MicroRNAs and Trans-acting siRNA Pathways in Apple (*Malus x domestica* Borkh.) and Peach (*Prunus persica*)

Rui Xia

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Horticulture

Eric P. Beers (Committee Chair)

Zongrang Liu

Anthony K. Wolf

Bingyu Zhao

May, 2013

Blacksburg, Virginia, USA

Key Words: apple, peach, deep-sequencing, miRNA, tasiRNA, *TAS*, *MYB*, *PHAS*, phasiRNA, *PPR*, *NB-LRR*

MicroRNAs and Trans-acting siRNA Pathways in Apple (*Malus x domestica* Borkh.) and Peach (*Prunus persica*)

Rui Xia

Abstract

The unveiling of small RNA (sRNA)-mediated gene regulatory pathways has profoundly shaped our understanding of the complexity of gene regulation. In eukaryotes, sRNAs have been found to control cellular metabolism, growth and differentiation, to maintain genome integrity, and to combat viruses and mobile genetic elements. To gain insight into the roles of small RNAs in apple and peach, we conducted sRNA-seq, computational analysis and molecular experiments to genome-widely characterize their microRNAs (miRNAs) and trans-acting siRNA (tasiRNA) pathways.

We identified totally 75 miRNAs or families, including 23 conserved, 10 less-conserved and 42 apple-specific ones, and 118 miRNA target genes in apple. Two classical trans-acting siRNA (tasiRNA) pathways, miR390-*TAS3* and miR828-*TAS4*, were characterized with similar but unique tasiRNA biogenesis profiles and target specificities. Importantly, miR159, miR828 and miR858 can collectively target up to 81 *MYB* genes potentially involved in diverse aspects of plant growth and development. In contrast to the location of the miR159 target site in a sequence-divergent region, the target sites of miR828 and miR858 are located in the region encoding the conserved R3 repeat domain of MYB proteins. 10 out of the 19 miR828-targeted *MYB*s undergo the biogenesis of various phased siRNA (phasiRNA), which potentially regulate diverse genes outside the *MYB* family. In peach, totally 94 miRNAs or families and 80 target genes were identified. Similar pathways of tasiRNA (miR828-*TAS4* and miR390-*TAS3*) or phasiRNA (miR828-*MYB*-siRNA) processing were also characterized in peach.

Taking advantage of reverse computation and public available deep-sequencing data, we demonstrated that the miRNA-*TAS*-*PPR*-siRNA pathway is a highly dynamic and widespread feature of eudicots. Nine eudicot plants, representing six different plant families, have evolved similar tasiRNA pathways to instigate phasiRNA production from *PPR* genes, which are triggered by different 22-nt miRNAs, including miR7122,

miR1509, and fve-PPRtri1/2 and through distinct mechanistic strategies, like miRNA direct-targeting or indirect-targeting through *TAS*-like genes, one-hit or two-hit, or even two layers of tasiRNA-*TAS* interactions. We found that the *MIRNA* genes of these miRNA triggers show great identity with the *Arabidopsis MIR173*, implying a common origin of this group of miRNAs (super-miR7122). Combined results from phylogenetic analyses and conservation extent profiling revealed that the super-miR7122 was potentially evolved from another miRNA superfamily (super-miR4376), which probably originated from the miR390. Additionally, the miR482/2118-*NB-LRR*-siRNA pathway was found to be conserved, but evolved with distinct features, in apple and peach.

Taken together, widespread and complex miRNA and tasiRNA regulatory networks have been adapted in apple and peach. They add another crucial layer of regulation on gene activity and stability, and must exert essential functions in all aspects of plant life.

Acknowledgement

I would like to express my sincere appreciation to Zongrang Liu and Eric P. Beers for their guidance and mentorship during my thesis work, Tony Wolf, Bingyu Zhao, Roger Harris for support and advice. A special thanks to Rongcai Yuan, who gave me this invaluable study opportunity, in memory of his care and financial support during the first year of my study. Many thanks to the following people: Hong Zhu, Christopher Dardick, Ann M. Callahan, Yongqiang An, Blake Meyers for discussion and suggestion, Chengsong Zhao, and Guangdu Gao for technique assistances, Dennis Bennett, Carrie Whitaker, Jinjin Zhang, and Zhifeng Wen for experimental assistances. Thanks also to my family and friends, above all my lovely wife Jing Xu and daughter Linsey Xia.

Dedication

This dissertation is dedicated to my wife Jing Xu, my daughter Linsey Xia, my parents Anming Xia and Yingming Xia, my parents-in-law Jianhua Xu and Yan Chen, my brother Wusheng Xia's family and my sister Huiyan Xia's family for their love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1
General introduction

The discovery of small [~20-24 nucleotide (nt)] regulatory noncoding RNAs has been reshaping our understanding of RNA molecules in the last two decades. These small RNAs (sRNAs) regulate the genome function at some of the most important levels, including heterochromatin formation, chromosome segregation, gene transcription, RNA processing, RNA stability, and gene translation (Baulcombe, 2004; Carthew and Sontheimer, 2009). The effects of sRNAs at the transcriptional or post-transcriptional level are general inhibitory, and the corresponding regulatory mechanisms are therefore collectively referred to as RNA silencing, where sRNAs always serve as specificity determinant, directing effector proteins to target nucleic acid molecules via base-pairing interactions (Carthew and Sontheimer, 2009; Voinnet, 2009).

1.1 Small RNAs

In eukaryotes, the primary RNAs are derived from the decoding of genetic information stored in the DNA genome by DNA-dependent RNA polymerases which convert the DNA into RNA (transcription). Eukaryotes normally possess three nuclear RNA polymerases (Pols) to decipher their genomes, with Pol I specified in the transcription of primary transcripts which mature into different catalytic RNA of ribosomes, Pol II transcribing thousands of protein-coding genes, long non-coding RNAs and miRNAs, and Pol III transcribing several classes of smaller RNAs, such as 5S ribosomal RNA, transfer RNA, small regulatory RNAs and short interspersed nuclear elements (Haag and Pikaard, 2011). Remarkably, plants have evolved two additional Pols, Pol IV and V, which orchestrate non-coding RNA-mediated gene silencing processes. Pol IV is required for 24-nt siRNA production and Pol V for siRNA-mediated gene silencing of transposons and other repeated elements (Haag and Pikaard, 2011).

Upon transcription, primary RNAs of specific features are routed into the RNA silencing pathway, which share four consensus biochemical steps in plants: (1) induction by double-stranded RNA (dsRNA), (2) dsRNA processing into 18-25-nt long sRNAs, (3) 3'-o-methylation of sRNA, and (4) sRNA incorporation into effector complexes that

associate with partially or fully complementary target RNA or DNA (Chapman and Carrington, 2007; Voinnet, 2009). These common steps are involved with several core components, including RNA-dependent RNA polymerase (RDR) proteins, Dicer-like ribonucleases (DCLs), dsRNA binding proteins (DRBs), sRNA methyltransferases and Argonaute (AGO) proteins.

1.2 Core components of RNA silencing pathways in plants

1.2.1 RNA-dependent RNA polymerase proteins (RDR)

RDRs are defined by the presence of a conserved RNA-dependent RNA polymerase catalytic domain. In eukaryotes, there are three major clades of RDRs, RDRα, RDRβ, and RDRγ, and only RDRα and RDRγ genes have been identified in plants (Zong et al., 2009). The *Arabidopsis* possesses six *RDR* genes (*RDR1-6*), including three *RDRα* genes, *RDR1/2/6*, which involve in diverse RNAi processes, and three *RDRγ* genes, *RDR3/4/5*, whose function awaits clarification. (Willmann et al., 2011). RDR1, RDR2, and RDR6 convert single-stranded RNAs (ssRNAs) to double-stranded RNAs (dsRNAs) that are subsequently processed into different types of siRNAs targeting specific endogenous loci. RDR2 is responsible for the biogenesis of the most abundant endogenous siRNAs in plants---heterochromatin siRNAs (hc-siRNAs). RDR2 is also required for the production of a class of long siRNAs (23-27-nt) that originate from miRNA-producing stem-loop precursor (Chellappan et al., 2010). RDR6 is involved in the biogenesis of two additional types of siRNAs: trans-acting siRNAs and natural antisense siRNA (nat-siRNAs) (Peragine et al., 2004; Vazquez et al., 2004; Allen et al., 2005; Borsani et al., 2005; Yoshikawa et al., 2005). The major role of RDR1, not as much studied as did RDR2/6, is producing and amplifying exogenous, virus-derived siRNAs in infected plants (Wang et al., 2010).

1.2.2 Dicer-like ribonucleases (DCL)

The dsRNAs, formed by RDRs, are sebsequently processed into specifically sized sRNA duplexes by Dicer-like (DCL1-4) proteins, which is a type of ribonuclease III-like enzymes. Arabidopsis has four specialized DCL proteins, which catalyze the formation of miRNA (DCL1), 22-nt (DCL2), 24-nt (DCL3) and 21-nt (DCL4) siRNAs from selected dsRNA precursors. DCL1 involves in the accurate processing of ~21-nt miRNAs from

their foldback precursors (Park et al., 2002; Reinhart et al., 2002). DCL2 plays roles in formation of natural antisense siRNAs and in transitive silencing of transgene transcript (Borsani et al., 2005; Mlotshwa et al., 2008). DCL3 functions to process RDR2-dependent dsRNA precursors to produce 24-nt-long siRNAs acting in silencing of transposable elements and heterochromatin formation (Cao and Jacobsen, 2002; Xie et al., 2004). DCL4 generates 21-nt trans-acting siRNAs (tasiRNAs) and exogenous viral sRNAs (Xie et al., 2005; Yoshikawa et al., 2005). DCL2 and DCL3 also play a surrogate role of DCL4 in antiviral RNA silencing (Bouche et al., 2006; Garcia-Ruiz et al., 2010).

1.2.3 dsRNA binding proteins (DRB),

Since DCL proteins do not possess the property of RNA binding, its processing of sRNAs from dsRNAs requires a partner that is a double-stranded RNA binding protein (DRB). Arabidopsis has a small family of five DRBs (DRB1-5) (Hiraguri et al., 2005). DRB1, also known as HYL1, is involved in the DCL1-mediated miRNA processing from the foldback precursor, and it probably recruits DCL1 through binding to the miRNA/miRNA* duplex and guides the selection of the miRNA strand (Han et al., 2004; Eamens et al., 2009; Yang et al., 2010). DRB4 works together with DCL4 involving in the biosynthesis of tasiRNAs, exogenous viral siRNA, as well as some miRNAs (Adenot et al., 2006; Bouche et al., 2006; Rajagopalan et al., 2006; Nakazawa et al., 2007; Garcia-Ruiz et al., 2010). DRB2, DRB3, and DRB5 were recently found to function in a non-canonical miRNA pathway with DRB2 participated in the biogenesis of selected miRNAs and DRB3 and DRB5 mediated target gene silencing of DRB2-associated miRNAs (Eamens et al., 2012b; Eamens et al., 2012a).

1.2.4 Small RNA methyltransferase

After the generation of small RNA duplexes from dsRNAs, these duplexes are channeled into a 2'-*O*-methylation process mediated by methyltransferases. HUA ENHANCER1 (HEN1) was first identified in Arabidopsis as a methyltransferase that methylates both miRNA and siRNA duplexes (Yu et al., 2005; Yang et al., 2006). HEN1 homologues, prevalently present in other organisms, were found to methylate sRNAs in other plants, piRNAs in animals and AGO2-associated sRNAs in Drosophila (Horwich et al., 2007; Kirino and Mourelatos, 2007; Ohara et al., 2007; Abe et al., 2010; Kamminga

et al., 2010). 2'-*O*-methylation on the 3' terminal ribose protects sRNAs from 3'-5' degradation and 3' uridylation, serving as a major mechanism modulating the turnover rate of sRNAs (Li et al., 2005; Yu et al., 2005; Kamminga et al., 2010).

1.2.5 Argonaute proteins (AGO)

The name of ARGONAUTE (AGO) stemmed from the founding member of the *Arabidopsis thaliana* AGO1, null mutant of which has tubular shaped leaves resembling the tentacles of a argonaute squid (Bohmert et al., 1998). Shortly after the discovery of the Arabidopsis AGO1, the vital role of AGO proteins in small RNA-directed silencing was characterized in numerous organisms (Mallory and Vaucheret, 2010). Based on their functional domains, eukaryotic AGO proteins can be divided into two major groups: the AGO and the PIWI subfamilies (Hutvagner and Simard, 2008; Vaucheret, 2008). Numbers of AGO gene range from 0 in yeast *Saccharomyces cerevisiae* to as many as 27 in the nematode worm *Caenorhabditis elegans*, with Drosophila melanogaster and human expressing five and eight AGO genes, respectively (Mallory and Vaucheret, 2010). All of these AGO proteins fall into both subfamilies, while plant AGO proteins are all restricted to the AGO subfamily. In Arabidopsis there are 10 AGO proteins, which is phylogenetically classified into three major clades, the AGO1/5/10 clade; the AGO2/3/7 clade; and the AGO4/6/8/9 clade (Vaucheret, 2008). The working body of RNA silencing complexes is defined by an AGO protein and an associated sRNA, which acts as a sequence-specific guides leading AGO proteins to perfectly or partially complementary RNAs (Mallory and Vaucheret, 2010). Not only can AGO proteins function as Slicer proteins, but also regulate target genes through translational inhibition, likely depending the degree of complementarity between a small RNA and its target gene. In Arabidopsis, transcriptional gene silencing (TGS) and posttranscriptional gene silencing (PTGS) result from the action of specialized AGO protein and specialized sRNAs. TGS complexes involve AGO proteins from the AGO4/6/9 clade, which typically associate with 24-nt sRNAs. By contrast, PTGS complexes involve the cleavage-competent AGO1 and AGO7 and their associated 21- to 22-nt sRNAs (Mallory and Vaucheret, 2010). AGO1 and AGO10 are required for translational control of several miRNA targets (Brodersen et al., 2008). Recently, AGO10 was found to specifically sequester miR165/166 to regulate shoot meristem development (Zhu et al., 2011). Besides the sRNA length, 5' nucleotide

identity of sRNAs contribute to the sorting of sRNAs into AGO proteins. AGO1 prefers 21-22-nt sRNA with a 5' U whereas AGO2 selectively associate with 21-nt 5' A sRNAs; the majority of AGO4/6/9 loaded sRNAs are 24-nt long and have a 5' initial A, whereas AGO5 preferentially associates with 24-nt sRNA beginning with a 5' C (Mi et al., 2008; Havecker et al., 2010).

1.3 Biogenesis and function of small RNAs in plants

Based on the biogenesis and function properties, sRNAs in plant cells can generally be divided into two major categories: microRNAs (miRNAs) and small interfering RNAs (siRNAs).

1.3.1 microRNAs

Plant miRNAs are predominantly 20-22-nt small RNAs processed from primary *MIRNA* transcripts (pri-miRNA) transcribed by Pol II from long non-coding *MIRNA* genes. The pri-miRNA contains a region of imperfect self-complementarity that folds into a stem-loop RNA helix (Voinnet, 2009). The stem-loop structure, the source of dsRNAs, is processed by the DCL1 with the aid of two other proteins, the C2H2-zinc finger protein SERRATE (SE) and the double -stranded RNA-binding protein HYPONASTIC LEAVES1 (HYL1/DRB1), which could facilitate the recognition and routing of pri-miRNA and the loading or positioning of DCL1 (Han et al., 2004; Grigg et al., 2005; Dong et al., 2008). The DCL1 cleavage of both arms of the stem-loop helix releases the miRNA/miRNA* duplex, which possesses a 2-nt overhang at both ends. The first cleavage of most MIRNAs is occurred at the loop-distal side of the stem-loop with a distance of ~15-nt from an unpaired region (Mateos et al., 2010; Song et al., 2010; Werner et al., 2010), although the MIR159/319 family is cleaved loop-side first (Addo-Quaye et al., 2009; Bologna et al., 2009).The mature miRNA/miRNA* duplex is then methylated at both 3' ends by the S-adenosyl methionine-dependent methyltransferase HEN1 and transferred to the cytoplasm with the aid of HASRTY protein (Yu et al., 2005). The final step of miRNA maturation is the separation of the miRNA and miRNA*and incorporation of the mature miRNA into an AGO effector protein. The differential

thermostability of the miRNA/miRNA* duplex ends and the HYL1/DRB1 contributes to the selection of mature miRNA strand (Takeda et al., 2008; Eamens et al., 2009).

The major role of plant miRNAs is regulating their target genes through mRNA slicing, although translational repression of target genes by miRNAs is widely present (Voinnet, 2009). These two layers of regulation, protein production and RNA stability control, often coincide spatially or temporally (Voinnet, 2009). However, the functional analysis of plant miRNAs has been focused on their slicing roles so far. Based on the conservation extent, plants miRNAs can be grouped into conserved, and lineage/species-specific miRNAs (Jones-Rhoades et al., 2006; Axtell and Bowman, 2008). There are 23 conserved miRNAs (or families) which is present in the common ancestor of spermatophyte (Jones-Rhoades et al., 2006; Cuperus et al., 2011). These conserved miRNAs are usually of high expression in cells and targeting pivotal transcription factors playing critical roles in diverse developmental stages or metabolic pathways. For instances, the sequential action of miR156 and miR172 coordinates the developmental phase transition of Arabidopsis (Wu et al., 2009), and miR165/166 regulate shoot apical meristem development through targeting class III homeodomain-leucine zipper (HD-ZIP III) transcription factor (Zhu et al., 2011). By contrast, the lineage/species-specific miRNAs are of a relatively low abundance, and are thought to be less functionally essential, thus less well-characterized (Jones-Rhoades et al., 2006).

1.3.2 Small interfering RNAs

Another class of sRNAs, small interfering RNAs (siRNA), accounts for the majority of the sRNA population in plants. Most of siRNA are of 24-nt in length and derived from repetitive genomic elements including centromeres, transposons and retrotransposons, which are targets of DNA methylation and heterochromatin formation, thus this class of 24-nt siRNAs are referred to as heterochromatin siRNA (hc-siRNAs) or repeat-associated siRNAs (ra-siRNAs) (Rajagopalan et al., 2006; Zhang et al., 2006; Law and Jacobsen, 2010). First Pol IV generates single-stranded RNA (ssRNA) transcripts from these repetitive loci (Pontier et al., 2005). ssRNAs are then converted to dsRNAs by RDR2, which are subsequently chopped into 24-nt siRNA duplexes by DCL3 (Law and Jacobsen, 2010). Mature siRNAs are incorporated into the AGO4-family protein complex to direct RNA-dependent DNA methylation (RdDM)of target DNA loci. The siRNA mediated

RdDM are associated with diverse critical biological processes, including transposition control, heterochromatin establishment, and post-transcriptional gene regulation (Matzke et al., 2009; Allen and Howell, 2010; Law and Jacobsen, 2010).

1.3.3 Trans-acting siRNAs

Apart from the predominant 24-nt siRNA, plants also produce a 21-nt class of siRNA, trans-acting siRNAs (tasiRNA). The biogenesis of tasiRNAs relies on a miRNA cleavage on their non-coding transcripts of trans-acting siRNA genes (*TAS*), after which one of the cleaved fragments is stabilized by SUPPRESSOR OF GENE SILENCING3 (SGS3) and converted into dsRNA by RDR6 (Peragine et al., 2004; Yoshikawa et al., 2005). DCL4 is recruited to the miRNA cleavage site and sequentially processes the dsRNA into 21-nt tasiRNA duplexes, thus all the tasiRNAs are in phase with the cleavage site (Allen et al., 2005; Yoshikawa et al., 2005; Axtell et al., 2006). Like miRNA, tasiRNAs are incorporated into AGO effector complexes to exert slicing activity in *trans* to target genes.

22-nt miRNAs trigger secondary siRNA production from their targeted transcripts, and the miRNA triggers are often with a 5'-end U, implying an association of the AGO1 complex (Chen et al., 2010; Cuperus et al., 2010). To date, four classical *TAS* gene families have been well characterized in Arabidopsis, including the *TAS1/2*, *TAS3*, and *TAS4* targeted by miR173, miR390, and miR828, respectively (Allen et al., 2005; Axtell and Bartel, 2005; Yoshikawa et al., 2005; Rajagopalan et al., 2006). After the miR173-guided cleavage in *TAS1/2*, the poly(A) end fragment is routed into dsRNA formation and ensuing production of tasiRNAs targeting several PPR genes which might serve roles in cytoplasmic male sterility (Yoshikawa et al., 2005; Howell et al., 2007). Similarly, the miR828 directs the tasiRNA biogenesis from the poly(A) side after its cleavage in *TAS4*, and one of its derived tasiRNAs regulate several MYB genes associated with anthocyanin biogenesis (Rajagopalan et al., 2006; Luo et al., 2011). Distinct from the 22-nt miR173 and miR828 which trigger siRNA production with only one target site bearing in the target gene (one hit), the 21-nt miR390 instigates the tasiRNA production from *TAS3* through a different "two-hit" mechanism, with the *TAS3* gene bearing two miR390 target sites, a noncleavable 5' site and a cleavable 3' site (Axtell et al., 2006). After the cleavage of the 3' site, the cap-end fragment, instead of the poly(A)-end fragment, is converted to

dsRNAs and ensuing tasiRNA is produced in a 3'-5' direction, thus all the tasiRNAs are in phase with the 3' cleavage site (Allen et al., 2005; Axtell et al., 2006). Two tasiRNAs (also termed as tasiARFs) target auxin responsive factors, which play essential roles in leaf patterning and organ development (Allen et al., 2005; Axtell et al., 2006; Fahlgren et al., 2006).

Additionally, protein-coding genes are also channeled into secondary siRNA production by 22-nt miRNAs or tasiRNAs (Howell et al., 2007; Wang et al., 2011; Zhai et al., 2011). These secondary siRNAs are collectively referred to as phased siRNAs (phasiRNAs), some of which may target genes in *trans* (tasiRNAs) or may act in *cis* (casiRNAs), but the exact function for most phasiRNAs remains unexplored (Zhai et al., 2011).

1.3.4 Other types of sRNAs

Mant other types of sRNAs have also been reported in plants. Natural anti-sense siRNAs (nat-siRNAs) are processed from dsRNAs derived from convergent transcription of adjacent genes, which are induced by abiotic and biotic stresses or in specific developmental stages (Borsani et al., 2005; Zhang et al., 2012). The biogenesis of nat-siRNAs requires DCL1 or DCL2, RDR6, and PolIV in Arabidopsis (Borsani et al., 2005). Recently a novel class of 21-nt sRNAs is found to be involved in DNA double-strand breaks (DSBs) repair, and are referred to as DSB-induced sRNA (diRNAs). Its biogenesis requires the PI3 kinase, PolIV, and DCL proteins in Arabidopsis (Wei et al., 2012).

1.4 High-throughput next generation sequencing technology

As early as 1993, long before the mechanism of RNA silencing was uncovered, the first small RNA, now known as miRNA lin-4, was identified in *Caenorhabditis elegans* by forward genetic studies (Lee et al., 1993). The identification of a second miRNA, also through forward genetics in *C. elegans*, let-7 prompted the search of other small RNAs, and tens of miRNAs were found in several animal species through direct cloning. Since then direct cloning as well as bioinformatics prediction have revealed a large complement of miRNAs in diverse animal and plant species (Chen, 2009). After the first successful application of high-throughput next-generation DNA sequencing (NGS, also termed as deep-sequencing) technology for sRNA sequencing (Lu et al., 2005), the characterization

of sRNAs has been booming to all kind of species and this NGS technology has become the authentic approach to genome-widely profile the sRNA population in cells.

Compared to the traditional Sanger sequencing protocols, the NGS is of advantage in many aspects. The NGS is of extremely large capacity, i.e., it can conduct numerous sequence determinations in parallel without prior knowledge of sequence information. Apart from providing nucleotide-order information, the results of NGS can reflect the relative expression level of every single sequence. Moreover the NGS is more labor-, cost-, and time-effective. It can finish the resequencing of a human genome within hours. Therefore, the NGS has been used for many applications, including human genome resequencing, de nove assemblies of bacterial and low eukaryotic genomes, transcriptome cataloguing (RNA-seq), small RNA repertoire characterization (sRNA-seq), genome-wide profiling of epigenetic marks (methyl-seq) and chromatin structure (ChIP-seq), and species classification and new gene discovery by metagenomic studies (Metzker, 2010).

1.5 References

Abe, M., Yoshikawa, T., Nosaka, M., Sakakibara, H., Sato, Y., Nagato, Y., and Itoh, J. (2010). WAVY LEAF1, an ortholog of Arabidopsis HEN1, regulates shoot development by maintaining MicroRNA and trans-acting small interfering RNA accumulation in rice. Plant Physiol **154,** 1335-1346.

Addo-Quaye, C., Snyder, J.A., Park, Y.B., Li, Y.-F., Sunkar, R., and Axtell, M.J. (2009). Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. RNA **15,** 2112-2121.

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

Allen, E., and Howell, M. (2010). miRNAs in the biogenesis of trans-acting siRNAs in higher plants. Semin Cell Dev Biol **21,** 798 - 804.

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

Axtell, M., and Bartel, D. (2005). Antiquity of microRNAs and their targets in land plants. Plant Cell **17,** 1658 - 1673.

Axtell, M., and Bowman, J. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci **13,** 343 - 349.

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

Baulcombe, D. (2004). RNA silencing in plants. Nature **431,** 356-363.

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

Bologna, N.G., Mateos, J.L., Bresso, E.G., and Palatnik, J.F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. EMBO J **28,** 3646-3656.

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

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

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science **320,** 1185 - 1190.

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

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell **136,** 642-655.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nature reviews. Genetics **8,** 884-896.

Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W., and Jin, H. (2010). siRNAs from miRNA sites mediate DNA methylation of target genes. Nucleic Acids Res **38,** 6883-6894.

Chen, H., Chen, L., Patel, K., Li, Y., Baulcombe, D.C., and Wu, S. (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proceedings of the National Academy of Sciences.

Chen, X. (2009). Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol **25,** 21 - 44.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and functional diversification of miRNA genes. Plant Cell **23,** 431-442.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol **17,** 997-1003.

Dong, Z., Han, M.H., and Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A **105,** 9970-9975.

Eamens, A.L., Kim, K.W., and Waterhouse, P.M. (2012a). DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in Arabidopsis thaliana. Plant signaling & behavior **7,** 1224-1229.

Eamens, A.L., Kim, K.W., Curtin, S.J., and Waterhouse, P.M. (2012b). DRB2 Is Required for MicroRNA Biogenesis in <italic>Arabidopsis thaliana</italic>. PLoS ONE **7,** e35933.

Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.B., and Waterhouse, P.M. (2009). The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. RNA **15,** 2219-2235.

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

Garcia-Ruiz, H., Takeda, A., Chapman, E.J., Sullivan, C.M., Fahlgren, N., Brempelis, K.J., and Carrington, J.C. (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. Plant Cell **22,** 481-496.

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

Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature reviews. Molecular cell biology **12,** 483-492.

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

Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H., and Baulcombe, D.C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell **22,** 321-334.

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

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Current biology : CB **17,** 1265-1272.

Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C. (2007). Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell **19,** 926-942.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nature reviews. Molecular cell biology **9,** 22-32.

Jones-Rhoades, M., Bartel, D., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol **57,** 19 - 53.

Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J **29,** 3688-3700.

Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nat Struct Mol Biol **14,** 347-348.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics **11,** 204-220.

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

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

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

Luo, Q.J., Mittal, A., Jia, F., and Rock, C.D. (2011). An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in *Arabidopsis*. Plant Molecular Biology.

Mallory, A., and Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. Plant Cell **22,** 3879-3889.

Mateos, J.L., Bologna, N.G., Chorostecki, U., and Palatnik, J.F. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. Current biology : CB **20,** 49-54.

Matzke, M., Kanno, T., Claxinger, L., Huettel, B., and Matzke, A.J.M. (2009). RNA-mediated chromatin-based silencing in plants. Curr. Opin. Cell Biol. **21,** 367-376.

Metzker, M.L. (2010). Sequencing technologies [mdash] the next generation. Nature reviews. Genetics **11,** 31-46.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell **133,** 116-127.

Mlotshwa, S., Pruss, G.J., Peragine, A., Endres, M.W., Li, J., Chen, X., Poethig, R.S., Bowman, L.H., and Vance, V. (2008). DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis. PLoS One **3,** e1755.

Nakazawa, Y., Hiraguri, A., Moriyama, H., and Fukuhara, T. (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.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., and Suzuki, T. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. Nat Struct Mol Biol **14,** 349-350.

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

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H., and Poethig, R. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNA in *Arabidopsis*. Genes Dev **18,** 2368 - 2379.

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

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

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

Song, L., Axtell, M.J., and Fedoroff, N.V. (2010). RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. Current biology : CB **20,** 37-41.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant & cell physiology **49,** 493-500.

Vaucheret, H. (2008). Plant ARGONAUTES. Trends Plant Sci **13,** 350-358.

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

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell **136,** 669-687.

Wang, X.B., Wu, Q., Ito, T., Cillo, F., Li, W.X., Chen, X., Yu, J.L., and Ding, S.W. (2010). RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci U S A **107,** 484-489.

Wang, Y., Itaya, A., Zhong, X., Wu, Y., Zhang, J., van der Knaap, E., Olmstead, R., Qi, Y., and Ding, B. (2011). Function and evolution of a MicroRNA that regulates a *Ca^{2+}-ATPase* and triggers the formation of phased small interfering RNAs in tomato reproductive growth. Plant Cell **23,** 3185-3203.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. Cell **149,** 101-112.

Werner, S., Wollmann, H., Schneeberger, K., and Weigel, D. (2010). Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. Current biology : CB **20,** 42-48.

Willmann, M.R., Endres, M.W., Cook, R.T., and Gregory, B.D. (2011). The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. The Arabidopsis book / American Society of Plant Biologists **9,** e0146.

Wu, G., Park, M., Conway, S., Wang, J., Weigel, D., and Poethig, R. (2009). The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. Cell **138,** 750 - 759.

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

Xie, Z., Johansen, L., Gustafson, A., Kasschau, K., Lellis, A., Zilberman, D., Jacobsen, S., and Carrington, J. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol **2,** 642 - 652.

Yang, S.W., Chen, H.Y., Yang, J., Machida, S., Chua, N.H., and Yuan, Y.A. (2010). Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. Structure **18,** 594-605.

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

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

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. Science **307,** 932-935.

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., González, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., and Meyers, B.C. (2011). MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. **25,** 2540-2553.

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

Zhang, X., Xia, J., Lii, Y.E., Barrera-Figueroa, B.E., Zhou, X., Gao, S., Lu, L., Niu, D., Chen, Z., Leung, C., Wong, T., Zhang, H., Guo, J., Li, Y., Liu, R., Liang, W., Zhu, J.K., Zhang, W., and Jin, H. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol **13,** R20.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.H., Liou, L.W., Barefoot, A., Dickman, M., and Zhang, X. (2011). Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell **145,** 242-256.

Zong, J., Yao, X., Yin, J., Zhang, D., and Ma, H. (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. Gene **447,** 29-39.

Chapter 2

Apple miRNAs and tasiRNAs with novel regulatory networks

Rui Xia, Hong Zhu, Yongqiang An, Eric P. Beers, Zongrang Liu

Published in

Genome Biology

BioMed Central

236 Gray's Inn Road

London WC1X 8HB

United Kingdom

2012, **13**:R47 doi:10.1186/gb-2012-13-6-r47

http://genomebiology.com/2012/13/6/R47

Xiaet al.Genome Biology2012,13:R47
http://genomebiology.com/2012/13/6/R47

Genome Biology

RESEARCH Open Access

Apple miRNAs and tasiRNAs with novel regulatory networks

Rui Xia[1,2,3], Hong Zhu[2,3], Yong-qiang An[4], Eric P Beers[2] and Zongrang Liu[2,3*]

Abstract

Background: MicroRNAs (miRNAs) and their regulatory functions have been extensively characterized in model species but whether apple has evolved similar or unique regulatory features remains unknown.

Results: We performed deep small RNA-seq and identified 23 conserved, 10 less-conserved and 42 apple-specific miRNAs or families with distinct expression patterns. The identified miRNAs target 118 genes representing a wide range of enzymatic and regulatory activities. Apple also conserves two*TAS*gene families with similar but unique *trans*-acting small interfering RNA (tasiRNA) biogenesis profiles and target specificities. Importantly, we found that miR159, miR828 and miR858 can collectively target up to 81*MYB*genes potentially involved in diverse aspects of plant growth and development. These miRNA target sites are differentially conserved among*MYB*s, which is largely influenced by the location and conservation of the encoded amino acid residues in MYB factors. Finally, we found that 10 of the 19 miR828-targeted*MYB*s undergo small interfering RNA (siRNA) biogenesis at the 3 cleaved, highly divergent transcript regions, generating over 100 sequence-distinct siRNAs that potentially target over 70 diverse genes as confirmed by degradome analysis.

Conclusions: Our work identified and characterized apple miRNAs, their expression patterns, targets and regulatory functions. We also discovered that three miRNAs and the ensuing siRNAs exploit both conserved and divergent sequence features of*MYB*genes to initiate distinct regulatory networks targeting a multitude of genes inside and outside the *MYB* family.

Background

The discovery of RNA interference in the late 1990s [1] prompted a revolution in RNA biology, and the unveiling of small RNA (sRNA)-mediated gene regulatory pathways has profoundly shaped our understanding of the complexity of gene regulation. In eukaryotes, sRNAs have been found to control cellular metabolism, growth and differentiation, to maintain genome integrity, and to combat viruses and mobile genetic elements [2]. These regulatory sRNAs have been classified into at least six groups, including microRNAs (miRNAs), heterochromatic small interfering RNAs (hc-siRNAs), *trans*-acting small interfering RNAs (tasiRNAs), natural antisense small interfering RNAs (nat-siRNAs), repeat-associated small interfering RNAs (ra-siRNAs), and in metazoans, the piwi-interacting RNAs (piRNAs) [3-7].

* Correspondence: Zongrang.liu@ars.usda.gov
[2]Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
Full list of author information is available at the end of the article

miRNAs are derived from single-stranded RNA precursors that are transcribed by RNA polymerase II to generate self-complementary fold-back structures (stem-loop or hairpin) processed subsequently by DICER-like 1 (DCL1) in association with other protein factors [6,8]. Distinct from miRNA biogenesis, small interfering RNAs (siRNAs) are generated from long double-stranded RNAs that are converted from single-stranded RNAs by plant RNA-dependent RNA polymerases (RDRs), which usually give rise to transcript-wide, distinct siRNA species from both strands dependent on the choice of DCL proteins involved [8]. Biogenesis of predominant small 21-nucleotide siRNAs requires RDR6 and DCL4 while that of predominant 24-nucleotide siRNAs requires RDR2 and DCL3 [9,10]. miRNAs and siRNAs are incorporated into different RNA-induced silencing complexes (RISCs) [11], where one of the Argonaute (AGO) factors catalyzes sequence-specific endonucleotic cleavage of targeted gene transcripts [12,13] in the cases of miRNAs and 21-nucleotide siRNAs [11], or translational repression occasionally for some miRNAs

[14], or induction of DNA methylation in the case of 24-nucleotide siRNAs [15,16]. The biogenesis of tasiRNAs exploits both miRNA and 21-nucleotide siRNA biogenesis pathways, and requires all the factors necessary for miRNA and 21-nucleotide siRNA production, including DCL1, RDR6 and DCL4 as well as other protein factors [5,17]. In *Arabidopsis*, the transcript from a *trans*-acting siRNA (*TAS*) gene is first cleaved by one of three specialized miRNAs (miR173, miR390 and miR828), and then either the 3-cleaved (in the case of miRNA828, miR173) or the 5 -cleaved transcript fragments (in the case of miR390) are converted into double-stranded RNAs by RDR6 and subsequently diced into phased 21-nucleotide siRNAs by DCL4 to generate multiple but distinct tasiRNA species, some of which are found to further guide sequence-specific cleavage of their targeted gene transcripts through the RISC [5,17-20]. To date, only four *TAS* gene families have been identified in *Arabidopsis* and their biogenesis has been extensively characterized [21].

In plants, miRNAs are the second most abundant sRNAs [22], acting as powerful endogenous regulators. For example, many distinct miRNAs target transcripts encoding an array of transcription factors that control plant development and phase transition in *Arabidopsis*, maize and woody species [23-25], while others are involved in stress response and disease resistance [26-28]. In humans, it is estimated that at least 30% of genes are regulated by miRNAs [29], further underscoring their fundamental importance. Whether a similar proportion of plant genes are subjected to miRNA-mediated regulation is unknown; however, a large number of miRNAs have been identified, characterized and reported in diverse plant species, including *Arabidopsis* [30,31], rice [32], maize [33], poplar [34,35], grape [36], soybean [37], orange [38] and peanut [39]. The latest release of published miRNAs (miRBase 17) contains over 15,000 miRNA gene loci in over 140 species, and over 17,000 distinct mature miRNA sequences [40]. Like many gene regulatory systems, miRNAs show both conservation and diversity among plant lineages. Many miRNAs are conserved in angiosperms or even embryophyta [41], while a significant number of miRNAs or miRNA families show species-specificity, reflecting their fast evolving, functionally diverging natures [11,41-43].

Apple (*Malus domestica*) is a major temperate fruit crop worldwide. Its fruit is a widely consumed and rich source of phytochemicals, which may play a key role in reducing chronic disease risk in humans [44]. As a perennial species, apple undergoes many distinct developmental programs and inducible responses during its life cycle that cannot be easily replicated or investigated in annual model species such as *Arabidopsis*. For example, apple requires a long period of juvenility (5 to 7 years) before flowering [45] and its reproductive cycle lasts for nearly a year, as fruit forms from flower buds initiated during the previous summer. Its fruit development, which spans the spring, summer and fall seasons, comprises fruit enlargement, color changes, texture improvement and ripening, all of which are directly relevant to crop productivity and quality [45]. That apple trees produce fruit over a period spanning several decades is another important consideration for investigation of plant longevity. Thus, apple trees represent an important model for investigating the fundamental biology of a wide range of specialized strategies and programs to adapt or respond to seasonal and perhaps climatic changes as well as biotic and abiotic stress challenges while implementing multiple coordinated developmental events necessary for perennial fruit production. In addition to its importance as a new genomic model for tree fruit and Rosaceae studies, the discovery of genetic mechanisms that regulate fruit development and quality or stress responses and disease resistance could enhance the molecular breeding of apple for horticulturally important traits.

Although several groups have recently reported bioinformatic prediction and identification of a few miRNA families for apple [46,47], nearly all those reported are conserved miRNAs. Whether apple has evolved specific miRNAs and unique regulatory networks, which genes they target, if any, and what biological processes they regulate remain largely unknown. In this study, as part of a long-term goal to elucidate the role of miRNAs in apple, we exploited deep sequencing and computational and molecular analyses to comprehensively identify the conserved and apple-specific miRNAs and their targets, and characterized their expression in various tissues. We also delineated novel miRNA- and tasiRNA-mediated regulatory networks that modulate a large number of genes inside and outside the *MYB* family, which has not been reported in other species.

Results

Identification of conserved and less-conserved miRNAs in apple

A total of 59 million reliable reads were obtained from deep sequencing of leaf-, root-, flower- and fruit-derived sRNA libraries, and most of these reads (about 78% for redundant reads and 65% for unique reads) have at least one perfect match to the apple genome (Table S1 in Additional file 1). sRNAs from each library or tissue shared a more or less similar distribution pattern (Figures S1a-f in Additional file 2), with 24-nucleotide sRNAs (over 40%) being the most abundant followed by 21-nucleotide sRNAs, consistent with earlier findings in *Arabidopsis* [30], tomato [48], and soybean [37]. Those reads (46 million) that perfectly mapped to the apple genome were subjected to further analyses for miRNA identification. As a result, we identified 23 miRNAs or

families (Figure 1a) that are conserved in both angiosperms and coniferophyta lineages [41]; we refer to these as conserved miRNAs in this study. These miRNAs bore a canonical stem-loop structure in their precursors (Table S2 in Additional file 1). Expression levels of the conserved miRNAs, as reflected by normalized reads (reads per million genome-matched reads (RPM)), showed a great variation among families in all four tissues. The highest read abundance (166,000 RPM) was detected for miR156 and was 5 to 16 times more than other relatively abundant miRNA families, including miR165/166, miR167, miR396, and miR159, whose total abundance ranged from 10,000 to 30,000 RPM (Figure 1a; Table S3 in Additional file 1). Although lower expression (between 1,500 and 4,000 RPM) was observed for the miR162, miR164, miR168, miR172 and miR399 families, their overall expression level was 3 to 20 times greater than any of the remaining 12 conserved miRNA families (Figure 1a; Table S3 in Additional file 1). Almost all miRNAs showed, to various degrees, differential expression among the tissues analyzed, with the greatest variation observed for miR156, which was expressed at an abundance of more than 150,000 RPM in root but only 184 RPM in fruit (Figure 1a). Most of the conserved miRNA families comprised multiple species of different mature sequences (≤2 mismatches) and distinct length predominance (Tables S3 and S4 in Additional file 1).

We also identified ten miRNAs or families that have a canonical stem-loop structure (Figure 1b; Table S2 in Additional file 1) and were reported in some plant species or families but are not widely conserved in both angiosperm and coniferophyta lineages [41]. They are referred to as less-conserved miRNAs in this study. Compared to the conserved miRNAs, most of the less-conserved miRNAs exhibited relatively lower expression, with the most notable exception being miR535, which was expressed at an abundance of more than 20,000 RPM (Figure 1b). However, these less-conserved miRNAs, like the conserved miRNAs, were differentially regulated among the tissues examined. For example, leaf- and root-biased expression was observed for miR535, while flower-biased expression was apparent for miR828 (Figure 1b). Surprisingly, miR4376 exhibited

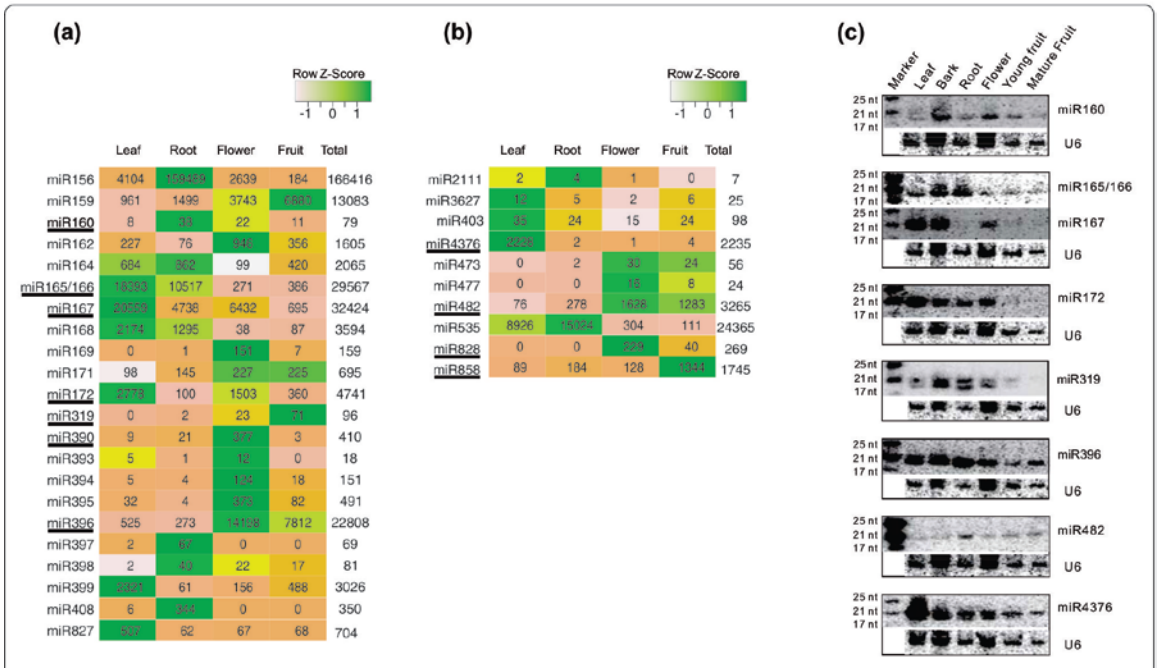


Figure 1 Analysis of the identified conserved and less-conserved miRNA expression in apple tissues. (a, b) Heat map and read abundance of the conserved (a) and less-conserved (b) miRNAs in four tissues. The read abundance was normalized and expressed as reads per million (RPM) of genome-matched reads in each tissue. Only reads having two or fewer mismatches to miRNA sequences were counted for this analysis. Read abundance is also denoted by color, as illustrated at the top-right corner of each panel. miRNAs selected for validation by RNA gel blot analysis are underlined.**(c)**RNA gel blot analysis of selected miRNAs in different tissues. Sizes of small RNA markers are indicated at left in nucleotides (nt). Blotting results from probing and reprobing the same filter were grouped together. The apple*U6*gene serves as a loading control.

virtually exclusive expression in leaf tissues where 2,228 RPM were detected in comparison with less than 4 RPM in other tissues.

To validate miRNA RPM data, we performed RNA gel blot analysis for selected miRNAs representing conserved, less-conserved and apple-specific (discussed below) examples in six different tissues, four of which (leaf, root, flower and fruit) could be compared to sRNA sequencing data (Figures 1c, 2, and 3c). We found that while blotting results for some miRNAs - miR828, miR858, miR390 (Figure 3c) and miR4376 (Figure 1c) - were reflective of the relative abundances of sequenced RNAs from these four tissues, many others displayed varying degrees of divergence between the two analyses. For example, miR172 RPM values and blot signals for leaf and flower were in agreement, while the blot signal for fruit, which should be nearly four-fold higher than for root, based on RPM values, was barely detectable. Additionally, miR396, which showed relative blot signal strength that was high for vegetative and low for reproductive tissues, revealed the opposite pattern through RNA sequencing (Figure 1a). At present we do not know why some RNA sequencing values were corroborated by RNA blots while others were not. However, contradictions between *in vivo* RNA levels and sequencing results for miRNAs have been previously reported for*Arabidopsis* [30] and grapevine [36]. Since the hybridization signal of RNA gel blotting is proportional to gene transcript abundance in general, the bias introduced against or for certain sequences or sequence motifs during either library construction or sequence amplification or deep sequencing may have contributed to the observed deviations.

Apple-specific miRNAs

Since numerous family- or species-specific miRNAs considered to be of a more recent evolutionary origin [11] have been identified in other species, apple is likely to have evolved unique miRNAs as well. After excluding sRNA reads homologous to known miRNAs (two or fewer mismatches, miRBase 17) and other non-coding sRNAs (Rfam 10), the remaining 20- to 22-nucleotide sRNAs were subjected to rigorous secondary structural analysis of their precursors using RNAfold software [49]. Those precursors with a canonical stem-loop structure were further analyzed through a series of stringent filter strategies to ensure that they met common criteria established by the research community [50,51]. A total of 42 miRNA candidates derived from 75 loci (Tables 1 and S5 in Additional file 1) met the screening criteria, of which 21 had miRNA star (miRNA*) sequences identified from the same libraries, while the other 21 had no miRNA* identified (Table 1). We considered the 21 candidates with miRNA* sequences as apple novel miRNAs and the remaining 21 without miRNA* sequences as apple miRNA candidates. Collectively, we term them apple-specific miRNAs. Of the

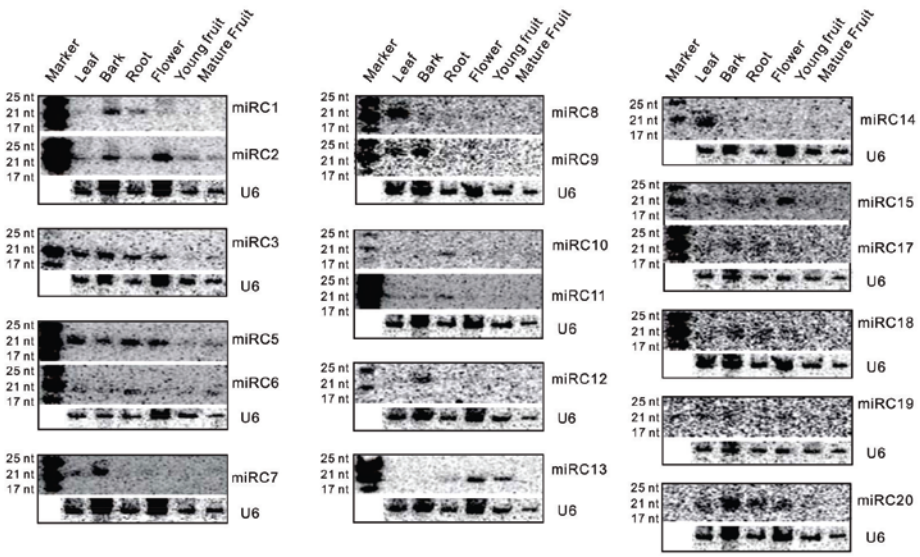


Figure 2Confirmation of apple-specific miRNA expression by RNA gel blot analysis. RNA gel blot analysis was performed as described in Materials and methods with exposure time varying from a few hours to two to three days for some low-abundance miRNAs (miR18, miR19, miR20). The blotting results from the same filter were grouped together. Sizes of small RNA markers are indicated at left in nucleotides (nt). The apple*U6*gene serves as a loading control.

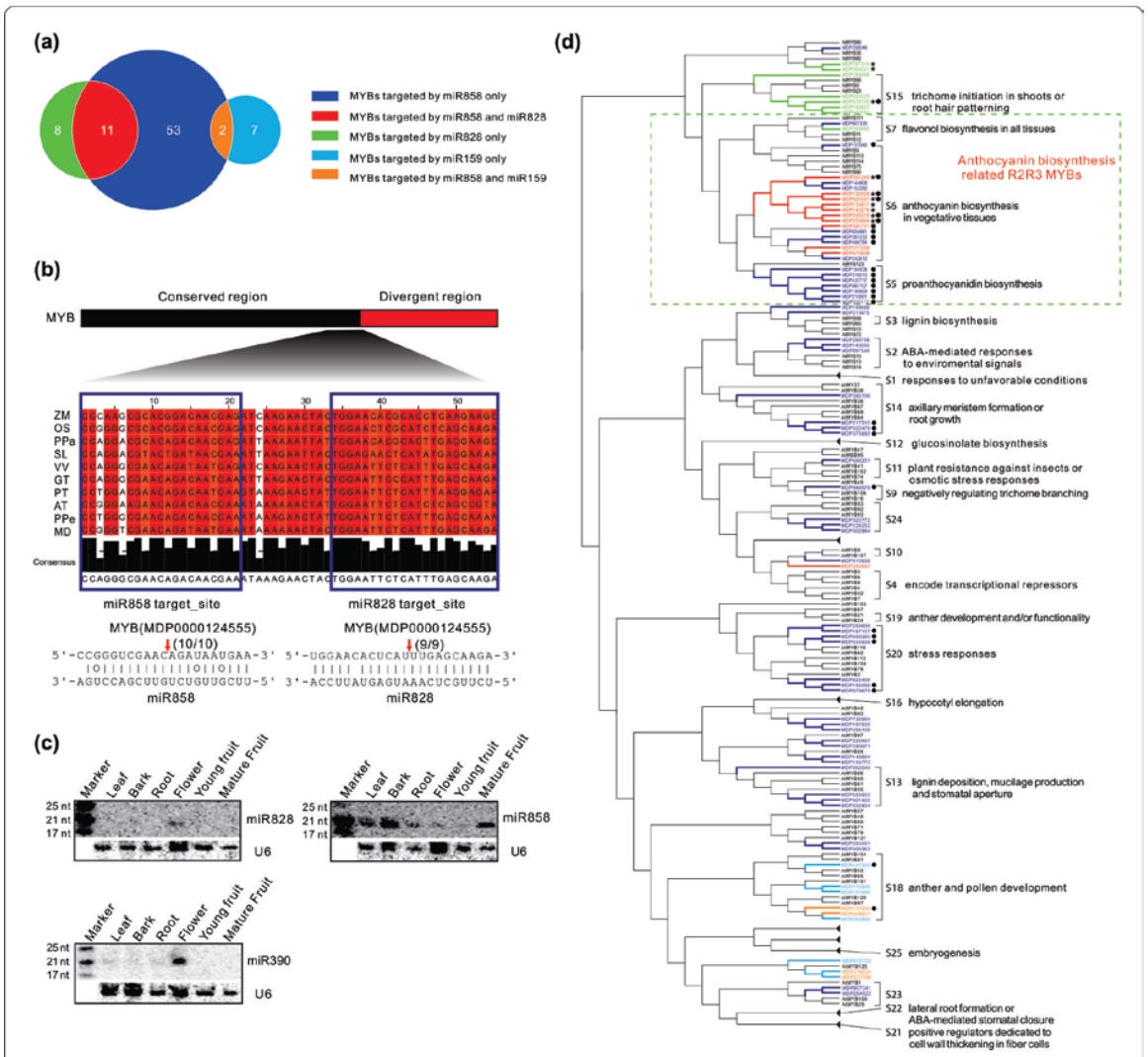


Figure 3 Complex *MYB* regulatory network mediated by miR159, miR828, and miR858. **(a)** Numbers of *MYB* genes targeted by apple miR159, miR828 and miR858. **(b)** Conservation of the region co-targeted by miR828 and miR858 among diverse plant species, including *Zea mays* (ZM; BT088210), *Oryza sativa* (OS; NM_001050499), *Physcomitrella patens* (PPa; XM_001781987), *Solanum lycopersicum* (SL; AK322126), *Vitis vinifera* (VV; XM_002269386), *Gossypium tomentosum* (GT; EU249416), *Populus trichocarpa* (PT; XM_002315854), *Arabidopsis thaliana* (AT; NM_116126), *Prunus persica* (PPe; ppa016135m), and *Malus domestica* (MD; MDP0000124555). The sequence-conserved region in the *MYB* coding region is shown in black while the sequence-divergent region is shown in red. Pairing of miR828 and miR858 with their target sites in a representative *MYB* transcript is illustrated below. **(c)** RNA gel blot analysis of expression of miR390, miR828 and miR858 in various tissues. **(d)** Phylogenetic analysis and functional relationship between apple and *Arabidopsis* MYB factors. The tree was inferred using the neighbor-joining method and 1,000 bootstraps with MYB amino acid sequences and Clustal X2 software. The subgroup and function annotation were designated as described previously [56,58]. Subclans consisting of only *Arabidopsis* MYBs were collapsed are and denoted with black triangles. The *MYB* genes that are targeted by a specific miRNA or co-targeted by two miRNAs are denoted in a specific color as designated in (a). miR828-targeted *MYB*s capable of producing secondary siRNAs are marked with a star, and *MYB*s whose cleavage was confirmed by degradome analysis are marked with filled circles. ABA, abscisic acid; nt, nucleotide.

42 apple-specific miRNAs, 32 belong to the 21-nucleotide class of miRNAs and 10 to the 22-nucleotide class (Table 1). In general, the apple-specific miRNAs were much less abundant compared to the conserved miRNAs in all tissues examined. For example, only miRC1 displayed total read abundance above 20,000 RPM, while 33 of the 42 miRNA candidates yielded levels below 100 RPM (Table 1). This low level expression was further confirmed by

Table 1 Novel or candidate miRNAs found in apple (excerpted[a])

Name	miRNA sequence	Contig[b]	Len	Str	Match position	miRNA* sequence[c]	Normalized reads				
							Leaf	Root	Flower	Fruit	Total
miRC1	ACAGGGAAGAGGTAGAGCATG	MDC006505.260	21	-	3,797	ATGCACTGCCTCTTCCCTGGC	1,091	23,246	0	12	24,349
miRC2	ACCTAGCTCTGATACCATGAA	MDC018599.370	21	+	9,369	TGTGGTATCAGGACTATGTTA	893	261	5,364	1,823	8,341
miRC3	CTACCGATGCCACTAAGTCCCA	MDC017130.228	22	+	6,809	GGACTTAGTAGCTCGGTGA	266	247	1,567	671	2,751
miRC4	TGTTATATTGTCAGATTGTCA	MDC019554.272	21	-	10,154	ACAGTCTGACAATATAACGTG	1,035	502	6	0	1,543
miRC5	AATGGAAGGGTAGGAAAGAAG	MDC006350.123	21	-	2,997	TCTTTCCTATCCCTCCCATTCC	1,021	324	144	0	1,490
miRC6a	TCCTCTTGGTGATCGCCCTGT	MDC009272.709	21	-	4,184	AGGGTGATTAACAAAGGGATG	359	279	336	319	1,294
miRC6b	TCCTCTTGGTGATCGCCCTGC	MDC006081.961	21	-	466	AGGGTGGTTACCAATGGGATG	122	218	0	46	385
miRC7	TTATACAGAGAAATCACGGTCG	MDC009778.59	22	-	4,735	ACCGTGTTTTTCTGTATAAAG	511	624	108	27	1,270
miRC8	AAGAGCGGGATGTGTAAAAGG	MDC001018.301	21	+	2,490	CTTTTACCTATCCCATTCTGT	248	0	0	0	248
miRC9	TCTGTCGTGGGTGAGATGGTGC	MDC011178.406	22	-	16,699	TTCATCTCTCCTCGACAGAAG	137	90	0	4	231
miRC10	GAATTCCTTCTCCTCTCCTTT	MDC026449.10	21	-	340	AGGAGGGAGAGAGGGTTTTAC	0	0	166	7	173
miRC11	CACCAATATCAACTTTATTTG	MDC005581.168	21	+	2,657	AATAAAGTTGATATTGGTGTG	6	12	44	38	100
miRC12	TCTGTCGAAGGTGAGATGGTGC	MDC003092.251	22	+	8,878	TTCATCCCTCCTCGACTGAAG	10	64	0	0	73
miRC13	ATCCAACGAAGCAGGAGCTGA	MDC009318.175	21	+	3,804	AGCTGCTGACTCGTTGGTTCA	0	0	18	42	60
miRC14	CGAACTTATTGCAACTAGCTT	MDC008558.180	21	+	6,507	CAAGCTAGTTGTAATAAGTTC	24	0	13	0	38
miRC15	AAAGTATCAAGGAGCGCAAAG	MDC014075.204	21	-	26,060	TTGCGTTCCACTGATTCTTTCG	8	8	9	2	27
miRC16	CATCTGGGTCGTTCAAATTTA	MDC009589.307	21	-	1,975	AATTTGAACGGCCCAGATGGG	10	12	0	0	22
miRC17	ATCATGCGATCCCTTCGGACG	MDC005391.194	21	+	11,429	TCCAAAGGGATCGCATTGATCT	11	0	0	0	11
miRC18	ATACTCATCGAATTTGTCATA	MDC012422.128	21	+	849	TGACAAATTGGATGAGTATTC	3	0	4	0	8
miRC19	TGGGATGTTGGTATGGTTCAA	MDC016463.170	21	-	5,187	GAGCCGTGCCAATATCACAGT	0	7	0	0	7
miRC20	TGAAGAGAAGAGCGTTGTTTGG	MDC001494.456	22	-	33,010	TGACAGCCTCTTCTTCTCATG	6	0	0	0	6
miRC21	ATCATTAACACTTAATAACGA	MDC006081.432	21	+	219	TTATTAAGTGTTAATGATTGG	0	0	0	2	2
miRC22	CCATATGTCCCTCCATATACT	MDC016302.308	21	+	5,368	No star found	0	0	72	8	80
miRC23	AATGATGATCAAACAACCCTT	MDC020884.221	21	+	4,086	No star found	0	0	46	2	48
miRC24	TGAACTTGGCTGAATGTGGACT	MDC001394.253	22	-	378	No star found	22	19	0	0	40
miRC25	TTTCGGAACCACTTACACCCA	MDC017130.228	21	+	9,764	No star found	0	0	24	0	24
miRC26	TCCCCAAAACCCCTCATTCCAA	MDC017130.228	22	+	9,967	No star found	0	0	15	0	15
miRC27	TTGCCAAAGGAGATCTGCTCAG	MDC019485.283	22	-	7,997	No star found	6	0	0	6	12
miRC28	TGCATTTGCACCTGCACTTGT	MDC007946.169	21	-	672	No star found	0	0	9	3	11
miRC29	CAAAGCTTTTAATATCAGTCGA	MDC018873.313	22	-	10,447	No star found	0	0	5	6	10
miRC30	TCCCTCAAGGGCTTCCAATATT	MDC004268.215	22	+	9,006	No star found	0	0	0	10	10
miRC31	TCCATAATTTTTCCAGATCAA	MDC005072.383	21	+	9,885	No star found	0	0	8	0	8
miRC32	TGGTTTGGTTGGAAAACGGCT	MDC006935.286	21	-	27,531	No star found	0	7	0	0	7
miRC33	AATTAGGCTGGCATTAGACAA	MDC009589.310	21	+	3,312	No star found	6	0	0	0	6
miRC34	TGGTGATAGGATAGTTGGAAG	MDC010150.221	21	-	29,453	No star found	6	0	0	0	6
miRC35	TACTGTTATAATGGCATTCCC	MDC001086.52	21	+	12,332	No star found	5	0	0	0	5
miRC36	CTCAATTTGAACGCGTGGCTA	MDC015454.116	21	+	3,773	No star found	0	5	0	0	5
miRC37	TGGCCTTGGTGGAAGAGATCC	MDC000614.265	21	-	5,670	No star found	0	4	0	0	4
miRC38	TGGGCCTGGTCAGGAGGATCC	MDC018501.179	21	-	2,535	No star found	0	4	0	0	4
miRC39	TTAAATACAAGCAGGAGCTCT	MDC011810.169	21	+	42,284	No star found	0	3	0	0	3
miRC40	CACCTGGGACTTGCAGCCATG	MDC009540.166	21	+	1,015	No star found	3	0	0	0	3
miRC41	CATCCGAATTACCAATAACTG	MDC009746.72	21	+	18,725	No star found	0	0	3	0	3
miRC42	ATAGATGGAAGCTACCAACCC	MDC013676.252	21	+	6,707	No star found	0	0	0	3	3

[a]Detailed information is listed in Table S5 in Additional file 1. [b]Only one matched locus is shown in this table. Others are included in Table S5 in Additional file 1. [c]The star sequence may have multiple variants. Only the best star sequence is shown in this table. Len, length; Str, strand.

RNA gel blot analysis showing that signal was detectable for only 18 of 42 apple-specific miRNAs (Figure 2). Almost all of the apple-specific miRNAs exhibited differential expression among tissues (Table 1 and Figure 2). For example, miRC1, miRC2, miRC5, miRC6, miRC9, miRC14, miRC15, miRC17, miRC18 and miRC20 showed preferential accumulation in either one or two tissues while miRC8 was exclusively expressed in leaf (Table 1

and Figure 2). As reported above for conserved and less-conserved miRNAs, RPM values for selected apple-specific miRNAs corresponded to relative signal intensity observed in RNA gel blots in some cases (miRC1 and miRC2), but several cases of divergence were observed as well. For example, miRC3 was the second-most abundant miRNA in fruit, miRC7 the most abundant in root and miRC10 was exclusively expressed in flower (Table 1), but RNA gel blots showed no or barely detectable signals for these three miRNAs in those tissues (Figure 2). As noted above for conserved and less-conserved miRNAs, RNA blotting revealed that the majority of the tested miRNAs were abundant in bark tissue from young seedlings, while very few were highly expressed in fruit (Figure 2).

Targets of known and apple-specific miRNAs

To identify gene targets for the known (both conserved and less-conserved) and apple-specific miRNAs reported here, we performed degradome sequencing to generate a total of 21 million short reads representing 5 ends of uncapped, poly-adenylated RNAs. About 65% of the unique reads can be perfectly aligned to the apple transcriptome [52]. These reads were subsequently screened and analyzed with the software Cleaveland 2.0 [53,54]. A total of 118 targets that fell into 5 categories (0 to 4) were identified (Table 2; Table S6 in Additional file 1), with 62 targets for 14 of the 23 conserved, 38 for 5 of the 10 less-conserved, and 18 for 8 of the 42 apple-specific miRNAs or families (Table 2; Table S6 in Additional file 1).

Among these targets for the conserved miRNA families, 13 fell into in category 0, which represented the most abundant degradome tags corresponding to the cleavage site and matching cognate transcripts, and 25 of them into category 2, whose cleavage abundance was higher than the median but below the maximum. The number of identified gene targets varied for different miRNAs, ranging from one to nine (Table 2; Table S6 in Additional file 1), but those miRNAs that targeted members of a gene family usually had more targets. For example, miR156 could target nine members of the squamosa promoter-binding-like protein family, and miR167 targeted six members of the auxin response factor (ARF) family (Table 2; Table S6 in Additional file 1). Although most of the genes (54 of 62) identified were the conserved targets for these miRNAs across a wide range of plant species, a few of them (8 of 62) had not been reported in other species. For example, miR319, which is known to target *TCP4* in other species, was found to target two genes coding for GDP-mannose 3,5-epimerase. Similarly, miR396, which exclusively targeted several members of the growth regulating factor (*GRF*) gene family in plants also targeted five IAA-amino acid hydrolase genes, three replicate factor C subunit 1 genes and one TIR-NB-LRR resistance gene. It was noted that a few identified apple-specific gene targets fell into category 4, which represents a low confidence group and might need to be further validated experimentally. Of the 38 targets identified for five less-conserved miRNAs or families, a single target was found for miR2111, miR3627, and miR535 (Table 2; Table S6 in Additional file 1). The remaining 35 targets identified were shared by miR828 and miR858, with the former targeting four *MYB* genes and *MdTAS4* and the latter targeting up to 30 genes, including 24 coding for MYB factors, 2 coding for mate efflux proteins and 3 coding for lipases (Table 2; Table S6 in Additional file 1). miR828 and miR858 have been shown to target *MYB*s in other species but their target number was very limited [36,55]. Finding an unusually large number of *MYB* targets for miR828 and miR858 suggests that they gained more diverse and broad regulatory roles in apple.

Gene targets were also identified for eight apple-specific miRNAs. Of the 18 gene targets identified, two belonged to category 0 and seven to category 2, while the remaining were classified into category 3 or 4 (Table 2; Table S6 in Additional file 1). Most of the apple-specific miRNAs, unlike their conserved counterparts, had relatively fewer gene targets with a higher alignment score. The apple-specific miRNAs, like conserved ones, targeted genes with diverse functions. For example, miRC5 targeted a gene coding for ARO4 protein while miRC42 targeted a gene encoding mitogen-activated protein kinase 2. miRC25 and miRC29 each targeted two members of gene families that code for cytochrome P450 and oligopeptide transporter 2, respectively. Further, miRC10 targeted up to six members of the translation initiation factor 2 subunit beta gene family and one protein kinase gene. Hence, these apple-specific miRNAs may be involved in regulation of an array of metabolic and biological processes and signaling pathways.

Three miRNAs target an unexpectedly large number of *MYB* genes in apple

The *MYB* gene family represents one of the largest families in plants, and some of its members are regulated by miRNAs [56]. In *Arabidopsis*, miR159, miR828 and miR858 were either predicted or confirmed to target at least 13 *MYB* genes [56,57]. Our degradome analysis confirmed they collectively targeted 29 *MYB*s (Table 2; Table S6 in Additional file 1, and Figure S2a, b in Additional file 2), which raised a question of how many genes these miRNAs actually targeted because the degradome analysis in this study identified less than 40% of the targets for the conserved miRNAs and an even lower percentage for the less-conserved and apple-specific miRNAs. To address the possibility that some *MYB* gene targets were missed during degradome analysis, possibly due to inactive or low levels of target gene expression in the plant tissues analyzed, we performed target prediction analysis in over 400

Table 2 Example targets for apple miRNAs (or families)[a]

miRNA	Target	AS[b]	RN_reads[a] (TPB)	Category[d]	Target gene annotation
Conserved targets for conserved miRNAs					
miR156	MDP0000146640	1	112.47	0	Squamosa promoter-binding-like protein
miR159	MDP0000147309	4.5	457.13	0	Transcription factor GAMYB
miR164	MDP0000121265	2.5	2,674.73	0	NAC domain-containing protein
miR165/166	MDP0000426630	3	212.24	0	Homeobox-leucine zipper protein
miR167	MDP0000137461	4	62.58	2	Auxin response factor
miR168	MDP0000161046	0	32.65	2	Argonaute protein
miR171	MDP0000151144	2.5	16.33	3	Scarecrow-like protein
miR172	MDP0000281079	3	212.24	0	Ethylene-responsive transcription factor RAP
miR319	MDP0000243495	3	16.33	2	Transcription factor TCP4
miR390	CN490861[e]				*MdTAS3*
miR393	MDP0000203334	2.5	2.72	3	Auxin signaling F-box protein
miR395	MDP0000121656	3.5	225.84	2	3-Phosphoadenosine 5-phosphosulfate synthase
miR396	MDP0000204597	4	32.65	0	Growth regulating factor (GRF)
tasiARFs	MDP0000179650	1	24.49	2	Auxin response factor
Other non-conserved targets for conserved miRNAs					
miR319	MDP0000296691	4.5	2.72	4	GDP-mannose 3,5-epimerase
miR396	MDP0000454027	4	5.44	4	IAA-amino acid hydrolase ILR1-like 6
miR399	MDP0000253476	4	21.77	3	Unknown
Targets for other known miRNAs					
miR2111	MDP0000416146	4	17.69	2	Unknown
miR3627	MDP0000941000	4	27.21	2	Amino acid transporter
miR535	MDP0000185769	4	27.21	3	Cysteine protease
miR828	MDP0000124555	1	69,205.65	0	MYB transcription factor
miR828	CN490819 (*MdTAS4*)	2.5	2,065.23	0	Non-coding mRNA
miR858	MDP0000140609	3	1,379.54	0	MYB transcription factor
miR858	MDP0000161125	4.5	351.01	0	Mate efflux family protein
miR858	MDP0000726382	4.5	2.72	3	Lipase family protein
Targets for apple-specific miRNAs					
miRC3	MDP0000485327	2	326.52	0	Unknown protein
miRC5	MDP0000863013	5	16.33	3	ARO4 (ARMADILLO REPEAT ONLY 4)
miRC6	MDP0000410000	5	16.33	2	Uncharacterized protein
miRC10	MDP0000294365	4	185.03	0	Protein kinase
miRC16	MDP0000737474	5	38.09	2	Uncharacterized protein
miRC25	MDP0000306273	4.5	10.88	2	Cytochrome P450 86B1
miRC29	MDP0000279462	4.5	13.60	2	Oligopeptide transporter 2
miRC42	MDP0000286564	5	27.21	2	Mitogen-activated protein kinase kinase 2

[a]A detailed list is included in Table S6 in Additional file 1. [b]The alignment score (AS) threshold was set to 4.5 and 5 for known and apple-specific miRNAs, respectively. [c]RN_reads, repeat normalized reads; TPB, transcripts per billion. [d]Category 0: > 1 raw read at the position, abundance at position is equal to the maximum on the transcript, and there is only one maximum on the transcript. Category 1: > 1 raw read at the position, abundance at position is equal to the maximum on the transcript, and there is more than one maximum position on the transcript. Category 2: > 1 raw read at the position, abundance at position is less than the maximum but higher than the median for the transcript. Category 3: > 1 raw read at the position, abundance at position is equal to or less than the median for the transcript. Category 4: only 1 raw read at the position. [e]Cleavage was predicted based on the distribution of tasiRNAs.

putative apple*MYB*s and identified an additional 8, 15 and 42 *MYB* genes with a cleavage-favorable alignment score (≤5) for miR159, miR828 and miR858, respectively. Thus, a total of nine*MYB*s for miR159, 19 for miR828 and 66 for miR858 were found, bringing the total number of *MYB*s potentially regulated by these miRNAs to 81 (Figure 3a; Table S7 in Additional file 1). We also found that miR858 shared 11 targets with miR828 and two with miR159 (Figure 3a; Table S7 in Additional file 1), but no common target was identified for miR828 and miR159.

MYB proteins are divided into four classes, 1R, 2R (R2R3), 3R (R1R2R3) and 4R, depending on the number of adjacent repeats homologous to R1, R2 and R3 in the animal c-Myb [56], but most MYBs in plants belong to

the R2R3 class [58,59], many of which share a very similar genomic organization and protein structure with a conserved region at the 5 end and a divergent one at the 3 end (Figure 3b). Out of the 81 *MYB* genes that we confirmed or predicted as miRNA targets, 67 belonged to the R2R3 class. The miR159 target site was found to locate in the sequence-divergent region, while the miR858 and miR828 target sites both mapped to a 55-nucleotide region in the conserved coding region upstream of the divergent region, and the two sites were separated by a 12-nucleotide fragment with the position of the miR858 target site at the 5 end and that of miR828 at the 3 end (Figure 3b). The dual cleavage by miR858 and miR828 was confirmed in one of the targeted *MYB*s (MDP0000124555) by RNA ligation-mediated 5 rapid amplification of cDNA ends (RLM-5-RACE) analysis (Figure 3b). Strikingly, this 55-nucleotide fragment encompassing the miR858 and miR828 targeted sequences and 12-nucleotide spacer was found to be highly conserved across a wide range of dicots and monocots (Figure 3b). The finding that miR828 and miR858 co-targeted a group of *MYB* genes prompted us to examine whether they were co-expressed or differentially regulated among apple tissues. Figures 1b and 3c show that miR828 and miR858 exhibit a distinct expression pattern that was generally corroborated by both RNA gel blots and RNA sequencing. miR828 was specifically expressed in flower while miR858 accumulated in all tissues tested, but was found to be most abundant in mature fruit (Figure 3c), suggesting that miR828 and miR858 differentially regulated their co-targeted *MYB*s in different tissues.

Potential functions of the miRNA-targeted *MYB*s in apple

In *Arabidopsis* , the R2R3 *MYB* gene family comprises 25 subgroups and includes many members that have been functionally characterized and conserved between divergent species [56-58]. These previous characterizations could be instrumental for deciphering the function of apple *MYB*s and the possible regulatory roles of these *MYB*-targeting miRNAs in apple. We performed phylogenetic analysis for those 81 miRNA-targeted apple *MYB*s with *Arabidopsis* R2R3 *MYB*s to investigate their potential functions (Figure 3d). Six of the nine miR159-targeted *MYB*s were placed into *MYB* subgroup 18 involved in anther and pollen development, while the remaining three were close to subgroup 25, which is associated with embryogenesis in *Arabidopsis* (Figure 3d). Hence, miR159 may regulate male organ and embryo development and growth in apple. The 19 miR828-targeted *MYB*s were related to three subgroups: S6, S7 and S15 (Figure 3d). Subgroups S6 and S7 were shown to directly or indirectly control anthocyanin biosynthesis in plant tissues, while S15 *MYB*s are involved in regulating trichome initiation and root hair patterning. Notably, most of the miR828-targeted *MYB*s are linked with primary and secondary metabolism related to anthocyanin production and color development. The 66 *MYB*s targeted by miR858 represent at least 14 subgroups shown to regulate diverse biological processes and metabolism pathways relevant to cell wall formation, lignification, anthocyanin biosynthesis, cell fate and identity, plant development and response to biotic and abiotic stresses in *Arabidopsis*. Nine of the ten *MYB*s co-targeted by miR828 and miR858 cluster together within subgroup 5, which is involved in the regulation of proanthocyanidin biosynthesis (Figure 3d). Thus, the roles of miR858-mediated regulation of *MYB*s in apple are predicted to be much broader than those for either miR828 or miR159. Of the 81 *MYB*s analyzed, the 29 *MYB*s confirmed as targets by degradome analysis fell into at least seven subgroups (S6, S5, S9, S14, S15, S18 and S20), with the majority of the confirmed *MYB*s clustered with the S5 and S6 groups, which are primarily involved in anthocyanin biosynthesis (Figure 3d).

The co-targeting sequence of miR828 and miR858 is located in the region encoding the conserved R3 repeat domain of MYB proteins

That miR828 and miR858 targeted substantially different numbers of *MYB* genes despite the adjacent location of their target sites prompted us to examine conservation profiles of their target sequences at both the amino acid and nucleotide levels (Figure 4). We found that the 18 amino acid polypeptide encoded by the 55-nucleotide sequence that bears both miR828 and miR858 target sites was located in the conserved R3 DNA binding domain of MYB factors (Figure 4a). Homology searching against the whole apple proteome using the 18 amino acid polypeptide obtained a total of 251 apple MYB factors containing this signature sequence, with 209 belonging to the R2R3 group and 4 and 38 belonging to the R1R2R3 and R3 groups, respectively (Figure 4a). The R3 domain consists of three α-helices (Figure 4a, c), and the third helix (H3) in each MYB repeat domain makes direct contact with its DNA target with the assistance of the first and second helices (H1 and H2) in basic helix-loop-helix (bHLH) motif folding (Figure 4a, c) [60,61]. Among the three helices in the R3 domain, the H3 helix that encompasses ten amino acid residues was most conserved among all MYB factors analyzed (Figures 4c; Figure S3a in Additional file 2). Of the 18 amino acid residues, the first seven (1 to 7) encoded by the 21-nucleotide miR858 target site were located in the highly conserved region covering three amino acid residues upstream and four amino acid residues at the 5 end of H3 while most of the last seven (12 to 18) encoded by the miR828 target site were located in the much less conserved region downstream of H3 (Figure 4c, d). Similar homology searching in

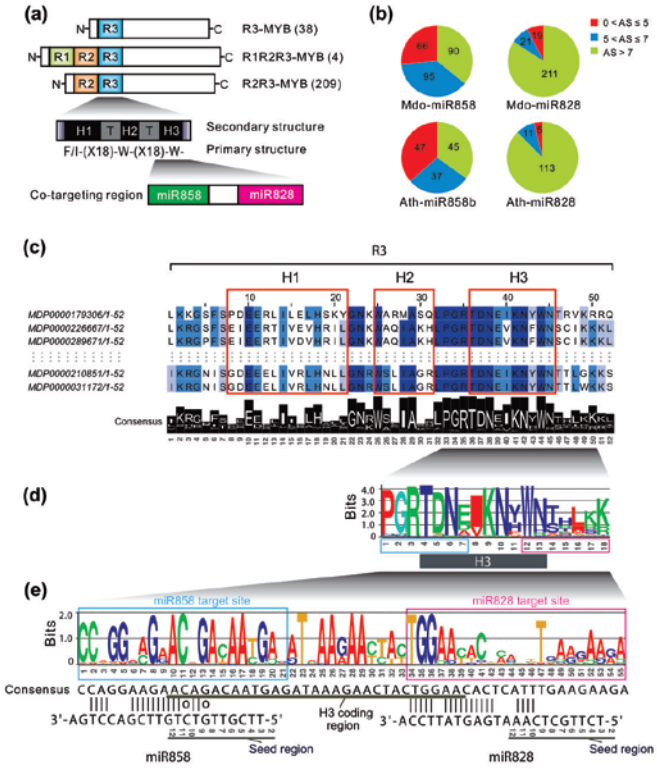


Figure 4The co-targeting sequence of miR858 and miR828 is located in the region encoding the R3 repeat domain of MYB proteins. **(a)**Schematic diagram of MYB proteins, and the location of the co-targeting region of miR828 and miR858. The MYBs containing a single or multiple R domains (R1, R2 and R3) are indicated. H, helix; T, turn; W, tryptophan; X, amino acid.**(b)**Distribution of alignment score (AS) between miRNAs (miR828 and miR858) and their target sites among 251 apple (upper pie charts) and 129*Arabidopsis MYB*genes (lower pie charts). AS categories:≤5, cleavage-favorable; 5 < AS≤7, less cleavage-favorable; > 7, cleavage-unfavorable.**(c)**Conservation of the H3 helix and its flanking sequences among 251 apple MYB proteins. The conservation of amino acid residues in each position is denoted with various colors: dark color indicates a high level; light color indicates a low level of conservation.**(d)**Conservation of 18 amino acid residues encoded by the miR828 and miR858 co-targeting region. The bit value representing the conservation in each position is indicated.**(e)**Conservation of the 55-nucleotide co-targeting region in 251 apple*MYB*genes. The pairing between the consensus sequence and each miRNA is illustrated. miRNA target sites, the H3 coding region and the miRNA seed region are also marked.

Arabidopsis found that 129 MYBs, including 124 R2R3 and 5 R1R2R3 MYBs, bear a similar signature sequence in the R3 domain (Figure S3b in Additional file 2). Correspondingly, the miR858 target site was found to be more conserved than the miR828 target site at the nucleotide level in both apple and*Arabidopsis* (Figure 4e; Figure S3b in Additional file 2). This difference was particularly pronounced in a region (positions 10 to 20 in the miR858 target site, and 44 to 54 in the miR828 target site) that specifically pairs with the miRNA seed region (positions 2 to 12) (Figure 4e; Figure S3b in Additional file 2). Since pairing between the miRNA seed region and corresponding target site is critical for miRNA cleavage [62], the level of sequence conservation in this region could impact miR828- and miR858-targeted *MYB* populations.

Based on alignment scores ≤5, 66 and 19 apple *MYB*s were identified to be targeted by miR858 and miR828, respectively (Figure 3a). Given that many target sites with high alignment scores > 5 have been proven to be cleavable [63,64], the actual number of MYB targets is likely to be larger than what we reported in Figure 3a. Therefore, we further analyzed the alignment score distribution profiles for these two miRNAs among all the 251 apple *MYB*s, and found 95 of the 251 *MYB*s with a less cleavage-favorable alignment score (> 5 and≤7) and 90 with a cleavage-unfavorable alignment score (> 7) with miR858 (Figure 4b, top). In contrast, 211 of the 251*MYB*s showed a cleavage-unfavorable alignment score (> 7) with miR828 while only a very small portion of them had a cleavage-favorable or less cleavage-favorable alignment score (≤7) (Figure 4b, top). A similar pattern was observed among

129 *Arabidopsis MYB*s (Figure 4b, bottom). These results imply that the targeting capacity of miR858 and miR828 in apple and *Arabidopsis* might be even broader than those reported in Figure 3a, especially for miR858.

tasiRNA biogenesis pathways with unique features evolved in apple

To date, only four *TAS* families (*AtTAS1-4*) and three miRNAs (miR173, miR828, and miR390) that target *TAS* transcripts and trigger tasiRNA production have been reported and well characterized in *Arabidopsis* [5,18-20,30]. Both miR390 and miR828 were identified in apple (Figure 1a, b), and showed highest expression specifically in flower as detected by RNA blot and RNA sequencing methods (Figures 1a, b and 3C). A *TAS4* homolog, *MdTAS4*, was found in apple (Figure 5e), and degradome analysis showed that miR828 cleaved *MdTAS4* (Figure S4E in Additional file 2). Sequencing data showed that abundant 21-nucleotide sRNAs were produced along the 3 cleaved *MdTAS4* transcript, and most of those sRNAs belonged to the first (miR828 target site) and second register while some of them fell into the 12th register (Figure S4E in Additional file 2). Analysis of siRNA abundance in four libraries showed that *MdTAS4*-derived tasiRNAs primarily accumulated in flower tissues (Figure S4E in Additional file 2), which is in agreement with the flower-biased expression of miR828 in apple (Figure 3c). In *Arabidopsis*, siR81(-), one of the *AtTAS4*-derived siRNAs, was shown to target *AtMYB75*, *AtMYB90* and *AtMYB113*, which are associated with anthocyanin biosynthesis [30,65]. Our analysis also predicted that apple *TAS4*-siR81(-) potentially targeted at least three *MYB* homologs (data not shown), and degradome data confirmed that apple *TAS4* -siR81(-) targeted an additional gene (MDP0000225680) coding for a bHLH transcription factor (Figure 5e), which is also involved in the regulation of anthocyanin biosynthesis in apple [66]. Thus, apple *TAS4*-siR81(-) is likely to target both *MYB* and *bHLH* genes.

The apple genome was rich in *TAS3* homologs, and at least two *TAS3* gene families, termed *MdTAS3-1* and *MdTAS3-2*, were identified. *MdTAS3-1* has at least three homologs (Figure 5b). *MdTAS3-1a* and *MdTAS3-1b* share about 98% sequence identity, but each share only 80% sequence identity with *MdTAS3-1c*. Our deep sequence data show that both *MdTAS3-1a/b* and *MdTAS3-1c* have two miR390 cleavage sites flanking phased-tasiRNA generation regions (Figure 5a). The 5 target site bears a conserved mismatch in the tenth position, as its counterpart does in *Arabidopsis*, and should be non-cleavable, while the siRNA distribution data suggest that the 3 site could be cleavable and likely sets the phasing for the production of 21-nucleotide tasiRNAs (Figure S4a, b in Additional file 2). The two conserved tasiARFs are flanked by a constant tasiRNA-generation region at the 3 end and a divergent region at the 5 end (Figure 5a). These features are consistent with canonical features characterized for *AtTAS3* in *Arabidopsis* [20]. Interestingly, tasiRNAs from each member displayed leaf- and flower-biased accumulation despite the nearly exclusive expression of miR390 in flower (Figures 1a, 3c and 5d; Figure S4f in Additional file 2).

The *MdTAS3-2* family exhibited similar and distinct features relative to the *MdTAS3-1* family. The two *MdTAS3-2* homologs, *MdTAS3-2a* and *MdTAS3-2b*, are relatively short compared to the *MdTAS3-1* family and share about 85% sequence identity (Figure 5b). Like the *MdTAS3-1* family, both *MdTAS3-2a* and *MdTAS3-2b* transcripts have two miR390 cleavage sites flanking an approximate 190-nucleotide region for 21-nucleotide phased siRNA production (Figure S4c, d in Additional file 2). In contrast to *MdTAS3-1*, *MdTAS3-2* family transcripts encode only one tasiARF, and there was no mismatch in the tenth position of the 5 miR390 target site (Figure 5c). Moreover, these short *TAS3* genes are absent in *Arabidopsis* but conserved in many dicots sharing the presence of the dual miR390 target sites and production of a single tasiARF (Figure 5c). Unlike *MdTAS3-1*, the regions flanking the *MdTAS3-2* tasiARF are orientationally reversed, with a constant region at the 5 end and a divergent region at the 3 end (Figure 5c). Based on the siRNA distribution pattern, the 3 site of miR390 is predicted to set the phase for siRNA generation (Figure S4c, d in Additional file 2). Similar to *MdTAS3-1*, *MdTAS3-2*-derived tasiRNAs also preferentially accumulated in both leaf and flower tissues but with much less abundance (Figure 5d; Figure S4f in Additional file 2).

Earlier studies showed that *Arabidopsis TAS3* -derived tasiARFs directly target several *ARF* genes [67], which were proposed to act as suppressors in the auxin signaling pathway [68]. Our degradome data show that they guide the cleavage of at least three apple *ARF* transcripts (Figure S2c in Additional file 2). Homologous sequence alignment revealed that the cleaved apple *ARF*s are closely related to *AtARF2*, *AtARF3* and *AtARF4*.

miR828-activated, *MYB* transcript-derived siRNAs and their gene targets

The possibility that miR828-cleaved *TAS4* RNA fragments could be channeled into tasiRNA biogenesis [30,65] led us to examine whether all miR828-cleaved *MYB* transcripts are also subjected to tasiRNA biogenesis. With direct searching against small sequencing libraries, we were able to map a large number of sRNA reads to the coding regions of ten miR828-targeted *MYB*s. These *MYB*s share similar genomic organization with the location of the miR828 target site in the third exon just before the divergent region where siRNA biogenesis occurred (Figure 5f). Further analysis showed that the generated siRNAs were in phase with the miR828 cleavage site, and the siRNA

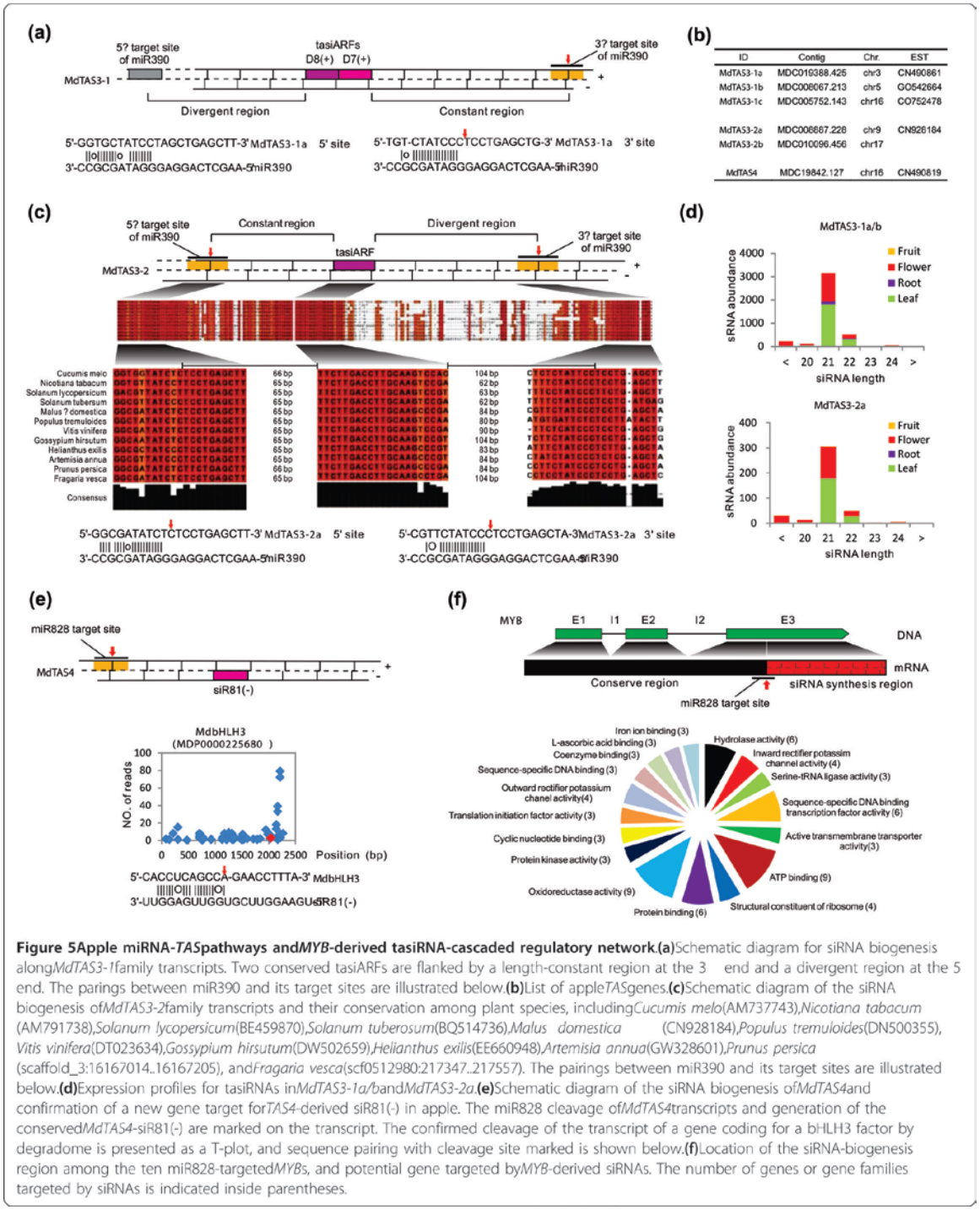


Figure 5 Apple miRNA-*TAS* pathways and *MYB*-derived tasiRNA-cascaded regulatory network. **(a)** Schematic diagram for siRNA biogenesis along *MdTAS3-1* family transcripts. Two conserved tasiARFs are flanked by a length-constant region at the 3 end and a divergent region at the 5 end. The pairings between miR390 and its target sites are illustrated below. **(b)** List of apple *TAS* genes. **(c)** Schematic diagram of the siRNA biogenesis of *MdTAS3-2* family transcripts and their conservation among plant species, including *Cucumis melo* (AM737743), *Nicotiana tabacum* (AM791738), *Solanum lycopersicum* (BE459870), *Solanum tuberosum* (BQ514736), *Malus domestica* (CN928184), *Populus tremuloides* (DN500355), *Vitis vinifera* (DT023634), *Gossypium hirsutum* (DW502659), *Helianthus exilis* (EE660948), *Artemisia annua* (GW328601), *Prunus persica* (scaffold_3:16167014..16167205), and *Fragaria vesca* (scf0512980:217347..217557). The pairings between miR390 and its target sites are illustrated below. **(d)** Expression profiles for tasiRNAs in *MdTAS3-1a/b* and *MdTAS3-2a*. **(e)** Schematic diagram of the siRNA biogenesis of *MdTAS4* and confirmation of a new gene target for *TAS4*-derived siR81(-) in apple. The miR828 cleavage of *MdTAS4* transcripts and generation of the conserved *MdTAS4*-siR81(-) are marked on the transcript. The confirmed cleavage of the transcript of a gene coding for a bHLH3 factor by degradome is presented as a T-plot, and sequence pairing with cleavage site marked is shown below. **(f)** Location of the siRNA-biogenesis region among the ten miR828-targeted *MYB*s, and potential gene targeted by *MYB*-derived siRNAs. The number of genes or gene families targeted by siRNAs is indicated inside parentheses.

generation pattern varied among the ten *MYB*s (Figure S5 in Additional file 2). Interestingly, despite the apparent flower-biased expression of miR828 (Figure 3c), distinct patterns of accumulation of phased siRNAs were observed among the ten miR828-targeted *MYB*s (Figure S5 in Additional file 2). We found that a total of more than 100 phased, sequence-distinct 21-nucleotide siRNA species were produced from the cleaved 3 transcripts of these *MYB*s (Figure 5f). To ascertain whether these siRNAs were able to guide the cleavage of other gene transcripts, we carried out degradome analysis with a stringent alignment score (≤4.5) and found that they could potentially target as many as 77 genes, including six *MYB* genes, and a diverse array of other genes encoding proteins such as potassium transporters, protein kinases, hydrolases, oxidoreductases, transcription factors and DNA-, protein- and ion-binding proteins (Figure 5f; Table S8 in Additional file 1).

Discussion

Apple miRNAs with conserved as well as new gene targets

A recent study reported 16 conserved and less-conserved miRNAs in apple based on bioinformatics prediction using EST sequences [47], which is far more limited compared to those identified in other plant species [30,35,36]. In this study, we employed deep sequencing and computational analyses to identify 33 known (23 conserved and 10 less-conserved) miRNA families and 42 apple-specific (21 novel and 21 candidate) miRNAs (Figure 1 and Table 1; Table S3 in Additional file 1), which provides, to date, the most comprehensive list of identified miRNAs in apple. The majority of these miRNAs displayed tissue-specific expression (Figure 1), which is consistent with a general scenario in which miRNAs are differentially regulated in fruit trees [36,38,46,47] and other species [30,35,69], although additional work is needed to resolve examples of apparently divergent results from RNA gel blotting versus sRNA sequencing, as noted in this report and elsewhere [30,36]. It is known that miRNAs are involved in regulation of leaf morphology and polarity, lateral root formation, hormone signaling, phase transition, flowering time, floral organ identity and reproduction, anthocyanin production and stress and pathogen response [8,65,70,71]. In apple, we have identified a total of 100 gene targets for 19 of the 33 known miRNAs using degradome analysis (Table 2; Table S6 in Additional file 1), and the majority of these targets are conserved in plant species, indicating broad conservation of the known miRNA regulatory roles in plants. However, a few of the known miRNAs, including miR319 and miR396, were found to target additional genes in apple that have not been previously reported, while others like miR828 and miR858 target an unexpectedly large number of *MYB* genes. Hence, while these known miRNAs conserve their gene targets, they also appear to have an expanded target gene population in apple.

Although many newly evolved miRNAs that may exhibit weak expression, imperfect processing and lack of targets are believed to serve no biological function, many of them have been shown to target and regulate specific genes or gene families in various species [36,64,72]. Eight of the 42 apple-specific miRNAs or candidates were also found to target specific genes, implicating these miRNAs in the control of signal transduction cascades, secondary metabolism and protein translation (Table 2; Table S6 in Additional file 1). Our inability to detect gene targets for the remaining apple-specific miRNAs or candidates may be due to a low level of expression or the stress- or developmentally inducible nature of their target genes.

Apple *TAS* gene families with unique features and target specificity

In *Arabidopsis*, four *TAS* families targeted by three miRNAs, including miR173, miR390 and miR828, have been characterized [5,17-20,30,63]. The miR390-*TAS3* pathway is highly conserved in the plant kingdom [20], and the miR828-*TAS4* pathway is widely represented in dicot species [65], but no miR173-*TAS1*/*TAS2* pathway has been found in other species besides *Arabidopsis*. In this study, we showed that apple conserved the miR390-*TAS3* and miR828-*TAS4* pathways with expanded features. Apple possesses an additional *MdTAS3-2* family that comprises two loci and transcribes short mRNA species with distinct structural organization of the siRNA generation region, which bears only one characteristic functional tasiARF, instead of the two in *MdTAS3-1* transcripts (Figure 5c). The *TAS3-2* family is not present in *Arabidopsis* but is widely conserved in many other dicot species (Figure 5c) [73]. Despite an extra *TAS3* family in apple, the derived tasiARFs were found to target similar *ARF* genes homologous to the *Arabidopsis* genes *AtARF2*, *AtARF3* and *AtARF4* (Figure S2c in Additional file 2), which negatively regulate auxin signaling [68]. Whether these genes are differentially targeted by *MdTAS3-1* or *MdTAS3-2* tasiARFs is difficult to determine. Nevertheless, the existence of more *TAS3* genes with distinct expression patterns could enable the auxin signal to be fine-tuned within a specific cell, tissue or developmental context.

One interesting feature of the apple miR828-*TAS4* pathway is that its derived tasiRNA, *MdTAS4*-siR81(-), targets an additional gene. In *Arabidopsis*, *AtTAS4*-siR81(-), a conserved siRNA derived from the phased siRNA production of *AtTAS4*, is shown to target at least three *MYB*s that positively regulate anthocyanin accumulation in response to environmental stresses [30,65,74]. In apple, besides the three predictable *MYB*s, *MdTAS4*-siR81(-) also targets a bHLH transcription factor (*MdbHLH3*) that interacts with

MdMYB10 (MDP0000259614) to regulate anthocyanin biosynthesis in apple [66].

Unique miR828-activated, siRNA-cascaded gene regulatory network and its potential biological function

One of the interesting findings in this study is that miR828 potentially targets up to 19 *MYB*s in apple, 10 of which are subjected to siRNA biogenesis, with production of over 100 diverse siRNA species from the diverged region of *MYB*s (Figure 5f). In *Arabidopsis*, miR828, which indirectly targets *AtMYB113* through *AtTAS4*-siR81(-), also directly targets *AtMYB113*[65]. Up-regulation of miR828, *AtTAS4* and *AtTAS4*-siR81(-) is correlated with that of their three direct or indirect targets (*PAP1/AtMYB75*, *AtMYB90* and *AtMYB113*) under phosphate (Pi) and nitrogen deficiency conditions [74], which appears to contradict the anticipated negative role of miR828 in regulation of anthocyanin production. Interestingly, the elevated expression of *PAP1/AtMYB75* (and possibly *AtMYB90* and *AtMYB113*) induces miR828 and *AtTAS4* expression, presumably through binding of these *MYB*s to the 5 *cis*-elements in *MIR828* and *AtTAS4* promoter regions [74]. Such autoregulatory feedback was proposed to maintain proper anthocyanin production under stress conditions [65,74]. Conceivably, the miR828-activated siRNA biogenesis in seven of the ten targeted *MYB*s that relate to anthocyanin accumulation in apple would reinforce this feedback regulation to ensure proper color appearance in a specific tissue or apple fruit during development. However, identification of over 70 genes as targets for the *MYB* -derived siRNAs suggests that they may function beyond the feedback regulation of anthocyanin accumulation (Figure 5f; Table S8 in Additional file 1). The identified targeted genes are predicted to code for proteins regulating diverse functions ranging from hydrolase, oxidoreductase and kinase activities, and iron transport to DNA-, ATP-, and co-enzyme-binding activities (Figure 5f; Table S8 in Additional file 1), suggesting that this miRNA-activated, *MYB*-dependent and siRNA-cascaded gene regulation might orchestrate major physiological or biochemical or secondary metabolism switches associated with anthocyanin production and the pigmentation process.

miRNAs as master regulators to regulate a large number of *MYB*s through targeting of conserved sequences

The finding that three miRNAs potentially target up to 81 different *MYB*s indicates that miRNAs can, like transcription factors, serve as master regulators to modulate expression and function of a large number of genes in plants. This unique regulatory network is primarily based on the high degree of sequence pairing between miRNAs and their targeting sites as well as the availability of the miRNA target sites among the *MYB* population. *MYB* genes typically share a conserved 5 region and diverge at their 3 end [56]. Conceivably, the miRNA-targeted sites residing in highly conserved functional domains would be necessarily preserved among the *MYB* population relative to those located in the divergent region. Consistent with this prediction, miR828 and miR858 target sites, which overlap the conserved R3 region, are found in more *MYB*s than the miR159 target site located in the divergent region (Figure 3a). Similarly, the miR858 target site, which overlaps the more highly conserved 5 end of the H3 domain, is conserved in more *MYB*s than the miR828 site, which overlaps the less conserved 3 end of the H3 domain (Figure 4c-4e). Thus, sequence conservation and divergence of the miRNA target sites could directly impact the miRNA targeted gene population within a gene family.

Although the footprint of the 55-nucleotide sequence encompassing both miR828 and miR858 target sites is detected in dicot and monocot species (Figure 3b), miR828 and miR858 emerged only in dicot species [41], indicating that miR828- and miR858-mediated regulation of *MYB* genes is a feature of dicot species, consistent with our finding that miR828 and miR858 target a large number of *MYB*s in both apple and *Arabidopsis* (Figure 3a). It is not clear why this regulatory network specifically occurs in dicots while large *MYB* families exist in monocots as well [75]. Currently, it is known that *MYB*s are differentially regulated by transcription factors [76] as well as a variety of post-translation interactions or modifications. Our new discoveries related to miRNA-mediated regulation of a multitude of *MYB*s strengthens our understanding of how apple and other dicot species integrate transcriptional, post-transcriptional and post-translational regulatory mechanisms to achieve exquisite spatiotemporal regulation of each member of the *MYB* family.

Intriguingly, miR858 was found to co-target 11 *MYB* s with miR828 and two *MYB*s with miR159 (Figure 3a), raising the question of whether the convergence of two miRNAs upon the same *MYB* genes is an evolutionary coincidence or conveys some biological significance. The latter possibility is favored by the fact that 10 out of 11 *MYB*s co-targeted by miR828 and miR858 are related to regulation of anthocyanin biosynthesis. miR828 and miR858 may either redundantly reinforce each other s silencing function or differentially regulate anthocyanin accumulation in various apple tissues. The detection of differential expression of miR828 and miR858 among various tissues appears to support their different regulatory roles (Figure 3c).

The targeting of multiple members of gene families by one or a few miRNAs is not unique to the *MYB* family. Recently, a similar regulatory strategy was reported for *NB-LRR* defense genes in *Medicago*, where three miRNAs

collectively target over 70 *NB-LRR* loci [77]. Since plant genomes have evolved many large gene families with unique sequence conservation features, such regulatory strategies could be conceivably adopted by various species to modulate large groups of genes. Further characterization of this mechanism for regulating multi-gene families among different species could provide insight regarding both their evolution and function.

Conclusions

We carried out extensive characterization of miRNAs, their targets and expression in apple and provide a comprehensive list of miRNAs identified. We show that apple conserves and has evolved a variety of miRNAs with distinct expression patterns, and these miRNAs target dozens of apple genes with a wide range of functions. The discovery of an additional short *MdTAS3* family suggests that miR390 and tasiARFs may play more complicated roles in the auxin signaling pathway. More importantly, we reveal the existence of two similar but distinct regulatory networks in apple: direct miRNA targeting of a large number of *MYB*s and miR828-activated and *MYB*-derived siRNA-cascaded targeting of 77 genes primarily outside the *MYB* family, which has not yet been reported in other species.

Materials and methods

Plant material

Malus domestica of Golden delicious, grafted on M.111 rootstock, was selected in an orchard located in the Alson H Smith Agricultural Research and Extension Center for tissue collection of leaf, flower and fruits. Root and bark tissues were collected from rapidly growing, two-year-old seedlings. Fruits were harvested at 15 and 120 days after anthesis (DAA).

RNA preparation and small RNA sequencing

Total RNA from different tissues was extracted using the Plant RNA Purification Reagent (Invitrogen, Grand Island, NY, USA). sRNA quantity and quality was evaluated by the Agilent 2100 Bioanalyzer. RNA samples of RNA integrity number (RIN) above 8 were sent to BGI (Hong Kong, China) for sRNA and degradome sequencing using standard protocols on the SOLID sequencing system or Illumina Hiseq 2000 platform.

Small RNA data analysis

Small RNA libraries were constructed and sequenced for four apple tissues. The GenBank Gene Expression Omnibus (GEO) accession number for the sequencing data is GSE36065. All the sequencing data were first processed by removing the 3 adaptor sequence using CLC Genomic Workbench 4.9 (CLC bio, Aarhus, Denmark). Any sequences without adaptor matches were excluded from further analyses. Reads homologous to non-coding RNAs and conserved miRNAs were removed by BLATN alignment against the Rfam 10 [78] and mature miRNAs collected in miRBase (release 17) [79], allowing up to two mismatches. The remaining sRNAs were subjected to new miRNA identification. Read mapping was conducted using Bowtie [80], and Vienna RNA package [49] was used for the secondary structure prediction of sRNAs. Only those sRNAs (20- to 22-nucleotide) with a good stem-loop structure (no more than four mismatches, and no more than one central bulge) and a miRNA/miRNA* pair accounting for more than 75% reads matching to the precursor locus were considered as potential miRNAs (Additional files 3 and 4). Detailed screening critera were set up according to Meyers *et al.* [51]. The total number of reads perfectly matching the apple genome in a given library was used for the normalization of read abundance, which was denoted as RPM (reads per million reads). Apple genome sequences were retrieved from the Genome Database for Rosaceae [52]. The R package was used for the construction of heat maps.

Multiple alignment, phylogenetic analysis and Gene Ontology annotation

Multiple alignment was conducted using CLUSTAL X2, with the coloration based on the residue identity (above 60%) [81]. All the apple *MYB* targets for miR828, miR858, and miR159 were predicted by Targetfinder 1.6 with the alignment score no more than 5. Amino acid sequences of 126 R2R3 and 5 R1R2R3 MYB factors in *Arabidopsis* were retrieved from TAIR [82] and the phylogenetic tree was inferred using the neighbor-joining method and 1,000 bootstraps with putative full-length sequences using CLUSTAL X2 [81]. The subgroup and function annotation were designated according to Dubos *et al.* [56]. Gene Ontology annotation for the target genes of the *MYB*-derived siRNAs was performed using Blast2GO [83] with the default settings.

The *MYB*s, including 251 from apple and 129 from *Arabidopsis*, were retrieved by using the conserved 18 amino acid sequence corresponding to the co-targeting region (PGRTDNEIKNYWNTHLKK) to blast against the Apple Genome V1.0 predicted peptides [52] and the *Arabidopsis* TAIR10 proteins [82] with an e-value of 100. The classification of MYB subfamilies (R3, R2R3, R1R2R3) was based on the quantity of conserved R-repeats identified by NCBI Conserved Domain Search [84]. The consensus nucleotide sequences of the co-targeting region were obtained by counting the most frequently appearing residue at the corresponding position. The alignment score of each target site was calculated according to the scoring algorithm established by Allen *et al.* [18]: mismatches and single-nucleotide bulges or gaps were assessed by a penalty of 1 while GU base pairs were assessed by a penalty of 0.5; and the penalty score from mismatches, bulges, gaps and GU

pairs for positions 2 through 13 was doubled. Sequence logos were produced by GENIO/logo [85].

RNA gel blot

For RNA gel blot analysis, 25 to 50μ g of total RNA from apple leaf, bark, root, flower, young fruit and mature fruit was separated on 15% denaturing polyacrylamide gel and transferred to Amersham HybondTM-NX membranes (GE Healthcare, Waukesha, WI, USA). RNA was cross-linked using EDC (N-(3-dimethylaminopropyl)-N -ethyl-carbodiimide hydrochloride (Sigma, St Louis, MO, USA). The probes of 21-nucleotide DNA oligonucleotides (Table S9 in Additional file 1) that are reverse complementary to apple-specific miRNA candidates were labeled with P^{32}-gamma-ATP by T4 polynucleotide kinase (NEB, Ipswich, MA, USA). A miRNA Marker Probe (21-nucleotide; NEB, Ipswich, MA, USA) was used for sRNA size determination. The prepared membrane filters were hybridized at 42 overnight, then washed twice at 55C with washing buffer containing 2 SSC and 2% SDS. Membranes were then exposed to phosphorscreens and scanned with a Typhoon TRIO Variable Mode Imager (GE Healthcare). Membrane exposure time was adjusted, dependent on signal intensity.

Degradome analysis

For degradome sequencing, mixed RNAs with equal amounts from leaf, root, flower and fruit tissues were used. After adaptor-trimming and genomic mapping as done for the sRNA data, the Cleaveland pipeline 2.0 [55] was optimized to analyze the degradome sequencing data in collaboration with Targetfinder 1.6 [86]. The alignment score threshold was set to 4.5 for conserved and less conserved miRNAs (except for two *ARF* targets of miR167 and two *MYB* targets of miR858, for which the score was 5) and to 5 for novel and candidate miRNAs. The apple consensus gene set and the annotation information of miRNA target genes were retrieved from Genome Database for Rosaceae (GDR). Degradome data were normalized to transcripts per billon (TPB).

RLM-5-RACE

Following the manufacturer s instructions for the First-Choice RLM-RACE Kit (Ambion, Austin, TX, USA), 2 μg of total RNA isolated from apple flower was used for ligating 5 RNA adaptors at 15 overnight. Two specific primers (Table S9 in Additional file 1) were designed to conduct nested PCRs, and PCR products were cloned to the pGEM-easy vector (Promega, Madison, WI, USA) and sequenced by Bechman Coulter Genomics (Danvers, MA, USA).

Additional material

Additional file 1: Supplemental Tables S1 to S9. Table S1: reads statistics in four libraries. Table S2: read length distribution in each conserved miRNA family. Table S3: homologous sequences for known miRNAs. Table S4: known miRNAs with good stem-loop structure predicted. Table S5: detailed list of novel and candidate miRNAs found in apple. Table S6: targets of apple miRNAs (or families; detailed list). Table S7:*MYB*genes targeted by miR828, mi858, and miR159. Table S8: targets of*MYB*-derived phased siRNAs. Table S9: RNA gel blotting probes and RLM-5-RACE primers.

Additional file 2: Supplemental Figures S1 to S5. Figure S1: diversity and size distribution of redundant and unique sRNAs. Figure S2: distribution of 21-nucleotide phasing siRNAs along apple*TAS*genes. Figure S3: multiple alignment of R3 repeat domain for 251 apple and 129*Arabidopsis*MYBs. Figure S4: T-plots for targets of miR828, miR858 and tasiARF. Figure S5: distribution analysis of siRNAs derived from the ten miR828-targeted*MYB*genes.

Additional file 3: Predicted secondary structures of the apple-specific miRNAs. This file contains all the secondary stem-loop structures for the apple-specific miRNAs. The miRNA and miRNA* sequences are denoted in red and green, respectively.

Additional file 4: Mapping plots of novel apple miRNAs. This file contains all the mapping plots illustrating the read distribution along the precursor region of novel apple miRNAs.

Abbreviations

ARF: auxin response factor; bHLH: basic helix-loop-helix; DCL: Dicer-like protein; miRNA: microRNA; PCR: polymerase chain reaction; RDR: RNA-dependent RNA polymerase; RISC: RNA-induced silencing complex; RLM-5-RACE: RNA ligation-mediated 5 rapid amplification of cDNA ends; RPM: reads per million genome-matched reads; siRNA: small interfering RNA; sRNA: small RNA; tasiRNA: trans-acting small interfering RNA.

Acknowledgements

We thank Dr Tony Wolf for financial support to RX; Dr Ann Callahan for critical review of the manuscript; Dr Christopher Dardick for suggesting data mining; Mr Dennis Bennett for experimental assistance; Dr Guangtu Gao for computational analysis; Charles Addo-Quaye and Michael Axtell for providing the Cleaveland Software; and Noah Fahlgren, Christopher M Sullivan, Kristin D Kasschau, and James C Carrington for the Targetfinder software.

Author details

[1]Alson H Smith Agricultural Research and Extension Center, Department of Horticulture, Virginia Polytechnic Institute and State University, Winchester, VA 22602, USA. [2]Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. [3]Appalachian Fruit Research Station, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV 25430, USA. [4]Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Donald Danforth Plant Science Center, St Louis, MO 63132, USA.

Authors contributions

RX and ZL initiated the research. RX, ZL, EB and YA designed the experiments. RX performed the computational analyses. HZ, RX and ZL carried out the biological experiments. RX, EB and ZL interpreted the results and prepared the manuscript. All authors have read and approved the manuscript for publication.

Competing interests

The authors declare that they have no competing interests.

Received: 10 February 2012 Revised: 30 May 2012
Accepted: 15 June 2012 Published: 15 June 2012

References

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC:**Potent and specific genetic interference by double-stranded RNA in*Caenorhabditis elegans*.***Nature*1998,**391**:806-811.
2. Jinek M, Doudna JA:**A three-dimensional view of the molecular machinery of RNA interference.***Nature*2009,**457**:405-412.
3. Hofmann NR:**MicroRNA evolution in the genus*Arabidopsis*.***Plant Cell* 2010,**22**:994.
4. ODonnell KA, Boekel JD:**Mighty piwis defend the germline against genome intruders.***Cell*2007,**129**:37-44.
5. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P:**Endogenous trans-acting siRNAs regulate the accumulation of*Arabidopsis*mRNAs.***Mol Cell*2004,**16**:69-79.
6. Voinnet O:**Origin, biogenesis, and activity of plant microRNAs.***Cell*2009, **136**:669-687.
7. Siomi H, Siomi MC:**On the road to reading the RNA-interference code.** *Nature*2009,**457**:396-404.
8. Chen XM:**Small RNAs and their roles in plant development.***Annu Rev Cell Dev Biol*2009,**25**:21-44.
9. Vazquez F, Legrand S, Windels D:**The biosynthetic pathways and biological scopes of plant small RNAs.***Trends Plant Sci*2010,**15**:337-345.
10. Xie ZX, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC:**Genetic and functional diversification of small RNA pathways in plants.***PLoS Biol*2004,**2**:642-652.
11. Jones-Rhoades MW, Bartel DP, Bartel B:**MicroRNAs and their regulatory roles in plants.***Annu Rev Plant Biol*2006,**57**:19-53.
12. Hannon GJ:**RNA interference.***Nature*2002,**418**:244-251.
13. Meister G, Tuschl T:**Mechanisms of gene silencing by double-stranded RNA.***Nature*2004,**431**:343-349.
14. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O:**Widespread translational inhibition by plant miRNAs and siRNAs.***Science*2008,**320**:1185-1190.
15. Zilberman D, Cao XF, Jacobsen SE:**ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation.***Science*2003, **299**:716-719.
16. Qi YJ, He XY, Wang XJ, Kohany O, Jurka J, Hannon GJ:**Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation.***Nature*2006,**443**:1008-1012.
17. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS:**SGS3 and SGS2/ SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in*Arabidopsis*.***Gene Dev*2004,**18**:2368-2379.
18. Allen E, Xie ZX, Gustafson AM, Carrington JC:**microRNA-directed phasing during trans-acting siRNA biogenesis in plants.***Cell*2005,**121**:207-221.
19. Yoshikawa M, Peragine A, Park MY, Poethig RS:**A pathway for the biogenesis of trans-acting siRNAs in*Arabidopsis*.***Gene Dev*2005,**19**:2164-2175.
20. Axtell MJ, Jan C, Rajagopalan R, Bartel DP:**A two-hit trigger for siRNA biogenesis in plants.***Cell*2006,**127**:565-577.
21. Allen E, Howell MD:**miRNAs in the biogenesis of trans-acting siRNAs in higher plants.***Semin Cell Dev Biol*2010,**21**:798-804.
22. Lu C, Tej SS, Luo SJ, Haudenschild CD, Meyers BC, Green PJ:**Elucidation of the small RNA component of the transcriptome.***Science*2005, **309**:1567-1569.
23. Lauter N, Kampani A, Carlson S, Goebel M, Moose SP:**microRNA172 down-regulates glossy15 to promote vegetative phase change in maize.***Proc Natl Acad Sci USA*2005,**102**:9412-9417.
24. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS:**The sequential action of miR156 and miR172 regulates developmental timing in*Arabidopsis*.***Cell*2009,**138**:750-759.
25. Wang JW, Park MY, Wang LJ, Koo YJ, Chen XY, Weigel D, Poethig RS: **MiRNA control of vegetative phase change in trees.***PLoS Genet*2011,**7**: e1002012.
26. Axtell MJ, Bowman JL:**Evolution of plant microRNAs and their targets.** *Trends Plant Sci*2008,**13**:343-349.
27. Mallory AC, Vaucheret H:**Functions of microRNAs and related small RNAs in plants.***Nat Genet* 2006,**38** Suppl:S31-36.
28. Sunkar R, Chinnusamy V, Zhu JH, Zhu JK:**Small RNAs as big players in plant abiotic stress responses and nutrient deprivation.***Trends Plant Sci* 2007,**12**:301-309.
29. Lewis BP, Burge CB, Bartel DP:**Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.***Cell*2005,**120**:15-20.
30. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP:**A diverse and evolutionarily fluid set of microRNAs in*Arabidopsis thaliana*.***Gene Dev*2006, **20**:3407-3425.
31. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP:**MicroRNAs in plants.***Gene Dev*2002,**16**:1616-1626.
32. Zhu QH, Spriggs A, Matthew L, Fan LJ, Kennedy G, Gubler F, Helliwell C:**A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains.***Genome Res*2008,**18**:1456-1465.
33. Zhang LF, Chia JM, Kumari S, Stein JC, Liu ZJ, Narechania A, Maher CA, Guill K, McMullen MD, Ware D:**A genome-side characterization of microRNA genes in maize.***PLoS Genet*2009,**5**:e1000716.
34. Barakat A, Wall PK, Diloreto S, Depamphilis CW, Carlson JE:**Conservation and divergence of microRNAs in Populus.***BMC Genomics*2007,**8**:481.
35. 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.
36. Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, Dalmay T, Burgyan J: **Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis.***Plant J*2010, **62**:960-976.
37. Song QX, Liu YF, Hu XY, Zhang WK, Ma BA, Chen SY, Zhang JS: **Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing.***BMC Plant Biol*2011,**11**:5.
38. Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X:**Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type.***BMC Genomics*2010,**11**:246.
39. Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, Li AQ, Li MJ, Li CS, Zhang BH, Wang XJ:**Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea*L.).***BMC Plant Biol*2010,**10**:3.
40. Kozomara A, Griffiths-Jones S:**miRBase: integrating microRNA annotation and deep-sequencing data.***Nucleic Acids Res*2011,**39**:D152-D157.
41. Cuperus JT, Fahlgren N, Carrington JC:**Evolution and functional diversification of miRNA genes.***Plant Cell*2011,**23**:431-442.
42. Allen E, Xie Z, 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.
43. Axtell MJ, Bartel DP:**Antiquity of microRNAs and their targets in land plants.***Plant Cell*2005,**17**:1658-1673.
44. Boyer J, Liu R:**Apple phytochemicals and their health benefits.***Nutrition J* 2004,**3**:5.
45. Ferree DC, Warrington IJ:**Apples: Botany.***Production and Uses*New York: CABI; 2003.
46. Varkonyi-Gasic E, Gould N, Sandanayaka M, Sutherland P, MacDiarmid RM: **Characterisation of microRNAs from apple (*Malus domestica*Royal Gala) vascular tissue and phloem sap.***BMC Plant Biol*2010,**10**:159.
47. Yu HP, Song CN, Jia QD, Wang C, Li F, Nicholas KK, Zhang XY, Fang JG: **Computational identification of microRNAs in apple expressed sequence tags and validation of their precise sequences by miR-RACE.***Physiol Plant* 2011,**141**:56-70.
48. Moxon S, Jing RC, Szittya G, Schwach F, Pilcher RLR, Moulton V, Dalmay T: **Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening.***Genome Res*2008,**18**:1602-1609.
49. Hofacker IL:**Vienna RNA secondary structure server.***Nucleic Acids Res* 2003,**31**:3429-3431.
50. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen XM, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T, Matzke M, Ruvkun G, Tuschl T:**A uniform system for microRNA annotation.***RNA*2003,**9**:277-279.
51. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen XM, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi YJ, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhui JK:**Criteria for annotation of plant microRNAs.***Plant Cell*2008,**20**:3186-3190.
52. Genome Database for Rosaceae.. [http://www.rosaceae.org/projects/ apple_genome].
53. Ma Z, Coruh C, Axtell MJ:**Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the*Arabidopsis*genus.***Plant Cell* 2010,**22**:1090-1103.
54. Addo-Quaye C, Miller W, Axtell MJ:**CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets.***Bioinformatics*2009, **25**:130-131.

55. 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.
56. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L:**MYB transcription factors in*Arabidopsis*.***Trends Plant Sci*2010,**15**:573-581.
57. Feller A, Machemer K, Braun EL, Grotewold E:**Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.***Plant J*2011,**66**:94-116.
58. Stracke R, Werber M, Weisshaar B:**The R2R3-MYB gene family in *Arabidopsis thaliana*.***Curr Opin Plant Biol*2001,**4**:447-456.
59. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM:**Expansion and diversification of the Populus R2R3-MYB family of transcription factors.** *Plant Physiol*2009,**149**:981-993.
60. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y:**Solution structure of a specific DNA complex of the MYB DNA-binding domain with cooperative recognition helices.***Cell* 1994,**79**:639-648.
61. Williams CE, Grotewold E:**Differences between plant and animal MYB domains are fundamental for DNA binding activity, and chimeric MYB domains have novel DNA binding specificities.***J Biol Chem*1997, **272**:563-571.
62. Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP:**MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5 region.***EMBO J* 2004,**23**:3356-3364.
63. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC:**Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in*Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting.***Plant Cell*2007,**19**:926-942.
64. Li Y-F, Zheng Y, Addo-Quaye C, Zhang L, Saini A, Jagadeeswaran G, Axtell MJ, Zhang W, Sunkar R:**Transcriptome-wide identification of microRNA targets in rice.***Plant J*2010,**62**:742-759.
65. Luo QJ, Mittal A, Jia F, Rock CD:**An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in*Arabidopsis*.***Plant Mol Biol*2011.
66. Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC: **Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10.***Plant J*2007,**49**:414-427.
67. Williams L, Carles CC, Osmont KS, Fletcher JC:**A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the*Arabidopsis ARF2 ,ARF3*, and*ARF4*genes.***Proc Natl Acad Sci USA*2005, **102**:9703-9708.
68. Guilfoyle TJ, Hagen G:**Auxin response factors.***Curr Opin Plant Biol*2007, **10**:453-460.
69. Sunkar R, Zhou X, Zheng Y, Zhang W, Zhu JK:**Identification of novel and candidate miRNAs in rice by high throughput sequencing.***BMC Plant Biol* 2008,**8**:25.
70. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JDG:**A plant miRNA contributes to antibacterial resistance by repressing auxin signaling.***Science*2006,**312**:436-439.
71. Sunkar R, Zhu JK:**Novel and stress-regulated microRNAs and other small RNAs from*Arabidopsis*.***Plant Cell*2004,**16**:2001-2019.
72. Zhang WX, Gao S, Zhou XF, Xia J, Chellappan P, Zhou XA, Zhang XM, Jin HL:**Multiple distinct small RNAs originate from the same microRNA precursors.***Genome Biol*2010,**11**:R81.
73. Krasnikova MS, Milyutina IA, Bobrova VK, Ozerova LV, Troitsky AV, Solovyev AG, Morozov SY:**Novel miR390-dependent transacting siRNA precursors in plants revealed by a PCR-based experimental approach and database analysis.***J Biomed Biotechnol*2009,**2009**:952304.
74. Hsieh LC, Lin SI, Shih ACC, Chen JW, Lin WY, Tseng CY, Li WH, Chiou TJ: **Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis*by deep sequencing.***Plant Physiol*2009,**151**:2120-2132.
75. Jia L, Clegg MT, Jiang T:**Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes.***Plant Physiol*2004,**134**:575-585.
76. Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C, Smeekens S, Tonelli C, Paz-Ares J, Weisshaar B:**Towards functional characterisation of the members of the R2R3-MYB gene family from*Arabidopsis thaliana*.***Plant J*1998, **16**:263-276.
77. Zhai JX, Jeong DH, De Paoli E, Park S, Rosen BD, Li YP, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC:**MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs.***Gene Dev* 2011,**25**:2540-2553.
78. **Rfam 10.**. [http://www.sanger.ac.uk/resources/databases/rfam.html].
79. **miRbase.**. [http://www.mirbase.org/].
80. Langmead B, Trapnell C, Pop M, Salzberg SL:**Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol*2009,**10** :R25.
81. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG:**Clustal W and clustal X version 2.0.***Bioinformatics* 2007,**23**:2947-2948.
82. **TAIR.**. [http://www.arabidopsis.org].
83. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M:**Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.***Bioinformatics*2005,**21**:3674-3676.
84. **NCBI Conserved Domain Search.**. [http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi].
85. **GENIO/logo.**. [http://www.biogenio.com/logo/logo.cgi].
86. **Targetfinder 1.6.**. [http://carringtonlab.org/resources/targetfinder].

doi:10.1186/gb-2012-13-6-r47
Cite this article as:*Xiaet al.*:Apple miRNAs and tasiRNAs with novel regulatory networks.*Genome Biology*201213:R47.

Chapter 3

Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs

Hong Zhu, Rui Xia, Bingyu Zhao, Yong-qiang An, Chris D Dardick, Ann M Callahan and Zongrang Liu

Contribution: Rui Xia did two sRNA library construction, helped with RNA isolation, conducted all the bioinformatics analysis, and assisted with figure preparation and manuscript writing.

Published in
BMC plant biology
BioMed Central
236 Gray's Inn Road
London WC1X 8HB
United Kingdom
2012, 12:149. doi: 10.1186/1471-2229-12-149.
http://www.biomedcentral.com/1471-2229/12/149

Zhuet al. BMC Plant Biology2012,12:149
http://www.biomedcentral.com/1471-2229/12/149

BMC Plant Biology

RESEARCH ARTICLE Open Access

Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs

Hong Zhu[1,3], Rui Xia[1,2,3], Bingyu Zhao[1], Yong-qiang An[4], Chris D Dardick[3], Ann M Callahan[3] and Zongrang Liu[1,3*]

Abstract

Background: MicroRNAs (miRNAs) have recently emerged as important gene regulators in plants. MiRNAs and their targets have been extensively studied in*Arabidopsis*and rice. However, relatively little is known about the characterization of miRNAs and their target genes in peach (*Prunus persica*), which is a complex crop with unique developmental programs.

Results: We performed small RNA deep sequencing and identified 47 peach-specific and 47 known miRNAs or families with distinct expression patterns. Together, the identified miRNAs targeted 80 genes, many of which have not been reported previously. Like the model plant systems, peach has two of the three conserved*trans*-acting siRNA biogenesis pathways with similar mechanistic features and target specificity. Unique to peach, three of the miRNAs collectively target 49*MYB*s, 19 of which are known to regulate phenylpropanoid metabolism, a key pathway associated with stone hardening and fruit color development, highlighting a critical role of miRNAs in the regulation of peach fruit development and ripening. We also found that the majority of the miRNAs were differentially regulated in different tissues, in part due to differential processing of miRNA precursors. Up to 16% of the peach-specific miRNAs were differentially processed from their precursors in a tissue specific fashion, which has been rarely observed in plant cells. The miRNA precursor processing activity appeared not to be coupled with its transcriptional activity but rather acted independently in peach.

Conclusions: Collectively, the data characterizes the unique expression pattern and processing regulation of peach miRNAs and demonstrates the presence of a complex, multi-level miRNA regulatory network capable of targeting a wide variety of biological functions, including phenylpropanoid pathways which play a multifaceted spatial-temporal role in peach fruit development.

Keywords:miRNA, Deep sequencing,*Prunus persica*, Pre-miRNA processing,*Trans*-acting siRNA, MYB

Background

There are many mechanisms by which plants regulate gene expression to ensure normal development and appropriate responses to both biotic and abiotic signals. One regulatory mechanism involves endogenous small RNA (sRNA) molecules, 20~24-nt in length [1,2], which act by silencing gene expression. In plants, sRNAs have been classified based on their biogenesis, including microRNAs (miRNAs), heterochromatic siRNAs (hc-siRNAs),*trans*-acting siRNAs (tasiRNAs) and natural antisense siRNAs (nat-siRNAs) [1,3-6]. TasiRNA biogenesis from *TAS* loci depends on miRNA-directed cleavage of their transcripts [4,7,8] and three tasiRNA pathways have been characterized in *Arabidopsis* [7,9]. Although miRNAs only constitute a small fraction in the sRNA population [10,11], the miRNA-guided post-transcriptional gene regulation is one of the most conserved and well-characterized gene regulatory mechanisms [6,10,12]. Increasing evidence shows that miRNAs negatively regulate their target genes, which function in a wide range of biological processes, including organogenesis, signal transduction and stress responses [13,14].

* Correspondence:zongrang.liu@ars.usda.gov
[1]Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[3]Appalachian Fruit Research Station, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV 25430, USA
Full list of author information is available at the end of the article

Chapter 4

MicroRNA superfamilies descended from miR390 and their roles in secondary siRNA biogenesis

Rui Xia,[a,b,c] Blake C. Meyers,[d] Zhongchi Liu,[e] Eric P. Beers,[a] Songqing Ye,[e,1] and Zongrang Liu[a,b,2]

[a] Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA

[b] Appalachian Fruit Research Station, Agricultural Research Service, United States Department of Agriculture, Kearneysville, West Virginia 25430, USA

[c] Alson H Smith Agricultural Research and Extension Center, Department of Horticulture, Virginia Polytechnic Institute and State University, Winchester, Virginia 22602, USA

[d] Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19717, USA.

[e] Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, USA

Running Title: miRNA superfamily descendants of miR390

Submitted to The Plant Cell

Abstract

TasiRNAs are a major class of small RNAs performing essential biological functions in plants. The first-reported tasiRNA pathway of miR173-*TAS1/2* produces tasiRNAs regulating a set of *PPR* genes and has been characterized only in *Arabidopsis* to date. Here we demonstrate that the miRNA-*TAS*-*PPR*-siRNA pathway is a highly dynamic and widespread feature of eudicots. Nine eudicot plants, representing six different plant families, have evolved similar tasiRNA pathways to instigate phasiRNA production from *PPR* genes, which are triggered by different 22-nt miRNAs, including miR7122, miR1509, and fve-PPRtri1/2, and through distinct mechanistic strategies exploiting miRNA direct-targeting or indirect-targeting through *TAS*-like genes (*TASL*), one-hit or two-hit, or even two layers of tasiRNA-*TASL* interactions. Intriguingly, although those miRNA triggers display high sequence divergence caused by the occurrence of frequent point mutations and splicing shifts, their corresponding *MIRNA* genes show pronounced identity with the *Arabidopsis MIR173*, implying a common origin of this group of miRNAs (super-miR7122). Further analyses reveal that super-miR7122 may have evolved from a newly defined miR4376 superfamily, which probably originated from the widely conserved miR390. The elucidation of this novel evolutionary path expands our understanding of the course of miRNA evolution, especially for relatively conserved miRNA families.

4.1 Introduction

Small RNAs (sRNAs), typically 20-24 nucleotides (nt) long, are important gene and chromatin regulators in plants. MicroRNAs (miRNAs) are 20-22-nt RNAs that negatively regulate target genes through homology-directed mRNA cleavage or translational inhibition at the post-transcriptional level (Bartel, 2004; Voinnet, 2009). Another class of sRNAs, small interfering RNAs (siRNAs), account for the majority of the sRNA repertoire in cells and are implicated in a variety of processes, including combatting viral and transposon replication, establishment and maintenance of heterochromatin, and post-transcriptional gene regulation (Baulcombe, 2004; Jones-Rhoades et al., 2006; Matzke et al., 2009; Law and Jacobsen, 2010). Trans-acting siRNAs (tasiRNAs) are a class of 21-nt siRNAs which direct sequence-specific cleavage of their target genes, as do miRNAs. TasiRNA biogenesis from trans-acting siRNA transcripts (*TAS*) requires an initial cleavage of the mRNA transcripts by a specific miRNA, after which one of the cleaved products is made double-stranded by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), and subsequently diced by DICER-LIKE ENZYME 4 (DCL4) into 21-nt species; these 21-nt secondary siRNAs have a characteristic phased pattern initiating at the miRNA cleavage site (Vazquez et al., 2004; Allen et al., 2005; Yoshikawa et al., 2005). Some of the phased siRNAs (phasiRNAs) may further target their parental genes in *cis* or other genes in *trans* , distinguishing them as either casiRNAs (cis-acting RNAs) or tasiRNAs, respectively (Zhai et al., 2011).

To date, four *TAS* gene families have been well characterized in *Arabidopsis*, including the “one-hit” *AtTAS1/2/4* initiated by a single target site of a 22-nt miRNA (“1_{22}”, miR173 or miR828) and the “two-hit” *AtTAS3* with two target sites of the 21-nt miR390 (“2_{21}”) (Yoshikawa et al., 2005; Axtell et al., 2006; Rajagopalan et al., 2006). MiR173-targed *TAS1/2* produce tasiRNAs targeting a few pentatricopeptide repeat-containing proteins (*PPRs*) (Yoshikawa et al., 2005; Howell et al., 2007), miR828-*TAS4*-derived siRNA81(-) targets genes coding MYB transcription factors (MYBs) (Rajagopalan et al., 2006; Luo et al., 2012; Xia et al., 2012; Zhu et al., 2012) while miR390-*TAS3*-spawned tasiARFs regulate genes coding for auxin responsive factors (ARFs) (Williams et al., 2005; Fahlgren et al., 2006; Xia et al., 2012). Although miR828 was originally identified as a trigger of phasiRNA production solely for the non-coding

TAS4 in *Arabidopsis*, it was later found to trigger phasiRNA production in a few MYB-encoding genes as well (Xia et al., 2012; Zhu et al., 2012). In fact, phasiRNA production has been demonstrated to occur in numerous protein-coding genes, including those coding for nucleotide-binding site leucine-rich repeat proteins (NB-LRRs) (Klevebring et al., 2009; Zhai et al., 2011; Shivaprasad et al., 2012), MYBs (Xia et al., 2012; Zhu et al., 2012), PPRs (Chen et al., 2007; Howell et al., 2007), Ca^{2+}-ATPase (Wang et al., 2011), and TIR/AFB2 (transport inhibitor response 1 / auxin signaling F-box protein 2) auxin receptors (Si-Ammour et al., 2011). In most cases, the miRNA triggers are 22-nt and act through the 1_{22} mode, while a few of them, such as miR156 and miR172, are 21-nt and act together to function (like miR390) via the 2_{21} mode (Zhai et al., 2011). Although originally described for 21-nt miRNAs, the two-hit mechanism can also employ 22-nt miRNAs, as in the case of miR1509 in *Medicago truncatula* that triggers phasiRNA biogenesis through binding and cleavage at two sites in its target gene (Zhai et al., 2011). Intriguingly, but apparently rarely, 22-nt tasiRNAs can also serve as effective triggers of phasiRNAs, as in the case of a predominantly 22-nt *TAS2*-derived tasiRNA 3' D6(-) (also termed as tasiR2140) that is capable of initiating phasiRNA biogenesis in *PPR*s, while one of the *TAS1c*-derived 21-nt tasiRNAs works together with miR173 to delimit the boundary of phasiRNA production along the cleaved transcript (Yoshikawa et al., 2005; Howell et al., 2007; Chen et al., 2010; Rajeswaran et al., 2012). Evidently, both miRNA and tasiRNA can exploit different mechanisms to precisely trigger and define phasiRNA production in specific gene regions.

In contrast to the broad conservation of the miR390-*TAS3* pathway in seed plants and the miR828-*TAS4* pathway in eudicots, the miR173-*TAS1/2* pathway has not been reported in species beyond *Arabidopsis* (Axtell et al., 2006; Cuperus et al., 2011; Luo et al., 2012),. Several models have been proposed for the evolution of plant *MIRNA*s (Allen et al., 2004; Felippes et al., 2008; Piriyapongsa and Jordan, 2008), however, a complete view of how most individual *MIRNA*s originated and evolved is still lacking, especially for the relatively conserved *MIRNA*s. In this study, we characterized a widely-adapted miRNA-*TAS*-*PPR*-siRNA regulatory circuit in eudicots. The initial miRNA triggers of *PPR* phasiRNA production share a common ancestor with miR173. We also demonstrate

that these miRNA triggers are evolutionarily related to the highly conserved miR390 family, revealing a novel miRNA evolutionary history.

4.2 Results

4.2.1 *PHAS* gene identification with reverse computation

Plant genomes are rich in coding genes or non-coding genomic loci capable of producing 21-nt siRNAs (Howell et al., 2007; Johnson et al., 2009; Zhai et al., 2011; Zhang et al., 2012). Twenty-two nucleotide miRNAs can trigger the production of secondary phasiRNAs, which are generally 21-nt in length and in phase with the miRNA cleavage site, from the 5' end of the 3' cleavage product of their targeted transcripts (Chen et al., 2010; Cuperus et al., 2010). Both coding genes and non-coding transcripts capable of phasiRNA production (*PHAS* genes) have been identified in many species but efficient identification of these *PHAS* genes and their miRNA trigger remains challenging. We developed customized computation pipelines for genome-wide identification of *PHAS* genes and their corresponding miRNA triggers. The process, named as reverse computation, is illustrated in Figure 1A, and includes a p-value-based strategy similar to that of Chen et al. (2007), to first retrieve *PHAS* genes and then identify their corresponding triggers.

We first tested our pipelines in peach using its annotated transcriptome. From the four peach deep-sequencing sRNA libraries available, we were able to identify 350 *PHAS* gene candidates, of which 263 (about 75%) protein-coding genes were capable of secondary siRNA production with a p-value ≤ 0.001 (Figure 1B), a threshold proposed by Chen et al. (2007), including 94 *NB-LRR* defense genes, 20 *PPR* genes, 10 kinase-like genes, and other genes coding for transcription factors (*MYB*, *NAC*, *TIR*/*AFB*) or sRNA biogenesis related genes (*SGS*, *AGO* and *DCL*) (Table 1; see Supplemental Table 1 online). Members of these gene families have been reported to produce secondary siRNAs in other plants, like Medicago and soybean (Zhai et al., 2011). Additional *PHAS* genes encode other proteins like UDP-glucosyl transferase and the dehydration responsive protein RD22 (Table 1). Two *TAS* genes, miR390-targeted *TAS3* and miR828-targeted *TAS4* which produce tasiRNAs (Zhu et al., 2012), were retrieved with an

extremely low p-value in a genome-wide profiling of non-coding *PHAS* loci (Table 1). Our reverse computation analysis also helped identify triggers for these *PHAS* genes, including miR482/2118 as the trigger for *NB-LRR*s, miR828 for *MYB*s, miR393 for *TIR/AFB*s, and miR168 for *AGO*s (Table 1). These results verified our computation pipelines as a valid and efficient means for the identification of *PHAS* genes and their miRNA triggers.

4.2.2 miR7122 as a trigger for phasiRNA production in both *PPRs* and *TAS*-like loci in peach

Our initial analysis showed that 23 peach *PPR* genes produced phasiRNAs (including three with p-value > 0.001; see Supplemental Table 2 online). MiR161, miR400 and miR173 directly or indirectly trigger the phasiRNA production of *PPR* genes in *Arabidopsis* (Axtell et al., 2006; Chen et al., 2007; Howell et al., 2007). In searching for homologs of these miRNAs, however, using the current criterion for miRNA families (≤ 4 base differences) (Meyers et al., 2008), these miRNAs appears to be absent from peach (Zhu et al., 2012), suggesting different miRNAs might trigger the *PPR* phasiRNAs. To address this, we profiled all the possible sRNA complementary sites for the 23 *PPR* genes and identified a confident target site [alignment score (AS) ≤ 5] in 10 out of the 23 *PPR*s for ppe-miR7122, with cleavage in three putative *PHAS PPR*s supported by PARE data (see Supplemental Table 2 online)(Zhu et al., 2012). Detailed examination of the mapping profile revealed that the miR7122 cleavage site set the phase of the phasiRNA production (Figure 1D). In peach, two miR7122 variants (a/b) were identified, both of which possess all the features required to trigger phasiRNA biogenesis: 22-nt in length, 5' terminal U, and a bulge-bearing stem-loop precursor (Figure 1C) (Chen et al., 2010; Manavella et al., 2012; Zhu et al., 2012).

In *Arabidopsis, PPRs* are targeted not only by miR161, miR400 and miR173, but also by *TAS1/2*-derived tasiRNAs. To ascertain whether ppe-miR7122 also indirectly targets *PPR*s through trans-acting siRNAs, we performed a genome-wide search for ppe-miR7122 targeted genomic loci and found that two loci, separated in the genome by about 4500 bp, were cleaved by ppe-miR7122 and generated robust siRNAs (Figure 2). These loci, which share 78% sequence identity and appear to be non-protein coding, are

likely to represent *TAS*-like genes, hereafter referred to as *PpTASL1* and *PpTASL2*. The majority of tasiRNA generated at the poly(A) proximal end fell into the second phase (Figure 2A, see Supplemental Figure 1A online), and several phased siRNAs, as expected, were predicted to target *PPRs* in *trans* (Figure 2B; see Supplemental Table 3 online). In particular, two tasiRNAs [*PpTASL1*-3' D2(-) or *PpTASL2*-3' D2(-)], predominantly 22-nt in length, potentially target 14 *PHAS PPR*s (AS ≤ 5) and trigger subsequent phasiRNA production, based on the siRNA matching pattern (see Supplemental Figure 1B online). Six out of these 14 *PHAS PPR*s have either no or a less cleavage-favorable (AS > 5) target site for ppe-*MIR7122* (see Supplemental Table 2 online), indicating synergistic and complementary roles of these *TASL*-derived tasiRNAs with the ppe-miR7122. Similar 22-nt tasiRNA triggering of siRNA production in *PPRs* and reinforcement of the silencing effect have also been reported in the miR173-*TAS2-PPR* pathway in *Arabidopsis* (Yoshikawa et al., 2005; Howell et al., 2007; Chen et al., 2010). Thus, as illustrated in Figure 3 (category I), the ppe-miR7122 in peach is, like miR173 in *Arabidopsis*, able to trigger siRNA production from *PPR* genes by direct targeting or through an additional indirect *TASL*-tasiRNA path.

4.2.3 Apple miR7122 triggers the phasiRNA production of *PPR* genes only through tasiRNAs with a two-hit 2_{22} mode

The identification of a homologous miR7122 (mdm-miR7122, Figure 1C) in our previous study in apple (Xia et al., 2012) prompted us to evaluate whether an analogous miR7122-(*TASL*)-*PPR*-siRNA pathway was conserved in apple. After examining the apple small RNA data, 15 *PPR*s (three with p-value > 0.001) were retrieved from our phasing analysis (see Supplemental Table 4 online), indicating that apple *PPR*s are channeled into phasiRNA biogenesis as well. However, reverse computation failed to identify any complementary target sites for mdm-miR7122 within these *PHAS PPR* genes. Instead, two 22-nt sRNAs located tandemly in the apple genome were found to be potential triggers. Analysis of sequence folding (e.g. secondary structure) failed to identify possible stem-loop structures for the precursor locus, suggesting these two sRNAs are unlikely to be miRNAs, for which biosynthesis requires a canonical stem-loop structure. The question of how these two sRNAs originated inspired us to examine the

sRNA population matched to the precursor locus. We found that extremely abundant siRNAs were produced from this locus (Figure 4A). Trigger identification for this locus revealed that mdm-miR7122 was the potential trigger of siRNA production. Curiously, two target sites of mdm-miR7122, 792 bp apart, were found in this locus, and siRNAs predominantly 21-nt in length with some 22-nt siRNAs (Figure 4E) were profusely produced only from the region bordered by the two mdm-miR7122 target sites (Figure 4A). Both target sites are predicted to be cleavable based on the sequence-pairing, but we only confirmed the cleavage of the 5' site using 5'-RLM-RACE, which was consistent with the bulk of siRNAs in phase with the 5' cleavage site (Figure 4A and 4B; see Supplemental Figure 1C online). However, the majority of 21 tasiRNAs fell into the third register (a two nucleotide shift), instead of the first register that is in phase with the 5' cleavage site (see Supplemental Figure 1C online). This out-of-phase phenomenon was also observed for other *TAS* or *PHAS* genes, perhaps due to slippage of the Dicer activity or reinforcement of *cis*-acting tasiRNAs (Axtell et al., 2006; Zhai et al., 2011; Rajeswaran et al., 2012; Xia et al., 2012). Since the nucleotide sequence of this *PHAS* locus appears not to code a polypeptide of any significant length or homology, we denoted it as *MdTASL1*.

In total, 233 siRNA species were produced from *MdTASL1*, 123 of which had both 21-nt and 22-nt variants (Figure 4F); over half of these siRNAs (135 out of 233) were predicted to be capable of targeting *PPR* genes in apple (AS $\leq$ 5, Figure 4F; see Supplemental Table 5 online). Deeper sequence analyses uncovered that *MdTASL1* has a duplicated 83-nt sequence (R1 and R2) 78-nt away from the 5' cleavage site (Figure 4B), and the two 22-nt tasiRNAs triggering siRNA production from *PPR* genes are located in tandem within this repetitive region (Figure 4B and 4C). One is in phase with the 5' cleavage site in the R1 repeat, while the other is in phase in the R2 repeat (Figure 4C). Accordingly, we designated these two siRNAs as 3' D6(-) and 3' D11(-) in accordance with their phase position (Figure 4C). They potentially co-target a 44-nt region in *PPR* genes which encodes a peptide of 14 amino acids spanning the adjacent parts of two PPR repeats (Figure 4D); the 3' D11(-)-mediated cleavage of three *PPR* genes was confirmed by the PARE data (see Supplemental Table 5 online). Taken together, we conclude that, distinct from the ppe-miR7122 which targets *PPR* directly or indirectly through *TAS*-like

genes, mdm-miR7122 triggers *PPR* phasiRNA production solely by targeting *MdTASL1* through a two-hit 2_{22} mode (Figure 3, category III).

4.2.4 The MiRNA-*TAS*-*PPR*-siRNA pathway is more robust in strawberry

To ascertain whether analogous pathways are conserved in Rosaceae species, we conducted sRNA deep-sequencing and similar reverse computation analyses in strawberry. Nineteen *PPR* genes (see Supplemental Table 6 online, including eight *PPR*s not annotated accurately and one with p-value > 0.001) were shown to produce phasiRNAs; two 22-nt miRNAs with critical structural features (22-nt in length, 5' terminal U, and a bulge-bearing stem-loop precursor) were identified as potential triggers. Pairwise comparison of their sequences revealed that they share 16 identical nucleotides with a 4-nt mutual shift (Figure 5A), but both show high sequence divergence (> 7 base differences) relative to the miR7122 conserved in peach and apple. We designated these as the strawberry *PPR* triggers 1 and 2 (fve-PPRtri1/2). The high similarity of these two trigger miRNAs showed an 18-base overlap of their target sites, with the target site of fve-PPRtri1 shifted 4-nt toward the 5' end of target genes relative to fve-PPRtri2 (Figure 5B). Apart from the *PPR* genes, fve-PPRtri1/2 appeared to target five other putative coding genes and 18 non-coding *TAS*-like loci in the strawberry genome, suggesting fve-PPRtri1/2 have gained targets outside the *PPR* family and thus are involved in a more extensive *trans*-acting network (Figure 5B; see Supplemental Table 7 online). It is noteworthy that five *TAS*-like loci are clustered in the same scaffold (scf05113146), with three residing within a ~9 kb region (see Supplemental Figure 2 and Supplemental Table 7 online). They all show low similarity to *PPR* genes, and several phasiRNAs derived from these *TAS*-like loci have the potential to target *PPR* genes, indicative of a *trans*-acting role in line with that of *TAS1/2* in *Arabidopsis*. Thus, as shown in Figure 3 (category I), fve-PPRtri1/2 directly or indirectly trigger the phasiRNA production of *PPR* genes, like ppe-miR7122, but in a much more robust way.

4.2.5 Legume plants evolved a similar miRNA-*TAS*-*PPR*-siRNA pathway with distinct mechanistic modes.

Regulation of *PPR* transcripts via secondary siRNA production has been reported for several plant species, including *Arabidopsis* (Howell et al., 2007), soybean (Zhai et al., 2011), and poplar (Klevebring et al., 2009), indicating a broad existence for this pathway. Zhai et al. (2011) reported several *PHAS PPR* genes, but their miRNA triggers are currently unknown. To explore whether the pathway(s) by which *PPR* genes produce phasiRNAs in soybean shares features with those in the Rosaceae, we analyzed publicly available sRNA data from soybean. This identified a regulatory pathway consisting of two layers of *TASL*-tasiRNA interactions (Figure 5C; see Supplemental Figure 3 online). First, a 22-nt miR1509 targets a transcript derived from an intergenic locus (*GmTASL-L1*) and initiates a first layer of siRNA biogenesis via a possible 2_{22} mode. Subsequently, a 22-nt tasiRNA [*GmTASL-L1*-D2(-)] from this locus triggers a second layer of siRNA biogenesis from another intergenic non-coding locus *GmTASL-L2*, previously reported by Zhai et al. (2011) (Figure 5C; see Supplemental Figure 3 online). *GmTASL-L2* generates 21 or 22-nt phased tasiRNAs, some of which are able to target *PPR* genes and subsequently initiate siRNA production (Figure 5C; see Supplemental Table 8 online). Eight of 13 *PHAS PPR*s have a target site complementary to gma-miR1508, for which cleavage in five *PPR*s was confirmed in the PARE data and also resulted in phasiRNA production (see Supplemental Figure 3 and Supplemental Table 8 online). A similar two-layer pathway was also identified in *Medicago truncatula*, another legume species (Figure 5C; see Supplemental Figure 3 online), although we failed to identify gma-miR1508 homologues for that species. An additional layer of complexity in soybean and Medicago is that the tasiRNA *TASL-L1*-D2(-) is capable of directly targeting *PPR* genes and initiating subsequent siRNA production, with the cleavage of a Medicago *PHAS PPR* gene (Medtr1g043080) by *MtTASL-L1*-D2(-) confirmed by PARE data (see Supplemental Table 8 online). In summary, *PPR* phasiRNA biogenesis can be initiated indirectly by miR1509 through one or two layers of *TASL*-tasiRNA interactions in both soybean and Medicago, with the miR1508-*PPR*-siRNA pathway present only in soybean (Figure 3, category IV).

4.2.6 miRNA triggers share a common origin and the miRNA-*TAS*-*PPR*-siRNA pathway is conserved in many other species.

The functional similarity of miR173, miR7122, fve-PPRtri1/2, and miR1509 as triggers of the production of tasiRNAs targeting *PPR* genes raised the question of whether these miRNA triggers are evolutionarily related. To address this question, we performed a multiple alignment using the foldback sequences of eight relevant *MIRNA* genes (Figure 6A). Surprisingly, we found that with the exception of the *ath-MIR173/mtr-MIR1509b* pair for which the identity was only 52%, all interspecies pairwise comparisons revealed high levels of identity, ranging from 60% to 85% (Figure 6B). In particular, the miRNA and miRNA* as well as their adjacent sequences are highly conserved (Figure 6A; see Supplemental Figure 4A online). Three of the eight *MIRNA*s show different processing positions with a 5-nt shift toward the 5' end occurring for ath-miR173 and fve-PPRtri2, and a 1-nt shift for fve-PPRtri1 (Figure 6A). Our results strongly suggest that these eight *MIRNA* genes, including *MIR173, MIR7122*s, *fve-PPRTRI1/2*, and *MIR1509*s, evolved from a common ancestor.

To examine the conservation of these miR7122-related miRNAs and their functions in plants, we used publicly available deep-sequencing data to retrieve miR7122 homologues from five additional species including tobacco (*Nicotiana tabacum*, nta-miR7122), grape (*Vitis vinifera*, vvi-miR7122), citrus (*Citrus sinensis* csi-miR7122), tree cotton (*Gossypium arboreum*, gar-miR7122), and lettuce (*Lactuca sativa*, lsa-miR7122), all of which are of the same splicing pattern as ppe-miR7122 without any miRNA shift (Figure 6C). Using reverse computation, we identified the miR7122-*TAS-PPR*-siRNA pathway in these species, except in tree cotton and lettuce due to the absence of their sequenced genomes. In citrus, in addition to the miR7122 homologue (a 22-nt isoform of csi-miR3954; derived from a different *MIRNA* gene), we identified another miRNA trigger (csi-miRC1, see Supplemental Figure 5 online), which is predicted to directly target a *PHAS PPR* gene (see Supplemental Table 9 online). Similarly, we identified in poplar another 22-nt miRNA as the trigger of the *TAS-PPR*-siRNA pathway, ptc-miR6427, instead of a miR7122 homologue (see Supplemental Table 9 online). Evidently, regulation of *PPR* genes mediated by miR7122 homologues is widespread in plants but with great mechanistic plasticity and diversity. Taken together, these pathways appear to fall into four categories as outlined in Figure 3.

Next, to examine the degree of homology shared by *PHAS PPR*s from different species, a total of 91 *PHAS PPR* gene sequences retrieved from eight species were subjected to phylogenetic analysis. These *PPR*s generally clustered together into species-specific subclasses (see Supplemental Figure 6A online), and the target site for miRNA or tasiRNAs appeared to be randomly distributed along or outside the *PPR* repeat-coding region (see Supplemental Figure 6B online), implying that the strategy of targeting *PPR*s by miRNAs or tasiRNAs may have evolved relatively quickly after the divergence of these plant species, consistent with the extraordinary evolutionary plasticity of *PPR* genes in plants (O'Toole et al., 2008).

4.2.7 miR7122 shares, to various degree, sequence identity with many other miRNAs

The identification of the homologous relationship among miR7122-homologous miRNAs indicates that miRNAs of a common ancestor and similar function are often divergent (> 4 bases) or shifted along the precursor (> 2 bases). Therefore, we applied more permissive search parameters to identify close relatives of miR7122. We found that miR7122 shares substantial identity with miR4376, especially in grape, in which vvi-miR4376 (see Supplemental Figure 7 online) and vvi-miR7122 share 17 nucleotides with 60% sequence identity detected over the length of the *MIRNA* foldback sequences (see Supplemental Figure 4B online). A deeper search in miRbase uncovered several other miRNAs that shared identity with miR7122 or miR4376, including the widely conserved miR390, and the lineage/species-specific miRNAs such as miR391, miR1432, miR5225 and miR3627. Multiple alignments revealed a 13-nt core sequence highly conserved among these miRNAs (Figure 6D), suggestive of a potential common origin. The core sequence, which shows relatively more sequence divergence among the miR7122 homologues (including miR7122, miR1509, miR173, and miR3954, Figure 6D), is not consistently positioned within the miRNAs from different families, for instance, miR1432 shifted three nucleotides in the 3' direction while miR173 shifted five nucleotides in the 5' direction, compared to miR390 (Figure 6D), indicating the common occurrence of sequence mutation and miRNA shifts during evolution within this group of plant miRNAs.

4.2.8 Potential evolution of MIR7122s from MIR390s via MIR4376s

Next, we asked how these miRNA families evolved in the plant kingdom. The sequence similarity described above for these miRNAs provided a strong clue that they may have evolved from a common, ancestral *MIRNA* gene. To test this possibility, we performed phylogenetic analysis for all the *MIRNA* genes from these families. Given that only the sequences of miRNA and miRNA* in the foldback sequences are relatively conserved during evolution (Jones-Rhoades et al., 2006; Fahlgren et al., 2010; Ma et al., 2010) and that frequent miRNA sequence shift was observed in this group of miRNAs (Figure 6D), to obtain a confident multiple alignment which is the requirement of accurate phylogenetic analysis, we used the sequences consisting of the miRNA, miRNA* and five nucleotides of their flanking sequences at each end to perform alignment and subsequent phylogenetic tree construction.

As shown in the unrooted tree in Figure 7A, the *MIRNA* genes of miR390, a highly conserved miRNA family of great sequence consistency (≤ 2 nucleotide differences) and of extremely conserved biological function in plants, represented a basal clade and a separate group independent of other *MIRNA* genes. As expected, *MIR7122* homologues constituted one of the two major clades other than *MIR390* members, with *MIR7122* genes dispersed among the *MIR173* and *MIR1509* taxa (Figure 7A). We designated this clade as the superfamily of *MIR7122* (super-*MIR7122*). Unexpectedly, *MIRNA* genes of miR391 members, previously classified into the miR390 family (Xie et al., 2005; Cuperus et al., 2011), were grouped into a separate clade with *MIR4376*, *MIR1432*, *MIR3627* and *MIR5225*. Pairwise sequence comparisons revealed that miR391 showed greater identity with miR4376 compared to miR390, especially for mdm-miR391 and gma-miR4376, which has only two nucleotide differences (Figure 7A; see Supplemental Figure 4C online). We identified a complementary site for ath-miR391 in the *Arabidopsis* gene coding for Ca^{2+}-ATPase 10 [*AtACA10* (AT4G29900)] (see Supplemental Table 10 online), which is also suggestive of the closer evolutionary relationship of *MIR391* to *MIR4376*, because miR4376 has been proven to target *AtACA10* homologous genes in members of the Solanaceae (Wang et al., 2011). The monocot-specific miR1432 also possesses high degree of sequence identity with the miR4376 (Figure 6D), which was reported to be present in eudicots in a lineage-specific manner (Wang et al., 2011),

implying miR1432 is likely to be the variant of miR4376 in monocots. This is further supported by the observation that osa-miR1432 is also predicted to target a gene coding for Ca^{2+}-ATPase (Os04g51610) (see Supplemental Table 10 online) (Lu et al., 2008; Sunkar et al., 2008). Similarly, miR5225 and miR3627 showed a high degree of sequence identity with miR4376 and were predicted to target *AtACA10* homologues (see Supplemental Figure 4D and 4E and Supplemental table 10 online). These observations support the accuracy of our phylogenetic tree analysis (Figure 7A), thereby classifying *MIR391*, *MIR1432*, *MIR4376*, *MIR3627* and *MIR5225* as members of the *MIR4376* superfamily (super-miR4376).

Overall, super-*MIR7122* showed a close relationship with super-*MIR4376*, which was more closely related to *MIR390* (Figure 7A). Therefore, we propose that super-*MIR7122* evolved from super-*MIR4376*, which originated from the *MIR390* family (Figure 7A).

4.2.9 miR7122 and the*TAS*-*PPR*-siRNA pathway emerged only in eudicot plants

To further explore the proposed evolutionary path from *MIR390*, we examined the extent of conservation of the three (super-)families using extensive publicly-available data (small RNAs and genomes). As previously reported, miR390 is present in the common ancestor of all embryophytes (land plants) (Cuperus et al., 2011) (Figure 7B). For the miR4376 superfamily, we were able to identify 24 homologues (representing miR4376, miR1432, miR3627, miR5225) in 16 plant species besides those deposited in miRbase (see Supplemental Figure 7 online). The most ancient super-miR4376 homologue was identified from *Picea abies*, a Pinaceae plant belonging to the Coniferophyta. Two super-miR4376 homologues were also retrieved from *Amborella trichopoda*, a basal flowering plant (angiosperm). Thus, the super-*MIR4376*s were probably present in the common ancestor of all spermatophyta (seed plants) (Figure 7B). However, super-*MIR7122* emerged much late and only detected in the common ancestor of eudicots (Figure 7B), in agreement with the restricted appearance of the *PPR*-siRNA pathway in the eudicots. Thus, our results further revealed *MIR390* as the common ancestor of both super-*MIR4376* and super-*MIR7122*, and confirmed the delineated evolutionary course of super-*MIR7122* from *MIR390* through super-*MIR4376*.

Accordingly, we inferred the evolutionary history for these three (super-)families as that the *MIR4376* superfamily was probably originated from *MIR390* after the divergence of seed plants at ~360 Mya (million years ago), after which the *MIR1432* family evolved from *MIR4376* following the origin of monocots at ~140 Mya (Figure 7C), and the *MIR7122* superfamily arose after the split of the eudicots from its nearest relative at ~110 Mya (Figure 7C).

4.3 Discussion

We identified a miRNA-*TASL*-*PPR*-siRNA pathway in many eudicots, which is commonly triggered by the miR7122 superfamily. We also elucidated the potential evolutionary path of this miRNA superfamily, thereby expanding our understanding of the process of miRNA evolution.

4.3.1 22-nt sRNAs trigger secondary siRNA production

We identified a group of miRNAs, including miR7122, fve-PPRtri1/2, and miR1509, which are able to trigger secondary siRNA production in either *TASL* genes or *PPR*s. We also discovered their homologous relationship with miR173 in *Arabidopsis* and classified them all together into a miRNA superfamily, super-miR7122. All the members of the super-miR7122, including the well-characterized miR173, are of 22-nt in length and an initial nucleotide “U” (Figure 6C); their miRNA/miRNA* duplexes are of asymmetrical structure with a bulge in the mature miRNA strand (see Supplemental Figure 8 online). The initial nucleotide “U” is vital for loading the miRNA to AGO1, the major effector protein directing RNA target cleavage (Mi et al., 2008; Chen et al., 2010). Even when shift in the processing sites of the mature miRNA has occurred, as for fve-PPRtri1/2 and miR173 (Figure 6A), the initial nucleotide “U” was preserved, consistent with the importance of AGO1-loading for secondary siRNA production (Montgomery et al., 2008b). Although sequence mutation frequently occurred for this group of miRNAs (only four nucleotides are completely conserved, Figure 6C), leading to nucleotide and position variation in the single bulge in the mature miRNA strand (see Supplemental Figure 8 online), the asymmetric miRNA/miRNA* duplex persisted, in agreement with

the requirement of an asymmetric structure for processing miRNAs of 22-nt in length, an important feature for a canonical miRNA trigger capable of initiating secondary siRNA production (Chen et al., 2010; Cuperus et al., 2010; Manavella et al., 2012).

The production of 22-nt tasiRNAs was found to be prevalent in plants, and many of these 22-mers were capable of triggering secondary siRNA production in either *TASL* genes or *PPR*s in different species. This was described for the *TAS2* in *Arabidopsis* (Yoshikawa et al., 2005; Chen et al., 2007; Howell et al., 2007), *PpTASL1/2* in peach, *MdTASL1* in apple, *GmTASL-L1/2* in soybean and *MtTASL-L1* in Medicago, as well as *NtTASL1/2* in tobacco (see Supplemental Table 9 online). This observation is consistent with the importance of the 22-nt length in triggering secondary siRNA production, but it is hard to reconcile with a recent proposal that an asymmetric bulge in the miRNA-duplex structure is necessary for secondary siRNA production, regardless of miRNA or miRNA* length (Manavella et al., 2012), because tasiRNAs are derived from double-stranded RNAs which are converted from miRNA-cleaved single-stranded RNAs, therefore are perfectly paired and presumably free of any asymmetric bulge (Peragine et al., 2004; Vazquez et al., 2004; Allen et al., 2005; Yoshikawa et al., 2005). Further, miR828, which is able to trigger siRNA production in *TAS4* transcripts and *MYB* genes (Rajagopalan et al., 2006; Xia et al., 2012; Zhu et al., 2012), is produced from a symmetrical duplex without an asymmetric bulge. Thus it appears that mechanisms underlying miRNA triggering of phasiRNA production in plants are far more complex.

4.3.2 miR7122-mediated regulation of *PPR* genes is ubiquitous but mechanistically complex in eudicots

We were able to identify the miR7122-*TAS*-*PPR*-siRNA pathway in nine different eudicots representing six plant families, indicating that this pathway is widely present in eudicots. For those pathways (as summarized in Figure 3), the initial miRNA triggers are homologous to each other and the processing of *PPR* genes into phasiRNA is conserved among different species. However, the intermediate components forming these regulatory circuits are highly dynamic from species to species. The miR7122 homologues can either target the *PPR* genes directly or act through *TASL* genes to initiate the phasiRNA production from *PPR* genes; the *TASL*-tasiRNA interaction can be multi-layered, as in

legume plants, and the miRNA triggers function through either 1_{22} or 2_{22} modes to activate *TASL* or *PPR* phasiRNA biogenesis. We found that two-layers of *TASL*-tasiRNA interactions are integrated into a single *trans*-acting circuit in soybean and Medicago, i.e., miR1509 acts through the *TASL*-L1-tasiRNA-D2(-) (a first layer) and *TASL*-L2-tasiRNAs (a second layer) to initiate phasiRNA production from *PPR* genes (Figure 5C). This second layer could have an advantage of adding more regulatory flexibility to the trans-acting circuit. While most of the *TASL* or *PHAS PPR* genes are triggered via 1_{22}-acting miRNAs, we were able to identify three 2_{22} *PHAS* loci including *MdTASL1*, *GmTASL-L1* and *MtTASL-L1*. In contrast to the first 2_{22} *PHAS* gene (Medtr7g012810) reported in Medicago in which two target sites are cleavable by miR1509 and phasiRNAs are produced only from the region flanked by the two sites (Zhai et al., 2011), we identified cleavage at only the 5' target site of miR1509 in *GmTASL-L1* and *MtTASL-L1* while phasiRNA production was also detected in the region downstream from the 3' site (see Supplemental Figure S3 online). A similar phenomenon was observed for a *PHAS NB-LRR* gene in tomato, with two target sites of the trigger miR482 (5' cleavable target site and 3' noncleavable site) and siRNA produced all the way through the 3' site (Shivaprasad et al., 2012). For *MdTASL1*, although we failed to detect the cleavage of the 3' site, phasiRNAs were produced only from the region flanked by the two target sites, and low abundant siRNAs were found to be in phase with the 3' cleavage site, implying that the *MdTASL1* is more similar to the Medtr7g012810 in which both sites are cleavable. Comparing these two types of 2_{22} loci, it is conceivable that two target sites might be of unequal importance, as indicated in *MdTASL1* where the 5' mdm-miR7122 target site may have more functional importance than the 3' site since it is the 5' site that sets the phase for the biogenesis of the two functional essential tasiRNA triggers of *PPR* siRNA production. Thus, the less important site may become nonfunctional or even lost over time, unlike the classic *TAS3* 2_{21} triggers in which the non-cleaving 5’ site is required (Axtell et al., 2006; Montgomery et al., 2008a). An alternative explanation is that the 3' site of *MdTASL1* may serve as a protein-binding site to set the boundary of siRNA production, like the 5' non-cleavable miR390 site in *AtTAS3* which is essential for miR390:AGO7 complex binding to terminate the RDR6-mediated synthesis of complementary RNA (Rajeswaran and Pooggin, 2012).

In addition to miR7122 homologues, other miRNAs able to trigger *PPR* phasiRNA production have been identified in some species, including miR400 and miR161 in *Arabidopsis* (Howell et al., 2007), csi-miRC1 in citrus, and miR1508 in soybean. Intriguingly, miR1508 is missing from Medicago, a close relative of soybean, and miR6427, instead of a miR7122 homologue, was identified as the trigger of *PPR* phasiRNA biogenesis in poplar, illustrating both the dynamism and prevalence of the miRNA-*TASL-PPR*-siRNA pathway in eudicots.

4.3.3 Evolution of the *MIR7122* superfamily from the widely conserved *MIR390* via the *MIR4376* superfamily is driven by target variation

We demonstrated that the miRNA triggers of the *PPR*-siRNA pathway (miR7122, miR1509, fve-PPRtri1/2 and miR173) in different species are of a common origin and belong to the same miR7122 superfamily. Unlike other miRNA (super-) families targeting a large family of *PHAS* genes, members of the miR7122 superfamily are of great sequence diversity, which could be attributed to two causes: (1) processing shifts, particularly apparent in miR173 and fve-PPRtri2, in which the miRNA sequence shifts as much as five nucleotides toward the 5' end (Figure 6A and C); and (2) point mutations, i.e. only eight nucleotide positions were highly conserved (over 90% identity) across all of the species examined (Figure 6C). It is known that miR828 and the miR482/2118 superfamily directly cleave their target genes (*MYB*s and *NB-LRR*s, respectively) instigating phasiRNA production without an intermediate role of *TAS* genes. These two miRNA (super-) families share the common feature of targeting highly conserved regions in their target genes (Zhai et al., 2011; Xia et al., 2012). However, the target sequences for miR7122 homologues and their ensuing tasiRNAs show high sequence diversity and are randomly distributed within *PHAS PPR* genes, which have low levels of sequence similarity in any given genome (see Supplemental Figure 6). In contrast to the *MYB* and *NB-LRR* families which are defined by highly-conserved functional domains, the large *PPR* family typically consists of hundreds of members and is characterized by two to 35 tandem repeats of a highly degenerate, 35 amino acid motif (Lurin et al., 2004; Shivaprasad et al., 2012). The *PPR* coding regions are generally tolerant to

nonsynonymous mutation and share little identity among family members at the nucleotide level. This dynamic nature of *PPR* genes could drive the rapid evolution of miRNA or other sRNA triggers for phasiRNA production to achieve proper regulation of gene function, which could account for the significant sequence divergence observed in the super-miR7122 and the requirement of a variety of tasiRNAs via one or two layer(s) of regulatory complexity. In fact, the sequence diversity of non-coding *TASL* genes would provide more flexibility for the derived phasiRNAs or tasiRNAs to meet rapid sequence change in the *PPR* genes in order to down-regulate them effectively. Hence, we propose that the rapid variation of targeting sequences in *PPR* or *TASL* genes could eventually drive the evolution of miRNA triggers (Figure 8), leading to the high sequence diversity within the super-miR7122. The change of targeting sequences caused by point mutation (① in Figure 8) would result in the switch of a functional, mature miRNA (from "A" to "C"), and give rise to the preferential processing of the functionally advantageous variant "C", substituting the loss-of-function "A" variant (Stage I to Stage II). Continued selection would further reinforce the preferential processing of "C", ultimately generating a new miRNA (Stage III).

We also elucidated a novel miRNA evolutionary pathway, in which the miR7122 superfamily is potentially evolved from another superfamily (super-miR4376), which originated from the conserved family miR390 (Figure 7). These three miRNA (super-) families regulate distinct genes of diverse biological functions, with the miR390 regulating several auxin responsive factors through *TAS3*-derived tasiARFs and the miR4376 superfamily targeting genes coding for Ca^{2+}-ATPase (Axtell et al., 2006; Wang et al., 2011), while the miR7122 superfamily regulates *PPR* genes. Undoubtedly, diversification of target genes and selection pressure to diversify the proteins they encode can diversify the miRNA targeting sequences, serving to drive miRNA evolution. Therefore, the *MIRNA* evolution from *MIR390* to super-*MIR7122* via super-*MIR4376* is as well driven by the target variation (Figure 8). Briefly, after the initial gene duplication, one of the duplicated *MIRNA* genes is preserved to maintain the original essential function, while the other would be subjected to extensive mutation over time to create different miRNA variants, some of which could match or acquire new gene targets (② in Figure 8). The new function for the miRNA variant "C", initially gained by chance, may

confer adaptive advantages under conditions of stress or reproductive isolation and stimulate the preferential processing of variant "C" (Stage II). This neofuctionalization process would eventually lead to the fixation of a new miRNA (Stage III). This model of "target-driven miRNA evolution" also fits to the super-miR4376 homologues, whose target sites are not well conserved, instead are of relatively high diversity in their target genes (see Supplemental Table 10 online) (Wang et al., 2011).

In both inter- and intra- superfamily evolution, sequence mutations would evenly occur in the foldback sequence of *MIRNA* genes, but fewer mutations would be retained in the miRNA/miRNA* duplex than other regions of *MIRNA* foldback sequences because of a higher purifying selection pressure (Fahlgren et al., 2010; Ma et al., 2010). This is apparent in the low sequence identity outside of the 13 bp core sequence among all the miRNAs from these three (super-) families (Figure 6D). Mutations outside of the core sequence are not without effect, however, as they would impact the formation of secondary stem-loop structure, e.g. the loop position and stem length (Song et al., 2010), possibly causing shifts in processing sites and giving rise to various miRNA variants. But only those mutations beneficial to the preferential processing of the advantageous new miRNA ("C" in Figure 8) would be selectively preserved. In the long term, the sequence similarity between the newly spawned MIRNA and its ancestral *MIRNA* would be gradually diluted due to mutation in the foldback sequences and miRNA processing changes.

The identification of two plant miRNA superfamilies, super-miR7122 and super-miR4376, and the discovery of the evolutionary path from *MIR390* to super-*MIR7122* via super-*MIR4376* provide strong evidence for a sustained, deep evolutionary history of these plant miRNAs. We have linked one of the most conserved plant miRNAs (miR390) to narrowly-conserved or lineage-specific miRNAs (miR173). Variation of target sequences has served as one of the major driving forces in this divergence. To date, the origin of most plant *MIRNA*s, especially highly conserved *MIRNA*s, remains largely unknown, although several evolutionary routes have been proposed. This includes the generation of new *MIRNA* genes from inverted gene duplication of their target genes (Allen et al., 2004), random foldback sequences (Felippes et al., 2008), or miniature inverted-repeat transposable elements (Piriyapongsa and Jordan, 2008). All of these

routes require the accumulation of unpaired bases in the foldback sequences, acquisition of DCL1 as a processing enzyme, and subsequent production of discrete miRNA species, all in the context of coevolution with the miRNA targets (Voinnet, 2009). By contrast, generating a new *MIRNA* from a pre-existed *MIRNA* would dramatically reduce the "cost" of adaption to the miRNA-biogenesis pathway, because the pre-existed *MIRNA* has already provided a "draft" of the structure for the new *MIRNA* - the canonical stem-loop structure and the DCL1-dependency required for miRNA biogenesis. Therefore, this route of spawning new miRNAs should be more efficient, resembling the "birth-and-death" cycle for many protein-coding genes (Nei and Rooney, 2005).

4.3.4 Possible function of the miR7122-*TAS*-*PPR*-siRNA pathway in plants

The wide conservation of the miR7122-*TAS*-*PPR*-siRNA pathway in eudicots suggests it assumes important biological functions. Most PPR proteins have nucleotide sequence-specific binding activity, and have been proposed to be molecular adaptors directing RNA processing complexes to target RNAs in mitochondria or chloroplasts (Lurin et al., 2004; Schmitz-Linneweber and Small, 2008). PPR proteins can be separated into two major classes based on the nature of their PPR motifs, the P and PLS class (Schmitz-Linneweber and Small, 2008). In plants, the P class of PPR proteins, to which phasiRNA production is restricted, may be involved in plant fertility, as discerned from roles of these proteins in cytoplasmic male sterility (CMS) (Howell et al., 2007; Schmitz-Linneweber and Small, 2008). To produce hybrid seed, nuclear "restorer" (*Rf*) genes, encoding P-class PPR proteins in most cases, are capable of preventing the expression of mitochondrial CMS-inducing genes, which encode hydrophobic CMS-specific polypeptide (Schmitz-Linneweber and Small, 2008). Intriguingly, we failed to identify a similar *PPR*-siRNA pathway in monocots, although most of these *Rf* genes seem to share common characteristics in plants, even from species as divergent as rice and *Arabidopsis* (Schmitz-Linneweber and Small, 2008). The reason for the absence of *PPR*-siRNA pathway is probably the lack of miRNA triggers in monocots, in which the super-miR4376 ancestor has evolved into a 21-nt miR1432 instead of a 22-nt miR7122 homologue. Thus, the biological function of the *PPR*-siRNA pathway may be substituted

by other sRNA-involved mechanisms in monocots, as suggested in a recent study that a non-coding RNA locus producing a 21 bp sRNA, likely a phasiRNA, was characterized as an important regulator of male sterility in rice (Ding et al., 2012; Zhou et al., 2012). Additionally, miR4376, a miRNA evolutionarily related to miR7122, is also involved in reproductive growth through regulating an *ACA10* gene and triggering phasiRNA production (Wang et al., 2011). Taken together, we speculate that the miR7122-*TAS*-*PPR*-siRNA pathway may play essential roles in the reproductive processes of plants, an important topic for future investigations.

4.4 Materials

4.4.1 Plant material and RNA isolation for strawberry

Plant tissue collection, RNA isolation and sRNA sequencing for peach (*Prunus persica* cv. Lovell) and apple (*Malus* × *domestica* cv. Golden delicious) were reported and described early in Zhu et al. (2012) and Xia et al (2012), respectively. Sequencing of small RNAs in Diploid strawberry (*Fragaria vesca*) was similarly performed. Total RNAs were isolated from 1 day-opening flowers of the inbred line 'Yellow Wonder 5AF7' grown in growth chambers (Hollender et al., 2012) using RNeasy Plant Mini Kit (Qiagen, Valencia, CA). Small RNAs were enriched by precipitating the supernatant passing through Qiashedder spin column. The recovered small RNA samples were used for library construction and deep sequencing in Genomics Resources Core Facility at Weill Cornell Medical College (New York, NY).

4.4.2 Bioinformatic analysis

Publicly available deep-sequencing data from peach, apple, soybean, medicago, grape and other species were downloaded from Genbank Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), and original SRA files were dumped into FASTQ files using the SRA Toolkit (http://www.ncbi.nlm.nih.gov/sra). FASTA toolkits (http://hannonlab.cshl.edu/fastx_toolkit/) was used for data processing, including format conversion, adaptor trimming and read collapsing. Read mapping was conducted using Bowtie (Langmead et al., 2009) with no mismatches allowed, and Vienna RNA package

(Hofacker, 2003) or Mfold (Zuker, 2003) was used for the secondary structure prediction of miRNA precursors. Multiple alignments were performed using Clustalw2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) with default settings, and viewed with Jalview (http://www.jalview.org/). Sequence logo (Figure 4D) was produced by GENIO/logo (http://www.biogenio.com/logo/logo.cgi). miRNA target prediction and confirmation were conducted by combining the utilization of Targetfinder (http://carringtonlab.org/resources/targetfinder) and Cleveland 2.0 (Addo-Quaye et al., 2009). The p-value plot (Figure 1B) was produced using the R package (http://www.r-project.org/). Small RNA abundance was normalized to RPM (reads per million), and PARE reads were normalized to TPTM (transcripts per ten million).

4.4.3 Computational calculation of p-value and phasing score

After mapping sRNA reads to the reference genome, unique sRNAs were denoted with their matching coordinates. 2-nt positive offset was added for sRNA matching to the antisense strand, because of the existing of 2-nt overhang at the 3' end of sRNA duplex. A transcriptome or genome wide search was performed using a 9-cycle sliding window (189 bp) with each shift of 3-cycle (63 bp), and windows were reported when ≥ 10 unique reads fell into a 9-cycle window, $\geq 50\%$ of matched unique reads were 21-nt in length, and with ≥ 3 unique reads fell into a certain register. Next reported windows with overlapping region were combined into a single longer window. Then a p-value was calculated for each windows based on the mapping results using an algorithm similar that of Chen (2007) with optimization.

$$\Pr(X = k) = \frac{\binom{20m}{n-k}\binom{m}{k}}{\binom{21m}{n}}$$

$$P\ value: p(k) = \sum_{X=k}^{m} \Pr(X)$$

where n = number of total unique 21-nt sRNAs matched within a window; k = number of unique 21-nt sRNAs fall into the maximum register ($k \geqslant 3$); and m = number of 21-nt phases within a window.

Phasing score were calculated using the algorithm developed by De Paoli (2009) within a 9-cycle window (phase cycle length was set to 21-nt).

$$phasing\ score = ln\left[\left(1 + 10 \times \frac{\sum_{i=1}^{9} P_i}{1 + \sum U}\right)^{k-2}\right], k \geq 3$$

ere n = number of phase cycle positions occupied by at least one 21-nt sRNA within a 9-cycle window; *P* = total number of reads for all 21-nt sRNAs falling into a given phase within a 9-cycle window; and *U* = total number of reads for all 21-nt sRNA falling out of the given phase.

Phasing score data were viewed using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

4.4.4 Deep sequencing data and reference sequence files

All sRNA deep sequencing and PARE data are accessible in the GenBank GEO with the corresponding accession number listed in Supplemental Table 11. Transcriptome and genome files for the Rosaceae plants (apple, peach, and strawberry) and other plants were downloaded from the Genome Database for Rosaceae (GDR, http://www.rosaceae.org) and the Phytozome (http://www.phytozome.net), respectively. MiRNA and foldback sequences were retrieved from the miRbase (version 19, www.mirbase.org).

4.4.5 Phylogenetic analysis

Nucleotide sequences, consisting of the miRNA, miRNA* and 5-nt flanking sequences at each end, of 95 *MIRNA* genes were aligned using the IUB DNA weight matrix by Clustalw2, with an open gap penalty of 10 and an extension gap penalty of 0.1 in pairwise alignments, an extension gap penalty of 0.2 in multiple alignments, and a gap distances penalty of 5. Manual refinement was conducted for the nucleotide alignment by Jalview (shown in Supplemental file 1). The phylogenetic tree was inferred using the Maximum Likelihood method based on the Tamura-Nei model by MEGA5 (Tamura et al., 2011). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Uniform rate variation among site was enabled. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise

distances estimated using the Maximum Composite Likelihood approach, and then selecting the topology with superior log likelihood value. All positions with less than 50% site coverage were eliminated. That is, fewer than 50% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 62 positions in the final dataset.

4.4.6 5'-RLM-RACE

Following the manufacturer's instructions for the First-Choice RLM-RACE Kit (Ambion, Austin, TX, USA), 5 μg of total RNA isolated from apple flower was used for ligating 5' RNA adaptors at 15°C overnight. Two specific primers (1st: 5'-CTGCCCTCCACCATTTCTTTG-3' and 2nd: 5'-AGCTAAATCACACGCTTCAAAC-3') were designed to conduct nested PCRs, and PCR products were cloned to the pGEM-easy vector (Promega, Madison, WI, USA) and sequenced by Bechman Coulter Genomics (Danvers, MA, USA).

Supplemental Data

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

Supplemental Figure 1 SiRNA distribution along *PpTASL1/2*, a peach *PPR* gene and *MdTASL1*.

Supplemental Figure 2 MiRNA-*TASL*-*PPR*-siRNA pathway is conserved in strawberry.

Supplemental Figure 3 Two layers of trans-acting interaction involved in the miR1509-*TAS*-*PPR*-siRNA pathway in soybean and medicago.

Supplemental Figure 4 Alignment of *MIRNA* genes.

Supplemental Figure 5 Stem-loop structure of miRC1 in citrus.

Supplemental Figure 6 Phylogenetic analysis of *PHAS PPR* genes and distribution of target sites of miRNAs or tasiRNAs along *PPR* repeats.

Supplemental Figure 7 Stem-loop structures of newly identified miRNA homologues of the miR4376 superfamily.

Supplemental Figure 8 Stem-loop structures of *MIR7122* homologues identified.

Supplemental Table 1 *PHAS* genes identified in peach.

Supplemental Table 2 *PHAS PPR* genes identified in peach.

Supplemental Table 3 *PpTASL1/2*-derived tasiRNAs predicted to target *PPR* genes.

Supplemental Table 4 *PHAS PPR* genes identified in apple.

Supplemental Table 5 *MdTASL1*-derived tasiRNAs predicted to target *PPR* genes.

Supplemental Table 6 *PHAS PPR* genes identified in strawberry.

Supplemental Table 7 *PHAS* genomic loci or other genes targeted by fve-PPRtri1/2 in strawberry.

Supplemental Table 8 *PHAS PPR* genes identified in soybean and medicago.

Supplemental Table 9 *PHAS PPR* genes or genomic loci related to *PPR*-siRNA pathways in other species.

Supplemental Table 10 Predicted target genes for miR391, miR1432, miR5225 and miR3627.

Supplemental Table 11 Deep sequencing data used for analysis.

Supplemental File 1 Multiple alignment of *MIRNA* excerpted sequences.

Supplemental File 2 Multiple alignment of deduced amino acid sequences of *PHAS PPR* genes.

Acknowledgements

We thank Dr. Tony Wolf and the Department of Horticulture at Virginia Tech for financial support to R.X. We also thank members of the Liu and Meyers labs for critical discussions and valuable suggestions.

Author contributions

R.X. and Zo. L. designed the research, analyzed the data, and wrote the article. B.C.M. analyzed the data and wrote the article. Zh. L. and S. Y. performed the research. E.P.B. designed the research and wrote the article.

4.5 References

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

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

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

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

Baulcombe, D. (2004). RNA silencing in plants. Nature **431,** 356-363.

Chen, H., Li, Y., and Wu, S. (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **104,** 3318-3323.

Chen, H., Chen, L., Patel, K., Li, Y., Baulcombe, D.C., and Wu, S. (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc. Natl. Acad. Sci. USA **107,** 15269-15274.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and functional diversification of miRNA genes. Plant Cell **23,** 431-442.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. Nat. Struct. Mol. Biol. **17,** 997-1003.

De Paoli, E., Dorantes-Acosta, A., Zhai, J., Accerbi, M., Jeong, D.H., Park, S., Meyers, B.C., Jorgensen, R.A., and Green, P.J. (2009). Distinct extremely abundant siRNAs associated with cosuppression in petunia. RNA **15,** 1965-1970.

Ding, J., Shen, J., Mao, H., Xie, W., Li, X., and Zhang, Q. (2012). RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. Mol. Plant **5,** 1210-1216.

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

Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C. (2010). MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. Plant Cell **22,** 1074-1089.

Felippes, F.F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of *Arabidopsis* thaliana microRNAs from random sequences. RNA **14,** 2455-2459.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. **31,** 3429-3431.

Hollender, C.A., Geretz, A.C., Slovin, J.P., and Liu, Z. (2012). Flower and early fruit development in a diploid strawberry, *Fragaria vesca*. Planta **235,** 1123-1139.

Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C. (2007). Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell **19,** 926-942.

Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.-L., Walbot, V., Sundaresan, V., Vance, V., and Bowman, L.H. (2009). Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res. **19,** 1429-1440.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. **57,** 19-53.

Klevebring, D., Street, N.R., Fahlgren, N., Kasschau, K.D., Carrington, J.C., Lundeberg, J., and Jansson, S. (2009). Genome-wide profiling of *Populus* small RNAs. BMC Genomics **10,** 620.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10,** R25.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. **11,** 204-220.

Lu, C., Jeong, D., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B.C., and Green, P.J. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc. Natl. Acad. Sci. USA **105,** 4951-4956.

Luo, Q.J., Mittal, A., Jia, F., and Rock, C.D. (2012). An autoregulatory feedback loop involving *PAP1* and *TAS4* in response to sugars in *Arabidopsis*. Plant Mol. Biol. **80,** 117-129.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16,** 2089-2103.

Ma, Z., Coruh, C., and Axtell, M.J. (2010). *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. Plant Cell **22,** 1090-1103.

Manavella, P.A., Koenig, D., and Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. Proc. Natl. Acad. Sci. USA **109,** 2461-2466.

Matzke, M., Kanno, T., Claxinger, L., Huettel, B., and Matzke, A.J.M. (2009). RNA-mediated chromatin-based silencing in plants. Curr. Opin. Cell Biol. **21,** 367-376.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., and Zhu, J.K. (2008). Criteria for annotation of plant microRNAs. Plant Cell **20,** 3186-3190.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. Cell **133,** 116-127.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008a). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in *TAS3* trans-acting siRNA formation. Cell **133,** 128-141.

Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H., and Carrington, J.C. (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. Proc. Natl. Acad. Sci. USA **105,** 20055-20062.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. Mol. Biol. Evol. **25,** 1120-1128.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H., and Poethig, R. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNA in *Arabidopsis*. Genes Dev. **18,** 2368-2379.

Piriyapongsa, J., and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. RNA **14,** 814-821.

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

Rajeswaran, R., and Pooggin, M.M. (2012). RDR6-mediated synthesis of complementary RNA is terminated by miRNA stably bound to template RNA. Nucleic Acids Res. **40,** 594-599.

Rajeswaran, R., Aregger, M., Zvereva, A.S., Borah, B.K., Gubaeva, E.G., and Pooggin, M.M. (2012). Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. Nucleic Acids Res. **40,** 6241-6254.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. **29,** 24-26.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. **13,** 663-670.

Shivaprasad, P.V., Chen, H.M., Patel, K., Bond, D.M., Santos, B.A., and Baulcombe, D.C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell **24,** 859-874.

Si-Ammour, A., Windels, D., Arn-Bouldoires, E., Kutter, C., Ailhas, J., Meins, F., and Vazquez, F. (2011). miR393 and secondary siRNAs regulate expression of the TIR1/AFB2 auxin receptor clade and auxin-related development of *Arabidopsis* leaves. Plant Physiol. **157,** 683-691.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., and Zhu, J.K. (2008). Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC Plant Biol. **8,** 25.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28,** 2731-2739.

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

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell **136,** 669-687.

Wang, Y., Itaya, A., Zhong, X., Wu, Y., Zhang, J., van der Knaap, E., Olmstead, R., Qi, Y., and Ding, B. (2011). Function and evolution of a MicroRNA that regulates a *Ca^{2+}-ATPase* and triggers the formation of phased small interfering RNAs in tomato reproductive growth. Plant Cell **23,** 3185-3203.

Williams, L., Carles, C.C., Osmont, K.S., and Fletcher, J.C. (2005). A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis ARF2*, *ARF3*, and *ARF4* genes. Proc. Natl. Acad. Sci. USA **102,** 9703-9708.

Xia, R., Zhu, H., An, Y.Q., Beers, E.P., and Liu, Z. (2012). Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol. **13,** R47.

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

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

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., González, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., and

Meyers, B.C. (2011). MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. **25,** 2540-2553.

Zhang, C., Li, G., Wang, J., and Fang, J. (2012). Identification of trans-acting siRNAs and their regulatory cascades in grapevine. Bioinformatics **28,** 2561-2568.

Zhou, H., Liu, Q., Li, J., Jiang, D., Zhou, L., Wu, P., Lu, S., Li, F., Zhu, L., Liu, Z., Chen, L., Liu, Y.G., and Zhuang, C. (2012). Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. Cell Res. **22,** 649-660.

Zhu, H., Xia, R., Zhao, B., An, Y.Q., Dardick, C.D., Callahan, A.M., and Liu, Z. (2012). Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol. **12,** 149.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31,** 3406-3415.

Chapter 5

Network of miR482/2118 regulating *PHAS NB-LRR* defense genes in apple and peach

5.1 Introduction

Plants have adapted diverse strategies to counter pathogen invasion. In essence, there are two types of immunity triggered by distinct innate immune receptors (Jones and Dangl, 2006). The primary immune response uses transmembrane pattern recognition receptors (PRRs) to respond to slowly evolving pathogen-associated molecular pattern (PAMPs), thus is regarded as PAMP-triggered immunity (PTI) (Chisholm et al., 2006; Jones and Dangl, 2006). The PTI can be suppressed by some pathogens through delivering effector proteins to the plant cell, in response to which the host plants can deploy surveillance proteins (R proteins) to either directly or indirectly monitor the presence of the pathogen effector proteins, triggering the effector-triggered immunity (ETI) (Chisholm et al., 2006; Jones and Dangl, 2006). ETI is an accelerated and amplified PTI response, usually resulting in a hypersensitive cell death response (HR) at the infection site (Jones and Dangl, 2006). The majority of R genes encode intracellular immune NB-LRR proteins, which are named after their characteristic nucleotide binding (NB) and leucine-rich repeat (LRR) domains.

Recently abundant phasiRNA were found to be produced from *NB-LRR* genes, with 22-nt miRNAs from the miR482/2118 superfamily identified as the common trigger by targeting at a region coding for the conserved P-loop motif (Zhai et al., 2011; Li et al., 2012; Shivaprasad et al., 2012). Some of the secondary siRNAs derived from *NB-LRR* genes are potentially regulating some immunity-related genes (Shivaprasad et al., 2012). Therefore, the miR482/2118-*NB-LRR*-siRNA pathway may not only act as a counter-counterdefense system that play crucial roles in immune responses upon pathogen infection, but also serve as a regulatory circuit to minimize plant fitness cost when infection pressure is low (Li et al., 2012; Shivaprasad et al., 2012). We found the miR482/2118-*NB-LRR*-siRNA pathway is also conserved in peach and apple with multiple members of the miR482/2118 superfamily and over 100 *PHAS NB-LRR* genes identified in both of them. We also found that certain miR482/2118 variants targeted putative non-coding genomic loci and triggered phasiRNA production in addition to *NB-LRR* genes.

5.2 Results

5.2.1 miR482/2118 homologs in apple and peach

The miR482/2118 superfamily is widely conserved in plants. The miR482/2118 was also found to be present in apple and peach. As described in Shivaprasad et al. (2012), there is a 2-nt mutual sequence shift between miR482 and miR2118, with the first position of miR2118 aligned to the third position of miR482. Following this rule, we identified four *MIR2118* genes and two *MIR482* genes in apple with three *MIR2118* genes (mdm-*MIR2118a/b/c*) and the two *MIR482* genes (mdm-*MIR482a/b*) generating miRNAs with the same sequence (Table 5-1). In peach, totally nine *MIR482/2118* genes were encoded in the genome, including six *MIR2118* genes and three MIR482 genes, two of which (ppe-*MIR482b/c*) generate an identical miRNA sequence (Table 5-1). In apple, the miR482/2118 family possesses sequence divergence in eight positions within the aligned region in addition to the 2-nt shift, with mdm-miR2118d showing higher identity to the mdm-miR482a/b (Fig. 5-1A). Similarly, sequence variation was observed in seven positions of the aligned region of the peach miR482/2118 superfamily (Fig. 5-1B). Surprisingly, six members of the peach *MIR482/2118* are closely located within a 7 kb region in the scaffold_1 of the peach genome, including five *MIR2118* genes and one *MIR482* gene (Table 5-1, Fig. 5-1C). All the mature miRNAs of these six *MIRNA* genes are located in the 3' arm of their stem-loop precursor (Fig. 5-1C).

5.2.2 PHAS NB-LRR genes in apple and peach

As aforementioned, the *NB-LRR* genes were identified as the largest *PHAS* gene family in peach with totally 93 *NB-LRR* genes processing secondary siRNAs (p-value < 0.001) (Table 5-2). Among them, 54 genes bear a complementary target site (AS ≤ 5) for at least one of the miRNA members of the peach miR482/2118 family (Table 5-2). The *PHAS NB-LRR* gene population is even larger in apple. Totally 295 apple *NB-LRR* genes were found to produce phased siRNAs (p-value < 0.001), 85 of which possess a target site of the mdm-miR482/2118 (Table 5-3).

5.2.3 Neofunctionalization of mdm-miR2118 and ppe-miR2118b

Most of the 22-nt miRNAs, capable of triggering phasiRNA production, are of an initial "U" (uridine) nucleotide, which is essential for the miRNA loading to the AGO1 effector, where miRNA-guided mRNA cleavage occurs (Montgomery et al., 2008; Chen et al., 2010; Cuperus et al., 2010). Intriguingly, the mdm-miR2118 (including mdm-miR2118), ppe-miR2118b, and ppe-miR2118f start with a "C" (cytosine) nucleotide, instead of a classical "U". Functional analysis found that the mdm-miR2118 is able to target a set of putative non-coding genomic loci (*TAS*-like genes) and initiate their phasiRNA production, in addition to serve as trigger for a subset of *NB-LRR* genes (Table 5-4). Some of these non-coding loci show a low degree of partial similarity to *NAC* genes, implying these loci may generate tasiRNAs regulating *NAC* genes. Similarly, the ppe-miR2118b also has putative non-coding targets, which produced abundant phasiRNAs as well. Sequence similarity to other known genes is hardly detected for those non-coding *PHAS* genes.

5.4 Discussion

The capacity of *NB-LRR* genes processing phasiRNAs was first reported in legume plants (Zhai et al., 2011), and then was demonstrated to be present more broadly in other plants (Li et al., 2012; Shivaprasad et al., 2012). The widely conserved miR482/2118 was identified as the major trigger, which targets *NB-LRR* genes at the region coding for the conserved P-loop motif (Zhai et al., 2011; Shivaprasad et al., 2012). In addition to the miR482/2118 superfamily, other miRNAs , probably young and species- or lineage-specific, have also evolved to regulate *NB-LRR* genes, including miR6019 and miR6020 in tobacco, miR1507 and miR2109 in medicago and soybean (Zhai et al., 2011; Li et al., 2012). Among them, the miR6019 and miR1507, in contrast to the 21-nt miR6020, are 22-nt in length and able to initiate phasiRNA production. Interestingly, the miR2109 is 22-nt long in medicago but predominantly 21-nt in soybean, indicating miRNA size diversification may allow flexibility in the degree of silencing of target genes due to differences in their ability to trigger phasiRNAs that amplify post-transcriptional silencing (Zhai et al., 2011; Li et al., 2012). The phasiRNAs derived from *NB-LRR* genes may potentially regulate other immunity-related genes in trans in addition to reinforcing the silencing of *NB-LRR* genes in cis (Shivaprasad et al., 2012). Moreover for some

miR482/2118 duplexes in apple and peach, the star strand (miR482/2118*) is of higher abundance than the corresponding miRNAs. Conceivably the strand switch may be of important functions in the entire regulatory network through unknown mechanism. Also we found mdm-miR2118 and ppe-miR2118b targeted non-coding *PHAS* genomic loci other than *NB-LRR* genes. It is possible that *TASL*-derived phasiRNAs provide another layer of gene regulation associated with plant immunity. Additionally, many truncated *NB-LRR* transcripts are produced in plant genomes (Meyers et al., 2003). They may serve as a miRNA competitor to compromise the silencing effect of miRNA triggers. In general, the miRNA-*NB-LRR*-siRNA pathway is a greatly complex regulatory network, playing crucial roles in the plant immune system.

5.5 References

Chen, H., Chen, L., Patel, K., Li, Y., Baulcombe, D.C., and Wu, S. (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proceedings of the National Academy of Sciences.

Chisholm, S.T., Coaker, G., Day, B., and Staskawicz, B.J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. Cell **124,** 803-814.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol **17,** 997-1003.

Jones, J.D., and Dangl, J.L. (2006). The plant immune system. Nature **444,** 323-329.

Li, F., Pignatta, D., Bendix, C., Brunkard, J.O., Cohn, M.M., Tung, J., Sun, H.Y., Kumar, P., and Baker, B. (2012). MicroRNA regulation of plant innate immune receptors. Proc. Natl. Acad. Sci. USA **109,** 1790-1795.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell **15,** 809-834.

Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H., and Carrington, J.C. (2008). AGO1-miR173 complex initiates phased siRNA formation in plants. Proc. Natl. Acad. Sci. USA **105,** 20055-20062.

Shivaprasad, P.V., Chen, H.M., Patel, K., Bond, D.M., Santos, B.A., and Baulcombe, D.C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell **24,** 859-874.

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., González, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., and Meyers, B.C. (2011). MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. **25,** 2540-2553.

Appendix I Tables

Table 4-1 *PHAS* genes/loci identified in peach (*Prunus persica***)**

Trigger miRNA	*PHAS* loci or gene families	Loci or gene number	p-value[a]	References
ppe-miR390	*TAS3*	1	0	Allen et al. 2005; Axtell et al. 2006; Zhu et al. 2012
ppe-miR828	*TAS4*	1	0	Rajagopalan et al. 2006; Zhu et al. 2012
ppe-miR828	*MYB*	5 (1)[b]	0	Xia et al. 2012; Zhu et al. 2012
ppe-miR482/2118	*NB-LRR*	94 (14)	0	Zhai et al. 2011
ppe-miR7122	*PPR*	20 (3)	0	Chen et al. 2007, 2010; Howell et al. 2007; Zhai et al. 2011;
ppe-miR393	*TIR /AFB*	2	0	Chen et al. 2010; Si-Ammour et al. 2011; Zhai et al. 2011
tasiARF	*ARF*	3	9.09E-06	Axtell et al. 2006;
ppe-miR408	*Laccase*	2	4E-10	
ppe-miR4376	*Ca^{2+}-ATPase*	3	0	Wang et al. 2011
NA[c]	*NAC*	6 (3)	3.67E-05	Zhai et al. 2011
ppe-miR168	*AGO*	2 (1)	0.000743	Chen et al. 2010; Zhai et al. 2011
NA	*SGS*	2	0	Zhai et al. 2011
NA	*DCL*	1	0.000386	Zhai et al. 2011
NA	*ATP TR*	5	0	Zhai et al. 2011
NA	*UDP-glucosyl transferase*	5 (3)	0.000533	
NA	*RD22*	17 (1)	4.5E-10	
NA	*Kinase-like*	10 (4)	5.97E-06	Zhai et al. 2011

Note: [a.] median p-value was used for a gene family with ≥ 2 *PHAS* genes identified; [b.] the value in parentheses indicates the number of additional *PHAS* genes with p-value > 0.001 in the corresponding gene family; [c.] NA: not available. ARF: auxin responsive factor; RD22: responsive to dehydration 22; ATP TR: ATP binding/protein binding/transmembrane receptor.

Table 5-1 miRNA members of miR482/2118 superfamily in apple and peach

ID	Sequence	Matching contig	Matching strand	Matching position
mdm-miR2118a	CUACCGAUGCCACUAAGUCCCA	MDC017130.228	+	6809
mdm-miR2118b	CUACCGAUGCCACUAAGUCCCA	MDC019300.125	-	1308
mdm-miR2118c	CUACCGAUGCCACUAAGUCCCA	MDC019300.66	-	1894
mdm-miR2118d	UUCCCAAGCCCGCCCAUUCCUA	MDC002407.840	+	1665
mdm-miR482a	UCUUUCCUAUCCCUCCCAUUCC	MDC006350.123	-	2936
mdm-miR482b	UCUUUCCUAACCCUCCCAUUCC	MDC000859.244	-	731
ppe-miR2118a	UUCCCAAGCCCGCCCAUUCCAA	scaffold_3	+	10579365
ppe-miR2118b	CUACCGAUUCCACCCAUUCCGA	scaffold_1	+	29644701
ppe-miR2118c	UUUCCGAAACCUCCCAUUCCAA	scaffold_1	+	29646139
ppe-miR2118d	UUGCCUAUUCCUCCCAUGCCAA	scaffold_1	+	29646838
ppe-miR2118e	UUGCCAACCCCGCCCAUUCCAA	scaffold_1	+	29648326
ppe-miR2118f	CUUCCCAAACCUCCCAUUCCUA	scaffold_1	+	29648613
ppe-miR482a	UCUUUCCUACUCCACCCAUUCC	scaffold_1	+	29651667
ppe-miR482b	UCUUUCCCAAUCCACCCAUGCC	scaffold_6	+	2760161
ppe-miR482c	UCUUUCCCAAUCCACCCAUGCC	scaffold_6	-	2718614

Table 5-2 *PHAS NBS-LRR* genes in peach

ID	Accession #	miRNA trigger	p-value	Annotation
1	ppa001384m	ppe-miR2118a	0	disease resistance protein (CC-NBS-LRR class)
2	ppa016635m	ppe-miR482a	0	disease resistance protein (CC-NBS-LRR class)
3	ppa021194m	ppe-miR482a	0	disease resistance protein (CC-NBS-LRR class)
4	ppa001212m	ppe-miR482a/b/c	0	disease resistance protein (CC-NBS-LRR class)
5	ppa001003m	ppe-miR482b/c	0	disease resistance protein (CC-NBS-LRR class)
6	ppa015274m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
7	ppa019824m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
8	ppa020450m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
9	ppa020745m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
10	ppa021741m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
11	ppa022876m	ppe-miR2118a	0	disease resistance protein (NBS-LRR class)
12	ppa000343m	ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
13	ppa000391m	ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
14	ppa019071m	ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
15	ppa021541m	ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
16	ppa026310m	ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
17	ppa026334m	ppe-miR2118c	0	disease resistance protein (NBS-LRR class)
18	ppb024266m	ppe-miR2118c/f	0	disease resistance protein (NBS-LRR class)
19	ppa018842m	ppe-miR2118d	0	disease resistance protein (NBS-LRR class)
20	ppa021551m	ppe-miR2118d	0	disease resistance protein (NBS-LRR class)
21	ppa024822m	ppe-miR2118d	0	disease resistance protein (NBS-LRR class)
22	ppa000373m	ppe-miR2118f	0	disease resistance protein (NBS-LRR class)
23	ppa000247m	ppe-miR482a	0	disease resistance protein (NBS-LRR class)
24	ppa000274m	ppe-miR482a	0	disease resistance protein (NBS-LRR class)
25	ppa000335m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
26	ppa000407m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
27	ppa014576m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
28	ppa017078m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)

29	ppa017126m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
30	ppa017399m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
31	ppa019097m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
32	ppa023090m	ppe-miR482a&ppe-miR2118a/e	0	disease resistance protein (NBS-LRR class)
33	ppa016138m	ppe-miR482a&ppe-miR2118d/e	0	disease resistance protein (NBS-LRR class)
34	ppa020719m	ppe-miR482a&ppe-miR2118d/e	0	disease resistance protein (NBS-LRR class)
35	ppa018378m	ppe-miR482b/c	0	disease resistance protein (NBS-LRR class)
36	ppa001315m	ppe-miR2118d	0	disease resistance protein (TIR-NBS-LRR class)
37	ppa000477m	ppe-miR2118e	0	disease resistance protein (TIR-NBS-LRR class)
38	ppb015618m	ppe-miR2118e	0	disease resistance protein (TIR-NBS-LRR class)
39	ppa016623m	ppe-miR482a	0	disease resistance protein (TIR-NBS-LRR class)
40	ppa017433m	ppe-miR482a	0	disease resistance protein (TIR-NBS-LRR class)
41	ppa023180m	ppe-miR482a	0	disease resistance protein (TIR-NBS-LRR class)
42	ppa024626m	ppe-miR482a	0	disease resistance protein (TIR-NBS-LRR class)
43	ppa015762m	ppe-miR2118a	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
44	ppa020740m	ppe-miR2118a	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
45	ppa001015m	ppe-miR2118d	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
46	ppa000953m	ppe-miR482a/b/c&ppe-miR2118d	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
47	ppa001007m	ppe-miR482b/c	1.00E-10	disease resistance protein (CC-NBS-LRR class)
48	ppa026003m	ppe-miR482a	4.38E-08	disease resistance protein (TIR-NBS-LRR class)
49	ppa015104m	ppe-miR482b/c	8.40E-08	disease resistance protein (NBS-LRR class)
50	ppa001182m	ppe-miR2118d	1.63E-06	disease resistance protein (CC-NBS-LRR class)
51	ppa015444m	ppe-miR2118a	3.53E-06	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
52	ppa022023m	ppe-miR482a	8.04E-05	disease resistance protein (TIR-NBS-LRR class)
53	ppa016120m	ppe-miR482a/b/c	0.000106	disease resistance protein (NBS-LRR class)
54	ppa022790m		0	disease resistance protein (CC-NBS-LRR class)
55	ppa023373m		0	disease resistance protein (CC-NBS-LRR class)

56	ppa024579m	0	disease resistance protein (CC-NBS-LRR class)
57	ppa025033m	0	disease resistance protein (CC-NBS-LRR class)
58	ppa026312m	0	disease resistance protein (CC-NBS-LRR class)
59	ppb019722m	0	disease resistance protein (CC-NBS-LRR class)
60	ppa000970m	0	disease resistance protein (NBS-LRR class)
61	ppa001008m	0	disease resistance protein (NBS-LRR class)
62	ppa015125m	0	disease resistance protein (NBS-LRR class)
63	ppa018247m	0	disease resistance protein (NBS-LRR class)
64	ppa019012m	0	disease resistance protein (NBS-LRR class)
65	ppa019037m	0	disease resistance protein (NBS-LRR class)
66	ppa019094m	0	disease resistance protein (NBS-LRR class)
67	ppa019200m	0	disease resistance protein (NBS-LRR class)
68	ppa022119m	0	disease resistance protein (NBS-LRR class)
69	ppa022439m	0	disease resistance protein (NBS-LRR class)
70	ppa024157m	0	disease resistance protein (NBS-LRR class)
71	ppa027039m	0	disease resistance protein (NBS-LRR class)
72	ppb015994m	0	disease resistance protein (NBS-LRR class)
73	ppa000524m	0	disease resistance protein (TIR-NBS-LRR class)
74	ppa015410m	0	disease resistance protein (TIR-NBS-LRR class)
75	ppa018131m	0	disease resistance protein (TIR-NBS-LRR class)
76	ppa022336m	0	disease resistance protein (TIR-NBS-LRR class)
77	ppa023459m	0	disease resistance protein (TIR-NBS-LRR class)
78	ppa024258m	0	disease resistance protein (TIR-NBS-LRR class)
79	ppa024336m	0	disease resistance protein (TIR-NBS-LRR class)
80	ppa025472m	0	disease resistance protein (TIR-NBS-LRR class)
81	ppa001076m	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
82	ppa016482m	0	RPM1 (RESISTANCE TO P. SYRINGAE PV MACULICOLA 1); nucleotide binding / protein binding
83	ppa015433m	0	RPP8 (RECOGNITION OF PERONOSPORA PARASITICA 8); nucleotide binding

84	ppa017506m	0	RPP8 (RECOGNITION OF PERONOSPORA PARASITICA 8); nucleotide binding
85	ppa018885m	0	RPP8 (RECOGNITION OF PERONOSPORA PARASITICA 8); nucleotide binding
86	ppa019683m	1.10E-09	disease resistance protein (NBS-LRR class)
87	ppa023712m	6.50E-09	disease resistance protein (CC-NBS-LRR class)
88	ppa016937m	8.50E-09	disease resistance protein (NBS-LRR class)
89	ppa015461m	5.26E-08	disease resistance protein (NBS-LRR class)
90	ppa025473m	2.80E-05	disease resistance protein (TIR-NBS-LRR class)
91	ppa024306m	0.000217	disease resistance protein (NBS-LRR class)
92	ppa027225m	0.000522	disease resistance protein (NBS-LRR class)
93	ppa016991m	0.000617	disease resistance protein (TIR-NBS-LRR class)

Table 5-3 *PHAS NBS-LRR* genes in apple

ID	Gene ID	miRNA trigger	p-value	Annotation
1	MDP0000292704	miR2118a/b/c	0	TMV_resistance_protein_N
2	MDP0000291494	miR2118a/b/c	0	TMV_resistance_protein_N
3	MDP0000160232	miR482a	0	Putative_disease_resistance_protein_RGA3
4	MDP0000190531	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
5	MDP0000129088	miR482a	0	TMV_resistance_protein_N
6	MDP0000378930	miR482a	0	TMV_resistance_protein_N
7	MDP0000811129	miR482a	0	Disease_resistance_protein_RPM1
8	MDP0000906293	miR482a	0	Disease_resistance_protein_RPM1
9	MDP0000552270	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
10	MDP0000415930	miR482a	0	Putative_disease_resistance_protein_RGA3
11	MDP0000210772	miR482a	0	TMV_resistance_protein_N
12	MDP0000704323	miR482a	0	Putative_disease_resistance_protein_RGA3
13	MDP0000138499	miR482a	0	Putative_disease_resistance_protein_RGA3
14	MDP0000160821	miR482a	0	Disease_resistance_protein_At4g27190
15	MDP0000258186	miR482a	0	TMV_resistance_protein_N
16	MDP0000279585	miR482a	0	TMV_resistance_protein_N
17	MDP0000279752	miR482a	0	TMV_resistance_protein_N
18	MDP0000464475	miR482a	0	Putative_disease_resistance_protein_RGA3
19	MDP0000295886	miR482a	0	TMV_resistance_protein_N
20	MDP0000472864	miR482a	0	TMV_resistance_protein_N
21	MDP0000233739	miR482a	0	Putative_disease_resistance_protein_RGA3
22	MDP0000734561	miR482a	0	Disease_resistance_protein_At4g27190
23	MDP0000772634	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
24	MDP0000266556	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
25	MDP0000312668	miR482a	0	Putative_disease_resistance_protein_RGA4

26	MDP0000161787	miR482a	0	TMV_resistance_protein_N
27	MDP0000223518	miR482a	0	Disease_resistance_protein_At4g27190
28	MDP0000148024	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
29	MDP0000472974	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
30	MDP0000309995	miR482a	0	Putative_disease_resistance_protein_RGA1
31	MDP0000286700	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
32	MDP0000119559	miR482a	0	Putative_disease_resistance_protein_RGA3
33	MDP0000136726	miR482a	0	TMV_resistance_protein_N
34	MDP0000191794	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
35	MDP0000410777	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
36	MDP0000174796	miR482a	0	Disease_resistance_protein_At4g27190
37	MDP0000316884	miR482a	0	Putative_disease_resistance_protein_RGA4
38	MDP0000260531	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
39	MDP0000157081	miR482a	0	Putative_disease_resistance_protein_RGA4
40	MDP0000304601	miR482a	0	Putative_disease_resistance_protein_RGA3
41	MDP0000166800	miR482a	0	Putative_disease_resistance_protein_RGA3
42	MDP0000167895	miR482a	0	Putative_disease_resistance_protein_RGA3
43	MDP0000182552	miR482a	0	TMV_resistance_protein_N
44	MDP0000637744	miR482a	0	TMV_resistance_protein_N
45	MDP0000173380	miR482a	0	TMV_resistance_protein_N
46	MDP0000732188	miR482a	0	Disease_resistance_protein_At4g27190
47	MDP0000174069	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
48	MDP0000213925	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
49	MDP0000724110	miR482a	0	TMV_resistance_protein_N
50	MDP0000810351	miR482a	0	Putative_disease_resistance_protein_RGA3
51	MDP0000287351	miR482a	0	TMV_resistance_protein_N
52	MDP0000168385	miR482a	0	Putative_disease_resistance_protein_RGA3
53	MDP0000195139	miR482a	0	Putative_disease_resistance_protein_RGA4

54	MDP0000211316	miR482a	0	Putative_disease_resistance_protein_RGA3
55	MDP0000214360	miR482a	0	TMV_resistance_protein_N
56	MDP0000232600	miR482a	0	Putative_disease_resistance_protein_RGA4
57	MDP0000275398	miR482a	0	Putative_disease_resistance_RPP13-like_protein_1
58	MDP0000715013	miR482a	0	Putative_disease_resistance_protein_RGA4
59	MDP0000131170	miR482b	0	Disease_resistance_protein_RGA2
60	MDP0000315988	miR482b	0	Disease_resistance_protein_RGA2
61	MDP0000315301	miR482b	0	Putative_disease_resistance_protein_RGA1
62	MDP0000311542	miR482b	0	Disease_resistance_protein_RGA2
63	MDP0000308801	miR482a/b	0	Putative_disease_resistance_protein_RPP8-like_protein_1
64	MDP0000901826	miR482b	0	Putative_disease_resistance_protein_RGA1
65	MDP0000301892	miR482b	0	TMV_resistance_protein_N
66	MDP0000182716	miR482b	0	Putative_disease_resistance_protein_RGA1
67	MDP0000267866	miR482b	0	Disease_resistance_protein_RGA2
68	MDP0000284699	miR482b	0	Disease_resistance_protein_RGA2
69	MDP0000173414	miR482a	7.00E-10	Disease_resistance_protein_RPS5
70	MDP0000164312	miR482b	5.35E-06	Disease_resistance_protein_RGA2
71	MDP0000307119	miR482a	6.16E-06	Putative_disease_resistance_RPP13-like_protein_1
72	MDP0000146388	miR482b	2.75E-05	Disease_resistance_protein_RGA2
73	MDP0000312606	miR482a	6.81E-05	Putative_disease_resistance_protein_RGA3
74	MDP0000215060	miR2118a/b/c	0.000202	TMV_resistance_protein_N
75	MDP0000376600	miR2118a/b/c	0.000202	TMV_resistance_protein_N
76	MDP0000573171	miR2118a/b/c	0.000202	TMV_resistance_protein_N
77	MDP0000738080	miR482a	0.00044	Putative_disease_resistance_protein_RGA4
78	MDP0000811127	miR482a	0.000566	Disease_resistance_protein_RPM1
79	MDP0000284479	miR482a	0.00057	TMV_resistance_protein_N
80	MDP0000302574	miR482a	0.000585	Putative_disease_resistance_protein_RGA4
81	MDP0000825919	miR482a	0.000585	Putative_disease_resistance_protein_RGA4

82	MDP0000220175	miR482a	0.000608	Putative_disease_resistance_protein_RGA3
83	MDP0000272830		0	Putative_disease_resistance_protein_RGA3
84	MDP0000553601		0	TMV_resistance_protein_N
85	MDP0000265485		0	Probable_disease_resistance_RPP8-like_protein_2
86	MDP0000300921		0	Probable_disease_resistance_RPP8-like_protein_2
87	MDP0000272762		0	Putative_disease_resistance_protein_RGA4
88	MDP0000634037		0	TMV_resistance_protein_N
89	MDP0000641906		0	Putative_disease_resistance_RPP13-like_protein_1
90	MDP0000458200		0	Putative_disease_resistance_protein_At1g59780
91	MDP0000296941		0	TMV_resistance_protein_N
92	MDP0000380952		0	Putative_disease_resistance_protein_At4g19050
93	MDP0000309287		0	Putative_disease_resistance_protein_RGA4
94	MDP0000264960		0	TMV_resistance_protein_N
95	MDP0000223201		0	Disease_resistance_protein_RGA2
96	MDP0000295375		0	Putative_disease_resistance_protein_RGA3
97	MDP0000198228		0	Putative_disease_resistance_RPP13-like_protein_1
98	MDP0000624125		0	TMV_resistance_protein_N
99	MDP0000209861		0	Putative_disease_resistance_protein_RGA3
100	MDP0000432746		0	Putative_disease_resistance_protein_At4g19050
101	MDP0000265756		0	TMV_resistance_protein_N
102	MDP0000197733		0	Probable_disease_resistance_RPP8-like_protein_2
103	MDP0000308635		0	Putative_disease_resistance_protein_RGA3
104	MDP0000210751		0	Putative_disease_resistance_protein_RGA1
105	MDP0000569556		0	TMV_resistance_protein_N
106	MDP0000157208		0	Putative_disease_resistance_RPP13-like_protein_1
107	MDP0000166759		0	Putative_disease_resistance_protein_RGA3
108	MDP0000309588		0	Disease_resistance_protein_RGA2
109	MDP0000715969		0	Putative_disease_resistance_protein_At3g14460

110	MDP0000218445	0	Putative_disease_resistance_protein_RGA1
111	MDP0000291677	0	TMV_resistance_protein_N
112	MDP0000263446	0	Putative_disease_resistance_protein_RGA4
113	MDP0000582553	0	Disease_resistance_protein_RGA2
114	MDP0000453695	0	Putative_disease_resistance_protein_At4g19050
115	MDP0000158933	0	Disease_resistance_protein_RGA2
116	MDP0000574827	0	Putative_disease_resistance_protein_RGA1
117	MDP0000157531	0	TMV_resistance_protein_N
118	MDP0000579550	0	TMV_resistance_protein_N
119	MDP0000399716	0	TMV_resistance_protein_N
120	MDP0000759303	0	Disease_resistance_protein_RPM1
121	MDP0000150893	0	Putative_disease_resistance_protein_RGA3
122	MDP0000735861	0	TMV_resistance_protein_N
123	MDP0000292984	0	Putative_disease_resistance_protein_RGA3
124	MDP0000584095	0	TMV_resistance_protein_N
125	MDP0000657897	0	Putative_disease_resistance_protein_At4g19050
126	MDP0000301983	0	Putative_disease_resistance_protein_RGA1
127	MDP0000309192	0	Putative_disease_resistance_protein_At1g50180
128	MDP0000294102	0	Disease_resistance_protein_RGA2
129	MDP0000317557	0	Disease_resistance_protein_RPM1
130	MDP0000426020	0	Disease_resistance_protein_RPM1
131	MDP0000152934	0	Putative_disease_resistance_protein_RGA1
132	MDP0000189702	0	Putative_disease_resistance_protein_RGA4
133	MDP0000184060	0	TMV_resistance_protein_N
134	MDP0000275483	0	Putative_disease_resistance_protein_RGA4
135	MDP0000457846	0	Putative_disease_resistance_RPP13-like_protein_1
136	MDP0000675339	0	TMV_resistance_protein_N
137	MDP0000173694	0	TMV_resistance_protein_N

138	MDP0000207179	0	Putative_disease_resistance_protein_RGA3
139	MDP0000306762	0	TMV_resistance_protein_N
140	MDP0000290186	0	Putative_disease_resistance_protein_RGA3
141	MDP0000662922	0	Disease_resistance_protein_RPM1
142	MDP0000578301	0	Disease_resistance_protein_At4g27190
143	MDP0000126186	0	Putative_disease_resistance_protein_RGA3
144	MDP0000205727	0	TMV_resistance_protein_N
145	MDP0000280705	0	Disease_resistance_protein_RGA2
146	MDP0000601725	0	Putative_disease_resistance_protein_RGA4
147	MDP0000241462	0	Disease_resistance_protein_RGA2
148	MDP0000135249	0	Putative_disease_resistance_protein_RGA3
149	MDP0000142444	0	Disease_resistance_protein_RPM1
150	MDP0000273281	0	Putative_disease_resistance_protein_RGA4
151	MDP0000725802	0	Putative_disease_resistance_protein_RGA3
152	MDP0000125338	0	Probable_disease_resistance_protein_At5g63020
153	MDP0000198998	0	Putative_disease_resistance_protein_RGA3
154	MDP0000240942	0	Putative_disease_resistance_protein_RGA3
155	MDP0000310628	0	Putative_disease_resistance_protein_RGA4
156	MDP0000178216	0	Putative_disease_resistance_RPP13-like_protein_1
157	MDP0000195674	0	Putative_disease_resistance_protein_RGA3
158	MDP0000260265	0	Putative_disease_resistance_RPP13-like_protein_1
159	MDP0000279176	0	TMV_resistance_protein_N
160	MDP0000305882	0	Disease_resistance_protein_RPM1
161	MDP0000591990	0	Putative_disease_resistance_RPP13-like_protein_1
162	MDP0000294230	0	Putative_disease_resistance_RPP13-like_protein_1
163	MDP0000708478	0	Disease_resistance_protein_RGA2
164	MDP0000172886	0	Disease_resistance_protein_RGA2
165	MDP0000286805	0	TMV_resistance_protein_N

166	MDP0000693359	0	TMV_resistance_protein_N
167	MDP0000816743	0	TMV_resistance_protein_N
168	MDP0000169431	0	Putative_disease_resistance_protein_RGA3
169	MDP0000202315	0	Disease_resistance_protein_RPM1
170	MDP0000247576	0	Disease_resistance_protein_RGA2
171	MDP0000317856	0	TMV_resistance_protein_N
172	MDP0000127009	0	Disease_resistance_protein_RGA2
173	MDP0000151729	0	Putative_disease_resistance_protein_RGA4
174	MDP0000171561	0	Putative_disease_resistance_protein_At3g14460
175	MDP0000227547	0	Disease_resistance_protein_RPM1
176	MDP0000455346	0	TMV_resistance_protein_N
177	MDP0000242361	0	Disease_resistance_protein_RPM1
178	MDP0000311735	0	Disease_resistance_protein_RPM1
179	MDP0000314069	0	Putative_disease_resistance_protein_RGA3
180	MDP0000442056	0	Disease_resistance_protein_RPM1
181	MDP0000231100	0	Putative_disease_resistance_protein_RGA4
182	MDP0000288587	0	TMV_resistance_protein_N
183	MDP0000410696	0	TMV_resistance_protein_N
184	MDP0000636042	0	Putative_disease_resistance_protein_RGA1
185	MDP0000252069	0	TMV_resistance_protein_N
186	MDP0000284876	0	Disease_resistance_protein_RPM1
187	MDP0000573412	0	TMV_resistance_protein_N
188	MDP0000159004	0	Disease_resistance_protein_RPM1
189	MDP0000166360	0	TMV_resistance_protein_N
190	MDP0000206335	0	Putative_disease_resistance_protein_RGA3
191	MDP0000268912	0	Probable_disease_resistance_protein_At5g63020
192	MDP0000294503	0	Putative_disease_resistance_protein_RGA1
193	MDP0000468697	0	Putative_disease_resistance_protein_RGA3

194	MDP0000652037	0	Putative_disease_resistance_protein_RGA4
195	MDP0000718279	0	Putative_disease_resistance_RPP13-like_protein_1
196	MDP0000737010	0	TMV_resistance_protein_N
197	MDP0000569038	0	Putative_disease_resistance_protein_RGA3
198	MDP0000668041	0	TMV_resistance_protein_N
199	MDP0000901201	0	TMV_resistance_protein_N
200	MDP0000154841	0	Putative_disease_resistance_protein_RGA3
201	MDP0000158225	0	Putative_disease_resistance_RPP13-like_protein_1
202	MDP0000223442	0	Putative_disease_resistance_protein_At4g19050
203	MDP0000292810	0	Putative_disease_resistance_protein_RGA4
204	MDP0000310094	0	Probable_disease_resistance_protein_At5g66900
205	MDP0000831734	0	TMV_resistance_protein_N
206	MDP0000852369	0	Putative_disease_resistance_protein_RGA4
207	MDP0000860172	0	TMV_resistance_protein_N
208	MDP0000129379	0	Disease_resistance_protein_RGA2
209	MDP0000165280	0	Putative_disease_resistance_protein_RGA4
210	MDP0000173893	0	TMV_resistance_protein_N
211	MDP0000211709	0	TMV_resistance_protein_N
212	MDP0000243301	0	Disease_resistance_protein_RGA2
213	MDP0000249539	0	Putative_disease_resistance_protein_RGA4
214	MDP0000263006	0	Putative_disease_resistance_protein_RGA1
215	MDP0000278375	0	Putative_disease_resistance_protein_RGA1
216	MDP0000284090	0	TMV_resistance_protein_N
217	MDP0000296620	0	Disease_resistance_protein_RPM1
218	MDP0000318995	0	Putative_disease_resistance_RPP13-like_protein_1
219	MDP0000430035	0	TMV_resistance_protein_N
220	MDP0000540010	0	Putative_disease_resistance_protein_At3g14460
221	MDP0000120040	0	Putative_disease_resistance_RPP13-like_protein_1

222	MDP0000137113	0	Disease_resistance_protein_RPM1
223	MDP0000156373	0	Putative_disease_resistance_RPP13-like_protein_1
224	MDP0000162814	0	Putative_disease_resistance_RPP13-like_protein_1
225	MDP0000216428	0	Putative_disease_resistance_RPP13-like_protein_1
226	MDP0000227461	0	Disease_resistance_protein_RPM1
227	MDP0000233740	0	Putative_disease_resistance_protein_RGA3
228	MDP0000258012	0	TMV_resistance_protein_N
229	MDP0000259527	0	Putative_disease_resistance_protein_RGA3
230	MDP0000291788	0	Putative_disease_resistance_RPP13-like_protein_1
231	MDP0000293114	0	Putative_disease_resistance_protein_RGA3
232	MDP0000304705	0	Putative_disease_resistance_protein_RGA4
233	MDP0000307215	0	Putative_disease_resistance_RPP13-like_protein_1
234	MDP0000404538	0	TMV_resistance_protein_N
235	MDP0000560245	0	TMV_resistance_protein_N
236	MDP0000602346	0	TMV_resistance_protein_N
237	MDP0000675405	0	TMV_resistance_protein_N
238	MDP0000705431	0	Putative_disease_resistance_protein_RGA4
239	MDP0000778327	0	Putative_disease_resistance_protein_RGA4
240	MDP0000861936	0	Putative_disease_resistance_protein_RGA3
241	MDP0000252903	4.00E-10	Putative_disease_resistance_protein_RGA3
242	MDP0000204916	3.20E-09	Putative_disease_resistance_RPP13-like_protein_1
243	MDP0000551160	5.00E-09	TMV_resistance_protein_N
244	MDP0000703006	5.50E-09	Disease_resistance_protein_RGA2
245	MDP0000303507	2.27E-08	Putative_disease_resistance_protein_At3g14460
246	MDP0000318389	2.82E-08	Disease_resistance_protein_RPM1
247	MDP0000292029	4.68E-08	TMV_resistance_protein_N
248	MDP0000749757	6.41E-08	TMV_resistance_protein_N
249	MDP0000200491	1.57E-07	Putative_disease_resistance_protein_RGA4

250	MDP0000277725	2.59E-07	Putative_disease_resistance_RPP13-like_protein_1
251	MDP0000436696	2.88E-07	Putative_disease_resistance_RPP13-like_protein_1
252	MDP0000531361	5.84E-07	TMV_resistance_protein_N
253	MDP0000140010	6.34E-07	Putative_disease_resistance_protein_RGA3
254	MDP0000216448	8.09E-07	Putative_disease_resistance_protein_RGA3
255	MDP0000272435	1.32E-06	Disease_resistance_protein_RGA2
256	MDP0000197558	1.33E-06	Disease_resistance_protein_RPM1
257	MDP0000129374	3.73E-06	Putative_disease_resistance_protein_RGA3
258	MDP0000532920	4.22E-06	TMV_resistance_protein_N
259	MDP0000257201	4.26E-06	TMV_resistance_protein_N
260	MDP0000284789	4.46E-06	Putative_disease_resistance_protein_RGA3
261	MDP0000715979	5.98E-06	Disease_resistance_protein_RGA2
262	MDP0000259862	8.71E-06	Putative_disease_resistance_protein_RGA3
263	MDP0000296048	1.03E-05	Putative_disease_resistance_protein_RGA1
264	MDP0000142781	1.08E-05	Putative_disease_resistance_protein_RGA3
265	MDP0000252913	1.15E-05	Disease_resistance_protein_RPM1
266	MDP0000289959	1.42E-05	Disease_resistance_protein_RPM1
267	MDP0000210537	1.46E-05	TMV_resistance_protein_N
268	MDP0000196621	1.57E-05	Putative_disease_resistance_protein_RGA1
269	MDP0000140978	1.88E-05	Putative_disease_resistance_protein_RGA4
270	MDP0000676773	2.20E-05	Putative_disease_resistance_protein_RGA4
271	MDP0000480725	2.24E-05	Putative_disease_resistance_protein_At1g50180
272	MDP0000493661	2.58E-05	Probable_disease_resistance_RPP8-like_protein_2
273	MDP0000140578	2.83E-05	TMV_resistance_protein_N
274	MDP0000478441	3.87E-05	TMV_resistance_protein_N
275	MDP0000136651	4.02E-05	TMV_resistance_protein_N
276	MDP0000151475	4.37E-05	Putative_disease_resistance_protein_At3g14460
277	MDP0000040083	5.23E-05	TMV_resistance_protein_N

278	MDP0000142273	7.30E-05	Putative_disease_resistance_protein_RGA4
279	MDP0000681569	8.35E-05	TMV_resistance_protein_N
280	MDP0000274330	0.000106	Disease_resistance_protein_RGA2
281	MDP0000304378	0.00011	Putative_disease_resistance_protein_RGA4
282	MDP0000486133	0.000125	TMV_resistance_protein_N
283	MDP0000667502	0.000133	Putative_disease_resistance_protein_RGA1
284	MDP0000403796	0.000178	TMV_resistance_protein_N
285	MDP0000437040	0.000183	TMV_resistance_protein_N
286	MDP0000308485	0.000229	Putative_disease_resistance_protein_RGA3
287	MDP0000149577	0.000396	TMV_resistance_protein_N
288	MDP0000827008	0.000397	TMV_resistance_protein_N
289	MDP0000751862	0.0004	TMV_resistance_protein_N
290	MDP0000130346	0.000411	Putative_disease_resistance_RPP13-like_protein_1
291	MDP0000187390	0.000411	Putative_disease_resistance_protein_RGA1
292	MDP0000516713	0.000417	Disease_resistance_protein_RPM1
293	MDP0000189848	0.00044	Putative_disease_resistance_RPP13-like_protein_1
294	MDP0000192700	0.000489	Putative_disease_resistance_protein_RGA3
295	MDP0000719767	0.000782	TMV_resistance_protein_N

Table 5-4 *PHAS* loci targeted by mdm-miR2118

miRNA	*PHAS* loci	Align-score	Target-site position	Strand	Target-site sequence	p-value
mdm-miR2118a/b/c	MDC013847.449:8650..11009	2	8703-8724	1	UGGGAUUUGGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC017425.176:8127..9509	2	8165-8186	1	UGGGAUUUGGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC007882.217:1853..3206	2	1861-1882	1	UGGGAUUUGGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC021042.366:7100..10015	2	9958-9979	-1	UGGGAUUUGGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC005296.140:6219..6918	2	6715-6736	-1	UGGGAGUUAGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC007579.331:5..901	3	841-862	-1	UAGGAUUUGGUGGCAUCGGUAA	0
mdm-miR2118a/b/c	MDC010300.238:13303..13649	2	13367-13388	1	UGGGAUUUGGUGGCAUCGGUAA	1.77E-06
mdm-miR2118a/b/c	MDC019637.224:17388..17815	2.5	17770-17791	-1	UGGGAGUUGGUGGCAUCGGUAA	3.22E-05

Appendix II Figures

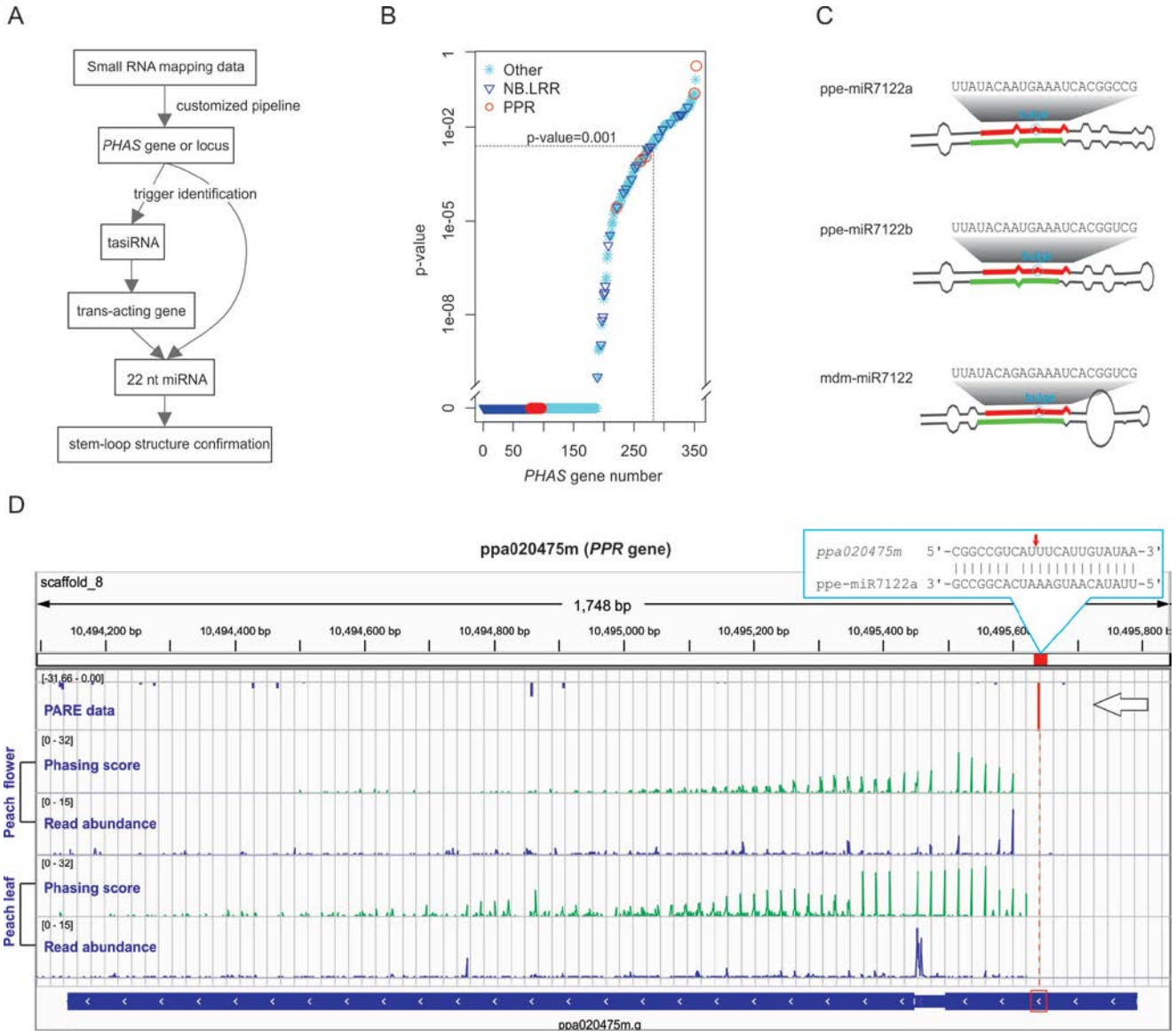


Figure 4-1 Identification of *PHAS PPR* genes and the trigger miRNA in peach.

(A) Workflow of identifying *PHAS* genes/loci and miRNA trigger (reverse computation).

(B) p-value distribution of all putative *PHAS* genes in peach. The p-value threshold of 0.001 is indicated with a dotted line.

(C) Stem-loop structures of the miR7122 homologues in peach (ppe-miR7122a and b) and apple (mdm-miR7122). Positions of miRNA and miRNA* sequences are marked in red and green, respectively; position of bulge in each stem-loop structure is marked with a light-blue circle.

(D) Phasing score and read abundance distribution along a *PHAS PPR* gene in peach flower and leaf tissues viewed together with PARE data in the Integrative Genomics Viewer (IGV). Phasing score was calculated based on the mapping results of 21-nt sRNAs (see METHODS). Paring of miR7122 and target site is denoted in the upper-right corner with the cleavage site marked with a small red arrow; grey gridlines show the 21-nt phasing pattern set up by the miR7122 cleavage site. The gene transcription direction and miR7122 cleavage site confirmed by PARA data are denoted by a big arrow and a red bar, respectively.

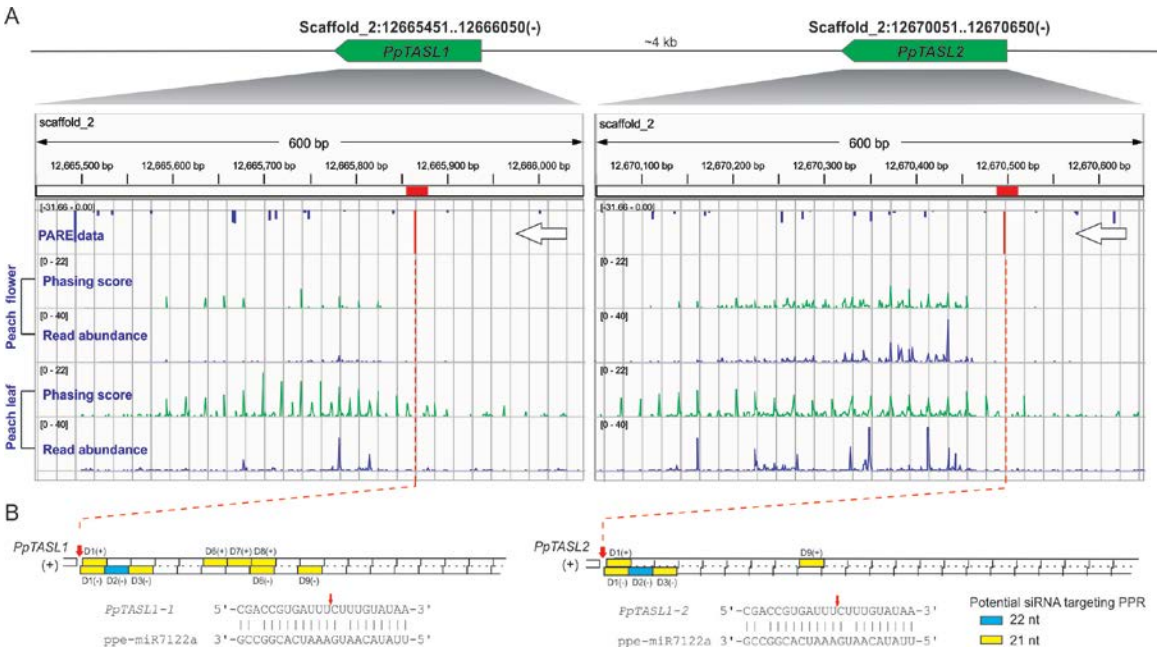


Figure 4-2 TasiRNAs derived from miR7122-targeted *PpTAS1/2* reinforce the silencing effect of *PPR* genes in peach.

(A) Genomic configuration of *PpTASL1/2*, and their phasing score and read abundance distribution in peach flower and fruit tissues viewed together with PARE data in IGV. Phasing score was calculated based on the mapping results of 21-nt sRNAs (see METHODS). Grey gridlines show the 21-nt phasing pattern set up by the miR7122 cleavage site. The gene transcription direction and miR7122 cleavage site confirmed by PARA data are denoted by big arrows and red bars, respectively.

(B) Diagrams illustrating the pattern and position of secondary siRNAs (green squares for 22 nt and yellow for 21nt) potentially targeting *PPR* genes. The miR7122 cleavage site (red bar) confirmed by PARE data set the phase for the phasiRNA production in both *PpTASL1* and *PpTASL2*. Generated siRNAs are numbered in order (D1, 2, 3, etc.) with strand information indicated in parentheses ("+" for plus strand and "-" for minus strand). MiR7122-mRNA parings are denoted below with the cleavage site marked with a small red arrow.

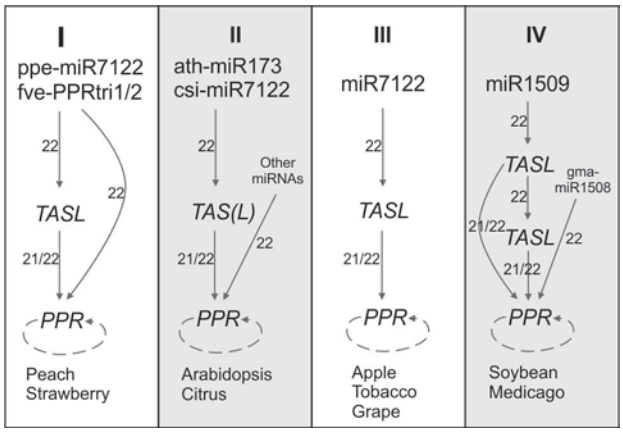


Figure 4-3 Diverse miR7122-*TASL*-*PPR*-siRNA pathways identified.

Four categories of miR7122-*TAS(L)*-*PPR*-siRNA pathways identified in nine plants are summarized. The dotted arc around the "*PPR*" represents the ability of processing secondary siRNAs. The length of trigger sRNA is indicated near the arrows. Category I: in peach and strawberry, *PPR* phasiRNA production was triggered by ppe-miR7122 and fve-PPRtri1/2, respectively, through direct targeting or the intermediation of *TASL* genes; Category II: in *Arabidopsis* and citrus, *PPR* phasiRNA generation was triggered either by miR173 and csi-miR7122, respectively, through the intermediation of *TAS(L)* genes, or by directly targeting of other miRNAs; Category III: in apple, tobacco, and grape, *PPR* siRNA production was triggered by miR7122 solely through the intermediation of *TASL* genes; Category IV, in soybean and Medicago, *PPR* phasiRNA generation was triggered by miR1509 through one or two layers of *TASL*-tasiRNA interactions or by direct targeting of gma-miR1508 in soybean as well.

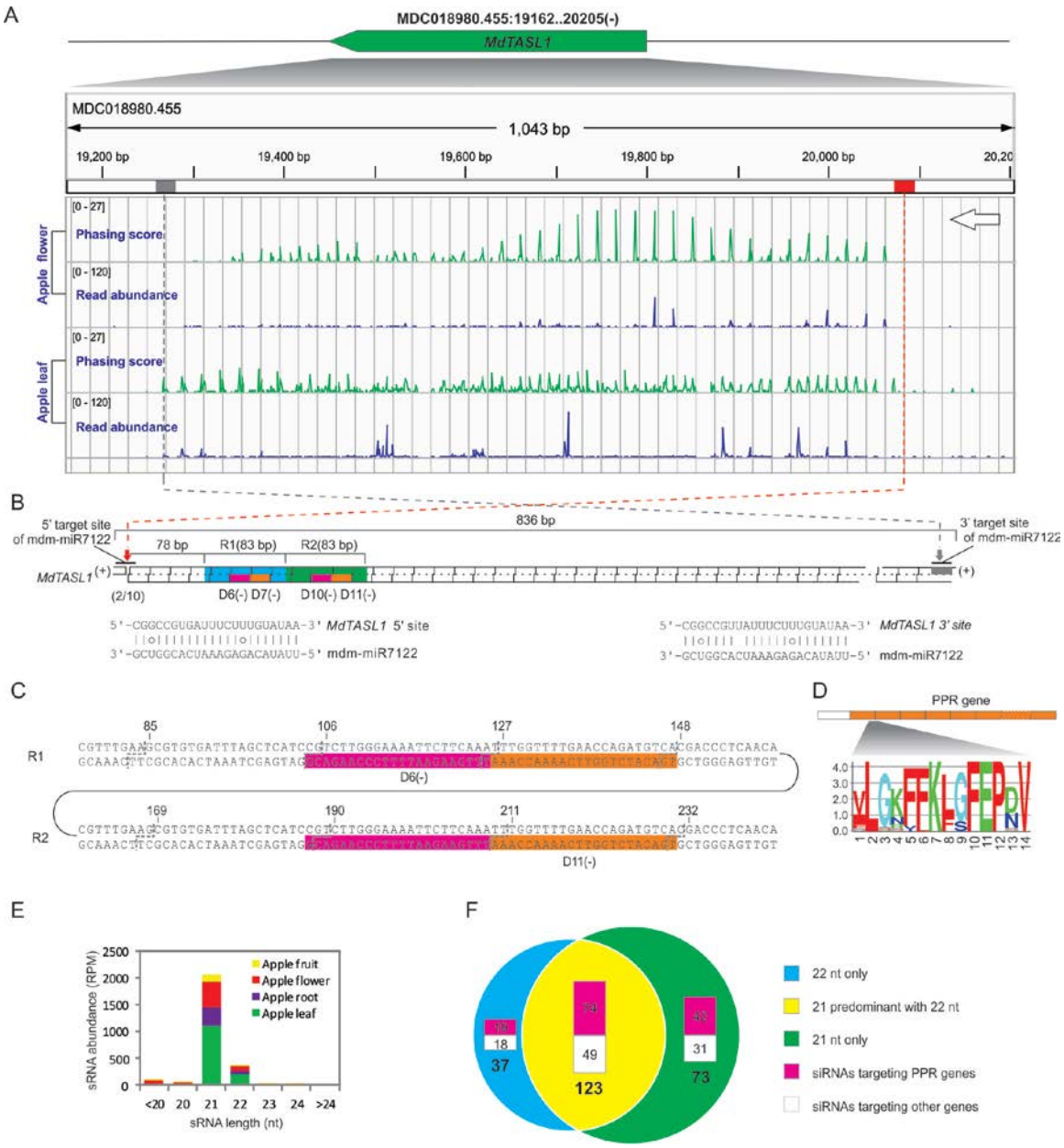


Figure 4-4 miR7122-*TASL-PPR*-siRNA pathway identified in apple with distinct features.

(A) Genomic configuration of *MdTASL1* and its phasing score and read abundance distribution in apple flower and fruit tissues viewed in IGV. Phasing score was calculated based on the mapping results of 21-nt siRNAs (see METHODS). Position of two mdm-miR7122 target sites are denoted in red (5' site) and grey (3' site) boxes; grey gridlines show the 21-nt phasing pattern set up by the 5' cleavage site of mdm-miR7122.

(B) A diagram illustrating the pattern of secondary siRNA biogenesis initiated by the 5' target site. Parings of mdm-miR7122with target sites are denoted below with the cleavage site marked by small red (5' site) and grey (3' site) arrows, respectively; two repetitive regions (R1 and R2) and tasiRNAs that trigger *PPR* phasiRNA production are individually labeled and indicated in boxes of different colors. Result of 5'-RLM-RACE is included in parentheses right below the 5' target site; the numbers separated by the backslash indicate the number of sequenced DNAs with 5'-end corresponding to the mdm-miR7122 cleavage site (10^{th} position) and the number of total sequenced DNAs.
(C) Double-stranded sequences of the two identical 83-bp regions (R1 and R2) with sites in phase with the 5' cleavage site indicated with grey dotted lines and the two identical tasiRNAs that trigger *PPR* phasiRNA production marked in pink or orange.
(D) Position and conservation logo of the amino acid sequence encoded by the sequence targeted by the two trigger tasiRNAs. PPR repeat is denoted by orange boxes.
(E) Length distribution of mdm-miR7122-triggered, *MdTASL1*-derived phasiRNA population.
(F) Categories of *MdTASL1*-derived phasiRNAs and their function. SiRNAs are classified according to their size and their ability to target *PPRs*.

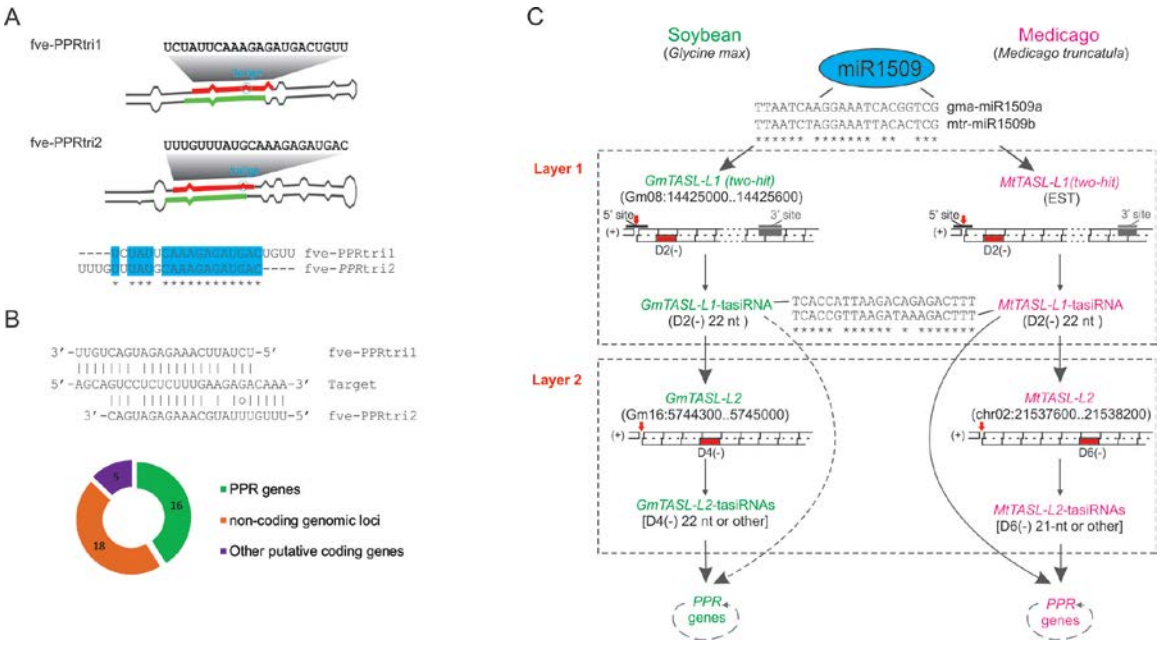


Figure 4-5 miRNA-*TASL*-*PPR*-siRNA pathway identified in strawberry and legume plants.

(A) Stem-loop structures of strawberry miRNA triggers (fve-PPRtri1 and fve-PPRtri2) of *PPR* phasiRNA production. miRNA and miRNA* are marked in red and green, respectively, and bulges in stem-loop structures are indicated with a light-blue circle. Sequence comparison of fve-PPRtri1/2 is shown below.

(B) Illustration of sequence co-targeted by fve-PPRtri1 and fve-PPRtri2. Numbers of genes potentially targeted by fve-PPRtri1/2 is illustrated in a doughnut chart below.

(C) A diagram illustrating the miR1509-*TASL*-*PPR*-siRNA pathway with one or two layers of *TASL*-tasiRNA interaction in soybean and Medicago. *TASL* genes and pattern of their phasiRNA biogenesis are individually denoted. The 22 nt tasiRNAs that are capable of targeting or triggering phasiRNA production in *TASL* transcripts or *PPRs* are marked in red. The potential direct cleavage of *PPR* transcripts by the 22 nt tasiRNA, *GmTASL-L1*-D2(-), is denoted with dotted line. The dotted arc around the "*PPR* genes" represents the ability of processing secondary siRNAs.

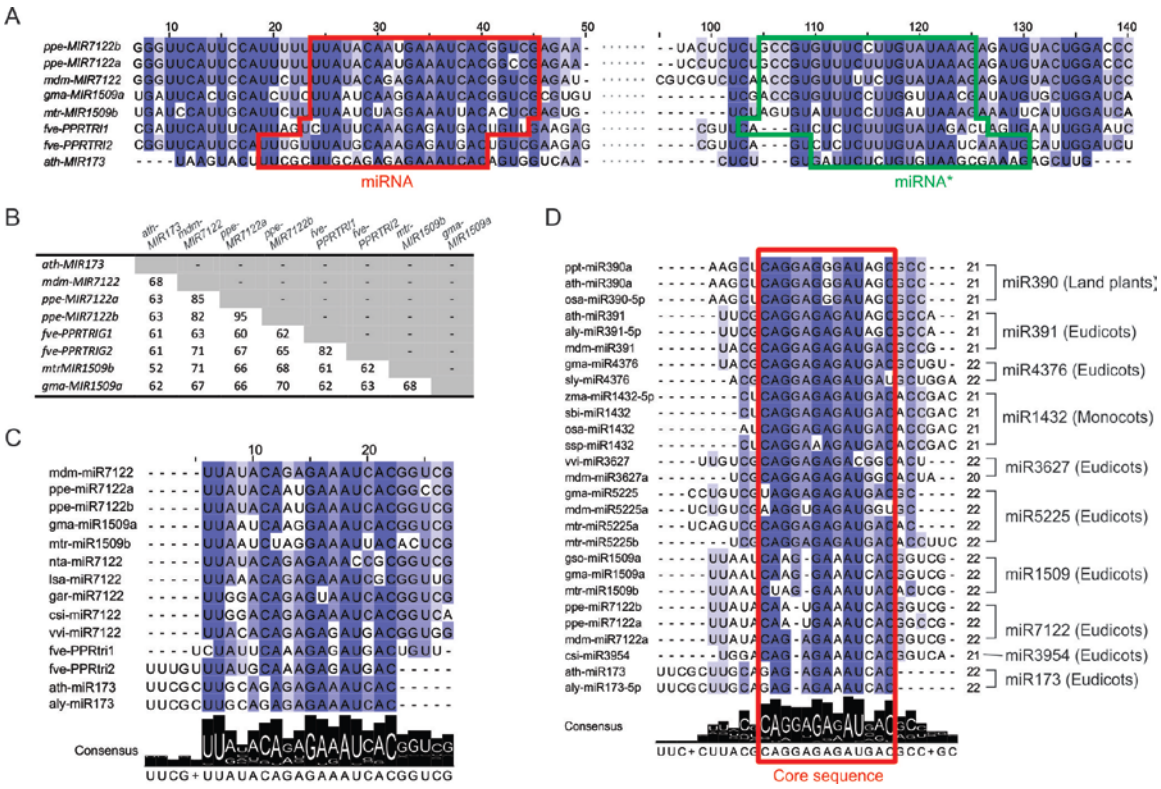


Figure 4-6 miRNA triggers share a common origin and show sequence similarity to many other miRNAs.

(A) Conservation profile of foldback sequences of eight *MIRNA* genes with miRNA and miRNA* marked in red and green boxes, respectively. The sequence of the 61~94 bp region is denoted as dots and full view of the alignment can be found in Supplemental figure 4A online. Conservation degree of each nucleotide along *MIRNA* transcripts, as calculated by frequency, is correspondingly represented by colors, with dark color standing for a high level of conservation and light color for a low level.

(B) Pairwise alignment scores (percentage) for each *MIRNA* gene.

(C) Conservation profile of all the miR7122 homologues identified. Consensus sequence and conservation logo are included below.

(D) Conservation profile of miR7122-related miRNAs. All the miRNA sequences were retrieved from miRbase (version 19). Position of the core sequences is marked with a red box.

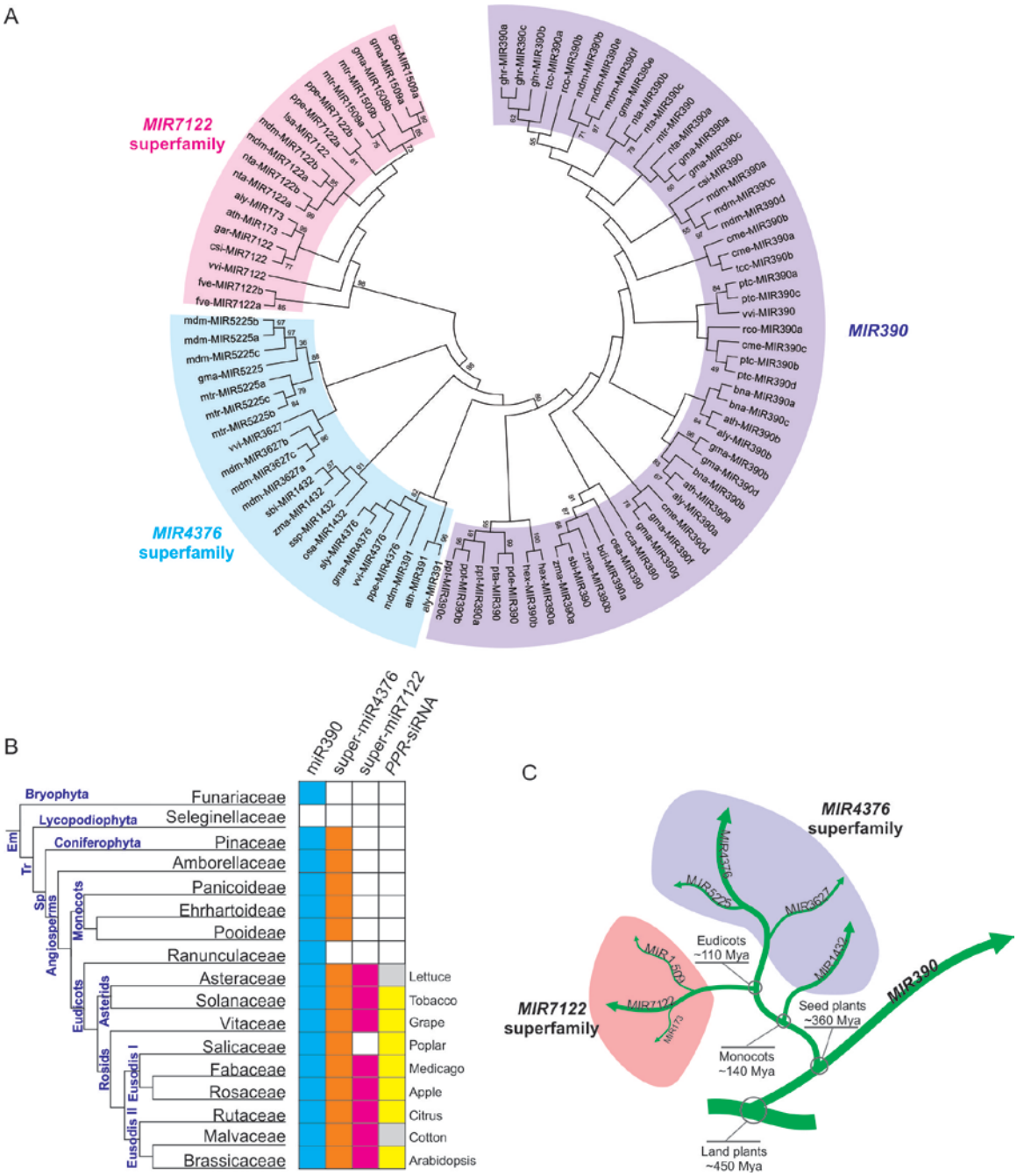


Figure 4-7 *MIR7122*s are potentially evolved from *MIR390*s via *MIR4376*s.

(A) Phylogenetic analysis of *MIR7122*-related *MIRNA* genes. Excerpted nucleotide sequences encompassing the miRNA, miRNA* and five nucleotide flanking sequences of them were aligned using CluastalW2 (see alignment in Supplemental file 1 online). The bootstrap consensus tree was inferred using the Maximum Likelihood method and 1,000 bootstraps with bootstrap values (above 50) indicated near the nodes (see detailed steps in METHODS). Each major group is denoted with distinct colors.

(B) Conservation extent of miR7122-related miRNAs and *PPR*-siRNA pathway in plants. Boxes are highlighted if a miRNA or pathway was identified in at least one species for each of the plant families listed. Plant taxonomy was designated according to the Angiosperm Phylogeny Group classification III (APGII), as described by Cuperus et al (2010). Em: Embryophyta; Tr: tracheophyta; Sp: spermatophyta. Grey squares indicate that the existence of *PPR*-siRNA pathway cannot be determined due to the absence of sequenced genome.

(C) A diagram illustrating the possible evolutionary history of *MIR7122*-related *MIRNA* genes. Branch width represents the antiquity of the corresponding *MIRNA* gene family. MYA: millions years ago.

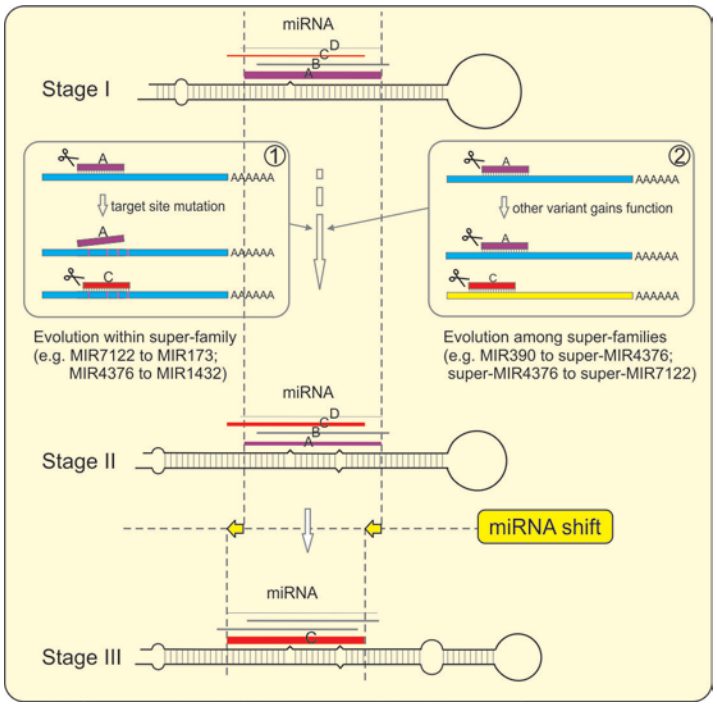


Figure 4-8 Model proposed for miRNA evolution via duplication and neofunctionalization.

For a given *MIRNA* gene, many miRNA variants are typically processed apart from the most abundant one, which we referred as mature miRNA (Stage I). This mature miRNA "A" is the functional one able to cleave the target gene, and the variants are normally of low-level abundance and nonfunctional (Stage I). Occasionally, for miRNAs from the same superfamily (①), like super-miR7122 and super-miR4376, mutations occurred in the target site make the original mature miRNA "A" unable to cleave its target gene, but one of the variant "C" happens to complementarily pair with the target gene and cleave it. Or for miRNAs from different superfamilies (②), the variant "C" gains function to target another gene different from the authentic target gene of the mature miRNA "A". Conceivably, the functional variant switch (from "A" to "C" in ①) or the newly-gained function (②), conveying adaptive advantages, would give rise to the preferential processing of the variant "C", and the "A" becomes less abundant (Stage II). After a long time, the variant "C" is evolved as the predominant mature miRNA, the previous "A" is almost lost (Stage III). During the long-time evolutionary process, more frequent mutations happens in regions of *MIRNA* genes other than the miRNA/miRNA* duplex because of less selection pressure imposed on those regions. Point mutations in

the foldback sequences, occurring as well in the miRNA/miRNA* region but at a much lower frequency, cause the change of the stem-loop structure and subsequent splicing shift, leading to the miRNA sequence shift.

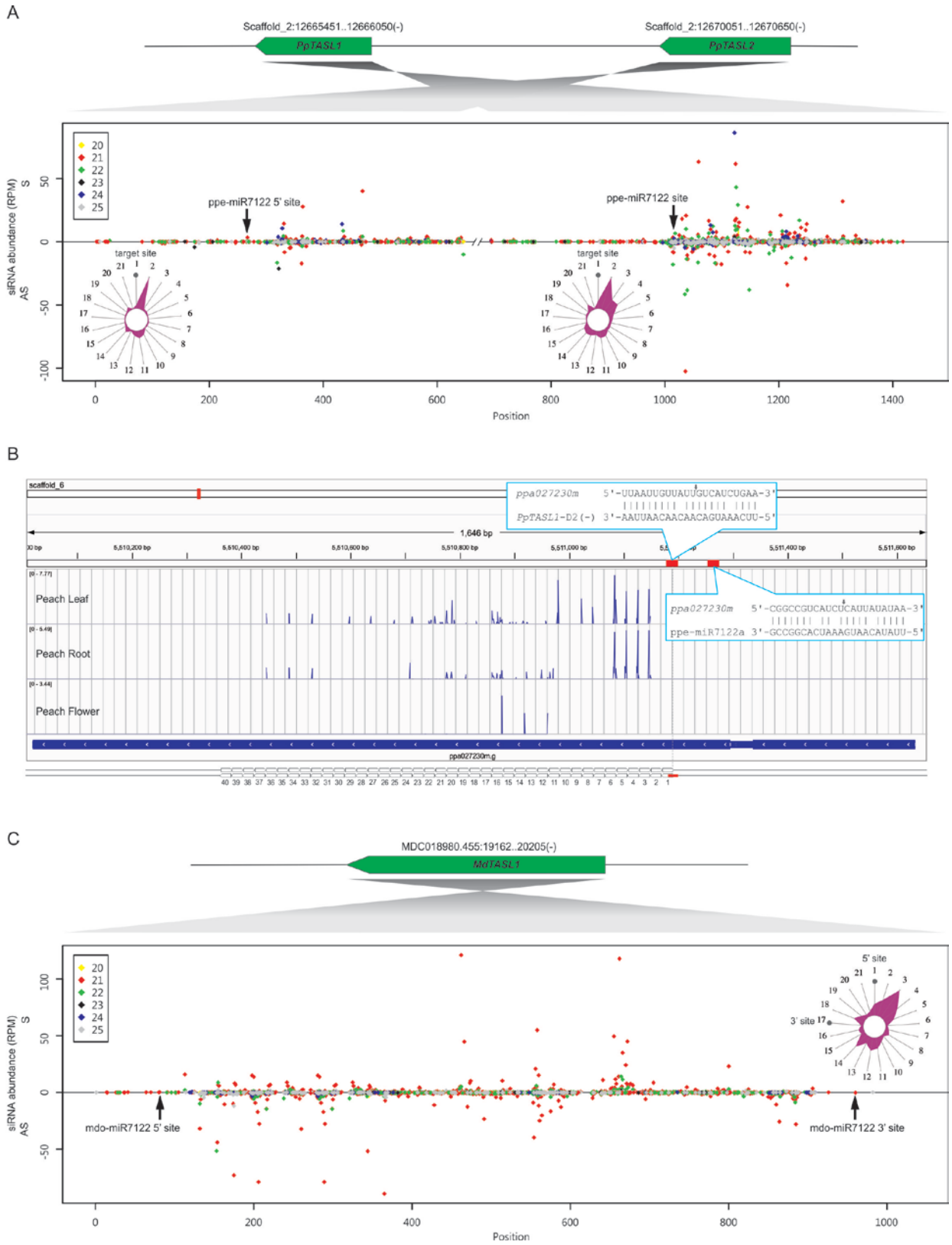


Figure 4-S1 siRNA distribution along *PpTASL1/2*, a peach *PPR* gene and *MdTASL1*

(A/C) Dot plots and radar charts illustrating the siRNA production of *PpTASL1/2* (A) and *MdTASL1* (C). Target sites of miR7122 are indicated by small black arrows. Radar chart displays the abundance of reads corresponding to each of the 21 possible phasing registers, with the 5' end of the miRNA trigger guided cleavage of the target site defined as register 1. The total number of small RNAs mapping to that register is plotted as relative distance from the center. The transcript direction is denoted by big arrows on the upper-right corner.

(B) Phasing score distribution along a peach *PHAS PPR* gene (ppa027230m). Parings of miRNA or tasiRNA and their target sites are denoted; grey gridlines show the 21-nt phasing pattern set up by the cleavage site of the *PpTASL1*-D2(-). A diagram showing the siRNA production pattern after the potential cleavage guided by the *PpTASL1*-D2(-) is included below.

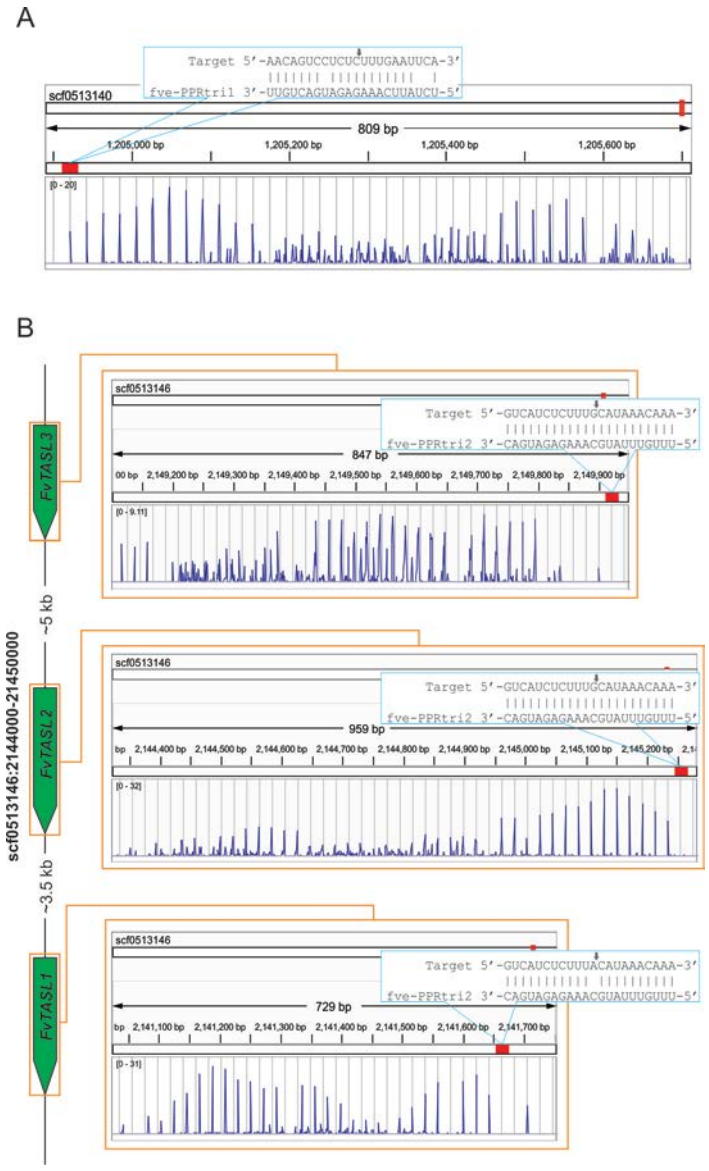


Figure 4-S2 MIR7122-(*TAS*)-*PPR*-siRNA pathway is conserved in strawberry

Phasing score distribution along a *PHAS PPR* gene (A) and three *TAS*-like loci (B) in strawberry. Pairings of fve-PPRtri1 or 2 and target sites are indicated correspondingly; genomic configuration of three *TAS*-like loci is denoted on the right of panel B; grey gridlines show the 21-nt phasing pattern set up by the cleavage site of fve-PPRtri1 or 2. The transcript direction is denoted by big arrows on the left or right side.

Figure 4-S3 Two layers of trans-acting interaction involved in the miR1509-*TASL*-*PPR*-siRNA pathway in soybean and medicago.

Phasing score distribution along *TASL-L1*s, *TASL-L2*s and *PPR*s are viewed together with PARE data in IGV. Parings of the trigger (miRNA or tasiRNA) and the target site are denoted for each gene with cleavage site marked with a small red arrow; grey gridline show the 21-nt phasing pattern set up by the cleavage site of the corresponding trigger. Uncleaved target site of *TASL-L1*s are indicated in grey boxes. Diagrams showing the

siRNA production pattern after the cleavage guided by trigger miRNAs or tasiRNAs are included below; tasiRNAs serving as triggers of next-layer TASL or PPR genes are highlighted in yellow. Red bars in the PARE data track indicates the values corresponding to the cleavage site of trigger miRNAs or tasiRNAs. The transcript direction is denoted by big arrows on the left or right side.

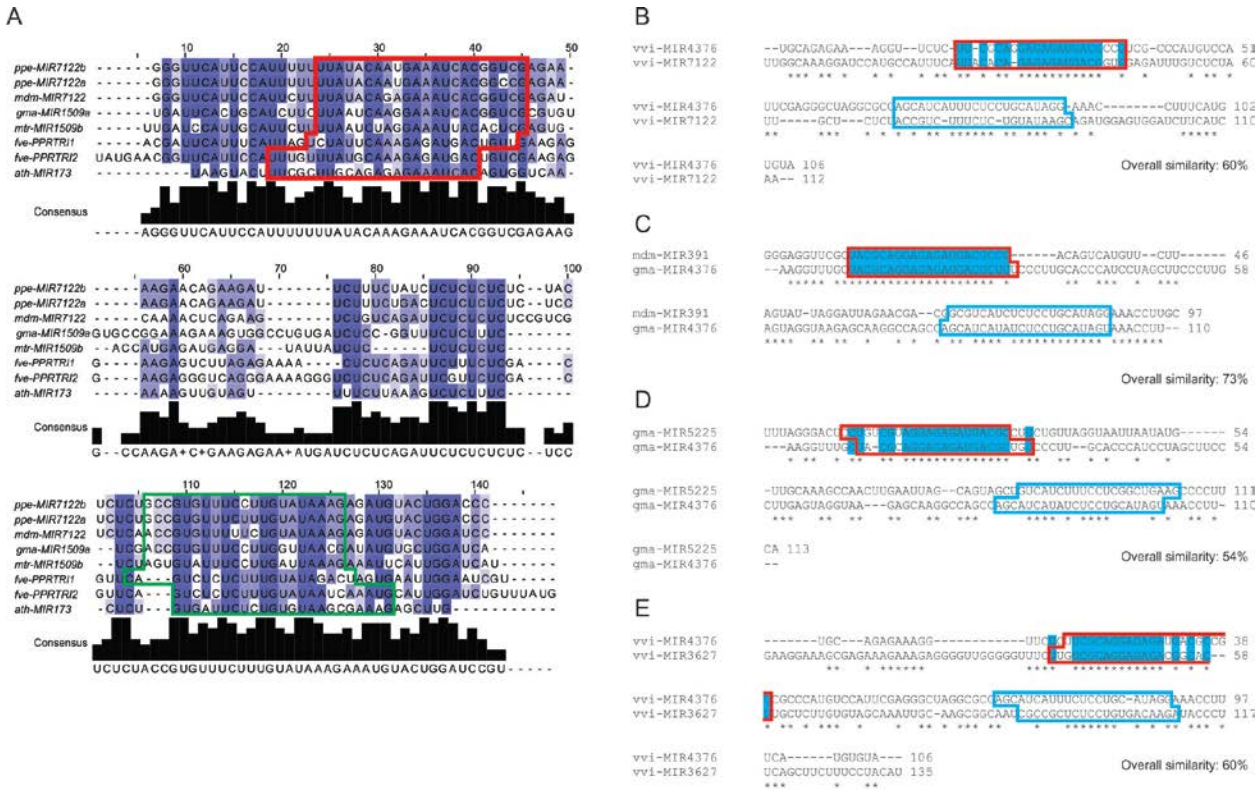


Figure 4-S4 Alignment of *MIRNA* genes.

(A) Full view of multiple alignment of foldback sequences of eight *MIRNA* genes, with miRNA and miRNA* marked in red and green boxes, respectively.

(B-E) Pairwise alignment of foldback sequences of *MIRNA* genes. Sequences of miRNA and miRNA* are highlighted in red or light-blue boxes, respectively; identical nucleotides within the miRNA aligned region are marked in light-blue. Overall similarity for each pair is included below each panel.

Figure 4-S5. Stem-loop structure of miRC1 in citrus.

Stem-loop structure was produced by RNAfold. MiRNA and miRNA* are highlighted with red and green lines, respectively, with their corresponding sequences denoted beside.

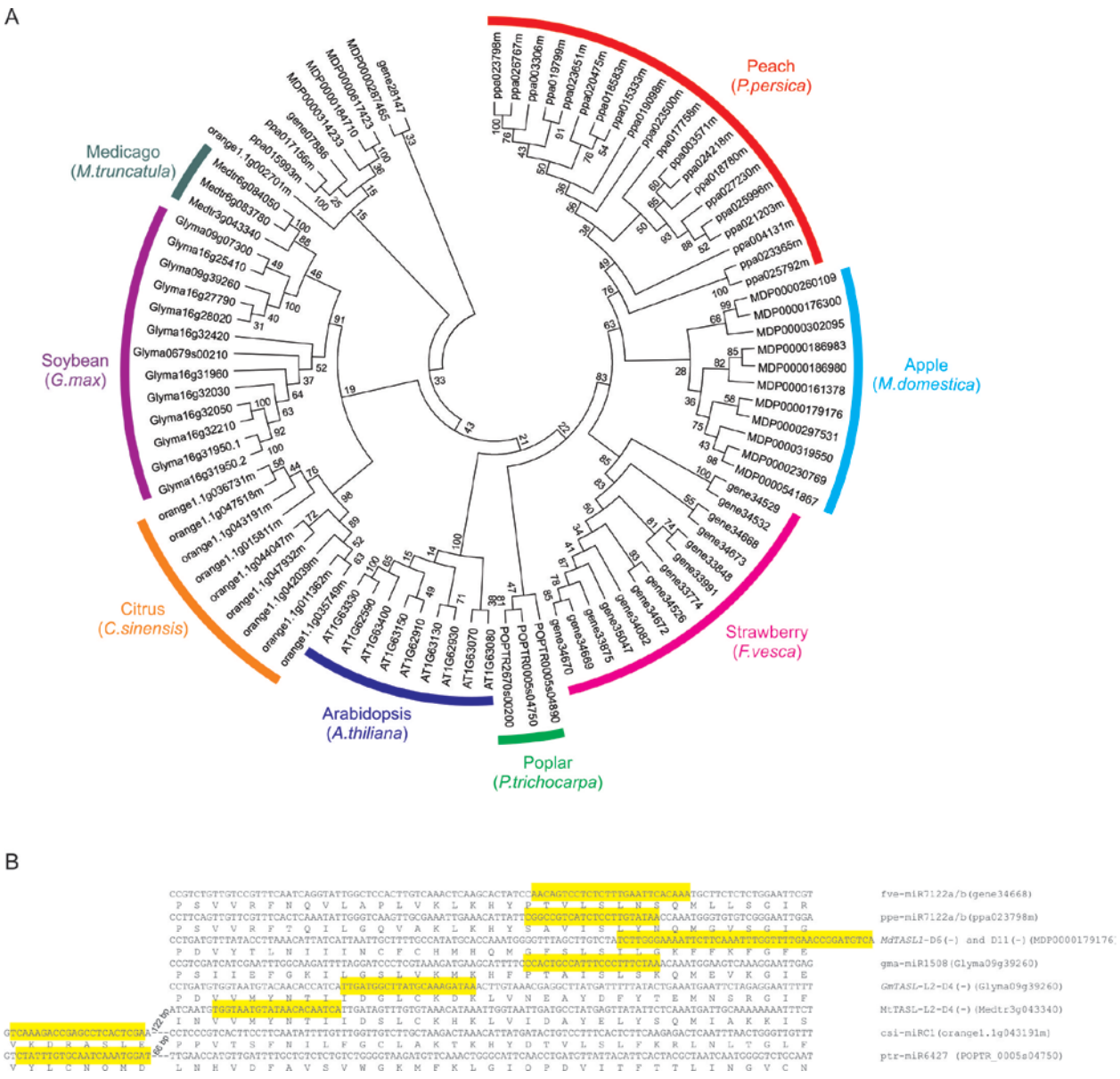


Figure 4-S6 Phylogenetic analysis of *PHAS PPR* genes and distribution of target sites of miRNAs or tasiRNAs along *PPR* repeats.

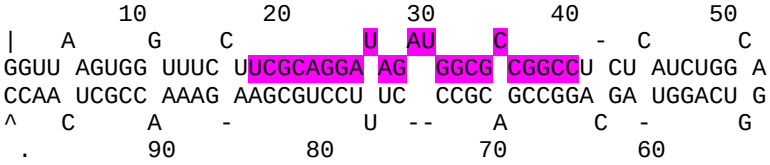
(A) Phylogenetic analysis of *PHAS PPR* genes identified in various plants. Multiple alignment of *PPR* genes was performed with Clustalw2 with similar settings with alignment results shown in Aupplemental file 2. Phylogenetic tree analysis for *PPR* genes was conducted using the neighbor-joining method by MEGA5. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap

replicates are collapsed. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 91 amino acid sequences. All positions with less than 50% site coverage were eliminated. There were a total of 560 positions in the final dataset. **(B)** Distribution of target sites of miRNAs or tasiRNAs along *PPR* repeats. Target sites of miRNAs or tasiRNAs are marked in yellow.

Picea abies

Pab-miR391 (23 bp)

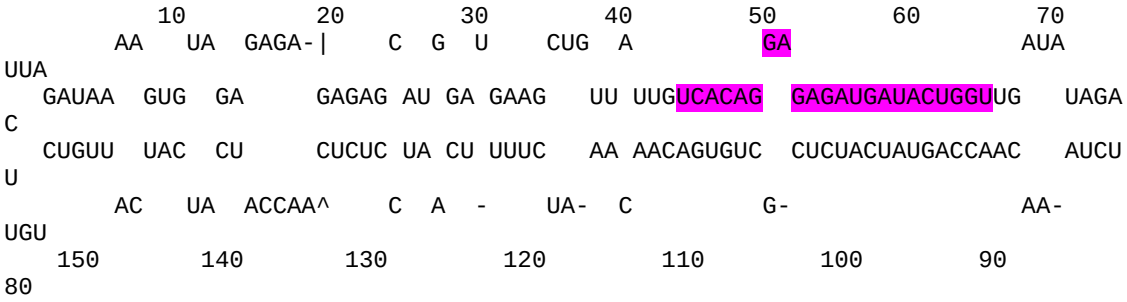
Initial ΔG = -58.10



Amborella trichophoda

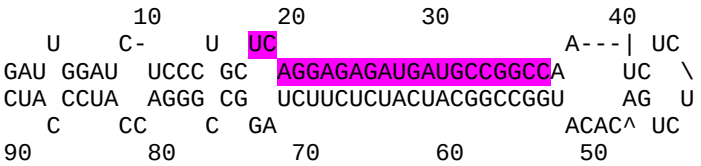
Atr-miR4376a (22 bp)

Initial ΔG = -57.40



Atr-miR4376b (21 bp)

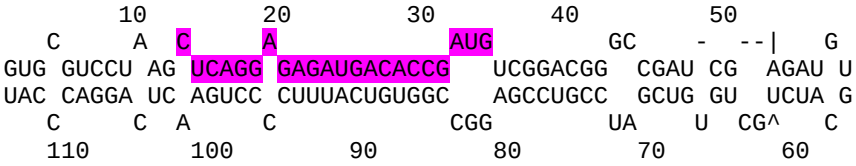
Initial ΔG = -55.90



Panicum virgatum

Pvi-miR1432 (22 bp)

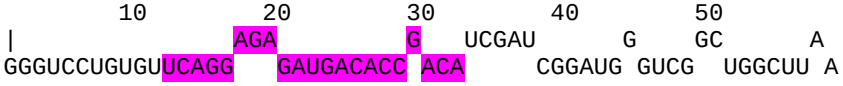
Initial ΔG = -63.50

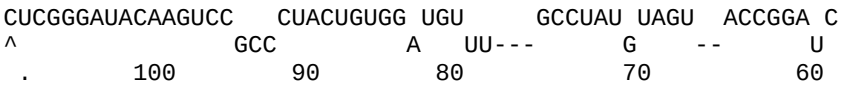


Hordeum vulgare

hvu-miR1432 (21 bp)

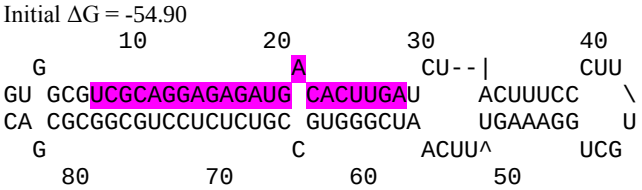
Initial ΔG = -66.00





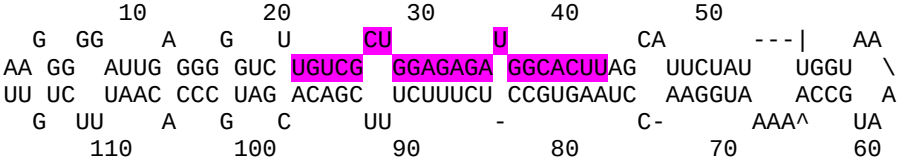
Mimulus guttatus

Mgu-miR5225 (22 bp)



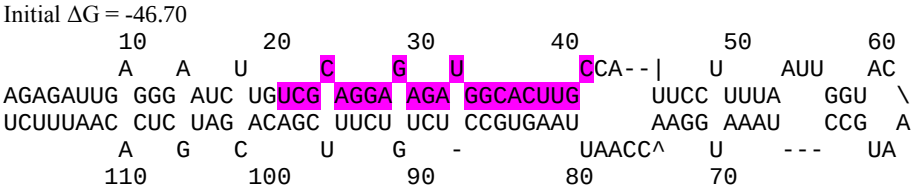
Nicotiana tabacum

Nta-miR3627 (22 bp)



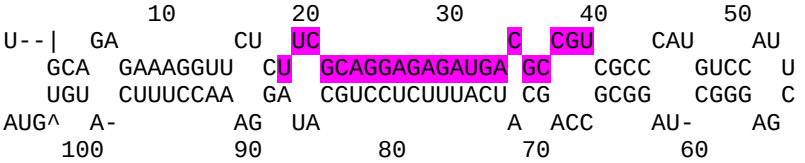
Solanum lycopersicum

Sly-miR3627

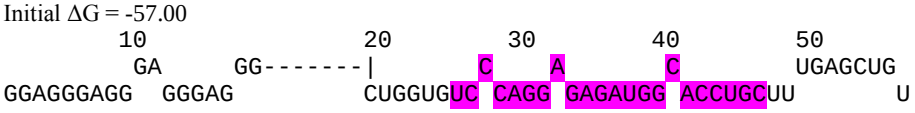


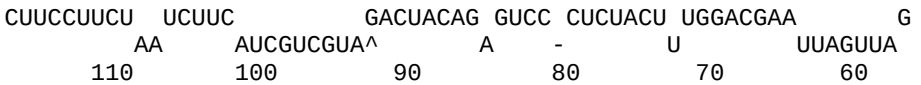
Vitis vinifera

Vvi-miR4376 (22 bp)



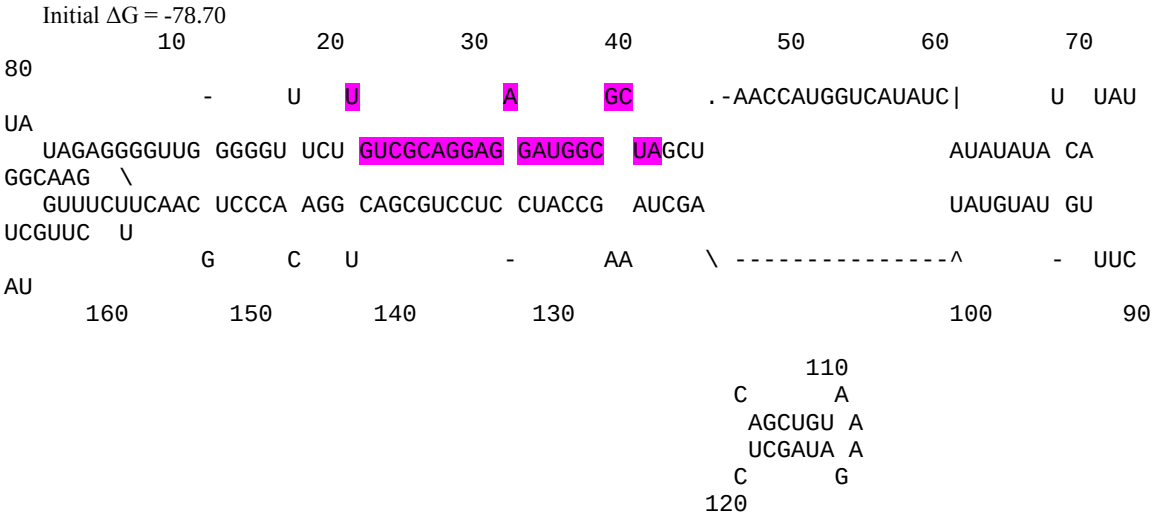
Vvi-miR5225 (22 bp)





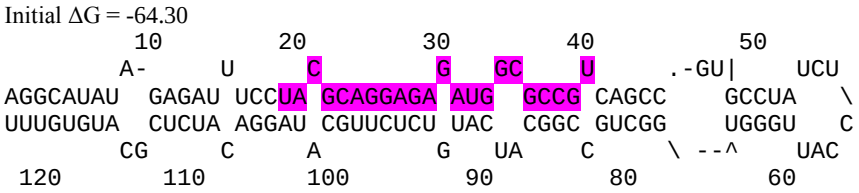
Ppulus trichocarpa

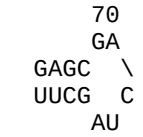
Ptr-miR3627 (22 bp)



Carica papaya

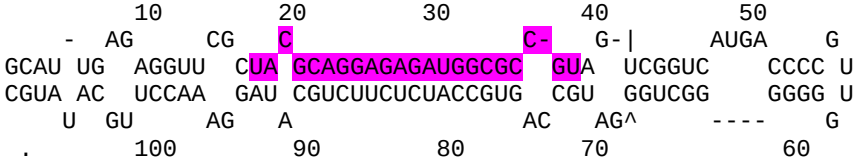
Cpa-miR4376 (22 bp)



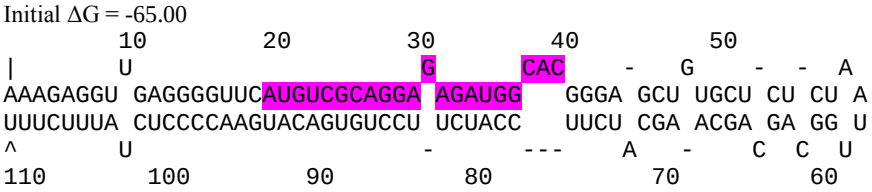


Prunus persica

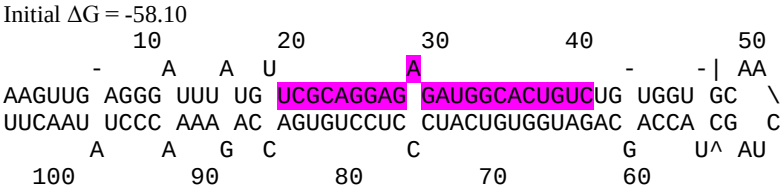
Ppe-miR4376 (22 bp)



Ppe-miR3627a (21 bp)

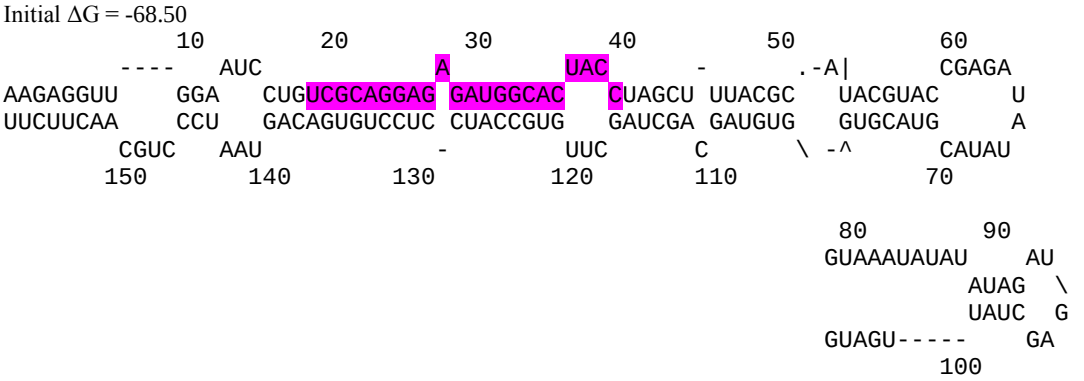


Ppe-miR3627b (*Prunus persica*, 22 bp)

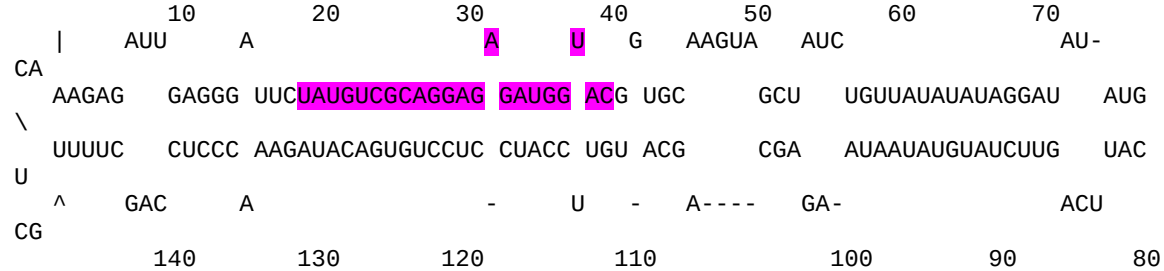


Fragaria vesca

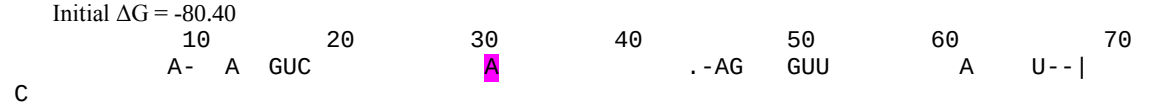
Fve-miR3627a (22bp)

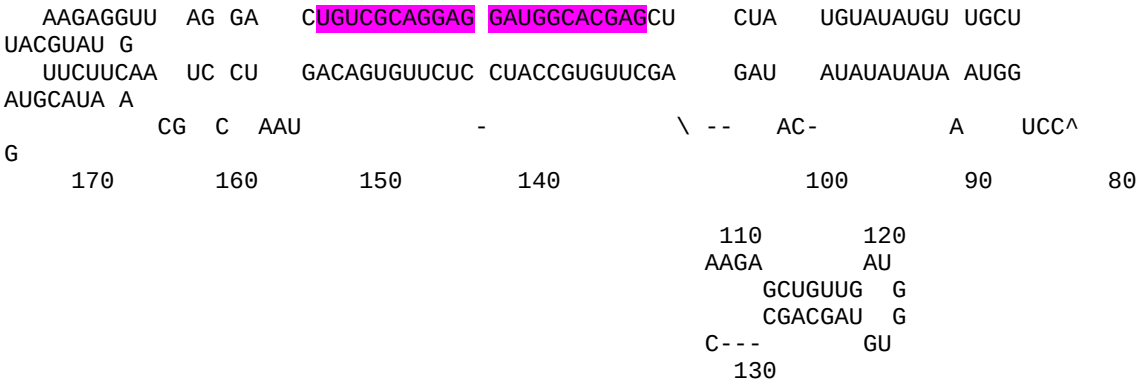


Fve-miR3627b (22bp)

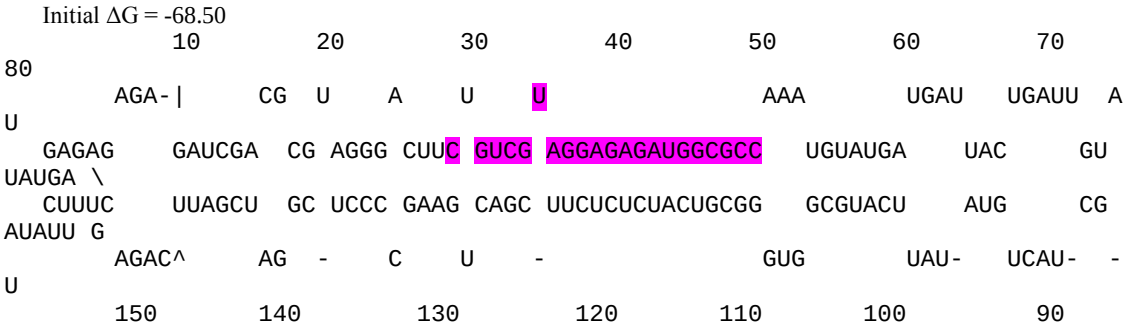


Fve-miR3627c (22bp)



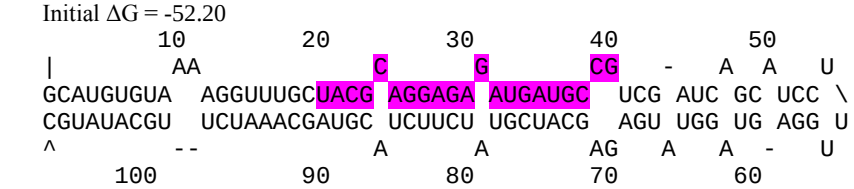


Fve-miR5225 (22bp)



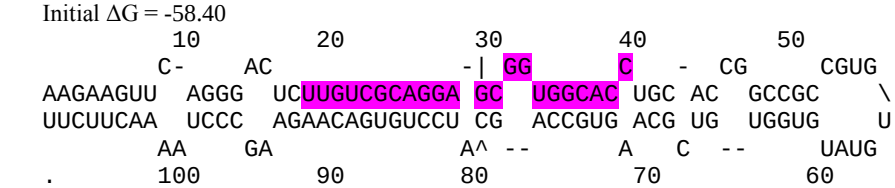
Medicago truncatula

Mtr-miR391 (21 bp)

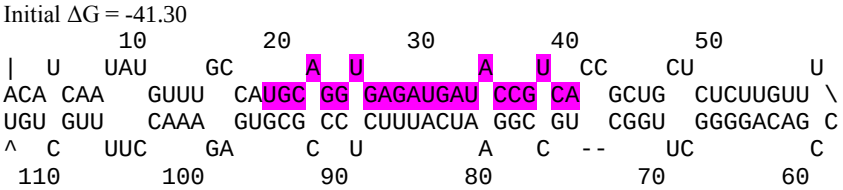


Citrus sinensis

Csi-miR3627 (22 bp)



Csi-miR5225 (22 bp)



Brassica napa

Bna-miR391 (21 bp)

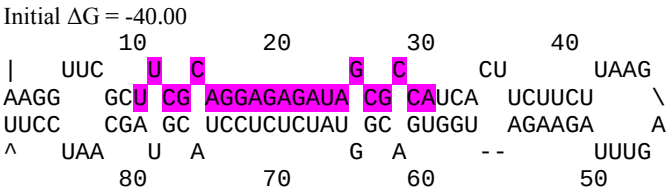


Figure 4-S7 Stem-loop structures of newly identified miRNA homologues of the miR4376 superfamily.

Structures were predicted using Mfold. MiRNA names were assigned according to the most similar miRNA hit retrieved by blasting the new miRNA sequence against miRbase (version 19). Mature miRNAs are highlighted in pink.

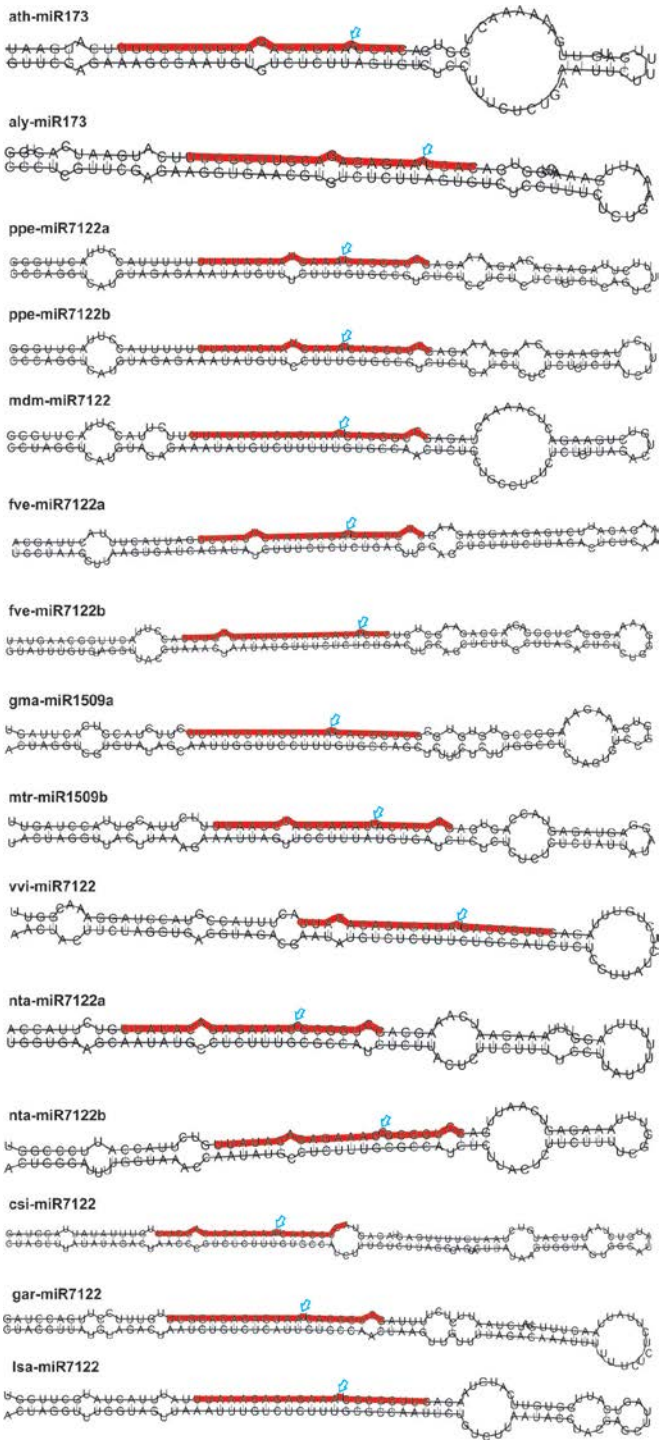


Figure 4-S8 Stem-loop structures of MIR7122s identified.

Stem-loop structures were produced by RNAfold. Mature miRNAs are marked with red lines; position of bulges in miRNA/miRNA* duplex is indicated with light-blue arrows.

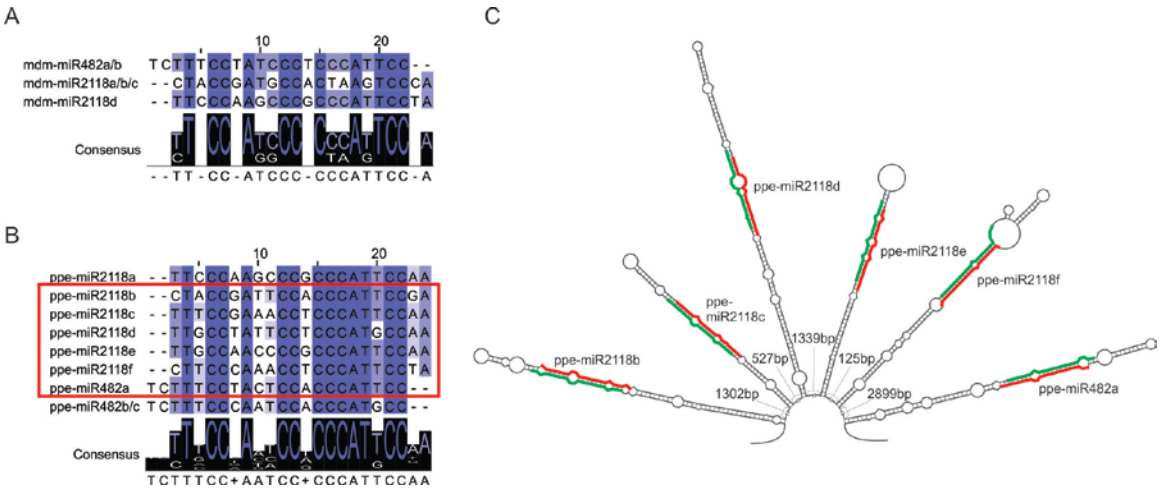


Figure 5-1 miR482/2118 in apple and peach

(A) miRNA alignment of the miR482/2118 superfamily in apple. Nucleotide conservation in each position is denoted with various colors with dark color corresponding to a high level conservation and light color to a low level. Consensus sequence and conservation logo are included below.

(B) miRNA alignment of the miR482/2118 superfamily in peach. Six miR482/2118 members clustered in the same scaffold are marked in the red box.

(C) Stem-loop structure of six miR482/2118 members. miRNA and miRNA* position are highlighted in red and green, respectively. The distances between two close miRNAs are indicated.

THE END