Degradome Sequencing-based Identification of PhasiRNAs Biogenesis Pathways in Oryza Sativa

Lan Yu  
College of Life Science, Huzhou University

Rongkai Guo  
Shanghai Institution of Plant Physiology and Ecology, Chinese Academy of Sciences

Yeqin Jiang  
College of Life Science, Huzhou University

Xinghuo Ye  
College of Life Science, Huzhou University

Zhihong Yang  
College of Life Science, Huzhou University

Yijun Meng  
College of Life and Environmental Sciences, Hangzhou Normal University

Chaogang Shao  (shaocg@zjhu.edu.cn)  
College of Life Sciences, Huzhou University, Huzhou 313000, PR. China  https://orcid.org/0000-0003-1799-2766

Research article

Keywords: Oryza sativa, phased small interfering RNAs, precursor, degradome sequencing, regulatory network

DOI: https://doi.org/10.21203/rs.3.rs-52306/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Background: The miRNA-derived secondary phased small interfering RNAs (phasiRNAs) participate in post-transcriptional gene silencing and play important roles in various bio-processes in plants. In rice, two miRNAs, miR2118 and miR2275, were mainly responsible for the triggering of 21-nt and 24-nt phasiRNAs biogenesis, respectively. However, compared to other plant species, relatively fewer phasiRNA biogenesis pathways have been discovered in rice, which limits the comprehensive understanding of the mechanism of phasiRNA biogenesis and the miRNA derived regulatory network.

Results: In this study, we performed a systematical searching for phasiRNA biogenesis pathways in rice. As a result, five novel 21-nt and 24-nt PHAS loci and their corresponding miRNA/sRNA triggers were identified. Moreover, the expression of corresponding miRNA/sRNA triggers in wild-type and mutants were also analyzed. Recent reports discovered that, the processing of 21-nt phasiRNAs mainly dependent on OsDCL4, and OsDCL3 is necessary for biogenesis of 24-nt phasiRNAs. As a result, fourteen 21-nt and nineteen 24-nt PHAS loci candidates were passed through the filtering procedures (Additional file 1: Table S1, Additional file 2: Table S2). In this study, we perform a systematically searching for novel PHAS loci as well as their miRNA guide gene silencing in plants can be amplified by producing secondary siRNAs. In silencing amplification, the cleavage fragments of miRNAs are converted into double-stranded RNA (dsRNA) by RNA-dependent RNA polymerase 6 (RDR6) and Suppressor of Gene Silencing 3 (SGS3) and further processed by Dicer-like (DCL) proteins in different phased manners to produce different classes of phased small interfering RNAs (phasiRNAs) with specialized molecular functions [3].

Conclusions: These results substantially extended the information of phasiRNA biogenesis pathways and their regulatory function in rice.

Background

RNA silencing is a conserved mechanism that regulates various bio-processes in eukaryotes. Two types of endogenous small RNAs, small interfering RNAs (siRNAs) and microRNAs (miRNAs), are found highly abundant in higher plant [1]. In plant, miRNAs loaded on an ARGONATE (AGO) protein to form an effector complex called the RNA-induced silencing complex (RISC), which can repress gene expression by slicing target mRNAs [1]. In recent years, researchers found some miRNA-guide gene silencing in plants can be amplified by producing secondary siRNAs. In silencing amplification, the cleavage fragments of miRNAs are converted into double-stranded RNA (dsRNA) by RNA-dependent RNA polymerase 6 (RDR6) and Suppressor of Gene Silencing 3 (SGS3) and further processed by Dicer-like (DCL) proteins in different phased manners to produce different classes of phased small interfering RNAs (phasiRNAs) with specialized molecular functions [3].

Mining of novel miRNA-phasiRNA pathways and their regulatory networks in plants has attracted wide attention, because it has been demonstrated that the miRNA triggered phasiRNAs play crucial regulatory roles in regulating genome stability, plant development, and various biotic and abiotic stress responses [1, 4]. As an important economic crop, rice is one of the popular subjects for scientific investigation. Different algorithm and software tools have been employed for mining the novel miRNA-phasiRNA pathways and exploring the miRNAs’ extended regulatory networks in rice for the purpose of improving plant development and yield [5], ability of resistance to environmental stress [6], antivirus [7], etc. To data, two miRNAs, miR2118 and miR2275, were found mainly responsible for the triggering of 21-nt and 24-nt phasiRNAs biogenesis, respectively [1], and most published reports mainly focus on the in-depth investigation of miR2118-phasiRNA and miR2275-phasiRNA pathways’ regulatory functions [8, 9]. However, comparing to other plant species, relative fewer phasiRNA biogenesis pathways have been discovered in rice, which limits the comprehensive understanding of the mechanism of phasiRNA biogenesis and the miRNA derived regulatory network.

Considering the mining of phasiRNA biogenesis pathways mainly based on the high throughput sequencing (HTS) data, we believe that, with the rapid increasing amount of HTS data from different samples of rice in recent years, considerable number of novel phasiRNA biogenesis pathways could be discovered. Therefore, in order to fully understand the mechanism of phasiRNA biogenesis and the miRNA derived regulatory network in rice, it is necessary to continue the mining work of novel phasiRNA biogenesis pathways.

In our previous work, we developed an approach for mining novel phasiRNA biogenesis pathways in plant based on the degradome-supported signatures on phasiRNA precursors and the revised ta-siRNA biogenesis model [10]. In this study, we perform a systematically searching for novel PHAS loci as well as their miRNA or small RNA (sRNA) triggers in different tissues of rice. The results revealed five novel phasiRNA biogenesis pathways for producing 21-nt phasiRNA and five novel pathways for 24-nt phasiRNA production. The targets of novel phasiRNAs were further identified. These results provide substantial extend information of mirRNA/sRNA-phasiRNA regulatory network in rice.

Results

Identification of novel phasiRNA biogenesis pathways in Oryza sativa

The sRNA HTS data set from different rice samples were employed as inputs and cDNA sequences as alignment reference for searching PHAS loci capable of producing 21-nt or 24-nt phasiRNAs. As a result, fourteen 21-nt and nineteen 24-nt PHAS loci candidates passed through the filtering procedures (Additional file 1: Table S1, Additional file 2: Table S2).

Recent reports discovered that, the processing of 21-nt phasiRNAs mainly dependent on OsDCL4, and OsDCL3 is necessary for biogenesis of 24-nt phasiRNAs in rice [1]. Therefore, the 21-nt PHAS loci candidates were further verified by evaluating their phasiRNA productivity difference between wild-type and osdcl4-1 mutant. The productivity of 24-nt phasiRNA from corresponding precursors in both wild-type and osdcl3-1 mutant were analyzed for identification of 24-nt PHAS loci. Additionally, the expression of corresponding miRNA/sRNA triggers in wild-type and mutants were also analyzed.

As a result, five novel 21-nt PHAS loci along with their corresponding miRNA/sRNA triggers were identified (Fig. 1, Table 1, Additional file 3: Figure S1, Additional file 4: Figure S2). We identified that the transcripts of LOC_Os01g57968.1 and LOC_Os05g43650.1 capable of generating phasiRNAs in seedling. The transcripts of LOC_Os02g16750.1, LOC_Os04g25740.2 and LOC_Os06g30680.1 were able to produce phasiRNAs in panicle. Additionally, the LOC_Os06g30680.1-derived phasiRNAs were also found in panicle under drought stress (Additional file 3: Figure S1). Two known 21-nt PHAS loci
(LOC_Os12g42380.1 and LOC_Os12g42390.1) were also uncovered by our screening procedure (Additional file 5: Table S3), which have been identified as two parts of a long non-coding RNAs [11].

For 24-nt phasedRNA biogenesis pathways, five novel PHAS loci along with their sRNA triggers were identified (Fig. 2, Table 1, Additional file 6: Figure S4). The LOC_Os01g37325.1-derived phasedRNAs were detected in panicle, under control and drought stress condition (Fig. 2, Additional file 6: Figure S3). The generated 24-nt phasedRNAs from the transcripts of LOC_Os02g20200.1, LOC_Os02g55550.1, LOC_Os04g45834.2 and LOC_Os09g14490.1 were detected in seedling.

In addition, for all these newly found PHAS loci, only the biogenesis of LOC_Os04g25740.2-derived phasedRNAs were triggered by a known miRNA, miR2118f. The phasedRNAs generated from the other newly found PHAS loci, to our knowledge, were all triggered by novel sRNAs (Table 1).

### Table 1

**Novel PHAS loci in Oryza sativa**

| PHAS loci  | PhasiRNA production region | Small RNA trigger ID | Small RNA trigger | Small RNA trigger binding sites | PARE cutsites |
|------------|-----------------------------|----------------------|-------------------|---------------------------------|---------------|
| 21-nt PHAS loci |  |  |  |  |  |
| LOC_Os01g57968.1 | 361–1765 | OSsRNA-1 | GCUUUUUGAACUUUUCAUUU | 424–444 | 435 |
| LOC_Os02g18750.1 | 188–920 | OSsRNA-2 | UUUUUUGCCAUCUGUAACUGG | 176–197 | 188 |
| LOC_Os04g25740.1 | 1908–2159 | osa-miR2118f | UUCCUGAUGCACCCUAUCUUA | 1875–1896 | 1887 |
| LOC_Os05g43650.1 | 1494–1620 | OSsRNA-3 | GAUCAUAUAAUAAUAAUGG | 1528–1549 | 1540 |
| LOC_Os06g30680.1 | 62–208 | OSsRNA-4 | UUCCUGAGCGCGCUAUCCAU | 50–71 | 62 |
| 24-nt PHAS loci |  |  |  |  |  |
| LOC_Os01g37325.1 | 1565–1760 | OSsRNA-14 | AAAAGUAGAUGGAUGCGGAGAC | 1676–1697 | 1688 |
| LOC_Os02g20200.1 | 4856–5052 | OSsRNA-15 | UAGAUUCUGUCGAGAAGG | 4873–4894 | 4885 |
| LOC_Os05g46580.1 | 620–208 | OSsRNA-16 | AGCCAUGCUAGUCUAAGGG | 5007–5027 | 5018 |
| LOC_Os02g55550.1 | 905–1101 | OSsRNA-17 | UAGAUUCUGUCGAGAAGG | 922–943 | 934 |
| LOC_Os04g45834.2 | 1051–1307 | OSsRNA-18/19 | UAUAUAAUAUAUAUAAUAGUGUC | 1103–1124 | 1115 |
| LOC_Os09g14490.1 | 4585–4757 | OSsRNA-20 | UGAAUGCUACGGAGGAAGG | 4578–4599 | 4590 |

#### Analysis of the regulatory function of novel phasedRNAs generated from 21-nt PHAS loci

21-nt phasedRNAs have been revealed function in trans-regulation of target genes by cleaving mRNAs in plant, these phasedRNAs were named as trans-action siRNAs (ta-siRNAs).

In order to identify novel ta-siRNAs generated from the newly found 21-nt PHAS loci, all the 21-nt phasedRNAs were systematically "predicted" by computer based on the modified model of ta-siRNA biogenesis [12]. All the detectable phasedRNAs were then employed for target prediction based on miRU algorithm and verified by using degradome HTS data (see details in "materials and methods"). As a result, we discovered ten novel ta-siRNAs which generated from three newly found 21nt PHAS loci (LOC_Os02g18750.1, LOC_Os05g43650.1 and LOC_Os06g30680.1), respectively, they mediated forty sRNA-target interactions (Table 2, Fig. 3, Additional file 8: Figure S5). Some targets of these ta-siRNAs were found playing important roles in plant cellular signaling cascades (LOC_Os02g39380.1) [13]. Some targets were involved in plant growth and development (LOC_Os01g37325.1, LOC_Os02g20200.1, LOC_Os04g45834.2, LOC_Os05g43650.1, LOC_Os06g30680.1, LOC_Os09g14490.1, LOC_Os06g47850.1, LOC_Os11g41860.1, LOC_Os11g41860.2 and LOC_Os05g46580.1) [14–19]. And some targets related to plant defense and stress response (LOC_Os09g12230.1, LOC_Os04g38450.1 and LOC_Os04g49160.1) [20–22].

Although, the transcript of LOC_Os12g42380.1 has been identified as part of an lncRNA phasedRNA precursor [11], one novel LOC_Os12g42380.1-derived ta-siRNAs was found based on our revised ta-siRNA biogenesis model [12]. LOC_Os12g42380.1 (1414)21 5'D7(+) targeted to a NAD dependent epimerase/dehydratase gene (LOC_Os07g47700.1) (Table 1, Addition file 8: figure S5), suggesting it might be involved in plant growth, development and environmental stress [23, 24].

PhasedRNAs generated from the transcripts of LOC_Os02g18750.1, LOC_Os06g30680.1 and LOC_Os12g42380.1 were detected in panicle, and the LOC_Os05g43650.1–derived phasedRNAs were detected in seedling, which suggest the requirement of these phasedRNAs in different stages of development. Combining with the annotation information of ta-siRNA target genes (Table 2) and their function verification by searching relative references. We believed that, the OSsRNA-2-LOC_Os02g18750.1-phasedRNA, OSsRNA-3-LOC_Os05g43650.1-phasedRNA, OSsRNA-4-LOC_Os06g30680.1-phasedRNA and OSsRNA-5-LOC_Os12g42380.1-phasedRNA pathways might play crucial regulatory roles in rice growth, development.

The regulatory network of the phasedRNAs pathways that mentioned above were constructed based on the target information (Fig. 4).
| PhasiRNA ID | phasiRNA sequence | Targets | Target annotation | miRU start-ending | PARE cutsites |
|-------------|-------------------|---------|-------------------|-------------------|--------------|
| LOC_Os02g18750.1(189)21 3'D26 (+) | UGUGCCACGUCAACACCACA | LOC_Os03g40260.1 | Regulator of chromosome condensation domain containing protein | 1676–1696 | 1687 |
| LOC_Os02g18750.1(192)21 3'D25 (+) | GCGCACUGCGCGACGUGU | LOC_Os02g39380.1 | OsCML17 - Calmodulin-related calcium sensor protein | 343–363 | 354 |
| LOC_Os02g18750.1(204)21 3'D13 (+) | UCGACUUCGCCGCUCGC | LOC_Os02g39090.1 | expressed protein | 802–823 | 814 |
| LOC_Os05g43650.1(1540)21 3'D2(+) | UCAAUAUGAUGUGAAAAUG | LOC_Os01g15520.1 | expressed protein | 1248–1268 | 1259 |
| | | LOC_Os01g34620.8 | OsGrx_S15.1 - glutaredoxin subgroup II | 500–520 | 511 |
| | | LOC_Os03g50070.1 | DUF1295 domain containing protein | 1195–1215 | 1206 |
| | | LOC_Os04g38450.1 | gamma-glutamyltranspeptidase 1 precursor | 2137–2157 | 2148 |
| | | LOC_Os04g49160.1 | zinc finger, C3HC4 type domain containing protein | 1093–1113 | 1104 |
| | | LOC_Os05g03574.1 | expressed protein | 648–668 | 659 |
| | | LOC_Os06g23274.1 | zinc finger, C3HC4 type, domain containing protein | 4632–4652 | 4643 |
| | | LOC_Os06g47850.1 | zinc finger family protein | 97–117 | 108 |
| | | LOC_Os08g19114.1 | expressed protein | 2050–2070 | 2061 |
| | | LOC_Os08g40440.1 | dihydroflavonol-4-reductase | 1315–1335 | 1326 |
| | | LOC_Os09g12230.1 | ubiquitin-conjugating enzyme | 1021–1041 | 1032 |
| | | LOC_Os09g27500.1 | cytochrome P450 | 1714–1734 | 1725 |
| | | LOC_Os11g41860.1 | OsFBX429 - F-box domain containing protein | 1030–1050 | 1041 |
| | | LOC_Os11g41860.2 | OsFBX429 - F-box domain containing protein | 973–993 | 984 |
| | | LOC_Os12g12950.1 | expressed protein | 1071–1091 | 1082 |
| LOC_Os05g43650.1(1540)21 3'D2(+) | UUUUCACAUUCUAUUGAUG | LOC_Os02g45650.1 | peptidase | 1760–1780 | 1771 |
| LOC_Os05g43650.1(1542)21 3'D1(+) | AUAUGAAUCAGACAUAUAUAU | LOC_Os02g05810.1 | expressed protein | 1330–1350 | 1341 |
| | | LOC_Os02g05810.2 | expressed protein | 1324–1344 | 1335 |
| | | LOC_Os02g52900.2 | glutaredoxin 2 | 2034–2054 | 2045 |
| | | LOC_Os02g53000.2 | lysM domain-containing GPI-anchored protein precursor | 1340–1360 | 1351 |
| | | LOC_Os04g44590.1 | expressed protein | 651–671 | 662 |
| | | LOC_Os04g44590.5 | expressed protein | 445–465 | 456 |
| | | LOC_Os05g41190.1 | expressed protein | 1026–1046 | 1037 |
| | | LOC_Os05g41190.2 | expressed protein | 1082–1102 | 1093 |
Analysis of the RNA directed DNA methylation (RdDM) regulated promoters of novel 24-nt phasiRNAs

RdDM is an important regulatory event involved in repressive epigenetic modifications that can trigger transcriptional gene silencing. In order to analyze the novel 24-nt phasiRNA mediated RdDM in rice, all the known promoter sequences in rice were employed for scanning of the target sites of novel phasiRNAs generated from the newly found five 24-nt PHAS loci. The DNA methylation status of target promoter was further analyzed by utilizing the bisulfite-seq data sets of panicle (GSM4232038) and root (GSM4232039) of rice. As a result, a promoter of LOC_Os02g40860.1 gene was found targeted by five -LOC_Os01g37325.1-derived phasiRNAs (Table 3).

As CG and CHG methylation contexts are maintained by DNA methyltransferases and histone modifications, while CHH methylation is associated with 24-nt siRNA guided RdDM[25]. The CHH methylation status of promoter was found significantly higher in panicle than that in other tissues (Fig. 5), which consistent with the finding that LOC_Os01g37325.1-derived phasiRNAs only expressed in panicle tissue (Additional file 2: Table S2). Therefore, the results suggesting a methylation mediated transcriptional silencing of the promoter of LOC_Os02g40860.1.

LOC_Os02g40860.1 encodes a Casein kinase I1 (OsCKI1) protein, which belongs to the CKIs protein family. CKIs have been identified highly conserved in eukaryotes, they are believed involving in a variety of important biological events, since they have a wide substrate specificity in vitro[26]. The expression level of LOC_Os02g40860.1 in root and panicle tissues of rice have been analyzed by utilizing the RNA-seq libraries (panicle: GSM4230036 and GSM4230037; root: GSM4230038 and GSM4230039) which contributed by Zhao et al[27]. As shown in Fig. 5C, the expression level of LOC_Os02g40860.1 was significantly lower in panicle than that in root, therefore we speculated that, the biogenesis of LOC_Os01g37325.1-derived phasiRNAs are specific required in panicle for the regulation of the transcriptional level of LOC_Os02g40860.1. In another word, the OSsRNA-14-LOC_Os01g37325.1-phasiRNA pathway might involve in the development of panicle in rice.
Methods

We performed degradome-based screening of novel phasiRNA biogenesis pathways in rice. Five novel 21-nt phasiRNA biogenesis pathways and five novel 24-nt phasiRNA biogenesis pathways were identified. In addition, two known 21-nt phasiRNA biogenesis pathways were also identified. Further analysis on the targets of the detectable novel phasiRNAs with 21-nt and 22-nt length revealed total eleven novel phasiRNAs involving in forty-one siRNA–target interactions and suggest these phasiRNAs might play important roles in rice growth and development (Table 1, Additional file 8: Figure S5). These results demonstrated the effectiveness of degradome-based screening in mining novel phasiRNA biogenesis pathways and substantially extend the information of phasiRNA biogenesis pathways in rice. We believed that, more novel phasiRNA biogenesis pathways might be identified if extend our approach to other plant species.

Discussion

In recent years, researches on *Oryza sativa* have shown that 21- or 24-nt phasiRNAs distribute in genomic clusters and preferentially accumulate in reproductive tissues [28]. To date, considerable PHAS loci have been discovered in rice[29]. According to the previous reports, the miR2118 and miR2275 triggered 21-nt and 24-nt phasiRNA biogenesis pathways seem to be the primary phasiRNA production sources in rice. However, considering there are rich sources for phasiRNA production in other plant species, we speculated that the miRNA-phasiRNA pathways have not been fully discovered in rice, it is necessary to continue the mining work for better understanding the mechanism of phasiRNA biogenesis and the miRNA derived regulatory network. In this study, we performed a degradome-based screening of novel PHAS loci in rice by utilizing sRNA libraries from different tissues. As a result, five novel 21-nt PHAS loci and five novel 24-nt PHAS loci were identified, and only one of these PHAS loci was triggered by known miRNA, miR2118, which confirmed our suspicion. In addition, the novel 21-nt PHAS loci, LOC_Os05g43650.1, is a miniature inverted-repeat transposable element (MITE) gene [30], and two 24-nt PHAS loci, LOC_Os01g37325.1 and LOC_Os02g20200.1, are two retrotransposon genes, which indicated that the transcripts of transposons and retrotransposons are capable of producing secondary siRNAs. Same phenomenon has also been reported by Creasey et al. in *Arabidopsis* [31, 32]. Transposons and retrotransposons are ubiquitous in plants and play important roles in plant gene and genome evolution [33]. According to the information of targets of phasiRNAs, the OsRNA-3-LOC_Os05g43650.1-phasiRNA and OsRNA-14-LOC_Os01g37325.1-phasiRNA pathways are required for the development of rice. We speculated that, the transcripts of transposon and retrotransposon might also function as important sources of phasiRNA in plants. Further exploration of this kind of phasiRNA biogenesis pathways could benefits the in-depth investigation of their biogenesis mechanism and the miRNA/sRNA directed regulatory networks in plants.

Recently, some protein-coding genes were identified as phasiRNA precursors, which draws a lot of attention [29, 34]. Among the newly identified PHAS loci, LOC_Os04g45834.2 encoding a DUF584 domain containing protein. These protein family has been proved that involved in leaf senescence in plant [35]. LOC_Os09g14490.1 encoding a TIR-NBS type disease resistance protein, which has been identified in resistance to multiple viruses in plant [36–38]. LOC_Os02g55550.1 encoding a F-box/LRR-repeat protein 14, the LRR-repeat protein is involved in plant immune response [39]. These genes have been proved to play important roles in plants, however, their capability of producing secondary phasiRNAs suggest they might be involved in much more complex function than we expected. Systematically investigation of the temporal and spatial expression specificity of phasiRNAs generated from the transcripts of protein-coding genes in our future work might gain insight into these phasiRNAs biogenesis requirement mechanism.

In this study, besides those identified PHAS loci and other 21-nt PHAS loci candidates (table S1a and S1b), two cDNA sequences, LOC_Os09g00999.1 and LOC_Os09g01000.1, which were able to produce plenty of Dicer-independent secondary siRNAs in almost every tissues of rice have attracted our attention. We further employed the phasiRNAs generated from LOC_Os09g00999.1 and LOC_Os09g01000.1 for target prediction and identification. The results indicated plenty of siRNA-target interaction pairs were discovered (data not shown). We speculated that this might be a novel pattern of secondary siRNAs biogenesis pathways. Further investigation of Dicer-independent secondary siRNAs biogenesis pathways in plant could provide more strong evidence of this biogenesis pattern, which also provide meaningful information of the small RNA regulatory mechanism in plant.

Conclusions

We performed degradome-based screening of novel phasiRNA biogenesis pathways in rice. Five novel 21-nt phasiRNA biogenesis pathways and five novel 24-nt phasiRNA biogenesis pathways were identified. In addition, two known 21-nt phasiRNA biogenesis pathways were also identified. Further analysis on the targets of the detectable novel phasiRNAs with 21-nt and 22-nt length revealed total eleven novel phasiRNAs involving in forty-one siRNA–target interactions and suggest these phasiRNAs might play important roles in rice growth and development (Table 1, Additional file 8: Figure S5). These results demonstrated the effectiveness of degradome-based screening in mining novel phasiRNA biogenesis pathways and substantially extend the information of phasiRNA biogenesis pathways in rice. We believed that, more novel phasiRNA biogenesis pathways might be identified if extend our approach to other plant species.

Methods

Data source
The *Oryza sativa* small RNA high-throughput sequencing (sRNA HTS) data sets of seedling, root, shoot and panicle samples under control and stress conditions, the sRNA HTS data sets of wild type, *osdc14* and *osdc13* mutants and the degradome sequencing data sets (Additional file 5: Table S3) were retrieved from GEO (Gene Expression Omnibus; http://www.ncbi.nlm.nih.gov/geo/). The cDNAs, full-length genomic sequences of *Oryza sativa* were retrieved from PlantGDB (http://plantgdb.org/XGDB/phplib/). The promoter sequence of *Oryza sativa* were retrieved from PlantProm DB (http://linux1.softberry.com/). All the high-throughput sequencing data were pre-processed before use, the data of each library was normalized in RPM (reads per million) as described in our previous report[40].

**Identification of phasiRNA biogenesis pathways in Oryza sativa**

The phasiRNA loci identification criteria were established based on the revised trans-acting siRNA (ta-siRNA) ta-siRNA biogenesis model as we reported previously [12]. The screening of PHAS loci in rice was followed by four steps: (1) cDNA-genome sequences-derived 21-nt phased duplexes were computational predicted by “phase processing”, each of these duplexes has a 2nt overhang at 3’-end. (2) Each of these duplexes was separated into two increments and used for matching with small RNAs from small RNA high throughput sequencing data set of Rice seedling. A potential phasiRNA production region shall contains at least 5 tandem “processing” duplexes and each of these duplexes shall contains detectable phasiRNA from sense strand (plus siRNA) and/or antisense strand (minus siRNA). (3) Degradome high throughput sequencing (HTS) libraries were employed for systematically scanning the degradome-supported cleavage signatures on the screened possible phasiRNA production regions as we described in our previous work[12], and maintain the PHAS loci candidates with cleavage signatures which located in the phasiRNA production region. (4) The sRNAs bound to the PHAS loci were analyzed by using miRU algorithm[41], and the sRNA cleavage sites on those loci were further verified by using degradome sequencing libraries. The degradome-supported cleavage site of a sRNA trigger shall reside within 10 to 11-nt from the 5’ end of the binding site[42].

Phasing scores of phasiRNA regions were calculated based on the formula which contributed by Zheng et al [29]: Phasing score

\[
\text{Phasing score} = \log \left( 1 + 10^{\frac{N}{1 + U}} \right)^{w-2}
\]

where \(N\) represents the number of phase register occupied by at least one unique 21-nt/24-nt small RNA within a five-phase register window, \(p\) represents the total number of reads for all 21-nt/24-nt small RNA falling into a given phase in a given window, \(U\) represents the total number of unique reads for all 21-nt/24-nt small RNA falling out of a given phase.

**Identification of phasiRNA-target interaction based on degradome sequencing**

The expressed novel phasiRNAs generated from 21-nt PHAS loci were predicted based on previously modified model of ta-siRNA biogenesis in plant[12]. The predicted phasiRNAs were recruited for target prediction by using miRU with default parameters[41], and followed by degradome sequencing-based verification, as described previously[40, 43].

**Gene expression level analysis**

The sequences of each RNA-seq data set were employed for removing adapter, and then be normalized in RPM (reads per million) as we described previously [40]. The sequences were then mapped to the reference cDNA sequences, and each gene expression level was calculated by the total RPM of mapped sequences. The gene expression were profiled with the log2 transformed values of abundance with Heml [44].

**Identification of 24nt phasiRNA target**

The promoter sequences in rice were obtained from PlantProm (www.softberry.com). In order to identify the potential 24-nt phasiRNA target sites in promoter sequences, BLAST analysis was performed for finding the location of the complementary sequence of 24-nt phasiRNA with no mismatch [45]. The promoters possessed phasiRNA binding sites were remained as potential target promoters. As each of the downloaded promoter sequence containing partial mRNA sequence, we identified the corresponding potential target genes by mapping the partial mRNA sequence to cDNA sequences. The DNA methylation status of potential target promoters were analyzed by utilizing the bisulfite-seq data sets of panicle (GSM4232038) and root (GSM4232039) of rice. The expression specificity of phasiRNA in different tissues should consistent with the occurring of increasing methylation of the target promoter.

The DNA methylation analysis of promoters were performed according to the method of Zhao et al [27]. The sequences of bisulfite sequencing libraries was processed by adapter cut and normalization in RPM (reads per million) as we described previously [40]. Then the sequences were mapped to the potential promoter sequences, and the uniquely mapped sequences were used for further DNA methylation level analysis. The DNA methylation level of each cytosine was obtained by calculation of the total coverage of individual cytosines in RPM.

**Abbreviations**

phasiRNAs  
phased small interfering RNAs  
siRNAs  
small interfering RNAs  
miRNAs  
microRNAs  
AGO  
ARGINATE
RISC
RNA-induced silencing complex
dsRNA
double-stranded RNA
RDR6
RNA-dependent RNA polymerase 6
SGS3
Suppressor of Gene Silencing 3
DCL
Dicer-like
phasiRNAs
phased smallinterfering RNAs
PHAS loci
phasiRNA production precursors
sRNA
small RNA
ta-siRNA
trans-acting siRNA
RNA directed DNA methylation
RdDM

**Declarations**

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no competing interest.

**Funding**
This work was supported by National Natural Science Foundation of China, under Grants [numbers: 31801102, 31771457, 31571349], Zhejiang Provincial Natural Science Foundation of China, under Grants [numbers: LQ16C060001, LY17C190001], Public Welfare Technology Application Research Project of Zhejiang Province, under Grant [number 2016C33193], and Scientific Research Fund of Zhejiang Provincial Education Department, under Grant [number Y201533314].

**Availability of data and materials**
The small RNA high-throughput sequencing data sets and degradome sequencing data sets used for the study were retrieved from GEO (Gene Expression Omnibus; [http://www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). The accession IDs include: GSM455965, GSM816687, GSM816688, GSM816689, GSM816690, GSM816691, GSM816692, GSM816693, GSM816694, GSM816695, GSM816704, GSM816705, GSM816716, GSM816717, GSM816714, GSM816715, GSM816718, GSM816719, GSM816728, GSM816729, GSM816730, GSM816731, GSM816732, GSM816733, GSM816734, GSM816735, GSM816736, GSM816737, GSM816738, GSM816739, GSM816740, GSM562942, GSM562943, GSM562944, GSM562945, GSM520638, GSM520640, GSM434596, GSM455938, GSM455939, GSM476257, GSM4232038, GSM4232039, GSM4230036, GSM4230037, GSM4230038, GSM4230039.

**Authors’ contributions**
L.Y. accomplished the identification of phasiRNA biogenesis pathways, analysis of target gene transcriptional level and analysis of target promoter methylation status; R.G. accomplished the identification of phasiRNAs’ target and construction of the regulatory networks. L.Y. and R.G. wrote the main manuscript. L.Y, R.G, Y.M. and C.S. designed the experiments. Y.J., X.Y. and Z.Y. contributed to collection and pre-treatment of the data, and preparation of supplemental files. Y.M. and C.S. revised the manuscript. All of the authors reviewed the manuscript.

**Acknowledgements**
Not applicable.

**References**
1. 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.

2. Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol. 2012;19(6):586–93.

3. Fei Q, Xia R, Meyers BC. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. Plant Cell. 2013;25(7):2400–15.

4. Baulcombe D. RNA silencing in plants. Nature. 2004;431(7006):356–63.

5. Fan Y, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, et al. PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. Proc Natl Acad Sci U S A. 2016;113(52):15144–9.

6. Kumar K, Kumar M, Kim SR, Ryu H, Cho YG. Insights into genomics of salt stress response in rice. Rice. 2013;6(1):27.

7. Cui F, Zhao W, Luo L, Kang L. Rice Responses and Resistance to Planthopper-Borne Viruses at Transcriptomic and Proteomic Levels. Curr Issues Mol Biol. 2016;19:43–52.

8. Ma W, Chen C, Liu Y, Zeng M, Meyers BC, Li J, Xia R. 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. New Phytol. 2018;217(4):1535–50.

9. Liu Y, Wang Y, Zhu QH, Fan L. Identification of phasiRNAs in wild rice (Oryza rufipogon). Plant Signal Behav. 2013;8(8):e25079.

10. Yu L, Guo R, Jiang Y, Ye X, Yang Z, Meng Y, Shao C. Identification of novel phasiRNAs loci on long non-coding RNAs in Arabidopsis thaliana. Genomics. 2019;111(6):1668–75.

11. Huang J, Wang R, Dai X, Feng J, Zhang H, Zhao PX. A microRNA biogenesis-like pathway for producing phased small interfering RNA from a long non-coding RNA in rice. J Exp Bot. 2019;70(6):1767–74.

12. Yu L, Meng Y, Shao C, Kahrizi D. Are ta-siRNAs only originated from the cleavage site of miRNA on its target RNAs and phased in 21-nt increments? Gene. 2015, 569(1):127–135.

13. Boonburapong B, Buaboocha T. Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. BMC Plant Biol. 2007;7:4.

14. Rouhier N, Couturier J, Jacquot JP. Genome-wide analysis of plant glutaredoxin systems. J Exp Bot. 2006;57(8):1685–96.

15. Ma H, Zhao J. Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (Oryza sativa L.). J Exp Bot. 2010;61(10):2647–68.

16. Li X, Shahid MQ, Xia J, Lu Z, Fang N, Wang L, Wu J, Chen Z, Liu X. Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. BMC Genom. 2017;18(1):129.

17. Dong S, Zhang J, Sun D, Liu H, Yang Q, Wang H, Chen Z, Wang J. Identification of Magnaporthe oryzae-elicited rice novel miRNAs and their targets by miRNA and degradome sequencing. Eur J Plant Pathol. 2018;151(3):18.

18. Ding Y, Qu A, Gong S, Huang S, Lv B, Zhu C. Molecular identification and analysis of Cd-responsive microRNAs in rice. J Agric Food Chem. 2013;61(47):11668–75.

19. Vranova E, Coman D, Gruissem W. Structure and dynamics of the isoprenoid pathway network. Molecular plant. 2012;5(2):318–33.

20. Zhou B, Mural RV, Chen X, Oates ME, Connor RA, Martin GB, Gough J, Zeng L. A Subset of Ubiquitin-Conjugating Enzymes Is Essential for Plant Immunity. Plant Physiol. 2017;173(2):1371–90.

21. Shaar-Moshe L, Hubner S, Peleg Z. Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. BMC Plant Biol. 2015;15:111.

22. Campo S, Baldich P, Messegue J, Lalanne E, Coca M, San Segundo B. Overexpression of a Calcium-Dependent Protein Kinase Confers Salt and Drought Tolerance in Rice by Preventing Membrane Lipid Peroxidation. Plant Physiol. 2014;165(2):688–704.

23. Zeng F, Wu X, Qiu B, Wu F, Jiang L, Zhang G. Physiological and proteomic alterations in rice (Oryza sativa L.) seedlings under hexavalent chromium stress. Planta. 2014;240(2):291–308.

24. Zhu H, Kranz RG. A nitrogen-regulated glutamine amidotransferase (GAT1_2.1) represses shoot branching in Arabidopsis. Plant physiology. 2012;160(4):1770–80.

25. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics. 2010;11(3):204–20.

26. Liu W, Xu ZH, Luo D, Xue HW. Roles of OsCKI1, a rice casein kinase I, in root development and plant hormone sensitivity. The Plant journal: for cell molecular biology. 2003;36(2):189–202.

27. Zhao L, Xie L, Zhang Q, Quyang W, Deng L, Guan P, Ma M, Li Y, Zhang Y, Xiao Q, et al. Integrative analysis of reference epigenomes in 20 rice varieties. Nature communications. 2020;11(1):2658.

28. Johnson C, Kasprzewska A, Tennesen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res. 2009;19(8):1429–40.

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

30. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol Biol Evol. 2012;29(3):1005–17.

31. Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature. 2014;508(7496):411–5.
32. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 2009;136(3):461–72.

33. Lisch D. How important are transposons for plant evolution? Nat Rev Genet. 2013;14(1):49–61.

34. Meng Y, Shao C, Ma X, Wang H. Introns targeted by plant microRNAs: a possible novel mechanism of gene regulation. Rice (N Y). 2013;6(1):8.

35. Zheng X, Jehanzeb M, Habiba, Zhang Y, Li L, Miao Y. Characterization of S40-like proteins and their roles in response to environmental cues and leaf senescence in rice. BMC Plant Biol. 2019;19(1):174.

36. Xun H, Yang X, He H, Wang M, Guo P, Wang Y, Pang J, Dong Y, Feng X, Wang S, et al. Over-expression of GmKR3, a TIR-NBS-LRR type R gene, confers resistance to multiple viruses in soybean. Plant Mol Biol. 2019;99(1–2):95–111.

37. Zhou X, Liu J, Bao S, Yang Y, Zhuang Y. Molecular Cloning and Characterization of a Wild Eggplant Solanum aculeatissimum NBS-LRR Gene, Involved in Plant Resistance to Meloidogyne incognita. International journal of molecular sciences 2018, 19(2).

38. Fan JJ, Wang P, Xu X, Liu K, Ruan YY, Zhu YS, Cui ZH, Zhang LJ. Characterization of a TIR-NBS-LRR gene associated with downy mildew resistance in grape. Genetics molecular research: GMR. 2015;14(3):7964–75.

39. Matsushima N, Miyashita H. Leucine-Rich Repeat (LRR) Domains Containing Intervening Motifs in Plants. Biomolecules. 2012;2(2):288–311.

40. Yu L, Shao C, Ye X, Meng Y, Zhou Y, Chen M. miRNA Digger: a comprehensive pipeline for genome-wide novel miRNA mining. Scientific reports. 2016;6:18901.

41. Zhang Y. miRU: an automated plant miRNA target prediction server. Nucleic acids research 2005, 33(Web Server issue):W701-704.

42. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol. 2008;18(10):758–62.

43. German MA, Luo S, Schroth G, Meyers BC, Green PJ. Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nature protocols. 2009;4(3):356–62.

44. Deng W, Wang Y, Liu Z, Cheng H, Xue Y. Heml: a toolkit for illustrating heatmaps. PloS one. 2014;9(11):e111988.

45. Baev V, Naydenov M, Apostolova E, Ivanova D, Doncheva S, Minkov I, Yahubyan G. Identification of RNA-dependent DNA-methylation regulated promoters in Arabidopsis. Plant physiology biochemistry: PPB. 2010;48(6):393–400.

Figures
Figure 1

21-nt phasiRNAs generated from novel PHAS loci in Oryza sativa. Above are the degradome supported cleavage sites of sRNA triggers on PHAS loci (Target-plot profile graphs). The x axis represents the position on PHAS loci and the y axis represents the abundance of degradome signature. Below the Target-plot profile graphs are images showing 21-nt phasiRNAs generated from transcripts of LOC_Os01g57968.1 (A) and LOC_Os05g43650.1 (D) in seedling, LOC_Os02g18750.1 (B), LOC_Os04g25740.1 (C) and LOC_Os06g30680.1 (E) in panicle. The black arrows indicate the sRNA trigger cleavage sites, the x-axis represent the phasiRNA position mapped within the PHAS loci, the y-axis represent the read abundance (in RMP, reads per million)of the small RNAs mapped to the sense and antisense strands of PHAS loci. The phasing score of 21-nt PHAS windows were profiled in bottom graphs. The x-axis represent the position of the PHAS loci, y-axis represent the phasing scores.
24-nt phasiRNAs generated from novel PHAS loci in Oryza sativa. Above are the degradome supported cleavage sites of sRNA triggers on PHAS loci (Target-plot profile graphs). The x axis represents the position on PHAS loci and the y axis represents the abundance of degradome signature. Below the Target-plot profile graphs are images showing 24-nt phasiRNAs generated from transcripts of LOC_Os01g37325.1 (A) in root, LOC_Os02g20200.1 (B), LOC_Os02g55550.1 (C), LOC_Os04g45834.2 (D) and LOC_Os09g14490.1 (E) in seedling. The black arrows indicate the sRNA trigger cleavage sites, the x-axis represent the phasiRNA position mapped within the PHAS loci, the y-axis represent the read abundance (in RMP, reads per million) of the small RNAs mapped to the sense and antisense strands of PHAS loci. The phasing score of 24-nt PHAS windows were profiled in bottom graphs. The x-axis represent the position of the PHAS loci, y-axis represent the phasing scores.
Examples of degradome sequencing-based validation of the phasiRNA-target interactions. Four libraries of degradome sequencing data libraries (GSM434596 and GSM476257) were recruited for T-plot profiling. The IDs of the target transcripts and the corresponding phasiRNAs generated from the transcript of LOC_05005g43650.1 are listed on the top. The y axis measure the normalized reads (in RMP, reads per million) of the degradome signals, and the x axis represent the position of the cleavage signals on the target transcripts. The binding sites of the phasiRNA on their target transcripts were denoted by gray horizontal lines, and the dominant cleavage signals were marked by black arrows.

Figure 3
The regulatory networks of phasiRNA pathways in Oryza sativa. The OSsRNA-2-LOC_Os02g18750.1-phasiRNA (A), OSsRNA-3-LOC_Os05g43650.1-phasiRNA (B), OSsRNA-4-LOC_Os06g30680.1-phasiRNA (C) and OSsRNA-5-LOC_Os12g42380.1-phasiRNA (D) regulatory network were constructed by Cytoscape based on the validated phasiRNAs and their targets. The phasiRNAs are the gray nodes, the orange nodes represent the targets involving plant development, stress response, disease resistance or signaling transport. The blue nodes represent the expressed proteins with unknown functions.

Figure 5
DNA methylation status and expression analysis of target promoter. DNA methylation by CG, CHG and CHH context at the promoter of LOC_Os02g40860.1 in panicle (A) and root (B) were analyzed and profiled. X-axis represents the position on promoter sequence and Y-axis represents the abundance of CG, CHG or CHH. The expression level of LOC_Os02g40860.1 in panicle and root were also showed in a bar chart (C).

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- figureS5.pdf
- figureS4.pdf
- figureS3.pdf
- TableS3.xlsx
- figureS2.pdf
- figureS1.pdf
- TableS2.xlsx
- TableS1.xlsx