arabidopsis, despite the large number of targets predicted at scores appropriate for cleavage to occur. For example, miR2109 potentially targets numerous *NB-LRRs* in Arabidopsis, with scores comparable to targets validated in Medicago, and including high levels of complementarity in the 3' end of the miRNA (Figure 17). We utilized public RNA-seq data to measure *NB-LRR* transcript levels in Arabidopsis leaves, and found many abundant transcripts with good target scores yet produce no measured phasiRNAs in the presence of the Medicago miRNAs (Figure 18). From this, we infer that there are factors other than miRNA targeting or target abundance that impact phasiRNA production. One possibility is the nature of the miRNA-target pairing; the most abundant *NB-LRR* was At3g50950 (Figure 18), which has predicted interactions with miR1507 and miR2118b, yet no phasiRNAs in their presence.

Previous work in animals has shown that target site (including the 17 nt upstream and 13 nt downstream flanking sequences) accessibility is important for successful miRNA-target interactions during RNA silencing (Kertesz *et al.*, 2007),

with reduced secondary structure in the mRNA flanking miRNA target sites, in *Drosophila*, *C. elegans*, and *Arabidopsis* (Li *et al.*, 2012c; Li *et al.*, 2012d). To evaluate the role of RNA structure in our observations, we calculated the interaction energy between the miRNA and target sequences of the same regions in both *Arabidopsis* and *Medicago* using the Vienna RNA package (Gruber *et al.*, 2008). This projected a significantly higher interaction energy between miR2109 and its predicted targets in *Arabidopsis* than for validated targets in *Medicago* (Figure 15D). In contrast, for validated targets, we performed the same analysis for miR1507 and its validated targets: five in *Arabidopsis* (from Figure 15B) and 30 in *Medicago*. In this case, there was no significant difference in interaction energies (Figure 15E), indicating potentially similar accessibilities of these miR1507-targeted *NB-LRRs* in *Arabidopsis* and *Medicago*. Therefore, target accessibility may be a primary impediment for ectopic or *trans*-species function of miRNAs.

Figure 15. Ectopic expression of Medicago miRNAs in Arabidopsis.

A. Abundances of five miRNAs in overexpression lines of Arabidopsis, measured by sequencing. The Y-axis denotes the log₂ values of the average abundances of each miRNA from two independent T1 hemizygous lines (log₂ value of reads per million, log₂RPM); the X-axis indicates the miRNA overexpression line. Thus each bar shows the level of a particular miRNA in the line in which it is overexpressed. No control is shown as these miRNAs are all absent from wildtype Arabidopsis.

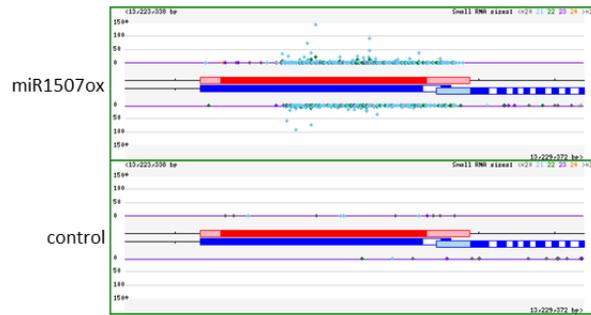
B. PhasiRNA abundance at Arabidopsis *PHAS* loci. Above, three *NB-LRRs* in wildtype Arabidopsis generate substantial levels of phasiRNAs (AT5G38850 phasiRNAs were described by Howell *et al.*, 2007), but these are unchanged in the tissues overexpressing Medicago miRNAs. Below, six Arabidopsis genes demonstrated increased levels of phasiRNAs resulting from overexpression of Medicago miRNAs. Columns indicate different transgenic lines (T1 hemizygous plants); abundances were calculated using the average of two replicates. Asterisks indicate loci significantly up-regulated in those lines compared to the empty vector control.

C. 5' RACE validation of miR1507 targets in Arabidopsis. At left, the PCR products of the RACE reaction; arrows indicate the expected band size. At right, the alignment of miR1507 and *RPS2* or *RPP8*; numbers above with arrowheads indicate the proportion of sequenced clones out of the total that corresponded to a 5' cleavage product at the designated position.

D. Interaction energy between miR2109 and its potential targets in Arabidopsis and Medicago. A lower interaction energy indicates a more stable pairing of the miRNA with its target. ** indicates the support for a difference between the two species, at a P-value < 0.01.

E. Interaction energy between miR1507 and its potential targets in Arabidopsis and Medicago.

A AT4G26090; annotated as “NB-ARC domain-containing disease resistance protein (RPS2)”



B AT5G36930; annotated as “Disease resistance protein (TIR-NBS-LRR class) family”

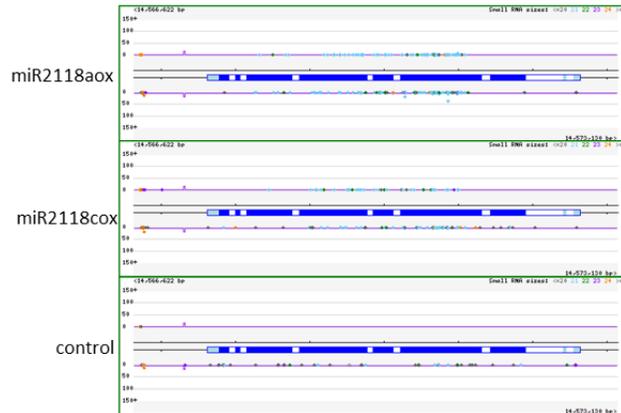


Figure 16. PhasiRNAs are triggered by ectopic expression of Medicago miRNAs in Arabidopsis.

Shown are two of the six Arabidopsis *NB-LRRs* demonstrating increased levels of phasiRNAs in the presence of Medicago miRNAs. These are examples of the impact of the Medicago miRNAs on Arabidopsis targets. In both panels, the upper image(s) shows a screenshot of our genome browser for the miRNA overexpression line (indicated at left), whereas the lower image shows the image for the empty vector control.

A. Increased phasiRNAs from AT4G26090 in the miR1507 overexpression line (upper panel).

B. Increased phasiRNAs from AT5G36930 in either the miR2118a overexpression line (upper panel) or the miR2118c overexpression line (middle panel).

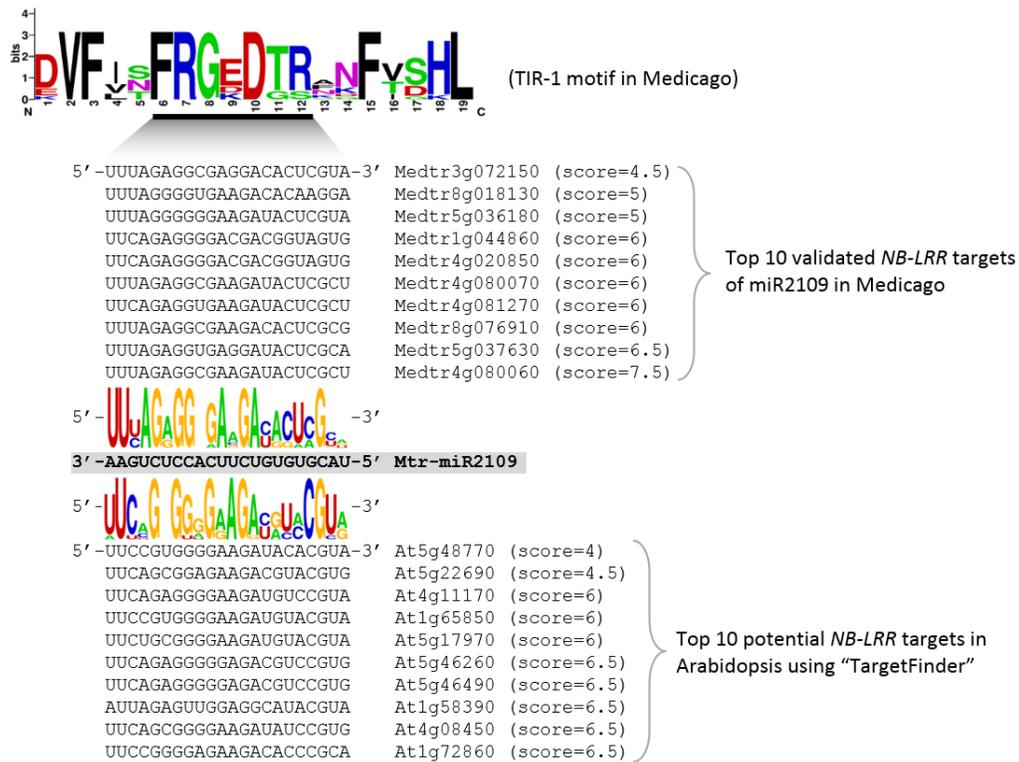


Figure 17. Validated and potential targets in Medicago and Arabidopsis for mtr-miR2109.

At top, the WebLogo graphic of the conservation of protein sequences for the top 10 *NB-LRRs* with increased phasiRNAs in the hairy-root overexpression experiments in Medicago using miR2109; the gene identifiers are listed just below, along with the sequences of the mRNA target sites in these genes. The consensus amino acids encoded at the target site are underlined in black, with numbers indicating the position within the larger TIR-1 motif (Meyers *et al.*, 1999). Below the Medicago *NB-LRR* list is shown the conservation of nucleotides at these sites (also from WebLogo), aligned to miR2109 (the absence of a letter indicates a complete lack of conservation). Below this, the alignment of miR2109 with the conserved nucleotides from the top 10 potential *NB-LRR* targets in Arabidopsis, selected from a rank order of all targets identified using "TargetFinder".

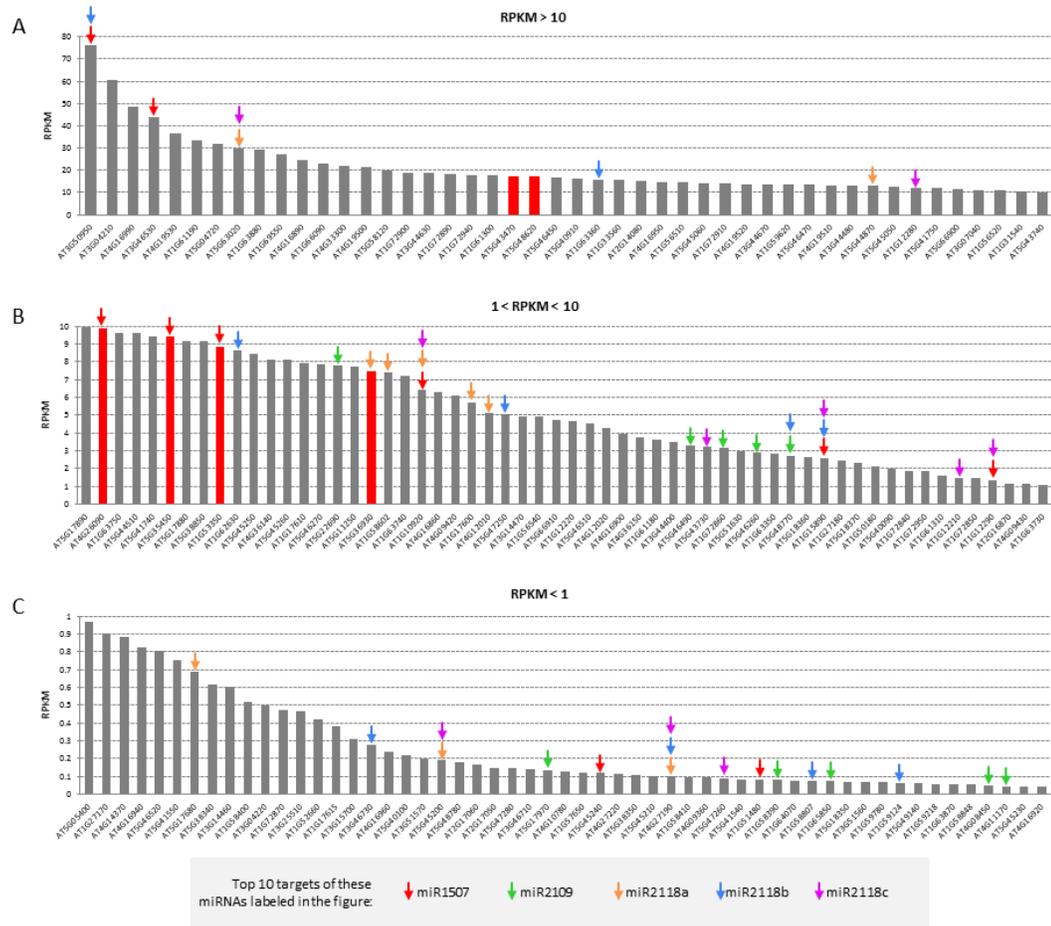


Figure 18. Abundance levels of the 159 NB-LRR genes in Arabidopsis three-week old leaves.

Using published RNA-seq data from Genbank (GEO ID: GSE36129), we extracted the abundance levels of all 159 Arabidopsis *NB-LRRs*, in Reads per Kilobase per Million (RPKM), and rank ordered them from highest to lowest abundance. The top 10 targets (ranked from best match to worst) of the five Medicago miRNAs tested in Arabidopsis are indicated with colored arrows (see legend below). For display purposes, the 159 genes were split into three panels: A. Genes exceeding 10 RPKM; B. Genes with counts between 1 and 10 RPKM; C. Genes with counts below 1 RPKM. In each panel, bars in red indicate the genes demonstrating increased phasiRNA levels from Medicago miRNA overexpression.

2.3 Discussion

Small RNAs and the associated RNA silencing machinery play crucial roles in many biological processes in plants. miRNAs function to direct silencing of their targets, and in some cases trigger the production of phasiRNAs from their transcripts in an expansion of the post-transcriptional regulatory network. In this study, we focused on *NB-LRR*-derived phasiRNAs modulated by several Medicago miRNAs; *NB-LRRs* constitute the majority of plant disease resistance genes, and comprise one of the largest gene families in plant genomes. Their transcriptional regulation and even the phenotypic or functional consequences of phasiRNA production are poorly understood, although presumably play important roles in the modulation of plant immune responses. PhasiRNAs generated from *NB-LRRs* are abundant in many plant species, including Medicago, so in this work, we investigated the factors that influence phasiRNA production, including miRNA-target interactions.

In order to observe the relationship between miRNAs and phasiRNAs, we modulated the expression of these miRNA triggers in transgenic hairy roots of Medicago, focusing on *pNLs* and their five Medicago triggers. These experiments addressed several limitations of analyses in wildtype tissues, including the following: (1) Members of a single miRNA family (miR2118 in this case) may be quite similar, making it difficult to assess whether a *PHAS* locus is triggered as a consequence of just one member of the family (which one?), or multiple members of the family (which ones?). Similarly, with distinct miRNA families potentially targeting the same

gene at different sites (miR1507, miR2109, or miR2118, in this case), it is not clear whether they specialize. (2) *NB-LRRs* are a difficult group of genes for which to validate cleavage, as their low expression levels make it difficult to confirm cleavage via PARE; the production of secondary siRNAs from miRNA targets may be a more sensitive “readout” of successful cleavage. Due to the targeting of encoded motifs conserved across the entire gene family, prior studies predicted miRNAs may interact with numerous *NB-LRRs* (Shivaprasad *et al.*, 2012). However, our experiments altering miRNA levels unexpectedly demonstrated that Medicago *NB-LRRs* tend to have a preferential interaction with a much smaller number of miRNAs than prediction programs would anticipate, even within the closely-related miR2118 family. In many cases, PARE signals in the miRNA overexpression lines, perhaps a semi-quantitative measurement, were greatly strengthened compared to the control regardless of the relatively low expression levels of *NB-LRRs*, while the specificity of the PARE data confirmed activity of individual miRNAs at the *NB-LRRs* target sites. We found it remarkable that the spectrum of targets, as measured by phasiRNAs, largely differed for relatively similar miRNAs. This highly restricted targeting perhaps provides an explanation for the proliferation of miRNAs targeting a single gene family, particularly pronounced for *NB-LRRs*, but perhaps also including those targeting PPRs in other genomes (Xia *et al.*, 2013). We could speculate that highly similar miRNAs may be co-expressed yet interact with distinct, non-overlapping subsets of their predicted targets.

From our bidirectional modulation of miRNAs (i.e. overexpressed or suppressed), we also observed that (1) better-paired miRNAs and targets yielded more robust levels of phasiRNAs, and (2) phasiRNA levels are largely controlled by the abundance of corresponding miRNA triggers; we then further explored other factors that may affect phasiRNA production or accumulation. For example, our analysis of endogenous levels of *PHAS* loci precursors demonstrated a strong correlation with phasiRNA levels. Thus, precursor variation may explain why the overall abundance of phasiRNAs varies substantially across *PHAS* loci, with some loci generating higher levels than others. Consistent with this, a recent analysis of hundreds of reproductive *PHAS* loci from maize anthers showed that these abundances can vary across orders of magnitude yet are highly reproducible (Zhai *et al.*, 2015). Among our other observations of factors contributing to different levels of phasiRNA production, we found cases of miRNAs targeting *NB-LRRs* with high penalty scores, tolerating more mismatches than typically considered permissible. In testing the role of pairing in phasiRNA production, we found that miRNA-mRNA pairing indeed affects phasiRNA production, and poor interactions reduced phasiRNA levels. In summary, we believe that the level of secondary siRNAs at a given *PHAS* locus reflects a complex yet reproducible interplay of trigger-target interactions, and both precursor and miRNA trigger abundances.

The large number of *NB-LRR* targets identified in this study allowed us to analyze miRNA-target interactions in detail. Such interactions have been analyzed by other methods, such as the Axtell lab's dual-luciferase-based reporter system (Liu *et*

al., 2014b), but our approach relied on measurements of phasiRNAs produced downstream of targeting, via an endogenous pathway, and thus we focused on interactions competent to trigger phasiRNA production. We found that nucleotides at the 3'-most positions of these 22-nt miRNAs are well-paired with their targets. Our experiments using transient expression assays in *N. benthamiana* showed that the 3' terminal nucleotides are important for phasiRNA production, as introduction of a single mismatch largely diminished phasiRNA production. This may explain why miR171a in *hen1* triggers phasiRNA production from its target (Zhai *et al.*, 2013); monouridylation of miR171a by URT1 creates an additional 22nd "U" paired with the "G" on the target transcript forming a G-U wobble, which is highly tolerated for triggers of phasiRNAs (Tu *et al.*, 2015). In contrast, we showed that the mismatch introduced at the 10th position has a minimal impact on phasiRNA production, inconsistent with data prior observations showing the critical importance of the central positions in miRNA targeting (Liu *et al.*, 2014b). However, they also observed (and left unexplained) a discrepancy between their analysis of 21-nt miR164 and 22-nt miR173: mismatches at the central 9th or 10th positions eliminate cleavage by miR164 in their transgene system, but miR173 in wildtype plants successfully cleaves and triggers phasiRNA production with either a 9th or 9th/10th position mismatch (Allen *et al.*, 2005). In fact, in our analysis, 9th position mismatches are relatively common for phasiRNA triggers. Consistent with these results, a recent study on miR159 showed that central mismatches are allowable in the silencing of target transcripts in *Arabidopsis*, suggesting that miRNAs may not require central complementarity for

their activity in plants (Li *et al.*, 2014a). In addition, our data suggest that pairing of the 3' terminal nucleotide of 22-nt miRNAs could induce a change, perhaps the conformation of AGO1, resulting in recruitment of the components for second strand synthesis (i.e. SGS3 and RDR6) and subsequent further processing by DCL4.

The widespread activity of a few miRNAs in triggering *pNLs* in Medicago led us to ask if these triggers function ectopically, in Arabidopsis, which has numerous predicted *NB-LRR* target sites for these miRNAs. Few Arabidopsis *NB-LRRs* produced phasiRNAs, however, an observation not attributable to (1) *NB-LRR* transcript abundance, (2) miRNA trigger abundance, or (3) miRNA-mRNA complementarity. We hypothesize that one reasonable explanation for the lack of successful interactions could be target site accessibility, supported by our calculation of interaction energy between miRNAs and the larger context of their complementary sequences in Arabidopsis (i.e. including regions flanking the target). This is consistent with prior work indicating that the flanking sequence around miRNA target sites shows evidence of selection to increase target accessibility (Gu *et al.*, 2012), and with analyses of mRNA secondary structure in *Drosophila*, *C. elegans*, and Arabidopsis (Li *et al.*, 2012c; Li *et al.*, 2012d). Our work also is consistent with an observation that the flanking sequence of miR159 target site impacts its efficacy in RNA silencing in Arabidopsis (Li *et al.*, 2014a). It is likely that *NB-LRRs* in Medicago have co-evolved with these miRNAs, potentially exerting selection pressure on sequences flanking target sites for accessibility by miRNAs. One implication of this observation is that optimal artificial miRNA (amiRNA) design should select regions containing low

interaction energy (for example, below -20kcal/mol) for flanking sequences of target sites.

2.4 Conclusions and Future Directions

The model legume *Medicago truncatula* generates phasiRNAs from many *PHAS* loci triggered by 22-nt miRNA. We investigated the mechanism of their biogenesis by modulating their miRNAs triggers. Sequencing data from transgenic tissues of *Medicago* showing that the abundance of phasiRNAs correlates with the levels of both miRNA triggers and the target mRNA transcripts. We identified sets of phasiRNAs and *PHAS* loci that predominantly and substantially increased in response to the overexpression of miRNA triggers. Using the validated targets from miRNA overexpression experiments, we found that in the miRNA-mRNA target pairing, the 3' terminal nucleotide (the 22nd position), but not the central position, is important for phasiRNA production. Mutating the single 3' terminal nucleotide dramatically diminishes phasiRNA production. Ectopic expression of *Medicago NB-LRR*-targeting miRNAs in *Arabidopsis* showed that only a few *NB-LRRs* are capable of phasiRNA production; our data indicate that this was likely due to target inaccessibility determined by sequences flanking target sites. These results suggest that target accessibility is an important component in miRNA-target interactions that could be utilized in target prediction, and the evolution of mRNA sequences flanking miRNA target sites may be impacted.

Arabidopsis and Medicago demonstrate divergent histories in their utilization of miRNAs; these paths may have differentially shaped their *NB-LRRs*. The encoded NB-LRR motifs targeted by a number of plant miRNAs are highly conserved and strongly influenced by selection on the resultant proteins; a requirement for target accessibility may have a weak but still important impact across a broader region of target genes. Arabidopsis has only two miRNAs (miR472 and miR825*) known to target just three *NB-LRRs* (Chen *et al.*, 2010; Howell *et al.*, 2007), whereas Medicago and many other plants have many more miRNAs targeting even hundreds of *NB-LRRs*. While miR825 is a very young miRNA (Bologna *et al.*, 2013), miR472 is a member of the ancient miR482/miR2118 family that has targeted *NB-LRRs* perhaps since the gymnosperms (Zhai *et al.*, 2011). Thus we may infer that even though Arabidopsis and Medicago genomes both include one or a few members of the miR472/miR482/miR2118 superfamily, they have utilized these miRNAs in different ways: *NB-LRRs* in Medicago are broadly accessible targets of the family, whereas in Arabidopsis, targeting seems restricted. The functional relevance of this divergence will need to be determined in future studies.

2.5 Materials and methods

Plant materials

Total RNA representing 2-month-old leaves, 3-week-old roots, flowers of mixed stages, seedlings collected 48 h post-germination, and 14 and 16-day old nodules were used for construction of small RNA libraries were collected from wild type *Medicago truncatula* Jemalong A17. The growth conditions of plants were described in Zhai et al. (2011). *A. thaliana* Col-0 and *N. benthamiana* plants were grown in the growth chamber with the light cycle of 16 h light / 8 h dark at 22 to 23°C.

Vector construction

MIRNA precursor sequences for miR1507, miR2109, and miR2118a/b/c were cloned by PCR from *Medicago* genomic DNA, confirmed by sequencing, and inserted into Gateway pDONR Vectors (Invitrogen, Carlsbad, CA). LR reactions were performed to transfer the miRNA precursor fragment into binary vector pGWB2 containing the 35S promoter upstream of the cloning site. These constructs were used for *Arabidopsis* (Col-0) transformation. The same vectors were used for hairy root transformation in *Medicago* with an additional DsRed gene inserted to select transformed roots via microscopy. The short tandem target mimic (STTM) constructs were synthesized (Integrated DNA Technologies, Coralville, Iowa) which were amplified by PCR and inserted into binary vector pRedRootII, which contains DsRed as a reporter gene.

Plant transformation

Hairy root transformation was conducted using *Medicago* seedlings according to the protocol by Chabaud et al. (2006). DsRed was used for the selection of

transformed hairy roots via fluorescence microscopy at the Delaware Biotechnology Institute Bioimaging Center (Newark, Delaware). We used the excitation and emission filters for Rhodamine on Zeiss M2BIO dissecting microscope to select transformed roots. Transformed roots were selected through fluorescence microscopy, and cut for in vitro cultures (Chabaud et al., 2002). A combination of antibiotics (Timentin 100mg/L, Carbenicillin 200mg/L) was used to suppress the growth of *A. rhizogenes* during tissue culture. Root tissues were harvested after a two-week culture at 25°C. At least five different transformed roots were pooled together as one biological replicate for RNA extraction and subsequent experiments. Transient expression assays on *N. benthamiana* were according to methods described before (Sparkes et al., 2006). *A. tumefaciens* carrying binary vectors was cultured overnight at 28°C, pelleted by centrifugation, and resuspended at OD₆₀₀ = 0.4. Equal amounts of bacterial suspensions were mixed to infiltrate for coexpression assays. Infiltrated areas on tobacco leaf were marked and collected for RNA extraction 48 hours after infiltration. Floral dip transformation of *Arabidopsis* was carried out using binary Gateway vectors that express miR1507, miR2109, and miR2118a/b/c respectively. *A. tumefaciens* GV3101 was used for both floral dip transformations in *Arabidopsis* and transient expression assays in *N. benthamiana* as described above. Floral dip transformation was conducted as previously described (Clough and Bent 1998).

RNA extraction and RNA gel blot hybridization

Total RNA samples in this study were all extracted using PureLink Plant RNA Reagent (Ambion, Foster City, CA) according to the manufacturer's protocol. Twenty mg total RNA samples from the leaf tissue of transient expression assays were loaded

into 15% denaturing polyacrylamide gels for electrophoresis. Non-radioactive small RNA gel blot hybridization was carried out using Locked Nucleoside Analogues (LNA; Exiqon, Woburn, MA) probes after transferring RNA to positively charged nitrocellulose membrane. The small RNA gel blot procedures were according to a previous study (Kim et al., 2010).

Small RNA, RNA-seq, and PARE library construction

Total RNA samples from miRNA overexpression lines of *Medicago*, target mimic lines of *Medicago*, and leaf tissues from 4-week old *Arabidopsis* miRNA overexpression T1 plants were used to construct small RNA libraries through TruSeq Small RNA Sample Preparation Kit (Illumina, Hayward, CA) according to manufacturer's manual. In addition to small RNA libraries, total RNA samples from empty vector control for miRNA overexpression in *Medicago* were also used for RNA-seq library construction via the TruSeq RNA Sample Preparation Kit v2 (Illumina). Total RNA samples from miRNA overexpression lines of *Medicago* were used for PARE library construction (German et al., 2009, Zhai et al., 2014). All libraries in this study were sequenced using the Illumina HiSeq 2500 at the Delaware Biotechnology Institute.

5' RACE assays

Total RNA samples from miR1507 overexpression lines of *Arabidopsis* were used for 5' RACE using GeneRacer Kit (Invitrogen). Gene specific primers were designed according to the manufacturer's instructions. Nested PCR products were analyzed on 1% agarose gel, and cloned using TA Cloning Kit (Invitrogen). Plasmids

were extracted from transformed single clones, and subjected to Sanger sequencing to confirm the mRNA cleavage sites.

Bioinformatics and statistical analysis

All sequencing data were first trimmed to remove the adapters, and then mapped to the genome using Bowtie (for small RNA and PARE) and TopHat (for RNA-seq) to the genomes (Langmead et al., 2009, Trapnell et al., 2009). Phasing analysis for PHAS loci identification in *Medicago* was performed using the same method and criteria as described by Zhai et al. (2011). PhasiRNA abundance from each *PHAS* loci was calculated by adding up normalized reads with the size of 21- and 22-nt. The abundance of reads that have multiple hits on the genome was calculated by dividing by the number of hits. Reads that mapped to more than 10 genomic locations were not included in the calculation. The Bioconductor (www.bioconductor.org) package “OutlierD” was used to identify the outliers of phasiRNA abundance from *PHAS* loci by comparing either miRNA overexpression with empty vector controls in both *Medicago* and *Arabidopsis*. Boxplots in this study were all generated using R (www.r-project.org). The Bioconductor package “edgeR” was used for differential analysis of small RNAs in miRNA overexpression lines in *Medicago*, and the R package “pheatmap” for heat maps in this study. PARE data analysis was according to the method described in the previous study (Zhai et al., 2011). Gene expression levels were represented by the value of RPKM (reads per kilobase per million reads) of each gene. The calculation of RPKM values was according to the formula: $RPKM = 10^9 NL^{-1} T^{-1}$ (N, number of reads mapped to the exons of a gene; L, length of exons of a gene; T, total number of reads mapped to the genome). Penalty scores of miRNA/target pairing was calculated according to

published rules (Fahlgren and Carrington 2010). The interaction energy between miRNA and target sequences was calculated by the Vienna RNA package “RNAup” (Gruber et al., 2008). The target sequence used for interaction energy calculation includes the target sequence of a miRNA, along with 17 nt upstream and 13 nt downstream of the target site (Kertesz et al., 2007).

Chapter 3

STUDIES OF phasiRNA BIOGENESIS AND PLANT miRNA FUNCTIONS VIA CRISPR/Cas9

3.1 Introduction

Plant miRNAs represent an important class of small RNAs derived from hairpin-structured pri-miRNAs that are transcribed by RNA polymerase II. Artificial regulation of a miRNA accumulation generally results in the misregulation of miRNA target genes, resulting in a variety of phenotypes that can facilitate functional studies of the miRNA. For example, overexpression of miR172 disrupts floral patterning by suppressing a subset of *APETELA2* (*AP2*) genes in *Arabidopsis* (Aukerman and Sakai, 2003; Chen, 2004). Overexpression of miR165 causes reduction of the class III homeodomain leucine-zipper (*HD-ZIP III*) genes, exhibiting pleiotropic developmental effects on plant development, such as apical meristem formation, establishment of organ polarity, and vascular tissue development (Zhou *et al.*, 2007). miRNA overexpression experiments showed that the phenotypes caused by miRNA over-accumulation mimics the phenotype of the loss-of-function mutants of target genes to some extent. In contrast, expression of the miRNA-resistant form of target genes is a way to mimic the loss-of-function of a miRNA. In this case, the binding site of a miRNA on the target mRNA is disrupted by substituting a few nucleotides while maintaining the encoded amino acids;

this is possible if the disrupted nucleotides correspond to synonymous substitutions within a codon, decoupling the regulation by a miRNA from the translation of the same protein. For example, the expression of miR160-resistant *ARF10*, *ARF16* and *ARF17* displayed pleiotropic developmental defects, such as smaller plant size, serrated leaf, aberrant flower, and reduced fertility (Liu *et al.*, 2007; Mallory *et al.*, 2005; Wang *et al.*, 2005). Expression of a miR398-resistant form of the copper/zinc superoxide dismutase *CSD2* leads to enhanced resistance to oxidative stress in Arabidopsis, compared to wildtype and *CSD2* overexpression lines, suggesting that miR398 is a negative regulator of resistance to plant oxidative stress (Sunkar *et al.*, 2006).

The finding of miR399 “target mimicry” by the noncoding gene *INDUCED BY PHOSPHATE STARVATION1 (IPS1)* in Arabidopsis provided novel insights into endogenous mechanisms of miRNA modulation (Franco-Zorrilla *et al.*, 2007). This discovery of a non-functioning target site that acts like a sponge to reduce the effective titer of the miRNA led to the development of artificial target mimics for plant miRNAs (Todesco *et al.*, 2010). An improved approach using a short tandem target mimic (STTM) of miRNAs showed higher efficiency in miRNA reduction *in vivo* (Yan *et al.*, 2012). Due to sequence similarities, transcribed STTMs sequester mature miRNAs representing all or nearly all of an entire *MIRNA* gene family, and subject them to the degradation pathway mediated by SMALL RNA DEGRADING NUCLEASEs (SDNs) (Yan *et al.*, 2012). *MIRNA* gene families, especially those conserved families, such as *MIR156*, *MIR166* and *MIR172*, contain large numbers of members encoded in a single plant genome, amplified and diversified during speciation and evolution (Cuperus *et al.*,

2011). Therefore, *MIRNA* paralogs may have distinct expression patterns resulting from diversified transcription and processing efficiency, and the study of individual members or single *MIRNA* genes seems necessary for functional studies of plant miRNAs. Artificial miRNAs (amiRNAs), a widely used technique for gene silencing in plants (Schwab *et al.*, 2006), was shown to be successful for suppression of either (a) all members of a miRNA family, or (b) an individual member. This can be done by the design of amiRNAs targeting either the mature miRNA, in the case of (a), or stem-loop regions in the case of (b) (Eamens *et al.*, 2011). However, the efficiency of silencing should always be considered for the methods mentioned above, as they may be less than 100% effective. Random T-DNA insertional gene knockout mutants have been widely used for gene functional studies in Arabidopsis, and null mutants would be acquired if the position of T-DNA insertion is at a critical nucleotide (Alonso *et al.*, 2003). However, due to the small size of *MIRNA* genes, or at least the most important hairpin, T-DNA insertion lines for *MIRNA* genes are rare, although functional analyses of *MIRNA* genes using T-DNA mutants have been reported (Beauclair *et al.*, 2010; Yang *et al.*, 2013).

The successful applications of CRISPR/Cas9 technology in plant genome editing were firstly reported in August, 2013, by different groups (Li *et al.*, 2013; Nekrasov *et al.*, 2013; Shan *et al.*, 2013). This makes it possible to generate miRNA mutants using CRISPR/Cas9-mediated site-directed genome-editing. In this chapter, I applied CRISPR/Cas9 for a functional study of (a) the miR160 family, and (b) the biogenesis of phasiRNAs; this was done by targeting Cas9 using a guide RNA (gRNA)

with the intent of altering miRNA precursor structures encoded in the Arabidopsis genome. We selected miR160 for study because: a) the miR160 family contains three members, including *MIR160a*, *MIR160b* and *MIR160c*, a moderate size among *MIRNA* gene families in Arabidopsis; b) miR160 loss-of-function mutants should have clear and specific leaf phenotypes, according to previous studies (Liu *et al.*, 2007; Mallory *et al.*, 2005; Wang *et al.*, 2005), and this would facilitate identification of Arabidopsis mutants at early vegetative stages; c) *MIR160* gene sequences contain multiple PAM sites recognized by Cas9, allowing more choices of target sites (data not shown).

3.2 Results

3.2.1 Genome editing of the passenger strand of *MIR160a* using a single guide RNA

A previous study showed that a miRNA with an asymmetric duplex structure tends to trigger secondary siRNA production from its targets (Manavella *et al.*, 2012). Such an asymmetric duplex structure would result from a bulge on either the miRNA or passenger (miRNA*) strand (Manavella *et al.*, 2012). However, data in other publications are not entirely consistent with the conclusion that duplex asymmetry triggers phasiRNA biogenesis; it may more likely be the 22-nt length of miRNAs that determines secondary siRNA production, as a 22-nt “tailed” version of miR171 (which is normally 21-nt) mediated by UTP:RNA uridylyltransferase 1 (URT1) in *hen1* background triggered secondary siRNA production (Tu *et al.*, 2015; Zhai *et al.*, 2013).

To examine factors that determine secondary siRNA production, i.e. bulge asymmetry versus 22-nt length, we designed a CRISPR/Cas9 vector with a single guide RNA (sgRNA) targeting the genomic sequence that encodes the miRNA* strand of *MIR160a* (Figure 19). miR160 is a 21-nt miRNA derived from pre-miR160 with a symmetric structure, so if a base pair is inserted into the genomic DNA coding for the miR160a* strand, the duplex will become asymmetric, perhaps resulting in a 22-nt miRNA*, which theoretically is sufficient to activate the biogenesis of phasiRNAs at the mRNA target of the miRNA (Manavella *et al.*, 2012). My experiment was designed to test this hypothesis. This type of mutation at the miR160* target of the sgRNA should

be frequent, as single nucleotide insertions and deletions are the most common events occurring at a double-stranded break (DSB) site (Li *et al.*, 2013). A bulge on the miRNA strand in a miR160/miR160* duplex might affect the pairing between miR160 and its target mRNAs, thus this design wouldn't test secondary siRNA production and I instead targeted the miRNA* strand.

I transformed this CRISPR/Cas9 vector into Arabidopsis Col-0 and screened the T0 seeds using antibiotics. Among the 17 T1 individual plants that I screened, seven of them showed successful genome editing confirmed by Sanger sequencing with an efficiency of ~ 41%. Surprisingly, six of them showed a single T/A base pair insertion, while only one (Line 13) showed a mixed sequencing result, suggesting that it may contain biallelic modifications or bear a modification constituting multiple base pairs (Figure 20). These successfully edited plants showed a moderate serrated leaf phenotype (data not shown), consistent with previously-described alterations of miR160 activity, as mentioned above. Genomic sequences of *MIR160b* and *MIR160c* were sequenced in these lines, and no evidence of off-target activity was found.

We next grew the genome-edited T2 plants from each individual T1 line (~20 to 30 plants per line) to observe and characterize the segregation of the phenotype and genotype. We observed that a total of four lines (Lines 9, 10, 11 and 13) did not show phenotype segregation, among which the T2 plants from lines 9 (*160a*-9*) and 10 (*160a*-10*) all showed a single T/A insertion (Figure 21A). Except for single T/A insertions, a few T2 plants of line *160a*-11* showed a single "A" insertion, while others showed a mixture of "T" and "A" (Figure 21A). It is possible that the Cas9 continued

to modify the plant genome in the T1 line before T2 segregation, which was reported previously (Feng *et al.*, 2014). Therefore, segregating out the Cas9 transgene in the CRISPR mutagenized plants would be necessary to maintain the genome integrity of the progeny. In addition to a single T/A insertion, by Sanger sequencing some T2 plants from the line *160a*⁻¹³* still showed indiscernible sequences, while others showed a five base-pair deletion (Figure 21A), suggesting biallelic genome modifications exist in line 13 in the T1 generation. This five base-pair deletion was likely able to disrupt pre-miR160a processing by DCL1, as the secondary structure of pre-miR160a was dramatically changed (Figure 21B).

I obtained homozygous T2 plants with a single “T” insertion (*160a*⁻⁹⁻²*) and 5 bp deletion (*160a*⁻¹³⁻¹⁶*) in the miR160a* strand but segregated out the Cas9 transgenes. Next, I examined more carefully and in the T3 generation the plant phenotypes derived from these two different modifications; I found that the degree of serration in the leaf was slightly higher for *160a*⁻¹³⁻¹⁶* plants compared with *160a*⁻⁹⁻²* (Figure 22A). I further quantified via qRT-PCR the relative expression levels of *ARF10*, *ARF16* and *ARF17*, the targets of miR160, in leaf tissue. I found that the transcript levels of all three genes were slightly higher in two mutants than the wildtype control. Moreover, the levels were higher in *160a*⁻¹³⁻¹⁶* compared to *160a*⁻⁹⁻²*, suggesting that the expression levels of *ARF* genes correlated with subtle but observable degrees of a serrated leaf phenotype (Figure 22B). I subsequently assessed the accumulation of *MIR160a* transcripts in these two mutants, and I found that the pri-miR160a accumulation in *160a*⁻⁹⁻²* and *160a*⁻¹³⁻¹⁶* was ~10-fold and ~40-fold

higher than the wildtype. This result indicated that in both mutants the processing of pri-miR160a by DCL1 might be interrupted by nucleotide insertions or deletions. To investigate whether secondary siRNAs were produced from miR160 target genes in *160a*⁻⁹⁻²* plants, I prepared small RNA libraries from the leaf tissue of this line, and I found that this line producing asymmetric miR160a/miR160a* duplexes does not produce secondary siRNAs at the target gene (Figure 23). These results suggest that the secondary structure of miRNA/miRNA* duplex may not determine the secondary siRNA production, unlike the conclusion of Manavella *et al.* (2012).

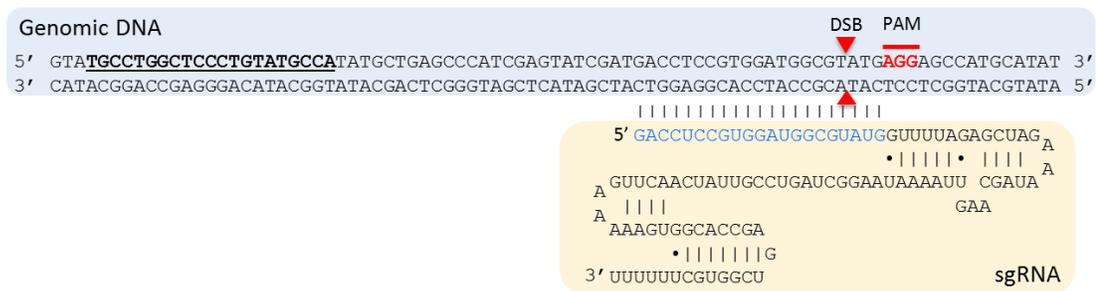


Figure 19. Single guide RNA design for miRNA*.

Double-stranded genomic DNA is shown in the blue box and the guide RNA (sgRNA) is shown in the yellow box. The sequence encoding the mature miR160 is in bold and underlined. The PAM site is indicated in red, and the potential site of a double-strand break (DSB) is specified by red arrows.

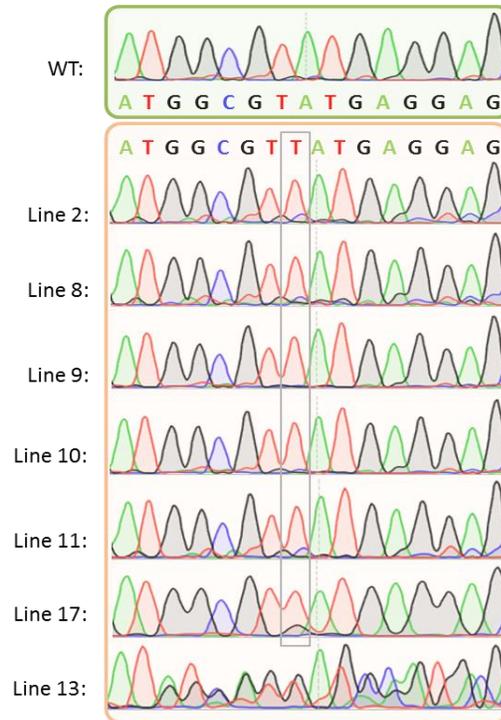


Figure 20. Sanger sequencing results of individual T1 plants transformed with CRISPR/Cas9 targeting the miRNA* strand of miR160a. Most plants showed a single T/A insertion (grey box), while Line 13 showed mixed PCR products, evident in the diverse peaks of the electropherogram.

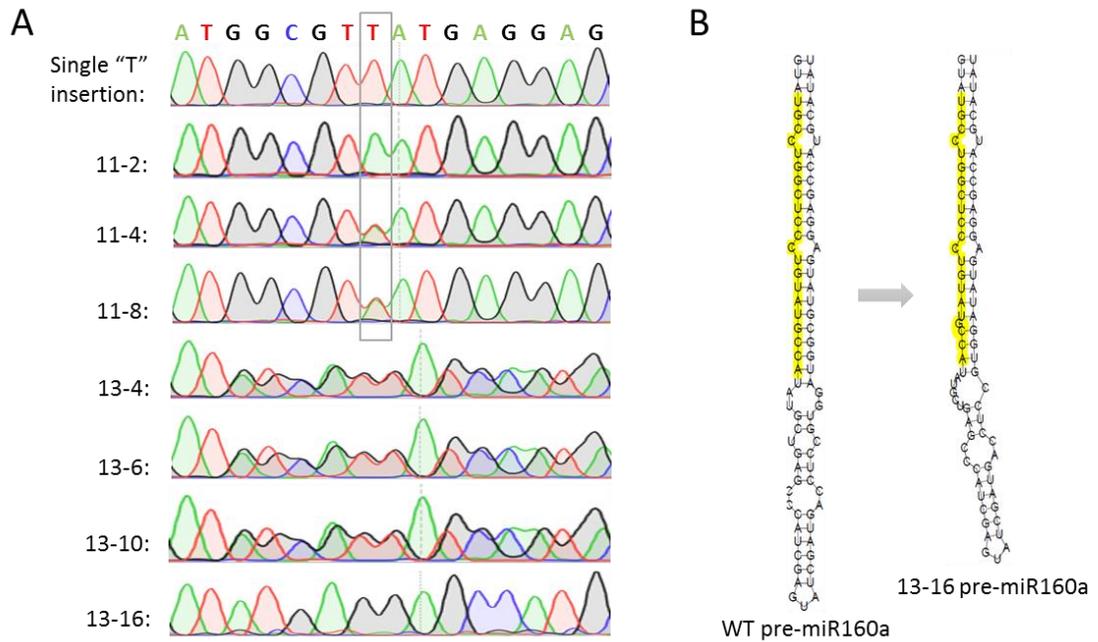


Figure 21. Sanger sequencing results of individual T2 plants transformed with CRISPR/Cas9 targeting the miRNA* strand of miR160a.

A. Line 11 showed mixed T/A and A/T insertions (grey box). Lines 13-4, 13-6, and 13-10 show a mix of the 5-bp deletion and a single base insertion. Line 13-16 (*160a**-13-16) showed a homozygous 5-bp deletion on the passenger strand of miR160a.

B. The predicted secondary structure of pre-miR160a in *160a**-13-16 is dramatically changed as a consequence of the 5-bp deletion (right side), compared to wildtype (left side).

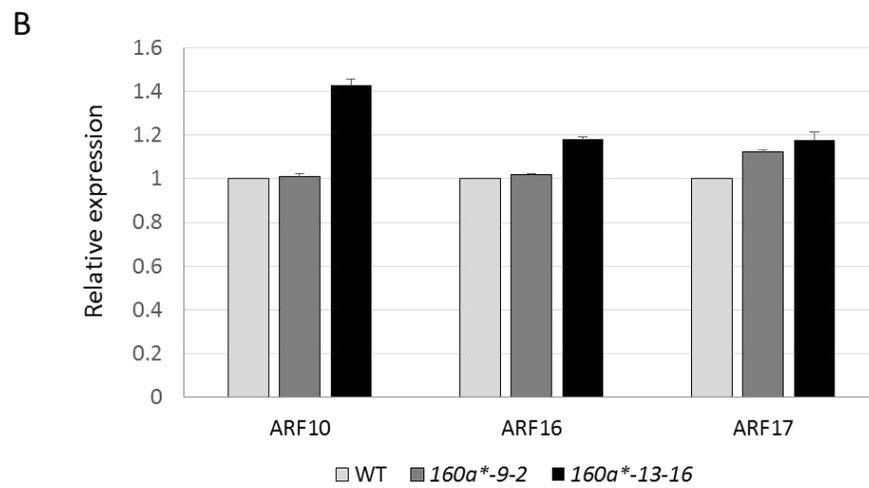


Figure 22. T3 plants of line *160a**-13-16 and *160a**-9. Plants (T3) of *160a**-13-16 showed a more severe phenotype than *160a**-9 (panel A); higher levels of miR160 target transcripts (measured by qRT-PCR, with two biological replicates) accumulated in *160a**-13-16 than *160a**-9 (panel B), including *ARF10*, *ARF16* and *ARF17*.

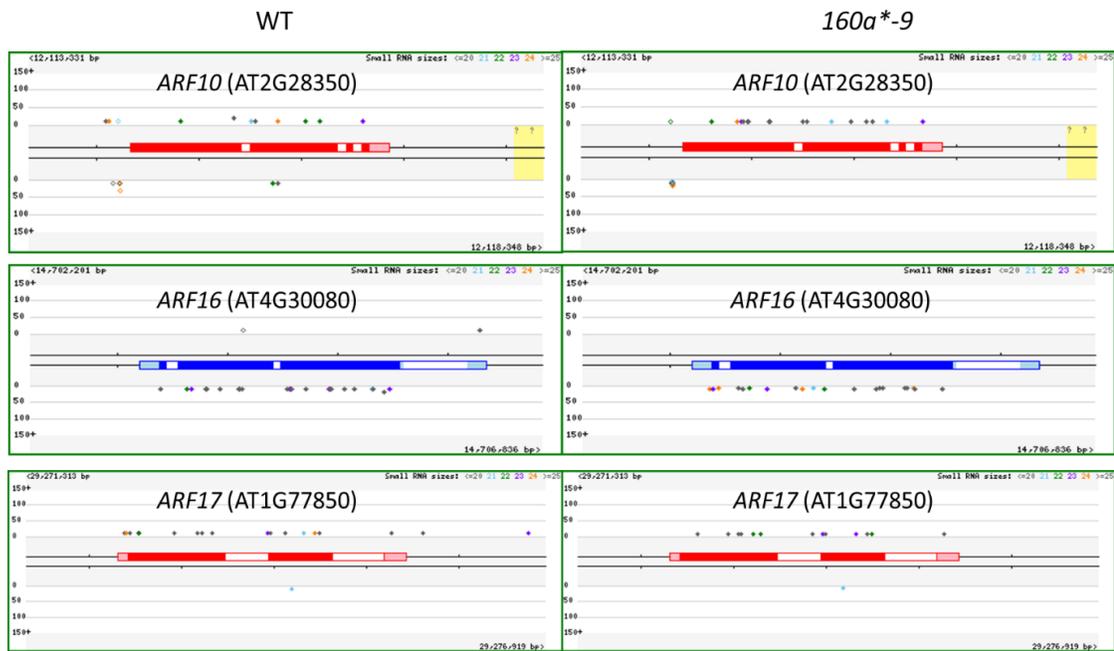


Figure 23. Small RNAs leaves of wildtype and *160a**-9 plants, mapped to target *ARF* genes to look for phasiRNA production. Screenshots of our lab's genome viewer indicate very low levels of small RNA generated from transcripts of *ARF10*, *ARF16* and *ARF17*, with no evidence of enrichment for 21-nt phasiRNAs.

3.2.2 Genome editing of the miRNA strand of *MIR160a* using a single guide

RNA

miR160 is an important regulator of *ARF* genes, important in Arabidopsis development. However, to date, the mutants of *MIR160* families have not been well-described, although a Ds insertion line (*foc*) for *MIR160a* was reported in Arabidopsis (Liu *et al.*, 2010). I designed three sgRNA CRISPR vectors individually targeting the miRNA strand of *MIR160a*, *MIR160b* and *MIR160c* genes and transformed these constructs into the Col-0 background (Figure 24). Sequencing results of the T1 plants showed that the overall efficiency of genome editing for *MIR160a*, *MIR160b* and *MIR160c* was 83%, 74% and 95%, respectively, suggesting that the CRISPR/Cas9 is highly successful in genome modifications in my experiments. Similar to the miR160a*-sgRNA experiment, a single T/A insertion was predominant for the genome-modified lines. However, unlike the case of protein-encoding genes in which a single nucleotide insertion will cause a shift of an open reading frame (“ORF”) during translation, in noncoding genes such as miRNAs, there is no ORF to disrupt. Therefore, considering the relatively low frequency of occurrence of multiple nucleotide modifications, I decided to use a double-gRNA (“dgRNA”) system to obtain genomic DNA deletion mutants, as deletion mutants are more likely give rise to null mutants of the miR160 family (see below).

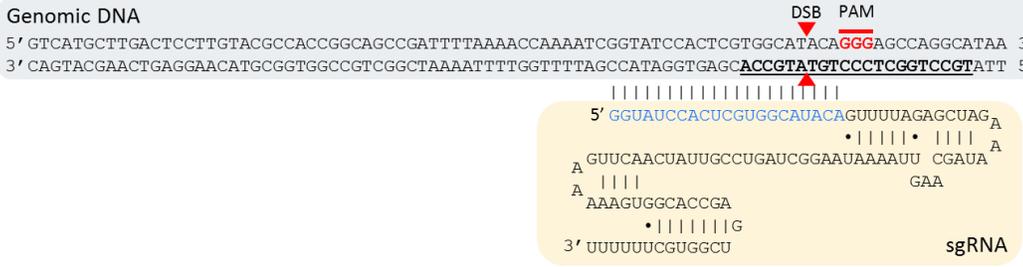
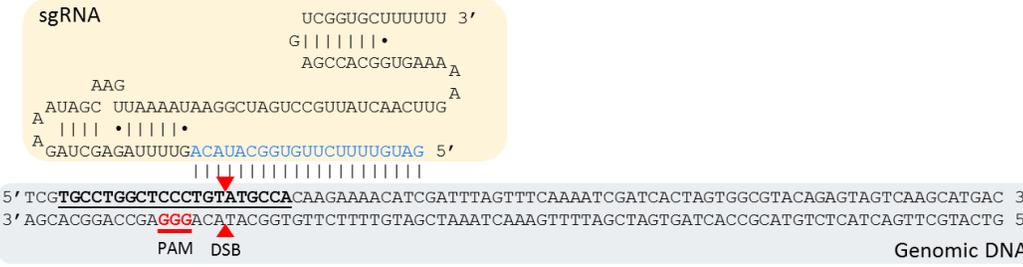


Figure 24. Single guide RNA design for the miRNA strand of *MIR160*. From top to bottom: *MIR160a*, *MIR160b*, *MIR160c*. Labels in this figure are as described in Figure 19.

3.2.3 Genome editing on *MIR160a* using double guide RNA

Considering the low frequency of genomic DNA fragment deletions created by single guide RNA vectors, I designed a double guide RNA (dgRNA) vector to test its efficiency for *MIR160a* and to see if I could get null mutants by a fragment deletion. The distance between two DSB sites is 49 bp (Figure 25A), a deletion of which could be detected by PCR using primers up- and down-stream of the DSB sites. Surprisingly, among the 66 T1 plants that I screened, >20 of them showed a lower band size compared to the wildtype plants (Figure 25B), indicating that genomic DNA fragment deletions frequently occurred in the transformed plants. I thereafter obtained homozygous T2 plants with 47 or 48 bp fragment deletions confirmed by Sanger sequencing (Figure 26A). I believe that these plants are likely null mutants, because deletion of 47 or 48 bp in the miR160a precursor will likely abolish or severely disrupt processing of the precursor by DCL1 (Figure 26B). These plants displayed severe developmental phenotypes. For example, the leaves were highly serrated (Figure 27A), and the petals were thinner and inward-curved (Figure 27B). I also observed that the siliques of the homozygous plants were generally shorter than the wildtype siliques at the same developmental stages (Figure 27C), and thus I speculated that the fertility might also be reduced. I examined developing seeds in the siliques, and I found that the siliques of plants with a hemizygous deletion had seemingly randomly unfertilized ovules (Figure 27D), but at a very low rate. In contrast, the homozygous mutant siliques displayed severe defects in fertility (Figure 27D). A large number of the ovules were unfertilized, while some the developing seeds looked aberrantly small. I also observed that the

development of some embryos was aborted or they were dead at an early developmental stage, showing a brown color (Figure 27D). These results suggest that in embryo development, *MIR160a* has an important role in post-transcriptional control of aberrant *ARF* gene expression.

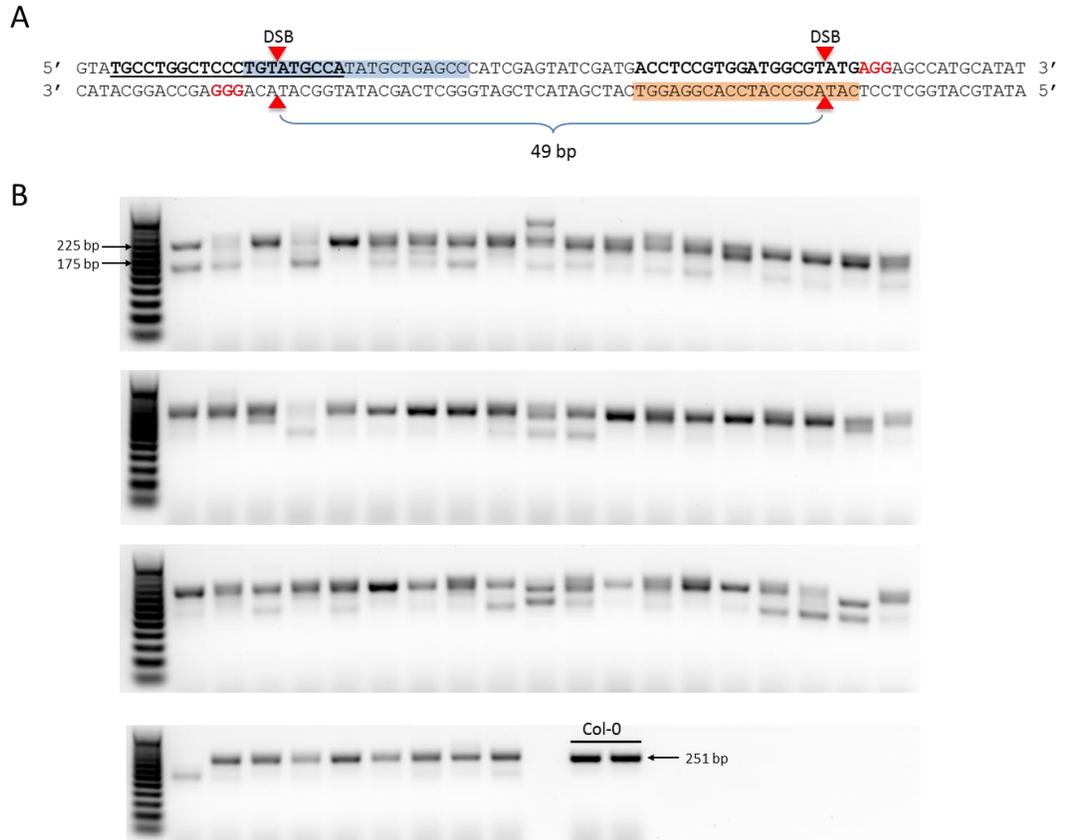


Figure 25. CRISPR/Cas9-mediated genome editing via double guide RNAs.

A. Double guide RNA design for a 49 bp fragment deletion on *MIR160a*. Blue and orange boxes indicate the sequences targeted by the two guide RNAs, with the predicted breakpoints designated by red triangles (DSBs).

B. PCR products amplified from the *MIR160a* gene using the genomic DNA of T1 plants carrying the double guide RNA transgene.

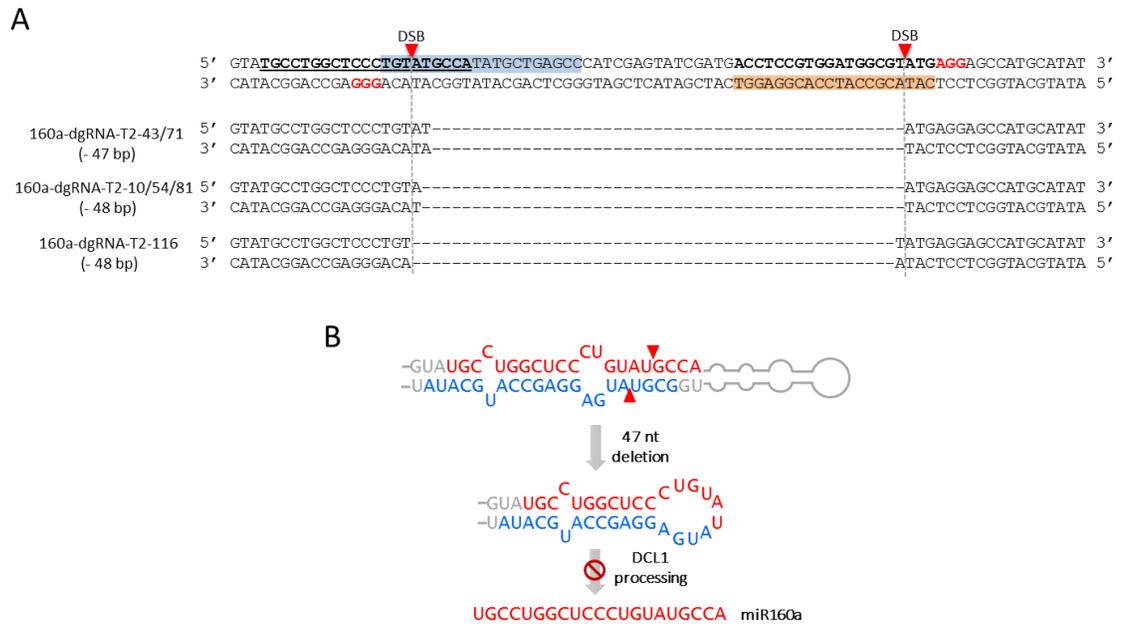


Figure 26. Fragment deletion on *MIR160a* generated by CRISPR/Cas9 with double guide RNAs.

A. Base pair deletions confirmed by Sanger sequencing. The mature miR160 sequence is underlined.

B. Predicted secondary structure of pre-miR160a containing a 47 nt deletion resulting from the double guide RNA, and the consequent lack of ability to be processed by DCL1.

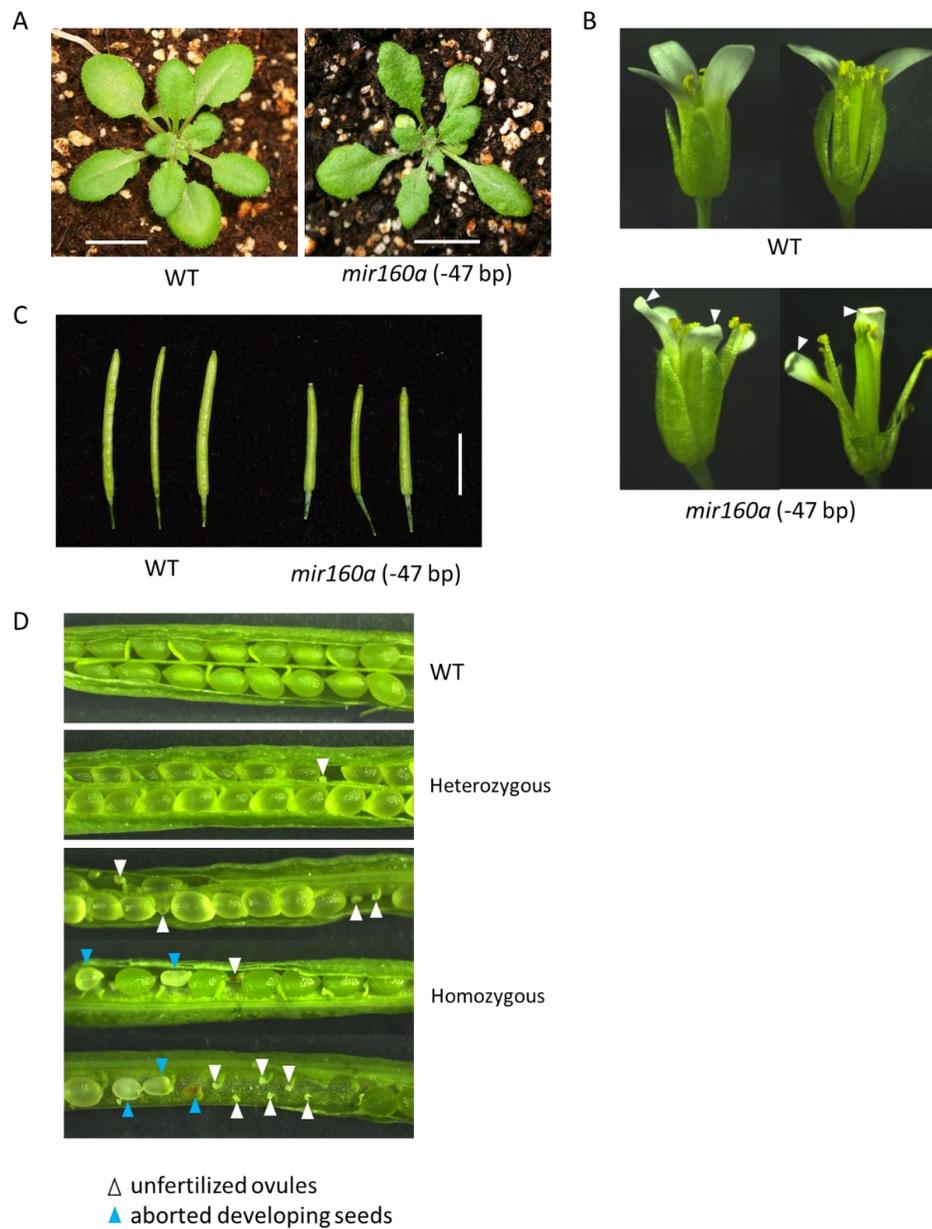


Figure 27. Mutants of *MIR160a* with a 47-bp fragment deletion display pleiotropic phenotypes in plant development. Serrated leaves (A, scale bar = 1 cm), inward curled and thin petals (B), and shorter siliques (C, scale bar = 0.5 cm), and reduced fertility (D).

3.3 Discussion

The previous study by Manavella *et al.* (2012) showed that the asymmetric duplex structure resulting from a bulge on either the miRNA or passenger (miRNA*) strand in a miRNA/miRNA* duplex is necessary and sufficient to trigger secondary siRNA production at the miRNA targets. However, more recent data are showing that the 22-nt length of miRNAs, instead of the asymmetric structure of miRNA/miRNA* duplex, determines secondary siRNA production. To date, there have been no reports demonstrating by modification of endogenous loci whether the ability to produce phasiRNAs can be endowed or abolished by changing the secondary structure of a miRNA; this was tested in the Manavella *et al.* (2012) paper only using transgenes. In this study, I designed a CRISPR/Cas9 vector with a single guide RNA targeting the miRNA* strand of the *MIR160a* gene. I obtained homozygous T2 plants with a single “T” insertion (*160a*-9*) at the miR160a* strand, and with the Cas9 transgene segregated away. Thus, I created via CRISPR/Cas9 plants with a single nucleotide insertion in the genome compared to the parental Col-0 line without a transgene. Small RNA sequencing data from these plants showed that secondary siRNAs were not generated from the targets of miR160, indicating that secondary siRNA production may not be determined by the asymmetric structure of a miRNA precursor.

CRISPR/Cas9 has widely been proven as an efficient technique for genome editing. Our experiments showed that CRISPR/Cas9 is also powerful in functional studies of plant non-coding genes, such as miRNAs. In our CRISPR experiments, the genome editing efficiency for single guide RNAs ranged from ~ 40% to 95%, mainly

causing a single nucleotide insertion. For CRISPR experiments using double guide RNAs via the floral dip method, the overall rate of fragment deletions (~50 bp) was up to ~30%, suggesting that fragment deletions can be achieved at a relatively high efficiency. Moreover, it is possible that some larger fragment deletions might also be achieved by the floral dip method when the guide RNA design is appropriate; or it might also be achievable by multiple guide RNAs to generate even longer fragment deletions by tandem double-strand breaks.

In our experiments, we found that the *MIR160a* gene with a 47 or 48 bp deletion caused a severe developmental phenotype, suggesting that miR160a plays a major role among the three loci in the miR160 family. However, we did not observe phenotypes, such as an “emerged inflorescence in silique”, that were previously ascribed to a loss of *MIR160a*, in the published case of the *foc* mutant (Liu *et al.*, 2010). It is possible that this difference is because the *foc* mutant is in the Landsberg erecta (Ler) background. In order to confirm that the *mir160a* mutants with the large fragment deletions are in fact null mutants, I am currently preparing small RNA libraries from these mutants to confirm whether or not the mature miR160a in these mutants is totally abolished.

3.4 Conclusions and Future Directions

I showed that CRISPR/Cas9 is a powerful tool for genome editing, targeted mutagenesis and functional characterization of miRNAs in Arabidopsis, particularly when coupled with the floral dip method of transformation. Generally, the genome editing efficiency mediated by single guide RNAs was from ~ 40% to 95%. For double guide RNA CRISPR experiments, the efficiency of the expected ~50 bp fragment deletions was around 30%. These results indicate that fragment deletions can be achieved at a relatively high efficiency in Arabidopsis via the floral dip method of transformation.

In the experiment editing the miR160* strand of *MIR160a* gene via CRISPR/Cas9, I obtained homozygous T2 plants with a single “T” insertion (*160a*-9*) at the miR160a* strand, converting the symmetric pre-miRNA into the asymmetric structure. Small RNA sequencing data from these plants proved that secondary siRNAs were not produced from the target transcripts of miR160. This result suggested that the asymmetric structure of a miRNA might not determine secondary siRNA production from the target transcripts; the corollary to this conclusion is that it is more likely that the 22-nt length of a miRNA is more important than the bulge structure for the activity of the miRNA in generating secondary siRNAs.

In this study, I generated *mir160a* mutants with fragment deletions via CRISPR/Cas9, which were likely null mutants. I am preparing small RNA libraries from these mutants to confirm whether or not the mature miR160a in these mutants was totally abolished, but the visible phenotype is consistent with a loss of miR160

activity. In addition, double guide RNA constructs will also be designed for *MIR160b* and *MIR160c*, so that we can obtain the null mutants of the miR160 family individually. These mutants will be very helpful in understanding why multiple copies of miRNAs are necessary in plants in the evolutionary perspective, and whether/how the expression of these miRNA members are specialized during plant development in the regulation of auxin response factors.

3.5 Materials and Methods

Plant materials

All the plants of *Arabidopsis thaliana* Col-0 in this study were grown in the growth chamber with the light cycle of 16 h light / 8 h dark at 22 to 23°C. Leaf materials were collected from the 3-week old plants.

Vector construction and transformation

The CRISPR/Cas9 vector was obtained from Zachary Nimchuk at the University of North Carolina at Chapel Hill. Basically, the vector contains a ubiquitin 10 promoter (UBQ10) driving the Cas9 gene, and U6 promoter driving the transcription of guide RNAs. Overlap-extension PCR was employed to amplify the U6:gRNA fragment by substituting the ~20 bp gene-specific sequences in the guide RNA using primers containing *PmeI* restriction sites. The amplified fragment was then cloned into the pCR-Blunt vector using the Zero Blunt PCR Cloning Kit (Invitrogen). *PmeI* was used to release the fragment out of the pCR-Blunt vector, followed by running an agarose gel and gel extraction. The binary vector containing a single *PmeI* site was linearized by *PmeI* digestion and treated with rSAP to prevent self-ligation. The purified U6:gRNA fragment and the linearized vector were then ligated by T4 DNA ligase (NEB) at 16° C overnight. The ligation product was then transformed into One Shot TOP10 chemically competent *E. coli* (Invitrogen) and then selected on LB plates with 50 ug/mL spectinomycin.

For the double guide RNA vector construction, we synthesized a fragment containing tandem U6:gRNA sequences from Invitrogen. The double guide RNA fragment was digested from the vector provided by the company using *PmeI*, and transformed into pCR-Blunt vector, followed by ligation into the binary vector as described above.

The plasmids containing a single U6:gRNA fragment or a double U6:gRNA fragment were confirmed by Sanger sequencing, and transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation. Arabidopsis plants at proper stages were transformed with *Agrobacterium tumefaciens* via the floral dip method as described before (Clough and Bent 1998).

Screening and genotyping

The transformed T0 seeds were screened on plates containing MS media with 50 ug/mL kanamycin. The selected T1 plants were transferred into soil and grown in the growth chamber. The genomic DNA was extracted from each individual plant and used as templates for PCR using primers both up- and down-stream the target sites of guide RNAs. For the T1 plants transformed with single guide RNA vectors, PCR products were sent for Sanger sequencing.

For T1 plants transformed with double guide RNA vectors, PCR products were firstly examined on an agarose gel to check if smaller bands may result from fragment deletions generated by CRISPR/Cas9. T2 plants were grown and further analyzed by PCR and gel running. Only the plants having a PCR product with a single band (~50

bp smaller than the wildtype) were selected for Sanger sequencing. We also used primers to amplify the Cas9 gene to genotype T2 plants in order to obtain transgene-free plants, which would likely be stable in genome contexts without the continuous activity of Cas9 and guide RNAs.

RNA extraction and quantitative RT-PCR

Total RNA samples in this study were extracted using PureLink Plant RNA Reagent (Ambion) according to the manufacturer's protocol. We treated 1 mg total RNA with DNase I and followed by reverse transcription by SuperScript III Reverse Transcriptase (Invitrogen). The reverse transcription products were diluted and used as templates for quantitative PCR using GoTaq qPCR Master Mix (Promega). We used the house keeping gene GAPDH as the internal control. The data analysis was conducted using the C_T method as described before (Schmittgen and Livak, 2008).

Small RNA library construction and data analysis

Total RNA was used as input for small RNA library construction using TruSeq Small RNA Sample Preparation Kit (Illumina) according to manufacturer's manual. Small RNA libraries were sequenced using the Illumina platforms. All the sequencing data were firstly trimmed to remove adapters, and then mapped to the Arabidopsis genome using Bowtie (Langmead *et al.*, 2009). Sequencing data from different libraries were normalized as described before (McCormick *et al.*, 2011), and for visualization, the data were loaded into the genome viewer of the Meyers Lab.

Chapter 4

SUMMARY AND FUTURE WORK

(This chapter has been published previously as Fei *et al.* (2016), modified to meet the formatting requirements of the dissertation.)

Small RNAs are a type of non-coding RNA which have a variety of biological functions. Plant small RNAs can be divided into several categories according to their distinct biogenesis pathways (Axtell, 2013), in every case functioning in gene silencing at the level of either transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS) (Brodersen and Voinnet, 2006). Host small RNAs, such as microRNAs (miRNAs) and small interfering RNAs (siRNAs), have been described to play crucial roles in plant disease resistance (Katiyar-Agarwal and Jin, 2010; Padmanabhan *et al.*, 2009). Plant miRNAs are processed from transcripts forming a stem-loop secondary structure, transcribed from *MIR* genes (Voinnet, 2009). A subset of plant miRNAs has been shown to target *NLR* genes regulating plant immunity (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Zhai *et al.*, 2011). As described above, some of the *NLR*-targeting miRNAs are capable of triggering the production of phasiRNAs from their cleaved target mRNAs, a capability variously attributed either to the 22-nt length of the miRNA triggers or to an asymmetric bulge in the region of the mRNA precursor processed into the miRNA-miRNA* duplex (Chen *et al.*, 2010; Cuperus *et al.*, 2010;

Manavella *et al.*, 2012). Our data from CRISPR/Cas9-mediated genome editing on *MIR160a* suggest that such an asymmetric duplex may not result in secondary siRNA production; instead, a 22-nt length of miRNA, is typically necessary for triggering phasiRNA production.

In Arabidopsis, treatment with the pathogen-associated molecular pattern (PAMP) known as flg22 induces the accumulation of miR393, which in turn targets F-box auxin receptors, including TIR1, AFB2, and AFB3 (Navarro *et al.*, 2006). This repression of auxin signaling correlates with enhanced disease resistance in plants, reflecting an enhancement of PTI (Navarro *et al.*, 2006). Apart from miRNA roles in PTI via hormonal regulation, there are numerous miRNAs that directly target transcripts from *NLRs*, a class of genes predominantly involved in effector-triggered immunity (ETI) (Figure 28). Moreover, many or most of these miRNAs are 22-nt and trigger the production of phasiRNAs from *NLRs* (Fei *et al.*, 2013). In the legume *Medicago truncatula*, several hundred *NLRs* are targeted by just five miRNAs (miR1507, miR2109, and miR2118a/b/c) at sequences encoding conserved motifs of R proteins, triggering widespread phasiRNAs (Zhai *et al.*, 2011). Due to the close evolutionary relationship with *Medicago*, the regulatory network for *NLRs* in soybean utilizes a similar but somewhat expanded repertoire of miRNAs (Arikiti *et al.*, 2014). In Solanaceous species, including tomato, potato, and tobacco, *NLR*-targeting miRNAs have also been well-characterized as triggering abundant phasiRNAs, including the miRNAs miR482, miR5300, miR6019, and miR6027 (Li *et al.*, 2012b; Shivaprasad *et al.*, 2012). Perennial woody plants have also been reported to employ the miR482/2118

superfamily and other miRNA families to repress *NLRs*. For example, in peach, about 94 *NLR* genes were identified as *PHAS* loci, which are predominantly triggered by the miR482 family (Zhu *et al.*, 2012). A recent study identified a novel miRNA that targets an *R* gene in apple (Ma *et al.*, 2014); this *R* gene is expressed at a higher level in the resistant than the susceptible cultivar, and interestingly, agrobacteria-mediated infiltration of the *R* gene in the leaf of susceptible apple cultivar enhanced plant immunity against the fungal pathogen *Alternaria alternata* f.sp. *mali* (Ma *et al.*, 2014). In spruce, poplar and grape, a large proportion of *NLR* genes produce 21-nt phasiRNAs (Kallman *et al.*, 2013). In a recently published study in spruce, it was shown that spruce *NLRs* are targeted by both the conserved miR482/2118 superfamily and a large number of other miRNAs (Xia *et al.*, 2015), indicating that *NLRs* are targeted by a variety of miRNA families in different plants. Indeed, there is substantial variation across species in both the presence/absence of these miRNAs, and in their breadth of their targets (and of the resulting phasiRNAs). Arabidopsis has just two miRNAs, miR472 and miR825*, that target only a few *NLRs* (Chen *et al.*, 2010; Howell *et al.*, 2007). In the grasses, a member of the miR482 superfamily, miR2118, has a distinct and specialized role in reproductive tissues as a trigger of 21-nt phasiRNAs from long non-coding RNAs, instead of from *NLR* transcripts (Song *et al.*, 2012; Zhai *et al.*, 2015). This spatially-restricted pattern of miR2118 is inconsistent with its role in *R* gene regulation seen in most eudicots, in which *NLR*-derived phasiRNAs are observed in vegetative tissues (Arikiti *et al.*, 2014; Li *et al.*, 2012b; Shivaprasad *et al.*, 2012; Zhai *et al.*, 2011). Do the grasses lack miRNA-mediated regulation of *R* genes? Apparently not, as miR9863 was

recently identified in both barley and wheat, and shown to target *MLA* genes, a class of CC-type *NLRs* (Liu *et al.*, 2014a). Interestingly, the 22-nt variant of miR9863 more efficiently suppresses *MLA1* than 21-nt miR9863, presumably via the *cis* activity of *MLA* phasiRNAs (Liu *et al.*, 2014a). Coupling the sequencing of ever more plant genomes with detailed molecular studies of plant defenses, we are sure to learn more about the diversity and roles of *NLR*-targeting miRNAs.

Despite a growing number of studies, we still lack a clear and incontrovertible understanding of the functional importance of the role of miRNAs and phasiRNAs in *R*-gene regulatory networks. Yet, clues are starting to emerge. For example, a recent study demonstrated that tomato miR482 and miR5300, the latter a member of the miR482/2118 superfamily, target four *R* genes that play a role in disease resistance to the wilt fungus *Fusarium oxysporum* (Ouyang *et al.*, 2014). Individual knock-downs in tomato of these four *R* genes via virus-induced gene silencing (VIGS) rendered a resistant cultivar susceptible to *F. oxysporum* (Ouyang *et al.*, 2014). Combining this study and the work by Shivaprasad *et al.* (2012), the miR482 superfamily has a demonstrated role to suppress a wide range of *R* genes that confer resistance to viral, bacterial, and fungal pathogens. Earlier work in *Medicago* demonstrating a handful of miRNAs can trigger phasiRNAs from more than 100 targets, resulting in phasiRNAs with an even more greatly expanded set of related targets led to hypothesis that these miRNAs are ‘master regulators’ of the *NLR* family of *R*-genes. But the basis for the variation across plants in the extent of this regulatory network is puzzling.

Since phasiRNAs function as negative regulators of *NLR* genes, loss-of-function mutants in the phasiRNA biogenesis pathway should exhibit enhanced ETI-based resistance to some pathogens. Indeed, consistent with this, Arabidopsis mutants of both *rdr6* and miR472 (a variant of the miR482 family found in Arabidopsis) displayed enhanced ETI mediated by RPS5 to the *P. syringae* DC3000 strain carrying AvrPphB (Boccaro *et al.*, 2014). These results suggest that phasiRNA biogenesis from *NLRs* may negatively regulate ETI. Boccaro *et al.* (2014) identified a number of *CC-NB-LRRs* (*CNLs*) targeted by NLR-derived phasiRNAs resulting from miR472 activity, showing that these miR472-triggered phasiRNAs act in *cis* and *trans* to suppress disease resistance genes until necessary, constituting an ETI enhancement switch (Boccaro *et al.*, 2014). Surprisingly, this study also showed that RDR6 negatively regulates PTI, because the expression levels of *WRKY22*, *WRKY29* and *FRK1* (PAMP-responsive genes) were significantly higher in *rdr6* compared to wild type (Boccaro *et al.*, 2014). In addition, increased callose deposition was observed in *rdr6*. These boosted PTI responses likely contributed to enhanced resistance against the pathogen *P. syringae* in *rdr6*, as quantified by bacterial titer (Boccaro *et al.*, 2014). Interestingly, it was also observed that *RDR6* expression decreased rapidly upon flg22 treatment (Boccaro *et al.*, 2014); therefore, it is likely that plants swiftly inhibit the RDR6-mediated RNA silencing pathway to strengthen host immune responses when sensing PAMPs. However, there is likely an underlying signaling pathway mediated by pattern recognition receptors that down-regulates *RDR6* expression; the components of this pathway and the molecular mechanism by which RDR6 suppresses PTI responses

remain to be determined. In the future, it will be informative to replicate assays of phasiRNA function by reducing or eliminating RDR6 activity in species with even more extensive sets of *NLR*-targeting miRNAs and phasiRNAs than *Arabidopsis*.

Host small RNAs and the RNA biogenesis machinery have well-described roles in plant disease resistance and plant-microbe interactions (Katiyar-Agarwal and Jin, 2010; Pelaez and Sanchez, 2013). During an evolutionary ‘arms race’ between pathogens and their plant hosts, the secretion of suppressors of RNA silencing to promote host susceptibility has proven an effective strategy (Pumplin and Voinnet, 2013). The first bacterial suppressor of RNA silencing (BSR), AvrPtoB from *P. syringae*, was demonstrated to suppress the transcription of *MIR393*, enhancing PTI via modulation of hormone signaling (Navarro *et al.*, 2008) (Figure 28). In contrast, the BSR AvrPto does not alter *MIR* transcription, suggesting perhaps an inhibitory role in pri-miRNA processing, while the effector HopT1-1 was shown instead to interrupt translational repression mediated by miRNAs (Navarro *et al.*, 2008). A study in tomato showed that miR482 levels decreased upon infection by *P. syringae*, suggesting that BSRs may interfere with either pri-miR482 transcription or processing (Shivaprasad *et al.*, 2012) (Figure 28). Levels of a control miRNA with no known role in defenses (miR168) were not impacted upon pathogen infection, suggesting a specific inhibition of miR482 (Shivaprasad *et al.*, 2012). Alternatively, plant recognition of effectors could result in transcriptional inhibition of *MIR482*; however, it is unlikely that PAMP-mediated signaling causes miR482 reduction, because infection by a *P. syringae hrcC* mutant (mentioned above) also reduced miR482 levels (Shivaprasad *et al.*, 2012).

Intriguingly, the decrease in mature miR482 was also produced by inoculation with the fungal pathogen *F. oxysporum* in a resistant but not susceptible tomato cultivar (Ouyang *et al.*, 2014), suggesting a possible pathway for miR482 suppression. Hence, it is possible that resistant host genotypes have evolved to recognize pathogen effectors and reduce levels of miR482 and miR482-triggered phasiRNAs, thereby increasing levels of *R* genes and enhancing ETI (Figure 28). Alternatively, some effectors may activate transcription of *NLR*-targeting miRNAs to attenuate ETI. For example, strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) secrete transcription activator-like (TAL) effectors to specifically upregulate *HEN1*, encoding a methyltransferase which stabilizes small RNAs via 3' methylation in rice (Li *et al.*, 2005; Moscou and Bogdanove, 2009). Presumably these effectors induce disease susceptibility via *HEN1* activation, although the connection between the stabilization of host miRNAs/siRNAs and defenses is not yet clear. In summary, pathogens have evolved a variety of suppressors of RNA silencing which act in diverse ways to enhance plant susceptibility.

Current data indicates that plant miRNAs, together with the phasiRNAs they trigger, are important regulators of plant immunity in both PTI and ETI. As described above, miR393 plays a role in enhancing PTI by repressing auxin signaling, while miR398 functions in PTI in both *Arabidopsis* and rice (Li *et al.*, 2014b; Li *et al.*, 2010). Yet more remains to be explored, such as how miR398b-mediated RNA silencing pathway affects pathogen resistance. The most extensive, yet still poorly understood, set of miRNAs is miR482/2118 and other miRNAs that generate phasiRNAs from *NLRs*; these miRNAs appear to be regulators of ETI in plants by suppression of *R* genes,

presumably as ‘master regulators’ - a small number of miRNAs regulating an enormous family of genes. PhasiRNAs may attenuate plant immunity either in *trans* by targeting other *R* genes or genes in other families, or in *cis* to target the genes that generate the phasiRNAs, thereby enhancing the suppression efficiency of the miRNA. Or another way to think about this is that relief of small RNA suppression could boost plant immunity. As mentioned above, RDR6, a key protein in the biogenesis of phasiRNAs, plays a role in resistance responses, although with an underlying mechanism that remains unclear. The miR482/2118 family is a conserved miRNA family regulating *NLRs* in a wide range of plant species, and it is probably the most complex family in a large set of miRNAs that target different regions of *NLRs*. Interestingly, in addition to miR482/2118 family in dicots, the miR9863 family seems to be restricted to the *Triticeae* (Liu *et al.*, 2014a). Therefore, a question that remains to be solved by future studies is to understand whether and how other monocots that lack *NLR*-targeting miRNAs regulate immunity at the level of *NLR* transcripts.

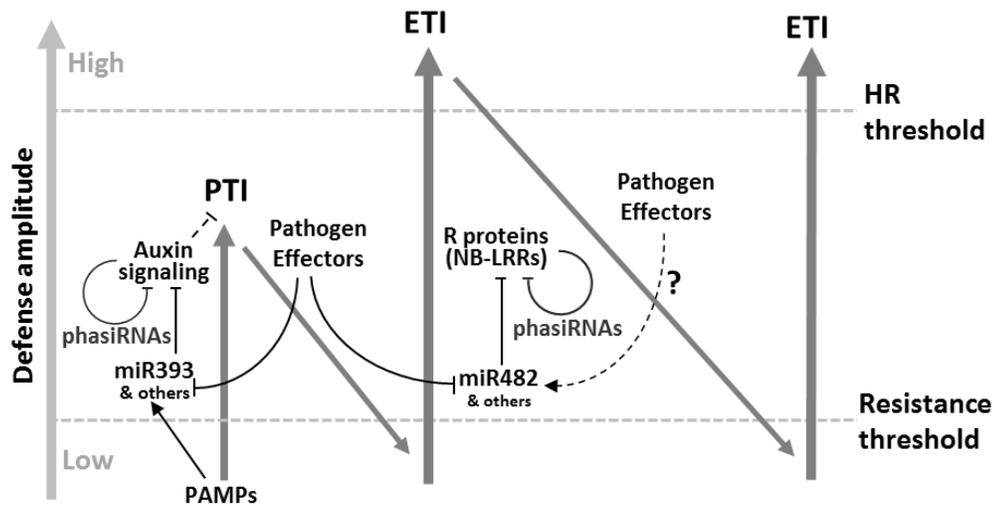


Figure 28. The integration of miRNAs and phasiRNAs in the “zig-zag-zig” model of the plant immune system.

The original model by Jones & Dangl (2006) describes a stepped, evolutionary model for defenses that describes the quantitative nature and molecular evolution of disease resistance in plants. In a variation on that model, PAMPs induce the accumulation of miRNAs that participate in PTI via hormone crosstalk. For example, miR393, which targets genes that are involved in auxin signaling (TIR1, AFB2, and AFB3) is induced upon treatment of flg22. The repression of auxin signaling during infection enhances host PTI by hormone crosstalk. PhasiRNAs are triggered by miR393, which enhances the activity this miRNA by targeting genes involved in the auxin signaling pathway. Effectors from pathogens can suppress the levels of plant miRNAs, such as miR393, to enhance susceptibility. However, the miR482 family, a negative regulator of plant R genes, can also be repressed upon detection of effectors to enhance ETI. Some miRNAs can trigger of phasiRNA biogenesis from R genes, and these phasiRNAs may function synergistically with miRNAs either in cis or trans to suppress R gene transcript levels. Some effectors may also promote miRNA stability or production by targeting the cellular machinery involved in small RNA biogenesis or turnover. Some aspects of this model are speculative (marked with “?”), for example, whether effectors may specifically activate miR482 expression to attenuate host ETI.

REFERENCES

- Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouche N, Gascioli V, Vaucheret H.** 2006. DRB4-dependent *TAS3* trans-acting siRNAs control leaf morphology through AGO7. *Current Biology* **16**, 927-932.
- Allen E, Xie Z, Gustafson AM, Carrington JC.** 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207-221.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR.** 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653-657.
- Arif MA, Fattash I, Ma Z, Cho SH, Beike AK, Reski R, Axtell MJ, Frank W.** 2012. DICER-LIKE3 activity in *Physcomitrella patens* *DICER-LIKE4* mutants causes severe developmental dysfunction and sterility. *Mol Plant* **5**, 1281-1294.
- Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdes-Lopez O, Prince S, Musket TA, Nguyen HT, Stacey G, Meyers BC.** 2014. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell* **26**, 4584-4601.
- Aukerman MJ, Sakai H.** 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* **15**, 2730-2741.
- Axtell MJ.** 2013. Classification and Comparison of Small RNAs from Plants. *Annual Review of Plant Biology* **64**, 137-159.
- Axtell MJ, Jan C, Rajagopalan R, Bartel DP.** 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565-577.
- Beauclair L, Yu A, Bouche N.** 2010. microRNA-directed cleavage and translational repression of the copper chaperone for superoxide dismutase mRNA in *Arabidopsis*. *Plant J* **62**, 454-462.
- Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ.** 1994. RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856-1860.
- Boccaro M, Sarazin A, Thiebauld O, Jay F, Voinnet O, Navarro L, Colot V.** 2014. The *Arabidopsis* miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathog.* **10**, e1003883.

- Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF.** 2013. Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* **23**, 1675-1689.
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK.** 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**, 1279-1291.
- Brodersen P, Voinnet O.** 2006. The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**, 268-280.
- Chabaud M, Boisson-Dernier A, Zhang J, Taylor CG, Yu O, Barker DG.** 2006. Agrobacterium rhizogenes-mediated root transformation. *The Medicago truncatula Handbook, Version November.*
- Chen HM, Chen LT, Patel K, Li YH, Baulcombe DC, Wu SH.** 2010. 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A* **107**, 15269-15274.
- Chen HM, Li YH, Wu SH.** 2007. Bioinformatic prediction and experimental validation of a microRNA-directed tandem *trans*-acting siRNA cascade in *Arabidopsis*. *Proc Natl Acad Sci U S A* **104**, 3318-3323.
- Chen X.** 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* **303**, 2022-2025.
- Chitwood DH, Nogueira FT, Howell MD, Montgomery TA, Carrington JC, Timmermans MC.** 2009. Pattern formation via small RNA mobility. *Genes Dev* **23**, 549-554.
- Chitwood DH, Timmermans MC.** 2010. Small RNAs are on the move. *Nature* **467**, 415-419.
- Cho SH, Coruh C, Axtell MJ.** 2012. miR156 and miR390 regulate tasiRNA accumulation and developmental timing in *Physcomitrella patens*. *Plant Cell* **24**, 4837-4849.
- Cooley MB, Pathirana S, Wu H-J, Kachroo P, Klessig DF.** 2000. Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* **12**, 663-676.
- Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, Takeda A, Sullivan CM, Gilbert SD, Montgomery TA, Carrington JC.** 2010. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nature Structural & Molecular Biology* **17**, 997-1003.
- Cuperus JT, Fahlgren N, Carrington JC.** 2011. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**, 431-442.
- de la Luz Gutierrez-Nava M, Aukerman MJ, Sakai H, Tingey SV, Williams RW.** 2008. Artificial *trans*-acting siRNAs confer consistent and effective gene silencing. *Plant Physiol* **147**, 543-551.
- Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MC.** 2014. Genome-Wide Analysis of leafbladeless1-Regulated and Phased Small RNAs Underscores the Importance of the TAS3 ta-siRNA Pathway to Maize Development. *PLoS Genet.* **10**, e1004826.

- Dunoyer P, Schott G, Himber C, Meyer D, Takeda A, Carrington JC, Voinnet O.** 2010. Small RNA duplexes function as mobile silencing signals between plant cells. *Science* **328**, 912-916.
- Eamens AL, Agius C, Smith NA, Waterhouse PM, Wang MB.** 2011. Efficient silencing of endogenous microRNAs using artificial microRNAs in *Arabidopsis thaliana*. *Mol Plant* **4**, 157-170.
- Fahlgren N, Carrington JC.** 2010. miRNA target prediction in plants. *Plant MicroRNAs*: Springer, 51-57.
- Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington JC.** 2006. Regulation of AUXIN RESPONSE FACTOR3 by *TAS3* tasiRNA affects developmental timing and patterning in *Arabidopsis*. *Curr Biol* **16**, 939-944.
- Fei Q, Xia R, Meyers BC.** 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**, 2400-2415.
- Felippes FF, Weigel D.** 2009. Triggering the formation of tasiRNAs in *Arabidopsis thaliana*: the role of microRNA miR173. *EMBO Reports* **10**, 264-270.
- Feller A, Machemer K, Braun EL, Grotewold E.** 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant Journal* **66**, 94-116.
- Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, Wang Z, Zhang Z, Zheng R, Yang L, Zeng L, Liu X, Zhu JK.** 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc Natl Acad Sci U S A* **111**, 4632-4637.
- Filipowicz W, Bhattacharyya SN, Sonenberg N.** 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9**, 102-114.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J.** 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* **39**, 1033-1037.
- Fukudome A, Kanaya A, Egami M, Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T.** 2011. Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of *Arabidopsis* Dicer-like 4. *RNA* **17**, 750-760.
- Garcia D, Collier SA, Byrne ME, Martienssen RA.** 2006. Specification of leaf polarity in *Arabidopsis* via the trans-Acting siRNA pathway. *Curr. Biol.* **16**, 933-938.
- Gascioli V, Mallory AC, Bartel DP, Vaucheret H.** 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Current Biology* **15**, 1494-1500.
- German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R.** 2008. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**, 941-946.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL.** 2008. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70-W74.

- Gu W, Wang X, Zhai C, Xie X, Zhou T.** 2012. Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. *Mol. Biol. Evol.* **29**, 3037-3044.
- Guo Y-L, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D.** 2011. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiol.* **157**, 757-769.
- Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC.** 2010. The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**, 321-334.
- Heisel SE, Zhang Y, Allen E, Guo L, Reynolds TL, Yang X, Kovalic D, Roberts JK.** 2008. Characterization of unique small RNA populations from rice grain. *PLoS One* **3**, e2871.
- Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE.** 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**, 721-725.
- Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC.** 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.
- Hsieh LC, Lin SI, Shih AC, Chen JW, Lin WY, Tseng CY, Li WH, Chiou TJ.** 2009. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiol* **151**, 2120-2132.
- Hunter C, Willmann MR, Wu G, Yoshikawa M, de la Luz Gutierrez-Nava M, Poethig SR.** 2006. *Trans*-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* **133**, 2973-2981.
- Jin H, Martin C.** 1999. Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Mol Biol* **41**, 577-585.
- Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH.** 2009. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res* **19**, 1429-1440.
- Kallman T, Chen J, Gyllenstrand N, Lagercrantz U.** 2013. A significant fraction of 21 nt sRNA originates from phased degradation of resistance genes in several perennial species. *Plant Physiol* **162**, 741-754.
- Katiyar-Agarwal S, Jin H.** 2010. Role of small RNAs in host-microbe interactions. *Annu. Rev. Phytopathol.* **48**, 225-246.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E.** 2007. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278-1284.
- Komiya R, Ohyanagi H, Niihama M, Watanabe T, Nakano M, Kurata N, Nonomura K.** 2014. Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J.* **78**, 385-397.
- Kozomara A, Griffiths-Jones S.** 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68-73.

- Krasnikova MS, Milyutina IA, Bobrova VK, Ozerova LV, Troitsky AV, Solovyev AG, Morozov SY.** 2009. Novel miR390-dependent transacting siRNA precursors in plants revealed by a PCR-based experimental approach and database analysis. *J Biomed Biotechnol* **2009**, 952304.
- Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, 1.
- Law JA, Jacobsen SE.** 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**, 204-220.
- Lee TF, Gurazada SG, Zhai J, Li S, Simon SA, Matzke MA, Chen X, Meyers BC.** 2012. RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats. *Epigenetics* **7**, 781-795.
- Li F, Orban R, Baker B.** 2012a. SoMART: a web server for plant miRNA, tasiRNA and target gene analysis. *Plant J* **70**, 891-901.
- Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B.** 2012b. MicroRNA regulation of plant innate immune receptors. *Proc Natl Acad Sci U S A* **109**, 1790-1795.
- Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, Valladares O, Yang J, Bambina S, Sabin LR, Murray JI, Lamitina T, Raj A, Cherry S, Wang LS, Gregory BD.** 2012c. Global analysis of RNA secondary structure in two metazoans. *Cell Rep* **1**, 69-82.
- Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD.** 2012d. Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* **24**, 4346-4359.
- Li J, Reichel M, Millar AA.** 2014a. Determinants beyond both complementarity and cleavage govern microR159 efficacy in *Arabidopsis*. *PLoS Genet.* **10**, e1004232.
- Li J, Yang Z, Yu B, Liu J, Chen X.** 2005. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501-1507.
- Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J.** 2013. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* **31**, 688-691.
- Li Y, Lu YG, Shi Y, Wu L, Xu YJ, Huang F, Guo XY, Zhang Y, Fan J, Zhao JQ, Zhang HY, Xu PZ, Zhou JM, Wu XJ, Wang PR, Wang WM.** 2014b. Multiple rice microRNAs are involved in immunity against the blast fungus *Magnaporthe oryzae*. *Plant Physiol.* **164**, 1077-1092.
- Li Y, Zhang Q, Zhang J, Wu L, Qi Y, Zhou JM.** 2010. Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. *Plant Physiol.* **152**, 2222-2231.
- Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC.** 2010. An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol* **10**, 50.

- Liu J, Cheng X, Liu D, Xu W, Wise R, Shen QH.** 2014a. The miR9863 family regulates distinct *Mla* alleles in barley to attenuate NLR receptor-triggered disease resistance and cell-death signaling. *PLoS Genet.* **10**, e1004755.
- Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC.** 2007. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. *Plant J* **52**, 133-146.
- Liu Q, Wang F, Axtell MJ.** 2014b. Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. *Plant Cell* **26**, 741-753.
- Liu X, Huang J, Wang Y, Khanna K, Xie Z, Owen HA, Zhao D.** 2010. The role of floral organs in carpels, an Arabidopsis loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. *The Plant Journal* **62**, 416-428.
- Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I.** 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103.
- Ma C, Lu Y, Bai S, Zhang W, Duan X, Meng D, Wang Z, Wang A, Zhou Z, Li T.** 2014. Cloning and characterization of miRNAs and their targets, including a novel miRNA-targeted NBS-LRR protein class gene in apple (Golden Delicious). *Mol. Plant* **7**, 218-230.
- Mallory AC, Bartel DP, Bartel B.** 2005. MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* **17**, 1360-1375.
- Manavella PA, Koenig D, Weigel D.** 2012. Plant secondary siRNA production determined by microRNA-duplex structure. *Proc Natl Acad Sci U S A* **109**, 2461-2466.
- Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, Nussaume L, Crespi MD, Maizel A.** 2010. miR390, *Arabidopsis* TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell* **22**, 1104-1117.
- Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ.** 2009. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* **21**, 367-376.
- McCormick KP, Willmann MR, Meyers BC.** 2011. Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* **2**, 1.
- McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK.** 2015. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J.* **34**, 20-35.
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel**

- D, Zhu JK.** 2008. Criteria for annotation of plant microRNAs. *Plant Cell* **20**, 3186-3190.
- Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y.** 2008. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116-127.
- Molnar A, Melnyk CW, Bassett A, Hardcastle TJ, Dunn R, Baulcombe DC.** 2010. Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* **328**, 872-875.
- Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC.** 2008a. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in *TAS3* trans-acting siRNA formation. *Cell* **133**, 128-141.
- Montgomery TA, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, Sullivan CM, Alexander A, Nguyen G, Allen E, Ahn JH, Carrington JC.** 2008b. AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc Natl Acad Sci U S A* **105**, 20055-20062.
- Moscou MJ, Bogdanove AJ.** 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501.
- Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD.** 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312**, 436-439.
- Navarro L, Jay F, Nomura K, He SY, Voinnet O.** 2008. Suppression of the microRNA pathway by bacterial effector proteins. *Science* **321**, 964-967.
- Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S.** 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 691-693.
- O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I.** 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Molecular Biology and Evolution* **25**, 1120-1128.
- Ouyang S, Park G, Atamian HS, Han CS, Stajich JE, Kaloshian I, Borkovich KA.** 2014. MicroRNAs suppress NB domain genes in tomato that confer resistance to *Fusarium oxysporum*. *PLoS Pathog.* **10**, e1004464.
- Padmanabhan C, Zhang X, Jin H.** 2009. Host small RNAs are big contributors to plant innate immunity. *Curr. Opin. Plant Biol.* **12**, 465-472.
- Pelaez P, Sanchez F.** 2013. Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Front. Plant Sci.* **4**, 343.
- Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS.** 2004. *SGS3* and *SGS2/SDE1/RDR6* are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev* **18**, 2368-2379.
- Petroni K, Tonelli C.** 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science* **181**, 219-229.
- Pumplin N, Voinnet O.** 2013. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* **11**, 745-760.

- Rajagopalan R, Vaucheret H, Trejo J, Bartel DP.** 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* **20**, 3407-3425.
- Ren G, Chen X, Yu B.** 2012. Uridylation of miRNAs by hen1 suppressor1 in *Arabidopsis*. *Curr. Biol.* **22**, 695-700.
- Robinson MD, McCarthy DJ, Smyth GK.** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Schmittgen TD, Livak KJ.** 2008. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* **3**, 1101-1108.
- Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D.** 2006. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**, 1121-1133.
- Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C.** 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* **31**, 686-688.
- Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC.** 2012. A microRNA superfamily regulates Nucleotide Binding Site-Leucine-Rich Repeats and other mRNAs. *Plant Cell* **24**, 859-874.
- Si-Ammour A, Windels D, Arn-Bouloires E, Kutter C, Ailhas J, Meins F, Vazquez F.** 2011. miR393 and secondary siRNAs regulate expression of the TIR1/AFB2 auxin receptor clade and auxin-related development of *Arabidopsis* leaves. *Plant Physiol.* **157**, 683-691.
- Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C, Chu C, Wang XJ, Green PJ, Meyers BC, Cao X.** 2012. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant Journal* **69**, 462-474.
- Sunkar R, Kapoor A, Zhu JK.** 2006. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell* **18**, 2051-2065.
- Takos AM, Jaffe FW, Jacob SR, Bogs J, Robinson SP, Walker AR.** 2006. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol* **142**, 1216-1232.
- Talmor-Neiman M, Stav R, Klipcan L, Buxdorf K, Baulcombe DC, Arazi T.** 2006. Identification of *trans*-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. *Plant J* **48**, 511-521.
- Todesco M, Rubio-Somoza I, Paz-Ares J, Weigel D.** 2010. A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. *PLoS Genet* **6**, e1001031.
- Tu B, Liu L, Xu C, Zhai J, Li S, Lopez MA, Zhao Y, Yu Y, Ramachandran V, Ren G, Yu B, Li S, Meyers BC, Mo B, Chen X.** 2015. Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in *Arabidopsis*. *PLoS Genet* **11**, e1005119.
- Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R.** 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* **20**, 515-524.

- Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert JL, Bartel DP, Crete P.** 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell* **16**, 69-79.
- Voinnet O.** 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**, 669-687.
- Wang J-W, Wang L-J, Mao Y-B, Cai W-J, Xue H-W, Chen X-Y.** 2005. Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* **17**, 2204-2216.
- Wang X, Zhang S, Dou Y, Zhang C, Chen X, Yu B, Ren G.** 2015. Synergistic and Independent Actions of Multiple Terminal Nucleotidyl Transferases in the 3' Tailing of Small RNAs in *Arabidopsis*. *PLoS Genet.* **11**, e1005091.
- Wang Y, Itaya A, Zhong X, Wu Y, Zhang J, van der Knaap E, Olmstead R, Qi Y, Ding B.** 2011. Function and evolution of a microRNA that regulates a Ca²⁺-ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth. *Plant Cell* **23**, 3185-3203.
- Williams L, Carles CC, Osmont KS, Fletcher JC.** 2005. A database analysis method identifies an endogenous *trans*-acting short-interfering RNA that targets the *Arabidopsis* *ARF2*, *ARF3*, and *ARF4* genes. *Proc Natl Acad Sci U S A* **102**, 9703-9708.
- Willmann MR, Endres MW, Cook RT, Gregory BD.** 2011. The functions of RNA-dependent RNA Polymerases in *Arabidopsis*. *Arabidopsis Book* **9**, e0146.
- Wu L, Mao L, Qi Y.** 2012. Roles of DICER-LIKE and ARGONAUTE proteins in *TAS*-derived small interfering RNA-triggered DNA methylation. *Plant Physiol* **160**, 990-999.
- Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C, Qi Y.** 2010. DNA methylation mediated by a microRNA pathway. *Mol Cell* **38**, 465-475.
- Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z.** 2013. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA Biogenesis in Eudicots. *Plant Cell* **25**, 1555-1572.
- Xia R, Xu J, Arikait S, Meyers BC.** 2015. Extensive families of miRNAs and *PHAS* loci in Norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Mol. Biol. Evol.*
- Xia R, Zhu H, An YQ, Beers EP, Liu Z.** 2012. Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol* **13**, R47.
- Xie Z, Allen E, Wilken A, Carrington JC.** 2005. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **102**, 12984-12989.
- Yan J, Gu Y, Jia X, Kang W, Pan S, Tang X, Chen X, Tang G.** 2012. Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* **24**, 415-427.
- Yang L, Xu M, Koo Y, He J, Poethig RS.** 2013. Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of *MIR156A* and *MIR156C*. *Elife* **2**, e00260.

- Yoshikawa M, Iki T, Tsutsui Y, Miyashita K, Poethig RS, Habu Y, Ishikawa M.** 2013. 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proc Natl Acad Sci U S A* **110**, 4117-4122.
- Yoshikawa M, Peragine A, Park MY, Poethig RS.** 2005. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev* **19**, 2164-2175.
- Young ND, Debelle F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KA, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisine N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dubez AM, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong DH, Jing Y, Jocker A, Kenton SM, Kim DJ, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun JH, Najjar FZ, Nicholson C, Noirot C, O'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang BB, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA.** 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520-524.
- Yue JX, Meyers BC, Chen JQ, Tian D, Yang S.** 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol.* **193**, 1049-1063.
- Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC.** 2011. MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, *trans*-acting siRNAs. *Genes Dev* **25**, 2540-2553.
- Zhai J, Zhang H, Arikrit S, Huang K, Nan G, Walbot V, Meyers BC.** 2015. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3146-3151.
- Zhai J, Zhao Y, Simon SA, Huang S, Petsch K, Arikrit S, Pillay M, Ji L, Xie M, Cao X, Yu B, Timmermans M, Yang B, Chen X, Meyers BC.** 2013. Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. *Plant Cell* **25**, 2417-2428.
- Zhang C, Li G, Wang J, Fang J.** 2012a. Identification of *trans*-acting siRNAs and their regulatory cascades in grapevine. *Bioinformatics* **28**, 2561-2568.

- Zhang C, Ng DW, Lu J, Chen ZJ.** 2012b. Roles of target site location and sequence complementarity in *trans*-acting siRNA formation in *Arabidopsis*. *Plant Journal* **69**, 217-226.
- Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, Mo B, Chen X.** 2012. The *Arabidopsis* nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.* **22**, 689-694.
- Zhou G-K, Kubo M, Zhong R, Demura T, Ye Z-H.** 2007. Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in *Arabidopsis*. *Plant and Cell Physiology* **48**, 391-404.
- Zhu H, Xia R, Zhao B, An YQ, Dardick CD, Callahan AM, Liu Z.** 2012. Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biol* **12**, 149.

Appendix

PUBLICATIONS RELATED TO THIS DISSERTATION

Fei, Q., Xia, R. and Meyers, B.C. (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant Cell*, 25, 2400-2415.

Fei, Q., Li, P., Teng, C., Meyers, B.C. (2015) Secondary siRNAs from Medicago *NB-LRRs* modulated via miRNA-target interactions and their abundances. *The Plant Journal*. 83(3):451-65.

Fei, Q., Zhang, Y., Xia, R., Meyers, B.C. (2016) Small RNAs add zing to the zig-zag-zig model of plant defenses. *Molecular Plant-Microbe Interactions*. 29(3):165-9.