UNIVERSITY OF CALIFORNIA
RIVERSIDE

Identification and Functional Analysis of Bacterial-responsive Small RNAs and Their Targets in Plant Immunity

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Pathology

by

Shang Gao

December 2011

Dissertation Committee:
Dr. Hailing Jin, Chairperson
Dr. Shou-Wei Ding
Dr. Thomas Eulgem

The Dissertation of Shang Gao is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

The text of this thesis, in part or in full, is a reprint of the material as it appears in:

1. Book Chapter in Non Coding RNAs in Plants (2011)

(With kind permission from Springer Science+Business Media: <Non Coding RNAs in Plants, Host Small RNAs and Plant Innate Immunity, 2011, 21-34, Shang Gao and Hailing Jin, 1, Copyright 2011 Springer Science+Business Media>)

2. Genome Biology 11, R81 (2010)

The co-author, Dr. Hailing Jin directed and supervised the research which forms the foundation for this dissertation.
Shang Gao performed the biological experiments with figures 2, 3, 4, 5 and 6.
Dr. W. Zhang, J. Xia, X. Zhou and Dr. X. Zhou carried out the computational analyses.
Dr. P. Chellappan performed the biological experiment with figure 4.
Dr. X. Zhang participated in the biological experiment.

3. Plant Molecular Biology 75, 93 (2011)

The co-author, Dr. Hailing Jin directed and supervised the research which forms the foundation for this dissertation.
Shang Gao performed the biological experiments with figures 4 and 6.
Dr. W. Zhang, X. Zhou Z. Chen and Dr. X. Zhou carried out the computational analyses.
Dr. P. Chellappan performed the biological experiment with figure 5.
Dr. X. Zhang, N. Fromuth, G. Coutino and Dr. M. Coffey provided technical expertise.

I would like to express my greatest appreciation to my major advisor, Professor Hailing Jin, for her devoted guidance, constant encouragement and endless patience during the past five years. Without her guidance and persistent help, this dissertation would not have been possible.

I wish to express my warm and sincere thanks to my dissertation guidance committee members Dr. Shou-Wei Ding and Dr. Thomas Eulgem for providing valuable research advice, support and encouragement in the course of completing this dissertation.

I am grateful to all the past and current members of the Jin lab for sharing their experience and knowledge, and their unforgettable friendship. I would like to thank Dr. Surekha Katiyar-Agarwal for sharing me with a lot of experimental skills when I began my Ph.D. study in Jin lab. I also thank Eric Wu and Debbie Wong for their help as undergraduates in conducting my experiment. My gratitude also goes to all the faculty, staffs and fellow students in the department of Plant Pathology and Microbiology for their help, encouragement and friendship.

A huge thank you is going to my beloved husband and parents for their infinite love, support and patience. Without them, I would not have been able to accomplish what I did. Last but least, I want to thank to other members of my family and friends for their help and moral support.

ABSTRACT OF THE DISSERTATION

Identification and Functional Analysis of Bacterial-responsive Small RNAs and Their Targets in Plant Immunity

by

Shang Gao

Doctor of Philosophy, Graduate Program in Plant Pathology
University of California, Riverside, December 2011
Dr. Hailing Jin, Chairperson

Small RNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs), are geared to respond to the complexity of developmental and physiological processes and negatively regulated gene expression by guiding chromatin modification mRNA degradation, and translational repression of their targets.

Increasing evidence indicates that host small RNAs play a crucial role during the plant-pathogen interactions. In the first chapter of this thesis, we reviewed pathogen-responsive endogenous miRNAs and siRNAs in plants. In addition, we discussed the important components involved in the canonical RNA silencing pathways, including dsRNA-specific endoribonuclease (RNase) Dicer-like proteins (DCLs), RNA-dependent RNA polymerases (RDRs) and Argonaute (AGO) Proteins.

MiRNAs are short noncoding RNA molecules with 21-24 nucleotides and are found in most eukaryotic cells. In the second chapter of this thesis, we globally profiled plant miRNAs in response to infection of bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) in *Arabidopsis* and found 19 miRNA precursors that can yield multiple distinct miRNA-like RNAs in addition to miRNAs and miRNA*s. These miRNA

precursor-derived miRNA-like RNAs are often arranged in phase and form duplexes with approximately 2 nucleotides 3'-end overhang. These miRNA-like RNAs are potentially functional at least under certain conditions.

In the third chapter of this thesis, we identified 15, 27 and 20 miRNA families being differentially regulated after Pst DC3000 hrcC, *Pst* DC3000 *EV* and *Pst* DC3000 *avrRpt2* infections, respectively, through the deep sequencing results from 13 small-RNA libraries. Some bacteria-regulated miRNAs targets have been annotated and believed to be involved in plant hormone biosynthesis and signaling pathways, including auxin, abscisic acid, and jasmonic acid pathways, which suggests the regulation of plant innate immunity through fine-tuning of multiple plant hormone pathways. Moreover, we discussed the procedural differences between sequencing-based profiling and small RNA Northern blot and stated the possibilities that cause the inconsistency.

AtlsiRNA-1, as the first example of the long siRNAs identified in *Arabidopsis*, contributes to resistance against *Pseudomonas syringae* infection by degrading the negative regulator of disease resistance, *AtRAP*. How does *AtRAP* regulate defense response became the major focus of my study to elucidate the mechanisms of plant innate immunity. In the fourth chapter of this thesis, I further investigated the molecular mechanism of *AtRAP* by characterizing the subcellular localization of *AtRAP* and determining the function of C-terminal RAP domain in disease resistance.

Table of Contents

List of Fugures

CHAPTER IV Characterization of AtlsiRNA-1 target gene *AtRAP* and its RNA binding activities in *Arabidopsis thaliana*

List of Tables

CHAPTER 1

Host Small RNAs and Plant Innate Immunity

ABSTRACT

Small non-coding RNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs), are crucial gene expression modulators in most eukaryotes. Increasing evidence indicates that host small RNAs also play a critical role during the plant immune responses. In this chapter, we discuss the functions of these pathogen-responsive endogenous small RNAs and the silencing pathway components during pathogen–host interactions.

1. INTRODUCTION

Plants have evolved innate immune systems to protect themselves from invading microorganisms (Chisholm et al. 2006 ; Jones and Dangl 2006 ; Bent and Mackey 2007). Plant extracellular surface receptors, so-called pattern recognition receptors (PRRs), can recognize conserved microbe-associated molecular patterns (MAMPs) or pathogen-associated molecular patterns (PAMPs) and trigger general defense responses, which are referred to as PAMP-triggered immunity (PTI), mainly through activating MAP kinase cascade and WRKY transcription factors. Many microbes have developed countermeasures by delivering effector proteins into the plant cell and suppressing host PTI pathways. Many plant hosts have subsequently evolved resistance (R) genes to overcome the suppression of effectors by triggering so-called effector-triggered immunity (ETI). ETI responses are more robust and rapid than PTI. In ETI, the pathogens are often repelled by localized cell death at the site of infection, and the whole plants become totally immune to the pathogen. Host endogenous small RNAs have been recognized as important regulators in gene expression reprogramming during both PTI and ETI responses (Jin 2008; Voinnet 2008; Padmanabhan et al. 2009). In this chapter, we discuss the roles of host endogenous small RNAs and small RNA pathway components in plant immunity.

2. Small RNAs

2.1 miRNAs

microRNAs (miRNAs) are generated from single-stranded RNA precursors with hairpin structures by Dicer or Dicer-like (DCL) proteins, a class of RNase III type endoribonucleases. Emerging evidence indicates that some miRNAs function in plant defense responses against pathogen attacks. These miRNAs are upregulated or downregulated during pathogen infection to suppress negative regulators or to release positive regulators of immune responses, respectively (Fig. 1.1).

Arabidopsis miR393 was the first example of miRNA that regulates PTI in antibacterial defense (Navarro et al. 2006). miR393 is induced by a bacterial elicitor flagellin-derived MAMP flg22 and targets Auxin receptors, the F-box genes TIR1, AFB2, and AFB3. In miR393 overexpressing lines, the growth of bacterium Pseudomonas syringae pv. *tomato* (*Pst*) DC3000 was reduced, indicating that miR393-mediated suppression of auxin signaling contributes to PTI. *AFB1*, a third paralog of *TIR1*, is partially resistant to miR393-directed cleavage. The virulent Pst DC3000 can accumulate 20-fold higher in AFB1-Myc-overexpressing Arabidopsis in a tir1-1 mutant background compared to the *tir1-1* plants. This enhanced disease symptom is mainly caused by the constitutive overexpression of *AFB1*, which confirmed a negative role of auxin signaling in plant immune responses. However, no difference was observed when the same plants were inoculated with avirulent *Pst* DC3000 carrying a type-III effector gene, AvrRpt2, which can trigger race-specific resistance ETI. Thus, auxin signaling has antagonistic effect mainly in PTI, but not in ETI. Fahlgren et al. (2007) further confirmed that miR393 can be strongly induced by more than ten-fold at 3 h post inoculation (hpi) by nonpathogenic *Pst* DC3000 *hrcC*, a strain that was mutated in the type III secretion system (TTSS). In addition, miR160 and miR167 can also be upregulated by *Pst* DC3000 *hrcC* at 3 hpi by fiveand six-fold respectively, they both target Auxin-responsive factors (*ARF*) in auxin signaling pathway (Rhoades et al. 2002). Thus, at least three bacteria-induced miRNA families repress the auxin signaling and contribute to the PTI in plants. Auxin is a plant growth-promoting hormone that is antagonistic to salicylic acid (SA)-

mediated defense pathways. Upon perceiving the pathogen PAMPs, these miRNAs are induced to rapidly repress the auxin signaling and shift the energy from plant growth to defense responses. By profiling RNA silencing effector AGO1-bound small RNAs, Li et al. identified that miR158, miR160, miR161.2, miR169a, miR391, miR396a, miR399f, miR822, miR824, and miR1888 are also induced by flg22 (Li et al. 2010). Further study with the transgenic Arabidopsis that overexpressed these miRNAs confirmed that miR160, but not miR158a positively regulate PTI responses.

During the pathogen attacks, a group of miRNAs is also downregulated to accumulate certain target mRNAs, which may contribute positively to defense responses. miR398 is downregulated in response to avirulent strains of *Pst* (*avrRpm1*) or *Pst* (*avrRpt2*) at 12 hpi and continued until 24 hpi (Jagadeeswaran et al. 2009). Only a small reduction was observed for miR398 after infection of virulent *Pst* DC3000 at 24 hpi. The targets of miR398 are Cu/Zn superoxide dismutases 1 and 2 (*CSD1* and *CSD2*). Both targets were regulated by miR398 when Arabidopsis seedlings were exposed to high Cu2+ or high Fe3+ (Sunkar et al. 2006). However, during the Pst infection, only *CSD1* level was upregulated and negatively correlated with miR398 levels, but not *CSD2*. These data suggest that miRNAs are likely to regulate different subgroup of target genes under different conditions. It has been found that miR398 negatively regulated PAMP-induced callose deposition (Li et al. 2010). In addition, expression of miR156, miR168, and miR773 was also reduced upon flg22 treatment (Li et al. 2010). Stable transgenic plants overexpressing miR398b and miR773 showed enhanced susceptibility to *Pst hrcC*, which indicated their negative roles of these two miRNAs in PTI defense.

The profile of miRNA expression in response to virulent *Agrobacterium* (*A.*) *tumefaciens*-infection was different from that in the tumors induced by A. tumefaciens (Dunoyer et al. 2006; Pruss et al. 2008). The levels of several conserved miRNAs, such as miR171, and *Arabidopsis*-specific miRNAs, such as miR163, were reduced by about two-fold in the tumors. Moreover, the levels of miR393 and miR167 were repressed to the limit of detectable level in the tumors, which led to the derepression of the auxin signaling pathway and promote tumor growth. Interestingly, similar phenomenon was

also observed in human malignancies. miR-103/107 could attenuate miRNA biosynthesis by targeting Dicer to benefit the metastasis in human breast cancer (Martello et al. 2010). Inhibition of miR-103/107 opposes migration and metastasis of malignant cells. These studies suggest that global downregulation of miRNA may be a common feature of tumor growth, with only a few induced miRNAs.

2.2 siRNAs

Plants only contain several hundred miRNAs, which are largely conserved and limited in number as compared with siRNAs, which are not conserved and numerous in number. Several endogenous siRNAs have been identified to play an important role in both ETI and PTI (Fig. 1) (Katiyar-Agarwal et al. 2006, 2007).

Arabidopsis nat-siRNAATGB2 was the first example of plant endogenous siRNAs that regulates gene expression in ETI responses (Fig. 1.1) (Katiyar-Agarwal et al. 2006). nat-siRNAATGB2 is induced specifically by *Pst* DC3000 carrying effector *avrRpt2*. It was generated by Dicer-like1 (DCL1) from the overlapping region of a natural antisense transcript (NATs) pair and targets the antisense *PPRL* transcript for degradation. The biogenesis of this siRNA was dependent on HYL1, HEN1, RDR6, SGS3, and NRPD1a. Accumulation of this siRNA depends on the cognate R gene RPS2 and the downstream signaling component NDR1 gene. Transgenic plants overexpressing the siRNA target gene PPRL showed more growth of *Pst* DC3000 (*avrRpt2*) and delayed hypersensitive response (HR), which indicates that PPRL plays a negative role in antibacterial defense in ETI.

Arabidopsis lsiRNA-1 (Katiyar-Agarwal et al. 2007) is one the novel class of endogenous siRNAs identified by Northern blot analysis during the search for more pathogen-induced small RNAs. This new class of siRNAs is 30–40-nt in length and is mainly induced by bacterial infection or specific growth conditions. Like nat-siRNAATGB2, AtlsiRNA-1 is also strongly and specifically induced by *Pst* DC3000 (*avrRpt2*), and its biogenesis pathway requires *DCL1, HYL1, HEN1, HST1, AGO7, RDR6, NRPD1a*, and *NRPD1b*. The target gene of AtlsiRNA-1 encodes a RNA-binding

protein containing a RAP (RNA-binding domain abundant in Apicomplexans) domain. AtlsiRNA-1 employs a unique mechanism to repress AtRAP mRNA, which is by decapping and 5'-3' decay mediated by an exoribonuclease XRN4. *atrap* mutant shows enhanced resistance to both virulent *Pst* and avirulent *Pst* (*avrRpt2*).

According to the widely accepted "gene-for-gene" theory (Flor 1956), successful plant disease resistance is triggered by the recognition of the pathogen effector proteins by the cognate R proteins and activates a series of resistance response events, including HRs, leading to local cell death and restrict the pathogen proliferation at the infection zone (Martin et al. 2003). To counteract the continuous evolution of pathogen effectors, plants have evolved many R genes. These R genes are generally clustered in the genome and encode proteins with conserved motifs, which are believed as a consequence of segmental chromosome duplication and rearrangements (Baumgarten et al. 2003; Meyers et al. 2003). In Arabidopsis thaliana ecotype Columbia, RPP4 locus (first discovered in Landsberg erecta for recognition of Peronospora parasitica 5) comprises seven TIR-NBS-LRR class-R genes and is interspersed with three related and two unrelated genes (Noel et al. 1999; Yi and Richards 2007). Among this gene cluster, RPP4 confers resistance to two races of Hyaloperonospora parasitica , while SNC1 (for suppressor of npr1-1, constitutive 1) impairs resistance to both P. syringe and another race of H. parasitica (Stokes and Richards 2002 ; Zhang et al. 2003 ; Yang and Hua 2004), respectively. Yi et al. further extended Yang and Hua's findings that many paralogous R genes in this locus are positively regulated at the transcriptional level by SA amplification loop in SNC1 overexpression plants (Yi and Richards 2007). Moreover, these R genes can be cosuppressed by siRNAs generated at this locus. On the one hand, activated SNC1 can induce SA accumulation and defense responses that in turn inhibit cell growth, so RNA silencing might minimize the fitness cost associated with excessive SNC1 expression. On the other hand, enhanced transcript levels of SNC1 were observed in small RNA biogenesis-deficient mutants such as *dcl4, upf1*, and *ago1*, as well as in transgenic plants expressing P1/HC-Pro suppressor, which indicates that this locus is under the regulatory control of siRNAs.

3 Small RNA Pathway Components Play an Important Role in Plant Defense

3.1 miRNA and siRNA Biogenesis Pathways

miRNAs are derived from miRNA genes, which are transcribed by RNA polymerase II. The resulting single-stranded miRNA precursors could form stem-loop structure and be processed by DCL1 to generate miRNA /miRNA* duplex. DCL1 functions with two other proteins, HYL1 (HYPONASTIC LEAVES 1) and SE (SERRATE), which help ensure the dicing accuracy. HEN1 (HUA ENHANCER 1) methylates the small RNA duplex at the 3' end (Yu et al. 2005), which is a crucial step to stabilize small RNAs. The matured miRNAs are incorporated predominantly into AGO1 to induce silencing of the target mRNA bearing fully or partly complementary sequences. Conversely, siRNAs are derived from perfectly paired region of doublestranded RNA (dsRNA) precursors. These regions could be the product of RNA-dependent RNA polymerase (RDR) or the overlapping regions of NATs. In plants, trans -acting siRNAs (ta-siRNAs), heterochromatic siRNAs (hc-siRNAs) or repeat-associated siRNAs (ra-siRNAs), nat-siRNAs, and lsiRNAs are four types of siRNAs identified so far with distinct biogenesis pathways (Katiyar-Agarwal and Jin 2010). During the arms race between host and pathogens, viruses and bacteria have evolved countermeasures to suppress host RNAi machinery. Here, we discuss the roles of small RNA pathway components in plant defense responses and silencing suppressors the pathogens evolved to suppress these RNAi pathway components.

3.2 DCLs and Associated Proteins

Arabidopsis encodes four DCLs; DCL1 is the major enzyme for processing miRNAs, and all the four DCLs are involved in siRNA formation. Some miRNAs, such as miR393, participate in PTI (Navarro et al. 2006). It has been shown that *dcl1-9* mutant is more susceptible to *Pst* DC3000 *hrcC*, and the transcriptional level of basal defense marker gene WRKY30 was greatly reduced upon infection (Navarro et al. 2008).

Moreover, PAMP-induced callose deposition was reduced to 50–70% in *Pst hrcC*-infected *dcl1-9* mutant (Li et al. 2010). These results confirmed the importance of DCL1 in PTI.

In antiviral immunity, DCL4 is the primary sensor and generates 21-nt siRNAs from several RNA and DNA viruses (Blevins et al. 2006; Deleris et al. 2006; Fusaro et al. 2006; Diaz-Pendon et al. 2007). In the absence of DCL4, 22- and 24-nucleotide-long virus-derived siRNAs are produced by DCL2 and DCL3, respectively. DCL1 may play an indirect role as a negative regulator of DCL4 (Qu et al. 2008). Although a number of studies indicate that *dcl2/4* double mutant is susceptible to RNA viruses (Bouche et al. 2006; Deleris et al. 2006; Diaz-Pendon et al. 2007), it has been found that SA-mediated resistance is sufficient to inhibit the CMV and TMV (Tobacco mosaic virus) infection in *dcl2/3/4* triple mutant spraying with SA (Lewsey and Carr 2009).

Small RNA-binding proteins (DRBs) are cofactors of DCLs (Hiraguri et al. 2005; Nakazawa et al. 2007). HYL1, colocalized with DCL1 and SE in nuclear dicing bodies (Fang and Spector 2007; Fujioka et al. 2007; Song et al. 2007), is also an important component involved in antibacterial defense. hyl1-1 mutant is susceptible to Pst (avrRpt2) and fails to accumulate nat-siRNAATGB2 and AtlsiRNA-1 (Katiyar-Agarwal et al. 2006, 2007). dsRNA-binding 4 (*DRB4*) contributes to antiviral defense by interacting with DCL4 (Qu et al. 2008). Pathogens have also evolved silencing suppressors to target various proteins of the RNAi pathways. Viral silencing suppressor P6 protein of Cauliflower mosaic virus (CaMV, family Caulimoviridae) inhibits the biogenesis of 21-nt siRNAs by physically binding with DRB4 protein, which subsequently interferes with the activity of the major plant antiviral silencing factor DCL4 (Haas et al. 2008).

HEN1 contributes to both antibacterial and antiviral defenses. *hen1-1* mutant compromised the resistance to Pst DC3000 hrcC (Navarro et al. 2008). Viral RNA of CMV accumulates five-fold more in *hen1-1* than in wild type (Boutet et al. 2003). Therefore, small RNA methylation and stabilization are crucial for plant defenses. P126 protein from TMV, a RNA replicase, binds to duplex viral siRNAs and inhibits HEN1-

dependent methylation at the 3' end, which destabilizes the viral siRNAs and attenuates the silencing efficiency (Vogler et al. 2007).

3.3 RNA-Dependent RNA Polymerases

Arabidopsis genome encodes six RDRs. RDR1 and RDR6 are considered to be involved in the antiviral defense. The activity of tobacco (*Nicotiana tabacum* cv. *Xanthi*, nn genotype) NtRDR1 was increased in TMV-infected or SA-treated plants. NtRDR1 antisense transgenic plants accumulated more TMV than wild-type control plants, indicating NtRDR1 plays a positive role in antiviral defense (Xie et al. 2001). Similarly, Potato virus Y (PVY) can also induce NtRDR1, and plants with reduced NtRDR1 transcripts accumulate higher PVY. The induction of another three defenserelated genes (mitochondrial alternative oxidase, IVR, and ERF5) is also repressed in this plant (Rakhshandehroo et al. 2009). In *Arabidopsis*, both SA and compatible TMV can induce AtRDR1 gene expression. The authors suggest that the role of RDR1 in antiviral defense may not be through the RNAi pathway, but rather through the SA pathway (Yu et al. 2003). During the interaction between Arabidopsis and CMV-Δ2b (does not express the silencing suppressor protein 2b), viral siRNAs were mapped mostly to three viral RNA regions. Those mapped to the 5' ends are associated with RDR1, whereas those mapped to the 3' regions are associated with RDR6. Viral siRNAs were largely reduced in *rdr1/rdr6* double mutant. Thus, RDR1 and RDR6 work synergistically in *Arabidopsis* to amplify viral siRNAs, and RDR6 acts as a surrogate when RDR1 function is disrupted (Wang et al. 2010). However, mutated TuMV without silencing suppressor regained the ability to move systemically in *rdr1/rdr6* double mutant plants (Garcia-Ruiz et al. 2010). The induction of RDR1 was not sufficient to prevent the systemic spreading of PVY (Rakhshandehroo et al. 2009). Additional function of NtRDR1 was observed in Nicotiana bethamiana , which carries a nonfunctional endogenous NbRDR1 due to a 72-nt insertion (Yang et al. 2004). Transgenic *N. bethamiana* carrying tobacco NtRDR1 exhibits hypersusceptibility to Plum pox potyvirus (PPV) and several other viruses (Ying et al. 2010). The authors believe that transgenic NtRDR1 functions as a silencing suppressor to

inhibit RDR6-dependent posttranscriptional silencing induced by sense transgenes (S-PTGS). *N. bethamiana* is one of the best hosts for many viruses. The silencing suppressor activity of NtRDR1 may explain the natural loss-of-function variant in *N. bethamiana* (Ying et al. 2010).

Some viruses have evolved suppressors that target RDRs directly or indirectly. For example, the V2 protein of *Tomato yellow leaf curl geminivirus* (TYLCV) is a suppressor that outcompetes host SGS3, a component associated with RDR6, for substrate dsRNA recognition at 5' overhang and subsequently protects the viral dsRNAs from degradation (Glick et al. 2008; Elkashef and Ding 2009; Fukunaga and Doudna 2009).

3.4 Argonaute Proteins

Argonaute proteins are the effectors of RNAi machinery that associate with small RNAs and target complementary RNAs for transcriptional or posttranscriptional gene silencing (TGS). There are ten AGOs in Arabidopsis. AGO1 is involved in the biogenesis and function of most miRNAs. Two ago1 mutants (*ago1-25* and *ago1-27*) showed attenuated flg22-induced basal defense responses, including reduced callose deposition, and reduced accumulation of FRK1 and WRKY29 transcripts. Pretreatment of flg22 induces PTI and subsequently inhibits bacterial growth, but this inhibitory effect was attenuated in both ago1 mutants (Li et al. 2010). All of these results indicate that AGO1 is important for PTI against bacteriall infection. Bacteria have developed TTSS effectors, such as HopT1, to inhibit the activity of AGO1. In the SUC-SUL RNAi reporter line, expression of HopT1 inhibits the degradation of miRNA target transcripts, although the miRNA level remained the same. So, HopT1 can be considered as a bacteria-encoded suppressor of RNA silencing (BSRs) (Navarro et al. 2008).

AGO1 is also the target of many virus suppressors. Previous studies (Pazhouhandeh et al. 2006; Baumberger et al. 2007; Bortolamiol et al. 2007) have confirmed that P0 from *beet Western yellow polerovirus* (BWYV) contains an F-box-like domain and promotes proteolysis of AGO1. P0 does not interfere with preassembled

miRNA/siRNA-containing RISC, but acts upstream of AGO1 loading and prevents the formation of small RNA-AGO1 complexes by AGO1 degradation (Csorba et al. 2010). P38 of Turnip crinkle virus (TCV) mimics host-encoded glycine/tryptophan (GW)-containing proteins and sequesters AGO1 from binding to other endogenous GW/WG proteins (Azevedo et al. 2010). Host GW/WG proteins bind to AGO proteins via the GW/WG "AGO-binding hooks" and are required for RISC assembly and activity (Jin and Zhu 2010). P1 protein from *Sweet potato mild mottle virus* (SPMMV, family Potyviridae) is another example of GW/WG containing VSR inhibits RISC activity by direct AGO1 binding. The three WG/GW motifs located at N-terminal part of P1 are efficient "Ago hooks" to inhibit functional RISC assembling. Thus, P38 and P1 proteins encoded by viruses are examples of how pathogens usurp and mimic host regulatory pathways. P25 protein of *Potato virus X* (PVX, family Flexiviridae) strongly interacts with *Arabidopsis* AGO1 through the degradation of AGO1 via the proteasome pathway (Chiu et al. 2010). Plants are less susceptible to PVX and Bamboo mosaic virus when treated with MG132, a proteasome inhibitor. P25 also interacts with AGO2, AGO3, and AGO4, but not with AGO5 and AGO9 by transient assay in *N. benthamiana* leaves. Cucumber mosaic virus (CMV, family Bromoviridae) encoded a nucleus-localized 2b protein, which can physically interact with both AGO1 and siRNAs of RISC and inhibit its slicing action (Diaz-Pendon et al. 2007; Goto et al. 2007). In addition, 2b was also found to interact with AGO4, and the 2b-AGO4 complex was mainly in the nucleus but absent from the nucleolus where the AGO4 and Pol IV complex reside (Gonzalez et al. 2010).

AGO4 and AGO6 are closely related AGO proteins that mediate mainly TGS by DNA methylation and/or chromatin modification (Vaistij et al. 2002; Huettel et al. 2007; Matzke et al. 2009). They predominantly associate with 24-nt siRNAs and silence mainly transposons and repeated sequences. Several studies indicated that viral, bacterial, and fungal infections alter the genomic methylation level and change the expression of many genes (Finnegan et al. 1998; Stokes and Richards 2002; Pavet et al. 2006). Geminiviruses are single-stranded DNA viruses that replicate in cell nuclei via double-stranded DNA

intermediates that associated with host histone proteins (Pilartz and Jeske 1992, 2003). In transgenic plants carrying a geminivirus *Tomato leaf curl virus* (TLCV) gene C4, hypermethylation was observed at the asymmetric cytosines of the C4 transgene and its promoter region, and no C4 transcripts were detected (Bian et al. 2006). These results suggest that plants utilize DNA methylation as a defense mechanism against geminiviruses. Similar results were observed in two other geminiviruses, *Cabbage leaf curl virus* (CaLCuV) and *Beet curly top virus* (BCTV) (Raja et al. 2008). Asymmetric cytosine methylation and histone H3K9 methylation (Gene silencing marker) were reduced in RNAdirected DNA methylation (RdDM) pathway mutants, and AGO4 was also implicated in this process. These data suggest that the DNA genome of geminiviruses is silenced via RdDM pathway in plants. In addition to the antiviral function, AGO4 is also involved in defense response to *P. syringea* (Agorio and Vera 2007). *ago4-2* mutant is susceptible to both virulent *Pst* DC3000 and avirulent *Pst* carrying *avrRpm1* effector compare to the wild-type plants. However, whether this antibacterial function is through RdDM pathway is still not clear because other lossof-function mutants in the RdDM pathway did not show compromised resistance, including genes operating upstream of AGO4, RDR2, and DCL3, and genes operating downstream of AGO4, such as DRD1, CMT3, DRM1, and DRM2.

AGO7 is preferentially associated with miR390, which triggers the production of ta-siRNAs (Montgomery et al. 2008). AGO7 is also required for the accumulation. Of AtlsiRNA-1, suggesting its role in RPS2 -mediated ETI (Katiyar-Agarwal et al. 2007). *ago7* mutant (*zip-1*) attenuated the resistance to *Pst* DC3000 (*avrRpt2*), although it exhibits normal responses in PTI, including callose deposition, FRK1 expression, seedling growth inhibition, and oxidative burst analysis (Li et al. 2010). Moreover, AGO1 and AGO7 have been found to contribute to viral RNA clearance (Qu et al. 2008). Small RNA pathway components also play an important role in defense responses against multicellular eukaryotic pathogens, such as fungi. Inoculation of the sgs mutants (*sgs1-1*, *sgs2-1* and *sgs3-1*) with different pathogenic strains of the Vascular fungi Verticillium all resulted in a similar increased susceptibility phenotype (Ellendorff et al. 2009), but no

altered defense was observed with the similar lifestyle vascular fungal pathogen Fusarium oxysporum f. sp. raphani . Further screening with other small RNA biogenesis mutants found that AGO7, DCL4, NRPD1a, RDR2, AGO1, HEN1, and HST1 all affect Verticillium defense. But those small RNA pathway components do not comply with one single RNA-silencing pathway among well-defined ones. This result suggests that plant defense against Verticillium may involve multiple small RNA-silencing pathways. Further in-depth analysis is needed to verify the specific roles of each component in the immune responses.

4 Conclusions and Future Prospects

Increasing evidence has demonstrated the importance of small RNAs and small RNA machinery in plant immune responses against various pathogens. They play essential roles in gene expression reprogramming and fine-tuning during host-microbial interaction. We predict that more pathogen-regulated host small RNAs and pathogen-derived silencing suppressors will be discovered. However, many important questions still remain. For example, what are the upstream signaling events that regulate small RNA biogenesis and metabolism in response to pathogen attack? What are the regulatory proteins that interact with small RNA pathway components in response to infection? Whether and how are these silencing signals transported? Can small RNAs and small RNA complex shuttle between hosts and pathogens? Increasing evidence suggests that pathogen infection alters host genomic chromatin structures. Whether and how long can these chromatin modifications maintain, and whether they can be transferred to the next generation? More exciting time will come when we will be able to answer these questions.

REFERENCES

Agorio A, Vera P (2007) ARGONAUTE4 is required for resistance to Pseudomonas syringae in Arabidopsis. Plant Cell 19:3778-3790

Azevedo J, Garcia D, Pontier D et al (2010) Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. Genes Dev 24:904-915

Baumberger N, Tsai CH, Lie M et al (2007) The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. Curr Biol 17:1609-1614

Baumgarten A, Cannon S, Spangler R et al (2003) Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics 165:309-319

Bent AF, Mackey D (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Ann Rev Phytopathol 45:399-436

Bian XY, Rasheed MS, Seemanpillai MJ et al (2006) Analysis of silencing escape of tomato leaf curl virus: an evaluation of the role of DNA methylation. Mol Plant Microbe Interact 19:614-624

Blevins T, Rajeswaran R, Shivaprasad PV et al (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34:6233-6246

Bortolamiol D, Pazhouhandeh M, Marrocco K et al (2007) The Polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. Curr Biol 17:1615-1621

Bouche N, Lauressergues D, Gasciolli V et al (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. EMBO J 25:3347-3356

Boutet S, Vazquez F, Liu J et al (2003) Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol 13:843-848

Chisholm ST, Coaker G, Day B et al (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803-814

Chiu MH, Chen IH, Baulcombe DC et al (2010) The silencing suppressor P25 of Potato virus X interacts with Argonaute1 and mediates its degradation through the proteasome pathway. Mol Plant Pathol 11:641-649

Csorba T, Lozsa R, Hutvagner G et al (2010) Polerovirus protein P0 prevents the assembly of small RNA-containing RISC complexes and leads to degradation of ARGONAUTE1. Plant J 62:463-472

Deleris A, Gallego-Bartolome J, Bao J et al (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science 313:68-71

Diaz-Pendon JA, Li F, Li WX et al (2007) Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. Plant Cell 19:2053-2063

Dunoyer P, Himber C, Voinnet O (2006) Induction, suppression and requirement of RNA silencing pathways in virulent Agrobacterium tumefaciens infections. Nat Genet 38:258-263

Elkashef S, Ding SW (2009) Possible new RNA intermediate in RNA silencing. Nat Chem Biol 5:278-279

Ellendorff U, Fradin EF, de Jonge R et al (2009) RNA silencing is required for Arabidopsis defence against Verticillium wilt disease. J Exp Bot 60:591-602

Fahlgren N, Howell MD, Kasschau KD et al (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2:e219

Fang Y, Spector DL (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. Curr Biol 17:818-823

Finnegan EJ, Genger RK, Peacock WJ et al (1998) DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol 49:223-247

Flor HH (1956) The complementary genic systems in Flax and Flax Rust. Adv Genet 8:29-54

Fujioka Y, Utsumi M, Ohba Y et al (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. Plant Cell Physiol 48:1243-1253

Fukunaga R, Doudna JA (2009) dsRNA with 5¢ overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. EMBO J 28:545-555

Fusaro AF, Matthew L, Smith NA et al (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep 7:1168-1175

Garcia-Ruiz H, Takeda A, Chapman EJ et al (2010) Arabidopsis RNA-dependent RNA

polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. Plant Cell 22:481-496

Glick E, Zrachya A, Levy Y et al (2008) Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. Proc Natl Acad Sci USA 105:157-161

Gonzalez I, Martinez L, Rakitina DV et al (2010) Cucumber mosaic virus 2b protein subcellular targets and interactions: their significance to RNA silencing suppressor activity. Mol Plant Microbe Interact 23:294-303

Goto K, Kobori T, Kosaka Y et al (2007) Characterization of silencing suppressor 2b of cucumber mosaic virus based on examination of its small RNA-binding abilities. Plant Cell Physiol 48:1050-1060

Haas G, Azevedo J, Moissiard G et al (2008) Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. EMBO J 27:2102-2112

Hiraguri A, Itoh R, Kondo N et al (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. Plant Mol Biol 57:173-188

Huettel B, Kanno T, Daxinger L et al (2007) RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. Biochim Biophys Acta 1769:358-374

Jagadeeswaran G, Saini A, Sunkar R (2009) Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis. Planta 229:1009-1014

Jin H (2008) Endogenous small RNAs and antibacterial immunity in plants. FEBS Lett 582:2679-2684

Jin H, Zhu JK (2010) A viral suppressor protein inhibits host RNA silencing by hooking up with Argonautes. Genes Dev 24:853-856

Jones JD, Dangl JL (2006) The plant immune system. Nature 444:323-329

Katiyar-Agarwal S, Jin H (2010) Role of small RNAs in host-microbe interactions. Annu Rev Phytopathol 48:225-246

Katiyar-Agarwal S, Morgan R, Dahlbeck D et al (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci USA 103:18002-18007

Katiyar-Agarwal S, Gao S, Vivian-Smith A et al (2007) A novel class of bacteria-induced

small RNAs in Arabidopsis. Genes Dev 21:3123-3134

Lewsey MG, Carr JP (2009) Effects of DICER-like proteins 2, 3 and 4 on cucumber mosaic virus and tobacco mosaic virus infections in salicylic acid-treated plants. J Gen Virol 90:3010-3014

Li Y, Zhang Q, Zhang J et al (2010) Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. Plant Physiol 152:2222-2231

Martello G, Rosato A, Ferrari F et al (2010) A MicroRNA targeting dicer for metastasis control. Cell 141:1195-1207

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23-61

Matzke M, Kanno T, Daxinger L et al (2009) RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol 21:367-376

Meyers BC, Kozik A, Griego A et al (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15:809-834

Montgomery TA, Howell MD, Cuperus JT et al (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133:128-141

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

Navarro L, Dunoyer P, Jay F et al (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436-439

Navarro L, Jay F, Nomura K et al (2008) Suppression of the microRNA pathway by bacterial effector proteins. Science 321:964-967

Noel L, Moores TL, van Der Biezen EA et al (1999) Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis. Plant Cell 11:2099-2112

Padmanabhan C, Zhang X, Jin H (2009) Host small RNAs are big contributors to plant innate immunity. Curr Opin Plant Biol 12:465-472

Pavet V, Quintero C, Cecchini NM et al (2006) Arabidopsis displays centromeric DNA

hypomethylation and cytological alterations of heterochromatin upon attack by Pseudomonas syringae. Mol Plant Microbe Interact 19:577-587

Pazhouhandeh M, Dieterle M, Marrocco K et al (2006) F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. Proc Natl Acad Sci USA 103:1994-1999

Pilartz M, Jeske H (1992) Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. Virology 189:800-802

Pilartz M, Jeske H (2003) Mapping of abutilon mosaic geminivirus minichromosomes. J Virol 77:10808-10818

Pruss GJ, Nester EW, Vance V (2008) Infiltration with Agrobacterium tumefaciens induces host defense and development-dependent responses in the infiltrated zone. Mol Plant Microbe Interact 21:1528-1538

Qu F, Ye X, Morris TJ (2008) Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. Proc Natl Acad Sci USA 105:14732-14737

Raja P, Sanville BC, Buchmann RC et al (2008) Viral genome methylation as an epigenetic defense against geminiviruses. J Virol 82:8997-9007

Rakhshandehroo F, Takeshita M, Squires J et al (2009) The influence of RNA-dependent RNA polymerase 1 on potato virus Y infection and on other antiviral response genes. Mol Plant Microbe Interact 22:1312-1318

Rhoades MW, Reinhart BJ, Lim LP et al (2002) Prediction of plant microRNA targets. Cell 110:513-520

Song L, Han MH, Lesicka J et al (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proc Natl Acad Sci USA 104:5437-5442

Stokes TL, Richards EJ (2002) Induced instability of two Arabidopsis constitutive pathogen-response alleles. Proc Natl Acad Sci USA 99:7792-7796

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

Vaistij FE, Jones L, Baulcombe DC (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative

RNA-dependent RNA polymerase. Plant Cell 14:857-867

Vogler H, Akbergenov R, Shivaprasad PV et al (2007) Modification of small RNAs associated with suppression of RNA silencing by tobamovirus replicase protein. J Virol 81:10379-10388

Voinnet O (2008) Post-transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility. Curr Opin Plant Biol 11:464-470

Wang XB, Wu Q, Ito T et al (2010) RNAi-mediated viral immunity requires amplification of virusderived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci USA 107:484-489

Xie Z, Fan B, Chen C et al (2001) An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. Proc Natl Acad Sci USA 98:6516-6521

Yang S, Hua J (2004) A haplotype-specific Resistance gene regulated by BONZAI1 mediates temperature-dependent growth control in Arabidopsis. Plant Cell 16:1060-1071

Yang SJ, Carter SA, Cole AB et al (2004) A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by Nicotiana benthamiana. Proc Natl Acad Sci USA 101:6297-6302

Yi H, Richards EJ (2007) A cluster of disease resistance genes in Arabidopsis is coordinately regulated by transcriptional activation and RNA silencing. Plant Cell 19:2929-2939

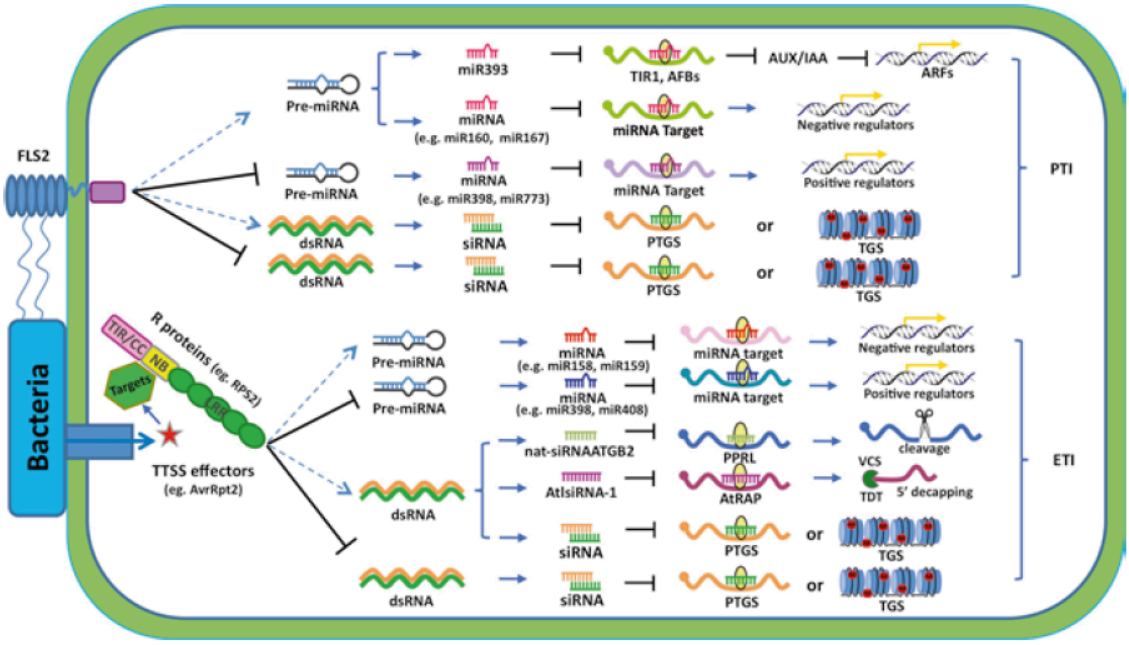
Ying XB, Dong L, Zhu H et al (2010) RNA-dependent RNA polymerase 1 from Nicotiana tabacum suppresses RNA silencing and enhances viral infection in Nicotiana benthamiana. Plant Cell 22:1358-1372

Yu D, Fan B, MacFarlane SA et al (2003) Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense. Mol Plant Microbe Interact 16:206-216

Yu B, Yang Z, Li J et al (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307:932-935

Zhang Y, Goritschnig S, Dong X et al (2003) A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. Plant Cell 15:2636-2646

Figure 1.1 miRNAs and siRNAs involved in the fine-tuning of plant immunity entail a combination of silencing and translational repression through PTI and ETI. Upper panel: after the elicitation of the extracellular domain of FLS2 receptor by bacterial flagellin, an unidentified signaling pathway will either upregulate the level of some miRNAs (eg. miR393, miR160 and miR167) and some siRNAs or downregulate the level of some other small RNAs (eg. miR398 and miR773). miR393 induction during PTI can suppress auxin receptor genes TIR1 and AFBs. The accumulated Aux/IAA inactivates auxin response factors (ARFs) and then suppresses the auxin responsive gene expression to enhance the basal defense. Lower panel: after the delivery of the bacteria TTSS effectors (eg. AvrRpt2) into the plant cell, the cognate R proteins (eg. RPS2) activates a suite of genes, which lead to up-regulation of some miRNAs (eg. miR158 and 159) and some endogenous siRNAs (eg. natsiRNAATGB2 and AtlsiRNA-1) or down-regulation of some small RNAs (eg. miR398 and miR408). Induction of nat-siRNAATGB2 and AtlsiRNA-1 during RPS2-mediated ETI specifically downregulate PPRL or AtRAP genes, respectively, through either target cleavage or mRNA degradation.



CHAPTER 2

Multiple distinct small RNAs originate from the same microRNA precursors

ABSTRACT

Background: MicroRNAs (miRNAs), which originate from precursor transcripts with stem-loop structures, are essential gene expression regulators in eukaryotes.

Results: We report 19 miRNA precursors in *Arabidopsis* that can yield multiple distinct miRNA-like RNAs in addition to miRNAs and miRNA*s. These miRNA precursor-derived miRNA-like RNAs are often arranged in phase and form duplexes with an approximately two-nucleotide 3'-end overhang. Their production depends on the same biogenesis pathway as their sibling miRNAs and does not require RNA-dependent RNA polymerases or RNA polymerase IV. These miRNA-like RNAs are methylated, and many of them are associated with Argonaute proteins. Some of the miRNA-like RNAs are differentially expressed in response to bacterial challenges, and some are more abundant than the cognate miRNAs. Computational and expression analyses demonstrate that some of these miRNA-like RNAs are potentially functional and they target protein-coding genes for silencing. The function of some of these miRNA-like RNAs was further supported by their target cleavage products from the published small RNA degradome data. Our systematic examination of public small-RNA deep sequencing data from four additional plant species (*Oryza sativa, Physcomitrella patens, Medicago truncatula and Populus trichocarpa*) and four animals (*Homo sapiens, Mus musculus, Caenorhabditis elegans and Drosophila*) shows that such miRNA-like RNAs exist broadly in eukaryotes.

Conclusions: We demonstrate that multiple miRNAs could derive from miRNA precursors by sequential processing of Dicer or Dicer-like proteins. Our results suggest that the pool of miRNAs is larger than was previously recognized, and miRNA-mediated gene regulation may be broader and more complex than previously thought.

BACKGROUND

MicroRNAs (miRNAs) are small regulatory RNAs that play a fundamental role in gene expression regulation in eukaryotes through mRNA cleavage, RNA degradation, translation inhibition, or DNA methylation [1-7]. miRNAs belong to a large repertoire of regulatory small RNAs, which also includes small interfering RNAs (siRNAs) [8-11]. Most miRNA genes (*MIR*) are transcribed by RNA polymerase II (Pol II) [12,13]. The resulting single- stranded miRNA precursors fold into stem-loop structures that can be recognized by RNase III-type enzymes, Drosha (as in animals) and Dicer or Dicer-like proteins (DCLs; as in plants), that sequentially cleave the precursors to liberate miRNA-miRNA* duplexes from the hairpins (miRNA* is a small RNA on the opposite arm of the miRNA in the hairpin with partial complementarity to the miRNA) [3,6,14]. The mature miRNAs are subsequently incorporated into Argonaute (AGO) family proteins, and then they target mRNAs through perfect or partially complementary base pairing [15]. miRNAs are normally more abundant than miRNA*s [3,6,14], but there are cases when miRNA* sequences are more abundant and can interact with AGO proteins to exert their function [16]; when the abundances of miRNAs and miRNA*s are comparable, they are called miR-5p and miR-3p, depending on their positions relative to the 5’ -end of the sequences [17,18]. *Arabidopsis* contains four Dicer-like proteins, DCL1 to DCL4. The biogenesis of *Arabidopsis* miRNAs depends mainly on DCL1, with that of a few relying on DCL4 [8,19]. *Arabidopsis* miRNAs are stabilized through 3’- end methylation by the RNA methyltransferase HEN1, which protects them from uridylation and subsequent RNA degradation [20,21].

In contrast to miRNAs, siRNAs are derived from double-stranded RNA molecules and have multiple sources of origin [6,8]. Four classes of siRNAs have been found in plants. The first class includes natural antisense transcript (nat)-siRNA which is derived from cis –natural antisense transcripts, the so-called nat-siRNAs. They are often induced by abiotic and biotic stresses, are generated by DCL1 and/or DCL2, and are often dependent on RNA-dependent RNA polymerase (RDR) 6 and Pol IV [22-25]. The

second class comprises endogenous trans -acting siRNAs (tasiRNAs), which are encoded by TAS genes [8]. miRNA-mediated cleavage of a TAS transcript serves as a template for RDR6 to synthesize a double-stranded RNA, which is subsequently cleaved into approximately 21-nucleotide phased tasiRNAs by DCL4. The third class of siRNAs comprises the heterochromatic siRNAs (hc-siRNAs) [10]. hc-siRNAs normally arise from transposon and repeat regions of the genome, and often silence mobile and repeat elements via DNA methylation and chromatin modification. The formation of hc-siRNAs requires DCL3, RDR2 and Pol IV. The fourth class comprises long siRNAs (lsiRNAs), which are 30 to 40 nucleotides in length [26]. The biogenesis of lsiRNAs requires DCL1 and is also partially dependent on RDR and Pol IV. Therefore, an effective way to distinguish miRNAs from various siRNAs is to examine the major distinctive components of their biogenesis. For example, the biogenesis of miRNAs does not require RDRs or Pol IV.

A structural property of miRNAs is that their precursors form foldback hairpin structures. One miRNAmiRNA* duplex is typically expected to arise from a miRNA precursor [3,14,27]. Nevertheless, some early work also observed additional small RNAs beyond miRNAs and miRNA*s, but such small RNAs were normally considered to be byproducts of Dicer activities and have never been systematically investigated [19,28 32]. Recent studies in animals identified miRNA-offset RNAs (moRNAs) in a chordate [33], human [34], and a herpesvirus [35], but the biogenesis and possible functions of these small RNAs remain to be determined. In a deep-sequencing-based study of small RNAs from bacterial-challenged *Arabidopsis thaliana*, we identified a substantial number of sequencing reads that can map perfectly onto many miRNA precursors even though they do not correspond to the mature miRNA or miRNA* sequences. Most of these small RNAs form pairing partners similar to miRNA-miRNA* duplexes with a two-nucleotide 3’ -end overhang and are arranged in phasing. Moreover, we found that they depend on the same biogenesis pathway as the known miRNAs. Furthermore, multiple lines of evidence suggest that some of these miRNA-like RNAs are authentic miRNAs. First, some of them are differentially expressed upon bacterial challenges, and some are more

abundant than their sibling miRNAs. Second, many of these miRNAlike RNAs can be associated with AGO proteins. Third, some of them have predicted protein-coding targets with similar functions, and several of their target cleavage products are present when performing parallel analysis of RNA ends (PARE) or in degradome data [36-38]. Fourth, expression analysis using Dicer mutants further supports that some of these miRNA-like RNAs silence their predicted target genes. Moreover, our systematic genome-wide survey of publically available small-RNA deep sequencing data shows that such miRNA-like RNAs broadly exist in plants (*Oryza sativa, Physcomitrella patens, Medicago truncatula* and *Populus trichocarpa*) and animals (*Homo sapiens, Mus musculus, Caenorhabditis elegans* and *Drosophila melanogaster*).

RESULTS

To study the role of small RNAs in response to bacterial challenge, we prepared 13 small-RNA libraries from *Arabidopsis* infected with various *Pseudomonas syringae pv. tomato* (*Pst*) DC3000 strains and sequenced them using the Illumina SBS deep-sequencing platform. Sequencing data were collected at 6 and 14 hours postinoculation (hpi) with 10 mM MgCl2 (mock), a type III secretion system mutated strain of *Pst* DC3000 *hrcC*, a virulent strain of *Pst* DC3000 carrying an empty vector (EV), and an avirulent strain of *Pst* DC3000 (*avrRpt2*). *Pst* DC3000 (*avrRpt2*) induces a hypersensitive response (HR) in *Arabidopsis* Col-0 that carries the cognate resistance gene RPS2 and leads to cell death symptoms (the hypersensitive response), usually at 15 to 16 hpi. Our samples were collected at 14 hpi, right before the hypersensitive response could be visualized. From a total of more than 24.6 million sequencing reads from all libraries, 13,985,938 reads perfectly matched the *Arabidopsis* genome and cDNAs, among which 2,578,531 were unique. After excluding reads shorter than 17-nucleotide and any that matched tRNAs, rRNAs, small nuclear RNAs (snRNAs), or small nucleolar RNAs (snoRNAs), the remaining reads were kept for further analysis. We detected the expression of 191 of the 207 *Arabidopsis* miRNAs listed in miRBase. The 13 libraries of sequencing reads have been deposited in the NCBI Gene Expression Omnibus (GEO)

database [GEO: GSE19694] and a summary of the sequencing data is given in Table S1 in Additional file 1.

Multiple distinct miRNA-like RNAs arise from a single miRNA precursor

A key observation from our sequencing data is that multiple unique small-RNA reads could be generated from the same miRNA precursor. Specifically, we found a substantial number of reads that originated from the double-stranded stem regions of many miRNA precursors and yet are not themselves the mature miRNA or miRNA* sequences. In some cases, the number of these small-RNA reads is comparable to or even greater than the number of reads mapped to the mature miRNA or miRNA* sequences. Furthermore, using a set of stringent criteria (see Materials and methods), we observed that many sequencing reads that map to a miRNA precursor were arranged in phase, in which unique small-RNA reads followed one another in tandem or sometimes were separated by a gap of 21 to 22 nucleotides along the precursor [39]. Figures 1, 2, 3, and 4 show this type of phasing pattern on the precursors of miR159a, miR169 m, miR319a/b, miR447, miR822 and miR839. It is important to note that more than one such miRNA-like RNA may appear in a fold-back structure. In total, using a minimum of 5 sequencing reads as a cutoff, we identified 35 miRNA-like RNAs from 19 miRNA precursors in 10 *Arabidopsis* miRNA families, including both evolutionarily conserved and young nonconserved miRNAs (Table 1). The sequences of the 35 newly identified small RNAs were also blasted against the *Pst* DC3000 genome, and no homologue with > 30% identity was found for any of them. This result means that at least 9.1% (19 of 207) of the known *Arabidopsis* miRNA precursors can produce this type of small RNA. Table 1 lists these miRNA precursors and the corresponding miRNA-like RNAs identified in our small RNA sequence libraries. As shown in the table, one precursor (that is, pre-miR822) can generate as many as ten distinct miRNA-like RNAs (with seven having more than five reads) from both sides of the stem-loop structure (Figure 3a). Additional file 2 displays the alignment of sequencing reads to these precursors, and Table 1 includes the numbers of sequencing reads of these miRNA-like RNAs. In the rest of this section, we

provide a slew of genomic and molecular evidence to show that many of these miRNA-like RNAs are authentic and functional miRNAs. Following miRNA nomenclature [17,18], we name these miRNA-like RNAs miR*n.k*, where integer *n* specifies a particular miRNA family and precursor (for example,159a for miR159a) and integer *k* denotes a specific miRNA or miRNA-like RNA. To minimize possible confusion, we reserve miR***n***.1 for the known miRNA, and name the newly identified miRNA-like RNAs as miR***n***.2, miR***n***.3, and so on, starting from the 5**'** -end of the miRNA precursor. Following the notation for miRNA*, the miRNA-like RNA opposite another miRNA-like RNA (miR***n.k***), but with a lower abundance than the latter, is labeled as miR***n.k*** *. However, if the abundances of miR***n.k*** and miR***n.k*** * are comparable, they are named as miR***n.k*** -5p and miR***n.k*** -3p, depending on their relative positions [17,18]. For example, the three miRNA-like RNAs on the miR159a precursor that passed our selection criteria are labeled as miR159a.2-5p, miR159a.3 and miR159a.2-3p, respectively, starting from the 5**'** -end of the precursor (Figure 1a).

The identified miRNA-like RNAs are generated by the miRNA biogenesis pathway

The newly identified miRNA-like RNAs and the known miRNAs share several common characteristics. First, an individual miRNA-like RNA often has a pairing partner on the opposite arm of the precursor fold-back structure, which is analogous to the pairing partnership of miRNA and miRNA*. More critically, such pairing partners typically have an approximately two-nucleotide 3**'**-end overhang, which reflects RNase III activities [39]. For example, miR159a.2-5p is paired with miR159a.2-3p with a two-nucleotide 3**'** -end overhang (Figure 1a). Similar examples can be found in the other miRNA precursor structures shown in Figures 1, 2, 3, and 4. Second and more importantly, these miRNA-like RNAs are generated by the same biogenesis pathway as the cognate miRNAs. We experimentally studied some of the miRNA-like RNAs on miR447a and miR822 precursors using various mutants of small RNA pathway components. As shown in Figure 2c, the accumulation of both miR447a (which was renamed as miR447a.1) and miR447a.3 depended on DCL1. The biogenesis of both

mature miR822 (that is, miR822.1) and miR822.2 depended on DCL4 (Figure 3c), which is consistent with previously published results [19]. Therefore, miR447a.3 and miR822.2 were generated by the same Dicer-like proteins as their cognate miRNAs. tasiRNAs are endogenous phased siRNAs generated by RDR6 and DCL4 [40]. miR447a.3 and miR822.2 did not require RDR (Figures 2c and 3c), which ruled out the possibility that these phased miRNA-like RNAs might be tasiRNAs. Furthermore, to determine whether these miRNA-like RNAs could be hc-siRNAs, we examined their accumulation in mutants of RDR2, DCL3 and the largest subunits of Pol IV (NRPD1) and Pol V (NRPE1), which are required for hc-siRNA formation and function [8,10,11,15]. As shown in Figures 2c and 3c, the production of miR447a.3 and miR822.2 did not need any RDR proteins, Pol IV, Pol V or DCL3. Therefore, these small RNAs were generated through the miRNA pathway by sequential DCL cleavages on the long hairpin stem regions; they are surely not siRNAs. Third, we examined the effect of HEN1 on these miRNA-like RNAs. In plants, small RNAs, including miRNAs, siRNAs and lsiRNAs, are methylated at their 3' -ends by HEN1 [21,26]. Methylation stabilizes the small RNAs and distinguishes them from RNA degradation products. The accumulation of miR447a.3 and miR822.2 was dependent on HEN1 (Figure 5), indicating that these small RNAs were methylated. Collectively, these results show that these miRNA-like RNAs are produced by the same miRNA pathway as their cognate known miRNAs.

The identified miRNA-like RNAs are differentially expressed

To investigate the potential functions of the newly identified miRNA-like RNAs in pathogen response, we examined the expression of some of them using Northern blotting. We found that many of the miRNA-like RNAs that we profiled, which have no homologue with identity > 30% in the bacterial genome, were differentially expressed under the challenge of different strains of *Pst*, and exhibited different expression patterns from their cognate miRNAs or miRNA*s (Figures 2d, 3d and 4c). As shown in Figure 2d, for instance, both miR447a.2-3p and miR447a.3 were strongly induced by the avirulent strain *Pst* (*avrRpt2*) and weakly induced by the non-pathogenic strain *Pst* DC3000 *hrcC*.

However, the virulent strain *Pst* DC3000 EV could induce only miR447a.3 but not miR447a.2-3p. Neither *Pst* DC3000 EV nor *Pst* DC3000 *hrcC* induced miR447a (that is, miR447a.1). In addition, miR447a.1 was expressed at a lower level than miR447a.2-3p and miR447a.3. Similarly, miR822.3 was induced by *Pst* DC3000 EV and *Pst* (*avrRpt2*) at 6 hpi, and by all three strains tested at 14 hpi, whereas miR822.2 was only induced by *Pst* (*avrRpt2*) at 14 hpi. miR822.3* was barely detected under these conditions (Figure 3d). miR839.2 and miR839.3 were only induced by *Pst* (*avrRpt2*) at 14 hpi and expressed at a very low level under other conditions, whereas miR839.1 was constitutively expressed at a similar level under these conditions (Figure 4c).

The identified miRNA-like RNAs may also be differentially expressed in different tissues. One such example can be seen by comparing the results for the miR839 precursor in Figure 4a with that in Figure 2b of [19]. The peak reads from the deep-sequencing data from [19] also exhibited a phasing pattern, which is in agreement with our deep-sequencing data (Figures 4a, b; Additional file 2). It is important to note that no sequencing read in our small-RNA libraries mapped to gap 2 in Figure 4a, whereas some sequencing reads at gap 2 were shown in Figure 2b in [19]. A major difference between the two deep-sequencing datasets is that total RNA was extracted from whole seedlings, flowers, rosette leaves, and siliques in [19], while we used only matured rosette leaves in our profiling.

As a final note on the expression levels, some of these miRNA-like RNAs can be more abundant than their cognate miRNAs (Table 1). For example, miR319b.2 has 491 reads while miR319b (that is, miR319b.1) has 30 reads (Table 1 and Figure 1d), which is a more than 10-fold difference. Similarly, both miR839.2 and miR839.3 have more reads than miR839 (that is, miR839.1) (Figure 4a). It is possible that some of the miRNA-like RNAs may be induced at certain developmental stages or under specific conditions to regulate gene expression.

The identified miRNA-like RNAs are potentially functional

We now present three pieces of evidence to show that many of the newly

identified miRNA-like RNAs have functional mRNA targets. First, most of these miRNAlike RNAs we identified can be associated with AGO proteins. In general, miRNAs are loaded onto AGO proteins to silence target genes by RNA cleavage, RNA degradation, or translation inhibition. Thus, we searched the *Arabidopsis* datasets of AGO-associated small RNAs [41,42] for the miRNA-like RNAs identified. We found that 25 (71.4%) of the 35 miRNA-like RNAs in 14 (73.7%) of the 19 precursors can be associated with 5 AGO proteins (AGO column in Table 1). In particular, 14, 15, 12, 8 and 11 miRNA-like RNAs can be associated with AGO1, AGO2, AGO4, AGO5 and AGO7, respectively. This result suggests that many of the identified miRNA-like RNAs can potentially function through AGO proteins. Further, the first nucleotide of a small RNA is critical for determining which AGO proteins it may associate with and may consequently dictate its mode of operation [41,42]. The first nucleotides of the unique sequences of the 35 miRNA-like RNAs are preferentially A (45.7% of the total) and U (42.9%), which account for nearly 90% of the total. As a comparison, 75.8% and 12.6% of the known *Arabidopsis* miRNAs start with U and A, respectively. Although the first nucleotides shifted from a preferential U in miRNAs to a nearly equal preference of U and A in the 35 miRNAlike RNAs, U and A are still the two dominant first nucleotides for miRNAs and the miRNA-like RNAs.

Second, many of the miRNA-like RNAs identified have putative mRNA targets that have coherent functions. We predicted their putative targets using the target prediction method in version 2 of the CleaveLand software for analyzing small RNA degradomes [43]. With an alignment score cutoff of 4.5, a total of 33 (94.3%) of the 35 miRNA-like RNAs identified have putative targets (Table S2 in Additional file 1). We reasoned that if these miRNA-like RNAs can silence their target genes, de-suppression of the targets might be expected in Dicer mutants, in which the miRNA-like RNAs would no longer be produced. Thus, we examined, using real-time RT-PCR, the expression of some of the predicted targets of miR169i.2-3p (At5g02710), miR169j.2 (At5g48300), miR447a.3 (At1g54710 and At1g06770), miR839.2 (At4g31210), and miR839.3 (At1g65960) in a *dcl1-9* mutant and in the wild type (Figure 6a), as well as a predicted

target of miR822.4-5p (At1g62030) in a *dcl4-2* mutant and in the wildtype (Figure 6b). Indeed, these targets were accumulated to a higher level in the mutants than in the wild type that we studied (Figures 6a,b). Further, because miR447a.3 and miR839.2 were induced by *Pst* (*avrRpt2*), we also examined the expression of their three target genes under the *Pst* (*avrRpt2*) treatment. As shown in Figure 6c, these targets were repressed during *Pst* (*avrRpt2*) challenge, showing a negative correlation with the expression of the corresponding miRNA-like RNAs. Furthermore, similar to most miRNAs, many miRNA-like RNAs identified can target multiple protein-coding genes (Table S2 in Additional file 1). In addition, some of the miRNA-like RNAs may have multiple targets with common or closely related functions. For example, miR775.2 targets two genes in the glycosyl hydrolase family. Different miRNA-like RNAs from the same miRNA precursor may have targets in the same gene family. One pronounced example is the miR822 precursor (Figure 4a). Three miRNA-like RNAs (miR822.3*, miR822.4-5p, and miR822.5), together with their cognate miR822 (miR822.1), can potentially target a total of 60 distinct DC1 domain containing proteins, some of which are targeted by multiple miRNA-like RNAs. Interestingly, miRNA-like RNAs from different miRNA families may also have targets in the same protein family. For example, miR159a.2-3p, miR169j.2, miR319a.2, miR447a.3, miR447b.3, miR822.4-5p, and miR839.2 all have targets in the leucine-rich repeat family. These relationships between the miRNA-like RNAs and their targets are reminiscent of miRNAs and their targets, and also allude to their possible origins of inverted gene duplication [30,44]. In short, our experimental and computational results indicate that the miRNA-like RNAs identified have the potential to silence their target genes, some of which have common or related functions.

Third, some miRNA-like RNAs can mediate target silencing by mRNA cleavage. Since the identified miRNA-like RNAs have the same characteristics as miRNAs and many can be associated with AGO proteins, we hypothesized that they might also directly cleave their mRNA targets. To test this hypothesis, we searched for, using version 2 of the CleaveLand degradome software [43], the small RNA target signatures of mRNA cleavage products in the data from *Arabidopsis* PARE or small RNA degradomes

collected by three labs from different tissues and under various conditions [36-38]. With an alignment-score cutoff of 4.5 and a *P* -value threshold of 0.2, we found small RNA cleavage products of seven mRNA genes targeted by six miRNA-like RNAs that we identified (miR159a.2-3p, miR169b.2, miR169i.2-3p, miR169j.2, miR822.4-5p, miR839.3; the PARE column in Table 1). Detailed information on these miRNA-like RNAs and their targets supported by the degradome data is in Table S3 in Additional file 1; the alignments of four of these pairs of miRNA-like RNAs and targets, along with another three pairs tested, are shown in Figure 6d. Furthermore, four of these six miRNA-like RNAs (miR159a.2-3p, miR169j.2, miR822.4-5p, and miR839.3) can also be associated with AGO proteins (Table 1), indicating that, mechanistically, these small RNAs can function through the canonical miRNA pathway. Indeed, the ablation of three of the six miRNA-like RNAs (miR169i.2-3p, miR169j.2, and miR839.3) in the *dcl1-9* mutant as well as miR822.4-5p in the dcl4-2 mutant led to elevated expression of some of their targets (Figures 6a,b). The relatively small number of the miRNA-like RNAs that have mRNA cleavage products may be due to two reasons. First, the miRNA-like RNAs were typically expressed at low abundance; thus, their cleavage products were too low to be detected. Second, different tissues were used in our experiments (mature leaves) and for PARE data collection (floral tissues, including the inflorescence meristem and early stage floral buds, and *EIN5* mutant). This tissue difference may also explain that no target cleavage product was detected even for four known miRNAs listed in Table 1 (miR447a.1/b.1, miR822.1 and miR839.1) while the expression of miRNAs and miRNA-like RNAs is often tissue-specific. Nevertheless, this degradome analysis provided evidence that some of the miRNA-like RNAs identified in our experiments can function through mRNA target cleavage.

Distribution of the miRNA-like RNAs on precursor foldback structures

A remarkable characteristic of the miRNA-like RNAs that we found in *Arabidopsis* is that they can appear on either side of a known miRNA-miRNA* duplex on a precursor hairpin and can be close to either the base or the loop of the hairpin. Two

or more miRNA-like RNAs can also reside on both sides of a miRNA-miRNA* duplex. A summary of the location distribution of the miRNA-like RNAs is given in the 'Position' column of Table 1, where a plus sign (+) means that miRNA-like RNAs appear exclusively between miRNA-miRNA* and the loop of the hairpin, a minus sign (-) indicates that miRNA-like RNAs occur exclusively between miRNAmiRNA* and the base of the hairpin, and ' +/-' means that there are miRNA-like RNAs on both sides of the miRNA-miRNA* duplex. As shown, among the 19 miRNA precursors identified, 7 harbored miRNA-like RNAs exclusively toward the loops of the hairpins. Examples include *MIR159a* , *MIR319a* and *MIR319b* in Figures 1a,c,d, respectively. This is consistent with the recent discovery that the DCL cleavage that produces mature miR159 and miR319 starts from the loop ends of their fold-back structures [45,46]. Another ten precursors produced miRNA-like RNAs near the hairpin bases - for example, *MIR169m* and *MIR447a* in Figures 1b and 2a. This miRNA-like RNA distribution is well supported by the conventional model of miRNA biogenesis [3,14], in which two subsequent DCL cleavage activities produce a miRNA-miRNA* duplex, first releasing the precursor miRNA and then liberating the duplex. The remaining two precursors - that is, *MIR822* and *MIR839* in Figures 3a and 4a - had miRNA-like RNAs on both sides of their cognate miRNA-miRNA* duplexes. These two precursors contain long stems, which are likely to be processed by continuous in-phase dicing activity of DCL1 or DCL4. This process is similar to the biogenesis of tasiRNAs by DCL4.

Another interesting observation is that not every slot in the phasing pattern of a precursor was filled by sequencing reads, as shown in *MIR319b* and *MIR839* (Figures 1d and 4a). Such a peculiar pattern of miRNA-like RNA expression suggests potential condition- or tissue-specific expression, and variable metabolism rates of these miRNA-like RNAs.

Conservation of sequential DCL cleavage of long MIR hairpins in eukaryotes

To determine whether the miRNA-like RNAs that we identified in *Arabidopsis* also exist in other plant species, we studied four additional plant organisms. Specifically,

we searched for miRNA precursors that are capable of generating multiple miRNA-like RNAs in publicly available small-RNA deep-sequencing datasets from *O. sativa* (rice), *P. patens* (moss), *M. truncatula* and *P. trichocarpa*; see Materials and methods for information on the sources of these datasets. In total, we identified 75, 37, 9 and 11 miRNA precursors, respectively, from these four plant species that harbor miRNA-like RNAs in addition to miRNAs. Table 2 lists these miRNA precursors and Additional files 3, 4, 5 and 6 contain the alignments of the miRNA-like RNAs to their corresponding miRNA precursors. Note that the sequencing data on rice and moss were from Solexa sequencing, while the data on *Medicago* and *Populus* were from a mixture of Solexa and 454 sequencing. Therefore, the sequencing depths for *Medicago* and *Populus* were not as deep as those for rice and moss, resulting in fewer miRNA-like RNA-bearing miRNA precursors identified in the latter two species. As shown in Table 2, several individual miRNA-like RNAs identified are conserved in multiple plants. In the five plant species we studied, miRNA-like RNAs appeared in two well-conserved miRNA families, that is, miR159 and miR319 (Table 2). miRNA-like RNAs appeared in miR319 precursors in all of these five plants, and miRNA-like RNAs occurred in miR159 precursors in all of five bar moss. miR159 and miR319 belong to the same *MIR* family based on their evolutionary origin [30,44], playing important roles in plant development [47]. Importantly, a close inspection showed that many individual miRNA-like RNAs on the miR159 and miR319 precursors are also highly conserved at the sequence level. Figure 6a displays the miR159a precursors in *Arabidopsis*, rice, *Medicago* and *Populus* and Figure 6b shows the miR319b precursors in *Arabidopsis*, rice, moss and *Medicago*. As shown, several miRNA-like RNAs, particularly those expressed at a relatively abundant level, are more conserved than their flanking sequences except their cognate miRNAs or miRNA*s. For example, the sequences of miR319b.2 and miR319b.2* have a comparable level of conservation to miR319b* (that is, miR319b.1*); specifically, miR319a, miR319a*, miR319b.2, and miR319b.2* have 19, 12, 12 and 11 identical nucleotides across the four species, respectively. Furthermore, most of these miRNA-like RNAs, including miR319b.2, miR159a.2-5p, and miR159a.2-3p, can be incorporated into

multiple AGO proteins in *Arabidopsis* (Table 1). We also examined the target conservation of the miRNA-like RNAs on miR159a and miR319b precursors across *Arabidopsis*, a dicotyledonous plant, and rice, a monocotyledonous plant. Both ath-miR159a.2-3p and osa-miR159a.2-3p have targets in the pentatricopeptide repeat (PPR) protein family (*At2G40720, Os2G05720, Os10G40920* and, *Os5G50690*). Collectively, all these results suggest that these highly conserved miRNA-like RNAs in plants are functional. To further understand such miRNA-like RNAs, we searched for miRNA-like RNAs in a large collection of small-RNA deep sequencing data from four animal species, *H. sapiens* (human), *M. musculus* (mouse), *C. elegans* and *D. melanogaster* (see Materials and methods). We identified 14, 27, 17 and 9 miRNA precursors in human, mouse, *C. elegans* and Drosophila, respectively, that can produce miRNA-like RNAs in addition to the known miRNAs and miRNA*s (Table 2). These miRNAlike RNA-bearing miRNA precursors are distributed in intergenic regions and introns, except has-miR34a and cel-miR354, which reside in 3' untranslated regions. These miRNA-like RNAs are immediately adjacent to the annotated miRNA-miRNA* duplexes, as shown by the alignments to their originating precursor hairpins in the four animals in Additional files 7, 8, 9, and 10, respectively. Different from the results for plants, no more than one small RNA pairing duplex was present beyond the known miRNA-miRNA* duplex in each animal miRNA precursor identified. This difference between plants and animals is mainly due to the relatively short miRNA precursors in animals. The newly identified miRNA-like RNAs in animals are also near the precursor bases. Notice that only miRNA-like RNAs on miR7 precursors in human and mouse are conserved (Table 2; Figure 7c; Additional files 7 and 8). It is also interesting to note that miR7 in Drosophila has diverged from that in human and mouse significantly (data not shown), and miR7 has not been reported in *C. elegans*. These results indicate a low conservation of miRNA-like RNAs based on the animal small-RNA sequencing data we examined.

DISCUSSION

We reported 19 miRNA precursors in *Arabidopsis* that are able to produce

multiple distinct miRNA-like RNAs with potential function. Our analysis of the 13 libraries of deep-sequencing data from *Arabidopsis* characterized several important features of these miRNA-like RNAs. First, two miRNA-like RNAs on opposite arms of a miRNA precursor hairpin usually form a duplex with a two-nucleotide 3'-end overhang, which is a key property of miRNA-miRNA* duplexes and reflects the activities of RNase III proteins [39]. Second, such miRNA-like RNAs are arranged in phase, again reflecting sequential dicing activities of some RNase III proteins. Third, the first nucleotide of the discovered miRNA-like RNAs has a strong preference (approximately 90%) for A and U, which is similar to the approximately 88% for all known miRNAs. Moreover, we obtained several lines of molecular evidence to support the notion that some of the miRNAlike RNAs identified are authentic and potentially functional miRNAs. First, these miRNA-like RNAs are generated through the miRNA pathway, but not the pathways for tasiRNAs or hc-siRNAs (Figures 2 and 3). Second, these miRNA-like RNAs are likely methylated (Figure 5). Third, 25 (71.4%) of the 35 miRNA-like RNAs identified were found in the pools of AGO-associated small RNAs (Table 1) [41,42], suggesting that they may potentially function through the AGO effectors. Fourth, several of these miRNA-like RNAs can induce target mRNA cleavage, and the cleavage products were present in the data from a PARE database or small-RNA degradome (Figure 6) [36-38]. Fifth, most of the miRNA-like RNAs identified have predicted targets, and many of them potentially target genes within the same gene family or with common functions (Table S2 in Additional file 1), which is similar to miRNAs. Sixth, a few of the miRNA-like RNAs are more abundant than cognate miRNAs (Figures 1 and 4), and some of the miRNA-like RNAs are differentially expressed under pathogen infections (Figures 2, 3 and 4), suggesting their potential regulatory functions in response to environmental stresses. The results of gene expression analysis using *dcl1* and *dcl4* mutants and wild type *Arabidopsis* plants showed that the loss of miR169i.2-3p, miR169j.2, miR447a.3, miR839.2, miR839.3, and miR822.4-5p can elevate the expression of the predicted target mRNAs that we tested (Figure 6). In addition, miR447a.3 and miR839.2 were induced by *Pst* (*avrRpt2*), and we also detected down-regulation of their targets after *Pst* (*avrRpt2*)

challenge (Figure 2.6).

Moreover, such miRNA-like RNAs exist in five evolutionarily distant plant species as well as four animal organisms we examined - *Arabidopsis*, rice, moss, *Medicago* and *Populus* as well as human, mouse, *C. elegans* and *Drosophila*. This result suggests that the mechanism underlying these miRNA-like RNAs must be conserved within and across plants and animals. The fact that these miRNA-like RNAs appear in moss, an ancient land plant, alludes to their possible evolutionary origin in plants. Some of these miRNA-like RNAs from conserved miRNA families, that is, miR159 and miR319 in plants, are conserved at the sequence level (Figure 2.7), which adds another layer of evidence that these miRNA-like RNAs are potentially functionally important in plants. In addition, the miRNA-like RNAs we identified appear to occur in evolutionarily 'old' miRNA precursors in both plants and animals based on the publicly available small-RNA sequencing data we examined.

We did not observe a clear correlation between deep sequencing read counts and Northern blot results in *Arabidopsis*, even for a couple of highly expressed known miRNAs. Similar observations were also made in some recent studies [48,49]. The inconsistency between sequencing read counts and Northern blot results is likely due to the differences between the two techniques. Sample preparation for deep sequencing involves two steps of RNA adaptor ligation and one step of PCR amplification, which introduce bias to the final data. Small RNAs with different 5' - and 3' -end structures or modification may have different efficiencies for ligation, and PCR tends to amplify highly abundant sequences more efficiently than less abundant ones. On the other hand, hybridization-based Northern blotting can hybridize any small RNAs with the same sequence regardless of their end modification or structure, although it may have a cross-hybridization side effect depending on the stringency of the hybridization conditions. In light of these observations, we thus relied on conventional Northern blot analysis with high stringency conditions to quantify the expression of the miRNA-like RNAs we studied. Some early studies, including those using deep sequencing, have also observed small RNAs beyond miRNAs or miRNA*s on some miRNA precursors in plants [19,28-

30] and animals [31,32]. For example, sequencing reads of miRNA-like RNAs were found on the ath-miR839 precursor [19]. Interestingly, such miRNA-like RNAs have also been found in a mirtron in mouse [50]. However, none of these small RNAs in plants and animals have been analyzed with regard to their biogenesis or potential functions. Chiang et al. [50] argued that their single discovery of a mirtron precursor in mouse was not a coincidence of spliceosome activity releasing the particular mirtron precursor. Indeed, our extensive data on *MIR* genes in four animal and five plant species convincingly show that continuous sequential cleavage activities are a common action for Dicer proteins in animal species or Dicer-like proteins in plant organisms. Another line of related work covers the moRNAs in a chordate, human and a herpesvirus [33-35]. Some of the miRNA-like RNAs that are immediately adjacent to miRNA or miRNA* and located near the base of the miRNA precursors may be classified as moRNAs. However, the moRNAs reported for the chordate *Ciona intestinalis* may be fundamentally different from the miRNA-like RNAs in the four animal genomes that we studied. In particular, the miRNA-like RNAs we identified are within the miRNA precursor sequences within the double-stranded RNA regions, and the ones we tested in *Arabidopsis* are dependent on DCL1 or DCL4. In contrast, many moRNAs identified in C. intestinalis- for example, those adjacent to ci-miR124-1 and ci-miR124-2 shown in Figure 2b of [33,34] - can span substantially beyond the double-stranded stem region and even outside the miRNA precursors, which cannot be recognized and processed by Dicer. Thus, although moRNAs are likely to be cut by Drosha at one end, how the other end of moRNAs is generated has yet to be determined [33-35]. More work is required to better understand moRNAs in animals, particularly their biogenesis and functions. In comparison, our study showed that the miRNA-like RNAs in *Arabidopsis* are produced by sequential DCL cleavage on the long hairpin stem regions and can be generated on both sides of the miRNA-miRNA* duplex. Many of them can be incorporated into AGO proteins, and mediate target mRNA cleavage. Some of them are differentially expressed in response to environmental stresses. All of these results suggested their potential biological functions.

CONCLUSIONS

To sum up the results from our in-depth genomic and molecular analysis on *Arabidopsis*, our results from an extensive survey of an additional four plant and four animal species, and some data from early studies on plants [19, 28-30] and animals [31, 32], it is evident that continuously sequential cleavage by Dicer or Dicer-like proteins is a common theme, rather than an exception, in plants and animals, which gives rise to phased miRNA-like RNAs from *MIR* genes. In *Arabidopsis*, DCL4 has been shown to be capable of such sequential cleavage to generate phased small RNAs such as tasiRNAs [8]. Our results suggest that DCL1, which generates most miRNAs and miRNAlike RNAs, may also act through sequential cleavage as DCL4 does. Further, a limited conservation of miRNA-like RNAs in evolutionarily 'old' *MIR* genes indicates that they are subject to evolutionary selection. The resulting small RNAs may be subject to different rates of degradation. Which miRNAs or miRNA-like RNAs can accumulate in response to developmental and/or environmental cues may be determined by their conservation and through tight regulation by post-biogenesis protection or degradation. For example, these miRNAs or miRNA-like RNAs can be stabilized by binding with AGO proteins or other RNA-binding proteins, or can be degraded by some exoribonucleases [51]. In summary, our results from extensive molecular experiments on *Arabidopsis* and a systematic examination of a large quantity of deep-sequencing datasets from five evolutionarily diverse plant species and four evolutionarily distant animal organisms suggest that the multiple distinct miRNA-like RNAs that we identified broadly exist in eukaryotic species and can be authentic and functional miRNAs. Our results further suggest that the pool of miRNAs is larger than was previously recognized, and miRNA-mediated gene regulation may be broader and more complex than previously thought.

MATERIALS AND METHODS

Plant materials, small-RNA library construction, and deep sequencing

Thirteen small-RNA libraries were prepared from 4- to 5-week-old short-day

grown *Arabidopsis* Col-0 inoculated with 10 mM $MgCl_2$ (mock), type III secretion system mutated strain *Pst* DC3000 *hrcC*, virulent strain *Pst* DC3000, and avirulent strain *Pst* DC3000 (*avrRpt2*). Bacteria infiltration was carried out on the *Arabidopsis* wild-type and mutant plants as described previously [24]. Infiltrated leaves were harvested at 6 and 14 hpi. A large number of *Arabidopsis* mutants (*dcl1-7, dcl1-9, dcl2-1, dcl3-1, dcl4-2, hen1, rdr2-1, rdr1-1, rdr6-15, nrpd1-3, nrpe1-11*) and their corresponding wild-type ecotypes were used in this study. Assays were performed on 4-week-old plants grown at 22°C with 12 h light.

Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA, USA) from infiltrated leaves and fractionated on 15% denaturing polyacrylamide gel. RNA molecules ranging from 18 to 28 nucleotidess were excised and ligated to 5**'**- and 3**'**-RNA adaptors using T4 RNA ligase followed by RT-PCR and gel purification as described in the instructions from Illumina [52,53]. The small RNA libraries were sequenced by Illumina Inc. and the University of California, Riverside (UCR) core facility.

Processing of deep sequencing data

Raw sequence reads were parsed to remove the 3' adaptors. The sequencing reads from each of the small RNA libraries, with adaptors trimmed, were mapped to the *Arabidopsis* nuclear, chloroplast and mitochondrial genome sequences and cDNA sequences, which were all retrieved from the TAIR database [54]. The reads that matched these sequences with no mismatches (the row labeled '*mapped*' in Table S1 in Additional file 1) were retained for further analysis. Sequencing reads were aligned to the precursors of the annotated *Arabidopsis* miRNAs in miRBase [55] with Novealign [56]. Those sequencing reads that could be mapped to a miRNA precursor with zero mismatches were retained for further analysis.

Small RNA deep-sequencing data on four additional plant and four animal species

We collected small-RNA deep-sequencing data, generated by Illumina/Solexa or 454

sequencing platforms, for four additional plants from the GEO database. In particular, we analyzed a total of 18 small-RNA sequencing datasets on *P. patens* (moss) [57]; six datasets by Solexa sequencing), *O. sativa* (rice [25]; and another two datasets, GEO accession number [GEO: GSE14462]; all by Solexa sequencing), *M. truncatula* (two datasets from Solexa sequencing [58] and three datasets from 454 sequencing [59], and *P. trichocarpa* (four datasets from 454 sequencing [60, 61]). Small RNA deep sequencing data were also collected for four animal species: *H. sapiens* (human [62, 63]; four datasets); *M. musculus* (mouse [32]; three datasets); *C. elegans* (C. elegans [64]; nine datasets) and *D. melanogaster* (Drosophila [16]; five datasets). The initial processing of these sequencing libraries followed the same steps as for the 13 *Arabidopsis* datasets. In summary, we collected, processed and analyzed a total of 52 small RNA deep-sequencing datasets in the current study.

Identifying miRNA-like RNAs and determining their phasing patterns

We first mapped sequence reads, allowing no mismatch, to miRNA precursors. We retained those that have reads arranged in block patterns for further analysis. To this end, we applied the Blockbluster software [33, 34] to first identify blocks of sequencing reads on the annotated miRNA precursors. The most abundant sequence read within each detected block was taken as the representative sequence for the block. The total number of sequence reads for the block is the sum of the copy number of the representative read and the copy numbers of other sequence reads that fall within the representative sequence and that overlap with the representative read with no more than three nucleotides beyond the representative sequence on either end. We allowed such overhangs to tolerate imprecise dicing activities of RNase III enzymes. The remaining blocks were further inspected; those spanning across two neighboring blocks were ignored. Those miRNA precursors that have blocks of reads arranged in phase were kept for further analysis. When reporting the results using our *Arabidopsis* deep sequencing data, we ignored such blocks that have less than five total sequencing reads to reduce potential false positive miRNA-like RNAs due to sequencing error.

Northern blot analysis of small RNAs

Total RNA was isolated using Trizol reagent (Invitrogen) from infiltrated leaves and 80 **μ**g was resolved on 15% denaturing polyacrylamide gel. RNA was crosslinked to membrane using EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (Sigma, St. Louis, MO, USA))) [65]. We used the locked nucleic acid (LNA) probe for small RNA Northern blot analysis to detect miR839, and DNA oligos for detecting miR447 and miR822. Blots were exposed to phosphorscreens, and scanned using PhosphorImager (Molecular Dynamics (Amersham Pharmacia Biotech Inc, Piscataway, NJ, USA)).

The LNA probes used for detection were: mirt839-AT1G67481.2, c+tac+tct+ttc+tgc+tca+cat+ga; mirt839-AT1G67481.1, g+gga+acg+atg+aaa+ggt+tgg+ta; and mirt839-AT1G67481.3, t+cag+cac+tat+cac+ggt+ttt+gca+aa.

We used the following oligo probes for miR447 and miR822: miR447a.1-anti, CAACAAAACATCTCGTCCCCAA; miR447a.2-3p-anti, AAGCCCCTTCTTATATC-GAGTC; miR447a.3-anti, AATACTACAATTTCTTCCATA; miR822.1-anti, CATGT-GCAAATGCTTCCCGCA; miR822.1*-anti, TGTAGAAAGCATTTGCACATG; miR822.3*-anti, AGAAAACAACATGCGTTACATCC; miR822.3-anti, GGGATGCA-ACGTATATTGTTTC; miR822.2-anti, ACGACCTTAAGTGTAAGTAGAT.

miRNA target cleavage product analysis and target prediction

We applied version 2 of the CleaveLand software [43] to the data from PARE analysis or small RNA degradomes [36-38] to characterize the signatures of mRNA cleavage products of the miRNA-like RNAs we studied. We only considered the target cleavage products with an alignment score from CleaveLand of no less than the cutoff threshold of 4.5. Putative targets of miRNA-like RNAs were also predicted by the miRNA target finding method in the CleaveLand software package. This target finding method used a scoring scheme that charges a penalty of one to a mismatch and a penalty of 0.5 to a wobble base pairing, and doubles these penalties in the seed region, which is

the two- to seven-nucleotide region near the 5'end of a miRNA.

Quantitative RT-PCR analysis of small RNA targets

For quantitative RT-PCR analysis, 5 μg of total RNA was used for synthesizing cDNA. DNA contamination was removed by using DNase I (Invitrogen). Amplification of miRNA 169f and miR447f targets (see AT numbers below) was carried out using a realtime PCR machine (iQ5, Bio-Rad, Hercules, CA, USA). The following oligos were used to perform the RT-PCR analysis: At5g02710 (miR169i.2-3p)-QRTF, TCGACAACCCTGATGATGTTGAGCTG; At5g02710 (miR169i.2-3p)-QRTR, CCTT-GTTGAATTTCTTTTCTTCAATCC; At5g48300(miR169j.2)-QRTF, AATGGGAGC-TGATTATTACGAGACTGC; At5g48300(miR169j.2)-QRTR, CTTGCACGTTGTCG-CTGTTTATGATC; At1g54710(miR447a.3)-QrtF, GGTATATGATAATTTTCACAGT-GTGTATACCAG; At1g54710(miR447a.3)-QrtR, ATCTTCGGGGTAAACCATACCA-TTGAAT; At1g06770(miR447a.3)-QrtF, CGAGAGTGAAAATGAGATAGAGAT; At1g06770 (miR447a.3)-QrtR, AGACAAGAACCATCGTAAAATCCT; At4g31210 (miR839.2)-QrtF, GTTATTCTAAAGTGTGGACCCTATGGGC; At4g31210 (miR839.2)-QrtR, CATCTTCAGGGTGGGTTCCCAGTG; At1g65960(miR839.3)-QRTF, TCATTCACTCTCAATTTCTCCAAGGGA; At1g65960(miR839.3)-QRTR, TTCTCTATACCTTCTTTGAGAACCAC; At1g62030 (miR822.4-5p)-QRTF, GCCTG-GTTCATCTTGGATAAGCTTCTCC; At1g62030(miR822.4-5p)-QRTR, TTGTAGAT-ATTCGCGATGACCAGC.

ABBREVIATIONS

AGO: Argonaute; DCL: Dicer-like protein; EV: empty vector; GEO: Gene Expression Omnibus; hc-siRNA: heterochromatic small interfering RNA; hpi: hours post-inoculation; lsiRNA: long small interfering RNA; MIR: miRNA gene; miRNA: microRNA; moRNA: miRNA-offset RNA; PARE: parallel analysis of RNA ends; Pol: RNA polymerase; RDR: RNA-dependent RNA polymerase; siRNA: small interfering RNA; tasiRNA: trans-acting small interfering RNA.

ADDITIONAL MATERIAL

Additional file 1: Tables S1, S2, and S3. Table S1: number of sequencing reads from 13 small RNA libraries of *Arabidopsis* seedlings at 6 and 14 hours after three pathogen infections, along with their corresponding mock infection as control. Shown in the table are the total number of reads (total), the number of reads that can map to the *Arabidopsis* genome and cDNA sequences (mapped), nuclear genome (nuclear), transposons and repeats (repeats/mobile), and cDNA sequences (cDNAs). No mismatch was allowed for the mapping in the table. The second number for each condition (column) is the percent of reads to the mapped reads. The sequencing reads are available in the GEO database under accession number [GEO: GSE19694]. Table S2: putative targets of the identified miRNA-like RNAs identified by the miRNA target finding method, developed by Michael Axtell, in the CleaveLand software for small RNA degradome analysis. Results are those with alignment scores above the cutoff threshold of 4.5. Table S3: six miRNA-like RNAs found in the current study and their mRNA targets identified in small RNA degradome data. In the table, the first two columns list the miRNAlike RNAs and their targets, the third to fifth columns list the three major quantitative measures of the results, that is, the alignment scores, the number of raw reads of target degradation products and P-values quantifying the enrichment of degradation products, respectively. The last column indicates if a pair of miRNA-like RNA and target was tested in the current study.

Additional file 2: Supplemental File S1. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in A. thaliana (ath). The sequencing data were collected in the current study and are available in GEO under accession number [GEO: GSE19694].

Additional file 3: Supplemental File S2. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in rice (osa). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 4: Supplemental File S3. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in

moss (ppt). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 5: Supplemental File S4. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in *Medicago* (mtr). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 6: Supplemental File S5. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in *Populus* (ptc). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 7: Supplemental File S6. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in H. sapiens (has). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 8: Supplemental File S7. This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in M. musculus (mms). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 9: Supplemental File S8.

This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in C. elegans (cel). The sequencing data were obtained from GEO; see Materials and methods for details.

Additional file 10: Supplemental File S9.

This is a file for sequencing reads mapped and aligned to miRNA precursors that can produce miRNA-sibling small RNAs (msRNAs) in D. melanogaster (dme). The sequencing data were obtained from GEO; see Materials and methods for details.

REFERENCES

1. Baulcombe D: RNA silencing in plants. Nature 2004, 431:356-363.
2. Mallory AC, Bouche N: MicroRNA-directed regulation: to cleave or not to cleave. Trends Plant Sci 2008, 13:359-367.
3. Kim VN: MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 2005, 6:376-385.
4. Ambros V: The functions of animal microRNAs. Nature 2004, 431:350-355.
5. Flynt AS, Lai EC: Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet 2008, 9:831-842.
6. Carthew RW, Sontheimer EJ: Origins and mechanisms of miRNAs and siRNAs. Cell 2009, 136:642-655.
7. Khraiwesh B, Arif MA, Seumel GI, Ossowski S, Weigel D, Reski R, Frank W: Transcriptional control of gene expression by microRNAs. Cell 2010, 140:111-122.
8. Chapman EJ, Carrington JC: Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 2007, 8:884-896.
9. Jin HL: Endogenous small RNAs and antibacterial immunity in plants. Febs Lett 2008, 582:2679-2684.
10. Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ: RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol 2009, 21:367-376.
11. Vazquez F: *Arabidopsis* endogenous small RNAs: highways and byways. Trends Plant Sci 2006, 11:460-468.
12. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: MicroRNA genes are transcribed by RNA polymerase II. EMBO J 2004, 23:4051-4060.
13. Zhou X, Ruan J, Wang G, Zhang W: Characterization and identification of microRNA core promoters in four model species. PLoS Comput Biol 2007, 3:e37.
14. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004, 116:281-297.
15. Vaucheret H: Plant ARGONAUTES. Trends Plant Sci 2008, 13:350-358.
16. Okamura K, Liu N, Eric C, Lai EC: Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. Mol Cell 2009, 36:431-444.
17. 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: Criteria for annotation of plant MicroRNAs. Plant Cell 2008, 20:3186-3190.
18. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T: A uniform system for microRNA annotation. RNA 2003, 9:277-279.
19. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis* thaliana. Genes Dev 2006, 20:3407-3425.
20. Li J, Yang Z, Yu B, Liu J, Chen X: Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. Curr Biol 2005, 15:1501-1507.
21. Yu B, Yang ZY, Li JJ, Minakhina S, Yang MC, Padgett RW, Steward R, Chen XM: Methylation as a crucial step in plant microRNA biogenesis. Science 2005, 307:932-935.

22. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 2005, 123:1279-1291.
23. Jin H, Vacic V, Girke T, Lonardi S, Zhu JK: Small RNAs and the regulation of cis-natural antisense transcripts in *Arabidopsis*. BMC Mol Biol 2008, 9:6.
24. Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, Zhu JK, Staskawicz BJ, Jin HL: A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci USA 2006, 103:18002-18007.
25. Zhou X, Sunkar R, Jin H, Zhu JK, Zhang W: Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in Oryza sativa. Genome Res 2009, 19:70-78.
26. Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H: A novel class of bacteriainduced small RNAs in *Arabidopsis*. Genes Dev 2007, 21:3123-3134.
27. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009, 11:228-234.
28. Lacombe S, Nagasaki H, Santi C, Duval D, Piégu B, Bangratz M, Breitler JC, Guiderdoni E, Brugidou C, Hirsch J, Cao X, Brice C, Panaud O, Karlowski WM, Sato Y, Echeverria M: Identification of precursor transcripts for 6 novel miRNAs expands the diversity on the genomic organisation and expression of miRNA genes in rice. BMC Plant Biol 2008, 8:123.
29. Kurihara Y, Watanabe Y: *Arabidopsis* micro-RNA biogenesis through Dicerlike 1 protein functions. Proc Natl Acad Sci USA 2004, 101:12753-12758.
30. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2007, 2:e219.
31. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC: Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res 2007, 17:1850-1864.
32. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs. Genes Dev 2008, 22:2773-2785.
33. Shi W, Hendrix D, Levine M, Haley B: A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol 2009, 16:183-189.
34. Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF: Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 2009, 25:2298-2301.
35. Umbach JL, Cullen BR: In-depth analysis of Kaposi's sarcoma-associated herpesvirus microRNA expression provides insights into the mammalian microRNA-processing machinery. J Virol 2010, 84:695-703.
36. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ: Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. Curr Biol 2008, 18:758-762.

37. Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR: A link between RNA metabolism and silencing affecting *Arabidopsis* development. Dev Cell 2008, 14:854-866.
38. German MA, Pillay M, Jeong DH, Hetawal A, Luo SJ, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ: Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 2008, 26:941-946.
39. MacRae IJ, Doudna JA: Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol 2007, 17:138-145.
40. Xie Z, Allen E, Wilken A, Carrington JC: DICER-LIKE 4 functions in transacting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis* thaliana. Proc Natl Acad Sci USA 2005, 102:12984-12989.
41. Mi SJ, Cai T, Hu YG, Chen Y, Hodges E, Ni FR, Wu L, Li S, Zhou H, Long CZ, Chen S, Hannon GJ, Qi YJ: Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5 ' terminal nucleotide. Cell 2008, 133:116-127.
42. Montgomery TA, Howell MD, Cuperus JT, Li DW, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC: Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 transacting siRNA formation. Cell 2008, 133:128-141.
43. Addo-Quaye C, Miller W: CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 2009, 25:130-131.
44. Allen E, Xie ZX, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis* thaliana. Nat Genet 2004, 36:1282-1290.
45. Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, Axtell MJ: Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. RNA 2009, 15:2112-2121.
46. Bologna NG, Mateos JL, Bresso EG, Palatnik JF: A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. EMBO J 2009, 28:3646-3656.
47. Chen X: A silencing safeguard: links between RNA silencing and mRNA processing in *Arabidopsis*. Dev Cell 2008, 14:811-812.
48. Fahlgren N, Sullivan CM, Kasschau KD, Chapman EJ, Cumbie JS, Montgomery TA, Gilbert SD, Dasenko M, Backman TW, Givan SA, Carrington JC: Computational and analytical framework for small RNA profiling by high-throughput sequencing. RNA 2009, 15:992-1002.
49. Willenbrock H, Salomon J, Søkilde R, Barken KB, Hansen TN, Nielsen FC, Møller S, Litman T: Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. RNA 2009, 15:2028-2034.
50. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP: Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev 2010, 24:992-1009.
51. Ramachandran V, Chen X: Degradation of microRNAs by a family of

exoribonucleases in *Arabidopsis*. Science 2008, 321:1490-1492.
52. Katiyar-Agarwal S, Jin H: Discovery of pathogen-regulated small RNAs in plants. Methods Enzymol 2007, 427:215-227.
53. Chellappan P, Jin H: Discovery of plant microRNAs and short-interfering RNAs by deep parallel sequencing. Methods Mol Biol 2009, 495:121-132.
54. TAIR9. [http://www.arabidopsis.org].
55. miRBase release 13.0. [http://microrna.sanger.ac.uk].
56. Novoalign version 2.04. [http://www.novocraft.com].
57. Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z, Frank W, Axtell MJ: Physcomitrella patens DCL3 is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. PLoS Genet 2008, 4:e1000314.
58. Szittya G, Moxon S, Santos DM, Jing R, Fevereiro MP, Moulton V, Dalmay T: High-throughput sequencing of *Medicago* truncatula short RNAs identifies eight new miRNA families. BMC Genomics 2008, 9:593.
59. Lelandais-Brière C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J, Crespi M: Genome-wide *Medicago* truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. Plant Cell 2009, 21:2780-2796.
60. Klevebring D, Street NR, Fahlgren N, Kasschau KD, Carrington JC, Lundeberg J, Jansson S: Genome-wide profiling of populus small RNAs. BMC Genomics 2009, 10:620.
61. Barakat A, Wall PK, Diloreto S, Depamphilis CW, Carlson JE: Conservation and divergence of microRNAs in *Populus*. BMC Genomics 2007, 8:481.
62. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008, 18:610-621.
63. Zhang H, Yang JH, Zheng YS, Zhang P, Chen X, Wu J, Xu L, Luo XQ, Ke ZY, Zhou H, Qu LH, Chen YQ: Genome-wide analysis of small RNA and novel microRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach. PLoS One 2009, 4:e6849.
64. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP: Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell 2006, 127:1193-1207.
65. Pall GS, Codony-Servat C, Byrne J, Ritchie L, Hamilton A: Carbodiimidemediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. Nucleic Acids Res 2007, 35:e60.

Figure 2.1 Four miRNA precursors that can generate multiple miRNA-like RNAs. (a-d). Three miRNA precursors with miRNA-like RNAs in the upper arms close to the loops of their hairpins (a, c, d), and a miRNA precursor with miRNA-like RNAs in the lower arm of its hairpin (b). Note that miRNA-miRNA* duplexes for miRNA-like RNAs, with approximately two-nucleotide 3'-end overhangs, appear on the miR159, miR169m and miR319b precursors. The previously annotated miRNAs were named as miR**n**.1 (see main text for detail) and those miRNA-like RNAs having less than four reads were not named. For clarity, miR169m.2* and miR319b.2* are also indicated though the numbers of reads mapped to them were below the cutoff threshold of 5.

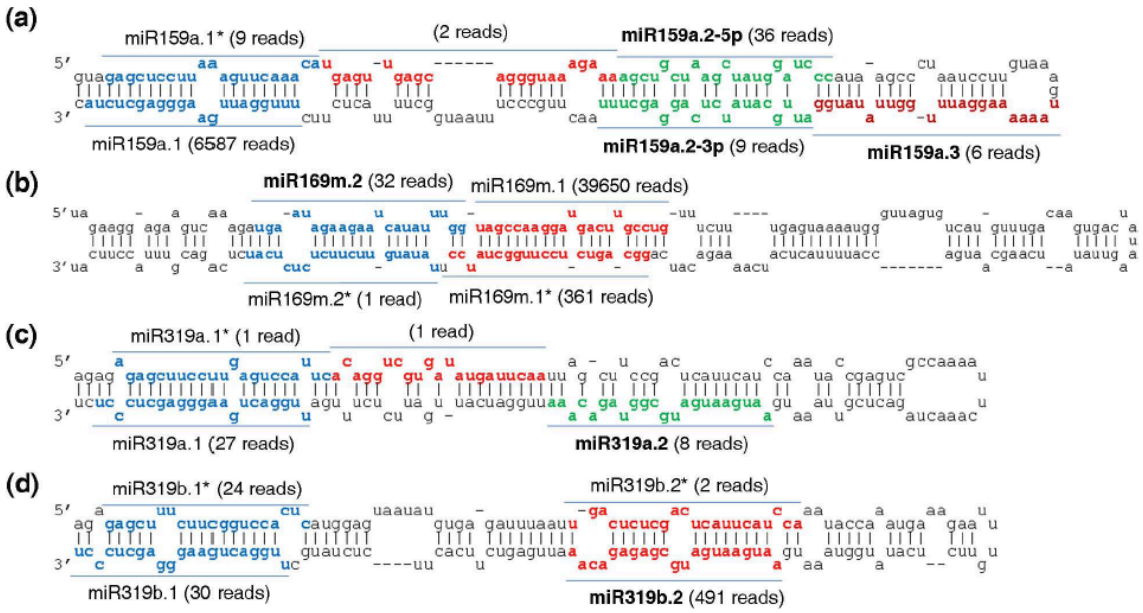


Figure 2.2 miRNA-like RNAs from the miR447a precursor.
(a) The precursor fold-back structure and sequencing reads mapped to miR447a.1 (that is, miR447a) and the individual miRNA-like RNAs. For clarity, miR447a.3* and miR447a.1* are indicated though the numbers of reads mapped to them were below the cutoff threshold of 5.
(b) Distribution of sequencing reads along the precursor.
(c) Expression of miR447a.1 and miR447a.3 in various *Arabidopsis* mutants of small RNA pathways.
(d) Expression of miR447a.3, miR447a.1, and miR447a.2-3p under the challenge of three *Pst* strains (*hrcC*, *avrRpt2* and EV) and in a mock control at 6 and 14 hpi.

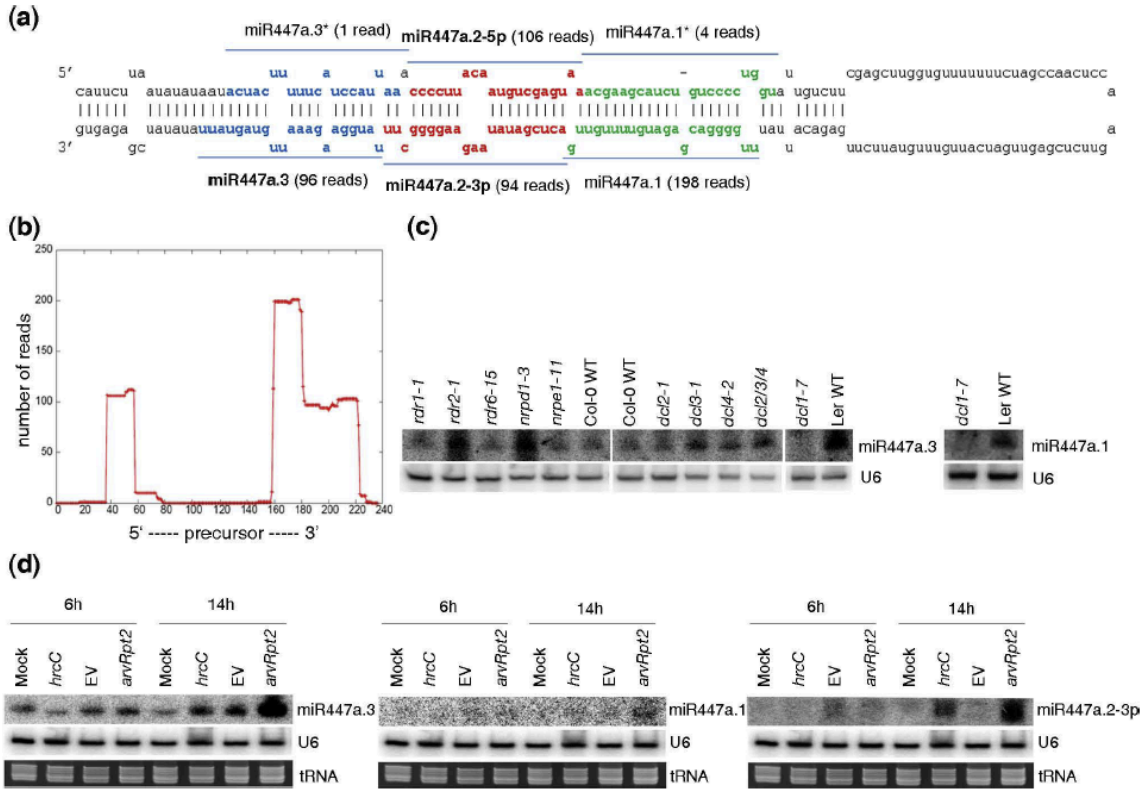


Figure 2.3 miRNA-like RNAs from the miR822 precursor.
(a) The precursor fold-back structure and sequencing reads mapped to miR822.1 (that is, miR822) and individual miRNA-like RNAs.
(b) Distribution of sequencing reads along the precursor. For clarity, miR822.5* is indicated though the number of reads mapped to it is below the cutoff threshold of 5.
(c) Expression of miR822.1 and miR822.2 in various *Arabidopsis* mutants of small RNA pathways.
(d) Expression of miR822.2, miR822.1, miR822.3, miR822.1* and miR822.3* under the challenge of three *Pst* strains (*hrcC*, *avrRpt2* and EV) and in a mock control at 6 and 14 hpi.

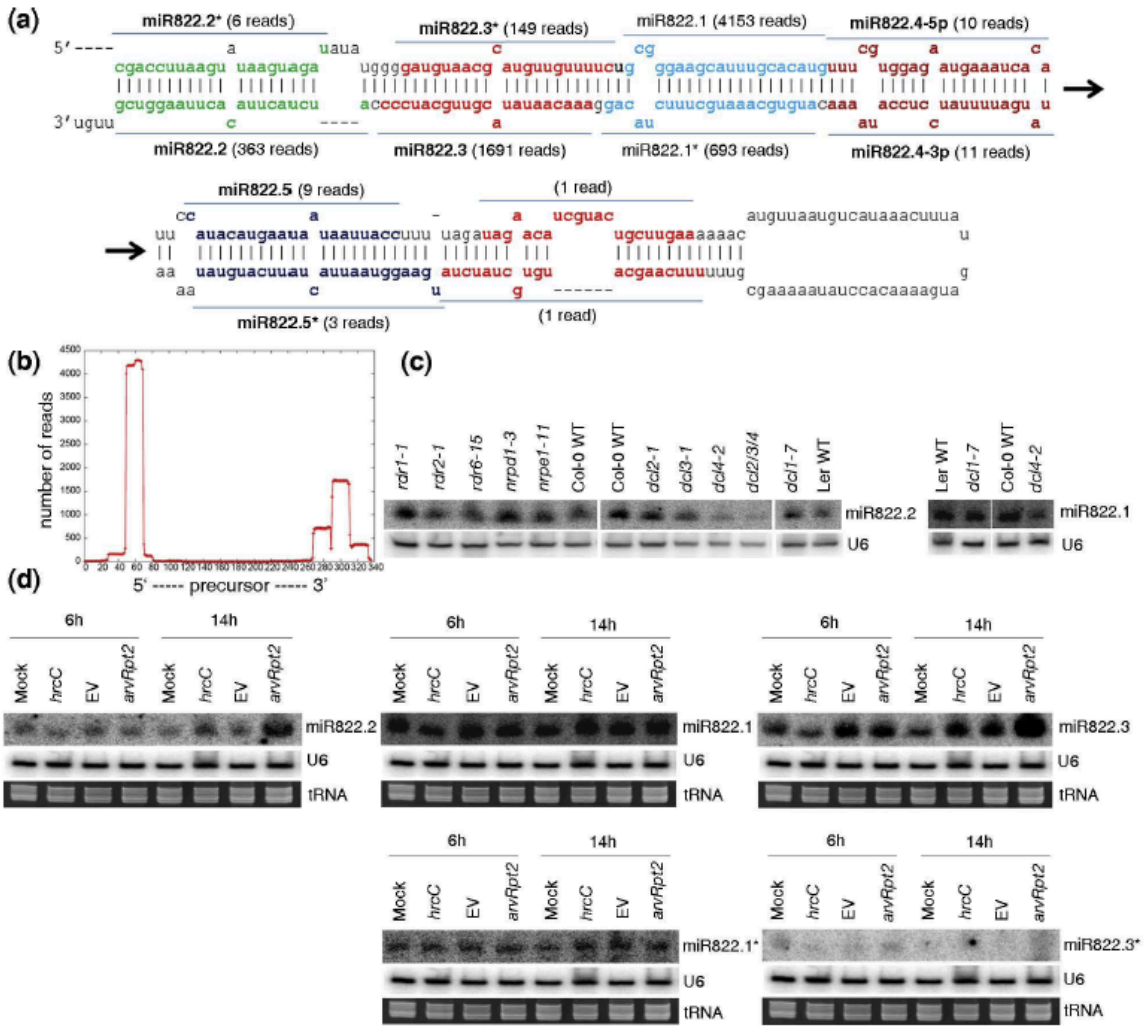


Figure 2.4 miRNA-like RNAs from the miR839 precursor.
(a) The precursor fold-back structure and sequencing reads corresponding to the individual miRNA-like RNAs miR839.1 (that is, miR839) and miR839.1* (that is, miR839*). Note that miR839.1*, miR839.2 and miR839.3 have more sequencing reads than miR839.1. For clarity, miR839.3* was is though the number of reads mapped to it is below the cutoff threshold of 5.
(b) Distribution of sequencing reads along the precursor.
(c) Expression of miR839.1, miR839.2 and miR839.3 under the challenge of three *Pst* strains (*hrcC*, *avrRpt2* and EV) and in a mock control at 6 and 14 hpi.

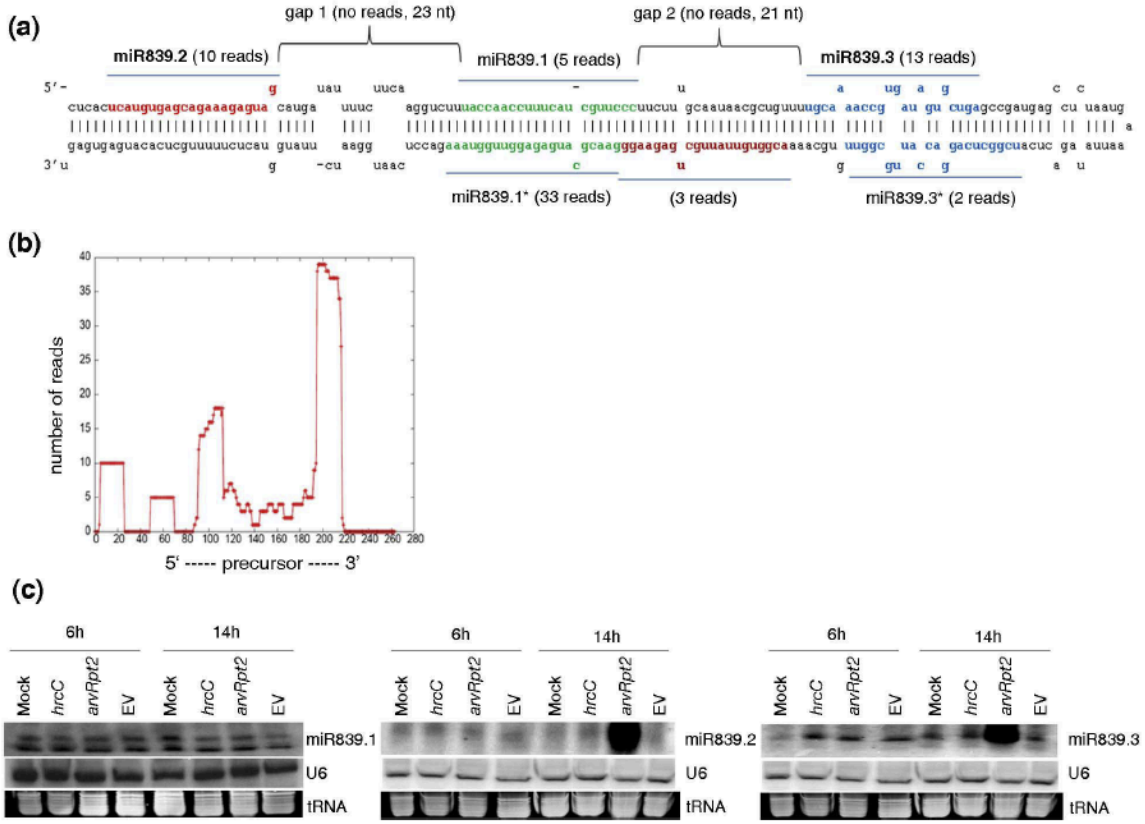


Figure 2.5 Accumulation of miR447a.3 and miR822.2 in a mutantof the small-RNA methyltransferase gene HEN1.
WT, wild type. U6, the control, shows sRNA equal loading.

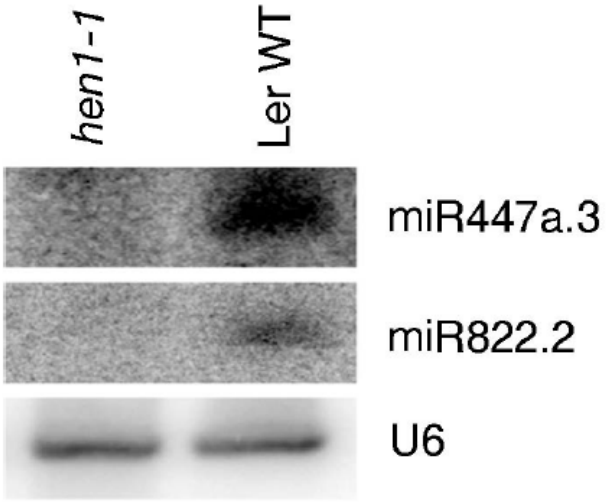


Figure 2.6 Negative correlation between the expression of selected miRNA-like RNAs and their targets.

(a) The expression of targets of miR169i.2-3p, miR169j.2, miR447a.3, miR839.2 and miR839.3 in a dcl1-9 mutant relative to that in the Ler wild type (WT), measured by realtime RT-PCR.

(b) The expression level of the miR822.4-5p target in a dcl4-2 mutant relative to the Col-0 wide type.

(c) The expression of two miR447a.3 targets and one miR839.2 target under the challenge of *Pst* DC3000 (*avrRpt2*) relative to that in the mock treatment. Actin was used as an internal control for delta Ct calculation. Error bars correspond to standard deviation data from three independent reactions. The experiments were repeated on three sets of biological samples and similar results were obtained.

(d) Alignments of selected miRNA-like RNAs and some of their mRNA targets whose expression was compared in Dicer mutants and in the wide type (a, b) and under bacterial infection and under mock infection (c). Included are alignment scores and P-values of target signatures if miRNA-like RNAs had target degradation products in the three small RNA degradome datasets. The arrows are the target cleavage sites detected in the degradome data.

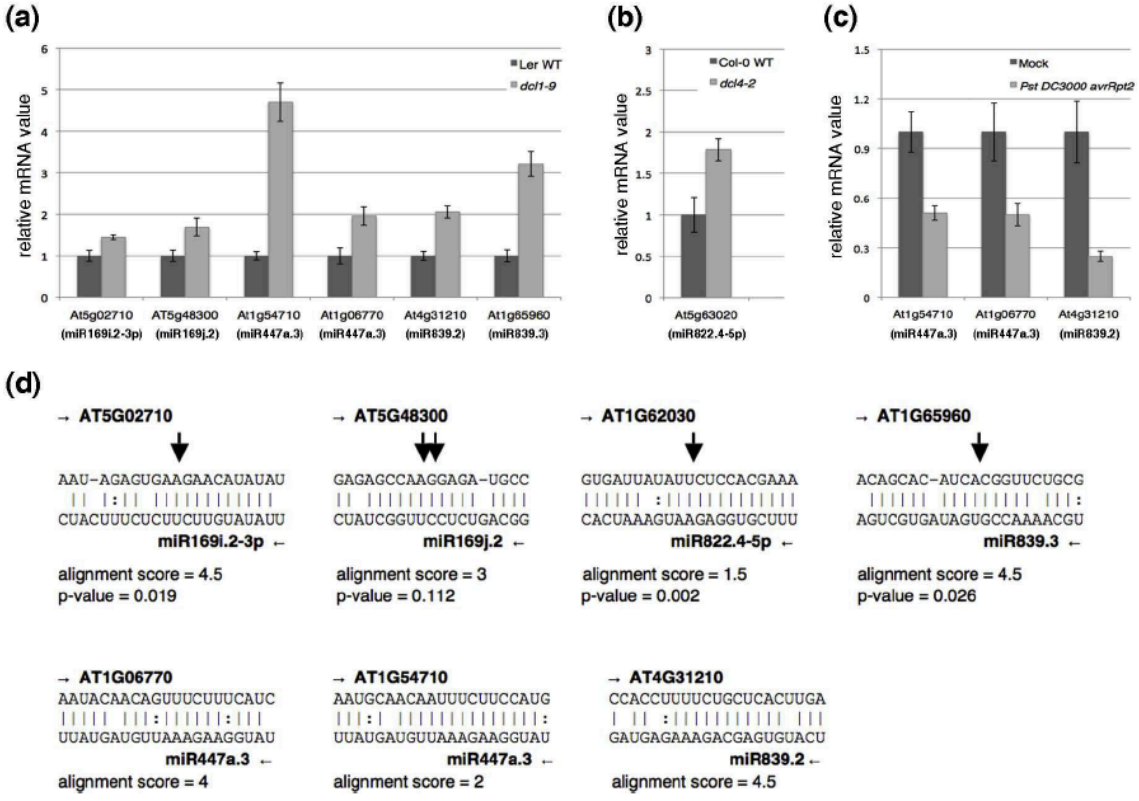


Figure 2.7 Conservation of miRNA-like RNAs and miRNAs. The newly identified miRNA-like RNAs are in red and the annotated miRNAs and miRNA*s are in blue. Also included are the consensus sequences.

(a) Conservation of miR159a.2-5p/miR159a.2-3p and miR159a.1/miR159a.1* (that is, miR159a/miR159a*) in four plants, A. thaliana (ath), O. sativa (osa), P. patens (ppt) and M. truncatula (mtr).

(b) Conservation of miR319.2/miR319b.2* and miR319.1/miR319b.1* (that is, miR319/miR319b*) in four plants, *Arabidopsis*, rice, *Medicago* and P. trichocarpa (ptc).

(c) Conservation of miR7-1 in human and miR7a-1 in mouse.

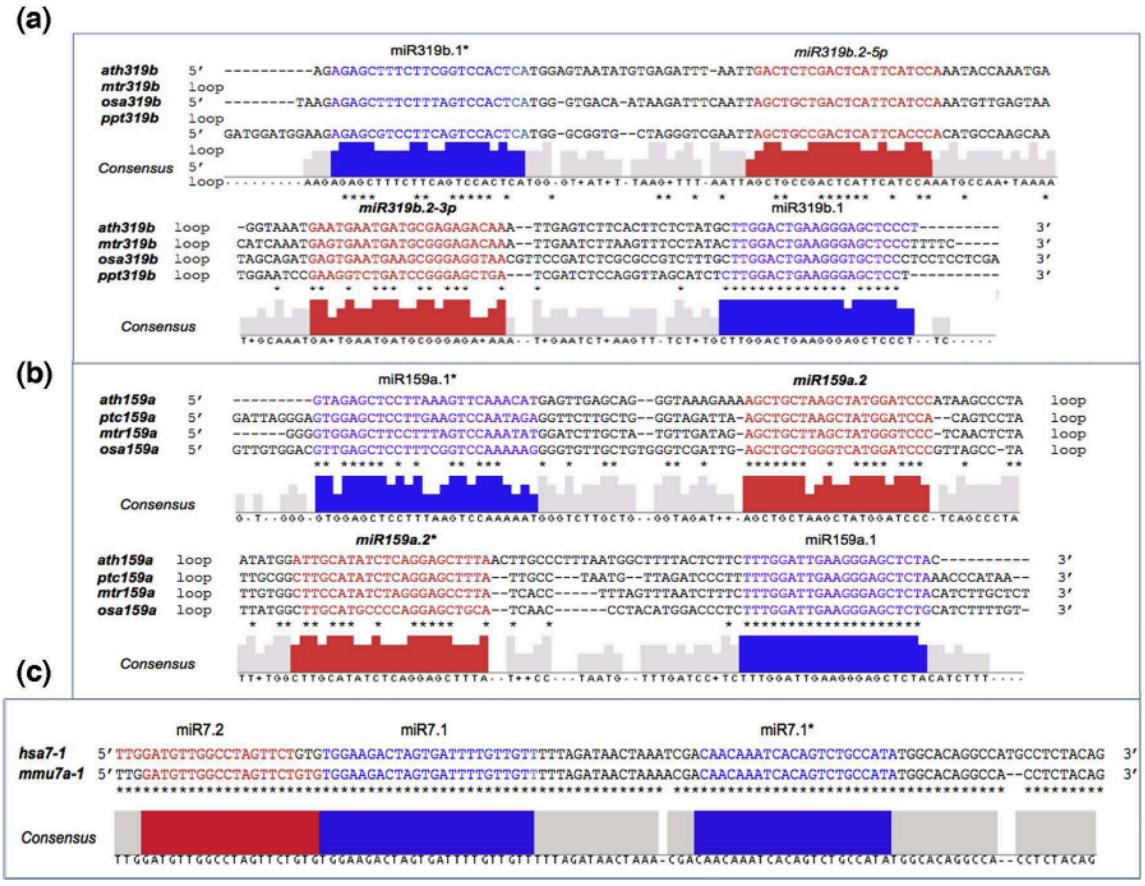


Table 2.1 Nineteen *Arabidopsis* miRNA precursors from 10 miRNA families generate a total of 35 miRNA-sibling RNAs

Nineteen *Arabidopsis* miRNA precursors (the MIR column) from 10 miRNA families generate a total of 35 miRNA-sibling RNAs (the miRNA ID and miRNA sequence columns). For a particular precursor, the positions of newly identified miRNA-like RNAs relative to the cognate miRNA are indicated in the **'**Position**'** column, where a plus sign (+) means that miRNA-sibling small RNAs (msRNAs) are in the upper arm of the hairpin close to the loop, a minus sign (-) indicates that they reside in the lower portion near the base of the hairpin, and **'**+/-**'** means that multiple msRNAs appear on both sides of the sibling miRNA, which is also included here. The cognate (known) miRNAs are named as miRn.1 (see the main text). The Argonaute proteins that a miRNA-like RNA can associate with are listed in the **'**AGO**'** column. The **'**PARE**'** column lists the mRNA genes for which a miRNA-like RNA has a corresponding small RNA target product in the three PARE/degradome datasets considered.

MIR	Loci	miRNA ID	miRNA sequence	Reads	AGO	PARE
miR159a	+	miR159a.1*	GAGCTCCTTAAAGTTCAAACA	9		
		miR159a.2-5p	**AGCTGCTAAGCTATGGATCCC**	36	2,7	
		miR159a.3	**TAAAAAAGGATTTGGTTATA**	6		
		miR159a.2-3p	**ATTGCATATCTCAGGAGCTTT**	9	1,2,7	At5g24620
		miR159a.1*	TTTGGATTGAAGGGAGCTCTA	6,587		
miR519b	+	**miR159b.2**	**AGCTGCTAAGCTATGGATCCC**	36	7	
		miR159b.2*	**ATGCCATATCTCAGGAGCTTT**	14	1,2,7	
		miR159b.1	TTTGGATTGAAGGGAGCTC *	2481		
miR168a	+	miR168a.1	TCGCTTGGTGCAGGTCGGGAA	266,020		
		miR168a.2	**ATTGGTTTGTGAGCAGGGATTGGAT**	10	2	
		miR168a.1*	CCCGCCTTGCATCAACTGAAT	1,497		
miR169b	-	**miR169b.2**	**TGAAGTGGAGTAGAGTATAATG**	7		At4g17420
		miR169b.1	CAGCCAAGGATGACTTGCCGG	4,444		
		miR169b.1*	GGCAAGTTGTCCTTCGGCTACA	8		
miR169f	-	**miR169f.2**	**TGAAGGAATAACGAATGGAAT**	108	1	
		miR169f.1	TGAGCCAAGGATGACTTGCCG	5,779		
		miR169f.1*	GCAAGTTGACCTTGGCTCTGC	2,505		
miR169i	-	**miR169i.2-5p**	**TGAATAGAAGAATCATATTTGG**	32		
		miR169i.1	TAGCCAAGGATGACTTGCCTG	44,477		
		miR169i.1*	GGCAGTCTCCTTGGCTATC	360		
		miR169i.2-3p	**TTATATGTTCTTCTCTTTCATC**	9		At5g02710
miR169j	-	miR169j.1	TAGCCAAGGATGACTTGCCTG	44,458		
		miR169j.1*	AATCTTGCGGGTTAGGTTTCA	9		
		miR169j.2	**GGCAGTCTCCTTGGCTATC**	224	4	At5g48300
miR169l	-	miR169l.1	TAGCCAAGGATGACTTGCCTG	44,392		
		miR169l.1*	AATCTTGCGGGTTAGGTTTCA	9		
		miR169l.2	**AGGCAGTCTCTTTGGCTATC**	366		
miR169m	-	**miR169m.2**	**TGAATAGAAGAATCATATTTGG**	32		
		miR169m.1	TAGCCAAGGATGACTTGCCTG	39,650		
		miR169m.1*	GGCAGTCTCCTTGGCTATC	361		
miR169n	-	**miR169n.2***	**TGGCGGAAAGCGTCATGTTTAG**	10	4	
		miR169n.1	TAGCCAAGGATGACTTGCCTG	44,458		
		miR169n.1*	AATCTTGCGGGTTAGGTTTCA	9		
		miR169n.2	**AGGCAGTCTCTTTGGCTATC**	366		
miR319a	+	**miR319a.2**	**AATGAATGATGCGGTAGACAAA**	8	1,2,4,5	
		miR319a.1	TTGGACTGAAGGGAGCTCCCT	27		
miR319b	+	miR319b.1*	GAGCTTTCTTCGGTCCACTC	28		
		miR319b.2	**AATGAATGATGCGAGAGACAA**	491	1,2	
		miR319b.1	TTGGACTGAAGGGAGCTCCCT	30		
miR447a	-	**miR447a.2-5p**	**ACCCCTTACAATGTCGAGTAA**	106	2,4,5	
		miR447a.1	TTGGGGACGAGATGTTTTGTTG	198		
		miR447a.2-3p	**ACTCGATATAAGAAGGGGCTT**	94	1,2,4,5,7	
		miR447a.3	**TATGGAAGAAATTGTAGTATT**	96	1,2,4,5,7	
miR447b	-	miR447b.1*	AGTAAACGAAGCATCTGTCCCC	8		
		miR447b.1	TTGGGGACGAGATGTTTTGTTG	198		
		miR447b.2	**ACTCGATATAAGAAGGGGCTT**	94	2,5,7	
		miR447b.3	**TATGGAAGAAATTGTAGTATT**	96	2,4,7	
miR775	+	miR775.1*	GCACTACGTGACATTGAAAC	8		
		miR775.2	**TTTGGTTTGTTCAAAGACATT**	10	5	
		miR775.1	TTCGATGTCTAGCAGTGCCA	3,136		
miR822	+/-	**miR822.2***	**CGACCTTAAGTATAAGTAGAT**	6		

Table 1 Nineteen *Arabidopsis* miRNA precursors from 10 miRNA families generate a total of 35 miRNA-sibling RNAs *(Continued)*

		miR822.3*	**GATGTAACGCATGTTGTTTTCT**	149	2,4,7	
		miR822.1	TGCGGGAAGCATTTGCACATGT	4,153		
		miR822.4-5p	**TTTCGTGGAGAATGAAATCAC**	10	1,4	At1g62030, At2g04680
		miR822.5	**CATACATGAATAATAATTACC**	9	1,5	
		miR822.4-3p	**TATGATTTTATCCTCCATAAAA**	11	5	
		miR822.1*	ATGTGCAAATGCTTTCTACAG	693		
		miR822.3	**AAACAATATACGTTGCATCCC**	1,691	1,2,4,7	
		miR822.2	**ATCTACTTACACTTAAGGTCG**	363	1,2,4,5	
miR839	+/-	**miR839.2**	**TCATGTGAGCAGAAAGAGTAG**	10	1	
		miR839.1	TACCAACCTTTCATCGTTCCC	5		
		miR839.3	**TGCAAAACCGTGATAGTGCTGA**	13	1,2,4,7	At1g65960
		miR839.1*	GAACGCATGAGAGGTTGGTAAA	33		
miR841	-	**miR841.2**	**TGTTCTTAAGTTGCTTGTGAA**	8	1	
		miR841.1	TACGAGCCACTTGAAACTGAA	59		
		miR841.1*	ATTTCTAGTGGGTCGTATTCA	3,904		
miR846	+	miR846.1*	CATTCAAGGACTTCTATTCAG	59		
		miR846.2	**AATTGGATATGATAAATGGTAA**	34		
		miR846.2*	**ACTTTTATCATATCCCATCAG**	18		
		miR846.1	TTGAATTGAAGTGCTTGAATT	37		

Table 2.2 miRNA precursors of miRNA-like RNA and miRNA families in plants and animals

Species	Number of precursors	Number of families	miRNA precursors
Arabidopsis	19	10	**159a/b**, 168a, **169**b/f/i/j/**l**/m/n, **319a/b**, 447a/b, 775, 822, 839, 841, 846
Rice	75	42	156i, **159a/b/c**/d/e/f, 169a, 171i, **319a/b**, 396e/f, 437, 441b, 444b/c/d, 445a/c/d/e/g/h/i, 806h, 807a, 809e/f, 812h/i/j, 820a/b/c, 821a/b/c, 1318, 1319, 1423b, 1427, 1428e/g, 1430, 1432, 1439, 1440, 1441, 1442, 1850, 1851, 1858b, 1862b/d, 1864, 1871, 1874, 1875, 1881, 1883b, 1884a/b, 2121b, 2122, 2123a/b, 2124a/b/c/e/f/g/h/i
Moss	37	27	160i, **319a/b/c**/e, 390a/b, 408b, 533a, 534a, 535d, 536d, 538a/c, 893, 896, 898b, 899, 902k, 904a/b, 1023a, 1030a/f, 1047, 1051, 1063b/d/g, 1068, 1069, 1070, 1072, 1219a/d, 2079, 2083
Medicago	9	7	**159a**, **169**d/**l**, **319a/b**, 1510b, 2087, 2088, 2089
Populus	11	5	**159a/b/c**, **319c**/d/f/g, 394a/b, 482, 1450
Human	14	14	**7-1**, 15b, 34a, 124-2, 181a-2, 187, 193a, 204, 221, 331, 498, 518c, 769, 1977
Mouse	27	21	**7a-1**, 20a, 24-2, 30b, 96, 101b, 138-1, 328, 329, 331, 337, 341, 434, 466l, 669a-1/-2/-3, 669b/e/m-1/m-2, 674, 685, 708, 1199, 2134-3
C. elegans	17	16	51, 60, 61, 63, 64, 70, 75, 78, 80, 232, 244, 257, 258-1/-2, 353, 354, 1817
Drosophila	9	9	33, 277, 279, 283, 964, 976, 988, 997, 999

miRNA precursors of miRNA-like RNA and miRNA families in five plant species (*Arabidopsis*, rice, moss, *Medicago* and *Populus*) and four animal organisms (human, mouse, *C. elegans* and *Drosophila*). Entries in bold are miRNA precursors that produce miRNA-like RNAs in more than one species.

Supplemental Data: http://genomebiology.com/2010/11/8/R81/additional

CHAPTER 3

MicroRNAs regulate plant innate immunity by modulating plant hormone networks

ABSTRACT

MicroRNAs (miRNAs) are key regulators of gene expression in development and stress responses in most eukaryotes. We globally profiled plant miRNAs in response to infection of bacterial pathogen *Pseudomonas syringae pv. tomato* (*Pst*). We sequenced 13 small-RNA libraries constructed from *Arabidopsis* at 6 and 14 hours post infection of non-pathogenic, virulent and avirulent strains of *Pst*. We identified 15, 27 and 20 miRNA families being differentially expressed upon *Pst* DC3000 *hrcC, Pst* DC3000 EV *and Pst* DC3000 *avrRpt2* infections, respectively. In particular, a group of bacterial-regulated miRNAs targets protein-coding genes that are involved in plant hormone biosynthesis and signaling pathways, including those in auxin, abscisic acid, and jasmonic acid pathways. Our results suggested important roles of miRNAs in plant defense signaling by regulating and fine-tuning multiple plant hormone pathways. In addition, we compared the results from sequencing-based profiling of a small set of miRNAs with the results from small RNA Northern blot and that from miRNA quantitative RT-PCR. Our results showed that although the deep-sequencing profiling results are highly reproducible across biological replicates, the results from deep sequencing may not always be consistent with the results from Northern blot or miRNA quantitative RT-PCR. We discussed the procedural differences between these techniques that may cause the inconsistency.

KEYWORDS

MicroRNAs · Post transcriptional regulation · Bacterial infection · Deep sequencing · Innate immunity · Hormone networks

INTRODUCTION

MicroRNAs (miRNAs) are small noncoding RNAs that exist in most eukaryotic genomes (Carthew and Sontheimer 2009; Bartel 2004; Baulcombe 2004). They play a fundamental role in almost all biological processes, including development, hormone signaling and response to biotic and abiotic stresses (Bartel 2004; Rubio-Somoza et al. 2009; Jin 2008; Sunkar et al. 2007; Padmanabhan et al 2009). miRNAs function through partial complementary base pairing to messenger RNAs (mRNAs), resulting mainly in post-transcriptional regulation of the targeted genes. Information of miRNA abundance under various conditions can thus provide insight to their regulatory functions.

Several studies have shown that small RNAs play a critical role in disease resistance responses (Jin 2008; Padmanabhan et al. 2009; Navarro et al. 2006; Katiyar-Agarwal et al. 2006, 2007; Gibbings and Voinnet 2010). miR393 can be induced by bacterial elicitor flg22 and positively contributes to pathogen associated molecular pattern (PAMP)-triggered immunity (PTI) by silencing auxin receptors and subsequently suppressing auxin signaling (Navarro et al. 2006). Auxin promotes plant growth, which provides carbon and nitrogen sources for biotrophic pathogens and increase their virulence and susceptibility. In addition, auxin can also promote pathogenesis by suppressing salicylic acid (SA)-mediated defense responses (Chen et al. 2007; Achard et al. 2007; Grant and Jones 2009; Spoel and Dong 2008). Small RNA profiling analysis using deep sequencing revealed that in addition to miR393, miR167 and miR160 were also induced by a non-pathogenic *Pseudomonas syringae* pv. *tomato (Pst*) DC3000 strain with a mutated type III secretion system *hrcC* (Fahlgren et al. 2007). Profiling of AGO1-bound small RNA after flg22 treatment revealed the role of miR160a, miR398b and miR773 in callose deposition, indicating their involvement in PAMP triggered immunity (Li et al. 2010). miR398 is involved in both biotic and abiotic responses (Sunkar et al. 2006; Jagadeeswaran et al. 2009). It is down-regulated by bacterial pathogen *Pst* DC3000 (*avrRpt2*) or *Pst* (*avrRpm1*). In *A. tumefaciens*, introduced T-DNA genes can trigger RNA silencing, which was then suppressed by an oncogenic protein. Accumulation of miR393 and miR167 was greatly reduced in C58-induced tumors in plants (Dunoyer et

al. 2006). Furthermore, induced host defense PR-gene expression and TMV resistance in *A. tumefaciens* were correlated with an increased expression of miR393 in the infiltrated area with C58 but not those with a disarmed control (Pruss et al. 2008). In *Pinus taeda*, the expression of 10 tested *Pinus taeda* miRNA (pta-miRNA) families were significantly repressed in the fusiform rust fungus (*Cronartium quercuum*) infected galled stem compared to healthy stem (Lu et al. 2007). Furthermore, the expression of ~82 plant disease-related transcripts was detected to be altered in response to miRNA regulation in pine (Lu et al. 2007). In *Brassica*, miR1885, which was predicted to target TIR-NBS-LRR class genes, was induced in Turnip mosaic virus infected plants (He et al. 2008).

In this paper, we studied the expression of the known miRNAs in *Arabidopsis* in response to infection of non-pathogenic, virulent and avirulent bacterium *Pst* strains using Illumina SBS high-throughput sequencing. Our results indicate that miRNAs contribute to plant immunity by regulating multiple hormone-signaling pathways and some stress-responsive genes. We also compared the results from sequencing-based profiling with those from small RNA Northern blot and miRNA quantitative RT-PCR, revealing some inconsistency between these general techniques. We discussed the procedural differences between these two methods that may lead to the inconsistency.

RESULTS

Identification of Pst-regulated miRNAs by deep sequencing

To examine the expression of *Arabidopsis* miRNAs in response to various strains of bacterial pathogen *Pst*, we performed a small-RNA profiling analysis using Illumina SBS deep sequencing. RNAs were isolated from 4-week-old short-day grown *Arabidopsis* Col-0 plants after 6 and 14 hours post-inoculation (hpi) with mock (10mM $MgCl_2$), a non-pathogenic strain *Pst* DC3000 *hrcC*, a virulent strain *Pst* DC3000 carrying an empty vector (EV), and an avirulent strain *Pst* DC3000 carrying an effector *avrRpt2* (eight libraries). We included biological replicates for mock 6-hpi, mock 14-hpi, *Pst* DC3000 *hrcC* 14-hpi and *Pst* DC3000 (*avrRpt2*) 14-hpi samples; we also sequenced one

of the *Pst* DC3000 (*avrRpt2*) 14-hpi samples twice to ensure the reproducibility of the sequencing results (five libraries in total). *Pst* DC3000 (*avrRpt2*) induces PTI triggered by the recognition of PAMPs by host PAMP recognition receptors. *Pst* DC3000 (EV) induces effector–triggered susceptibility (ETS) by injecting effector proteins into the plant cell to suppress PTI (Chisholm et al. 2006; Jones and Dangl 2006). *Pst* DC3000 (*avrRpt2*) induces effector-triggered immunity (ETI) and hypersensitive response (HR) in Col-0 carrying cognate resistance gene RPS2 and leads to cell death symptom usually at 15-hpi under the condition and concentration we used. Our samples were collected at 14-hpi before HR could be visualized. Sequencing reads with recognizable adaptor sequences removed were compared with Arabidopsis nuclear, chloroplast, mitochondrial genome sequences and cDNAs. From a total of more than 24.6 million sequence reads from all libraries, we obtained 13,985,938 total reads that have high sequencing quality (i.e., with no N's), contain a 30 sequencing adaptor sequence, and can map perfectly to the *Arabidopsis* genome and cDNA sequences with zero mismatches. This large collection of deep-sequencing data has been deposited to the NCBI/GEO databases (GSE19694). The data have been recently analyzed to identify *Arabidopsis* miRNA precursors that can harbor more than one authentic miRNAs (Zhang et al. 2010) and miRNA precursors that produce, at the same sites, miRNAs as well as siRNAs, some of which guide DNA methylation at their target loci (Chellappan et al. 2010). In the current study, we further analyze our deep sequencing data to understand the function of miRNAs in response to the challenge of bacteria infection.

Within the small-RNA population, 21- and 24-nt small RNA species are the two most abundant among all size classes (Figure 3.1a). Within the reads that can map to the known miRNAs, 80% have length of 21-nt (Figure 3.1c). More than 90% of these miRNA reads start with a uracil (Figure 3.1(d)), which is in shape contrast to the first nucleotide distribution of all small RNA reads (Figure 3.1b). This result is consistent with the current understanding of miRNA composition. We also examined these distributions for individual conditions, but no obvious differentce among them and the results in Fig 3.1 was observed (data not shown).

Although a detailed characterization of miRNA expression, particularly the absolute expression level of a miRNA, is useful, identification of differentially expressed (DE) miRNAs at the whole genome level in response to environmental stimuli or endogenous cues is often desirable for elucidating miRNA function in particular cellular processes. To identify DE miRNAs, we need to consider the reproducibility of sequencing data. We first sequenced one of the libraries, *Pst* DC3000 (*avrRpt2*) 14-hpi, twice. We compared the sequencing read numbers for miRNAs in the two technical duplicate libraries. The counts of reads in each library were normalized to the numbers of reads per million (RPM) to adjust for the differences in library sizes or sequencing depths (see Methods). As shown in Figure 3.2(a), the normalized numbers of reads for miRNAs in two technical duplicates for the *Pst* DC3000 (*avrRpt2*) 14-hpi sample were strongly correlated, with the correlation coefficient exceeded 0.94 and the slope of linear regression being 0.9766. We then sequenced biological replicates under 4 conditions. The normalized result on the biological replicates for mock 14-hpi (Figure 3.2(b)) was also highly correlated, with the correlation coefficient of 0.838 and the slope of linear regression of 0.809. The results on the other three pairs of biological replicates, i.e., that for mock 6-hpi, DC3000 *hrcC* 14-hpi and *Pst* DC3000 (*avrRpt2*) 14-hpi, were also well correlated (data not included), similar to that in Figure 3.2(b). Our results showed that the read counts of miRNAs from Illumina high-throughput sequencing are highly reproducible and the read counts of a miRNA in different small-RNA libraries are good indicators of the abundance of the miRNA in the corresponding biological samples.

We mapped the adaptor-trimmed sequence reads to the known miRNA loci to detect the expressed miRNAs. In this mapping process, we only considered perfect matches and those reads whose first nucleotide to be within 3-nt upstream or down-stream of the starting position of the annotated starting position of a miRNA to compensate the imperfect cleavage activities of Dicer-like proteins (DCLs). We detected the expression of Arabidopsis miRNAs in 86 families (Supplemental Table 1).

The sequence reads mapped to miRNAs were then normalized (see "Methods"). Based on the normalized read counts of miRNAs, DE miRNAs were identified using stringent criteria combining fold change and a Chi-square test across the infection conditions of interest and the respective mock treatments. Specifically, we used strict criteria of at least 10 raw sequence reads, at least threefold change between the infected and the mock, and a **P-**value for a Chi-square test no greater than 0.05 (see "Methods"). We identified a substantial number of DE miRNAs responding to pathogen infections. The DE miRNAs for the three infection conditions at the two time points and their fold changes are shown in Fig. 3.3. In particular, we identified 15, 27 and 20 miRNA families to be differentially expressed upon *Pst* DC3000 *hrcC*, *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) infections at either 6- or 14-hpi, respectively (Fig. 3.3). Overall, *Pst* DC3000 *hrcC* had the least effect on miRNA expression, whereas *Pst* DC3000 (EV) affected the expression of the most miRNAs among the three strains we studied (Fig. 3.3). Among all the miRNAs, miR158 was highly induced by all three strains at 14-hpi, and by *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) at 6-hpi (Fig. 3.3). miR158 targets several genes encoding PENTATRICOPEPTIDE REPEAT (PPR) domain-containing proteins, as well as FUCOSYLTRANSFERASE genes (AtFUT1 and AtFUT2) that encode glycosyltransferases for cell wall xyloglucan biosynthesis. PPR-domain containing proteins are mostly mitochondria or chloroplasts localized putative RNA binding proteins involved in RNA processing, metabolism, editing or translation. Their function in plant disease resistance remains to be explored. An untypical PPR-domaincontaining protein gene PPRL is a target of an endogenous nat-siRNA and negatively regulates *Pst* DC3000 (*avrRpt2*)-triggered ETI (Katiyar-Agarwal et al. 2006). Interestingly, miR403, which targets RNAi effector genes ARGONAUTE 2 (AGO2) and AGO3, was the only miRNA that was suppressed by all the three strains at 14 hpi, and weakly by *Pst* DC3000 (*avrRpt2*) and *Pst* DC3000 *hrcC* at 6 hpi (Fig. 3.3), this result suggests the potential role of AGO2 and AGO3 in plant immunity.

Notably, several miRNAs that regulate plant hormone signaling and biosynthesis were differentially expressed under bacterial infection. Firstly, the expression of a group

of miRNAs involved in auxin signaling pathways was largely altered. miR160 and miR167 that target *Auxin Response Factors* (*ARFs*) were both induced by all of the three *Ps* strains at 6-hpi (Figure 3.3). miR390 (miR390a), which triggers the production of TAS3 trans-acting siRNAs (ta-siRNAs) that also regulate another subset of *ARFs*, mainly *ARF3* and *ARF4*, was down-regulated at 6-hpi of virulent strain *Pst* DC3000 (EV) (Figure 3.3(b)). More than 30% of the 23 ARFs in *Arabidopsis* could be regulated by miR160, miR167, or miR390/TAS3 tasiRNAs, and function as positive or negative regulators of auxin signaling pathways (Guifoyle and Hagen 2007). In addition, miR393, which targets auxin receptors, *TIR1, AFB2,* and *AFB3*, was also differentially regulated by *Pst* DC3000 (EV), *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) at 6- or 14-hpi, respectively (Figure 3.3). Secondly, miR159 was slightly down-regulated by *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) at 6-hpi, but strongly up-regulated by *Pst* DC3000 (*avrRpt2*) at 14-hpi (Fig. 3.3b, c). miR159 targets transcription factors *MYB33, 65* and *101*, the homologous genes of the barley *GAMYB* that activates Gibberellin (GA)-signaling pathways (Reyes and Chua 2007; Millar and Gubler 2005). MYB33 and MYB101 act as positive regulators of ABA signaling pathways in Arabidopsis (Reyes and Chua 2007).

Furthermore, stress-related miRNAs were also differentially regulated by bacteria-challenges. miR398, which targets Copper/Zinc superoxide dismutase genes (*CSDs*), was repressed at 14-hpi of *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) (Figures 3.3a, c), which is consistent with an early result (Jagadeeswaran et al. 2009). *Pst* DC3000 *hrcC* triggers PTI and *Pst* DC3000 (*avrRpt2*) triggers ETI (Chisholm et al. 2006; Jones and Dangl 2006). Both immunity responses trigger oxidative burst resulting in the rapid accumulation of reactive oxygen species (ROS) (Lamb and Dixon 1997). Down-regulation of miR398 under the same conditions will lead to elevated level of *CSD* that help detoxify increased ROS. miR398 is also suppressed by flg22 (Li et al. 2010). Overexpression of miR398b negatively regulates callose deposition and exhibits enhanced susceptibility to virulent and non-pathogenic strains of Pst (Li et al. 2010). miR408 which was predicted to target copper protein plantacyanin, laccase copper

protein and copper ion binding protein genes, was first induced at 6-hpi (Figure 3.3) and then down-regulated at 14-hpi by *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) (Figures 3.3a, c). miR398 and miR408 are also differentially regulated by abiotic stresses (Trindade et al. 2010; Abdel- Ghany and Pilon 2008), suggesting their general roles in stress responses.

Expression validation using small RNA Northern blot analysis

To confirm the results we obtained from small-RNA deep sequencing, we performed a small-RNA Northern blot analysis to further validate several miRNAs that were detected as differentially expressed by deep sequencing. We found that the majority of small RNA Northern blot results were consistent with the deep sequencing data (Fig. 3.4). For example, at 6-hpi miR160 and miR408 were induced in all three treatments as determined by both methods (Figs. 3.3, 3.4a). miR408 was induced at 6-hpi in *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) as detected by both methods(Figs. 3.3a, c, 3.4a). Up-regulation of miR822 in *Pst* DC3000 (*avrRpt2*) and *Pst* DC3000 (EV) treatments observed in Northern blot (Fig. 3.4a) was consistent with the result from deep sequencing (Fig. 3.3b, c). At 14-hpi, miR158 was induced by all the three strains and the inductions shown in the data from the deep sequencing and Northern blot matched each other well. Similarly, miR159 and miR393 were induced whereas miR408 and miR822 were suppressed by *Pst* DC3000 (*avrRpt2*) (Fig. 3.4b), which were also consistent with the results from the deep sequencing data (Fig. 3.3c).

On the other hand, we also observed some discrepancies between the levels of miRNA expression measured by high-throughput sequencing and Northern blot (Figs. 3.3 and 3.4). For example, in the deep sequencing results of 6-hpi, miR159 was repressed by *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) (Fig. 3.3b, c), miR393 was down-regulated by all three infections (Fig. 3.3), and miR162 was induced by *Pst* DC3000 (EV) (Fig. 3.3 b). On the contrary, none of these miRNAs showed clear change under any of these conditions in the Northern blot analysis (Fig. 3.4). Similarly, our deep sequencing data showed that miR398 was induced by *Pst* DC3000 (EV) at 6-dpi (Fig. 3.3b), however, Northern blot analysis by Jagadeeswaran et al. did not detect any clear difference of

miR398 expression under the same condition (Jagadeeswaran et al. 2009). On the other hand, Northern blot analysis also detected some expression changes that were not detected by deep sequencing. For example, our small RNA Northern blot analysis revealed that miR319 and miR852 were clearly induced by *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) at 14-hpi (Fig. 3.4b), while the deep sequencing profiling failed to detect these differential expressions (Fig. 3.3). To further examine this observation, we selected three miRNAs (miR159, miR408 and miR319) and quantified them by miRNA real time quantitative RT-PCR (QRT–PCR) (Fig. 3.5). In consistent with the result from Northern blot and deep sequencing results, QRT–PCR results showed that miR159 was up-regulated and miR408 was downregulated in *Pst* DC3000 *hrcC*, *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) as compared with mock treatment at 14-hpi (Fig. 3.5a, b). miR319 showed an overall up-regulation in the treatments of *Pst* DC3000 *hrcC*, *Pst* DC3000 (*avrRpt2*) and *Pst* DC3000 (EV) as compared with the mock (Fig. 3.5c), which is consistent with the Northern blot analysis (Fig. 3.4), although it detected more induction of miR319 by *Pst* DC3000 *hrcC* and less induction of miR31 by *Pst* DC3000 (*avrRpt2*) than that in Northern blot analysis. The possible causes of discrepancies among these three techniques are included in ‘‘Discussion ’’ section.

Expression analysis of the targets of bacteria-regulated miRNAs

To determine whether the expression change of miRNAs in response to Pst infection altered the expression of their targets, we analyzed the public microarray dataset from the AtGenExpress Consortium (ME00331) collected from the same treatments as our small RNA libraries (see “Methods”), i.e., *Pst* DC3000 (EV)- and *Pst* DC3000 *hrcC* - infected plants at 6-hpi. We used miRNA targets that were previously validated to analyze the functions of some of the miRNAs that had a significant expression change between a treatment and the mock (see “Methods”). We focused on those miRNAs that had at least threefolds of expression change in an infection relative to the mock (Fig. 3.3) and their protein-coding targets that exhibited an anti-correlated expression with respect to that of the targeting miRNAs. The results are listed separately in Supplemental Table

3.2. The expression of 12 mRNA target genes, including miR160 and miR167 targets *ARF8, ARF10, ARF16*, and *ARF17*, miR393 targets *TIR1, AFB2* and *AFB3*, and miR159 targets *MYB33* and *MYB65*, was negatively correlated with their corresponding miRNAs under the infection of *Pst* DC3000 *hrcC* at 6-hpi (Supplemental Table 3.2a). Under the infection of *Pst* DC3000 (EV) at 6-hpi, 28 mRNA target genes had expression anti-correlated to the expression of their targeting miRNAs (Supplemental Table 3.2b). Again, miR160, miR167, miR393 that involved in auxin signaling pathway and miR159 that involved in ABA signaling pathway were on the list.

Although many miRNAs and targets exhibited similar anti-correlated expression patterns in both *Pst* DC3000 *hrcC* and *Pst* DC3000 (EV), a substantial number of miRNA/target pairs were infection specific. Specifically, miR166/*ATHB-15*, miR167/*ARF8*, miR393/*AFB2*, and miR393/At3G23690 (a bHLH protein) were specific to *Pst* DC3000 *hrcC* at 6-hpi, whereas ten miRNAs (i.e., miR158, miR161, miR162, miR167, miR171, miR394, miR396, miR402, miR775, and miR847) and their twenty targets listed in Supplemental Table 3.2b were anti-correlated specifically in *Pst* DC3000 (EV). These results suggest that there was specific regulation mediated by miRNAs in the non-pathogenic, compatible and incompatible interactions between bacteria and host, although largely overlapped patterns of miRNA expression were observed. In particular, miR162, which targets *DCL1*, and miR775, which targets a galactosyltransferase family protein, may play a rather specific role in compatible interaction. Future in-depth functional analysis on these miRNAs and their targets in plant defense responses may help address this question. It is already known that bacteria pathogens can suppress miRNA pathway by injecting certain effector proteins into the plant cell (Navarro et al. 2008). It is likely that the virulent strain *Pst* DC3000 (EV) utilizes an additional approach to suppress miRNA pathway by inducing miR162 and subsequently down-regulating *DCL1*, and thus promotes disease. Furthermore, not all the target genes were negatively regulated by corresponding DE miRNAs in a given condition. For example, miR167 was negatively correlated with 3 targets, including *ARF8*, under the infection of *Pst* DC3000 *hrcC* at 6-hpi (Supplemental Table 3.2a), but regulated *ARF6* instead in response to *Pst*

DC3000 (EV) (Supplemental Table 3.2b), suggesting that DE miRNAs may selectively regulate their targets under certain conditions and this type of selective regulation is prevalent. Indeed, this phenomenon of selective regulation of miRNAs was observed previously in other studies, e.g., that in (Rubio-Somoza et al. 2009; Jagadeeswaran et al. 2009). For example, miR398 was suppressed by *Pst* DC3000 (*avrRpt2*) or *Pst* (*avrRpm1*). Only one of the two targets, *CSD1*, was derepressed, but the other target, *CSD2*, was slightly repressed (Jagadeeswaran et al. 2009). Both *CSD1* and *CSD2* were shown to be regulated by miR398 under abiotic stresses (Sunkar et al. 2006). These results illustrated that a single miRNA can respond to both biotic and abiotic stresses and selectively regulate different targets under different conditions.

Furthermore, we quantified target mRNA expression levels by real-time QRT–PCR for the 14-hpi treatments on selected differentially regulated miRNAs from Northern and miRNA QRT–PCR (Fig. 3.6). The results revealed that miR158 target *AtFUT1* (At2g03220) and miR166 target *REVOLUTA* (At5g60690) were down-regulated in both *Pst* DC3000 (EV) and *Pst* DC3000 (*avrRpt2*) treated samples at 14-hpi (Fig. 3.6a,c). *MYB33* (At5g06100), a miR159 target, was down regulated in all three treatments at both 6- and at 14-hpi (Fig. 3.6b). *PLANTACYANIN* (*ARPN*, At2g02850), a target of miR408, was induced by all the three conditions at 14-hpi (Fig. 3.6e). These data showed a clear negative correlation between the expression levels of the mRNA targets and their corresponding miRNAs. In addition, *TCP4*, the target of miR319 was strongly down-regulated by Pst DC3000 (avrRpt2), but weakly repressed by *Pst* DC3000 (EV) (Fig. 3.6d). This result is consistent with the results from miRNA QRT–PCR and Northern blot where miR319 was induced by *Pst* DC3000 (*avrRpt2*). In summary, these results provided insightful information of miRNA-mediated gene regulation in bacterial pathogen infection.

DISCUSSION

miRNAs orchestrate hormone signaling pathways for defense response in plant

We identified a group of miRNAs that were differentially expressed after the infection of bacterial pathogen Pst using small RNA profiling by Illumina SBS highthroughput deep sequencing. Notably, many of these bacteria- regulated miRNAs manipulate the components of plant hormone signaling pathways, including auxin, ABA and jasmonic acid (JA) signaling pathways. Among them, miR160, miR167, miR390 and miR393 all regulate genes involved in the auxin signaling pathway, including different ARFs and auxin receptors. Auxin is an essential phytohormone for growth and development. Many biotrophic pathogens could synthesize auxin or auxin-like molecules to promote virulence (Navarro et al. 2006; Chen et al. 2007; Wang et al. 2007). *Pst* DC3000 (*avrRpt2*) (when expressed in a rps2 mutant background) promotes auxin signaling/biosynthesis (Chen et al. 2007). As a result, host plants have developed several counter measures, including miRNA-mediated gene regulation, to suppress auxin signaling and subsequently inhibit pathogen growth (Navarro et al. 2006; Wang et al. 2007). Our data suggested that miR160, miR167, miR390 and miR393 regulated the expression of a group of signaling genes that were involved in various steps of auxin signaling, and thus contributed to the inhibition of pathogen growth. miR393 was shown to be induced by bacterial elicitor flg22 to repress the expression of auxin receptor genes (Navarro et al. 2006). *Pst* DC3000 *hrcC* could induce miR160, miR167 and miR393 at 3-hpi (Fahlgren et al. 2007). We observed similar induction patterns of miR160 and miR167 at 6-hpi in our study. However, we did not observe clear changes of miR393 expression at 6- and 14-hpi after Pst treatment in our Northern blot analysis, suggesting that miR393 may regulate auxin signaling at an early stage of bacterial infection.

Our small RNA Northern blot analysis revealed that miR319 was induced by Pst DC3000 hrcC and Pst DC3000 (avrRpt2) at 14-hpi (Fig. 3.4). miR319 targets *TCP* (*TEOSINTE BRANCHED/CYCLOIDEA/PCF*) transcription factor family genes which directly regulate *LIPOXYGENASE2* (*LOX2*). *LOX2* encodes a chloroplast-localized enzyme that is responsible for the first step in the JA biosynthesis pathway (Schommer et al. 2008). Because JA signaling is usually antagonistic to salicylic acid (SA) signaling, while SA signaling is important for plant defense against biotrophic pathogens, including

Pst. Down-regulation of JA biosynthesis by miR319 through TCP genes in response to *Pst* DC3000 *hrcC* and *Pst* DC3000 (*avrRpt2*) would promote SA-mediated plant resistance responses.

Furthermore, miR159 was induced by *Pst* DC3000 (*avrRpt2*) at 14-hpi (Figs. 3.3c, 3.4). miR159 targets transcription factors *MYB33* and *MYB101*, both of which are positive regulators of ABA signaling pathways (Reyes and Chua 2007). Although ABA plays a positive role in antibacterial defense at early infection stage by regulating the pre-invasive stomata-based responses (Melotto et al. 2006). ABA was shown to attenuate SA-mediated resistance at later infection stages and can also suppress callose deposition in response to flg22 (de Torres-Zabala et al. 2007). Bacterial effectors, such as AvrPtoB, promote ABA biosynthesis to suppress SA signaling pathway (de Torres-Zabala et al. 2007). Induction of miR159 suppresses ABA signaling via *MYB33* and *MYB101*, which activates SA signaling pathway and subsequently promotes SA-mediated defense responses.

It is worth noting that, in many cases, the avirulent strain *Pst* DC3000 (*avrRpt2*) induces miRNAs to a higher level than by non-pathogenic *Pst* DC3000 (*hrcC*) and virulent strains *Pst* DC3000 (EV). Figure 3.4 shows several such examples, including miR158, miR159, miR166 and miR319. miR159 and miR319 suppress components involved in ABA and JA signaling pathways and subsequently promote SA-mediated defense signaling pathway. These results suggest that plants have more miRNAs involved in multiple hormone signaling pathways to trigger stronger and faster defense responses in ETI than in nonpathogenic strain and virulent strain-induced PTI. Similar observations were made in early mRNA profiling studies (Tao et al. 2003).

In summary, almost all the plant hormones, including SA, JA, ethylene, auxin, ABA, Gibberellin and cytokinin, have been implicated in playing a direct or indirect role in plant defenses (Grant and Jones 2009; Bari and Jones 2009; Spoel et al. 2007). Bacterial pathogens often produce hormones or hormone like molecules, such as auxin or coronatine, to interfere with host hormone homeostasis and signaling, which leads to the suppression of plant immune responses. Plants have developed successful defense

mechanisms against bacterial-triggered hormone perturbation. Our results suggested that a group of miRNAs work coordinately to promote defense responses by regulating and fine-tuning multiple host hormone signaling pathways in response to pathogen invasion.

Discrepancy between deep sequencing and Northern blot for miRNA profiling

Comparing with hybridization-based profiling methods, such as microarray, Northern blot, and PCR, highthroughput sequencing-based profiling methods have at least two advantages. First, they can not only profile the known genes and known small RNAs, but also detect and characterize novel transcripts, while hybridization-based methods can only provide information of the annotated genes. This advantage is important for identification of novel transcripts that are expressed under rare conditions, such as under a peculiar pathogen infection, or at low abundance. Second, sequencing-based methods can profile transcripts of single copy, and more importantly have a resolution of single nucleotide to detect even single nucleotide polymorphisms. This feature is essential for distinguishing individual members of a miRNA family that often differ by one or two nucleotides. In contrast, hybridization-based methods, such as microarray and Northern blot, could often detect a combined signal of individual members of the same miRNA families, due to the possible cross-hybridization that may not be able to discriminate a small number of mismatches between a probe and a sequence annealed. Thus hybridization-based methods measure the abundance of the intended miRNA families, not necessarily individual miRNAs.

The discrepancy between the results from Illumina deep sequencing and Northern blot and miRNA QRT–PCR as observed in the current study is perhaps due to the differences between the procedures of these two profiling techniques. A deep-sequencing method often uses small RNA adaptor ligation that may introduce a bias that some small RNA species are more likely to be ligated than others (Chellappan and Jin 2009; Katiyar-Agarwal and Jin 2007). Some miRNA modifications may also reduce the efficiency of ligation or even block ligation reactions, whereas small- RNA Northern blot is based on nucleotide hybridization, which could reveal all small RNAs that are reverse

complementary to the probes for hybridization regardless the modification. However, Northern blot will not be able to measure the exact size and sequence distribution at a locus. Furthermore, PCR amplification is the last step of the Illumina small RNA library preparation before sequencing. PCR is an inherently biased process because it tends to amplify highly abundant sequences more effectively than low abundant sequences. Therefore, PCR may exaggerate the differences between the high abundant and low abundant sequences. This discrepancy between the current sequencing-based approaches and hybridization-based approaches has been observed in a recent comparison of microarray-based miRNA profiling methods and Illumina deep sequencing (Git et al. 2010). The observation we made in the current study is also in agreement with the result from a comparison of two deep sequencing protocols for Illumina deep sequencing platform, one used RNA amplification and the other did not (Kapranov et al. 2010). The result in (Kapranov et al. 2010) showed that, although the amplification-based protocol in general preserves the general trend of abundances of small RNAs, the normalized abundances of individual small RNAs can vary up to 1-2 orders of magnitude in the amplification-based protocol versus the non-amplification-based protocol, revealing the biases introduced in amplification. In short, the sequencing-based and hybridization-based small-RNA profiling techniques have their own advantages and drawbacks. They may provide different measures of the same transcript in the same biological sample. Likewise, two miRNAs (or siRNAs) with similar numbers of sequencing reads may in fact differ substantially in their absolute abundances in a sample. Therefore, we must exercise extra caution when directly compare small RNA expression information using deep sequencing data and PCR-based assays or Northern blot hybridization.

METHODS

Plant materials, small-RNA library construction and deep sequencing

Arabidopsis Col-0 was used as the plant material in this study. Profiling experiments were performed on 4-week old Arabidopsis plants grown at 23°C with 12-h light. Infiltration of the following bacteria *Pst* DC3000 EV, *Pst* DC3000 *hrcC*, *Pst*

DC3000 (*avrRpt2*) and mock (10 mM $MgCl_2$) was carried out as described previously (Katiyar-Agarwal et al. 2006). The small RNA libraries for deep sequencing were constructed by 50 phosphate-dependent method as described (Chellappan and Jin 2009), and the libraries were sequenced at the Illumina Inc. The raw and processed sequencing reads are available at NCBI/GEO under the accession number GSE19694.

Small-RNA Northern blot analysis

For Northern blot analysis 80μg total RNA was resolved on 15% denaturing polyacrylamide gel. We used the following oligo probes to detect miRNAs. Blots were exposed to phosphorscreens and scanned using PhosphorImager (Molecular dynamics).

miR160-anti: tggcatacagggagccaggca

miR148-anti: gccagggaagaggcagtgcat

miR822-anti: catgtgcaaatgcttcccgca

miR159-anti: tagagctcccttcaatccaaa

miR162-anti: ctggatgcagaggtttatcga

miR393-anti: ggatcaatgcgatccctttgga

miR158-anti: tgctttgtctacatttggga

miR166-anti: ggggaatgaagcctggtccga

miR408-anti: gccagggaagaggcagtgcat

miR852-anti: cagaactaaggcgcttatctt

Quantitative RT-PCR to quantify miRNA expression levels

The expression levels of miRNAs were quantified by a modified quantitative RT-PCR from described (Varkonyi- Gasic et al. 2007). About 2 μg of DNase I (Invitrogen) treated total RNA was used for reverse transcription with artificially designed stem-loop specific RT-primers and oligo dT primer separately using Superscript III (Invitrogen). To capture individual miRNAs, specific stem-loop RT primers that were complementary to each miRNA by a six-nucleotide extension at the 3' end was used. The specificity of stem-loop RT primers to individual miRNAs is conferred by a six-nucleotide extension at the

3’ end. The reaction was incubated at 16°C for 30 min followed by 30 s at 30°C and 42°C and 1 s at 50°C of 60 cycles and finally 5 min at 85°C. Following RT, quantitative PCR in real time was performed using SensiMix SYBR (Quantace) using miRNA specific forward primer and universal reverse primers as listed below. The comparative threshold cycle (Ct) method was used for determining relative transcript levels (Bulletin 5279, Real-time PCR applications Guide, Bio-Rad). Actin was used as an internal control.

QmiR159:RT-primer
GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTAGAGC
miR159Fp:CGGCGGTTTGGATTGAAGGGA
qmiR319:RT-primer.
GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGGAGC
qmiR319Fp:TTGCGGTTGGACTGAAGGGAG
qmiR408:RT-primer.
GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGCCAGG
qmiR408Fp:TTGCGGATGCACTGCCTCTTC
qmiRUniversal Rp: GTGCAGGGTCCGAGGT
ActinF: AGTGGTCGTACAACCGGTATT
ActinR: GATGGCATGAGGAAGAGAGAA

Quantitiative RT-PCR to quantify miRNA targets expression levels

For QRT-PCR analysis, 5µg of total RNA was used for synthesizing cDNA. DNA contamination was removed by using DNase I (Invitrogen). Amplification of miRNA targets was carried out using real-time PCR machine (iQ5, Bio-Rad). The following oligos were used to perform the quantitative RT-PCR analysis.

miR158-ATFUT1(2g03220)QRTF: CCAAATGTTACTAACATCAATTC
miR158-ATFUT1(2g03220)QRTR: CGGACCACATCGCTTGTGAAGCTTT
miR159-MYB33(5g06100)QRTF: TAGATGGGCACGTATGGCTGCACATTTG

miR159-MYB33(5g06100)QRTR: TACTCTTGACTCCACTCAAGTGCCT
miR166-REVOLUTA(5g60690)QRTF: GGATTGCTCTCAATCGCAGAGG
miR166-REVOLUTA(5g60690)QRTR: CTCACAAACTGAGAAGCTGAAGC
miR319-TCP4(3g15030)QRTF:ACCATCACCTTCCTCCACCGGTTCACC
miR319-TCP4(3g15030)QRTR: AAATAGAGGAAGCAGAGGACGGCTTGTG
miR408-ARPN(2g02850)QRTF: TAGAGCCGGCGACGTTCTTGTGTT
miR408-ARPN(2g02850)QRTR: AAGTTATACGATCTTTGCCTGAAGTGTA

Preprocessing and normalization of sequence reads

The raw sequence reads were processed to remove reads with no 30 sequencing adaptors, of low quality (having at least an N), or shorter than 17-nt. The adaptor trimming was done by an in-house program that recursively searches for the longest subsequence of the adaptor appearing within a sequencing read. If a raw sequence read did not have a subsequence of the adaptor longer than 5-nt, it was considered as carrying no adaptor. The adaptor-trimmed reads that are longer than 16nt and can map to the *Arabidopsis* genome (TAIR version 9.0, from http://www.arabidopsis.org) and cDNAs (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_sequences/TAIR9_cdna_20090619) with zero mismatches were retained for furtheranalysis. The mapping of sequence reads to the genome and cDNAs were done using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml).

To facilitate identification of differentially expressed miRNAs, we normalized the numbers of sequencing reads in a small RNA library. For normalization, we assumed the total number of small RNAs in a cell to be a constant. To obtain the total number of reads of all small RNAs produced in the cell, we identified reads that can map to the Arabidopsis nuclear, chloroplast and mitochondrial genomes as well as the transcripts in available cDNA libraries since some small RNAs can originate from mRNA splicing junctions; we called this the number of total mapped reads. The normalized abundance of a unique small RNA is the ratio of the number of sequencing reads for the small RNA and the number of total mapped reads, multiplied by one million.

Identification of differentially expressed miRNAs

To identify differentially expressed miRNAs across bacterial infected and mock treated Arabidopsis plants, we first mapped the pre-processed sequencing reads to the genomic loci of the known miRNAs (using Bowtie) based on miRBase annotation (version 15, http://www.mirbase.org). We only considered miRNAs that had at least 10 raw sequencing reads in an infection and the mock conditions of interest in order to deal with noise in sequencing. In the mapping, we allowed up to 3-nt shift upstream and downstream from the annotated starting locus of a miRNA to compensate for possible variation in Dicer activities. We compared DC3000 EV 6-hpi, DC3000 *hrcC* 6-hpi and *Pst* DC3000 (*avrRpt2*) 6-hpi samples with the mock 6-hpi plant, and DC3000 EV 14-hpi, DC3000 *hrcC* 14-hpi and *Pst* DC3000 (*avrRpt2*) 14-hpi samples with the mock 14-hpi. Using normalized sequence reads, candidate differentially expressed (DE) miRNAs in an infection condition were chosen as those whose normalized abundance increased or decreased by at least threefolds with respect to the corresponding mock treatments. These candidate DE miRNAs were further tested using the Chi-square test with a threshold of 0.05. Only those miRNAs that passed the test were taken as DE genes. In the statistical test, the null hypothesis was that no difference for a miRNA between an infection and the mock treated conditions. The test involved comparing the observed frequency (number of reads) of a miRNA under the infection against its expected frequency in the mock condition.

Microarray data and processing, and miRNA target genes

Genome-wide mRNA expression profiling data from the international collaborative AtGenExpress project, available at TAIR (http://www.arabidopsis.org), were used in our study. Specifically, we used the microarray dataset for *Pst* DC3000 *hrcC* and *Pst* DC3000 (EV) collected at 6-hpi (TAIR Accession number ME00331), which perfectly matched two of the experimental conditions of our study. These datasets and the dataset of its corresponding control from the Affymetrix microarray platform were normalized using gcRMA in BioConductor (http://www.bioconductor. org).

The miRNA target genes we used were from those that were identified by Parallel Analysis of RNA Ends (PARE) and reported in the literature, which were collected by the Meyers lab at (PARE, http://mpss.udel.edu/common/web/ targets.php?SITE=at_pare) and (http://mpss.udel.edu/at/target.php).

REFERENCES

Abdel-Ghany SE, Pilon M (2008) MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. J Biol Chem 283(23):15932–15945

Achard P et al (2007) The plant stress hormone ethylene controls floral transition via DELLA-dependent regulation of floral meristemidentity genes. Proc Natl Acad Sci USA 104:6484–6489

Chellappan P et al (2010) siRNAs from miRNA sites mediate DNA methylation of target genes. Nucleic Acids Res [Epub ahead of print-online available]

Bari R, Jones JD (2009) Role of plant hormones in plant defence responses. Plant Mol Biol 69:473–488

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

Baulcombe D (2004) RNA silencing in plants. Nature 431(7006): 356–363

Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. Cell 136:642–655

Chellappan P, Jin H (2009) Discovery of plant microRNAs and shortinterfering RNAs by deep parallel sequencing. Methods Mol Biol 495:121–132

Chen Z et al (2007) Pseudomonas syringae type III effector AvrRpt2 alters Arabidopsis thaliana auxin physiology. Proc Natl Acad Sci USA 104:20131–20136

Chisholm ST et al (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124(4):803–814 de Torres-Zabala M et al (2007) Pseudomonas syringae pv. Tomato hijacks the Arabidopsis abscisic acid signalling pathway to cause disease. EMBO J 26:1434–1443

Dunoyer P, Himber C, Voinnet O (2006) Induction, suppression and requirement of RNA silencing pathways in virulent Agrobacterium tumefaciens infections. Nat Genet 38(2):258–263

Fahlgren N et al (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2:e219

Gibbings D, Voinnet O (2010) Control of RNA silencing and localization by endolysosomes. Trends Cell Biol 20(8):491-501

Git A et al (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA 16:991-1006

Grant MR, Jones JD (2009) Hormone (Dis)harmony moulds plant health and disease. Science 324:750–752

Guilfoyle TJ, Hagen G (2007) Auxin response factors. Curr Opin Plant Biol 10:453–460

He XF et al (2008) Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIRNBS-LRR class R gene-derived novel miRNA in Brassica. FEBS Lett 582:2445–2452

Jagadeeswaran G, Saini A, Sunkar R (2009) Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis. Planta 229:1009–1014

Jin HL (2008) Endogenous small RNAs and antibacterial immunity in plants. Febs Lett 582(18):2679–2684

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444(7117):323–329

Kapranov P et al (2010) New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. Nature 466:642–646

Katiyar-Agarwal S, Jin H (2007) Discovery of pathogen-regulated small RNAs in plants. Methods Enzymol 427:215–227

Katiyar-Agarwal S et al (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci USA 103(47):18002–18007

Katiyar-Agarwal S et al (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev 21(23):3123–3134

Lamb C, Dixon RA (1997) The oxidative burst in plant disease resistance. Annu Rev Plant Physiol Plant Mol Biol 48:251–275

Li Y et al (2010) Identification of microRNAs involved in pathogenassociated molecular pattern-triggered plant innate immunity. Plant Physiol 152:2222–2231

Lu S et al (2007) MicroRNAs in loblolly pine (Pinus taeda L.) and their association with fusiform rust gall development. Plant J 51:1077–1098

Melotto M et al (2006) Plant stomata function in innate immunity against bacterial

invasion. Cell 126:969–980

Millar AA, Gubler F (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell 17(3):705–721

Navarro L et al (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312(5772):436–439

Navarro L et al (2008) Suppression of the microRNA pathway by bacterial effector proteins. Science 321(5891):964–967

Padmanabhan C, Zhang X, Jin H (2009) Host small RNAs are big contributors to plant innate immunity. Curr Opin Plant Biol 12:465–472

Pruss GJ, Nester EW, Vance V (2008) Infiltration with Agrobacterium tumefaciens induces host defense and development-dependent responses in the infiltrated zone. Mol Plant Microbe Interact 21:1528–1538

Reyes JL, Chua NH (2007) ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. Plant J 49:592–606

Rubio-Somoza I et al (2009) Regulation and functional specialization of small RNA-target nodes during plant development. Curr Opin Plant Biol 12:622–627

Schommer C et al (2008) Control of jasmonate biosynthesis and senescence by miR319 targets. PLoS Biol 6(9):e230

Spoel SH, Dong XN (2008) Making sense of hormone crosstalk during plant immune responses. Cell Host Microbe 3(6):348–351

Spoel SH, Johnson JS, Dong X (2007) Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. Proc Natl Acad Sci USA 104:18842–18847

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

Sunkar R et al (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309

Tao Y et al (2003) Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae. Plant Cell 15:317–330

Trindade I et al (2010) miR398 and miR408 are up-regulated in response to water deficit in Medicago truncatula. Planta 231(3):705–716

Varkonyi-Gasic E et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:12

Wang D et al (2007) Salicylic acid inhibits pathogen growth in plants through repression of the auxin signaling pathway. Curr Biol 17:1784–1790

Zhang W et al (2010) Multiple distinct small RNAs originated from the same microRNA precursors. Genome Biol 11:R81

Figure 3.1 Distributions of lengths and the first nucleotides of small RNAs and miRNAs from the 13 small-RNA sequencing data.
a Length distribution of all sequencing reads (‘‘all’’) that have a 30 sequencing adaptor, are of high quality and can perfectly map to the Arabidopsis genome or cDNA sequences, and distribution of those reads in ‘‘all’’ that exclude miscellaneous noncoding RNAs, including tRNAs, rRNAs, snRNAs and snoRNAs (‘‘all no misc’’). Horizontal axis shows the sequence lengths and the vertical axis is the percent of sequences of a particular length.
b Distribution of the first nucleotides of ‘‘all’’ and ‘‘all no misc’’ reads in a, respectively.
c Length distribution of sequence reads that perfectly map to the known miRNAs.
d Distribution of the first nucleotides of reads mapped to the known miRNAs.

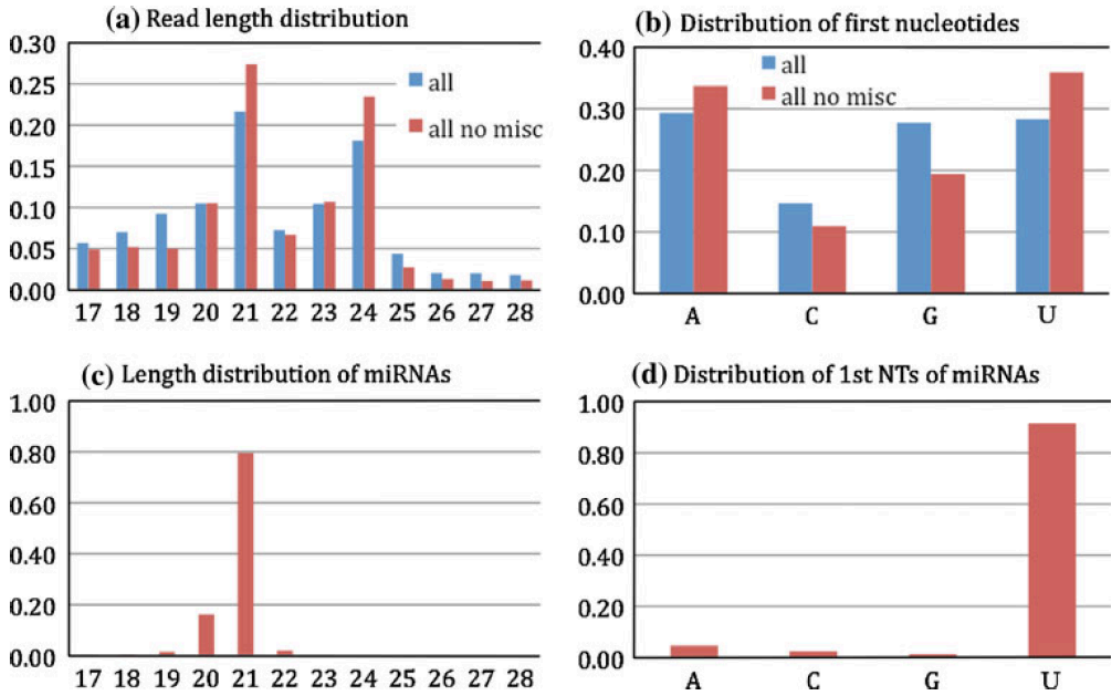


Figure 3.2 Correlation of miRNA abundances between a two technical duplicates for the *Pst* (*avrRpt2*) 14-hpi, and b two biological replicates for mock 14-hpi from Illumina high-throughput deep sequencing.

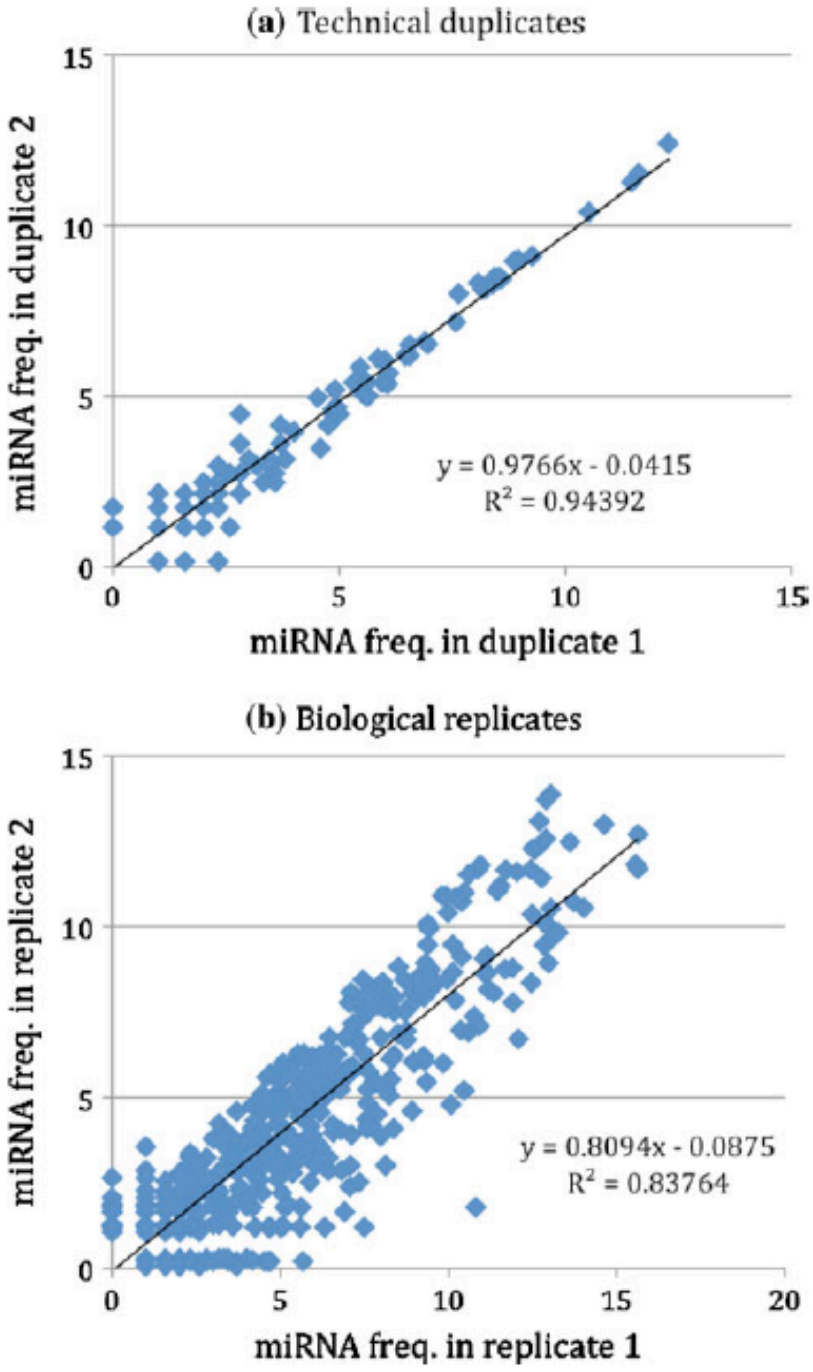


Figure 3.3 Differentially expressed miRNAs identified in three bacterial infections.
a type III secretion system mutated strain *Pst* DC3000 *hrcC*,
b virulent strain Pst DC3000 carrying an empty vector (EV), and c avirulent strain Pst DC3000 carrying effector avrRpt2, at 6- hpi and 14-hpi with respect to the respective mock treatments.
The bars shown in the figures are log (base 2) ratios of fold changes relative to the mock. miRNAs are sorted based on the fold changes at 14-hpi. FC stands for fold change

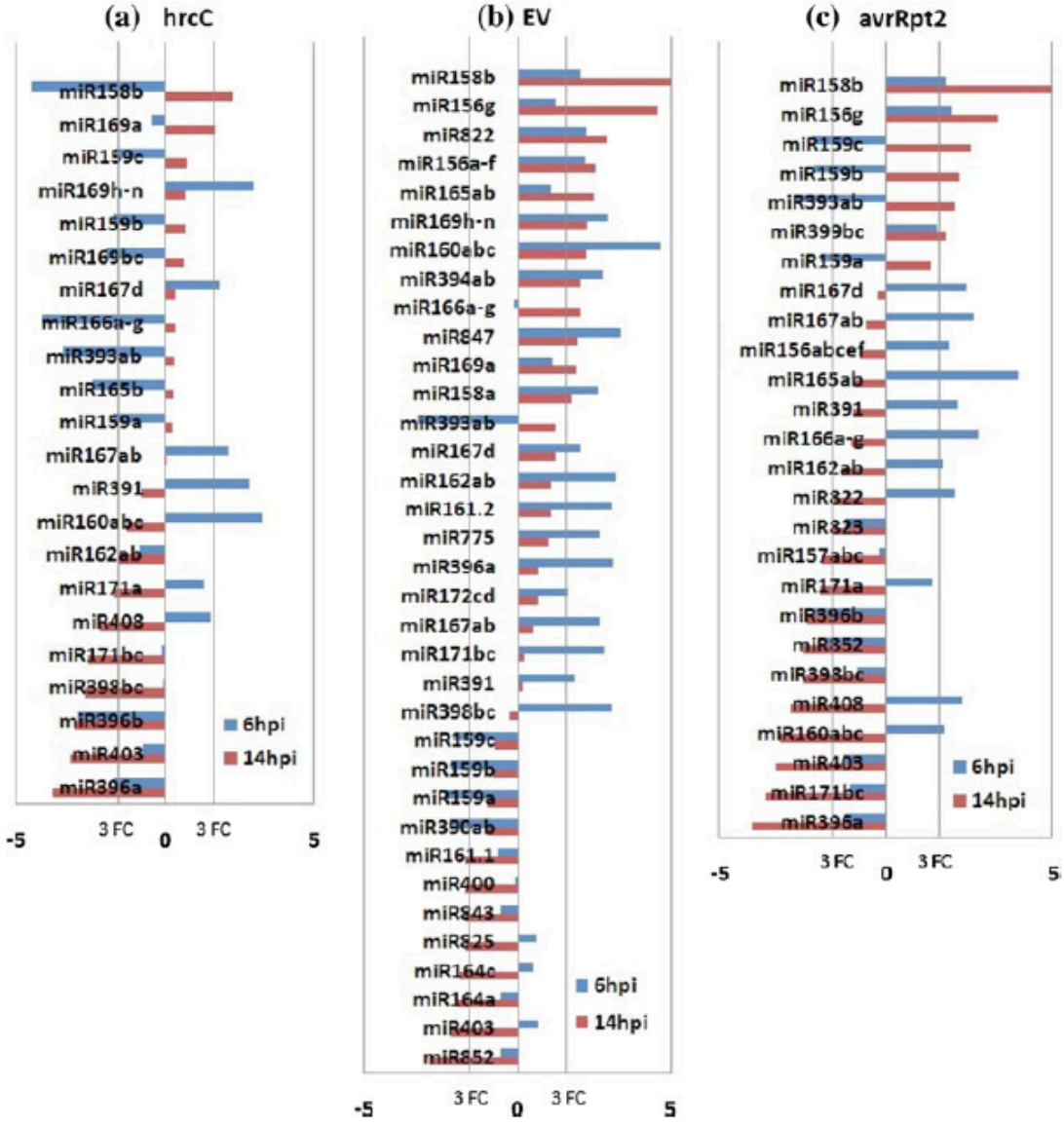


Figure 3.4 Small RNA Northern blot verification of some of the differentially expressed miRNAs detected by deep sequencing from the three bacterial pathogen infections and the mock controls at a 6-hpi and b 14-hpi. Similar results were obtained from two biological duplicates. U6 RNA was used as a control for measuring the relative amount of the bands (shown below each panel). Imagequant software version 2.1 was used for relative amount quantification (GE Healthcare Life Sciences)

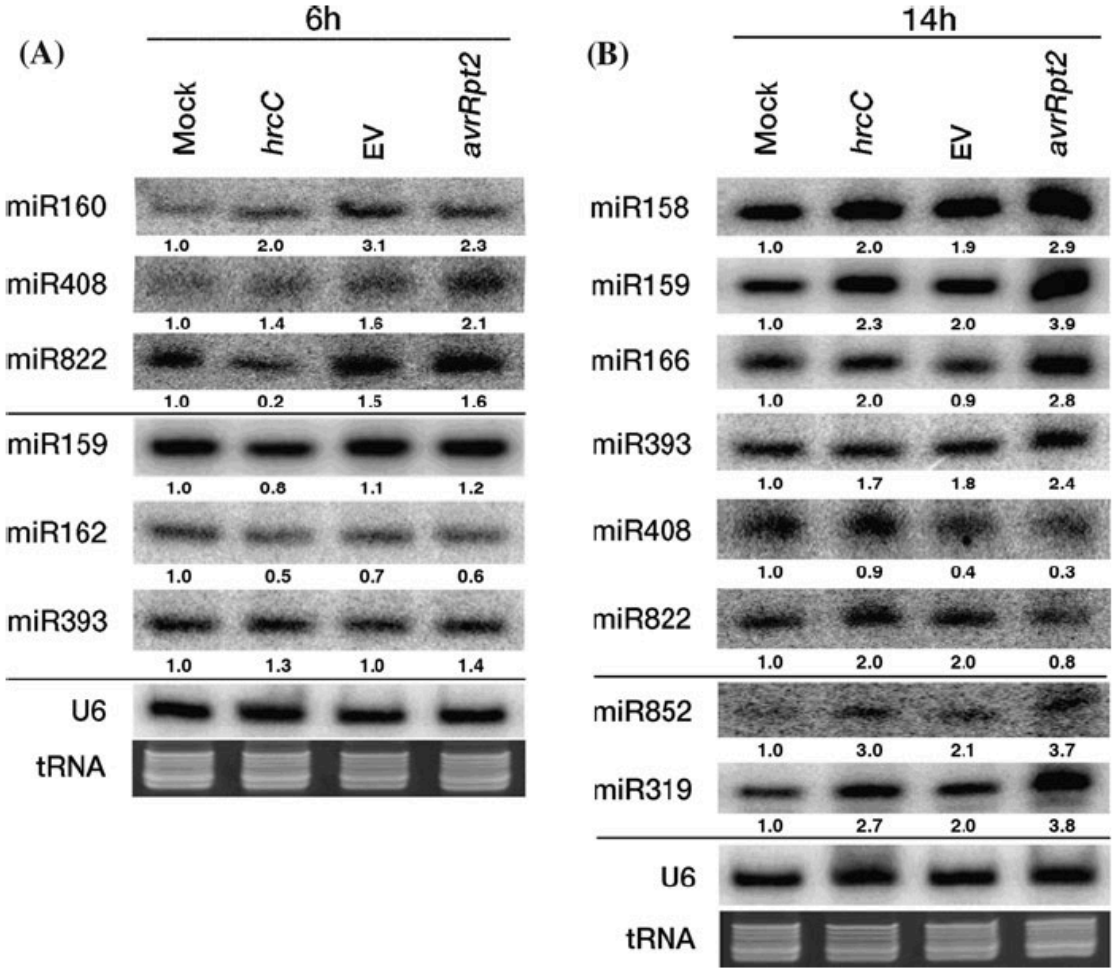


Figure 3.5 Relative expression levels of miR159, miR408 and miR319 in Pst DC3000 (EV), Pst DC3000 hrcC and Pst DC3000 (avrRpt2) compared to the mock by quantitative RTPCR. The comparative threshold cycle (Ct) method was used to determine relative transcript levels (Bulletin 5279, Real-time PCR applications Guide, Bio-Rad), using actin as an internal control. The expression level of a mock was set to 1, and the values for the other conditions are fold changes for those conditions. a miR159, b miR408, c miR319. Similar results were obtained from two biological replicates

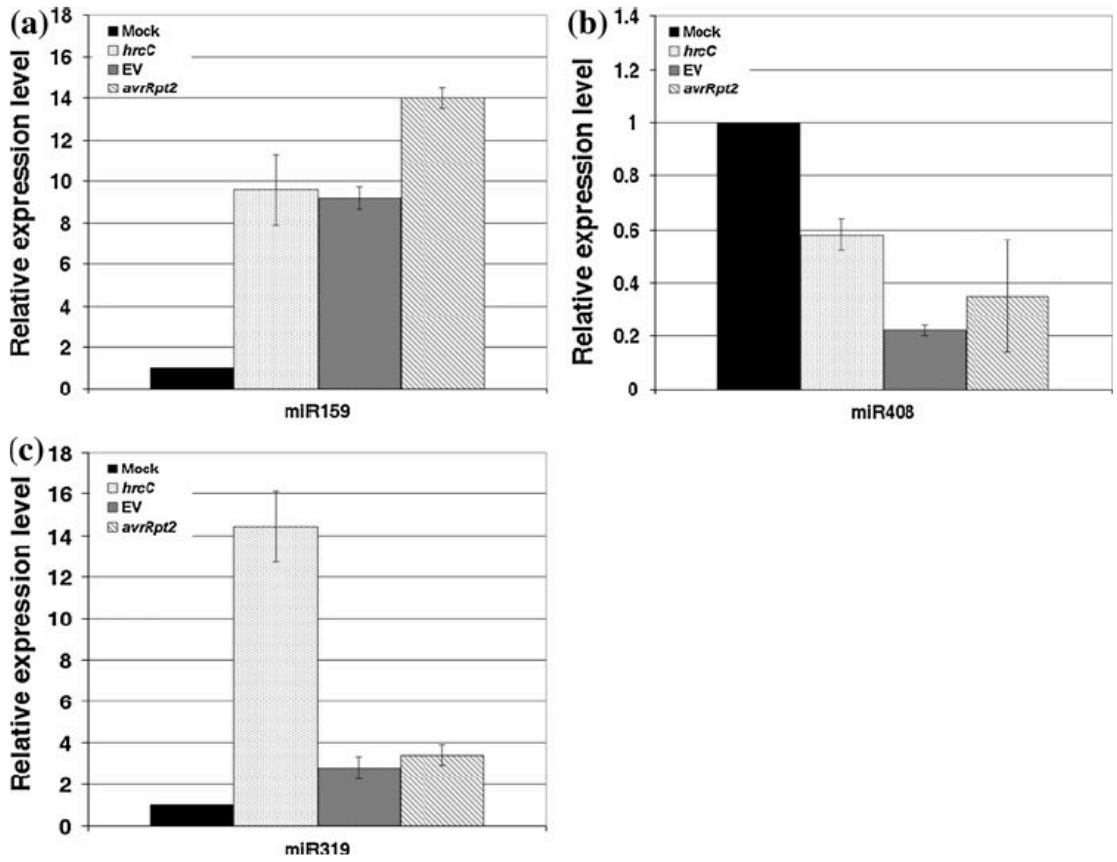
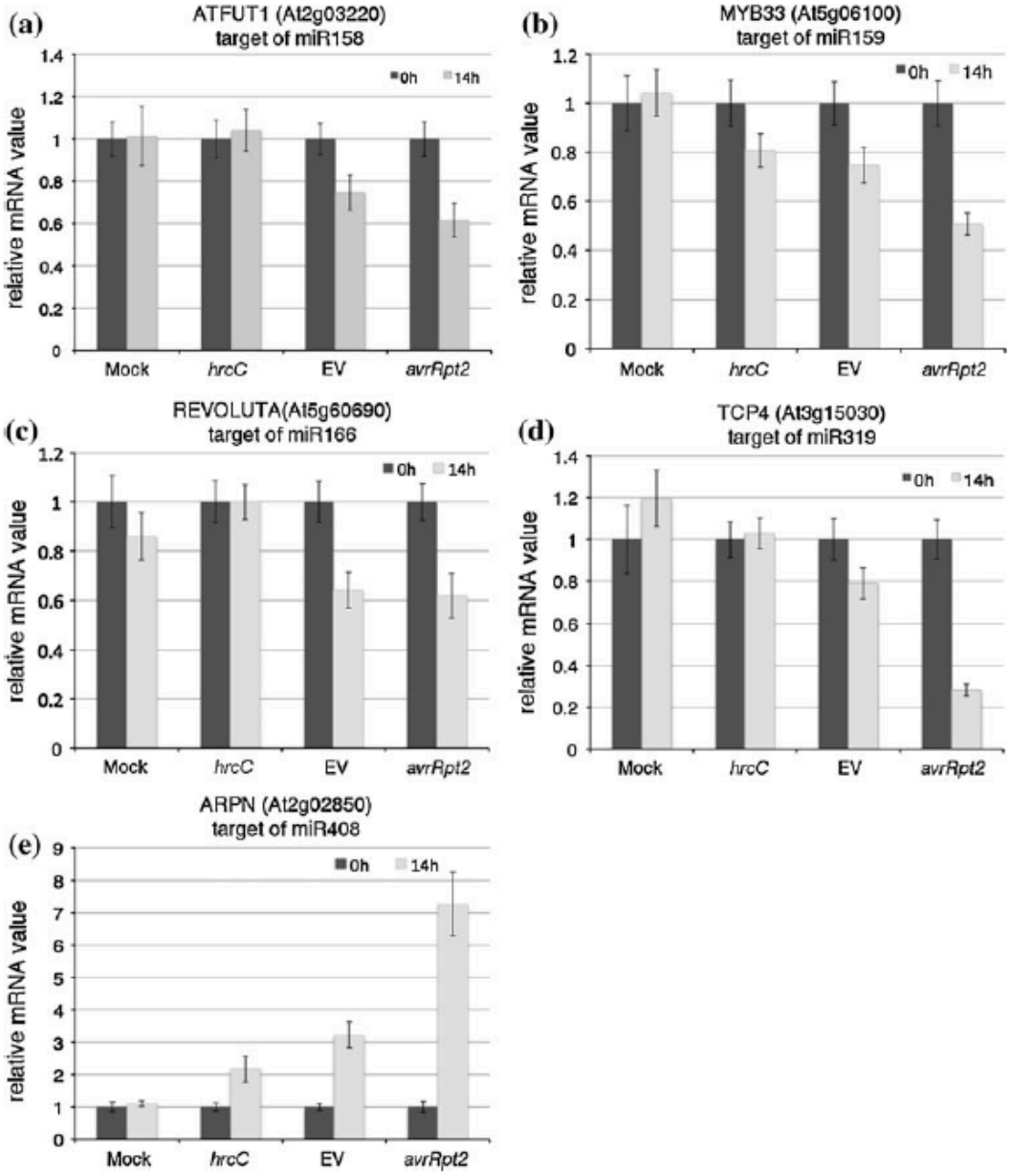


Figure 3.6 Expression analysis of miRNA target genes using quantitative real time RT-PCR. Relative expression levels of a At2g03220 (miR158 target), b At5g06100 (miR159 target), c At5g60690 (miR166 target) and d At3g15030 (miR319 target), e At2g02850 (miR408 target) at 14-hpi of Pst DC3000 hrcC, Pst DC3000 (EV), and Pst DC3000 (avrRpt2) infections as compared with the mock treatment. The comparative threshold cycle (Ct) method was used (Bulletin 5279, Real-time PCR applications Guide, Bio-Rad), with actin as an internal control. Similar results were obtained from two biological replicates.



Supplemental Table 3.1 miRNAs and their targets that have anti-correlated expression patterns after 6-hpi of (A) Pst DC3000 hrcC and (B) Pst DC3000 EV. Included here are miRNAs whose expression levels changed by at least 3 folds. miRNA expression was based on our deep sequencing profiling and mRNA expression data were from AtGenExpress accession number ME00331. Fold chnages are in log ratio (base 2).

(b) DC3000 EV at 6 hpi

miRNA	miRNA fold change vs Mock	Target gene	Targets fold change vs Mock	Gene Description
miR393ab	-3.289737	AT1G12820	0.066947	AFB3 (AUXIN SIGNALING F-BOX 3); auxin binding / ubiquitin-protein ligase
miR393ab	-3.289737	AT3G62980	1.116227	TIR1 (TRANSPORT INHIBITOR RESPONSE 1); auxin binding / protein binding / ubiquitin-protein ligase
miR393ab	-3.289737	AT4G03190	0.241918	GRH1 (GRR1-LIKE PROTEIN 1); auxin binding / protein binding / ubiquitin-protein ligase
miR159b	-2.180846	AT5G06100	0.209966	MYB33 (MYB DOMAIN PROTEIN 33); DNA binding / transcription factor
miR159b	-2.180846	AT3G11440	0.107287	MYB65 (MYB DOMAIN PROTEIN 65); DNA binding / transcription factor
miR171a	1.575249	AT2G45160	-1.058113	scarecrow transcription factor family protein
miR171a	1.575249	AT3G60630	-0.774517	scarecrow transcription factor family protein
miR402	1.625178	AT1G77230	-0.007632	tetratricopeptide repeat (TPR)-containing protein
miR167d	2.033289	AT1G30330	-0.244452	ARF6 (AUXIN RESPONSE FACTOR 6); transcription factor
miR158a	2.640428	AT1G64100	-0.225931	[AT1G64100, pentatricopeptide (PPR) repeat-containing protein];[AT1G64105, ANAC027 (Arabidopsis NAC domain containing protein 27); transcription factor]
miR167ab	2.655948	AT1G30330	-0.244452	ARF6 (AUXIN RESPONSE FACTOR 6); transcription factor
miR775	2.67785	AT1G53290	-0.310985	galactosyltransferase family protein
miR394ab	2.772362	AT1G27340	-0.591255	F-box family protein
miR171bc	2.779832	AT3G60630	-0.774517	scarecrow transcription factor family protein

miR161.2	3.050717	AT1G63330	-0.068122	[AT1G62590, pentatricopeptide (PPR) repeat-containing protein];[AT1G63330, pentatricopeptide (PPR) repeat-containing protein]
miR161.2	3.050717	AT1G06580	-0.011502	pentatricopeptide (PPR) repeat-containing protein
miR161.2	3.050717	AT1G62930	-0.064218	INVOLVED IN: biological_process unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: pentatricopeptide (PPR) repeat-containing protein (TAIR:AT1G63130.1); Has 29147 Blast hits to 6292 proteins in 192 species: Archae - 8; Bacteria - 22; Metazoa - 1258; Fungi - 858; Plants - 25291; Viruses - 0; Other Eukaryotes - 1710 (source: NCBI BLink).
miR161.2	3.050717	AT1G63130	-0.443663	pentatricopeptide (PPR) repeat-containing protein
miR161.2	3.050717	AT1G63080	-0.012654	pentatricopeptide (PPR) repeat-containing protein
miR161.2	3.050717	AT1G62860	-0.135915	pseudogene of pentatricopeptide (PPR) repeat-containing protein
miR161.2	3.050717	AT1G62590	-0.068122	[AT1G62590, pentatricopeptide (PPR) repeat-containing protein];[AT1G63330, pentatricopeptide (PPR) repeat-containing protein]
miR161.2	3.050717	AT1G64580	-0.036259	pentatricopeptide (PPR) repeat-containing protein
miR396a	3.102492	AT4G27180	-0.328723	ATK2 (ARABIDOPSIS THALIANA KINESIN 2); microtubule binding / microtubule motor
miR162ab	3.190719	AT1G01040	-0.162463	DCL1 (DICER-LIKE 1); ATP-dependent helicase/ double-stranded RNA binding / protein binding / ribonuclease III
miR847	3.32568	AT1G53340	-0.001588	DC1 domain-containing protein
miR160abc	4.660075	AT1G77850	-0.286314	ARF17 (AUXIN RESPONSE FACTOR 17); transcription factor

miR160abc	4.660075	AT4G30080	-0.346967	ARF16 (AUXIN RESPONSE FACTOR 16); miRNA binding / transcription factor
miR160abc	4.660075	AT2G28350	-0.941431	ARF10 (AUXIN RESPONSE FACTOR 10); miRNA binding / transcription factor

CHAPTER 4

Characterization of AtlsiRNA-1 target gene *AtRAP* and its RNA binding activities in *Arabidopsis thaliana*

ABSTRACT

AtlsiRNA-1 identified in *Arabidopsis* can be specifically induced by the bacterium *Pseudomonas syringae* DC3000 carrying the effector *avrRpt2* and silences its target *AtRAP* by mRNA degradation. *AtRAP* is a RAP-domain containing protein involved in disease resistance. Here we show that *AtRAP* is a single-copy gene in *Arabidopsis* and localizes in the chloroplast of the mesophyll cells. A mutation in *AtRAP* (*atrap-1*) showed reduced bacterial growth at 3 days post inoculation (dpi), while the plants overexpressing *AtRAP* showed increased susceptibility. I demonstrated that *AtRAP* could bind RNA molecules, while *AtRAP* without the C-terminal RAP domain lost its RNA binding activity. Taken together, our studies suggest that *AtRAP* is a negative regulator in *Arabidopsis* bacterial defense and the RNA binding activity is essential for its function.

INTRODUCTION

Small RNA pathways in Arabidopsis

There are at least four classes of endogenous siRNAs identified in *Arabidopsis*, including chromatin-associated siRNAs, trans-acting siRNAs (tasiRNAs), natural antisense transcript (NAT)-associated siRNAs (nat-siRNAs) and long siRNAs (lsiRNA). Small RNAs ranging from 30-40 nucleotides (nt) were identified as the fourth class of small RNA and mainly induced by bacterial infection or specific growth conditions. AtlsiRNA-1, as the first example, was strongly and specifically induced by *Pseudomonas syringae* pv. *tomato* (*Pst*) carrying effector *avrRpt2* (Katiyar-Agarwal et al., 2007). Unlike the dicer independent, 25-31 nt PIWI-interacting RNAs (piRNAs) identified in

animal germline cells, the biogenesis of AtlsiRNA-1 requires DCL1, HYL1, HEN1, HST1, AGO7, RDR6 and Pol IV, which is closely related to the known biogenesis pathways of siRNAs in plants. AtlsiRNA-1 is generated from *SRRLK/AtRAP* transcripts pair upon the induction and silenced *AtRAP* through decapping and XRN4-mediated 5'-to-3' degradation of the mRNA.

RAP domain containing proteins

The small RNA target gene At2g31890 was named *AtRAP*, because it encodes an expressed protein containing a RAP (RNA-binding domain abundant in Apicomplexans) domain at its C-terminus. RAP domains contain approximately 60 residues, consisting of multiple blocks of charged and aromatic residues. The domain is named RAP based on its inferred RNA-binding function.

The Apicomplexa are single celled lower eukaryotes and obligate intracellular protozoan parasites of animals. *Toxoplasma gondii*, responsible for the disease toxoplasmosis, has the broadest host range of the Apicomplexa as it infects virtually any warm-blooded vertebrate host. They are a large group of protists, most of which possess a unique organelle called an apicoplast, an apical complex structure involved in penetrating a host cell. Bioinformatic analysis revealed at least 15 unique RAP domain containing proteins in *Toxoplasma gondii* and 9 proteins in *Theileria annulata,* respectively.

Fas-activated serine/threonine phosphoprotein (FAST) is another example of a RAP protein found in *Homo sapiens*. Besides FAST, there are another 5 FAST kinase domain-containing proteins (FASTKD) that also share the FAST and RAP domains at their C-termini. Li *et al.* found that FAST is a survival protein and localizes at the outer mitochondrial membrane and the nucleus. Silencing of FAST by RNA interference will lead to apoptosis, while the overexpressed recombinant FAST inhibits Fas- and UV-induced apoptosis (Li et al., 2004). Further mutational analysis revealed that the N-terminus of FAST binds to the glutamine-rich carboxyl terminus of TIA-1 that promotes the expression of endogenous inhibitors of apoptosis, preventing stress-induced apoptosis. FAST, like TIA1, is also believed to function as a regulator of alternative

splicing of fibroblast growth factor receptor 2 (FGFR2) (Simarro et al., 2007), which promotes the inclusion of exon IIIb in an IAS1-dependent manner. Further domain deletion analysis reveals that the 188 amino acids (containing the RAP domain) at the C-terminus of FAST are important for the activation of exon IIIb, although they are dispensable for its survival activity. Recent results suggest FASTKD3 as an essential component of mitochondrial respiration that may modulate energy balance in cells exposed to adverse conditions by functionally coupling mitochondrial protein synthesis to respiration (Simarro et al., 2010).

Raa3 is another RAP domain containing protein that was identified in the green algae *Chlamydomonas reinhardtii.* *Raa3* is a nuclear gene and present in the soluble fraction of the chloroplast (Rivier et al., 2001). In *C. reinhardtii*, the chloroplast gene psaA consists of three exons and encodes a peptide of the photosystem I reaction center. Maturation of the psaA mRNA requires two steps of splicing *in trans* between the transcripts of exons 1, 2 and 3. The first psaA exon transcript was found in a 1,700 kDa complex, including tscA RNA and *Raa3*. *Raa3* was therefore believed to be involved in the first trans-splicing reaction. *Raa3* has 1,783 amino acid, contains an ActA domain, a PDX domain and one C-terminal RAP domain.

RESULTS

Molecular Characterization of *AtRAP*

From a previous study (Katiyar-Agarwal, et al. 2007), AtlsiRNA-1 is generated from the overlapping region of the NAT pair At2g31880 and At2g31890 upon the infection of *Pst* DC3000 (*avrRpt2*). AtlsiRNA-1 then triggers target gene (At2g31890) silencing by destabilizing its mRNA through decapping and XRN4-mediated 5'-to-3' degradation, rather than cleavage of target mRNA or translational inhibition. At2g31890 is considered an *AtRAP* because it encodes an expressed protein with a putative 59 aa RNA-binding domain at its C-terminal domain (Lee and Hong 2004). The RAP domain distributes across the last three exons among the five exons that are encoded (Fig. 4.1A).

The full CDS region is 2016 nt, and the gene product of *AtRAP* predicts a protein of 672 aa with approximate molecular masses of 75.7 kDa.

Lee *et al.* identified that the RAP domain protein was also encoded by the *Raa3* gene in *Chlamydomonas reinhardtii* (Cr) and by *OSJNBa0032H19.2* in *Oryza sativa* (Os) (Lee and Hong 2004). The RAP domain alignment was shown in Figure 4.1B.

Mutation in *AtRAP* shows retarded growth and a photobleached phenotype

To determine how the small RNA target gene, *AtRAP*, is involved in bacterial defense, we surveyed T-DNA insertions for *AtRAP* from the Salk Institute Genomic Analysis Laboratory Collection (Alonso et al., 2003) and chose the mutant alleles of *AtRAP*, SAIL_1223_C10.v1/CS_844807. T-DNA insertion was located in the middle of the third exon of *AtRAP*, downstream of the start codon (Fig. 4.1A). The homozygous *atrap-1* showed that no full-length transcripts are encoded using RT-PCR (Fig. 4.1C).

Compared to the Col-0 WT, *atrap-1* exhibits a growth-retarded phenotype (Fig. 4.2A), which starts after seeds germination. The cotyledons, rosette leaves and siliques (Fig. 4.2A and B) are pale green and smaller in size compared to the Col-0 WT and this photobleached phenotype is more obvious at younger stages during the first 4 weeks post-germination. Furthermore, *atrap-1* has a late flowering phenotype. Usually the Col-0 WT plants start to bolt at 4-5 weeks under the 12/12 hour light/dark cycle, while the *atrap-1* plants bolt about 2 weeks later than the wild type (Fig. 4.2C).

To uncouple the developmental phenotype from the disease resistance phenotype, we generated overexpression line in the *atrap* mutant background, which fully rescued the developmental phenotype. A full length *AtRAP* CDS region fused with a Flag epitope tag at the C-terminal was cloned under the control of the *Cauliflower mosaic virus* (CaMV) 35S promoter, then the resulting construct, 35S::*AtRAP*-Flag was transformed into the *atrap-1* mutant plants. All the T1 transgenic lines (*atrap-1*/35S::*AtRAP*-Flag) identified by PCR and western blot showed wild type phenotypes (Fig. 4.3A). In addition, western blotting with an anti-Flag antibody showed the presence of the *AtRAP* and Flag

fusion protein (Fig. 4.3B). These plants displayed enhanced susceptibility to *Pst* infection (see below).

AtRAP plays a negative role in the plant defense response against biotrophic pathogens

From the previous study, *AtRAP* mRNA can be degraded by the small RNA AtlsiRNA-1, immediately after the incompatible *Pst* (*avrRpt2*) infection, which suggests that *AtRAP* may play a role in basal defense. To further assess alterations of *atrap-1* in plant disease resistance, we infiltrated mutant plants with both virulent *Pst* (EV) and avirulent *Pst* (*avrRpt2*) strains of the bacterial pathogen *Pseudomonas syringae* DC3000, and measured the bacterial accumulation at 3 days post inoculation (dpi). Compared to the WT, *atrap-1* displayed enhanced resistance to both virulent and avirulent strains with less bacterial accumulation at 3 dpi (Katiyar-Agarwal, et al. 2007) (Fig. 4.4A). In addition to the reduction of pathogen growth in the *atrap-1* mutant, we also checked the hypersensitive response (HR) by inoculating the leaves with a high concentration (1×10^7 cfu/ ml) of *Pst* (*avrRpt2*). HR reaction usually takes place at 14 hpi in wild type plants, but an earlier HR (~8 hpi) was observed in *atrap-1* mutant plants (Fig. 4.4B). The early induction of HR can limit the bacterial growth more effectively compare to the wild type plant, which reflects the higher bacterial resistance levels of the *atrap-1* mutant. These results suggest that *AtRAP* negatively regulates the bacterial resistance response.

To further confirm these results, we also checked the alteration of the disease resistance by an *AtRAP* gain-of-function study, using overexpression lines. A full length *AtRAP* CDS region (without the UTR) fused with a Flag tag at the C-terminal region was cloned under the 35S promoter and then introduced into the mutant background. The transgenic lines did not have any visible phenotypes compare to WT (Fig. 4.3A). Two transgenic T1 lines (lines 7 and 10) with high and medium Flag expression levels, respectively, (Fig. 4.3B), were chosen and used for pathogen growth assays. In contrast to the *atrap-1* mutant results, both of the overexpression lines were more susceptible and exhibited decreased resistance to bacterial inoculation, both virulent and avirulent strains

(Fig. 4.4C). These collective observations indicate that *AtRAP* plays a negative role in plant defense against *Pst*, because the knockout and overexpression lines exhibited opposite reactions to both compatible and incompatible bacterial inoculation. The silencing of the *AtRAP* by atlsiRNA-1 could be considered a protection mechanism to promptly elevate the disease resistance in plants.

Subcellular Localization of *AtRAP*

Based on the AtGenExpress developmental series microarray data (http://www.weigelworld.org/resources/microarray/AtGenExpress/) (Schmid et al., 2005), *AtRAP* is predominantly expressed in photosynthetic tissues such as rosette leaves and in the vegetative shoot apex but are strongly repressed in roots, floral organs and later stages of seed development.

The subcellular localization of *AtRAP* was experimentally tested using yellow fluorescent protein (YFP) fused to the C-terminus of the *AtRAP* CDS region. The resulting DNA construct was subcloned into the pEarlyGate101 vector (Earley et al., 2006) downstream of the CaMV 35S promoter to give 35S::*AtRAP*-YFP. Since the C-terminal RAP domain is the only known domain with a putative RNA binding function, we also cloned the mutated *AtRAP* by deleting the RAP domain (residue 607 to 665), to produce 35S::*AtRAP*-Δ*RAP*-YFP. Both constructs were transiently expressed into *N. benthamiana* leaves using the empty vector as a control. The leaves were examined with a Leica SP2 confocal microscope at 48 hpi. The fluorescent images revealed that the transiently expressed full length *AtRAP* was localized exclusively in tobacco mesophyll chloroplasts characterized by their red autofluorescence (Fig. 4.5B), whereas the control 35S::YFP was detected everywhere (Fig. 4.5A). The RAP domain deleted version of *AtRAP* showed the same chloroplast localization of the fusion protein (Fig. 4.5C) as with intact RAP, which suggests that the C-terminal RAP domain is not involved in its subcellular localization.

To verify the *AtRAP* subcellular localization by using an independent experimental approach, the fluorescent signal were also checked in the *Arabidopsis*

transgenic plants. The native promoter region and genomic DNA region of *AtRAP* were fused to the GFP reporter gene by using the gateway destination vector pMDC107 (Curtis and Grossniklaus 2003), resulting in AtRAP::*AtRAP*-GFP, and the RAP domain deletion from *AtRAP*, AtRAP::*AtRAP*-Δ*RAP*-GFP, was also generated in the same process. Both constructs were transformed into *Arabidopsis* mutant *atrap-1* plants. The fluorescent images from T1 generation transgenic plants confirmed that both of these fused proteins were located in the mesophyll chloroplasts, although the fluorescent signal was relatively weaker compared to the 35S promoter constructs (Fig. 4.5D and E).

Collectively, these data demonstrate that *AtRAP* is localized in the mesophyll chloroplasts and the deletion of the C-terminal RAP domain does not interfere with the protein localization.

The C-terminal RAP domain is responsible for RNA binding and function

The construct with a native *AtRAP* promoter without the C-terminal RAP domain, AtRAP::*AtRAP*-Δ*RAP*-GFP, shows chloroplast localization in the *atrap-1* mutant background. However, these transgenic plants displayed the same photobleached phenotype as the knockout background (Fig. 4.6A). In contrast, expression of the wild type *AtRAP* in the mutant background can efficiently complement the *atrap-1* phenotypes under control of the same native *AtRAP* promoter (Fig. 4.6A). Thus, this phenomenon suggests that the 59 aa RAP domain (607-665 residue) is essential for *AtRAP* functionality.

Due to the proposed RNA binding activity of the C-terminal RAP domain of *AtRAP*, we expected to detect the RNA molecules by 5' end labeling with 32P-γ-ATP, in co-immunoprecipitation with the *AtRAP* fused with the GFP tag. Transgenic plants with either full length (*AtRAP*-GFP) or RAP domain deleted (*AtRAP*-Δ*RAP*-GFP) constructs were used to immunoprecipitate the protein (Fig. 4.6B). The RNA molecules bound to the protein was co-immunoprecipitated using an affinity gel system. Corresponding to our expectations, RNA molecules were detected as a strong autoradiography signal after

the radio-isotope labeled samples were electrophoresed through a 10% acrylamide-bis gel (Fig. 4.6C, Lane 5). The RNA was responsible for the signal rather than DNA, because the radioactive signal disappeared after the RNA sample was treated with RNaseA before the labeling (Fig. 4.6C, Lane 6). However, the RAP domain deletion construct shows much less RNA associated with it (Fig. 4.6C, Lane 1). These results suggest that the C-terminal RAP domain is vital for protein function; it contributes to the complementation of the mutant phenotype, disease defense and the *in vivo* RNA binding activity, although it is not involved in the protein localization.

Microarray data shows genes down/up-regulated in the *atrap-1* mutant are chloroplast associated

We profiled changes in gene expression in the *atrap-1* mutant at the mRNA level by using Affymetrix ATH1 microarrays. Three biological replicates were performed using 3-week-old rosette leaves from wild type and the *atrap-1* mutant plants. A total of 21 transcripts were down-regulated and 8 transcripts were up-regulated in the *atrap-1* mutant plants (Table 4.2 and 4.3). Seven out of the 21 down-regulated transcripts were chloroplast-associated genes.

Discussion

The subcellular localization of *AtRAP*

The nucleus, chloroplast and mitochondria are the three compartments that encode the genetic information in plants. With endosymbiotic origins, chloroplasts are believed to be derived from a cyanobacterium ancestor. Currently, chloroplasts only contain a relatively few protein-coding genes, between 87 and 183 known genes (Sugiura et al., 1998), because most of the prokaryotic genome has been transferred to the eukaryotic nuclear genome of the host cell during the long evolutionary process. Thus, the large number of nucleus-encoded proteins have to be synthesized in the cytoplasm and post-translationally imported back into the chloroplast (Leister 2005) to carry out

their function. Results from experiments using both the transient expression in tobacco and the transgenic *Arabidopsis* lines showed that, *AtRAP* localizes in the chloroplast of the mesophyll cells. Many chloroplast-localized proteins contain an N-terminal chloroplast transit peptide (cTP), which facilitates their translocation from the cytoplasm where they are synthesized to the chloroplast organelle (Peltier et al., 2000). However, the N-terminal regions of *AtRAP* does not display any sequence motifs that could be identified as putative chloroplast transit peptides by using the ChloroP 1.1 Server (http://www.cbs.dtu.dk/services/ChloroP/) (Emanuelsson et al., 1999) and TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2000). There are two classes of non-canonical chloroplast proteins identified so far (Armbruster et al., 2009). One is the chloroplast inner-envelope type and the other is an endoplasmic reticulum (ER)-dependent type. Tic32/IEP32 (Nada and Soll 2004) and ceQORH (chloroplast envelope Quinone OxidoReductase Homologue) (Miras et al., 2002; Ferro et al., 2003) are two examples of cp inner-envelope proteins that lack N-terminal cleavable presequences. *Arabidopsis* carbonic anhydrase 1 (CAH1) localizes in the chloroplast stroma and has a predicted signal peptide (SP) for ER translocation and is an example of the second type of non-canonical chloroplast proteins (Villarejo et al., 2005). The CAH1 was believed to translocate to the chloroplast through an ancestral secretory pathway. We did not find any homology between the *AtRAP* protein sequences with these non-canonical chloroplast proteins.

Analysis of *AtRAP* using the transmembrane prediction program TMpred (http://www.ch.embnet.org/software/TMPRED_forum.html) showed that *AtRAP* is a membrane-bound protein with 2 membrane-spanning domains located at residues 32 to 48 (outside to inside) and 320 to 339 (inside to outside). These two domains were also predicted in rice OSJNBa0032H19.2 at residues 10 to 36 (outside to inside) and 286 to 305 (inside to outside) (Fig. 4.7).

Further analysis is required to determine whether the *AtRAP* is localized in the chloroplast membrane, thylakoid membrane or stroma.

Function of RAP domain

The *AtRAP* protein was named because of the predicted C-terminal RNA binding domain. RNAs extracted from the full length *AtRAP* protein suggests its RNA binding activity. We failed to detect RNA from the RAP domain deleted *AtRAP* protein, which indicates that the C-terminal domain is indispensable for this RNA binding ability.

In addition, expression of the C-terminal RAP domain deleted *AtRAP* cannot complement the *atrap-1* mutant phenotype, although the deletion of the RAP domain does not change the localization of fluorescent protein. Moreover, the overexpression of the mutated *AtRAP* did not alter the plant disease resistance in the pathogen growth assay. These results suggest that the truncated *AtRAP* can localize at the same position as the wild type *AtRAP*, but its biological function is lost due to the missing RAP domain. Thus, without the RAP domain, *AtRAP* cannot complement the mutant phenotype, cannot bind RNA and cannot interfere with the disease resistance phenotype.

Microarray annotation of Genes that might be involved in the same biological function

The microarray results presented here indicate that 21 transcripts were down-regulated in the *atrap-1* mutant. This co-repressed expression in the mutant background suggests *AtRAP* might affect the same biological processes as these down-regulated genes.

Four genes out of 21 have been annotated to be in the chloroplast, including ABC1, a protein kinase superfamily protein (At5g05200), a dehydrin family protein (At1g54410), an aspartate-glutamate racemase family (At1g15410) and a NAD-dependent formate dehydrogenase (At5g14780). Interestingly, At1g54410 and At5g14780 display strong stress responses under cold, osmostic, salt, dry, UV, wound and heat treatment from the *Arabidopsis thaliana trans*-factor and *cis*-element prediction database (ATTED-II), which are quite similar to the responses of *AtRAP*.

Another down-regulated transcript, *AtGLK1* (At2g20570), encodes GOLDEN2-LIKE (GLK) transcription factors that are members of the GARP superfamily. Another

functional redundant gene, *AtGLK2*, has a highly conserved putative DNA-binding domain similar to *AtGLK1*. Interestingly, the double mutant plants (Atglk1.1/ Atglk2.1) with pale green coloration in all photosynthetic tissues have a phenotype quite similar to the *atrap-1* mutant plants. The gene pairs are believed to be required for chloroplast development (Fitter et al., 2002) because a reduction in granal thylakoids was observed in the double mutant. Moreover, the induction of AtGLK expression results in the enrichment of genes involved in chlorophyll biosynthesis (Waters et al., 2009). A recent study showed that BES1, a well-characterized transcription factor in the brassinosteroids (BRs) signaling pathway, could repress *AtGLK1* and *AtGLK2*, thereby inhibiting chloroplast biogenesis and/or function (Yu et al., 2011). BRs, as a class of phytohormones, function as signaling molecules in plants. The role of BRs is not only implicated in plant growth regulation (Azpiroz et al., 1998; Clouse and Sasse 1998), but also in alleviating biotic challenges of bacterial, fungal and viral pathogens (Krishna 2003; Nakashita et al., 2003). Being well defended may not always be the best strategy, so plants are able to quickly switch resources into the costly defense by diverting their resources away from growth and reproduction (Coley et al., 1985; Herms and Mattson 1992; Kempel et al., 2011). So BRs might elevate the plant disease resistance though a series of transcriptional repressions to reduce chloroplast biogenesis. The down-regulation of *AtGLK1* in the *atrap-1* plants is consistent with these evolutionary plant defense/fitness trade-offs, and this phenomenon is reminiscent of the PAMP-induced miR393 that contributes to plant resistance against *P. syringae* by repressing auxin signaling (Navarro et al., 2006; Voinnet 2008).

Only 7 transcripts were up-regulated in the *atrap-1* mutant plants. Three genes out of the 7 have been annotated to be located in the chloroplast, including *MRL1* (At4g34830), a FAD/NAD(P)-binding oxidoreductase family protein (At1g57770) and one chloroplast encoded DNA-directed RNA polymerase protein (AtCG00170). The *MRL1*, a regulator of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, contributes to the stability and/or maturation of the *rbcL* transcript in plants and green algae (Johnson et al., 2010). The gene At1g57770, was

reported to be involved in isoprenoid and chlorophyll metabolism in *Arabidopsis* (Lange and Ghassemian 2003). Meanwhile, this gene is positively regulated by HY5, a high hierarchical regulator of the transcriptional cascades for photomorphogenesis (Lee J. et al., 2007).

MATERIALS AND METHODS

Plant Materials

Arabidopsis plants were grown at 22°C under 12 h light per day. All mutants are in a Col-0 background. The T-DNA knock out line of *AtRAP* was obtained from the Salk Institute Genomic Analysis Laboratory.

Genotyping

For *atrap-1* genotyping, genomic DNA was amplified by Sail_1223_CIO-LP (5'-CGGAAAAGAACACTTGATTAACAAC-3'), Sail_1223_CIO-RP (5'-TGAGTT GTCAAAGAGGGCATC-3') and Sail LB3 (5'-GAATTTCATAACCAATCTCGATAC-AC-3') primers.

RT-PCR

Total RNA from leaves were reverse-transcribed using SuperScriptII (Invitrogen) and oligo-d(T) primers according to the manufacturers' instructions. Primers used are listed in Table 4.1.

Plasmid Construction

For generating Flag tagged *AtRAP* and RAP domain deletion expression lines, full-length *AtRAP* cDNA without the 3' UTR was amplified with the forward primer 5'-CACCATGGAGTGTGTAGTTCCATT-3' and reverse full length primer 5'-

TATGCAGCCGGTGAGAATCTCCCTCAA-3' or reverse RAP domain deletion primer 5'-TATGCAGCCGGTGAGAATTTTCTTCTCGACCAGA-3'. The fragments were cloned into the plant expression GATEWAY destination vector p35SGATFH with C-terminal Flag tag.

For generating YFP tagged *AtRAP* and RAP domain deletion expressing vectors, primers were used as the previously stated. Fragments were cloned into the plant expression vector pEarlyGate101 (Earley, et al. 2006). The plasmids 35S::*AtRAP*-YFP and 35S::*AtRAP*-Δ*RAP*-YFP transformed into *Agrobacterium tumefaciens* strain GV3101, then used for transient expression analysis in *N. benthamiana*.

For generating native promoter driven and GFP tagged *AtRAP* and RAP domain deletion expressing vectors, the *AtRAP* promoter and genomic DNA was amplified by Phusion® High-Fidelity DNA polymerase with cacc90PRMTR-F (5'-CACCGGG-CTAAGCCCATTAATACTAGTG-3') and 31890R-nostop (5'-TATGCAGCCGGTGA-GAATCTCCCT-3') primers and cloned into the entry vector, pENTR/D-TOPO (Invitrogen), which gives AtRAP::*AtRAP*-GFP. For AtRAP::*AtRAP*-Δ*RAP*-*GFP* plasmid construction, primers cacc90PRMTR-F (5'-CACCGGGCTAAGCCCATTAATACTA-GTG-3') and 31890R-ΔRAP-nostop (5'-TATGCAGCCGGTGAGAATTTTCTTCTCG-ACCAGA-3') were used. Then, the entry clones were introduced into the destination GFP6 fusion vector *pMDC107* (Curtis and Grossniklaus 2003) by LR reactions. These two constructs were transformed in the *A. tumefaciens* strain GV3101. *Arabidopsis* wild type Columbia or *atrap-1* plants were then transformed using the flower-dipping method. For selection of transgenic plants, T1 seedlings were selected for hygromycin resistance with Murashige-Skoog medium supplemented with 20 μg/ml hygromycin B.

Genomic DNA Extraction

One leaflet was ground with a disposable, plastic pestle, and 400 μl of extraction buffer (0.2M Tris-HCl (pH 7.5), 250 mM NaCl, 25 mM EDTA (pH 8.0), 0.5% SDS) was added and vortexed, the microcentrifuge tube was centrifuged at full speed in a

microcentrifuge for 10 min and the supernatant was transferred to a fresh tube, mixed with 300 µl of 100% isoproponal, and kept at -20°C for 30min. The tube was centrifuged for 10min at 13,000 rpm, the supernatant was discarded and the pellet was washed with 70% of Ethanol, and was dried for 5 min, then was dissolved in 50 µl TE buffer and 2 µl was used as template for genotyping PCR.

ATH1 Affymetrix Microarray Analysis

GeneChip arrays were hybridized according to the manufacturer's instructions (Affymetrix). Data analysis was done according to Horan et al. (2008). Normalization of raw intensities across all probe sets was performed in R using RMA algorithms. To calculate P-values for increases or decreases in expression, the Wilcoxon signed-rank test was applied to each pair of chips after normalization using the R&Bio Conductor. A P-value of <=0.05 in combination with a two-fold difference was used to define changes in gene expression.

Immunoprecipitation of *AtRAP*-Flag Fusion Protein

Approximately 1 gram of Arabidopsis tissue was ground into a fine powder with a mortar and pestle using liquid nitrogen. Ground tissue was homogenized with 5 ml/g of extraction buffer (20mM Tris/HCl (pH7.5), 300mM NaCl, 5mM $MgCl_2$, 0.5% Triton X-100, 5mM Dithiothreitol (DTT), 1 tablet/10mL Protein Inhibitor cocktail (Roche)) (Qi et al., 2005) and incubated for 10 min on ice. Cell debris was removed by centrifugation at 16,000g at 4°C for 10 min, and the supernatant was passed through the 100 µM cell strainer (BD Falcon) to further purify. Five milliliters of precleared extracts were aliquoted into 5 microcentrifuge tubes, then 20 µl of GFP antibody and 20 µl of Protein A Agarose (Roche Applied Science) was added to each tube, and all 5 tubes were rotated at 4°C for 6 hours using revolver rotator tube mixer (Denville Scientific, Inc). Immunoprecipitates were washed three times (10 min each) with extraction buffer. After combining all the beads from each tube (100 µl total), 10 µl was used for western blotting and the rest was used for RNA extraction.

Radioisotope labled RNA samples after Immunoprecipitation

RNA was extracted from GFP antibody and Protein A Agarose (Roche Applied Science) after immunoprecipitation by using TRIzol® Reagent (Invitrogen) and dissolved in 120 µl DEPC treated water. RNA (40 µl) was then treated with calf intestine Alkaline phosphatase (CIAP) (Invitrogen) for 30 min at 37°C to remove 5'-phosphate groups prior to 5'-end labeling of nucleic acids with T4 polynucleotide kinase, and the CIAP was deactivated the by phenol:chloroform:isoamyl alcohol (25:24:1). Purified RNA samples were aliquoted into two tubes and one tube was digested by RNaseA for 30 min at 37°C. Two sets of RNA samples were labeled with T4 polynucleotide kinase (NEB) with (γ-32P) ATP followed by purification using illustra MicroSpin™ G-25 Columns (GE Healthcare). RNA samples were then electrophoresed in polyacrylamide gels.

REFERENCES

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., 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, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301:653-657

Armbruster, U., Hertle, A., Makarenko, E., Zuhlke, J., Pribil, M., Dietzmann, A., Schliebner, I., Aseeva, E., Fenino, E., Scharfenberg, M., Voigt, C., Leister, D. (2009) Chloroplast proteins without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome? Mol Plant 2:1325-1335

Azpiroz, R., Wu, Y., LoCascio, J.C., Feldmann, K.A. (1998) An Arabidopsis brassinosteroid-dependent mutant is blocked in cell elongation. Plant Cell 10:219-230

Clouse, S.D., Sasse, J.M. (1998) BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. Annu Rev Plant Physiol Plant Mol Biol 49:427-451

Coley, P.D., Bryant, J.P., Chapin, F.S., 3rd (1985) Resource availability and plant antiherbivore defense. Science 230:895-899

Curtis, M.D., Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133:462-469

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. Plant J 45:616-629

Emanuelsson, O., Nielsen, H., von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978-984

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005-1016

Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., Rolland, N. (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. Mol Cell Proteomics 2:325-345

Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., Langdale, J.A. (2002) GLK gene pairs regulate chloroplast development in diverse plant species. Plant J 31:713-727

Herms, D.A., Mattson, W.J. (1992) The Dilemma of Plants: To Grow or Defend. The Quarterly Review of Biology 67:283-335

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D.B., Wollman, F.A., Vallon, O. (2010) MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. Plant Cell 22:234-248

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., Jin, H. (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev 21:3123-3134

Kempel, A., Schadler, M., Chrobock, T., Fischer, M., van Kleunen, M. (2011) Tradeoffs associated with constitutive and induced plant resistance against herbivory. Proc Natl Acad Sci U S A 108:5685-5689

Krishna, P. (2003) Brassinosteroid-Mediated Stress Responses. J Plant Growth Regul 22:289-297

Lange, B.M., Ghassemian, M. (2003) Genome organization in Arabidopsis thaliana: a survey for genes involved in isoprenoid and chlorophyll metabolism. Plant Mol Biol 51:925-948

Lee, I., Hong, W. (2004) RAP--a putative RNA-binding domain. Trends Biochem Sci 29:567-570

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., Deng, X.W. (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell 19:731-749

Leister, D. (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in Arabidopsis. Gene 354:110-116

Li, W., Simarro, M., Kedersha, N., Anderson, P. (2004) FAST is a survival protein that senses mitochondrial stress and modulates TIA-1-regulated changes in protein expression. Mol Cell Biol 24:10718-10732

Miras, S., Salvi, D., Ferro, M., Grunwald, D., Garin, J., Joyard, J., Rolland, N. (2002) Non-canonical transit peptide for import into the chloroplast. J Biol Chem 277:47770-47778

Nada, A., Soll, J. (2004) Inner envelope protein 32 is imported into chloroplasts by a novel pathway. J Cell Sci 117:3975-3982

Nakashita, H., Yasuda, M., Nitta, T., Asami, T., Fujioka, S., Arai, Y., Sekimata, K., Takatsuto, S., Yamaguchi, I., Yoshida, S. (2003) Brassinosteroid functions in a broad range of disease resistance in tobacco and rice. Plant J 33:887-898

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., Jones, J.D. (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436-439

Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., van Wijk, K.J. (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12:319-341

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

Rivier, C., Goldschmidt-Clermont, M., Rochaix, J.D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in Chlamydomonas reinhardtii. EMBO J 20:1765-1773

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., Lohmann, J.U. (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37:501-506

Simarro, M., Mauger, D., Rhee, K., Pujana, M.A., Kedersha, N.L., Yamasaki, S., Cusick, M.E., Vidal, M., Garcia-Blanco, M.A., Anderson, P. (2007) Fas-activated serine/threonine phosphoprotein (FAST) is a regulator of alternative splicing. Proc Natl Acad Sci U S A 104:11370-11375

Simarro, M., Gimenez-Cassina, A., Kedersha, N., Lazaro, J.B., Adelmant, G.O., Marto, J.A., Rhee, K., Tisdale, S., Danial, N., Benarafa, C., Orduna, A., Anderson, P. (2010)

Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. Biochem Biophys Res Commun 401:440-446

Sugiura, M., Hirose, T., Sugita, M. (1998) Evolution and mechanism of translation in chloroplasts. Annu Rev Genet 32:437-459

Villarejo, A., Buren, S., Larsson, S., Dejardin, A., Monne, M., Rudhe, C., Karlsson, J., Jansson, S., Lerouge, P., Rolland, N., von Heijne, G., Grebe, M., Bako, L., Samuelsson, G. (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat Cell Biol 7:1224-1231

Voinnet, O. (2008) Post-transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility. Curr Opin Plant Biol 11:464-470

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., Langdale, J.A. (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. Plant Cell 21:1109-1128

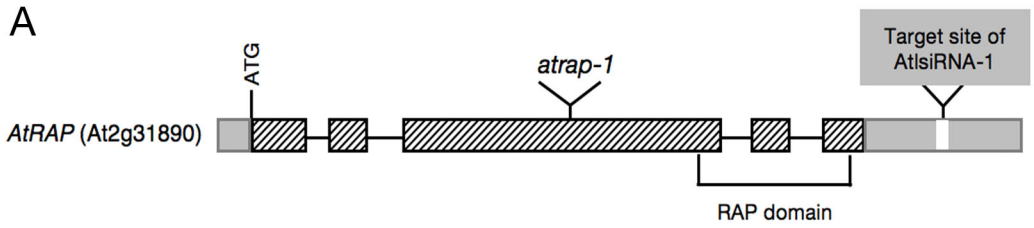
Yu, X., Li, L., Zola, J., Aluru, M., Ye, H., Foudree, A., Guo, H., Anderson, S., Aluru, S., Liu, P., Rodermel, S., Yin, Y. (2011) A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in Arabidopsis thaliana. Plant J 65:634-646

Figure 4.1 Structure of the AtRAP gene and similarity between *AtRAP* and other RAP domains containing homologs in eukartyotes

(A) A schematic diagram of the *AtRAP* gene showing the RAP domain in the 3-5 the exon

(B) RAP domain sequences are aligned with *OSJNBa0032H19.2* from rice and *Raa3* from chlamodomonas.

(C) Expression of *AtRAP* in wild type and *atrap-1* (CS_844807).



B

```
AtRAP                VALEIDGPTHFSRNSGLPL-GHTMLKRRYVAAAGWKVVSLSLQEWEEHEGSHEQLEYLRE 59
OSJNBa0032H19.2      LAFEIDGPSHFSRNLGTPL-GHTAFKRRYIAAAGWNLVSLSHQEWENLEGEFEQLEYLRR 59
Raa3                 LAVGAAAGGAVIRNSRWLLSGAGALRRRLLTHAGWLVVPVRERQWKDLRSAEQQRRVVRE 60
                     :*.   .   . **    * *   ::** :: *** :*.:  ::*:: ..  :* . :*.
```

C

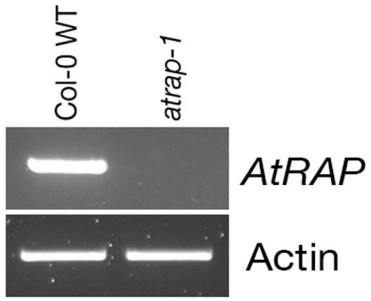


Figure 4.2 Phenotypes of atrap-1 mutant and the overexpression lines

(A) 4-week old Col-0 WT, *atrap-1* mutant and *AtRAP* overexpression line in the wild type background under the control of 35S promoter; *atrap-1* mutant shows growth-retarded and photobleached phenotype.

(B) siliques of Col-0 WT and *atrap-1* mutant from 8-week old plants

(C) 6-week-old Col-0 WT and *atrap-1* mutant, mutant plant shows late flowering phenotype

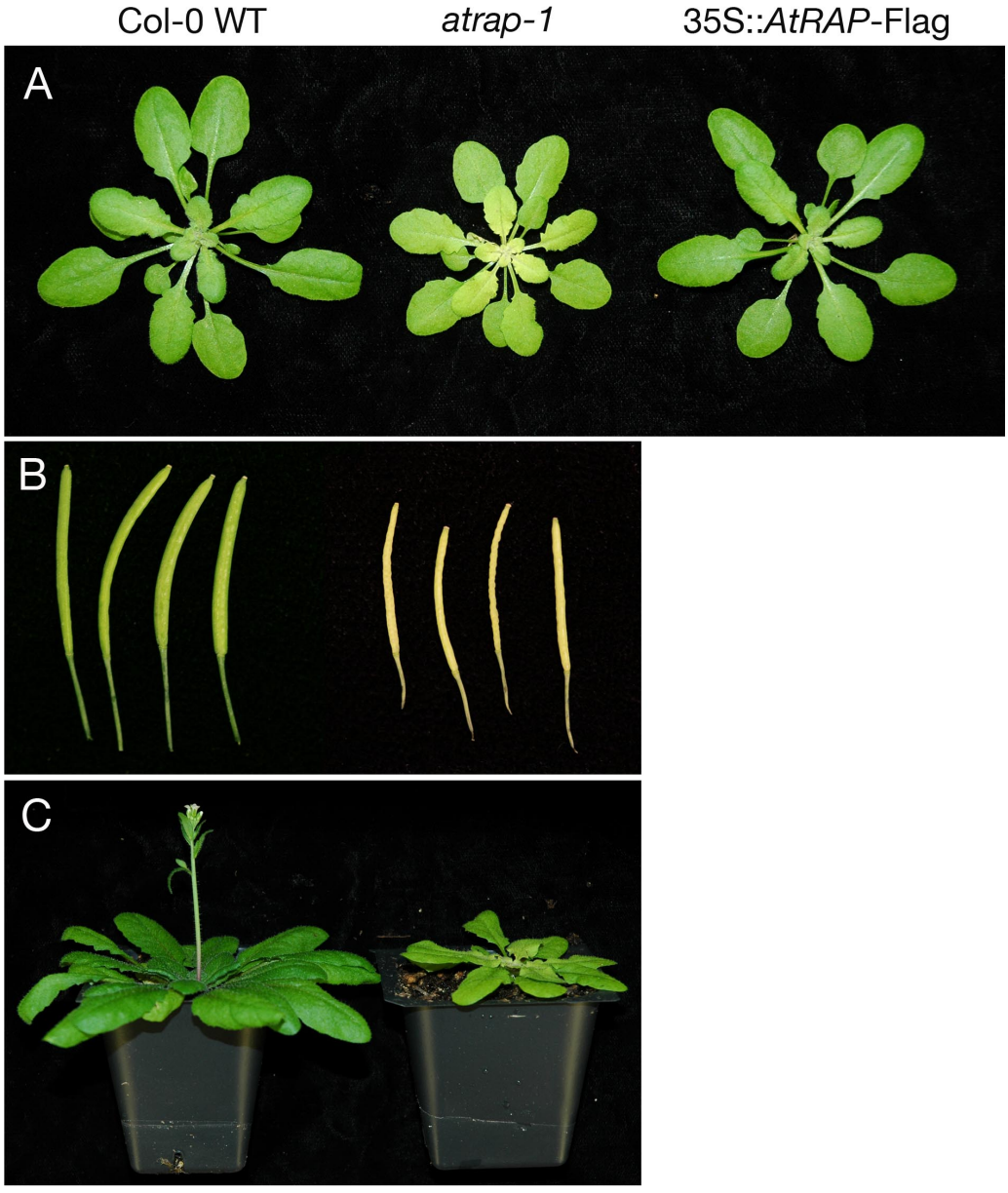

Col-0 WT
atrap-1
35S::AtRAP-Flag
A
B
C

Figure 4.3 Complementation of the atrap-1 mutant phenotype.

(A) Recovery of the phenotype in the complementary lines

(B) Western blotting of the complementary line

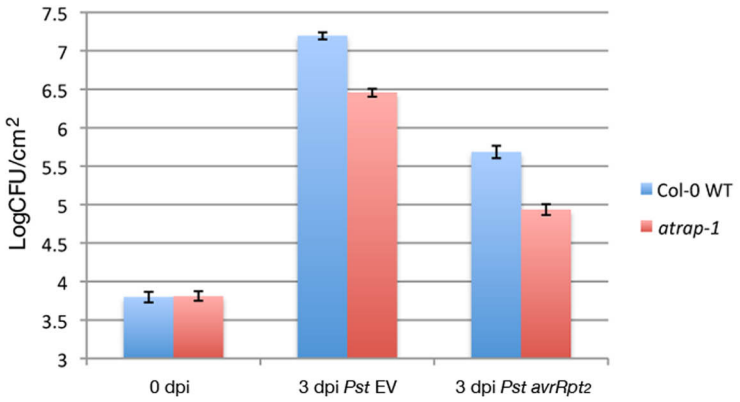
Figure 4.4 *atrap-1* mutant and overexpression lines displayed alterations in disease resistance to both virulent and avirulent strains of *Pseudomonas syringae* pv. *tomato* (*Pst*) in pathogen growth assay and HR reaction.

(A) *atrap-1* mutant shows resistance to *Pst* (EV) and *Pst* (*avrRpt2*) compared to wild type plants. Similar results were obtained from three biological replicates and error bars represent standard deviation of eight leaf discs. Student's t test was performed to determine the significant differences between mutants and wild type plants.

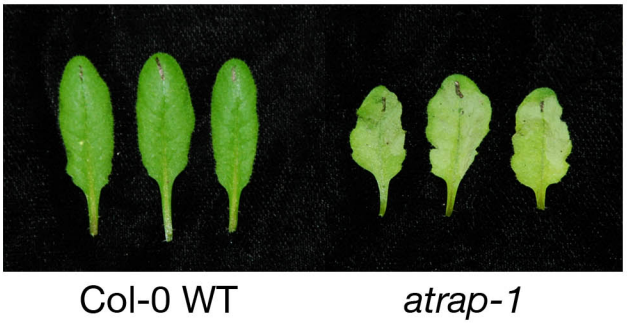
(B) *atrap-1* mutant shows early HR. Picture was taken at 8 hpi of *Pst* (*avrRpt2*) (1×10^7 cfu/ml)

(C) Two transgenic *Arabidopsis* lines (line 10 and 7) showed greater susceptibility to *Pst* (EV) and *Pst* (*avrRpt2*) than wild type plants. Similar results were obtained from three biological replicates and error bars represent the standard deviation of eight leaf discs. Student's t test was performed to determine the significant differences between mutants and wild type plants.

A


7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
LogCFU/cm2
0 dpi
3 dpi Pst EV
3 dpi Pst avrRpt2
Col-0 WT
atrap-1

B


Col-0 WT
atrap-1

C

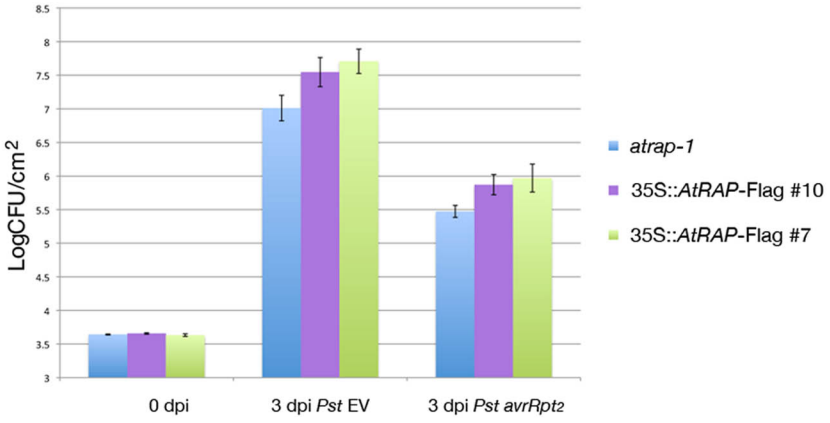

8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
LogCFU/cm2
0 dpi
3 dpi Pst EV
3 dpi Pst avrRpt2
atrap-1
35S::AtRAP-Flag #10
35S::AtRAP-Flag #7

Figure 4.5 Subcellular localization of *AtRAP* and *AtRAP-ΔRAP* are shown by transient expression in a stable transgenic *Arabidopsis* line in the *atrap-1* background.

1. Mesophyll chloroplast localization of AtRAP proteins fused to a YFP tag. Constructs consisted of either the control 35S::YFP, 35S::*AtRAP*-YFP or 35S::*AtRAP*-Δ*RAP*-YFP and were transiently expressed in *Nicotiana benthamiana* leaves. The fluorescence was analyzed using confocal laser scanning microscopy (SP2) at 48 hpi. Fluorescence images (left panel), light micrographs (middle panel) and merged images (right panel) were used to illustrate the different localizations of the proteins. The length of the bar corresponds to 150 μm.

(A) Cytosol localization of the YFP protein

(B) Mesophyll localization of the *AtRAP*-YFP

(C) Mesophyll localization of *AtRAP*-Δ*RAP*-YFP

2. Mesophyll chloroplast localization of *AtRAP* proteins fused to a GFP tag. Transgenic *Arabidopsis atrap-1* mutant expressing full length or a RAP domain deletion of the *AtRAP* under the control of a native promoter. The fluorescence was analyzed using Zeiss 510 confocal microscopy. Fluorescence images (left panel), light micrographs (middle panel) and merged images (right panel) were used to illustrate the different localizations of the proteins. The length of the bar corresponds to 10 μm.

(D) Mesophyll localization of the *AtRAP*-GFP

(E) Mesophyll localization of *AtRAP*-Δ*RAP*-GFP

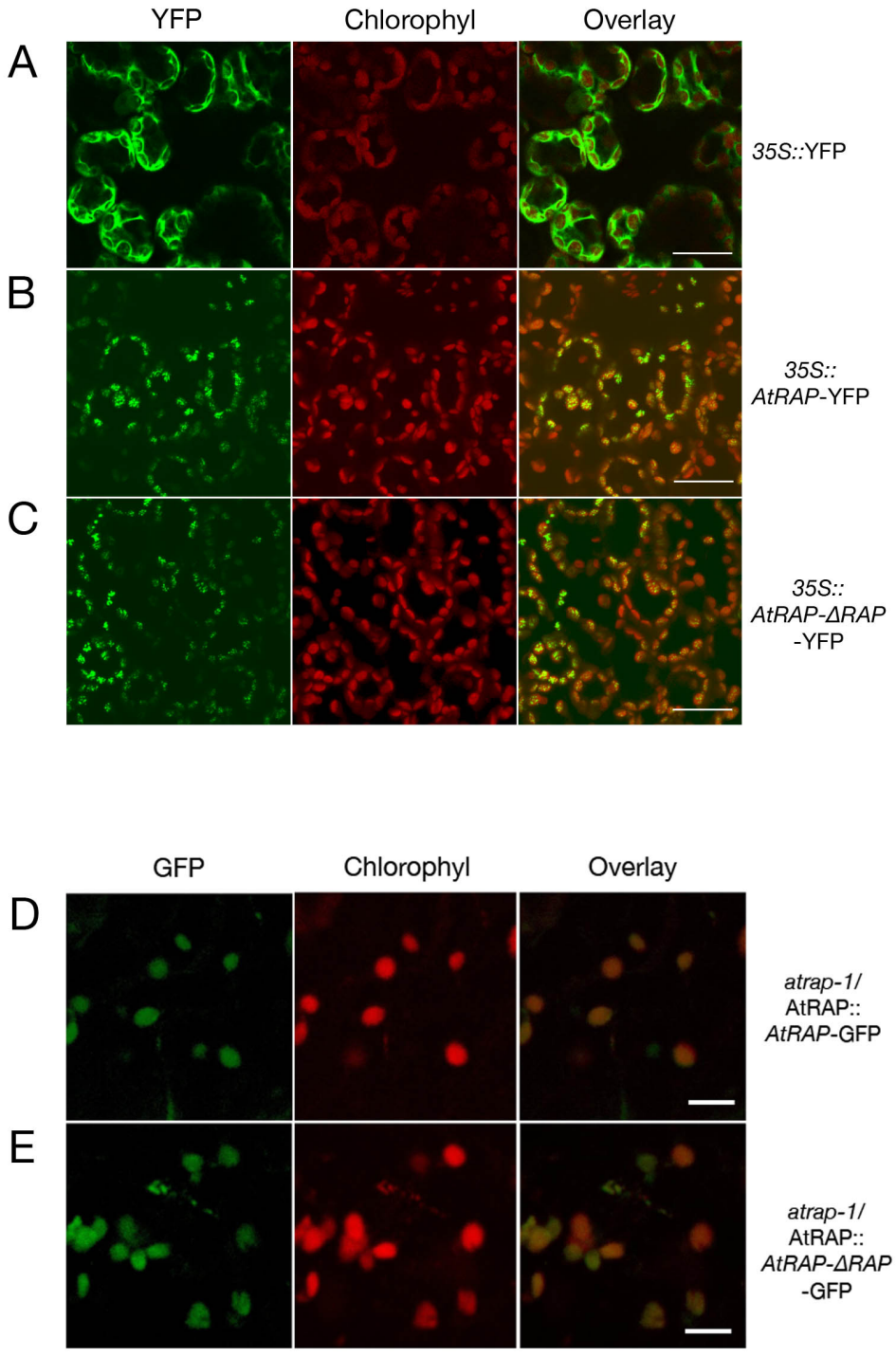

YFP
Chlorophyl
Overlay
A
35S::YFP
B
35S::
AtRAP-YFP
C
35S::
AtRAP-ΔRAP
-YFP
GFP
Chlorophyl
Overlay
D
atrap-1/
AtRAP::
AtRAP-GFP
E
atrap-1/
AtRAP::
AtRAP-ΔRAP
-GFP

Figure 4.6 Phenotypes of the transgenic plants with native promoters and RNA pull down result.

(A) Phenotypes of native promoter driven transgenic plants, from left to right: Col-0WT, *atrap-1*, pAtRAP::*AtRAP*-GFP in wild type background, pAtRAP::*AtRAP*-GFP in *atrap-1* background and pAtRAP::*AtRAP*-Δ*RAP*-GFP in *atrap-1* background

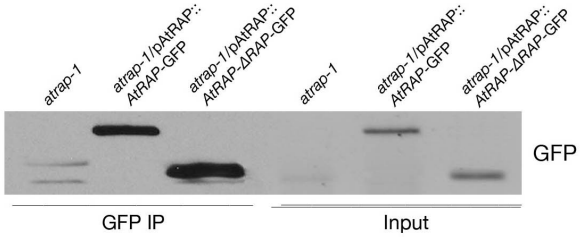
(B) Western blot showing that full length and deletion *AtRAP* proteins were recovered by GFP antibody and Protein A Agarose (Roche Applied Science). *atrap-1* extracts are shown as a control.

(C) RNAs extracted from full length and deletion *AtRAP* proteins were resolved in 10% acrylamide-bis gel. (Lane 1: RNAs extracted from deletion *AtRAP* construct; Lane 2: RNAs extracted from deletion *AtRAP* construct treated with RNase A; Lane 3: RNAs extracted from *atrap-1*; Lane 4: RNAs extracted from *atrap-1* treated with RNase A; Lane 5: RNAs extracted from full length *AtRAP* construct; Lane 6: RNAs extracted from full length *AtRAP* construct treated with RNase A)

A


Col-0 WT
atrap-1
Col-0 WT/
pAtRAP::AtRAP-GFP
atrap-1/
pAtRAP::AtRAP-GFP
atrap-1/
pAtRAP::AtRAP-Δ
RAP-GFP

B


atrap-1
atrap-1/pAtRAP::
AtRAP-GFP
atrap-1/pAtRAP::
AtRAP-ΔRAP-GFP
atrap-1
atrap-1/pAtRAP::
AtRAP-GFP
atrap-1/pAtRAP::
AtRAP-ΔRAP-GFP
GFP
GFP IP
Input

C

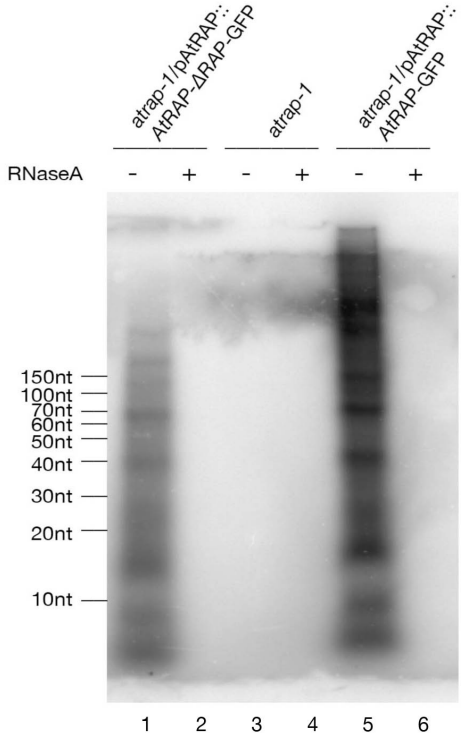

atrap-1/pAtRAP::
AtRAP-ΔRAP-GFP
atrap-1
atrap-1/pAtRAP::
AtRAP-GFP
RNaseA
- + - + - +
150nt
100nt
70nt
60nt
50nt
40nt
30nt
20nt
10nt
1 2 3 4 5 6

Figure 4.7 Transmembrane domain predictions of *AtRAP* and *OSJNBa0032H19.2* from Rice. The first transmembrane domain was labeled with "I", from residue 32 to 48 for *AtRAP* and from residue 10 to 36 for *OSJNBa0032H19.2*, respectively; the second transmembrane domain was labeled with "II", from residue 320 to 339 for *AtRAP* and from residue 286 to 305 for *OSJNBa0032H19.2*, respectively.

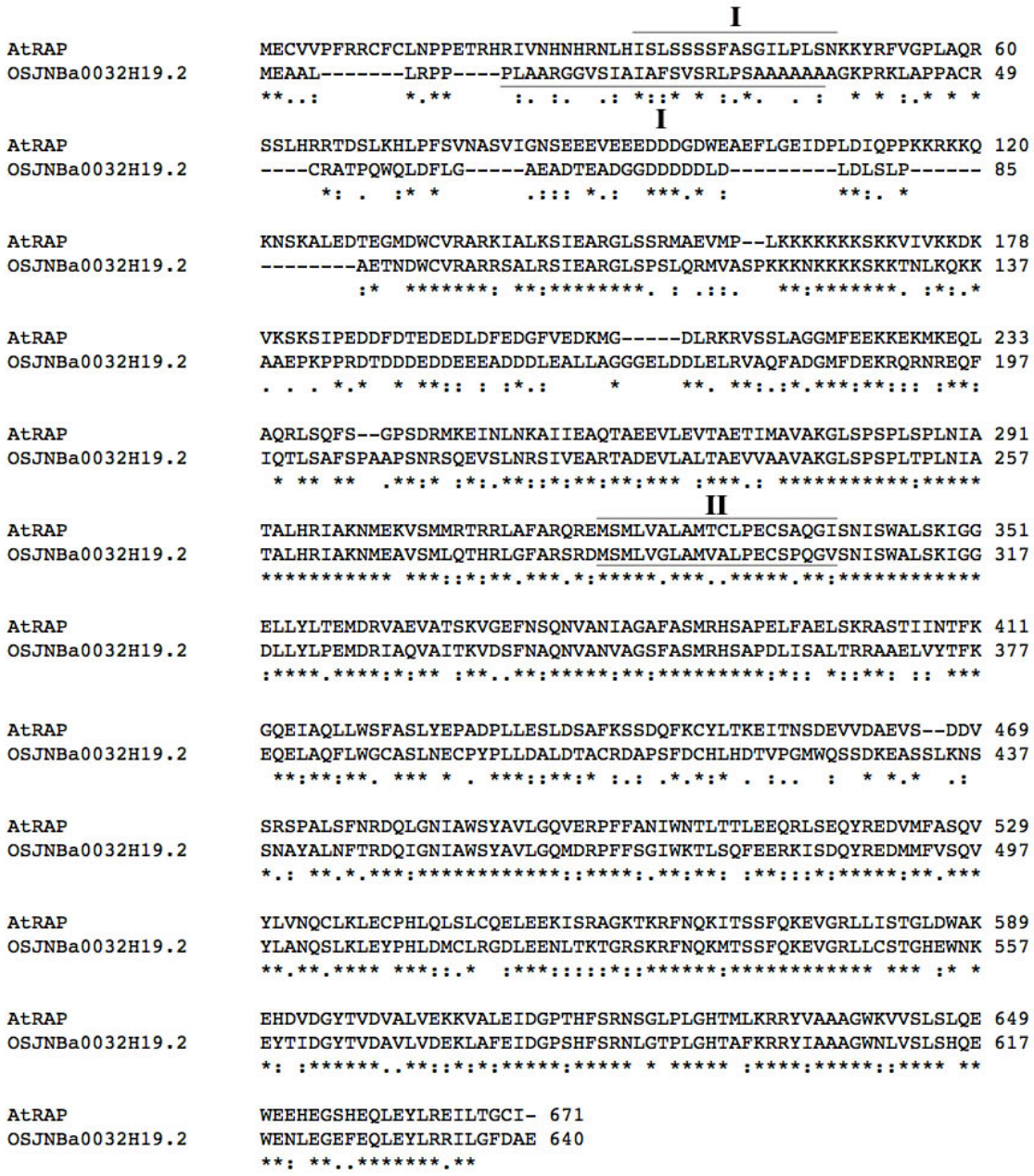

I
AtRAP MECVVPFRRCFCLNPPETRHRIVNHNHRNLHISLSSSSFASGILPLSNKKYRFVGPLAQR 60
OSJNBa0032H19.2 MEAAL-------LRPP----PLAARGGVSIAIAFSVSRLPSAAAAAAAGKPRKLAPPACR 49
I
AtRAP SSLHRRTDSLKHLPFSVNASVIGNSEEEVEEEDDDGDWEAEFLGEIDPLDIQPPKKRKKQ 120
OSJNBa0032H19.2 ----CRATPQWQLDFLG-----AEADTEADGGDDDDDLD---------LDLSLP------ 85
AtRAP KNSKALEDTEGMDWCVRARKIALKSIEARGLSSRMAEVMP--LKKKKKKKSKKVIVKKDK 178
OSJNBa0032H19.2 --------AETNDWCVRARRSALRSIEARGLSPSLQRMVASPKKKNKKKKSKKTNLKQKK 137
AtRAP VKSKSIPEDDFDTEDEDLDFEDGFVEDKMG-----DLRKRVSSLAGGMFEEKKEKMKEQL 233
OSJNBa0032H19.2 AAEPKPPRDTDDDEDDEEEADDDLEALLAGGGELDDLELRVAQFADGMFDEKRQRNREQF 197
AtRAP AQRLSQFS--GPSDRMKEINLNKAIIEAQTAEEVLEVTAETIMAVAKGLSPSPLSPLNIA 291
OSJNBa0032H19.2 IQTLSAFSPAAPSNRSQEVSLNRSIVEARTADEVLALTAEVVAAVAKGLSPSPLTPLNIA 257
II
AtRAP TALHRIAKNMEKVSMMRTRRLAFARQREMSMLVALAMTCLPECSAQGISNISWALSKIGG 351
OSJNBa0032H19.2 TALHRIAKNMEAVSMLQTHRLGFARSRDMSMLVGLAMVALPECSPQGVSNISWALSKIGG 317
AtRAP ELLYLTEMDRVAEVATSKVGEFNSQNVANIAGAFASMRHSAPELFAELSKRASTIINTFK 411
OSJNBa0032H19.2 DLLYLPEMDRIAQVAITKVDSFNAQNVANVAGSFASMRHSAPDLISALTRRAAELVYTFK 377
AtRAP GQEIAQLLWSFASLYEPADPLLESLDSAFKSSDQFKCYLTKEITNSDEVVDAEVS--DDV 469
OSJNBa0032H19.2 EQELAQFLWGCASLNECPYPLLDALDTACRDAPSFDCHLHDTVPGMWQSSDKEASSLKNS 437
AtRAP SRSPALSFNRDQLGNIAWSYAVLGQVERPFFANIWNTLTTLEEQRLSEQYREDVMFASQV 529
OSJNBa0032H19.2 SNAYALNFTRDQIGNIAWSYAVLGQMDRPFFSGIWKTLSQFEERKISDQYREDMMFVSQV 497
AtRAP YLVNQCLKLECPHLQLSLCQELEEKISRAGKTKRFNQKITSSFQKEVGRLLISTGLDWAK 589
OSJNBa0032H19.2 YLANQSLKLEYPHLDMCLRGDLEENLTKTGRSKRFNQKMTSSFQKEVGRLLCSTGHEWNK 557
AtRAP EHDVDGYTVDVALVEKKVALEIDGPTHFSRNSGLPLGHTMLKRRYVAAAGWKVVSLSLQE 649
OSJNBa0032H19.2 EYTIDGYTVDAVLVDEKLAFEIDGPSHFSRNLGTPLGHTAFKRRYIAAAGWNLVSLSHQE 617
AtRAP WEEHEGSHEQLEYLREILTGCI- 671
OSJNBa0032H19.2 WENLEGEFEQLEYLRRILGFDAE 640

Table 4.1 Primers used for RT-PCR

Primer name	5’ - 3’
PR-1 QRT Forward	GGAGCTACGCAGAACAACTA
PR-1 QRT Reverse	AGTATGGCTTCTCGTTCACA
β-tubulin forward	CAACGCTACTCTGTCTGTCC
β-tubulin reverse	TCTGTGAATTCCATCTCGTC
β-actin-2 Forward	AGTGGTCGTACAACCGGTATTGT
β-actin-2 Reverse	GATGGCATGAGGAAGAGAGAAAC
PDF1.2a QRT Forward	CTTGTTCTCTTTGCTGCTTTCGAC
PDF1.2a QRT Reverse	TAGTTGCATGATCCATGTTTG
PR-2 QRT Forward	TCTTCAACCACACAGCTGGACAAA
PR-2 QRT Reverse	CTTGTCGGCCTCCGTTTGACTGGA

Table 4.2 Genes up-regulated in *atrap-1*

Affimetrix code	Relative levels (WT/*atrap-1*)	AGI code	Gene description
253208_at	1.5466	AT4G34830	MRL1 (Pentatricopeptide repeat protein)
264698_at	1.8240	AT1G70200	RNA-binding (RRM)-containing protein
251638_at	1.9411	AT3G57490	RPS2D (40S ribosomal protein S2)
254907_at	2.3464	AT4G11190	disease resistance-responsive family protein
246411_at	2.3464	AT1G57770	FAD/NAD(P)-binding oxidoreductase family protein
244997_at	2.6193	ATCG00170	RNA polymerase beta' subunit-2
249645_at	2.9723	AT5G36910	thionin 2.2

Table 4.3 Genes down-regulated in *atrap-1*

Affimetrix code	Relative levels (WT/*atrap-1*)	AGI code	Gene description
254716_at	0.1347	AT4G13560	UNE15 (unfertilized embryo sac 15)
250824_at	0.1462	AT5G05200	Protein kinase superfamily protein
262357_at	0.3179	AT1G73040	jacalin lectin family protein
260869_at	0.3228	AT1G43800	acyl-(acyl-carrier-protein) desaturase
265948_at	0.3668	AT2G19590	ACC oxidase 1
260004_at	0.4208	AT1G67860	unknown protein
252612_at	0.4555	AT3G45160	unknown protein
254386_at	0.4634	AT4G21960	PRXR1 (Peroxidase superfamily protein)
263715_at	0.4643	AT2G20570	GLK1 (ARABIDOPSIS GOLDEN2-LIKE 1)
255626_at	0.4983	AT4G00780	TRAF-like family protein
245196_at	0.5058	AT1G67750	pectate lyase family protein
259001_at	0.5220	AT3G01960	unknown protein
259535_at	0.5271	AT1G12280	disease resistance protein (CC-NBS-LRR class)
262958_at	0.5329	AT1G54410	dehydrin family protein
264611_at	0.5543	AT1G04680	pectate lyase family protein
262582_at	0.5608	AT1G15410	aspartate-glutamate racemase family
245307_at	0.5688	AT4G16770	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein
255578_at	0.5757	AT4G01450	nodulin MtN21 family protein
247685_at	0.5917	AT5G59680	leucine-rich repeat protein kinase
246595_at	0.5983	AT5G14780	Encodes a NAD-dependent formate dehydrogenase
261368_at	0.6011	AT1G53070	legume lectin family protein