

RESEARCH

Open Access



Phased secondary small interfering RNAs in *Panax notoginseng*

Kun Chen^{1†}, Li Liu^{1,2†}, Xiaotuo Zhang^{3†}, Yuanyuan Yuan^{4†}, Shuchao Ren¹, Junqiang Guo³, Qingyi Wang³, Peiran Liao¹, Shipeng Li¹, Xiuming Cui^{1,5,6*}, Yong-Fang Li^{4*} and Yun Zheng^{2,3*}

From 16th International Conference on Bioinformatics (InCoB 2017) Shenzhen, China. 20-22 September 2017

Abstract

Background: Recent results demonstrated that either non-coding or coding genes generate phased secondary small interfering RNAs (phasiRNAs) guided by specific miRNAs. Till now, there is no studies for phasiRNAs in *Panax notoginseng* (Burk.) F.H. Chen (*P. notoginseng*), an important traditional Chinese herbal medicinal plant species.

Methods: Here we performed a genome-wide discovery of phasiRNAs and its host PHAS loci in *P. notoginseng* by analyzing small RNA sequencing profiles. Degradome sequencing profile was used to identify the trigger miRNAs of these phasiRNAs and potential targets of phasiRNAs. We also used RLM 5'-RACE to validate some of the identified phasiRNA targets.

Results: After analyzing 24 small RNA sequencing profiles of *P. notoginseng*, 204 and 90 PHAS loci that encoded 21 and 24 nucleotide (nt) phasiRNAs, respectively, were identified. Furthermore, we found that phasiRNAs produced from some pentatricopeptide repeat-contain (PPR) genes target another layer of PPR genes as validated by both the degradome sequencing profile and RLM 5'-RACE analysis. We also found that miR171 with 21 nt triggers the generations of 21 nt phasiRNAs from its conserved targets.

Conclusions: We validated that some phasiRNAs generated from PPRs and TASL genes are functional by targeting other PPRs in *trans*. These results provide the first set of PHAS loci and phasiRNAs in *P. notoginseng*, and enhance our understanding of PHAS in plants.

Keywords: PHAS, Phased small interfering RNA (phasiRNA), *Panax notoginseng*, High-throughput sequencing, Bioinformatics

Background

Plant small RNAs, with 21 to 24 nucleotides (nt), play crucial roles in a variety of biological processes, including development, stress responses, defense and epigenetic modifications [1, 2]. Based on the origin and biogenesis, small RNA in plants can be divided into two main

categories: microRNA (miRNA) and small interfering RNA (siRNA); siRNA can be further divided into three classes: heterochromatic siRNA (hc-siRNAs), natural antisense transcript siRNA (NAT-siRNA) and phased, secondary siRNA (phasiRNA) [3, 4]. The common feature of these small RNA is that they are generated by different DICER-LIKE (DCL) member from their long precursors, and then are incorporated into an effector complex to guide target recognition and subsequent silencing events in a sequence-specific manner [4, 5].

MicroRNAs are derived from their precursors, which are non-coding RNAs and can form hairpin-like secondary structures. The miRNA precursors are generated by RNA polymerase II (Pol II); RNase III enzyme DCL1 is responsible for the biogenesis of the mature

*Correspondence: sanqi37@vip.sina.com; yongfangli@gmail.com; zhengyun5488@gmail.com

†Equal contributors

¹Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650500, China

⁵Key laboratory of *Panax notoginseng* resources sustainable development and utilization of state administration of traditional Chinese medicine, Kunming 650500, China

Full list of author information is available at the end of the article

miRNAs via processing of miRNA precursors. Most miRNAs negatively regulate their target genes through homolog based mRNA cleavage or translation inhibition at post-transcriptional level [2, 6], however some miRNAs may activate their targets through different mechanisms [7–9]. Small interfering RNAs (siRNAs) are characterized by their biogenesis depending on different RNA-dependent RNA polymerase (RDR) member. hc-siRNAs are 24 nucleotides in length, generated by DCL3 activity [10]. hc-siRNAs play a crucial role in directing DNA methylation and histone modification in an AGO4-dependent pathway, and regulate target gene expression in transcription level [11, 12]. NAT-siRNAs derives from dsRNA precursors formed by bidirectional transcription of two partially overlapping genes [13–15]. PhasiRNA is a kind of secondary siRNA. Phase simple means these siRNAs are generated precisely in phased pattern initiated at a specific nucleotide. Its biogenesis requires an initiative cleavage on the phasiRNA precursor transcript (PHAS) by a specific miRNA in either “one-hit” or “two-hit” manner, then one of the cleaved product is made double stranded by RDR6 and then catalyzed by DCL4 into 21 nt siRNA in a phased pattern [16–20]. Some of the phasiRNAs may further target their parental genes in *cis* or other genes in *trans* [21].

PhasiRNAs can be generated from either long noncoding RNAs or coding genes. Arabidopsis miR173 (*TAS1* and *TAS2*), miR390 (*TAS3*), and miR828 (*TAS4*) can function as guides on non-coding primary transcripts to initiate tasiRNA1-2, tasiRNA3 and tasiRNA4 processing, respectively. These tasiRNAs can further target pentatricopeptide repeat (PPR) family member, auxin response factors (ARFs) and MYB transcription factor in *trans* manner [17, 19, 22, 23]. Among them, tasiRNA3 is much conserved in land plants [16, 17]. Recently, 21- and 24-nt phasiRNAs derived from long noncoding RNAs have been reported in rice and maize male reproductive organs, they are triggered by miR2118 and miR2275, and cleaved by DCL4 and DCL5 (also known as DCL3b), respectively [20, 24]. A noncoding PHAS locus, triggered by miR4392, was found accumulating preferentially in soybean anthers [25].

PhasiRNAs, triggered by miRNAs, are also generated from protein-coding loci in many plants, and first found in Arabidopsis [26]. It is worthy to point out that a significant number of PPR, nucleotide binding site-leucine-rich repeat (NB-LRR) and MYB family members are PHAS loci [21, 26–29]. PPR family is one of the largest families in Arabidopsis, containing at least 448 PPR-related genes. Many PPR-P clade transcripts have been identified as PHAS loci triggered by miR161.1, miR161.2, miR400 and tasiRNAs produced by miR173-TAS1/2 [3, 26]. PPR PHAS loci have been found in 9 eudicots, and triggered by miR7122, miR1509 and *fve*-PPRtri1/2 [3]. MYB

transcription factor are target by miR159/miR828/miR858 in apple, peach and lotus, MYB-derived phasiRNAs could target a variety of genes with different function [27, 29, 30]. NB-LRR is the largest family targeted by small RNA. Arikiti et al [25] found 500 PHAS loci in soybean, amazingly, 208 loci are NB-LRR genes; Zhai et al [21] found 114 PHAS loci in Medicago, among them, 79 loci are NB-LRR PHAS loci and 74 NB-LRR loci are targeted by miR1507, miR2109 and miR2118. NB-LRR PHAS loci are found in many plant species, indicating a conserved role of phasiRNA in regulation of NB-LRR [29, 31]. NB-LRR can be targeted by both miRNAs and phasiRNAs [21]. Besides these three gene families, AP2-like gene was found as PHAS loci triggered by miR156 and miR172 in Arabidopsis, DCL2 and Suppressor of Gene Silencing 3 (*SGS3*) were designed as PHAS loci triggered by miR1515 and miR2118, respectively, in Medicago and soybean [21]. *TAS3* derived tasiARFs are the only phasiRNAs that have been validated to target ARF genes in *trans* [16, 17, 28]. The functions of most phasiRNAs are still largely unknown [3].

To investigate phasiRNAs in *Panax notoginseng*, a precious traditional Chinese herb, we analyzed 24 small RNA profiles of *P. notoginseng*, and identified 204 PHAS loci generating 21 nt phasiRNAs, and 90 PHAS loci generating 24 nt phasiRNAs. Via integrating miRNAs and parallel analysis of RNA end (PARE) data, we found that some PPR genes generate phasiRNAs, moreover, we found these phasiRNAs could target other PPR loci in *trans* manner. Using RLM 5'-RACE technique, we further validated the cleavage on PPR transcripts induced by phasiRNAs. We present direct evidence that phasiRNAs originated from PPR genes are functional to regulate other PPR genes in *trans*. Our results also suggest that miR171, with 21 nt, could trigger phasiRNA generation in wide type plants, although existing study found that miR171 with additional nucleotide at 3' end could trigger phasiRNAs in a mutant.

Methods

Materials and small RNA sequencing profiles

We sequenced 7 small RNA sequencing profiles of *P. notoginseng* grown in Wenshan County, Yunnan, China (Additional file 1: Table S1). Two of these 7 samples was mixed RNA sample of the leaf and root. Other 5 samples were RNA samples extracted from roots of 5 different plants. The tissue samples collected were stored in liquid nitrogen immediately. The samples were kept at -80 °C until RNAs were extracted. Total RNAs were extracted from samples using the TRIzol reagent (Invitrogen, Thermo Fisher Scientific Inc., USA) according to the manufacturer's protocol. The integrities of the RNAs were checked using an ultraviolet spectrophotometer (Hoefer, MA, USA), based on the ratio of the optical density at 260 nm to that at 280 nm (OD₂₆₀/OD₂₈₀) and were also

assessed by electrophoresis in a denaturing formaldehyde agarose gel, based on visual comparison with the strength of 18S and 28S ribosomal RNAs. The RNA samples with OD_{260/280} between 1.8 and 2.0 were checked for the total quantities. Only samples with at least 20 μ g were chosen for preparation of sRNA sequencing libraries. The small RNAs of the samples were isolated from total RNAs and were sequenced using Illumina HiSeq 2000 sequencer. The 7 obtained small RNA sequencing profiles were deposited into the NCBI GEO database and are accessible with series ID GSE98118.

Because each plant sample may only express some parts of the whole gene set, it is very helpful to identify PHAS loci as completely as possible by analyzing sRNA profiles from multiple samples. Therefore, the 17 small RNA sequencing profiles downloaded from the NCBI SRA database with accession ID SRP082250 and the 7 profiles obtained in this study were combined to get unique sRNA sequences and normalized frequencies in each of the profiles.

Computational steps for identifying PHAS loci

We aligned the unique sequences in the small RNA libraries to the cDNA sequences of *P. notoginseng* (Zheng and Cui, unpublished) with SOAP2 [32]. Then, we examined the distribution of unique 21 and 24 nt sRNAs on cDNA sequences with a window of 210 nt or 240 nt (ten 21 nt or 24 nt) respectively. The positions of sRNAs on the anti-sense strand were considered in the same phase as sRNAs on the sense strand with a two-nucleotide positive offset because sRNA duplex have two-nucleotide over-hang at the 3'-end [21, 26, 28, 33]. Then, the *P*-values of windows under consideration were calculated with a Hypergeometric distribution as proposed previously [33],

$$P(X = k) = \sum_{X=k}^m \frac{\binom{20m}{n-k} \binom{m}{k}}{\binom{21m}{n}}, \quad (1)$$

where *k* was the number of phased unique 21 nt (or 24 nt) sRNAs in the window, *n* was the total number of unique 21 nt (or 24 nt) sRNAs in the window, and *m* was the number of phases. In this study, *m* was set to 10.

We set the window sizes of PHAS loci according to the sizes of reported PHAS loci and the distances between the complementary sites of miRNA triggers to the phasiRNAs in other species. As reported previously [16, 17, 29], the functional tasiRNAs derived from TAS3 are at the seventh or eighth 21 nt segments from the 3' complementary site of miR390. TAS4 derived functional and conserved tasiRNA, TAS4-siR81(-), is also very close to the miR828 complementary site [23, 29]. For other PHAS loci and TAS1/2 in *Arabidopsis*, it has not been reported that these PHAS loci encode conserved phasiRNAs. But

the phasiRNAs from these PHAS and TAS1/2 were very close to complementary sites of their miRNA triggers [17, 21]. Therefore, we set the window sizes of PHAS loci as ten 21 or 24 nt segments.

We calculated a phase score for each position of cDNA sequences using the method in [34]. For a window with at least three phased unique sRNAs, i.e., when $k \geq 3$,

$$PhaseScore = \ln \left(1 + 10 \times \frac{\sum_{i=1}^m P_i}{1 + \sum_{i=1}^m U_i} \right)^{k-2}, \quad (2)$$

where *m* was the number of phases in the window, *k* was the number of unique phased siRNAs in the window, *U_i* was the number of non-phased reads at the *i*th phase from the position, and *P_i* was the number of phased reads at the *i*th phase from the position. *m* was 10 in this study.

If a window has a *P*-value less than 0.05, it was extended 100bp at both 5'- and 3'-ends. Then we merged the overlapped windows. We next calculated the false discovery rates with the *P*-values of the merged windows with the method in [35]. We selected the merged windows with a maximal phase scores of 5 or more and multiple test corrected *P*-values less than 0.05 as putative PHAS loci. We named the predicted PHAS loci with its transcript ID and an increasing serial number. We reported the phased siRNAs of the predicted PHAS loci as phasiRNAs. We named the phasiRNAs of a PHAS loci by adding siR and a serial number to the name of the PHAS loci.

Degradome sequencing

The degradome of a *P. notoginseng* root sample in our previous study was retrieved from NCBI SRA database using the accession number SRP087606. Briefly, the root of a *P. notoginseng* plants grown in a shading greenhouse at Kunming University of Science and Technology (24° 51' 0" N, 102° 52' 2" E, altitude 1835 m), Kunming, Yunnan, China, was frozen in liquid nitrogen immediately after harvesting. During the experiment, the daily average temperature was 25 °C, the daily maximum difference in temperature was 10 °C and humidity was 60-80%. The total RNA from the root of a *P. notoginseng* plant was extracted using the TRIzol reagent (Invitrogen, Thermo Fisher Scientific Inc., USA) according to the manufacturer's protocol. The integrity of the RNA was checked with an ultraviolet spectrophotometry (Hoefer, MA, USA) and 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). The degradome of polyadenylated transcripts was sequenced using Illumina HiSeq 4000 sequencer. The obtained degradome sequencing profile was filtered to remove low quality reads that have low scored nucleotides (< 20). Then, the 3' adapters in the remaining reads were removed.

Identifying miRNA complementary sites on PHAS and phasiRNA targets

The degradome sequencing profile were downloaded from the NCBI SRA database with the accession number SRP087606. The raw degradome sequencing profile was filtered to remove low quality reads that have low scored nucleotides (< 20). Then, the 3' adapters in the remaining reads were removed. The unique sequences were obtained and the frequencies of the unique sequences were calculated. The SeqTar algorithm [36] was used to predict miRNA complementary sites on the original transcripts of PHAS loci. For conserved miRNAs, the targets that have at least one valid read, i.e., read started at the 9th to 11th positions of a miRNA binding site (as defined in [36]), or targets that have less than 4 mismatches were used for further analysis.

We identified the targets of phasiRNAs with the SeqTar algorithm [36]. Only targets that have at least one valid read and less than 4 mismatches were used for further analysis.

RLM 5'-RACE validations of some phasiRNA targets

Modified 5'-RACE assay was performed using the GeneRacer Kit (Invitrogen) to validate the predicted targets. Briefly, 5 μ g total RNA was ligated with GeneRacerTM RNA Oligo (5'-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3') and reverse transcription was performed using SuperScriptTM III Reverse Transcriptase and oligo dT(18) primer. The resulting cDNA was used as template for PCR amplification with GeneRacer 5' primer (5'-CGACTGGAGCACGAGGACACTGA-3') and a gene-specific reverse primer. A second nested PCR was performed using GeneRacer 5' nested primer (5'-GGACACTGCATGGACTGAAGGAGTA-3') and a gene specific nested primer. The amplified products were run on a 2% agarose gel, bands with expected size were

purified using QIAquick[®] Gel Extraction Kit (QIAGEN) and ligated onto T Vector pMD 19 (Simple)(Takara), following transformation and clone PCR, plasmids were isolated and subjected to Sanger sequencing. Gene specific primers used are listed in Additional file 1: Table S2.

Results and discussion

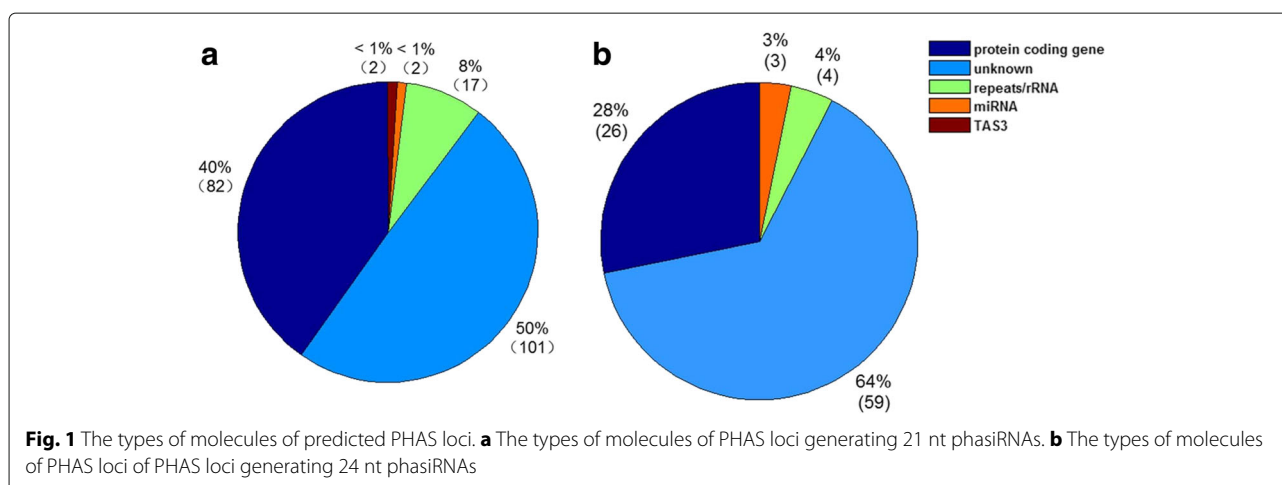
Samples and small RNA profiles of *P. notoginseng*

We collected 5 root samples and 2 whole plant samples of *P. notoginseng* grown in Wenshan County, Yunnan, China. The small RNAs of the samples were isolated from total RNAs and were sequenced using Illumina HiSeq 2000 sequencer (see Methods). The obtained sRNA profiles were combined with 17 root sRNA profiles from NCBI SRA database obtained with accession number SRP082250 to get unique sRNA sequences and their frequencies in each of the profiles.

Twenty-one nt PHAS loci in *P. notoginseng*

The combined sRNA profiles of 23 samples were aligned to a self-assembled transcriptome database of *P. notoginseng* using SOAP2 [32]. Then, we used a self-written program to predict PHAS loci and phasiRNAs from the alignment result of SOAP2 based on methods proposed previously [33, 34] (see Materials and methods for details). Then we merged overlapped PHAS loci. As listed in Additional file 1: Table S3, we totally identified 204 PHAS loci encoding 21 nt phasiRNAs with a combined criterion of a phase score ≥ 5 and a multiple-test corrected $P < 0.05$. These PHAS loci produced at least 2505 twenty-one nt phasiRNAs that were sequenced in the sRNA profiles used in the study (Additional file 1: Table S4).

The predicted PHAS were annotated by aligning their sequences to the NCBI Nucleotide Collection (nr/nt) database and the TIGR Plant Repeat database (details are given in Additional file 1: Table S3). 50% of PHAS



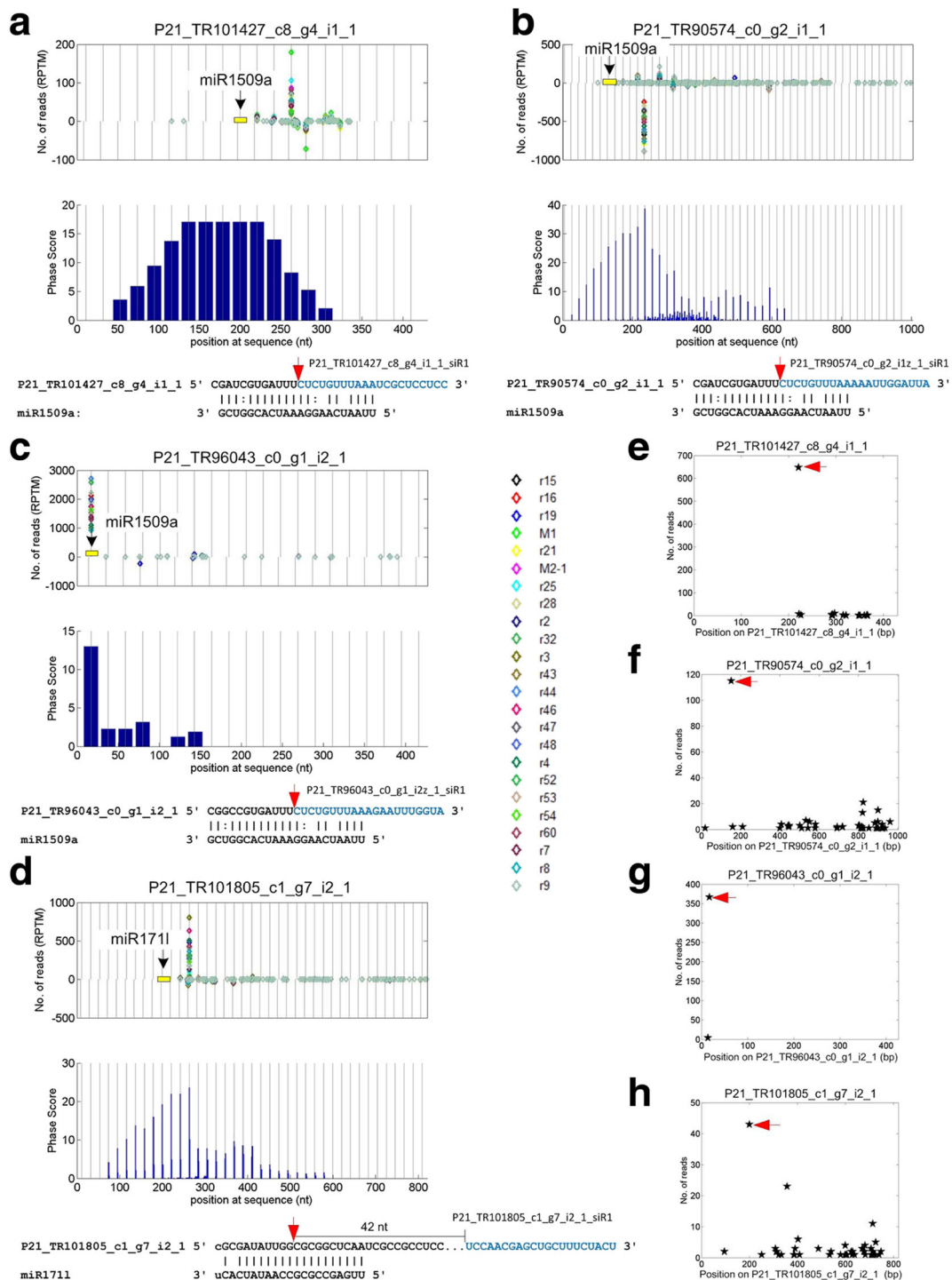
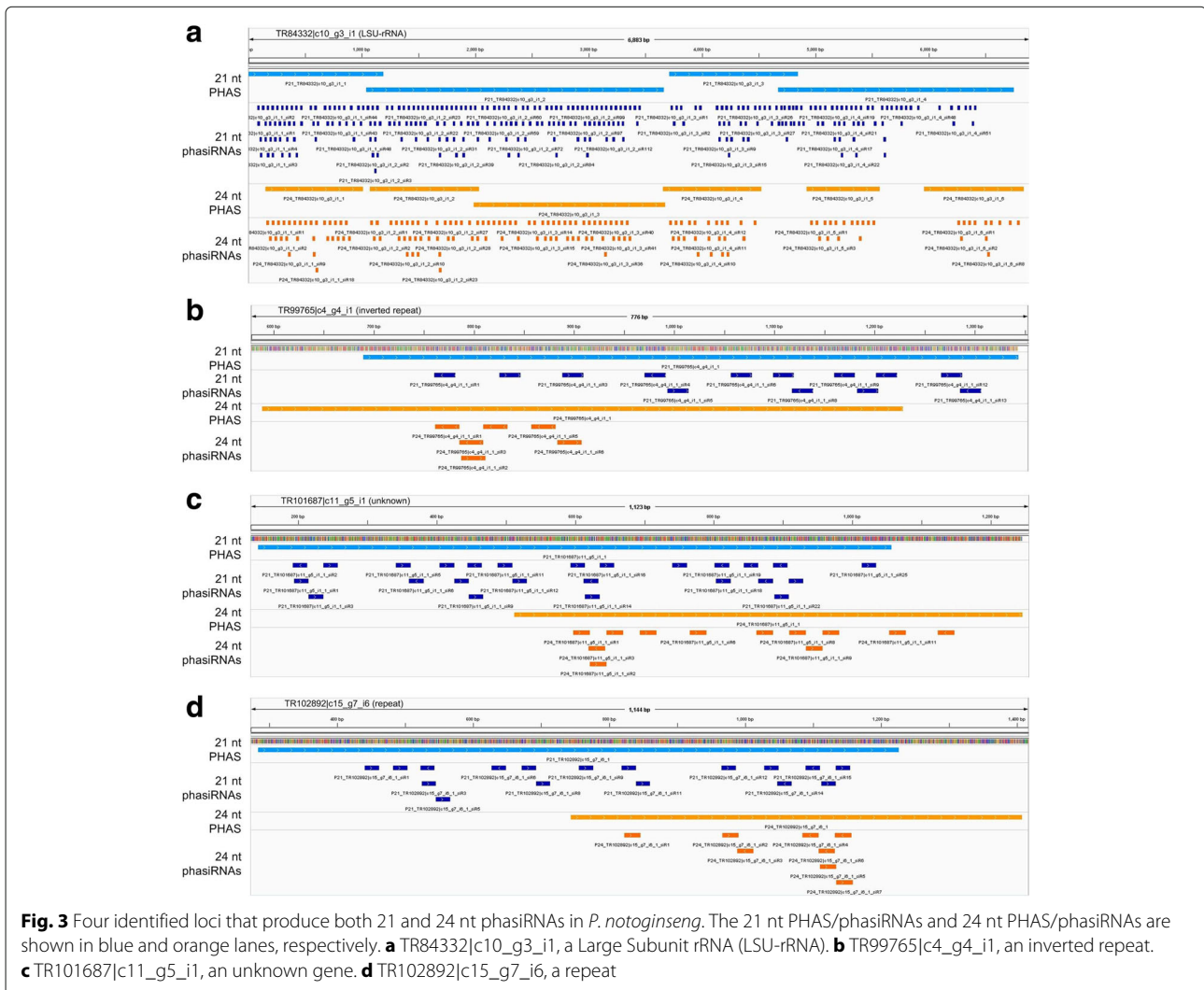


Fig. 2 The distributions of sRNA and degradome reads, and phase scores of four 21 nt PHAS loci identified. From Part (a) to (d), the phased positions from the position with the highest phase scores of the PHAS loci were marked with the vertical gray lines. The miRNA complementary sites were marked as the yellow boxes in the read distribution panel. MiRNAs complement to the sense and anti-sense strand of the predicted PHAS loci were pointed by miRNAs from above and under zero read line, respectively. The predicted miRNA binding sites were illustrated under the panels of phase scores. **a** P21_TR101427_c8_g4_i1_1 that is targeted by miR1509a. **b** P21_TR90574_c0_g2_i1_1 that is targeted by miR1509a. **c** P21_TR96043_c0_g1_i2_1 that is targeted by miR1509a. **d** P21_TR101805_c1_g7_i2_1 that is targeted by miR1711. **e** to **h** are the distributions of degradome reads for PHAS loci in Part (a) to (d), respectively. In Part (e) to (h), the arrows correspond to the positions pointed by the arrows of the same colors in the lower panels of Part (a) to (d), respectively

Table 1 The miRNA triggers that are reported to target PHAS loci

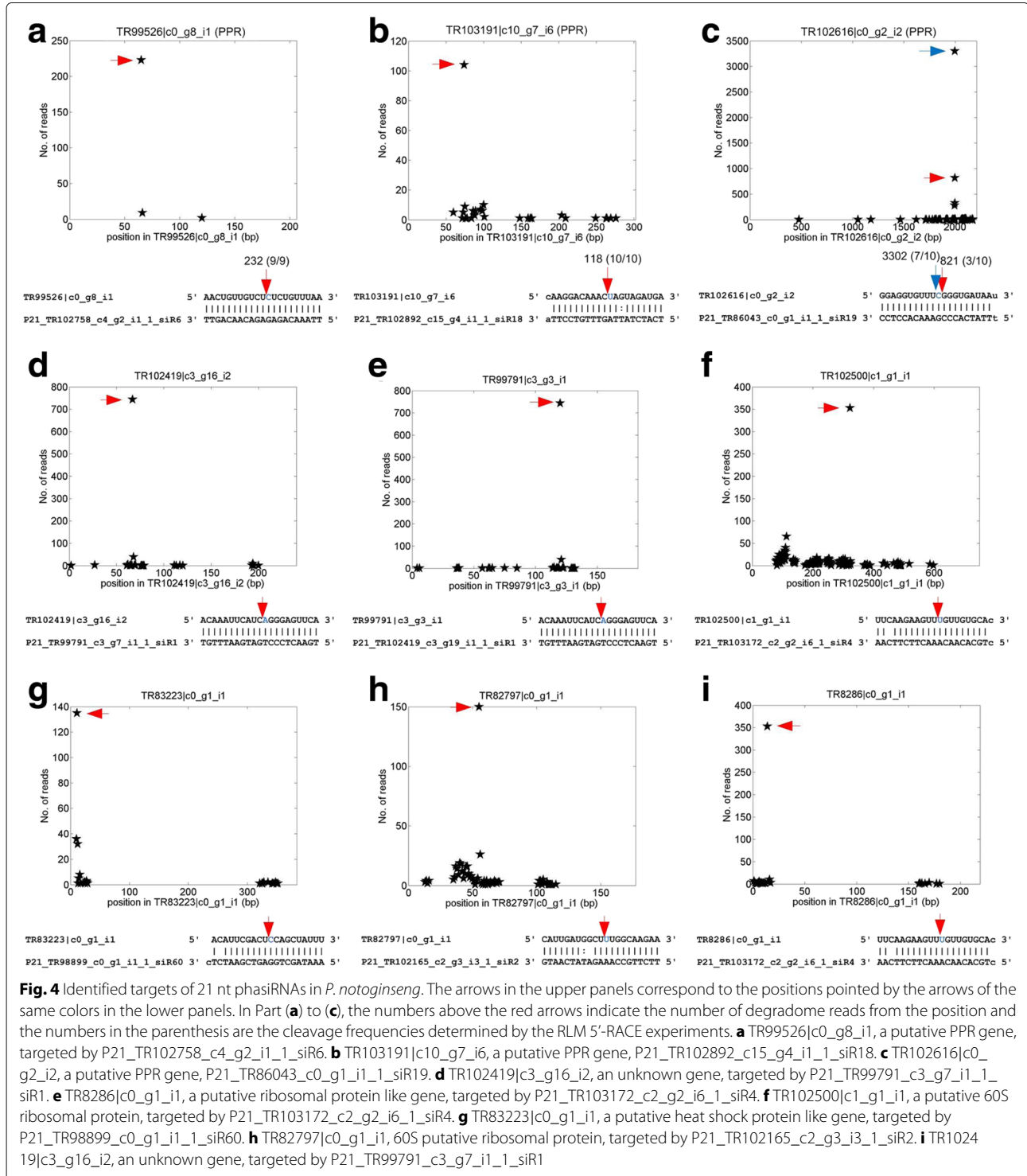
miRNA	Length (nt)	PHAS Loci	Species	References
miR161.1/2, miR400	21	PPR	<i>Arabidopsis</i>	[26]
miR171	21	SCL	<i>P. notoginseng</i>	[37] and this study
miR172	21	AP2	<i>Medicago</i>	[21]
miR173	22	TAS1/2	<i>Arabidopsis</i>	[17]
miR390	21	TAS3	land plants	[16, 17] and this study
miR828	22	TAS4	<i>Arabidopsis</i> , lotus	[23, 29]
miR828	22	MYB	apple, peach, lotus	[27–30]
miR1507, miR1515	22	DCL2	<i>Medicago</i> , soybean	[21]
miR2118, miR482, miR1507, miR2109	22	NB-LRR	legume, cotton, lotus, tomato, <i>P. notoginseng</i>	[20, 21, 29, 31, 39, 40] and this study
miR2118, miR2275	22	lncRNAs	rice, maize	[3, 20, 24]
miR5754	22	Protein kinase	<i>Medicago</i>	[21]
miR7122, miR1509, miR173	22	PPR	eudicots	[21, 26, 28] and this study



loci encoding 21 nt phasiRNAs are unknown genes (see Fig. 1a). The second largest type of PHAS loci encoding 21 nt phasiRNAs is protein coding genes, with 40% or 82 loci. Seventeen PHAS loci encoding 21 nt phasiRNAs were originated from rRNA or repeats. Two 21 nt PHAS

loci are TAS3 genes and two PHAS loci encoding 21 nt phasiRNAs could be mapped to MIRNAs.

Because miRNAs are critical in processing pathways of phasiRNAs, we identified miRNA triggers for these PHAS loci on both strands by using the degradome sequencing



profile and the SeqTar algorithm [36]. At least two of the 21 nt PHAS loci were NB-LRR genes, and were triggered by miR2118 (Additional file 1: Table S3). Two TAS3 loci were triggered by miR390 (Additional file 1: Table S3). miR1509a triggers 16 PHAS loci (Additional file 1: Table S3), of which most are unknown genes. Three of the miR1509 targeted PHAS loci are shown in Fig. 2a to c, where it is shown that miR1509a guides the generation of the phasiRNAs in these PHAS loci. Furthermore, from the degradome profile, we found that miR1509a induces the cleavage of these PHAS transcripts from the centers of its complementary sites (Fig. 2e to g). Four more miR1509 triggered PHAS loci are shown in Additional file 2: Figure S1a to S1d. The degradome profile also clearly shows that miR1509 induces strong cleavages on these transcripts to initiate the production of phasiRNAs (Additional file 2: Figure S1e to S1h).

miR171l, with 21 nt, triggers two PHAS loci from its conserved targets in the scarecrow-like transcription factor family (Additional file 1: Table S3). One of these two PHAS loci, P21_TR101805_c1_g7_i2_1, is shown in Fig. 2d. Together with the degradome profile (Fig. 2h), it is shown that miR171l triggers the generation of phasiRNAs from the center of its complementary site on P21_TR101805_c1_g7_i2_1.

More over, miR482 also triggers phasiRNAs by targeting an unknown gene and a hypothetical protein (Additional file 2: Figures S1i and S1j, respectively).

The miRNA triggers for generation of phasiRNAs are mainly 22 nt, except the miR161, miR400, miR172 and miR390 with 21 nt (Table 1). miR161 and miR400 were only reported in *Arabidopsis* [26]. A previous study reported that miR171 with an additional 3'-U triggers production of phasiRNAs in the *hen1* mutant, however, miR171 with 21 nt could not trigger phasiRNA production in the wide type plant [37]. But our results suggest that miR171, with 21 nt, could also trigger the generation of phasiRNAs by targeting one of its conserved targets (TR101805|c1_g7_i2) from the scarecrow-like (SCL) transcription factor family in *P. notoginseng* (Fig. 2d). To exclude the possibility that miR171 may have longer isoforms, we carefully examined the reads in the small RNA profiles and found the miR171 have as many as 20 isoforms that target the P21_TR101805_c1_g7_i2_1 (Additional file 2: Figure S2). But only four are highly expressed (Additional file 2: Figure S2), among which miR171b2-3p and miR171l have no 22 nt reads in the sequencing libraries, miR171f has a 22 nt read with less than 3 copies in most libraries, and miR171a2-3p targets a site 3 nt downstream of the miR171l site shown in Fig. 2d. Thus, miR171 members with 21 nt are the triggers for production of phasiRNAs from P21_TR101805_c1_g7_i2_1.

Twenty-four nt PHAS loci in *P. notoginseng*

In addition to 21 nt PHAS loci, we also found 90 PHAS loci (Additional file 1: Table S5) that generate over six hundred 24 nt phasiRNAs (Additional file 1: Table S6). Most of them, 66%, were unknown genes (Fig. 1b). 26 PHAS loci were from the protein coding genes (Fig. 1b). Three and four 24 nt PHAS loci are MIRNAs and rRNAs/repeats, respectively (Fig. 1b). Four 24 nt PHAS loci could also generate 21 nt phasiRNAs (Fig. 3). TR84332|c10_g3_i1 in Fig. 3a is a Large Subunit rRNA (LSU-rRNA). In previous studies, we also found that rRNAs could produce both 21 nt and 24 nt phasiRNAs in lotus and pineapple [29, 38], suggesting that it is a conserved pathway to generate phasiRNAs, both 21 and 24 nt, from rRNAs. Two of the other three PHAS loci that produce 21 nt as well as 24 nt phasiRNAs were repeats (Fig. 3b and d). An unknown gene also produces both 21 and 24 nt phasiRNAs (Fig. 3c).

PhasiRNA targets in *P. notoginseng*

We used the SeqTar algorithm [36] to identify targets of phasiRNAs by using a degradome profile of the *P. notoginseng* root. By keeping targets with at least 1 valid degradome read and have less than 4 mismatches, we found 5062 targets for 1699 twenty one nt phasiRNAs (Additional file 1: Table S7). In addition to 98 *cis*-targets, 4964 of these 5062 targets are *trans*-targets (Additional file 1: Table S7).

Some of the identified phasiRNA targets were shown in Fig. 4. As in Fig. 4a, a phasiRNA produced from a PPR loci, P21_TR102758_c4_g2_i1_siR6, targets another PPR gene (TR99526|c0_g8_i1) (Additional file 1: Table S3 and Table S7). Two phasiRNAs, P21_TR102892_c15_g4_i1_1_siR18 and P21_TR86043_

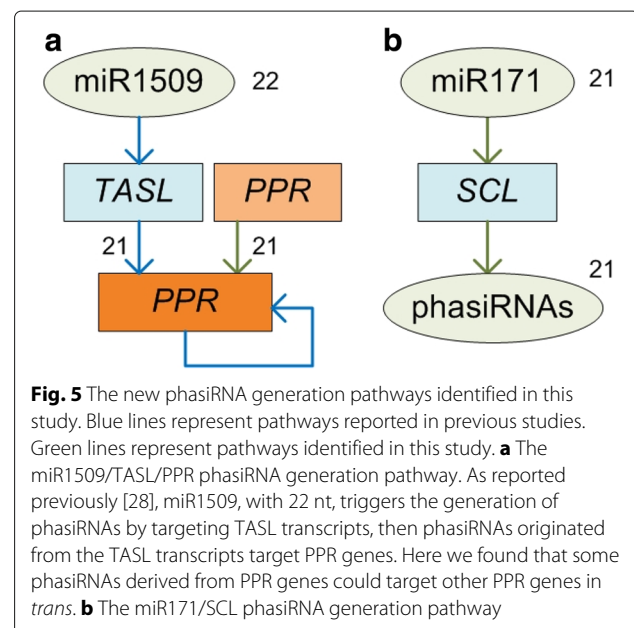


Fig. 5 The new phasiRNA generation pathways identified in this study. Blue lines represent pathways reported in previous studies. Green lines represent pathways identified in this study. **a** The miR1509/TASL/PPR phasiRNA generation pathway. As reported previously [28], miR1509, with 22 nt, triggers the generation of phasiRNAs by targeting TASL transcripts, then phasiRNAs originated from the TASL transcripts target PPR genes. Here we found that some phasiRNAs derived from PPR genes could target other PPR genes in *trans*. **b** The miR171/SCL phasiRNA generation pathway

c0_g1_i1_1_siR19, produced from two unknown genes targets two different PPR genes, TR103191|c10_g7_i6 and TR102616|c0_g2_i2, respectively (Fig. 4b and c). Both the degradome profile and our RLM 5'-RACE experiments verified that the three phasiRNAs target their PPR genes in *P. notoginseng* (Fig. 4a to c). P21_TR102892_c15_g4_i1_1 is targeted by miR1509 (Additional file 1: Table S3). It has been reported previously that miR1509 target TASL genes to produce 21 nt phasiRNAs that subsequently target PPR genes [28]. We thus guess that TR102892_c15_g4_i1 might be a TASL gene.

TR102419|c3_g16_i2 (an unknown gene), TR99791|c3_g3_i1 (a PPR gene), TR102500|c1_g1_i1 (a ribosomal protein), TR83223|c0_g1_i1 (a heat shock protein), TR82797|c0_g1_i1 (ubiquitin-60S ribosomal protein) and TR8286|c0_g1_i1 (a ribosomal protein) are also identified as targets of different phasiRNAs (Fig. 4d to i). Interestingly, TR102419|c3_g16_i2 and TR99791|c3_g3_i1 produce the same phasiRNA and may form a mutual regulation loop or self-regulation loops through phasiRNAs (Fig. 4d and e).

In soybean and *Medicago truncatula*, miR1509 indirectly initiates the generation of phasiRNAs from PPR genes by one or two-layers of TASL-tasiRNA interactions [28]. In comparison, our results suggest that in addition to miR1509/TASL/PPR pathway, some PPR genes not triggered by miR1509 could also generate phasiRNAs to repress other PPRs in *P. notoginseng*. Presumably, these PPR PHAS loci are triggered by other unknown miRNAs or siRNAs.

One hundred twenty-nine targets of sixty one 24 nt phasiRNAs were also identified and filtered with the same criteria as those for the 21 nt phasiRNAs (Additional file 1: Table S8). Thirteen and 116 of these targets are *cis*- and *trans*-targets, respectively. But base on the degradome, only a few 24 nt phasiRNAs induce clear cleavage on the targets (Additional file 1: Table S8), suggesting these 24 nt phasiRNAs might function in other means than cleaving targets in *trans*, as noticed previously [20].

Conclusion

We found 204 and 90 PHAS loci that generate 21 and 24 nt phasiRNAs, respectively, in *P. notoginseng*. We found that some 21 nt phasiRNAs generated from PPR genes could target other PPRs in *trans* (see the green line in Fig. 5a), indicating that some of the 21 nt phasiRNAs are functional. Furthermore, although most miRNA triggers of PHAS are 22 nt (Table 1), we demonstrated that miR171, with 21 nt, triggers the generation of phasiRNAs from its conserved SCL targets in wide type *P. notoginseng* (Fig. 5b). This result is different from previously report that an edited miR171 with 22 nt could trigger phasiRNAs

in a mutant plant [37]. These results provide the first set of PHAS loci and phasiRNAs in *P. notoginseng*, and enhance our understanding of PHAS in plants.

Additional files

Additional file 1: Supplementary Tables. This is an MS Excel file with multiple sheets. This file includes 8 supplementary tables. (XLS 5603 kb)

Additional file 2: Supplementary Figures. This is a pdf file. This file includes 2 supplementary figures. (PDF 1148 kb)

Funding

The publication costs of this article were supported by two grants (No. 31460295 and 31760314) of National Natural Science Foundation of China (<http://www.nsf.gov.cn/>) and a grant (No. SKLGE-1511) of the Open Research Funds of the State Key Laboratory of Genetic Engineering, Fudan University, China, to YZ; grants (No. 2013FA031 and 2014FA003) of the Science and Technology Bureau of Yunnan Province to XC; and a starting-up grant of Henan Normal University (No. 5101049470095) and Program for Innovative Research Team in Science and Technology in University of Henan Province (No. 17IRTSTHN017) to YL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The 7 obtained small RNA sequencing profiles were deposited into the NCBI GEO database and are accessible with series accession ID GSE98118.

About this supplement

This article has been published as part of *BMC Genomics* Volume 19 Supplement 1, 2018: 16th International Conference on Bioinformatics (InCoB 2017): Genomics. The full contents of the supplement are available online at <https://bmcbgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-1>.

Authors' contributions

YZ, YL, and XC conceived of and designed the research. YZ designed the computational pipeline for analyzing small RNA sequencing profiles and degradome profile. KC, LL, XZ predicted PHAS and phasiRNAs from the small RNA profiles. SR, JG, and QW predicted phasiRNA targets using the degradome profile. PL, SL and XC collected the plant samples and extracted RNAs for high throughput sequencing. KC, LL, XZ and YZ analyzed the results. YL and YY performed the RLM 5'-RACE experiments. YZ and YL wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650500, China. ²Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming 650500, China. ³Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China. ⁴College of Life Sciences, Henan Normal University, Xinxiang 453007, China. ⁵Key laboratory of *Panax notoginseng* resources sustainable development and utilization of state administration of traditional Chinese medicine, Kunming 650500, China. ⁶Provincial key laboratory of *Panax notoginseng* of Yunnan, Kunming 650500, China.

Published: 19 January 2018

References

- Chen X. Small RNAs in development—insights from plants. *Curr Opin Genet Dev.* 2012;22(4):361–7.
- Sunkar R, Li Y-F, Jagadeeswaran G. Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 2012;17(4):196–203.
- Fei Q, Xia R, Meyers BC. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell.* 2013;25(7):2400–15.
- Axtell MJ. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol.* 2013;64:137–59.
- Axtell MJ, Snyder JA, Bartel DP. Common functions for diverse small microRNAs can up-regulate translation. *Plant Cell.* 2007;19(6):1750–69.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004;116(2):281–97.
- Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. *Science.* 2007;318(5858):1931–4.
- Vasudevan S. Posttranscriptional upregulation by microRNAs. *Wiley Interdiscip Rev: RNA.* 2012;3(3):311–30.
- Xiao M, Li J, Li W, Wang Y, Wu F, Xi Y, Zhang L, Ding C, Luo H, Li Y, et al. MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol.* 2017;14(10):1326–34.
- Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol.* 2009;21(3):367–76.
- Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, Jacobsen SE. Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr Biol.* 2004;14(13):1214–20.
- Chan SW-L, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. RNA silencing genes control de novo DNA methylation. *Science.* 2004;303(5662):1336–1336.
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu J-K. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell.* 2005;123(7):1279–91.
- Han MH, Goud S, Song L, Fedoroff N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A.* 2004;101(4):1093–8.
- Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, 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 JK, author WZ, Jin H. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol.* 2012;13(3):20.
- Axtell MJ, Jan C, Rajagopalan R, Bartel DP. A two-hit trigger for siRNA biogenesis in plants. *Cell.* 2006;127(3):565–77.
- Allen E, Xie Z, Gustafson AM, Carrington JC. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell.* 2005;121(2):207–21.
- Allen E, Howell MD. miRNAs in the biogenesis of *trans*-acting siRNAs in higher plants. *Semin Cell Dev Biol.* 2010;21(8):798–804.
- Yoshikawa M, Peragine A, Park MY, Poethig RS. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* 2005;19(18):2164–75.
- Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C, et al. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J.* 2012;69(3):462–74.
- Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA, et al. MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, *trans*-acting siRNAs. *Genes Dev.* 2011;25(23):2540–53.
- Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert JL, Bartel DP, Crété P. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell.* 2004;16(1):69–79.
- Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* 2006;20(24):3407–25.
- Zhai J, Zhang H, Arikrit S, Huang K, Nan G-L, Walbot V, Meyers BC. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc Natl Acad Sci.* 2015;112(10):3146–151.
- Arikrit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdés-López O, Prince S, Musket TA, Nguyen HT, Stacey G, Meyers BC. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell.* 2014;26(12):4584–601.
- 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(3):926–42.
- Zhu H, Xia R, Zhao B, An YQ, Dardick CD, Callahan AM, Liu Z. Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biol.* 2012;12(1):149.
- Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell.* 2013;25(5):1555–72.
- Zheng Y, Wang S, Sunkar R. Genome-wide discovery and analysis of phased small interfering RNAs in chinese sacred lotus. *PLoS ONE.* 2014;9(12):113790.
- Xia R, Zhu H, An Y-q, Beers EP, Liu Z. Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol.* 2012;13(6):47.
- Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BA, Baulcombe DC. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell.* 2012;24(3):859–74.
- Li R, Yu C, Li Y, Lam TW, Yiu S-M, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics.* 2009;25(15):1966–7.
- Chen HM, Li Y-H, Wu SH. Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*. *PNAS.* 2007;104(9):3318–23.
- De Paoli E, Dorantes-Acosta A, Zhai J, Accerbi M, Jeong DH, Park S, Meyers BC, Jorgensen RA, Green PJ. Distinct extremely abundant siRNAs associated with cosuppression in petunia. *RNA.* 2009;15(11):1965–70.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
- Zheng Y, Li Y-F, Sunkar R, Zhang W. SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Res.* 2012;40(4):28–28.
- Zhai J, Zhao Y, Simon SA, Huang S, Petsch K, Arikrit S, Pillay M, Ji L, Xie M, Cao X, Yu B, Timmermans M, Yang B, Chen X, Meyers BC. Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. *Plant Cell.* 2013;25(7):2417–28.
- Zheng Y, Li T, Xu Z, Wai CM, Chen K, Zhang X, Wang S, Ji B, Ming R, Sunkar R. Identification of microRNAs, phasiRNAs and their targets in pineapple. *Trop Plant Biol.* 2016;9(3):176–86.
- Fei Q, Li P, Teng C, Meyers BC. Secondary siRNAs from *Medicago NB-LRRs* modulated via miRNA–target interactions and their abundances. *Plant J.* 2015;83(3):451–65.
- Zhu QH, Fan L, Liu Y, Xu H, Llewellyn D, Wilson I. miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS ONE.* 2013;8(12):84390.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

