Coupling of microRNA-directed phased small interfering RNA generation from long noncoding genes with alternative splicing and alternative polyadenylation in small RNA-mediated gene silencing

Wuqiang Ma[1,2], Chengjie Chen[1,2], Yuanlong Liu[1,2], Ming Zeng[3], Blake C. Meyers[4,5], Jianguo Li[1,2] and Rui Xia[1,2]

[1]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou 510642, China; [2]Guangdong Litchi Engineering Research Center, College of Horticulture, South China Agricultural University, Guangzhou 510642, China; [3]Modern Education and Technology Center, South China Agricultural University, Guangzhou 510642, China; [4]Donald Danforth Plant Science Center, St Louis, MO 63132, USA; [5]Division of Plant Sciences, University of Missouri – Columbia, 52 Agriculture Lab, Columbia, MO 65211, USA

Authors for correspondence:
Rui Xia
Tel: +86 20 38348652
Email: rxia@scau.edu.cn

Jianguo Li
Tel: +86 20 85288265
Email: jianli@scau.edu.cn

Received: *11 October 2017*
Accepted: *31 October 2017*

New Phytologist (2018) **217:** 1535–1550
doi: 10.1111/nph.14934

Key words: alternative splicing/polyadenylation, long terminal repeat (LTR) retrotransposon, microRNA (miRNA), miR482/2118, noncoding *PHAS* locus, phasiRNA, sRNA-mediated gene silencing.

Summary

- MicroRNAs (miRNAs) and phased small interfering RNAs (phasiRNAs) play vital regulatory roles in plant growth and development. Little is known about these small RNAs in litchi (*Litchi chinensis*), an economically important fruit crop widely cultivated in Southeast Asia.
- We profiled the litchi small RNA population with various deep-sequencing techniques and in-depth bioinformatic analyses.
- The genome-wide identification of miRNAs, their target genes, and phasiRNA-generating (*PHAS*) genes/loci showed that the function of miR482/2118 has expanded, relative to its canonical function. We also discovered that, for 29 *PHAS* loci, miRNA-mediated phasiRNA production was coupled with alternative splicing (AS) and alternative polyadenylation (APA). Most of these loci encoded long noncoding RNAs. An miR482/2118 targeted locus gave rise to four main transcript isoforms through AS/APA, and diverse phasiRNAs generated from these isoforms appeared to target long terminal repeat (LTR) retrotransposons and other unrelated genes. This coupling enables phasiRNA production from different exons of noncoding *PHAS* genes and yields diverse phasiRNA populations, both broadening and altering the range of downstream phasiRNA-regulated genes.
- Our results reveal the diversity of miRNA and phasiRNA in litchi, and demonstrate AS/APA as a new layer of regulation in small RNA-mediated gene silencing.

Introduction

Small RNAs (sRNAs) are 20–24-nt-long molecules that play critical regulatory roles at the transcriptional or post-transcriptional level (Bartel, 2004; Voinnet, 2009; Rogers & Chen, 2013; Reis *et al.*, 2015; Achkar *et al.*, 2016). According to the mechanism of their biogenesis and function, sRNAs are typically classified into two groups, microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Axtell, 2013). miRNAs are generated by the sequential processing of DICER-LIKE 1 (DCL1) on their precursors which form a stem-loop structure. By contrast, siRNAs are produced from double-stranded RNAs (dsRNAs) created by the activity of RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) (Chen, 2009; Axtell, 2013).

In plants, there is a special type of siRNA, the biogenesis of which relies on the function of an miRNA. After miRNA-mediated cleavage, one fragment of the cleaved target mRNA is converted into dsRNA by RDR6, and siRNAs with a determined size (21 or 24 nt) are produced in a sequential, head-to-tail manner according to the miRNA cleavage site (Allen *et al.*, 2005; Yoshikawa *et al.*, 2005). This special head-to-tail arrangement is called a 'phased' pattern, and the siRNAs are phased or in phase relative to the miRNA cleavage site. Therefore, these siRNAs are more generally termed phased siRNAs (phasiRNAs), although a subset of them is known as trans-acting siRNAs (tasiRNAs) because of their *in trans* function characterized when they were first discovered (Peragine *et al.*, 2004; Allen *et al.*, 2005).

miR482/2118 is one of the most well characterized miRNA families which triggers 21-nt phasiRNA production in plants. Members of the family probably evolved from the duplication of *NUCLEOTIDE BINDING LEUCINE-RICH REPEAT (NB-LRR)* genes (Zhang *et al.*, 2016). This family apparently has dual

functions in gymnosperms, each of which appears to be selectively retained in eudicots and monocots (Xia *et al.*, 2015a). In eudicots, miR482/2118 predominately targets hundreds of *NB-LRR* genes and induces abundant phasiRNA production (Zhai *et al.*, 2011; Shivaprasad *et al.*, 2012). This pathway plays essential roles in plant disease resistance. Arabidopsis loss-of-function mutants in *NB-LRR* targeting (mir472, a variant of the miR482 family found in Arabidopsis) or phasiRNA generation RDR6(rdr6), exhibit enhanced effector-triggered immunity (ETI)-based resistance to some pathogens (Boccara *et al.*, 2014). However, in grasses (not yet demonstrated widely in monocots), miR482/2118 mainly targets a large number of noncoding genes and triggers the production of phasiRNAs greatly enriched in inflorescences (Johnson *et al.*, 2009; Zhai *et al.*, 2015). While the biological function of the miR482/2118-noncoding-phasiRNA pathway remains largely unknown, at least one of these phased loci, *PHOTOPHERIOD-SENSETIVE GENIC MALE STERILITY 1* (*PMS1T*), regulates photoperiod-sensitive male sterility in rice (*Oryza sativa*) (Fan *et al.*, 2016).

Alternative splicing (AS) of mRNA precursors, a strategy for controlling gene expression and generating protein complexity, is a process coupled tightly with miRNA-mediated gene silencing (Syed *et al.*, 2012; Yang *et al.*, 2012b; Reddy *et al.*, 2013). AS modulates miRNA-mediated regulation of gene expression in multiple ways. First, AS leads to the generation of mRNA variants either including or excluding miRNA target sites. For example, in Arabidopsis, over 12% of miRNA-binding sites in target mRNAs are affected by AS (Yang *et al.*, 2012b). For the *SQUAMOSAL-PROMOTER BINDING PROTEIN LIKE 4* (*SPL4*) gene, a key regulator of the transition from the vegetative to the reproductive phase (Wu *et al.*, 2009) and a canonical target gene of miR156, only one of its three mRNA isoforms contains an miRNA-binding site for miR156 (Yang *et al.*, 2012b). Second, AS regulates the splicing of primary miRNAs (pri-miRNAs), thereby modulating the levels of miRNAs, which in turn regulates the stability of target mRNAs. Many *MIRNA* genes reside within introns of protein-coding genes and are coexpressed and processed with the host gene (Lu *et al.*, 2008; Yang *et al.*, 2012a). Lastly, AS also affects miRNA biogenesis indirectly by regulating the splicing of precursor mRNAs (pre-mRNAs) of critical miRNA processing genes, for instance, DCL1, a key endonuclease in miRNA biogenesis (Yang *et al.*, 2012b). AS may influence the generation of phasiRNAs from their precursor genes as well, as a large number of protein-coding genes were found to generate phasiRNAs from annotated introns alone or intron–exon junctions (Zheng *et al.*, 2015). Another strategy contributing to mRNA complexity in cells is alternative polyadenylation (APA), which generates distinct 3′ termini on mRNAs. APA also affects mRNA stability or translation through miRNA activity, as most APA sites are located in 3′ untranslated regions (UTRs), which contain many miRNA target sites, especially in mammals (Di Giammartino *et al.*, 2011; Tian & Manley, 2017). These two processes, AS and APA, are frequently interconnected, which can possibly facilitate the definition of 3′-terminal exons (Di Giammartino *et al.*, 2011; Tian & Manley, 2017).

In this study, we aimed to carry out a genome-wide exploration of the diversity of small RNAs in litchi (*Litchi chinensis* Sonn.), an economically important fruit crop widely cultivated in Southeast Asia. Multiple high-throughput sequencing strategies, including sRNA sequencing, parallel analysis of RNA ends (PARE) sequencing, and strand-specific RNA-sequencing (RNA-seq), were applied in combination with in-depth bioinformatics analyses using a variety of computational methods. We found that litchi, like many other perennial plants, has a large number of miRNAs and target genes yielding phasiRNAs. The miR482/2118 family appears to have expansive roles beyond the regulation of disease resistance genes (*NB-LRR* genes). We have revealed that miR482/2118-mediated phasiRNA biogenesis is coupled with AS and APA of target noncoding mRNA, thereby leading to the production of distinct phasiRNAs from different noncoding gene regions, which in turn regulate different downstream genes *in trans*. This demonstrates the role of AS/APA as an important, additional layer of regulation in sRNA-mediated gene silencing.

Materials and Methods

Plant material collection

Nine 'Huaizhi' litchi (*Litchi chinensis* Sonn.) trees with similar growth conditions, located in an orchard at ShenZhen, China, were selected for material collection in 2013. Each set of three trees was considered as a biological replicate. A total of six different tissues, namely the fruit pedicel, fruitlet, leaf, root, shoot and flower, were randomly collected from all parts of the tree, frozen immediately in liquid nitrogen, and stored at −80°C.

RNA extraction, sequencing of sRNAs, PARE and strand-specific RNA sequencing (ssRNA-seq)

Total RNA was extracted using Trizol (Ambion, Waltham, MA, USA) plus Fruit-mate (Takara, Tokyo, Japan) for polysaccharide/polyphenol removal, according to the manufacturers' instructions. For sRNA sequencing, equal amounts of total RNA from three replicates with RNA integrity number (RIN) ≥7.5 were pooled for each tissue. Then 10 μg of the mixed total RNA was used for sRNA library construction. In total, six sRNA libraries were constructed and sequenced on an Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA) at the Beijing Genomics Institute (BGI, Shenzhen, China).

A total of four PARE libraries were built, including mixed tissues (equal amounts of RNA from all tissues), pedicel, flower and fruitlet. For each PARE library, 100 μg of total RNA was used. Sequencing was also executed on an Illumina HiSeq 2000 platform, either at the Beijing Genomics Institute (BGI) or at RIBOBIO (Guangzhou, China).

Finally, two ssRNA-seq libraries were generated from the flower tissue. For each library, 10 μg of total RNA was used. Libraries were constructed using the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA). Sequencing of 150-bp paired-end reads was conducted on an Illumina HiSeq 4000 platform by Annoroad Genomics (Beijing, China).

miRNA annotation

The annotation of litchi miRNAs followed a workflow that was described previously (Xia *et al.*, 2013, 2015a). In brief, after read

quality control and adapter trimming, abundance collapsed reads were mapped to the litchi genome (http://litchidb.genomics.cn/page/species/index.jsp) with Bowtie1 (http://bowtie-bio.sourceforge.net/index.shtml) with no mismatch allowed. Next, sRNAs of 20–22 nt in length and ≤ 20 genomic matches were subjected to screening for stem-loop structures (≤ 4 mismatches and ≤ 1 bulge). All miRNA abundances were normalized to reads per ten million (RPTM). Identified miRNAs were divided into known miRNAs and novel miRNAs by homologous searching against the plant miRNAs deposited in miRBase (www.mirbase.org, v21).

Multiple alignment of miR482/2118 members was performed using Clustalo (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default settings and viewed with Jalview (Waterhouse *et al.*, 2009). Secondary structures of miR482/2118 precursors were folded using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and viewed using Varna (http://varna.lri.fr/index.php?lang=en&page=home&css=varna).

Identification of miRNA target genes

Targetfinder v.1.6 (Fahlgren & Carrington, 2010) and Cleaveland v.2.0 (Addo-Quaye *et al.*, 2009) were used to analyze PARE data. The annotated gene file of litchi (Litchi-gene-v3.cds) was used for target gene identification. Venn charts were produced using TBtools (https://github.com/CJ-Chen/TBtools). All PARE reads were normalized to tags per ten million (TPTM).

phasiRNA-generating (*PHAS*) locus identification

Genome-wide searching for *PHAS* loci was conducted as described previously (Xia *et al.*, 2013). Loci with a *P*-value < 0.001 were considered to be reliable *PHAS* loci. A potential miRNA trigger of *PHAS* loci was predicted using a procedure described previously as 'reverse computation' (Xia *et al.*, 2013). Coding and noncoding loci/genes were classified following a workflow reported previously (Xia *et al.*, 2015a).

Strand-specific RNA-seq analysis

After quality control of raw reads using Fastqc v.0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Fastx_toolkit v.0.0.3 (http://hannonlab.cshl.edu/fastx_toolkit/), the retained clean reads were mapped to the litchi genome using Star v.2.5 (Dobin *et al.*, 2013) with default parameters except alignIntronMax set to 20 000. Mapping results of two biological replicates were merged together and viewed using an integrative genomics viewer (IGV) (Robinson *et al.*, 2011).

Small RNA northern blotting

For each sample, 10 μg of total RNA was used for sRNA blotting following a biotin-label-based protocol (Liu *et al.*, 2014). DNA oligonucleotides, reverse-complementary to the target miRNAs, were labeled with the Pierce™ BioTIN 3′ end DNA Labeling Kit (ThermoFisher, Waltham, MA, USA) and used as probes. The Chemiluminescent Hybridization and Detection Kit (Cat. 89880; ThermoFisher) was used for hybridization and signal detection. U6 RNA was used as an internal reference. Probes used in this study are listed in Supporting Information Table S1.

Rapid amplification of 3′ cDNA ends (3′-RACE) and qRT-PCR

Rapid amplification of 3′ cDNA ends (3′-RACE) was performed using the 3′ Full RACE Core Set (with PrimeScript™ RTase; Takara). Quantitative reverse transcription–polymerase chain reaction (qRT-PCR) was conducted with ReverTra Ace qRT-PCR master mix with gDNA Remover (Toyobo, Osaka, Japan) for reverse transcription and the SYBR Green-PCR master kit (Thunderbird SYBR qPCR mix; Toyobo) for qPCR on a Light Cycler 480_II Real-Time PCR system (Roche). All primers are listed in Table S1.

Accession numbers

The sRNA, PARE and ssRNA-seq data sets generated in this study have been deposited in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under the accession number GSE98698.

Results

Main regulatory pathways of miRNAs are conserved in litchi

To obtain a comprehensive view of the miRNA repertoire in litchi, we constructed and sequenced six sRNA libraries from different tissues, namely leaf, fruitlet, flower, pedicel, shoot, and root, with an average of *c.* 20 million reads (Table S2). sRNA data were processed and mapped to the litchi genome (http://litchidb.genomics.cn/page/species/index.jsp), and miRNAs were annotated according to a well-established protocol (Arikit *et al.*, 2014; Xia *et al.*, 2015a). We identified members of all 22 miRNA families that are highly conserved in plants (Cuperus *et al.*, 2011), a few less conserved miRNAs, including miR482/2118, miR535, miR828 and miR858, and a few other known miRNAs which had only been reported in a few species previously, such as miR3954 and miR3627 (Fig. 1; Table S3). In addition, we also set out to identify novel litchi-specific miRNAs. Using high stringency criteria (listed in the Materials and Methods section), 40 miRNAs/families were characterized as novel miRNAs, and were termed Lc-miRN1 to miRN40 (Table S4). Most of them were observed at low abundance, compared with conserved miRNAs, which are evaluated by read abundance, except for miRN1 (Table S4). One-third of the novel miRNAs (13 miRNAs) are 22-nt in length (Table S4), a feature of miRNAs that typically triggers the production of phasiRNAs from their targets.

Next, we profiled the target genes of the miRNAs identified using PARE sequencing (Zhai *et al.*, 2014). Four PARE libraries were constructed and sequenced from tissues including the flower, fruit and pedicel, and pooled samples with an average of *c.* 20 million reads (Table S2). We identified 44 target genes for

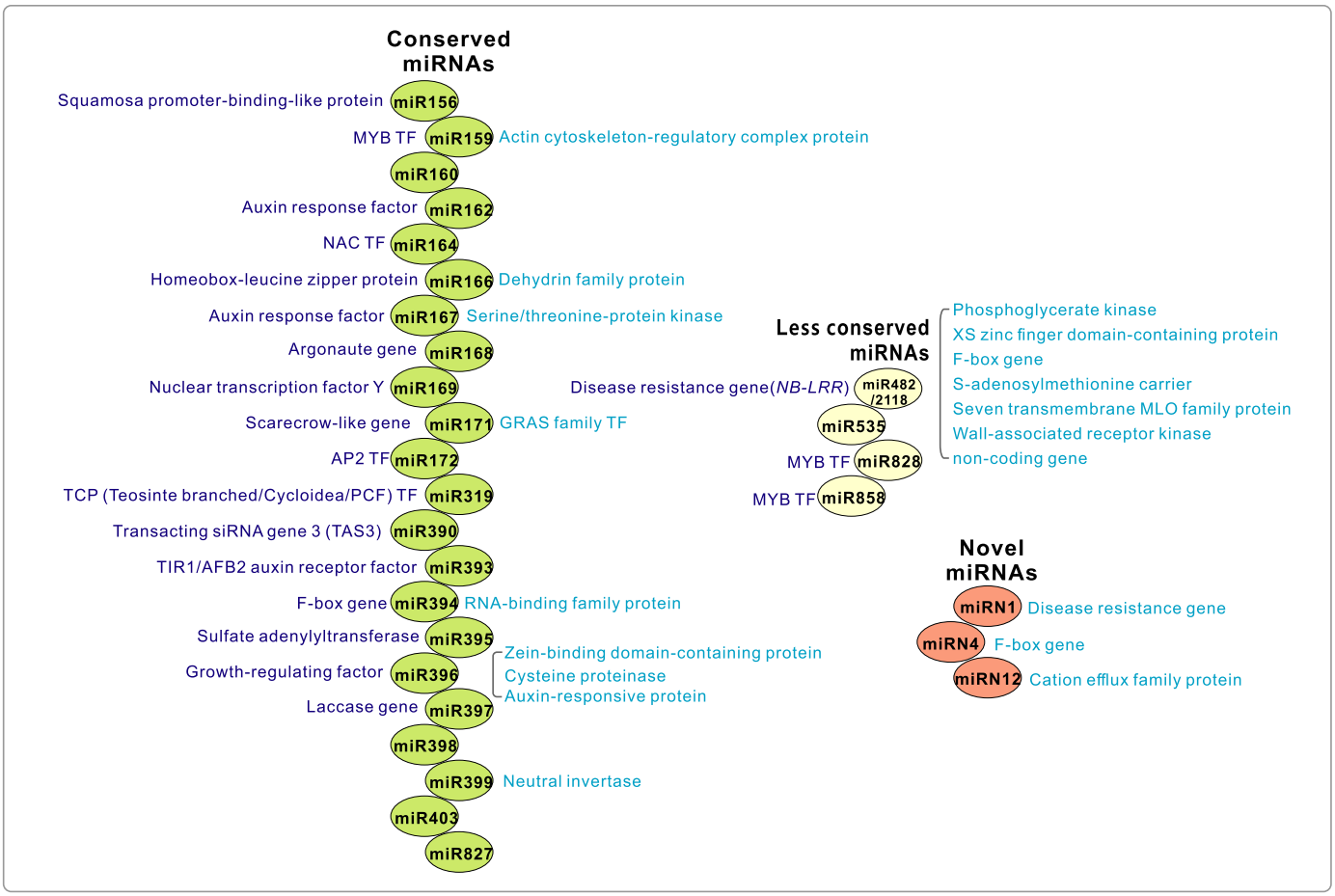


Fig. 1 Main microRNA (miRNA) regulatory pathways identified in litchi. Target genes of an miRNA family are listed with conserved target genes on the left (in dark blue), and newly identified target genes on the right (in light blue). TF, Transcription factor; MYB, myeloblastosis; NAC, (NAM, ATAF and CUC) proteins; GRAS, (GAI, RGA and SCR) proteins; TIR/AFB2, Transport inhibitor response 1/Auxin receptor F-box 2; AP2, Apetala 2.

13 novel miRNAs, and most cleavages of these target genes were supported by a weak PARE signal, except those targets of Lc-miRN1, Lc-miRN4 and Lc-miRN12 (Figs 1, S1a; Table S5). For the known miRNAs, target genes were identified for 18 out of the 22 conserved miRNA families, and all these miRNAs had conserved target genes identified (Table S6); that is, these miRNA–target relationships are widely conserved in plants. For example, miR156 targets *SPL* genes, and miR160 and miR167 target *AUXIN RESPONSIVE FACTOR* (*ARF*) genes (Table S6). All these results were robustly supported by PARE data (category = 0), indicating that the major miRNA regulatory pathways are completely conserved in litchi, and are probably involved in the same essential functions as in other plants (Fig. 1). In addition to the previously described conserved target genes, we identified novel target genes for most of these known miRNAs, but most of them were less supported by PARE data (category > 0; Table S6). However, the miR396 and miR482/2118 families were exceptional (Figs 1, S1b; Table S6). They had high-confidence target genes not previously observed in other species. For miR482/2118, in addition to a large number of well-described *NB-LRR* disease resistance genes, evidence of cleavage was detected for five other genes including those encoding a phosphoglycerate kinase, an F-box protein, an S-adenosylmethionine carrier, a seven transmembrane Mildew resistance locus O (MLO) family protein, and a wall-associated receptor kinase (category = 0; Figs S1b, 1; Table S6). Moreover, miR482/2118 targeted a few other unknown, possible noncoding genes (Table S6). These data suggested that the miR482/2118 function was greatly expanded in litchi, which is further discussed in more detail below (in Fig. 4).

PhasiRNAs were produced extensively from the litchi genome

PhasiRNAs are another major class of sRNAs, specifically produced in plants, and their production typically relies on a cleavage of precursor transcripts mediated by miRNAs. To explore the phasiRNA population in litchi, we conducted genome-wide analysis to identify the phasiRNA-generating genes or loci, and characterize their trigger miRNAs. We identified > 500 *PHAS* loci across the litchi genome, of which 420 are protein-coding genes. The trigger–gene pairs were representative of the major *PHAS* pathways found previously in plants, including miR482/2118–*NB-LRR* (disease resistance genes), miR828–*MYELOBLASTOSIS* (*MYB*), miR393–*TRANSPORT INHIBITOR RESPONSE 1/AUXIN RECEPTOR F-BOX 2*

(a)

Major *PHAS* pathways in litchi

miRNA trigger	Sequence	Targeted *PHAS* gene
miR167	UGAAGCUGCCAGCAUGAUCUUA	*Auxin responsive factor*
miR390	AAGCUCAGGAGGGAUAGCGCC	*Trans-acting siRNA gene 3 (TAS3)*
miR393	UCCAAAGGGAUCGCAUUGAUCC	*Transport inhibitor response 1/ Auxin receptor F-box proteins (TIR1/AFB)*
miR482/2118	UUGCCAAUUCCACCCAUGCCGU	*Disease resistance gene (NB-LRR genes) F-box gene Wall-associated receptor kinase gene Seven transmembrane MLO family gene noncoding genes*
miR828	UCUUGCUCAAAUGAGUAUUCCA	*MYB transcription factor*
miR3954	UUGGACAGAAAAAUCACGGUCA	*NAC transcription factor/noncoding gene*
Lc-miRN1	UCAUCGGAACACAAGCCCAUGC	*Disease resistance gene*

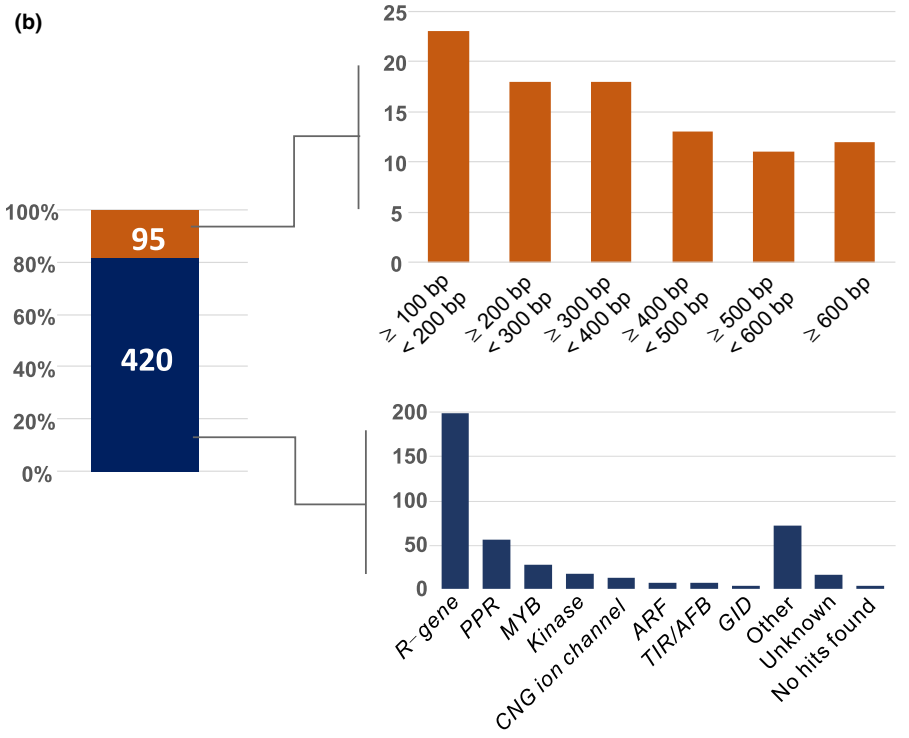


Fig. 2 Complex phasiRNA-generating (*PHAS*) networks in litchi. (a) Major *PHAS* pathway with microRNA (miRNA) trigger identified. (b) Classification of *PHAS* loci identified. Noncoding (orange) loci are grouped based on the length of the *PHAS* loci, and coding (dark-blue) genes were classified according to their functional annotation. R, disease resistance; PPR, pentatricopeptide repeat-containing; MYB, myeloblastosis; CNG, cyclic nucleotide-gated; ARF, auxin responsive factor; TIR/AFB, transport inhibitor response 1/auxin receptor F-box; GID, gibberellin insensitive dwarf; MLO, mildew resistance locus O; NB-LRR, nucleotide binding leucine-rich repeat gene; NAC, (NAM, ATAF and CUC) proteins.

(TIR1/AFB) and miR167–*ARF* (Fig. 2a). There were many *PENTATRICOPEPTIDE REPEAT-CONTAINING* (*PPR*) genes giving rise to abundant phasiRNAs, as in other plants (Xia *et al.*, 2013; Fig. 2b). In addition to these major conserved *PHAS* pathways, a great number of genes from small families, or single-gene families, were also routed into the pathway of phasiRNA biogenesis, indicating a complex utilization of networks of phasiRNAs in litchi. For instance, miR3954, which was previously reported only in citrus, targeted transcripts of genes encoding an *NAC* (*NAM, ATAF* and *CUC*) transcription factor to trigger phasiRNA generation in litchi (Fig. 2a). Finally, we also found the 22-nt miRNA (Lc-miRN1) that cleaved transcripts of *NB-LRR* disease resistance genes to yield phasiRNAs (Fig. 2a).

In addition to protein-coding genes, we also identified 95 predicted noncoding genes producing abundant phasiRNAs (Fig. 2b). Most of these noncoding *PHAS* loci are of short length based on sRNA data, consistent with a low likelihood of coding protein (Fig. 2b). Four paralogs of *TRANSACTING SIRNA GENE 3* (*TAS3*) widely conserved in land plants (Xia *et al.*, 2017), were found in litchi (Fig. S2). They were targeted by conserved miR390, and produced tasiRNA targeting *ARF* genes (tasiARFs) (Table S6). We also observed that miR482/2118 targeted, in addition to numerous *NB-LRR* disease resistance genes, a large number of predicted noncoding *PHAS* genes to trigger phasiRNA production (Tables S5, S7).

Diverse function of the miR482/2118 family in litchi

miR482/2118 is one of the most well-characterized miRNA families which trigger 21-nt phasiRNA production in plants. This miRNA family mainly targets either hundreds of *NB-LRR* genes in eudicots or a large number of noncoding genes in grasses, and

triggers secondary phasiRNA production in both cases (Johnson *et al.*, 2009; Zhai *et al.*, 2011, 2015; Shivaprasad *et al.*, 2012). In view of the diverse functions of the miR482/2118 family demonstrated in litchi, targeting both protein-coding and a large number of noncoding genes, we conducted more detailed analyses of the function of this miRNA family. In the litchi genome, we identified seven distinct miRNAs belonging to the miR482/2118 superfamily (Fig. 3a). These miRNAs included five of the miR482 type and two of the miR2118 type, derived from 17 stem-loop precursors which were named accordingly (Fig. 3a). Most members of the miR482/2118 family were abundant in all six tissues that we studied, except Lc-mR482i-j; many of them were equally abundant among these tissues (Fig. 3b). Interestingly, a few of stem-loop precursors clustered together in the genome in close proximity, such that they were probably transcribed as a single precursor gene (Figs 3c, S3). In total, four pairs of miR482/2118 members were found in the genome (Fig. 3c).

A total of 63 annotated genes were identified as target genes of the miR482/2118 family (Table S6). In addition to the large number of *NB-LRR*s, the targets included other new genes, such as those encoding a zinc-finger protein, kinase, gibberellin receptor and many others (Fig. 4a). The capacity of miR482/2118 to trigger phasiRNA generation from *NB-LRR*s was also observed in litchi (exemplified in Fig. 4b). From the large set of *PHAS* loci/genes identified, 38 loci contained a target site of miR482/2118 (Fig. 4b). Although we identified 63 miR482/2118 target genes using PARE analysis (Table S6), only six of them were confirmed as *PHAS* genes in our *PHAS* analyses (Fig. 4c; Table S7). This small amount of overlap was perhaps attributable to the low expression of target genes rendering phasiRNAs too weak to be detected by our *PHAS* analysis pipeline. Another probable reason for this low overlap is the incomplete and inaccurate gene annotation of the litchi genome, resulting in many genuine miR482/2118 target genes not being included in the current annotated model gene set (used for PARE analysis of miRNA targets).

miR482/2118-directed phasiRNA production is coupled with AS/APA

As our *PHAS* analysis indicated that the gene annotation of the litchi genome was incomplete, to obtain a more accurate and reliable gene annotation, deep ssRNA-seq was employed to profile mRNA transcripts. As we observed the largest number of *PHAS* loci in the sRNA library of flowers, we sequenced two biological replicates of ssRNA-seq libraries from the same flower tissue. ssRNA-seq analysis revealed that many of the *PHAS* loci contained long introns, which broke *PHAS* genes into several discontinuous *PHAS* regions. These breaks sometimes separated the *PHAS* region from their miRNA trigger sites. This phenomenon was extraordinarily prevalent for miR482/2118.

Our *PHAS* analyses revealed two *PHAS* loci from scaffold886 (scaffold886:83993-85112 and scaffold886:94907-95571), with a distance of *c.* 10 kb between them (Fig. 5a,b; Table S8). When we checked the sRNA data for this region, we found two other regions with abundant siRNA production between the two initially identified *PHAS* regions (Fig. 5c). Further analysis led us to conclude that most of the sRNAs mapped to these two middle regions were 21 nt long and the majority of them were phased, suggesting that these two regions are also *PHAS* loci (Figs 5c, S4). ssRNA-seq reads were exclusively mapped to the antisense strand, indicating that a primary transcript was transcribed in the antisense direction (Fig. 5e). The primary mRNA covered a *c.* 15-kb genomic region encompassing all the four sRNA-producing regions. The splicing pattern revealed by ssRNA-seq indicated that the mRNAs transcribed from this genomic locus (referred to as the scaffold886 locus) were complex and included a high degree of AS (Fig. 5d–f; described in more detail below in Fig. 6). There were diverse splicing patterns, especially for the splicing of the second intron; its splicing donor site was consistently at the same position but there were five different acceptor sites, giving rise to introns of dramatically different length (Fig. 5d,e,f). To verify these alternative spliced transcripts, we utilized RT-PCR and 3′ RACE to amplify possible transcripts derived from this locus. We successfully obtained at least seven transcript variants (described in more detail below in Fig. 6), with four main transcript isoforms (T1–4) with distinct polyadenylation sites (Fig. 5g), corresponding precisely to the major four splicing patterns (Fig. 5d). These four transcript variants covered the four main *PHAS* regions, denoted AS1–AS4 according to their genomic location (Fig. 5c), and confirmed that the four *PHAS* regions were derived from AS in conjunction with APA of this locus (Fig. 5g). T1–T4 consisted of three or four exons, with the first two exons shared by all of them. A target site of miR482/2118 was found in the second shared exon; it initiated phasiRNA production from all the four major transcripts over the region spanning the third (for T1–3) or third and fourth exons (for T4) (Fig. 5). The cleavage of the miR482/2118 target site was well supported by PARE data (Fig. 5b). Taken together, these findings indicate that the *PHAS* regions AS1–AS4 were derived from transcript isoforms generated from the same gene locus through AS/APA, and their phasiRNA production was triggered by the same miR482/2118 site, shared by all the four main transcripts T1–T4. This result demonstrates a novel mechanism coupling 22-nt miRNA-directed phasiRNA generation with the AS/APA of pre-mRNAs to control the genomic sequences generating phasiRNAs.

Complex transcript variants generated by AS/APA from the scaffold886 locus

Our ssRNA-seq data showed clearly that the mRNA structure of the scaffold886 locus is of great complexity. The splicing junctions detected using ssRNA-seq data were highly consistent in two biological replicates (Fig. S5) and with publicly released RNA-seq data (Fig. S6), and were generally validated by the PARE data (50-bp-long reads; Figs 6a, S5). To confirm the sRNA generation of this region, we also sequenced three more sRNA library replicates from litchi flowers (Fig. S7). Combining all the deep sequencing data (ssRNA-seq, PARE, and sRNA), we deduced that there were at least 14 potential transcript variants generated from this genomic locus, with seven of them verified by 3′-RACE (Fig. 6a). All these variants were of different exons

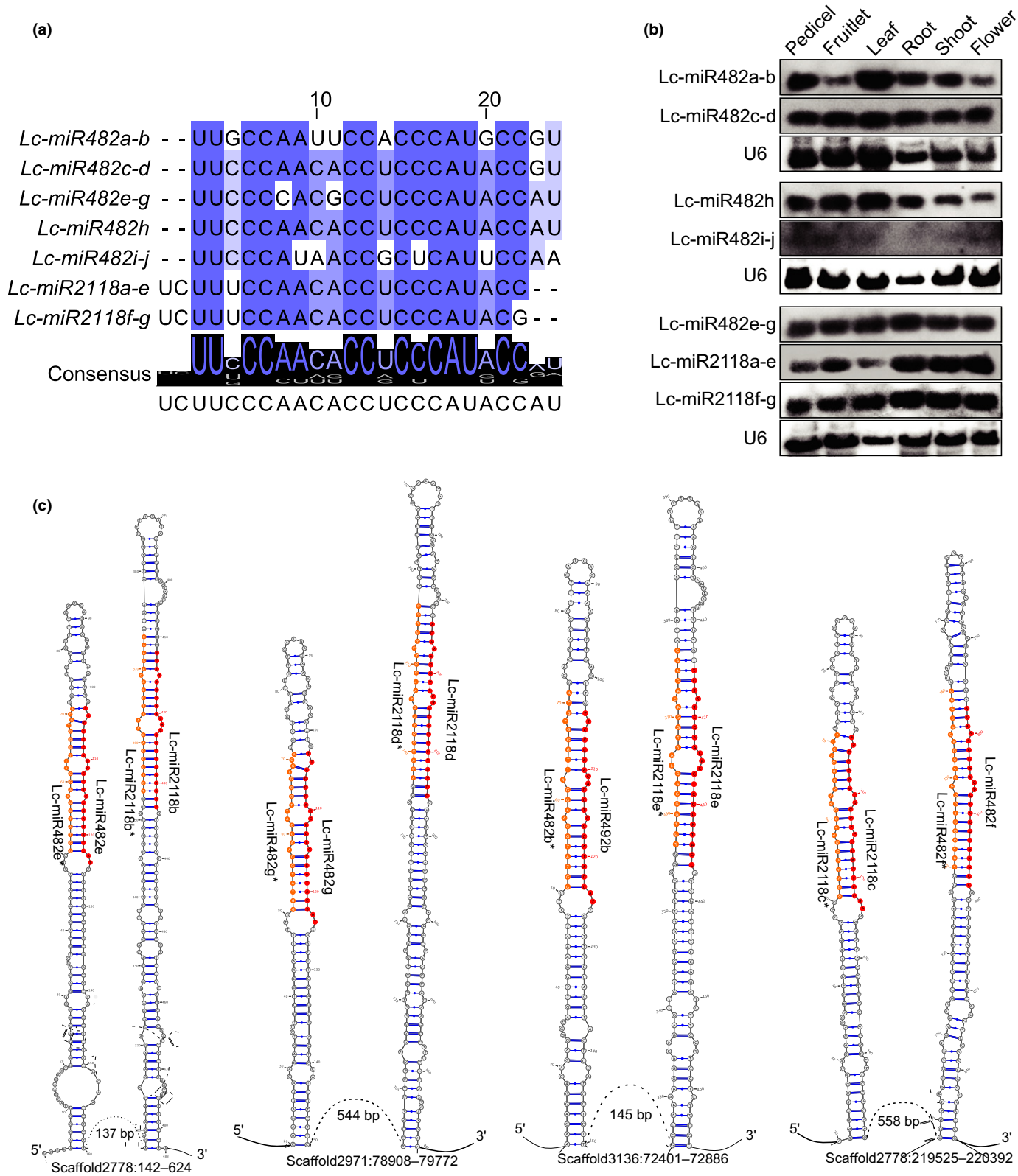


Fig. 3 microRNA482 (miR482)/2118 family in litchi. (a) Alignment of mature sequences for all members of the miR482/miR2118 family. The degree of conservation for each nucleotide along the miRNAs is represented by the color, with a dark color denoting a high level of conservation and a light color a low level. The consensus sequence of the alignment is displayed below with sequence logos. (b) Expression of all miR482/2118 members. U6 was used as a loading control. (c) Stem-loop structures of four miR482/2118 clusters in litchi. miRNA and miRNA* are marked in red and orange, respectively.

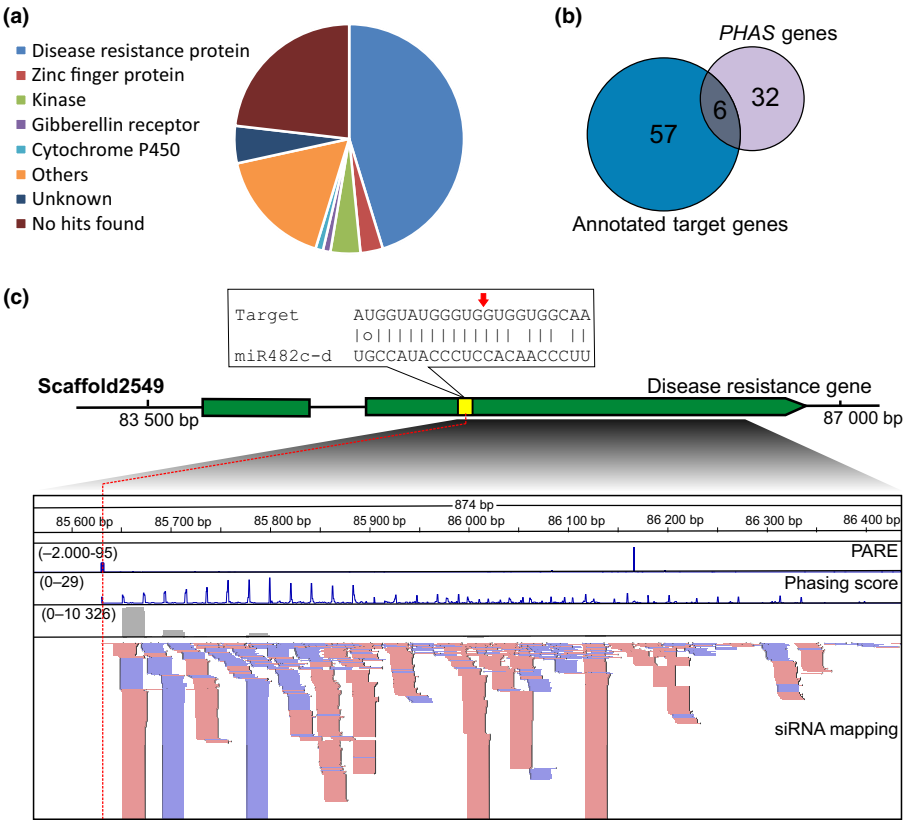


Fig. 4 Expanded function of microRNA482 (miR482)/2118 in litchi. (a) All target genes of miR482/2118 identified. (b) Six target genes of miR482/2118 verified by parallel analysis of RNA ends (PARE) (63 targets) were identified as phasiRNA-generating (*PHAS*) loci by PHAS analysis (38 loci). (c) miR482/2118 targets a disease resistance gene (*NB-LRR*), leading to the production of phasiRNAs. The PARE data, phasing score and mapping results of sRNA data are viewed in different tracks in an integrative genomics viewer (IGV). A diagram of the disease resistance gene is drawn above with the base pairing between miRNA and the target included at the top and the exact cleavage site indicated by a red arrow. Brick-red and blue lines in the sRNA track denote the sRNA reads matching the antisense and sense strands, respectively.

resulting from AS or had different polyadenylation sites caused by APA. For instance, transcripts V8, V9 and V10 have additional exons, which are spliced out in transcript V7 (T3), and the second intron of V3, V5 and V9 is longer than that of V2 (T1), V4 (T2), and V8, respectively. All these transcript variants were transcribed from the same transcription start site (TSS1; Fig. 6a). Another potential site (TSS2) was observed from the ssRNA-seq data, which is likely to give rise to a few more transcripts without the miR482/2118 target site (Fig. 6a), adding more complexity to the transcript variants of the genomic locus. Taken together, these findings suggest that AS/APA leads to the generation of a large number of transcript variants from the scaffold886 locus.

PhasiRNAs resulting from different AS/APA variants target different long terminal repeat (LTR) retrotransposons

In plants, phasiRNAs act *in cis* or *in trans* to regulate their target genes (Zhai *et al.*, 2011; Fei *et al.*, 2013). We thus investigated the potential targets of the phasiRNAs generated from the four *PHAS* regions of the scaffold886 locus (AS1–AS4), corresponding to the AS/APA transcript variants T1–T4. We first compared the sequence similarity of AS1–AS4 and found that AS1 showed low sequence similarity (< 55%) to the other three, which had relatively high mutual similarity (*c.* 70%; Fig. S8a). Multiple stop codons observed in the translated amino acid sequences and coding potential evaluation suggested that those four transcript variants (T1–T4) were unlikely to encode proteins (Fig. S8b), implying that they probably perform their function via the resulting phasiRNAs. Intriguingly, sequence similarity of various degrees to retrotransposon gag proteins was detected for all the *PHAS* regions, except AS1, when a homology search using BLASTX was conducted in GenBank. Next, knowing that AS2–AS4 probably have homology to retrotransposons, we performed a genome-wide annotation of retrotransposons and found that all three *PHAS* regions AS2–AS4 were part of LTR retrotransposon-like elements (Fig. 6a). However, while these elements contained the LTR regions and a partial gag protein sequence, none of them contained a complete *pol* gene which typically encodes enzymatic functions critical for the element translocation (Fig. 6a). Moreover, these LTR retrotransposon-like elements were not fully transcribed, as each of the T2–T4 transcripts covered only part of these elements. Thus, we deduced that the transcripts T2–T4 were more likely to be *PHAS* genes than LTR retrotransponsons, functioning to regulate the expression of LTR retrotransposons.

To test this hypothesis, we conducted PARE data analysis of the full set of LTR retrotransposons annotated in the litchi genome. Considering the high level of sequence similarity of LTR retrotransposons and the short length of PARE reads (50 bp), to distinguish between LTR retrotransposons that are genuinely expressed and cleaved by phasiRNAs and those that are highly similar in sequence but are transcriptionally not expressed or expressed at low levels, we applied rigorous criteria for our PARE analysis, requiring (1) only phasiRNAs (in exact phase to the miRNA cleavage site) with ≥ 10 raw reads; (2) expression of LTR retrotransposons confirmed by RNA-seq data; (3) PARE reads uniquely mapped to LTR retrotransposons; (4) good PARE data support with category ≤ 3. We were able to confirm 54 events of phasiRNA-directed cleavage on LTR retrotransposons (Table S9).

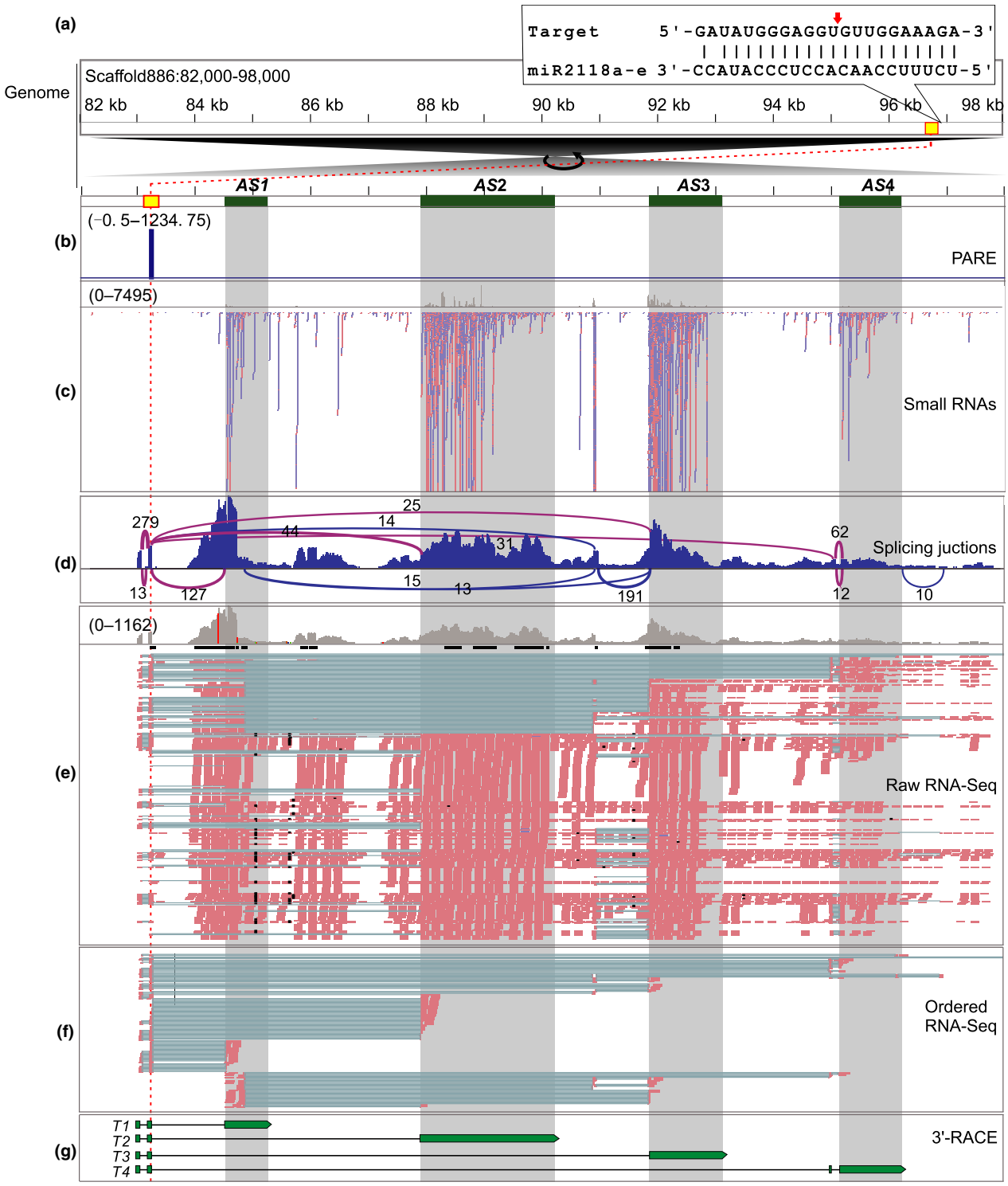
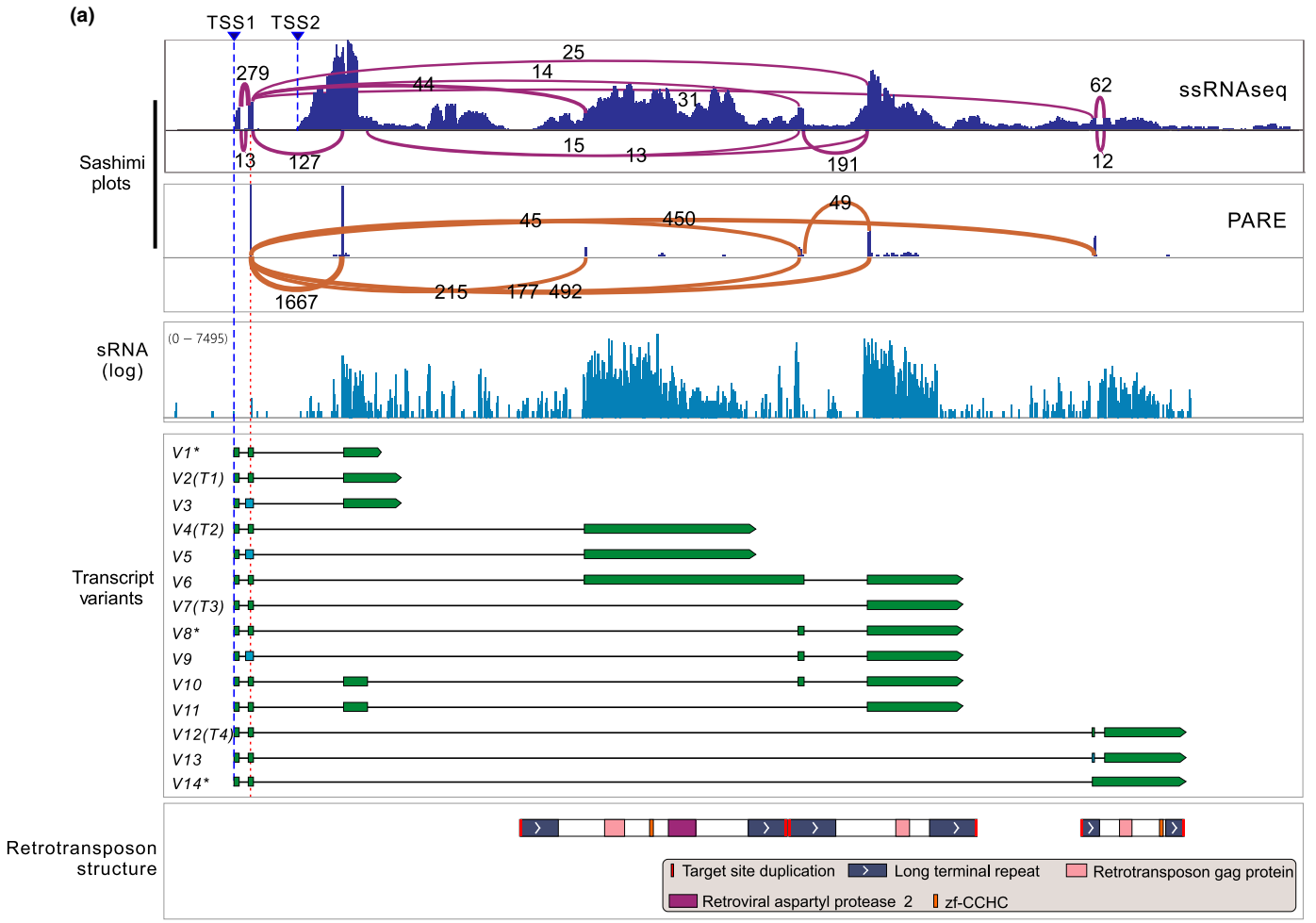


Fig. 5 microRNA482 (miR482)/2118-directed phasiRNA generation is coupled with the alternative splicing/polyadenylation of pre-mRNAs at the scaffold886 locus. Different sets of data are all viewed in an integrative genomics viewer (IGV). (a) Genomic view of the locus. The base pairing between miR482/2118 and the target is included at the top, with the exact cleavage site indicated by a red arrow. The round arrow denotes that the view of all the data below is turned around for a better presentation as the mRNA transcripts of the locus are all on the antisense strand. (b) Parallel analysis of RNA ends (PARE) data, with the blue bar indicating the cleavage site of miR482/2118. (c) Mapping results of small RNA (sRNA) data from the flower library. Brick-red and blue lines denote the sRNA reads matching the antisense and sense strands, respectively. (d) Splicing patterns deduced from the strand-specific RNA sequencing (ssRNA-seq) data. Numbers on each curve denote the number of ssRNA-seq reads supporting a given splicing pattern, and only splicing patterns with ≥10 raw RNA-seq reads supported are shown in the plot. Curves in purple mark the splicing patterns matching the four major transcript isoforms. (e) Mapping results of the ssRNA-seq data, with brick-red lines denoting the RNA-seq reads matched to the antisense strand and light-blue lines the intron regions. (f) Ordered mapping results of the ssRNA-seq data to better show the intron position. (g) mRNA transcripts amplified by Rapid amplification of 3′ cDNA ends (3′-RACE). The green boxes are exons and the middle lines are introns.

New Phytologist



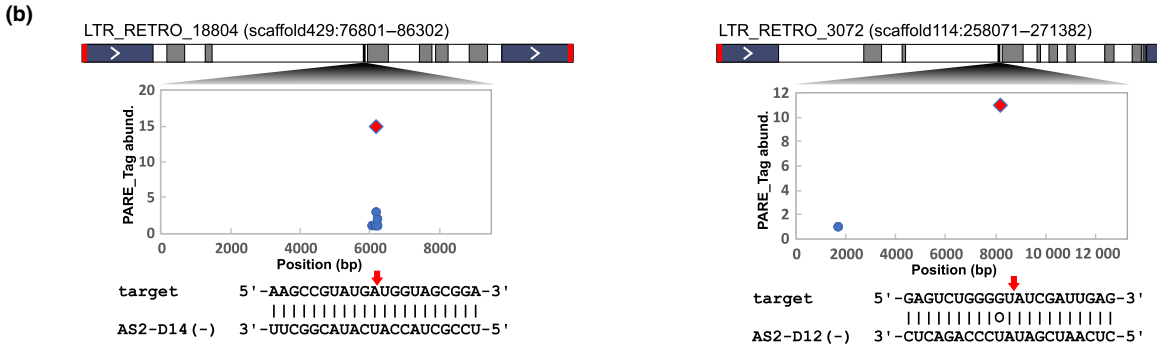


Fig. 6 Phased small interfering RNAs (phasiRNAs) generated from transcript variants of the scaffold886 locus target long terminal repeat (LTR) retrotransposons. (a) Transcript isoforms overlapped with annotated LTR retrotransposons. Sashimi plots show the splicing patterns deduced from strand-specific RNA sequencing (ssRNA-seq) and parallel analysis of RNA ends (PARE) data at the top, with the small RNA (sRNA) coverage data (log_{10} value) shown in the middle. In sashimi plots, numbers on each curve denote the number of ssRNA-seq reads supporting a given splicing pattern, and only splicing patterns with ≥10 raw RNA-seq or PARE reads supported are shown in the plot. Fourteen transcript variants deduced from deep sequencing data are shown below the sRNA coverage track, with those validated by RT-PCR/Rapid amplification of 3′ cDNA ends (3′-RACE) marked with an asterisk. The green boxes are exons and the middle lines are introns. Positions of LTR retrotransposons are shown at the bottom. Two transcription start sites (TSS1 and TSS2) are indicated on the top with small blue triangles plus blue dotted lines, and the target site of miR482/2118 is marked with a red dotted line. (b) T-plots for two LTR retrotransposons targeted by phasiRNAs generated from the scaffold886 locus. The structure of the LTR retrotransposons is denoted by a diagram above the T-plot. The *y*-axis indicates the abundance of cleavage PARE reads (reads per 10 million (RPTM)). The *x*-axis indicates the coding sequence nucleotide position (bp), with 1 marking the first nucleotide of the start codon. The red diamond denotes the cleavage site by the corresponding miRNA. The base pairing between miRNA and the target transcript is indicated beneath each plot, with the exact cleavage site indicated by a red arrow.

A few of these retrotransposons had high PARE data support (category = 0), as exemplified by the target gene of AS2-D14(-) or AS2-D12(-) in Fig. 6(b). These results verified that LTR retrotransposons are the primary target genes of the phasiRNAs derived from the AS/APA variants of the scaffold886 locus.

Although sequences of AS2–AS4 are very similar, their resulting phasiRNAs are probably different in sequence and have the capacity to regulate distinct target genes. To test this theory, we used all putative phasiRNAs from AS1–AS4 (21-nt in-phase siRNAs, and the sequences were extracted computationally) for target prediction against the annotated litchi transcriptome and LTR retrotransposons, and found that most of the target genes of phasiRNAs from different transcript isoforms are not shared among them (Fig. S8c), implying that phasiRNAs generated from different AS/APA variants are probably able to regulate different genes.

Spatiotemporal expression of transcripts covering the AS1–AS4 regions

Next, we compared the phasiRNA products of the scaffold886 locus among different tissues. Overall, phasiRNAs were more abundant in flower than other tissues (Fig. S9) and more phasiRNAs were generated from the AS2 and AS3 *PHAS* regions (Fig. 7a), indicating that the miR482/2118-mediated phasiRNA pathway may function more specifically in flowers to regulate LTR retrotransposons or other target genes. We also checked the expression level of transcript variants covering the AS1–AS4 phasiRNA-generating regions by qRT-PCR with primers specific to these regions (Fig. 7b). The transcript expression of AS1 and AS2 among different tissues agreed well with the accumulation of phasiRNAs generated from them, but transcripts of AS3 showed high expression in all studied tissues except root (Figs 7c, S9). Transcripts of AS4 were constitutively expressed at low levels in all tissues, in agreement with the relatively low production of phasiRNAs of AS4 (Figs 7b, S9). As transcript expression and sRNA accumulation were predominant in the flower tissue, we performed a more detailed evaluation of transcript expression during the flower development of litchi. Litchi flowers are dioecious; two sets of flower samples were collected after the flower sex was identifiable. For the male flower, expression of transcripts of AS1–AS4 gradually increased with male flower development and peaked at the MF4 stage, with generally no significant expression difference among transcripts of different AS regions (Fig. 7d). By contrast, transcripts of AS1–AS4 showed different

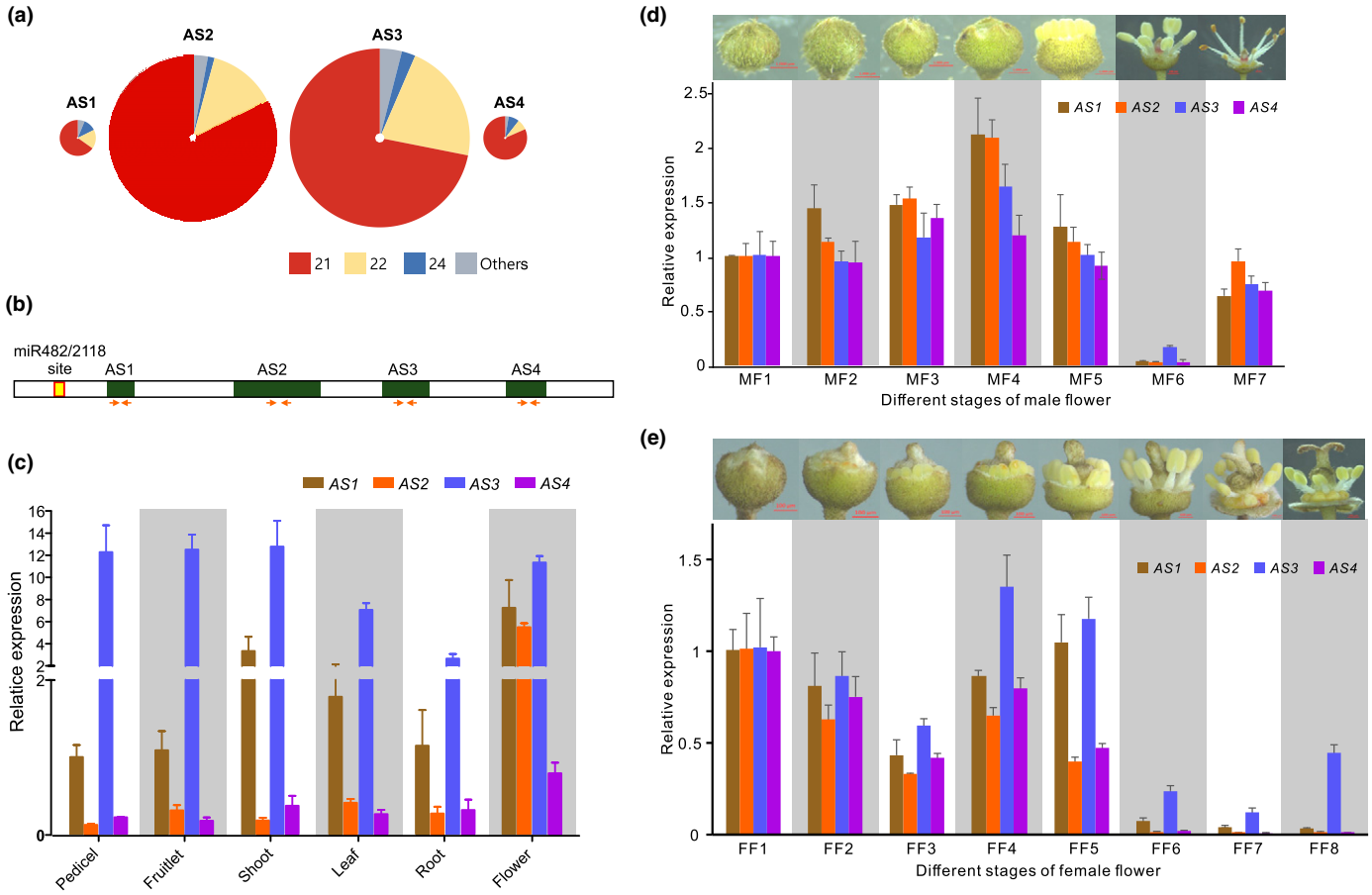


Fig. 7 Spatiotemporal expression of transcripts covering the four phasiRNA-generating regions. (a) Abundance of small RNAs (sRNAs) mapped to the AS1–AS4 regions in the flower library. (b) Primers specific to the AS1–AS4 regions were designed for quantitative RT-PCR. (c) Expression of transcripts covering the four phasiRNA-generating regions (AS1–AS4) in different tissues of litchi. Values are means ± SD of three replicates. The RNAs used were the same as the libraries of sRNAs. (d, e) Expression of transcripts covering the four phasiRNA-generating regions (AS1–AS4) during the development of male (d) and female (e) flowers in litchi. Photos of flower buds at each stage are included at the top. Values are means ± SD of three replicates.

expression in the female flower, with transcript expression of AS2 and AS4 generally decreasing with female flower development and that of AS1 and AS3 decreasing until the FF3 stage and increasing at the stages of FF4 and FF5, after which the expression of all transcripts, except those of AS3, was hardly detected (Fig. 7e). All these results suggested that the transcript variants covering phasiRNA-generating regions are probably expressed in a tissue- and stage-specific manner.

A large number of *PHAS* loci display AS/APA

The discovery of the role of AS/APA in phasiRNA production from the scaffold886 locus prompted us to carry out a broader survey of this phenomenon over the litchi genome. We found another 28 genomic loci that had AS/APA variants and produced a considerable amount of phasiRNAs (Tables 1, S10; Fig. S10).

Similar to the scaffold886 locus, we found another locus on scaffold3895, which generated phasiRNAs with the involvement of AS/APA (Fig. S11a–f). The two major phasiRNA-generating regions (AS1 and AS2 in Fig. S11b–f) show a significant level of sequence homology to transcripts encoding kinase protein (Fig. S11g), implying that the phasiRNAs generated from those two regions probably target homologous kinase genes. Indeed, a few kinase genes were verified as targets for the phasiRNAs generated from both AS1- and AS2-derived phasiRNAs (Table S11). Also, phasiRNAs generated from different *PHAS* regions (AS1 and AS2) or phases (AS2a and AS2b) seemed to have the potential to target a large number of different genes in spite of also targeting a few common genes (Fig. S11h), and AS/APA transcripts also showed tissue-specific expression (Fig. S11i), largely consistent with the accumulation pattern of sRNAs in different tissues (Fig. S12). Taken together, these findings indicate that the scaffold3895 locus produces diverse transcript isoforms and generates a range of phasiRNAs to regulate the expression of kinase genes or other genes.

Of the other 27 loci identified, more than half had sequence similarity to known proteins deposited in UniProt (www.uniprot.org), such as F-BOX PROTEIN, MEAL TOLERANCE PROTEIN and CYCLIC NUCLEOTIDE GATED CHANNEL 1 PROTEIN (Table 1). However, *in silico* translated amino acid sequences of most of the transcripts generated from those loci do not display significant open reading frames (ORFs). Thus, the main function of these loci is probably to regulate the expression of target genes through the generation of phasiRNAs. Interestingly, although most of these AS/APA *PHAS* loci were triggered by miR482/2118, a few were targeted by another miRNA, miR3954, suggesting that the pathway of phasiRNA generation coupled with AS/APA is not miRNA-specific; that is, any miRNA capable of triggering phasiRNA generation can probably function through this pathway.

Discussion

In this study, we profiled two major types of sRNA, miRNAs and phasiRNAs, in an economically important subtropical fruit tree, litchi. Our analyses revealed that this perennial tree species has complex miRNA and phasiRNA regulatory networks, and in particular the miR482/2118 family has evolved broader functions in addition to its well-conserved function. We also discovered that miRNA-mediated phasiRNA production is coupled with AS/APA of noncoding pre-mRNAs to generate more sequence-specific phasiRNAs, thereby regulating more diverse target genes. These results demonstrate that AS/APA is an important layer of regulation in sRNA-directed gene silencing.

Overview of miRNA and phasiRNA networks in litchi

In plants, with the development and wide employment of deep sequencing technologies, sRNAs have been studied not only in model plants, but also in more and more economic crops. Here we profiled the sRNA repertoire of litchi in different tissues. Compared with the recent study of litchi miRNAs (Yao *et al.*, 2015), we performed much deeper analyses with the sequencing of more libraries (six different tissue samples), the usage of the draft litchi genome sequence, and more thorough bioinformatic data mining. We found that the overall composition of the miRNA regulatory network in litchi is similar to that of other plants. The 22 conserved miRNA families and their corresponding target genes are present in litchi, consistent with a fundamental role in plant growth and development. However, several highly conserved miRNA families were found to have evolved new target genes in litchi, such as miR396 and miR482/2118, indicating that conserved miRNA families can both maintain a basic function and be co-opted for new target genes and putative functions. In addition to miRNAs, we revealed a complex phasiRNA network, with the involvement of a large number of coding or noncoding *PHAS* loci/genes. Although we were able to identify the most well-characterized (from other species) miRNA–*PHAS*–phasiRNA circuits, such as miR390–*TAS3*, miR393–*TIR1/AFB*, miR828–*MYB* and miR482/2118–*NB-LRR* (Fig. 2a), only a limited number of *PHAS* loci/genes had an identifiable miRNA trigger despite our exhaustive analyses of extensive sRNA and PARE data. Thus, we concluded that the mechanism of phasiRNA biogenesis may be more complicated than we have understood. Worthy of note is that miR3954, which was previously reported only in citrus (Xu *et al.*, 2010), is also found in litchi; in citrus, miR3954 affects flowering time through targeting two noncoding *PHAS* genes and a gene encoding an *NAC* transcription factor and triggering phasiRNA generation (Liu *et al.*, 2014, 2017). We found that miR3954 also targeted both noncoding genes and *NAC* genes in litchi, suggesting that miR3954 might play a role in the regulation of flowering in litchi.

Diverse function of the miR482/2118 family

miR482/2118 has become one of the most well-studied miRNAs in plants (Zhai *et al.*, 2011; Shivaprasad *et al.*, 2012; Fei *et al.*, 2013; Xia *et al.*, 2015a; Zhang *et al.*, 2016). It emerged in seed plants, and has dual functions in gymnosperms, which have probably been selectively retained in eudicots and monocots, in both cases, triggering phasiRNA generation from cognate target

Table 1 Alternative spliced/polyadenylated loci with phased small interfering RNA (phasiRNA) production in litchi

Trigger miRNA	Genomic coordinate	Alignment score	Target sequence	Trigger sequence	Annotation
miR482/2118 family					
Lc-miR482h	scaffold3578:17205-23340	4.5	GCGGAAUGGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	F-box protein
Lc-miR2118a-e	scaffold1190:693924-702577	3	GGAAUGGGAGGUUUUGGAAAGA	UCUUUCCAACACCUCCCAUACC	Metal tolerance protein
Lc-miR2118a-e	scaffold3124:29663-39191	1	GGAAUGGGAGGUGUUGGAAAGA	UCUUUCCAACACCUCCCAUACC	Metal tolerance protein
Lc-miR2118a-e	scaffold2175:4811-15729	2	GGUACGGGAGGUGUUGGGAAGA	UCUUUCCAACACCUCCCAUACC	Cyclic nucleotide gated channel 1
Lc-miR482h	scaffold2112:180437-190568	1	AUGGUACGGGAGGUGUUGGGAA	UUCCCAACACCUCCCAUACCAU	Cyclic nucleotide gated channel 1
Lc-miR482h	scaffold1688:354530-364765	4	AUGGUAUGGCAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	Sentrin-specific protease 1-like
Lc-miR482h	scaffold582:167917-176528	1	AUAGUAUGGGAGGUGUUGGGAA	UUCCCAACACCUCCCAUACCAU	Cysteine-rich receptor-like protein kinase 10
Lc-miR482h	scaffold886:80110-98980	5	UCGAUAUGGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	DNA/RNA polymerase superfamily protein
Lc-miR2118a-e	scaffold1831:123304-130958	2	GGUAUGGGAGGUGUUAGAAAGA	UCUUUCCAACACCUCCCAUACC	Alpha-L-arabinofuranosidase 1-like
Lc-miR482a-b	scaffold2549:75125-81654	4.5	ACGGUUUGGGUGGUGUUGGCAA	UUGCCAAUUCCACCCAUGCCGU	Disease resistance protein
Lc-miR482h	scaffold1883:524192-531523	4	ACAGUAUGGGAGGUAUUGGGAA	UUCCCAACACCUCCCAUACCAU	MLO-like protein 11
Lc-miR2118a-e	scaffold2393:40857-51254	3	GGAAUGGGAGGUGUUUGAAAGA	UCUUUCCAACACCUCCCAUACC	Protein PHR1-LIKE 1
Lc-miR2118a-e	scaffold868:267884-287885	3	GGAAUGGGAGGAGUUGGAAAGA	UCUUUCCAACACCUCCCAUACC	Hypothetical protein
Lc-miR2118a-e	scaffold1488:46064-56490	3	GGAAUGGGAGGUGUUUGAAAGA	UCUUUCCAACACCUCCCAUACC	Protein PHR1-LIKE 1
Lc-miR482h	scaffold2176:401778-419792	5	AUCGUACAGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	Phosphoglycerate kinase, cytosolic-like
Lc-miR482h	scaffold2295:444492-452423	4	ACAGUAUGGGAGGUAUUGGGAA	UUCCCAACACCUCCCAUACCAU	Retrovirus-related Pol polyprotein
Lc-miR2118a-e	scaffold3895:27377-41301	3	GGAAUGGGAGCUGUUGGAAAGA	UCUUUCCAACACCUCCCAUACC	Wall-associated receptor kinase-like 1
Lc-miR482h	scaffold919:354,809-372,259	2	AUGGUAUGGGACGUGUUGGGAA	UUCCCAACACCUCCCAUACCAU	Probable strigolactone esterase DAD2
Lc-miR482h	scaffold1093:384512-393217	5	AUGGUACGGAAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	No hits found
Lc-miR482h	scaffold2175:205772-214884	5	AUGGUACGGAAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	No hits found
Lc-miR482h	scaffold3466:1031-6575	3	AUGGAAUGGGAGGUGUGGGGAA	UUCCCAACACCUCCCAUACCAU	No hits found
Lc-miR482h	scaffold3828:49147-64375	4.5	GCGGAAUGGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	No hits found
Lc-miR482h	scaffold591:1269403-1291093	2	AUGGUAUGGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	No hits found
Lc-miR2118a-e	scaffold1190:644372-653171	3	GGAAUGGGAGGUUUUGGAAAGA	UCUUUCCAACACCUCCCAUACC	No hits found
Lc-miR482h	scaffold441:5670-15074	5	UCGAUAUGGGAGGUGUUGGAAA	UUCCCAACACCUCCCAUACCAU	No hits found
miR3954 family					
Lc-miR3954a/b	scaffold1190:178,567-188,405	2.5	UGAUCGUGAUUUCUCUGUCCAA	UUGGACAGAAAAAUCACGGUCA	Metal tolerance protein
Lc-miR3954a/b	scaffold1190:328,488-343,253	2.5	UGAUCGUGAUUUCUCUGUCCAA	UUGGACAGAAAAAUCACGGUCA	Chromosome transmission fidelity protein
Lc-miR3954a/b	scaffold962:706,455-713,323	2.5	UGAUCGUGAUUUCUCUGUCCAA	UUGGACAGAAAAAUCACGGUCA	Metal tolerance protein
Lc-miR3954a/b	scaffold962:16,183-22,246	2.5	UGAUCGUGAUUUCUCUGUCCAA	UUGGACAGAAAAAUCACGGUCA	Nucleotide binding protein

MLO, mildew resistance locus O; PHR1, phosphate response 1; DAD2, decreased apical dominance.

genes (Xia *et al.*, 2015a). In eudicots, miR482/2118 targets a large number of *NB-LRR* genes, but it targets hundreds or even thousands of noncoding genes in grasses (Nystedt *et al.*, 2013; Zhai *et al.*, 2015). In addition to these well-known functions, the miR482/2118 family has also evolved lineage-specific functions. For example, in strawberry (*Fragaria versa*), this miRNA family has 18 target genes unrelated to *NB-LRR* genes, including an *F-BOX* gene, a *CYCLIC PHOSPHODIESTERASE* gene, and a *TRANSPARENT TESTA12* gene (Xia *et al.*, 2015b). In a number of other species, lineage-specific targets of miR482/2118 have also been found, including apple (*Malus* × *domestica*) (Xia *et al.*, 2012), peach (*Prunus persica*) (Zhu *et al.*, 2012), and soybean (*Glycine max*) (Arikit *et al.*, 2014). In our current work in litchi, we found that miR482/2118 has diverse functions, targeting not only *NB-LRR* genes, but also many other genes not reported previously, such as genes encoding zinc-finger proteins and kinases. Several newly identified target genes are noncoding genes (or pseudogenes), and are dissimilar to those found in other species. Therefore, across many species, miR482/2118 has an active role in the gain and loss of target genes.

PhasiRNA generation coupled with AS/APA of pre-mRNAs

phasiRNAs are a major class of sRNAs in plants, and their generation relies on the activity of an miRNA trigger in either a 'one-hit' or 'two-hit' mode (Fei *et al.*, 2013). The latter is observed for only a few particular miRNAs (typically miR390), while the former is more pervasive and typical for 22-nt, 5′ 'U' miRNAs (Chen *et al.*, 2010; Cuperus *et al.*, 2010). Like other sRNAs, phasiRNAs function *in cis* or *trans* to regulate downstream genes via sequence complementarity. AS of pre-mRNA is a central mechanism of genetic regulation in higher eukaryotes, and it is a major source of mRNA sequence differences through which the protein diversity produced by the genome is increased (Black, 2003; Reddy *et al.*, 2013). APA of mRNA is another process contributing to mRNA complexity in cells, which is frequently interconnected with AS (Di Giammartino *et al.*, 2011; Tian & Manley, 2017). AS/APA can generate complex and diverse mRNA isoforms, as exemplified in Fig. 8(a). Here, we demonstrate that noncoding (or pseudogenic) *PHAS* loci in litchi have complex AS/APA patterns and abundant phasiRNA production, demonstrating that miRNA-directed phasiRNA generation is coupled with pre-mRNA AS/APA to increase the diversity of phasiRNAs, which in turn expands the capacity and complexity of phasiRNA targets. As exemplified by the scaffold886 locus, the four main mRNA variants generated through AS/APA give rise to distinct groups of phasiRNAs after miR482/2118 targeting; these phasiRNAs subsequently target a large number of LTR retrotransposons or other potential target genes (Figs 5, 6). Similarly, the scaffold3895 locus also appears to generate diverse phasiRNAs as a result of AS/APA, regulating kinase genes or other genes (Fig. S11). Thus, we model a mechanism by which AS/APA yields diverse phasiRNAs, using a common miRNA trigger located in an exon close to a splicing donor site with mRNA isoforms generated through different splicing acceptor sites in conjunction with different polyadenylation sites (Fig. 8a). This coupling of phasiRNA generation with AS/APA leads to the production of distinct populations of phasiRNAs (Fig. 8b). These phasiRNAs subsequently regulate groups of downstream genes, both related to and distinct from each other (Fig. 8c).

There are several conceivable benefits to the coupling of phasiRNA generation with AS/APA. The first is an increase in the diversity of phasiRNAs, and the capacity to regulate diverse downstream target genes. The second is the introduction of another control point modulating phasiRNA targets. AS/APA is a spatiotemporally controlled process. It can regulate the

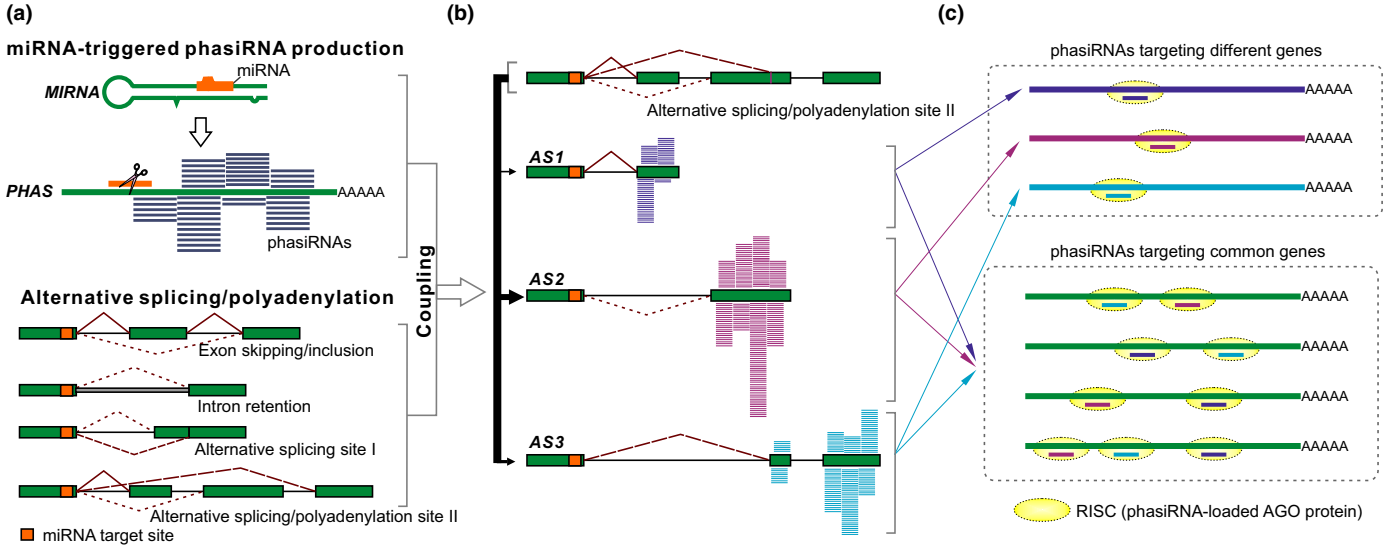


Fig. 8 A model coupling microRNA (miRNA)-directed phased small interfering RNA (phasiRNA) generation with alternative splicing (AS)/alternative polyadenylation (APA) in small RNA-mediated gene silencing. The coupling mechanism enables the production of diverse phasiRNA populations from different genomic regions, broadening the range of downstream phasiRNA-regulated genes. (a) miRNA-directed phasiRNA generation is coupled with AS/APA of pre-mRNAs. (b) Different populations of phasiRNAs are generated from the AS/APA transcript variants. (c) Distinct phasiRNAs yielded by AS/APA have a broad function, both targeting different genes and co-regulating common genes.

expression of each mRNA variant, and thereby the abundance of phasiRNAs from each mRNA variant, and ultimately the expression of target genes of resulting phasiRNAs. The third is the coordination of the expression of both related and unrelated genes, by inclusion or exclusion of exons that yield phasiRNAs from different mRNA variants. phasiRNA-generating regions of different mRNA variants share no or little sequence similarity (like the scaffold3895 locus), thus giving rise to distinct phasiRNAs to regulate unrelated target genes; the integration of phasiRNA production with AS/APA provides a regulatory connection between those unrelated targets. There are many demonstrated biotechnological applications of sRNA-mediated gene silencing (Price & Gatehouse, 2008; Auer & Frederick, 2009). The coupling of phasiRNA generation with AS/APA may add yet another tool to the genetic manipulation toolkit, providing a new way to conditionally regulate the expression of a diverse sets of genes, as we have demonstrated occurs in litchi.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (#31471859), the Outstanding Talent Program of the Ministry of Agriculture, the Chinese Thousand Young Talents Program, the Innovation Team Project of the Department of Education of Guangdong Province (2016KCXTD011), and the China Agricultural Research System (#CARS-33-11). Research in the Meyers lab is supported by the US National Science Foundation Plant Genome Research Program (#1339229). We thank members of the Xia and Li labs for helpful discussions.

Author contributions

R.X. and J.L. conceived the study. W.M., Y.L., J.L. and R.X. collected and generated the data. W.M., C.C., M.Z. and R.X performed the data analyses. W.M., B.C.M., J.L. and R.X. wrote the article.

ORCID

Rui Xia http://orcid.org/0000-0003-2409-1181

References

Achkar NP, Cambiagno DA, Manavella PA. 2016. miRNA biogenesis: a dynamic pathway. *Trends in Plant Science* **21**: 1034–1044.

Addo-Quaye C, Miller W, Axtell M. 2009. CleaveLand: A pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* **25**: 130–131.

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**: 207–221.

Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdés-López O, Prince S, Musket TA, Nguyen HT *et al.* 2014. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell* **26**: 4584–4601.

Auer C, Frederick R. 2009. Crop improvement using small RNAs: applications and predictive ecological risk assessments. *Trends in Biotechnology* **27**: 644–651.

Axtell MJ. 2013. Classification and comparison of small RNAs from plants. *Annual Review of Plant Biology* **64**: 137–159.

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

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* **72**: 291–336.

Boccara M, Sarazin A, Thiébeauld O, Jay F, Voinnet O, Navarro L, Colot V. 2014. The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathogens* **10**: e1003883.

Chen X. 2009. Small RNAs and their roles in plant development. *Annual Review of Cell and Developmental Biology* **25**: 21–44.

Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, Wu S-H. 2010. 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proceedings of the National Academy of Sciences, USA* **107**: 15269–15274.

Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, Takeda A, Sullivan CM, Gilbert SD, Montgomery TA, Carrington JC. 2010. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nature Structural & Molecular Biology* **17**: 997–1003.

Cuperus JT, Fahlgren N, Carrington JC. 2011. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**: 431–442.

Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Molecular Cell* **43**: 853–866.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Fahlgren N, Carrington JC. 2010. miRNA target prediction in plants. *Methods in Molecular Biology* **592**: 51–57.

Fan Y, Yang J, Mathioni SM, Yu J, Shen J. 2016. PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proceedings of the National Academy of Sciences, USA* **113**: 15144–15149.

Fei Q, Xia R, Meyers BC. 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**: 2400–2415.

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan G-LL, Walbot V, Sundaresan V, Vance V, Bowman LH. 2009. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Research* **19**: 1429–1440.

Liu Y, Ke L, Wu G, Xu Y, Wu X, Xia R, Deng X, Xu Q. 2017. miR3954 is a trigger of phasiRNAs that affects flowering time in citrus. *Plant Journal* **38**: 42–49.

Liu Y, Wang L, Chen D, Wu X, Huang D, Chen L, Li L, Deng X, Xu Q. 2014. Genome-wide comparison of microRNAs and their targeted transcripts among leaf, flower and fruit of sweet orange. *BMC Genomics* **15**: 695.

Lu C, Jeong D-H, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D *et al.* 2008. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proceedings of the National Academy of Sciences, USA* **105**: 4951–4956.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A *et al.* 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**: 579–584.

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes & Development* **18**: 2368–2379.

Price DRG, Gatehouse JA. 2008. RNAi-mediated crop protection against insects. *Trends in Biotechnology* **1**: 393–400.

Reddy ASN, Marquez Y, Kalyna M, Barta A. 2013. Complexity of the alternative splicing landscape in plants. *Plant Cell* **25**: 3657–3683.

Reis RS, Eamens AL, Waterhouse PM. 2015. Missing pieces in the puzzle of plant microRNAs. *Trends in Plant Science* **20**: 721–728.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**: 24–26.

Rogers K, Chen X. 2013. Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* **25**: 2383–2399.

Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BACM, Baulcombe DC. 2012. A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *Plant Cell* **24**: 859–874.

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS. 2012. Alternative splicing in plants – coming of age. *Trends in Plant Science* **17**: 616–623.

Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**: 18–30.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**: 669–687.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Wu G, Park MY, Conway SR, Wang J-WW, Weigel D, Poethig RS. 2009. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell* **138**: 750–759.

Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z. 2013. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell* **25**: 1555–1572.

Xia R, Xu J, Arikit S, Meyers BC. 2015a. Extensive families of miRNAs and PHAS loci in norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Molecular Biology and Evolution* **32**: 2905–2918.

Xia R, Xu J, Meyers BC. 2017. The emergence, evolution, and diversification of the miR390-TAS3-ARF tathway in land plants. *Plant Cell* **29**: 1232–1247.

Xia R, Ye S, Liu Z, Meyers BC, Liu Z. 2015b. Novel and recently evolved microRNA clusters regulate expansive F-BOX gene networks through phased small interfering RNAs in wild diploid strawberry. *Plant Physiology* **169**: 594–610.

Xia R, Zhu H, An Y, Beers EP, Liu Z. 2012. Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biology* **13**: R47.

Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X. 2010. Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. *BMC Genomics* **11**: 246.

Yang GD, Yan K, Wu BJ, Wang YH, Gao YX, Zheng CC. 2012a. Genomewide analysis of intronic microRNAs in rice and Arabidopsis. *Journal of Genetics* **91**: 313–324.

Yang X, Zhang H, Li L. 2012b. Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. *Plant Journal* **70**: 421–431.

Yao F, Zhu H, Yi C, Qu H, Jiang Y. 2015. MicroRNAs and targets in senescent litchi fruit during ambient storage and post-cold storage shelf life. *BMC Plant Biology* **15**: 181.

Yoshikawa M, Peragine A, Mee YP, Poethig RS. 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes and Development* **19**: 2164–2175.

Zhai J, Arikit S, Simon SA, Kingham BF, Meyers BC. 2014. Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods* **67**: 84–90.

Zhai J, Jeong D-HH, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA *et al.* 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes and Development* **25**: 2540–2553.

Zhai J, Zhang H, Arikit S, Huang K, Nan G-L, Walbot V, Meyers BC. 2015. Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proceedings of the National Academy of Sciences, USA* **112**: 3146–3151.

Zhang Y, Xia R, Kuang H, Meyers BC. 2016. The diversification of plant NBS-LRR defense genes directs the evolution of microRNAs that target them. *Molecular Biology and Evolution* **33**: 2692–2705.

Zheng Y, Wang Y, Wu J, Ding B, Fei Z. 2015. A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. *BMC Biology* **13**: 32.

Zhu H, Xia R, Zhao B, An Y, Dardick CD, Callahan AM, Liu Z. 2012. Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biology* **12**: 149.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Examples of target genes for known and novel miRNAs identified in litchi.

Fig. S2 *TAS3* genes identified in litchi.

Fig. S3 Transcription of *MIR482/2118* clusters shown in ssRNA-seq data.

Fig. S4 Phasing score of the four *PHAS* regions (AS1–AS4) in the scaffold886 locus.

Fig. S5 Splicing patterns of the scaffold886 locus are consistent between the two biological replicates of ssRNA-seq.

Fig. S6 Raw mapping results of ssRNA-seq, PARE and sRNA viewed together in IGV.

Fig. S7 Additional RNA-seq and sRNA data supporting the splicing pattern of the scaffold886 locus.

Fig. S8 Sequence analyses of AS1–AS4 phasiRNA of the scaffold886 locus.

Fig. S9 phasiRNA abundance of AS1–AS4 of the scaffold886 locus.

Fig. S10 Graphical views of alternatively spliced loci that are coupled with phasiRNA generation in litchi.

Fig. S11 phasiRNAs generated from alternatively spliced transcripts of the scaffold3895 locus target kinase or other genes.

Fig. S12 phasiRNA abundance of AS1–AS2 of the scaffold3895 locus

Table S1 Primers and probes used in this study

Table S2 Summary of sRNA and PARE libraries in litchi

Table S3 Known miRNAs identified in litchi

Table S4 Novel miRNAs identified in litchi

Table S5 Target genes of novel miRNAs in litchi

Table S6 Target genes of known miRNAs in litchi

Table S7 *PHAS* loci of miR482/1448 family identified in litchi

Table S8 *PHAS* loci identified in litchi

Table S9 LTR-transposon targets for the phasiRNAs from AS2–AS4 of the scaffold886 locus

Table S10 Alternatively spliced/polyadenylated loci coupled with miRNA-directed phasiRNA generation in litchi

Table S11 Target genes of the phasiRNAs from the scaffold3895 locus

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.