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)
Huzhou University https://orcid.org/0000-0003-1799-2766

Research article

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

Posted Date: November 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-52306/v2

Version of Record: A version of this preprint was published at BMC Genomics on January 30th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07406-7.

Abstract

Background: The microRNAs(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, compare 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.

Results: In this study, we performed a systematical searching for phasiRNA biogenesis pathways in rice. As a result, five novel 21-nt phasiRNA biogenesis pathways and five novel 24-nt phasiRNA biogenesis pathways were identified. Further exploration of the regulatory function of phasiRNAs revealed eleven novel phasiRNAs with 21-nt length targeting forty-one genes, most of which involving in the growth and development of rice. In addition, five novel phasiRNAs with 24-nt length were found targeting the promoter of an *OsCKI1* gene and causing higher methylation status in panicle, implying their regulatory function of the transcription of *OsCKI1*, and subsequently affect the development of rice.

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

Background

Small RNA-mediated RNA silencing is a conserved mechanism that regulates various bioprocesses in eukaryotes [1]. Two types of endogenous small RNAs, microRNAs (miRNAs) and small interfering RNAs (siRNAs), are highly abundant in higher plant.

The biogenesis of a miRNA in plants begins with the transcription of a primary miRNA (pri-miRNA). Next, an RNase III family of DICER-LIKE (DCL) enzyme, usually DCL1, assisted by HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE), sequentially processes the pri-miRNA into a precursor (pre-miRNA) and further cut into a miRNA/miRNA* short duplex. Usually, one strand (miRNA*) of the short duplex is degraded and the mature miRNA strand is incorporated into the RNA-induced silencing complex (RISC), which is highly complementary to the target gene and subsequently leads to the cleavage of the target mRNA and towards its degradation in plants[2]. In recent years, plenty of researches showed that, miRNAs play critical roles in diverse biological processes in plants, such as growth and development, stress response and plant metabolism[3, 4]. For example, OsmiR393a and OsmiR393b regulated rice primary root elongation and adventitious roots number via auxin signaling pathway [5]. The miR398 is proposed to be directly linked to the *Arabidopsis* stress regulatory network and regulates plant responses to oxidative stress, water deficit, salt stress, abscisic acid stress, ultraviolet stress, copper and phosphate deficiency, high sucrose and bacterial infection[6].

In terms of siRNAs, their biogenesis could be triggered either endogenously from species own genome or exogenous, such as external viruses or transgenes, and the precursor of siRNAs are usually long and double-stranded [7]. In recent years, researchers found the biogenesis of some siRNAs are “triggered” by miRNAs-mediated cleavage. Fragments resulted from mRNA cleavage are typically subjected to rapid degradation. However, a small proportion of the fragments will survive and further be processed into double-stranded RNA (dsRNA) by RNA-dependent RNA polymerase 6(RDR6) assisted with Suppressor of Gene Silencing 3 (SGS3). These double-stranded fragments will later be cleaved by Dicer-like (DCL) proteins in different phased manners to produce a series of 21- or 24-nt siRNAs, termed phased secondary small interfering RNAs (phasiRNAs)[8]

SiRNAs in 21-nt length regulates gene expression by cleaving complementary transcripts as the same to miRNA-mediated cleavage in plant. The best-characterized phasiRNAs are *TAS loci*-derived 21-nt *trans*-acting siRNAs (tasiRNAs) in *Arabidopsis*. Research indicated that miR173 targets *TAS1* and *TAS2* and resulted in production of tasiRNAs. Some of these tasiRNAs are further able to recognize target transcripts to produce tertiary phasiRNAs [9]. The *TAS1*- and *TAS2*-derived tasiRNAs were involved in regulation of stress responses, such as improvement of thermotolerance [10], maintaining the normal morphogenesis of flowers in plants under drought stress conditions [11]. The biogenesis of *TAS3-derived* tasiRNAs are triggered by the miR390 recognition [12]. These tasiRNAs target *ARF* family members which regulates various biological processes, including embryo development, developmental transitions, leaf morphology, flower and root architecture and stress responses [13, 14]. Besides, report showed the *TAS4*-derived tasiRNAs induced by miR828 regulated anthocyanin biogenesis by repressing *MYB* genes [15]. For siRNAs in 24-nt length, researches revealed that they are key players in triggering RNA-directed DNA methylation (RdDM)[16], which is the major small RNA-mediated epigenetic pathway that cause transposable element repression and transcriptional gene silencing (TGS) in plants [17]. For example, recent research discovered that the distribution of 24-nt siRNAs in rice gametes (sperm and egg) differ from each other, as well as from vegetative tissues, which further suggest a major difference in reprogramming of their genomes prior to fertilization[18].

Different algorithm and software tools have been employed not only in mining the novel miRNA-phasiRNA pathways, but in exploring the miRNAs’ extended regulatory networks [19]. Current research discovered two miRNAs, miR2118 and miR2275, were mainly responsible for the triggering of 21-nt and 24-nt phasiRNAs biogenesis, respectively [20]. And subsequent reports were then focused on the in-depth investigation of miR2118-phasiRNA and miR2275-phasiRNA biogenesis pathways and their biological functions [21, 22]. To our knowledge, about 56 phasiRNA precursors “PHAS loci”have been identified in rice. For PHAS loci in other economic crops, approximately 261 PHAS loci in *Zea mays*(maize), ~916 PHAS loci in *Setaria italica* (foxtail millet), ~201 PHAS loci in *Solanum tuberosum* (potato), and ~123 in *Solanum lycopersicum* (tomato) have been discovered, respectively[23]. In addition, recently reports revealed that in addition to those non-coding regions in genome, protein-coding genes could also be the PHAS loci in plants[23, 24], which implies a more complicated mechanism of plant phasiRNA biogenesis.

Due to the biological importance of phaiRNAs, mining of novel miRNA-phasiRNA pathways as well as functional cascade amplification have attracted wide attention. Since rice is an important economic crop as well as monocot model plant, mining novel phasiRNA pathways will not only benefit our understanding in post-transcriptional regulations in this organism, but also provide references across economic crops.

In our previous work, we discovered lots of siRNAs in a three-week-old seedling sample by using the corresponding sRNA high-throughput screening (HTS) datasets, which inspired us that, some of them might be phasiRNAs. In this study, we utilized our previously developed approach [25] for systematically mining of phasiRNA biogenesis pathways with these sRNA HTS datasets. In addition, considering some phasiRNAs expression might be tissue specific or stress dependent, we collected comparable sRNA HTS data sets published elsewhere using tissue-specific rice samples, which cultured under normal (control) or stress condition. The targets of novel phasiRNAs were further predicted and verified for providing substantial information of miRNA/sRNA-phasiRNA regulatory network in rice.

Results

Identification of novel phasiRNA biogenesis pathways in *Oryza sativa*

The sRNA HTS datasets from different rice samples were employed as inputs and rice 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 as well as the corresponding searching of sRNA triggers for phasiRNA production. (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[20].Therefore, we evaluated the abundance of 21-nt and 24-nt phasiRNAs generated from potential PHAS loci candidates by comparing the wild-type with *osdcl4* knockdown mutant (*osdcl4-1*)[26](for 21-nt phasiRNAs) and *osdcl3* knockdown mutant (*osdcl3-1*)[20](for 24-nt phasiRNAs), respectively.

As a result, five novel 21-nt PHAS loci and five novel 24-nt PHA loci along with their corresponding miRNA/sRNA triggers were identified. The information of PHAS loci gene ID, miRNA/sRNA trigger sequences, the miRNA/sRNA trigger binding sites and their mediated cleavage sites on PHAS loci which identified based on degradome-based HTS data (see details in “methods”) were listed in Table 1.The cleavage signature which detected by degradome sequencing, the abundance of phasiRNAs which generated from each phasing registers on the corresponding PHAS loci were profiled in Figure 1 and Figure 2. The abundance of phasiRNAs generated from these newly found 21-nt and 24-nt PHAS loci in wild type were relatively higher than that in *osdcl4-1* mutant and *osdcl4-1* mutant, respectively (Additional file 3: Figure S1 and Additional file 4: Figure S2).

Previously, we found lots of sRNAs generated in three-week-old seedling. Here, we tested whether these sRNA are phasiRNAs by using our mining method. As we expected, the phasiRNAs generated from two novel 21-nt PHAS loci, (*LOC_Os01g57968.1*：Gene ID of an expressed protein； and *LOC_Os05g43650.1*：Gene ID of an expressed protein；) and four novel 24-nt PHAS loci (*LOC_Os02g20200.1*：Gene ID of a retrotransposon protein；, *LOC_Os02g55550.1* (Gene ID of a F-box/LRR-repeat protein 14；, *LOC_Os04g45834.2* (Gene ID of DUF584 domain containing protein) and *LOC_Os09g14490.1* (Gene ID of TIR-NBS type disease resistance protein)) were identified in three-week-old seedling tissues (Table 1).

Since the triggering of phasiRNAs are sometimes tissue specific and stress dependent, a serial of sRNA HTS datasets of different rice samples were employed for mining novel PHAS loci in different tissue and stress condition. As shown in Figure 1, transcripts of 21-nt PHAS loci, *LOC_Os02g18750.1* (Gene ID of an expressed protein) and *LOC_Os04g25740.1* (Gene ID of an expressed protein), were able to produce 21-nt phasiRNAs in panicle under normal culture condition. *LOC_Os06g30680.1*(Gene ID of WD domain, G-beta repeat domain containing protein)-derived 21-nt phasiRNAs and *LOC_Os01g37325.1*(Gene ID of retrotransposon protein)-derived 24-nt phasiRNAs were detected in panicle both under normal and drought stress culture condition.

Two known 21-nt PHAS loci (*LOC_Os12g42380.1*(Gene ID of an expressed protein) and *LOC_Os12g42390.1*(Gene ID of hypothetical protein)) were also uncovered by our screening procedure (Additional file 1: Table S1), which have been identified as two parts of a long non-coding RNAs[27]. *LOC_Os12g42380.1*-derived phasiRNAs were detected in both seedling and panicle under normal, drought and salinity stress conditions, and in shoot they were only detected under salinity stress. *LOC_Os12g42390.1*-derived phasiRNAs were detected in shoot under normal culture condition, and panicle under drought culture condition. These results implied there are three alternative phasiRNA production regions within their lncRNA PHAS loci, their capability of phasiRNA production varies in different development stages and stress conditions.

To note, to our knowledge, for all these newly found PHAS loci, only the biogenesis of *LOC_Os04g25740.2*-derived phasiRNAs were triggered by a known miRNA, miR2118f. The rest of phasiRNAs were generated from first-time discovered PHAS loci, and were triggered by novel sRNAs (Table 1), which suggested that, these phasiRNA biogenesis pathways are not belong to the miR2118 or miR2257 mediated regulatory networks.

Table 1 Novel PHAS loci in *Oryza sativa*

Gene ID of the PHAS loci	Gene annotation	PhasiRNA production region	sRNA trigger ID	sRNA trigger sequence	Binding sites of sRNA trigger on gene of PHAS loci	Cleavage sites discovered by degradome on gene of PHAS loci
21-nt PHAS loci						
LOC_Os01g57968.1	expressed protein	361-1765	OSsRNA-1	GCUUUUUUGAACUUUUUCAUU	424-444	435
LOC_Os02g18750.1	expressed protein	188-920	OSsRNA-2	UUUUUUGGCAUUCUGUAACUUG	176-197	188
LOC_Os04g25740.1	expressed protein	1908-2159	osa-miR2118f	UUCCUGAUGCCUCCCAUUCCUA	1875-1896	1887
LOC_Os05g43650.1	expressed protein	1494-1620	OSsRNA-3	GAUUCAUUAACUUCAAUAUGAA	1528-1549	1540
LOC_Os06g30680.1	WD domain, G-beta repeat domain containing protein	62-208	OSsRNA-4	UUCCUGGAGCCGCUCAUUCCAU	50-71	62
24-nt PHAS loci						
LOC_Os01g37325.1	retrotransposon protein	1565-1760	OSsRNA-14	AAAAGUAGAUGGAUGCGGAGAC	1676-1697	1688
LOC_Os02g20200.1	retrotransposon protein	4856-5052	OSsRNA-15	UAGAUGCUGUCCUGAAAAGGUG	4873-4894	4885
			OSsRNA-16	AGCCAUGCUAGUCUAAGAGGG	5007-5027	5018
LOC_Os02g55550.1	F-box/LRR-repeat protein 14	905-1101	OSsRNA-17	UAGAUGCUGUCCUGAAAAGGUG	922-943	934
LOC_Os04g45834.2	DUF584 domain containing protein	1051-1307	OSsRNA-18/ OSsRNA-19	UUAAUAUUUAUAAUUAGUGUCU/ UUAAUAUUUAUAAUUAAUGUCC	1103-1124	1115
LOC_Os09g14490.1	TIR-NBS type disease resistance protein	4585-4757	OSsRNA-20	UAGAUGCUGUCCUGAAAAGGUG	4578-4599	4590

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

The tasiRNAs are those 21-nt phasiRNAs have regulatory function in trans-regulation of target genes by cleaving mRNAs in plant. In order to identify novel tasiRNAs generated from the newly found 21-nt PHAS loci, all the 21-nt phasiRNAs were systematically “predicted” based on the modified model of tasiRNA biogenesis [28]. All of the detectable phasiRNAs were then employed for target prediction based on miRU algorithm and verified by using degradome-based HTS data (see details in “methods”). The results indicated ten novel tasiRNAs were generated from three newly found 21nt PHAS loci (*LOC_Os02g18750.1*, *LOC_Os05g43650.1* and *LOC_Os06g30680.1*), respectively. These tasiRNAs mediated forty sRNA-target interactions (Table 2, Figure 3, Additional file 5: Figure S3).Among these targets, *LOC_Os02g39380.1* played important roles in plant cellular signaling cascades [29]. (*LOC_Os01g34620.8*, *LOC_Os02g52900.2*, *LOC_04g39600.1*, *LOC_08g40440.1*, *LOC_Os6g23274.1*, *LOC_Os06g47850.1*, *LOC_11g41860.1*,*LOC_11g41860.2* and *LOC_Os05g46580.1*) were involved in plant growth and development [30-35]. And *LOC_Os09g12230.1*, *LOC_Os04g38450.1* and *LOC_Os04g49160.1* were related to plant defense and stress response [36-38].

Although the transcript of *LOC_Os12g42380.1* has been identified as part of an lncRNA phasiRNA precursor[27], one novel *LOC_Os12g42380.1*-derived tasiRNAs was found based on our revised tasiRNA biogenesis model[28]. LOC_Os12g42380.1 (414)21 5'D7(+) targeted to a NAD-dependent epimerase/dehydratase gene (*LOC_Os07g47700.1*) (Table 1, Addition file 5: Figure S3), suggesting it might be involved in regulation of plant growth, development and environmental stress[39, 40].

Taken together, these results suggested the OSsRNA-2-*LOC_Os02g18750.1*-phasiRNA, OSsRNA-3-*LOC_Os05g43650.1*-phasiRNA, OSsRNA-4-L*OC_Os06g30680.1*-phasiRNA and OSsRNA-5-*LOC_Os12g42380.1*-phasiRNA pathways might play crucial regulatory roles in rice growth, development and stress response. In addition, the regulatory networks of the phasiRNAs pathways mentioned above were constructed based on the target information (Figure 4).

Table 2 Targets of novel tasiRNAs in *Oryza sativa*

TaisRNA ID	tasiRNA sequence	Targets	Target annotation	miRU start-ending	taisRN media cutsit
LOC_Os02g18750.1(189)21 3'D26 (+)	UGUGCCACGUCAACACCACCA	LOC_Os03g40260.1	Regulator of chromosome condensation domain containing protein	1676-1696	1687
LOC_Os02g18750.1(192)21 3'D25 (+)	GCGCCACUGCCGUCGACGUGU	LOC_Os02g39380.1	OsCML17 - Calmodulin-related calcium sensor protein	343-363	354
LOC_Os02g18750.1(204)21 3'D13 (+)	UCGACUUCGCCGCCUCGGCGC	LOC_Os02g39090.1	expressed protein	802-823	814
LOC_Os05g43650.1(1540)21 3'D2(+)	UCAAUAUGAAUGUGGAAAAUG	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(-)	UUUUCCACAUUCAUAUUGAUG	LOC_Os02g45650.1	peptidase	1760-1780	1771
LOC_Os05g43650.1(1542)21 3'D1(+)	AAUGAAUCUAGACAUAUAUAU	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
		LOC_Os05g51140.1	expressed protein	929-949	940

		LOC_Os05g51140.2	expressed protein	1586-1606	1597
		LOC_Os09g33930.1	farnesyltransferase/geranylgeranyltransferase type-1 subunitalph	1457-1477	1468
		LOC_Os09g33930.2	farnesyltransferase/geranylgeranyltransferase type-1 subunitalph	1454-1474	1465
		LOC_Os09g33930.3	farnesyltransferase/geranylgeranyltransferase type-1 subunitalph	1740-1760	1751
		LOC_Os09g33930.4	farnesyltransferase/geranylgeranyltransferase type-1 subunitalph	1453-1473	1464
		LOC_Os09g33930.5	farnesyltransferase/geranylgeranyltransferase type-1 subunitalph	1375-1395	1386
		LOC_Os12g37510.1	UDP-glucoronosyl and UDP-glucosyl transferase domain containing	1584-1604	1595
LOC_Os05g43650.1(1543)21 3'D2(-)	GCAUUUUCCACAUUCAUAUUG	LOC_Os02g48390.1	phosphoribosyl transferase	1758-1778	1769
LOC_Os05g43650.1(1543)21 3'D3(-)	UUCACAAUGUAAGUCAUUUUA	LOC_Os04g39600.1	fasciclin domain containing protein	1020-1040	1031
		LOC_Os07g01130.1	pentatricopeptide containing protein	4240-4260	4251
LOC_Os05g43650.1(1543)21 3'D1(+)	AUGAAUCUAGACAUAUAUAUC	LOC_Os12g40920.1	bZIP transcription factor domain containing protein	1312-1332	1323
LOC_Os06g30680.1(62)21 3' D2(+)	CAUGGACAACUUCCUGCACAG	LOC_Os05g46580.1	polyprenyl synthetase	1365-1385	1376
LOC_Os12g42380.1(414)21 5'D7(+)	UUUCUUCCAAGAGAGAGUAAG	LOC_Os07g47700.1	NAD dependent epimerase/dehydratase family domain containing protein	1753-1773	1764

Analysis of the RNA directed DNA methylation （RdDM）regulated promoters of novel 24-nt phasiRNAs

RdDM is an important regulatory event with regards to repressive epigenetic modification which triggers transcriptional gene silencing. In order to analysis the novel 24-nt phasiRNA mediated RdDM in rice, we focused on all the known promoter sequences for scanning the target sites of novel phasiRNAs generated from the newly found five 24-nt PHAS loci. The result indicated a promoter of *LOC_Os02g40860.1* gene was targeted by five *LOC_Os01g37325.1*-derived phasiRNAs (Table 3). Since *LOC_Os01g37325.1*-derived phasiRNAs were detected in panicle rather than in root tissue (Figure 2), we used the bisulfite-seq and RNA-seq datasets[41] of rice panicle and root for identification of the *LOC_Os01g37325.1*-derived phasiRNAs mediated DNA methylation on the target promoter and their role in transcriptional repression of target gene (*LOC_Os02g40860.1*), respectively. It is reported that CG and CHG methylation contexts are maintained by DNA methyltransferases and histone modifications, while CHH methylation is associated with 24-nt siRNA guided RdDM[16]. We discovered the CHH methylation status of promoter was relative higher in panicle than in root (Figure 5). In addition, the expression level of *LOC_Os02g40860.1* was relatively lower in panicle than in root. These results implied a methylation mediated transcriptional silencing of the promoter of *LOC_Os02g40860.1*.

For *LOC_Os02g40860.1*, it encodes a Casein kinase I1 (OsCKI1) protein belongs to the CKIs protein family. CKIs are highly conserved in eukaryotes, and they are involved in a variety of important biological events since they have a wide substrate specificity in *vitro* [42]. Taken together, we speculated that the OSsRNA-14- *LOC_Os01g37325.1*-phasiRNA pathway might play crucial roles for rice seedling and panicle development.

Table 3 The target promoter of LOC_Os01g37325.1-derived phasiRNAs

24-nt phasiRNAs_ID	PhasiRNAs_sequences	Binding_sites_on_promoter	Prmoter_location	Target_genes	Target annotation
LOC_Os01g37325.1(1684) 24 5'D12(+)	AUCAUGACUUGGGUAUUACGUUUC	111-134	chr2_24766608-24766807	LOC_Os02g40860.1	Casein kinase I1 (CKI1)
LOC_Os01g37325.1(1684) 24 5'D10(+)	AGUCCUGGUUUGAUAAGAUUGUAA	63-86			
LOC_Os01g37325.1(1684) 24 5'D9(+)	AGUAGAUUUAGGAAACCGAUACCG	39-62			
LOC_Os01g37325.1(1665) 24 5'D13(+)	ACUAGUUAUAGGGGAUAACUUAUA	154-177			
LOC_Os01g37325.1(1665) 24 5'D11(+)	GACUUGGGUAUUACGUUUCCCUGU	106-129			

Discussion

In recent years, researches focused on *Oryza sativa* have shown that 21- or 24-nt sRNAs distribute to genomic clusters[43]. To date, dozens of PHAS loci have been discovered in rice[23]. However, two miRNAs, miR2118 and miR2275 are mainly responsible for the triggering of 21-nt or 24-nt phasiRNAs biogenesis from these PHAS loci.

Considering there are rich sources for miRNA/sRNA-phasiRNA pathways in other plant species, we hypothesized that the miRNA-phasiRNA pathways have not been fully discovered in rice. Therefore, it is necessary to continue the mining for better understanding the mechanism of phasiRNA biogenesis and the miRNA-derived regulatory network. In our previous work, we found plenty of sRNAs with unknown function and origin from a sRNA HTS data set of three-week-old seedling tissue, and speculated some of them were phasiRNAs with regulatory functions. In this study, we performed a systematically searching of novel PHAS loci from rice cDNA by utilizing the same seedling dataset with our previous established mining approach [25].

As we expected, two novel 21-nt phasiRNA biogenesis pathways (OSsRNA-2-*LOC_Os01g57968.1*-phasiRNA and OSsRNA-3-*LOC_Os05g43650.1*-phasiRNA pathway) and four novel 24-nt phasiRNA biogenesis pathways (OSsRNA-15/OSsRNA-16-*LOC_Os02g20200.1*-phasiRNA, OSsRNA-17-*LOC_Os02g55550.1*-phasiRNA, OSsRNA-18/OSsRNA-19-*LOC_Os04g45834.2*-phasiRNA and OSsRNA-20-*LOC_Os09g14490.1*- phasiRNA pathway) were discovered. In addition, since the phasiRNAs are involved in regulation of plant growth and development, stress response, we integrated a serial of sRNA HTS datasets from different tissues (including two- week -old seedling samples) under normal and stress condition. As a result, three novel 21-nt phasiRNA biogenesis pathways (OSsRNA-2-*LOC_Os02g18750.1*-phasiRNA, osa-miR2118f-*LOC_Os04g25740.1*-phasiRNA and OSsRNA-4-*LOC_Os06g30680.1*-phasiRNA pathway) and one novel 24-nt phasiRNA biogenesis pathway (OSsRNA-2-LOC_Os01g37325.1-phasiRNA) were discovered. The mining results substantially extend the information of phasiRNA biogenesis pathways in rice. However, the six novel phasiRNAs biogenesis pathways that we discovered in three-week-old seedling were undetected in two-week-old seedling samples, which we speculated might be caused by the low expression level of phasiRNAs generated from these pathways in younger seedlings.

The novel 21-nt PHAS loci, *LOC_Os05g43650.1*, is a miniature inverted-repeat transposable element (MITE) gene [44], while for two 24-nt PHAS loci, *LOC_Os01g37325.1* and *LOC_Os02g20200.1*, they are two retrotransposon genes, which suggested that the transcripts of transponsons and retrotransponsons are capable of producing secondary siRNAs , which is consistent with the same phenomenon has been reported by Creasey *et al*. in *Arabidopsis*[45, 46].

According to the information of targets of phasiRNAs, the OSsRNA-3- *LOC_Os05g43650.1*-phasiRNA and OSsRNA-14- *LOC_Os01g37325.1*- phasiRNA pathways are required for the rice development. Transponsons and retrotransponsons are ubiquitous in plants and play important roles in plant gene and genome evolution [47]. We hypothesized that the transcripts of transponson and retrotransponson might also function as important sources of phasiRNA in plants. Further exploration of such phasiRNA biogenesis pathways could benefits the in-depth investigation of their biogenesis mechanism and the miRNA/sRNA directed regulatory networks in plants.

For those phasiRNAs generated from the transcripts of *LOC_Os01g57968.1*, *LOC_Os02g20200.1*, *LOC_Os02g55550.1*, *LOC_Os04g45834.2* and *LOC_Os09g14490.1*, none of their targets were identified. But considering these phasiRNAs were detected only in seedling tissue, it's still cannot rule out the possibility that these phasiRNA biogenesis pathways might be involved in rice seedling development. *LOC_Os04g45834.2* encodes a DUF584 domain containing protein. These protein family has been involved in leaf senescence in plant [48]. *LOC_Os09g14490.1* encodes a TIR-NBS type disease resistance protein, which has been identified in resistance to multiple viruses in plant [49-51]. *LOC_Os02g55550.1* encodes a F-box/LRR-repeat protein 14, which is involved in plant immune response [52]. 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 what we expected. Similarly, no targets of *LOC_Os01g57968.1*-derived phasiRNAs was identified, however, since these phasiRNAs only expressed in panicle tissue under normal culture condition, it might suggest the OSsRNA-1-*LOC_Os01g57968.1*-phasiRNA pathway might related to the rice panicle development. Thus, 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, 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 searching of 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). This might suggest a novel pattern of secondary siRNAs biogenesis pathways. Therefore, further investigation of Dicer-independent secondary siRNAs biogenesis pathways in plant might provide more strong evidence of this biogenesis pattern, and more meaningful information of the small RNA regulatory mechanism in plant.

Conclusions

We performed degradome-based screening of novel phasiRNA biogenesis pathways in rice. Besides two known 21-nt phasiRNA biogenesis pathways, five novel 21-nt phasiRNA biogenesis pathways and five novel 24-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 2, Additional file 1: Table S1, Additional file 2: Table S2 and Additional file 5: Figure S3). 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* sRNA HTS datasets of seedling, root, shoot and panicle samples under normal (control) and stress culture conditions, the sRNA HTS datasets of wild type, *osdcl4* and *osdcl3* mutants and the degradome sequencing data sets were retrieved from GEO (Gene Expression Omnibus; http://www.ncbi.nlm.nih.gov/geo/). The bisulfite-seq and RNA-seq datasets of panicle and root were contributed by Zhao et al [41]. All the HTS datasets employed for our study were listed in table 4.

Table 4 The datasets utilized for our study.

Datasets	Rice samples	GEO_number
sRNA HTS dataset from seedling	three week old seeding	GSM455965
sRNA HTS data sets from different tissues and culture condition (serier number GSE32973)	seedling_control (two-week-old seedling)	GSM816687,GSM816688 and GSM816689
	seedling_drought (two-week-old seedling)	GSM816690 and GSM816691
	seedling_salt (two-week-old seedling)	GSM816692 and GSM 816693
	seedling_cold (two-week-old seedling)	GSM816694 and GSM816695
	root_control	GSM816704 and GSM816705
	root_drought	GSM816716 and GSM816717
	root_salt and lack of potassium	GSM816714 and GSM816715
	shoot_control	GSM816718 and GSM816719
	shoot_salt and lack of potassium	GSM816728 and GSM816729
	panicle_control	GSM816730, GSM816731 and GSM816732
	panicle_drought	GSM816733,GSM816734 and GSM816735
	panicle_salt	GSM816736, GSM816737 and GSM816738
	panicle_cold	GSM816739 and GSM816740
	seedling (wt)	GSM562942
	panicle (wt)	GSM562943
	seedling (*osdcl4-1*)	GSM562944
	panicle (*osdcl4*-1)	GSM562945
	Seedling (wt)	GSM520638
	Seedling (*osdcl3*-1)	GSM520640
Degradome sequencing datasets	Seedling	GSM434596
	Seedling	GSM455938
	young inflorescences	GSM455939
	young inflorescences	GSM476257
bisulfite-seq datasets	panicle	GSM4232038
	root	GSM4232039
RNA-seq datasets	panicle	GSM4230036
	panicle	GSM4230037
	root	GSM4230038
	root	GSM4230039

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[53].

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 [28]. 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 2-nt 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 contain 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 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[28], 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[54], 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 [55]. (5) The phasing score of phasiRNA production from each PHAS loci candidate should above 1.

Calculation of phasing score

Phasing scores of phasiRNA regions were calculated based on the formula which contributed by Zheng et al [23]: Phasing score = $\ln\left[\left(1+10\times\frac{\sum_{i=1}^{5}pi}{1+\sum U}\right)^{n-2}\right]$, 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.

Identitification 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[28]. The predicted phasiRNAs were recruited for target prediction by using miRU with default parameters[54], and followed by degradome sequencing-based verification, as described previously[53, 56].

Gene expression level analysis

The sequences of RNA-seq datasets were mapped to the reference cDNA sequences, and each gene expression level was calculated by the total RPM of mapped sequences.

Identification of 24nt phasiRNA target

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 [57]. 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 datasets of panicle and root 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 developed by Zhao et al [41]. The sequences of bisulfite sequencing libraries 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: ARGONATE

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].

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. Hohn T, Vazquez F: **RNA silencing pathways of plants: silencing and its suppression by plant DNA viruses**. *Biochimica et biophysica acta* 2011, **1809**(11-12):588-600.
2. Mermigka G, Verret F, Kalantidis K: **RNA silencing movement in plants**. *Journal of integrative plant biology* 2016, **58**(4):328-342.
3. Yu D, Lu J, Shao W, Ma X, Xie T, Ito H, Wang T, Xu M, Wang H, Meng Y: **MepmiRDB: a medicinal plant microRNA database**. *Database : the journal of biological databases and curation* 2019, **2019**.
4. Singh A, Gautam V, Singh S, Sarkar Das S, Verma S, Mishra V, Mukherjee S, Sarkar AK: **Plant small RNAs: advancement in the understanding of biogenesis and role in plant development**. *Planta* 2018, **248**(3):545-558.
5. Bian H, Xie Y, Guo F, Han N, Ma S, Zeng Z, Wang J, Yang Y, Zhu M: **Distinctive expression patterns and roles of the miRNA393/TIR1 homolog module in regulating flag leaf inclination and primary and crown root growth in rice (Oryza sativa)**. *The New phytologist* 2012, **196**(1):149-161.
6. Zhu C, Ding Y, Liu H: **MiR398 and plant stress responses**. *Physiologia plantarum* 2011, **143**(1):1-9.
7. Guleria P, Mahajan M, Bhardwaj J, Yadav SK: **Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses**. *Genomics, proteomics & bioinformatics* 2011, **9**(6):183-199.
8. Yu Y, Zhang Y, Chen X, Chen Y: **Plant Noncoding RNAs: Hidden Players in Development and Stress Responses**. *Annual review of cell and developmental biology* 2019, **35**:407-431.
9. Fei Q, Xia R, Meyers BC: **Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks**. *The Plant cell* 2013, **25**(7):2400-2415.
10. Li S, Liu J, Liu Z, Li X, Wu F, He Y: **HEAT-INDUCED TAS1 TARGET1 Mediates Thermotolerance via HEAT STRESS TRANSCRIPTION FACTOR A1a-Directed Pathways in Arabidopsis**. *The Plant cell* 2014, **26**(4):1764-1780.
11. Matsui A, Mizunashi K, Tanaka M, Kaminuma E, Nguyen AH, Nakajima M, Kim JM, Nguyen DV, Toyoda T, Seki M: **tasiRNA-ARF pathway moderates floral architecture in Arabidopsis plants subjected to drought stress**. *BioMed research international* 2014, **2014**:303451.
12. Axtell MJ, Jan C, Rajagopalan R, Bartel DP: **A two-hit trigger for siRNA biogenesis in plants**. *Cell* 2006, **127**(3):565-577.
13. D'Ario M, Griffiths-Jones S, Kim M: **Small RNAs: Big Impact on Plant Development**. *Trends in plant science* 2017, **22**(12):1056-1068.
14. Xia R, Xu J, Meyers BC: **The Emergence, Evolution, and Diversification of the miR390-TAS3-ARF Pathway in Land Plants**. *The Plant cell* 2017, **29**(6):1232-1247.

15. Luo QJ, Mittal A, Jia F, Rock CD: **An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis**. *Plant molecular biology* 2012, **80**(1):117-129.
16. Law JA, Jacobsen SE: **Establishing, maintaining and modifying DNA methylation patterns in plants and animals**. *Nature reviews Genetics* 2010, **11**(3):204-220.
17. Liu Q, Ding C, Chu Y, Zhang W, Guo G, Chen J, Su X: **PlN24NT: a web resource for plant 24-nt siRNA producing loci**. *Bioinformatics* 2017, **33**(13):2065-2067.
18. Li C, Xu H, Fu FF, Russell SD, Sundaresan V, Gent JI: **Genome-wide redistribution of 24-nt siRNAs in rice gametes**. *Genome research* 2020, **30**(2):173-184.
19. 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):15144-15149.
20. 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**. *The Plant journal : for cell and molecular biology* 2012, **69**(3):462-474.
21. 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):1535-1550.
22. Liu Y, Wang Y, Zhu QH, Fan L: **Identification of phasiRNAs in wild rice (Oryza rufipogon)**. *Plant Signal Behav* 2013, **8**(8).
23. Zheng Y, Wang Y, Wu J, Ding B, Fei Z: **A dynamic evolutionary and functional landscape of plant phased small interfering RNAs**. *BMC biology* 2015, **13**:32.
24. Komiya R: **Biogenesis of diverse plant phasiRNAs involves an miRNA-trigger and Dicer-processing**. *Journal of plant research* 2017, **130**(1):17-23.
25. 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):1668-1675.
26. Liu B, Chen Z, Song X, Liu C, Cui X, Zhao X, Fang J, Xu W, Zhang H, Wang X *et al*: **Oryza sativa dicer-like4 reveals a key role for small interfering RNA silencing in plant development**. *The Plant cell* 2007, **19**(9):2705-2718.
27. 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**. *Journal of experimental botany* 2019, **70**(6):1767-1774.
28. 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.
29. 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.
30. Rouhier N, Couturier J, Jacquot JP: **Genome-wide analysis of plant glutaredoxin systems**. *J Exp Bot* 2006, **57**(8):1685-1696.
31. 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):2647-2668.
32. 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 Genomics* 2017, **18**(1):129.
33. 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**. *European Journal of Plant Pathology* 2018, **151**(3):18.
34. 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):11668-11675.
35. Vranova E, Coman D, Gruissem W: **Structure and dynamics of the isoprenoid pathway network**. *Molecular plant* 2012, **5**(2):318-333.
36. 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):1371-1390.
37. 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.
38. Campo S, Baldrich P, Messeguer 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.
39. 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.
40. Zhu H, Kranz RG: **A nitrogen-regulated glutamine amidotransferase (GAT1_2.1) represses shoot branching in Arabidopsis**. *Plant physiology* 2012, **160**(4):1770-1780.
41. Zhao L, Xie L, Zhang Q, Ouyang 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.
42. 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 and molecular biology* 2003, **36**(2):189-202.
43. Johnson C, Kasprzewska A, Tennessen 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):1429-1440.
44. 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):1005-1017.

45. 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):411-415.
46. 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):461-472.
47. Lisch D: **How important are transposons for plant evolution?***Nat Rev Genet* 2013, **14**(1):49-61.
48. 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.
49. 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.
50. 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).
51. 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 and molecular research : GMR* 2015, **14**(3):7964-7975.
52. Matsushima N, Miyashita H: **Leucine-Rich Repeat (LRR) Domains Containing Intervening Motifs in Plants**. *Biomolecules* 2012, **2**(2):288-311.
53. 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.
54. Zhang Y: **miRU: an automated plant miRNA target prediction server**. *Nucleic acids research* 2005, **33**(Web Server issue):W701-704.
55. 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):758-762.
56. 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):356-362.
57. 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 and biochemistry : PPB* 2010, **48**(6):393-400.

Figures

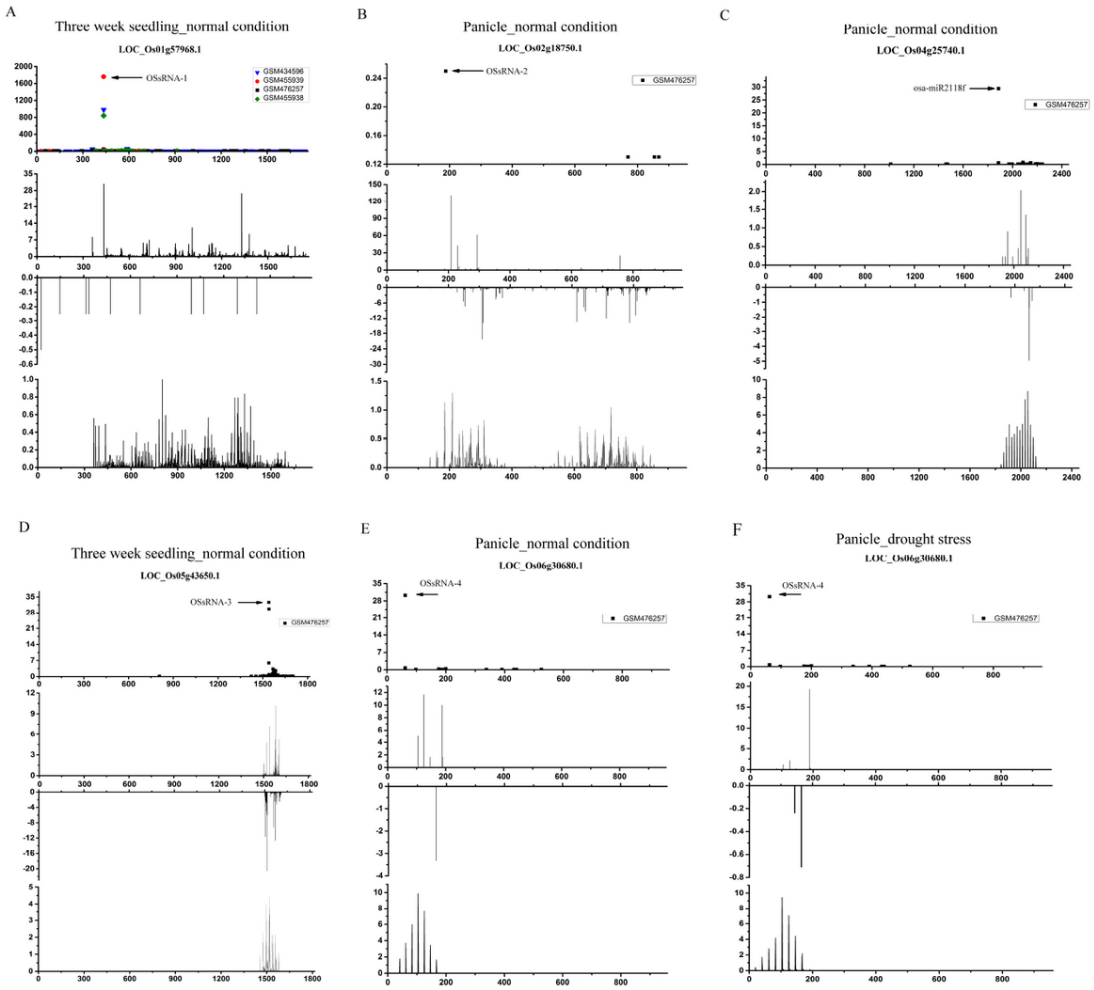

Figure 1

Identification of novel 21-nt phasiRNAs biogenesis pathways in Oryza sativa. (A) OSsRNA-1 induced phasiRNAs generation form the transcript of LOC_Os01g57968.1 in seedling, (B) OSsRNA-2 induced phasiRNAs generation form the transcript of LOC_Os02g18750.1 in panicle, (C) osa-miR2118f induced phasiRNAs generation form the transcript of LOC_Os04g25740.1 in panicle, (D) OSsRNA-3 induced phasiRNAs generation form the transcript of LOC_Os05g43650.1 in seedling, (E) OSsRNA-4 induced phasiRNAs generation form the transcript of LOC_Os06g30680.1 in panicle, and (F) OSsRNA-4 induced phasiRNAs generation form the transcript of LOC_Os06g30680.1 in panicle (drought stress). For each graph, degradome supported cleavage signature on PHAS loci were profiled above, four high throughput degradome sequencing datasets (GSM434596, GSM455938, GSM455939 and GSM476257 which were represented by triangle, diamond, circle and square with different colors, respectively) were employed for scanning the sRNA triggers' cleavage sites, which were marked by black arrows. The x axis represents the position on PHAS loci and the y axis represents the signature abundance. The abundance of 21-nt phasiRNAs which generated from the sense and antisense strand of PHAS loci in different samples were evaluated and profiled in middle images, the x axis represents the position on PHAS loci and the y axis represents the phasiRNA abundance. The phasing score of 21-nt PHAS windows were profiled at bottom, the x axis represents the position on PHAS loci and the y axis represents the phasing score.

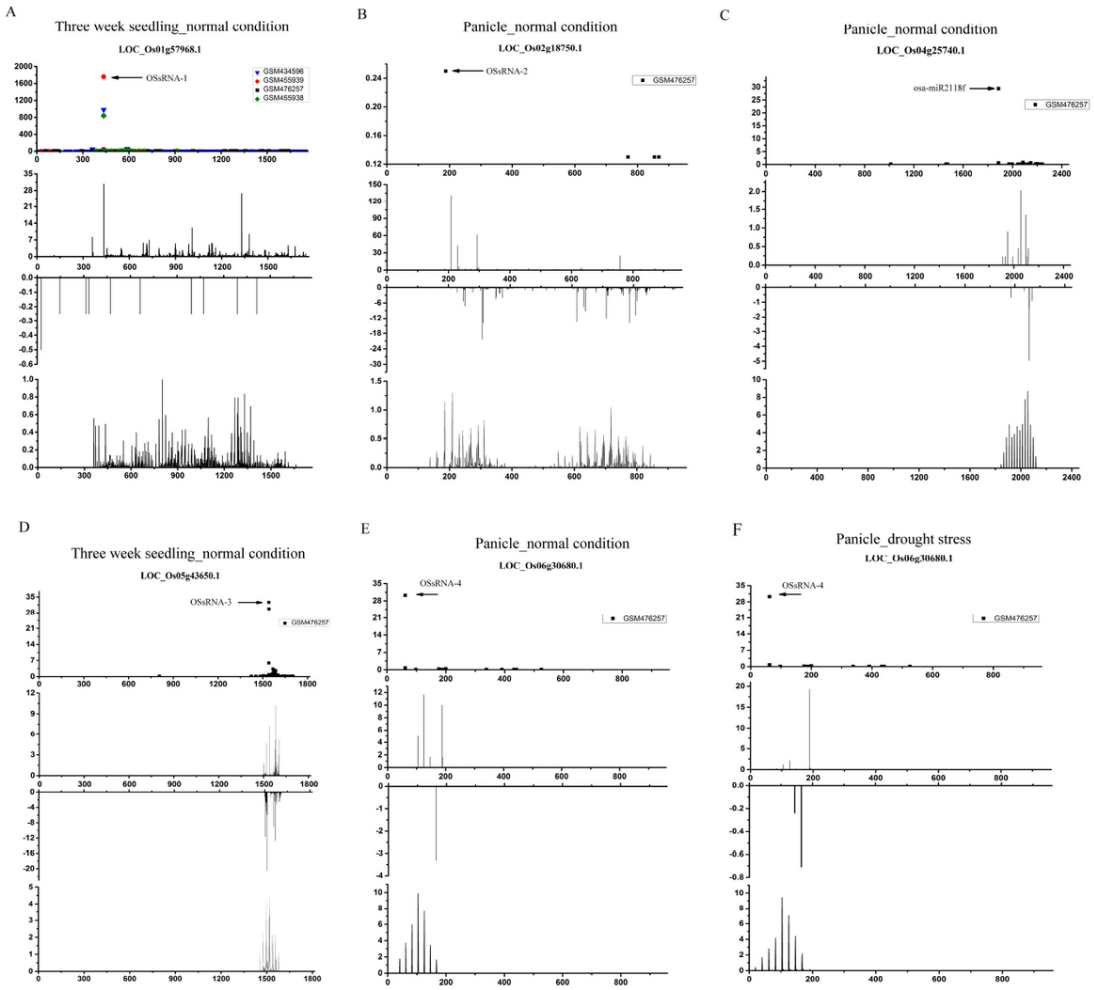

Figure 1

Identification of novel 21-nt phasiRNAs biogenesis pathways in Oryza sativa. (A) OSsRNA-1 induced phasiRNAs generation form the transcript of LOC_Os01g57968.1 in seedling, (B) OSsRNA-2 induced phasiRNAs generation form the transcript of LOC_Os02g18750.1 in panicle, (C) osa-miR2118f induced phasiRNAs generation form the transcript of LOC_Os04g25740.1 in panicle, (D) OSsRNA-3 induced phasiRNAs generation form the transcript of LOC_Os05g43650.1 in seedling, (E) OSsRNA-4 induced phasiRNAs generation form the transcript of LOC_Os06g30680.1 in panicle, and (F) OSsRNA-4 induced phasiRNAs generation form the transcript of LOC_Os06g30680.1 in panicle (drought stress). For each graph, degradome supported cleavage signature on PHAS loci were profiled above, four high throughput degradome sequencing datasets (GSM434596, GSM455938, GSM455939 and GSM476257 which were represented by triangle, diamond, circle and square with different colors, respectively) were employed for scanning the sRNA triggers' cleavage sites, which were marked by black arrows. The x axis represents the position on PHAS loci and the y axis represents the signature abundance. The abundance of 21-nt phasiRNAs which generated from the sense and antisense strand of PHAS loci in different samples were evaluated and profiled in middle images, the x axis represents the position on PHAS loci and the y axis represents the phasiRNA abundance. The phasing score of 21-nt PHAS windows were profiled at bottom, the x axis represents the position on PHAS loci and the y axis represents the phasing score.

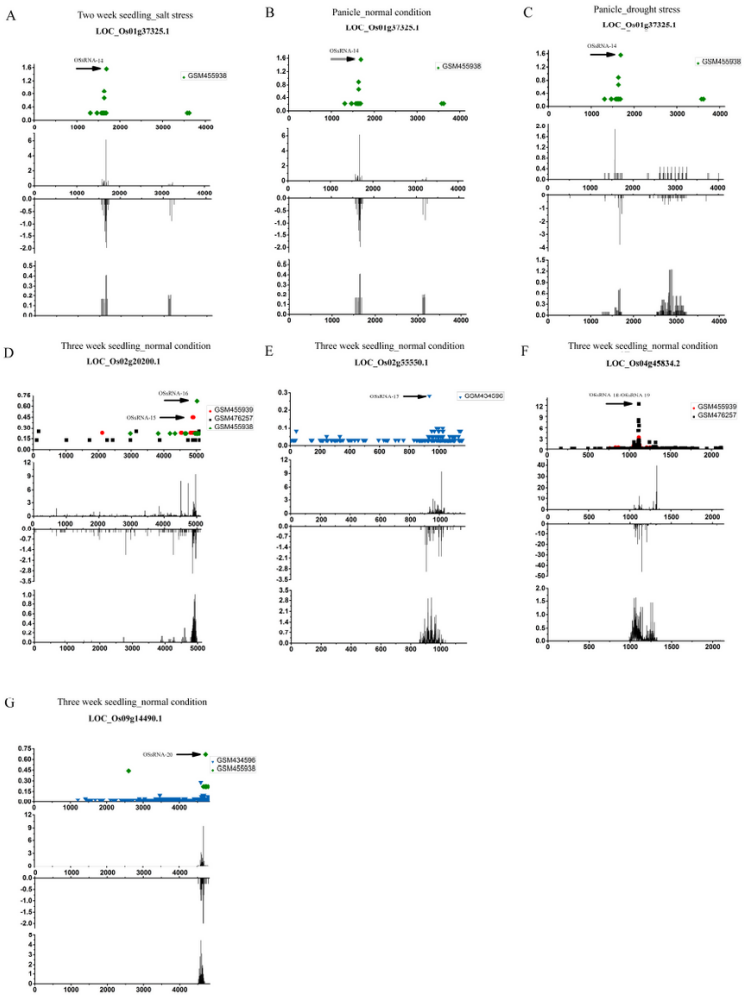

Figure 2

Identification of novel 24-nt phasiRNAs biogenesis pathways in Oryza sativa. (A)OSsRNA-14 induced phasiRNAs generation from the transcript of LOC_Os01g37325.1 in seedling (salt stress). (B) OSsRNA-14 induced phasiRNAs generation from the transcript of LOC_Os01g37325.1 in panicle. (C) OSsRNA-14 induced phasiRNAs generation form the transcript of LOC_Os01g37325.1 in panicle (drought stess). (D) OSsRNA-15 and OSsRNA-16 induced phasiRNAs generation form the transcript of LOC_Os02g20200.1 in seedling. (E) OSsRNA-17 induced phasiRNAs generation form the transcript of LOC_Os02g55550.1 in seedling. (F) OSsRNA-18 or OSsRNA-19 induced phasiRNAs generation form the transcript of LOC_Os04g45834.2 in seedling. (G) OSsRNA-20 induced phasiRNAs generation form the transcript of LOC_Os09g14490.1 in seedling. For each graph, degradome supported cleavage signature on PHAS loci were profiled above, four high throughput degradome sequencing datasets (GSM434596, GSM455938, GSM455939 and GSM476257 which were represented by triangle, diamond, circle and square with different colors, respectively) were employed for scanning the sRNA triggers' cleavage sites, which were marked by black arrows. The x axis represents the position on PHAS loci and the y axis represents the signature abundance. The abundance of 24-nt phasiRNAs which generated from the sense and antisense strand of PHAS loci in different samples were evaluated and profiled in middle images, the x axis represents the position on PHAS loci and the y axis represents the phasiRNA abundance. The phasing score of 24-nt PHAS windows were profiled at bottom, the x axis represents the position on PHAS loci and the y axis represents the phasing score.

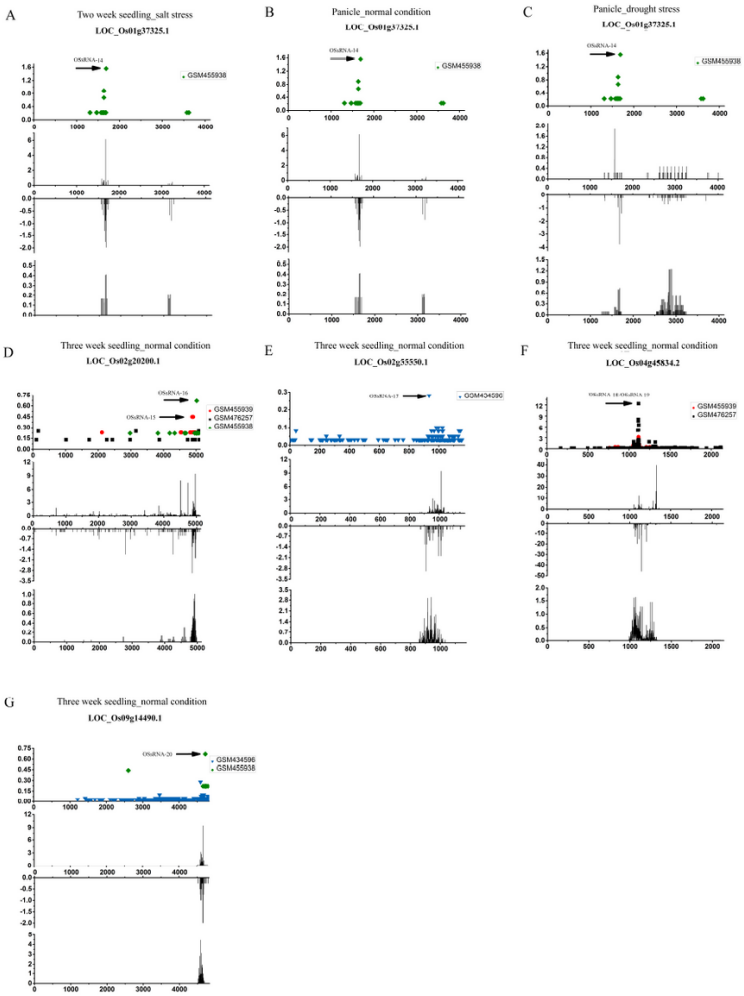

Figure 2

Identification of novel 24-nt phasiRNAs biogenesis pathways in Oryza sativa. (A)OSsRNA-14 induced phasiRNAs generation from the transcript of LOC_Os01g37325.1 in seedling (salt stress). (B) OSsRNA-14 induced phasiRNAs generation from the transcript of LOC_Os01g37325.1 in panicle. (C) OSsRNA-14 induced phasiRNAs generation form the transcript of LOC_Os01g37325.1 in panicle (drought stess). (D) OSsRNA-15 and OSsRNA-16 induced phasiRNAs generation form the transcript of LOC_Os02g20200.1 in seedling. (E) OSsRNA-17 induced phasiRNAs generation form the transcript of LOC_Os02g55550.1 in seedling. (F) OSsRNA-18 or OSsRNA-19 induced phasiRNAs generation form the transcript of LOC_Os04g45834.2 in seedling. (G) OSsRNA-20 induced phasiRNAs generation form the transcript of LOC_Os09g14490.1 in seedling. For each graph, degradome supported cleavage signature on PHAS loci were profiled above, four high throughput degradome sequencing datasets (GSM434596, GSM455938, GSM455939 and GSM476257 which were represented by triangle, diamond, circle and square with different colors, respectively) were employed for scanning the sRNA triggers' cleavage sites, which were marked by black arrows. The x axis represents the position on PHAS loci and the y axis represents the signature abundance. The abundance of 24-nt phasiRNAs which generated from the sense and antisense strand of PHAS loci in different samples were evaluated and profiled in middle images, the x axis represents the position on PHAS loci and the y axis represents the phasiRNA abundance. The phasing score of 24-nt PHAS windows were profiled at bottom, the x axis represents the position on PHAS loci and the y axis represents the phasing score.

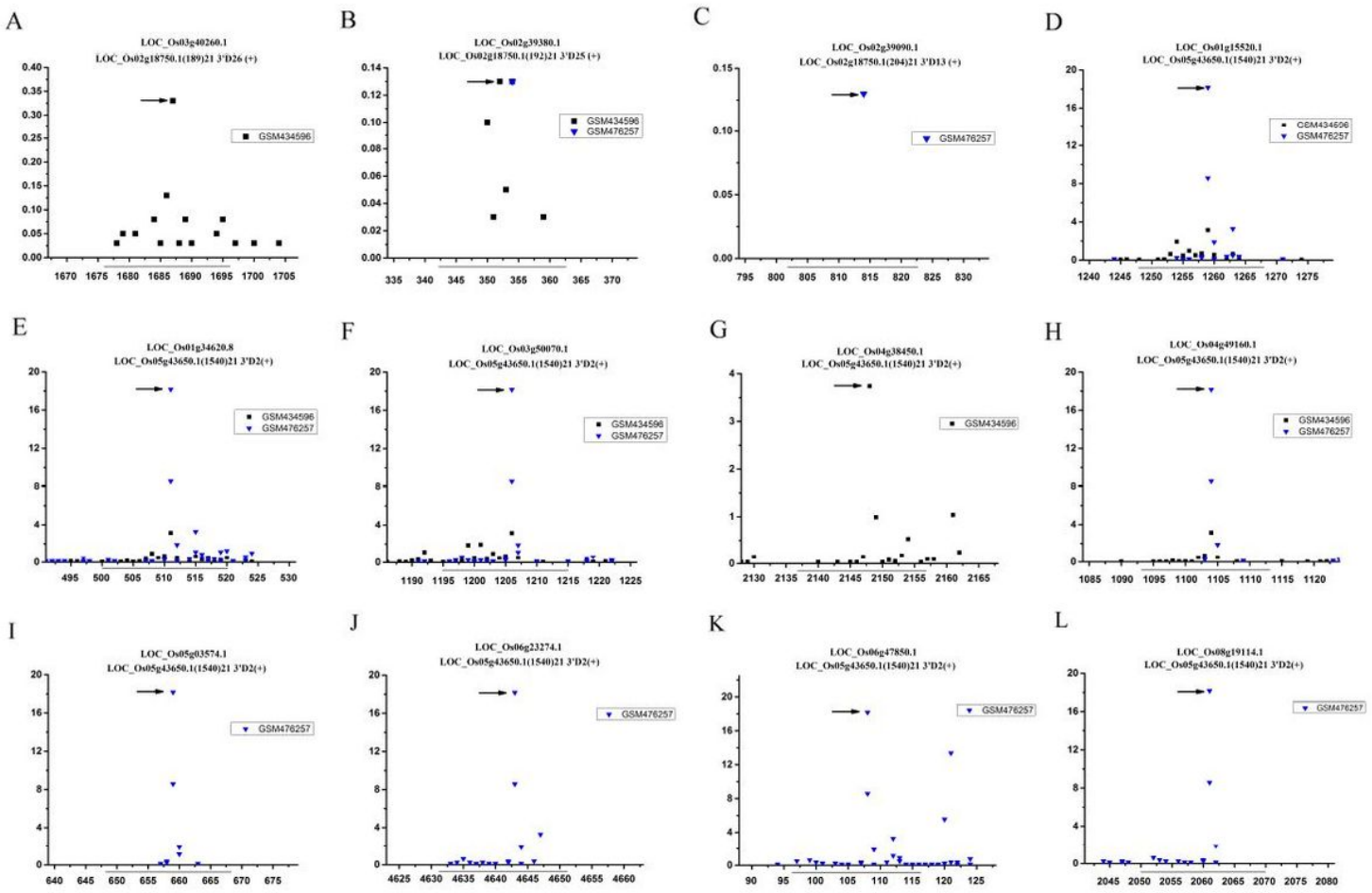

Figure 3

Examples of degradome sequencing-based validation of the phasiRNA-target interactions. Two 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_Os05g43650.1 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.

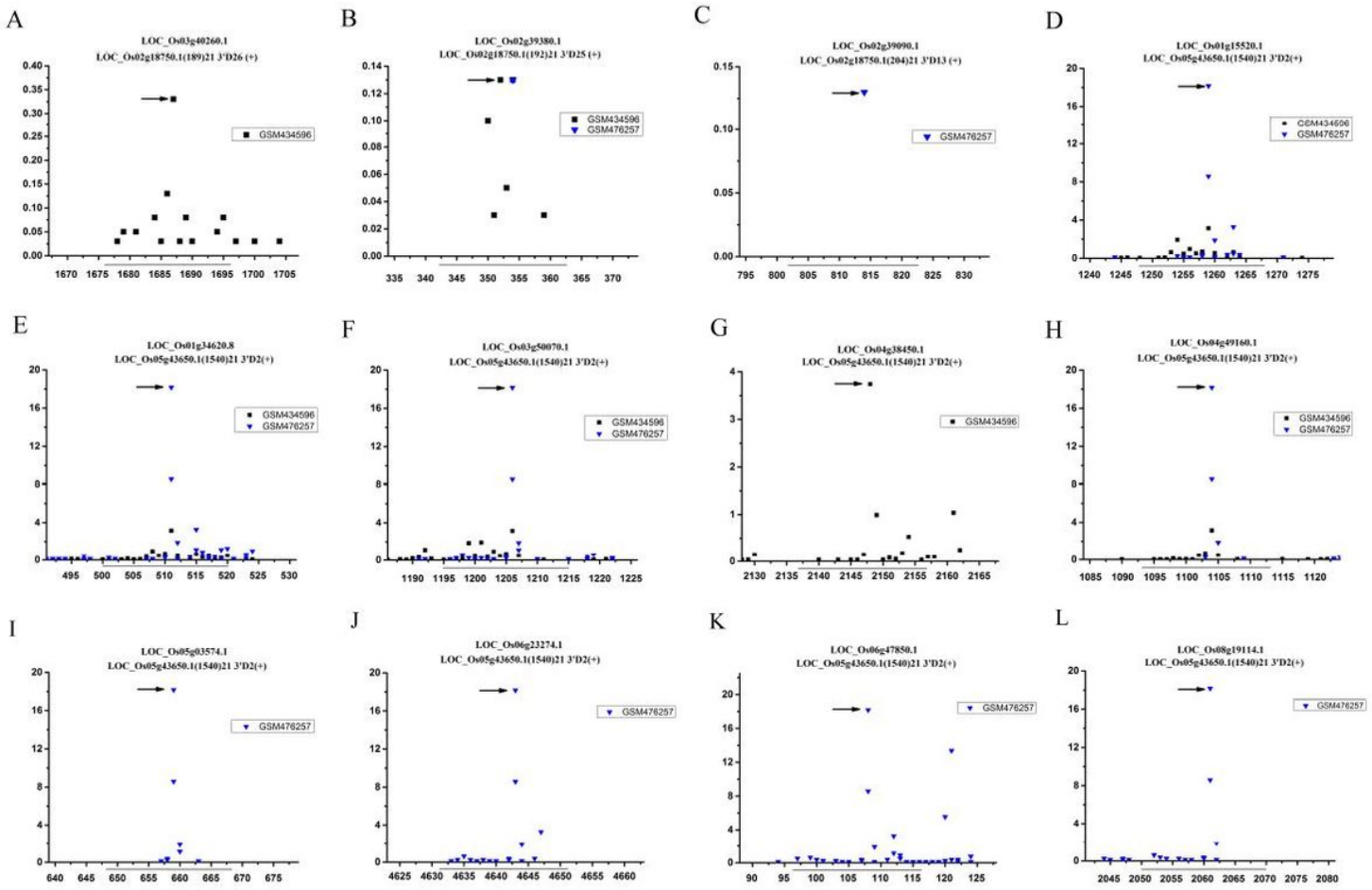

Figure 3

Examples of degradome sequencing-based validation of the phasiRNA-target interactions. Two 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_Os05g43650.1 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.

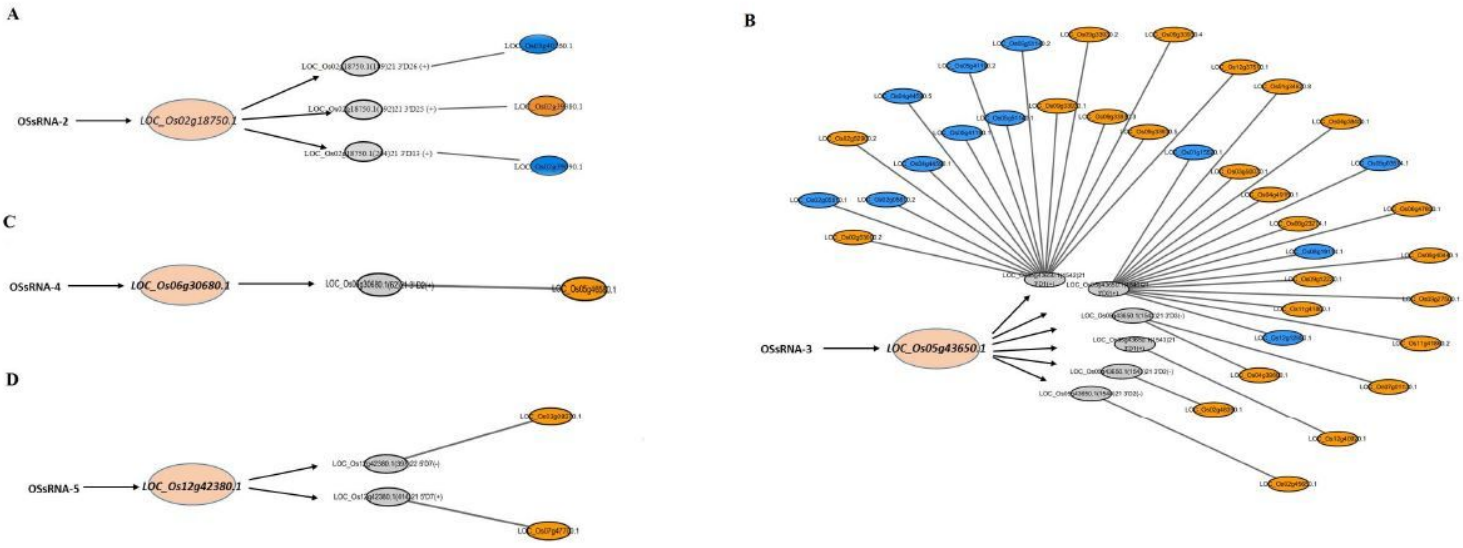

Figure 4

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.

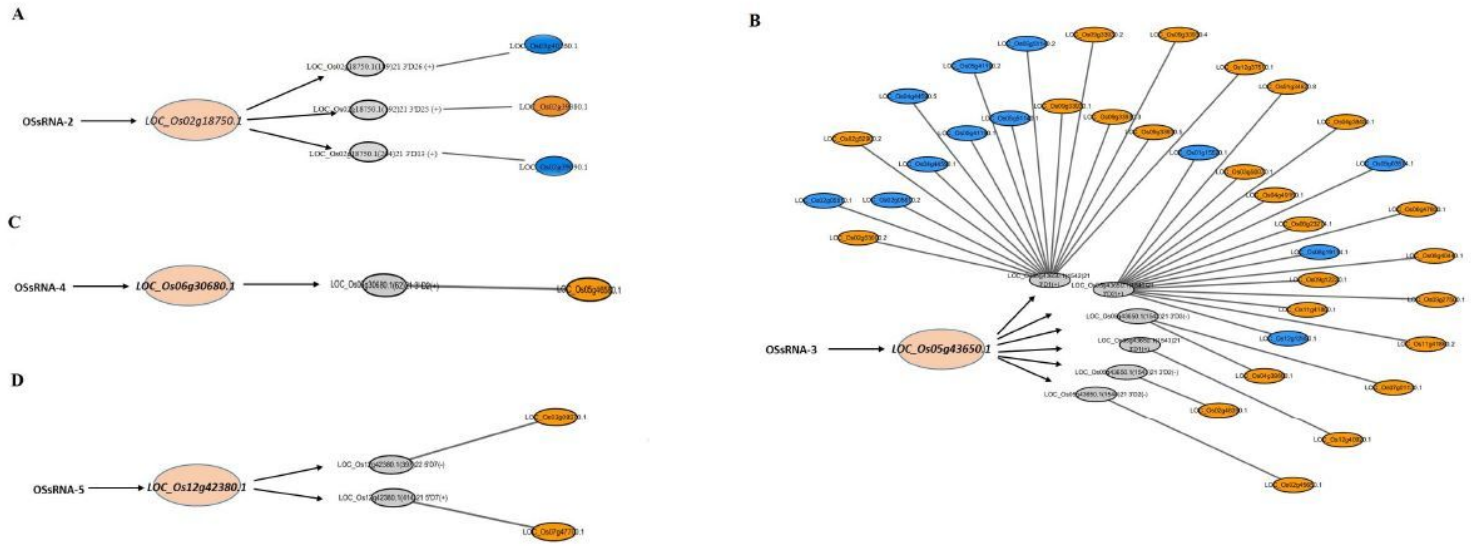

Figure 4

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.

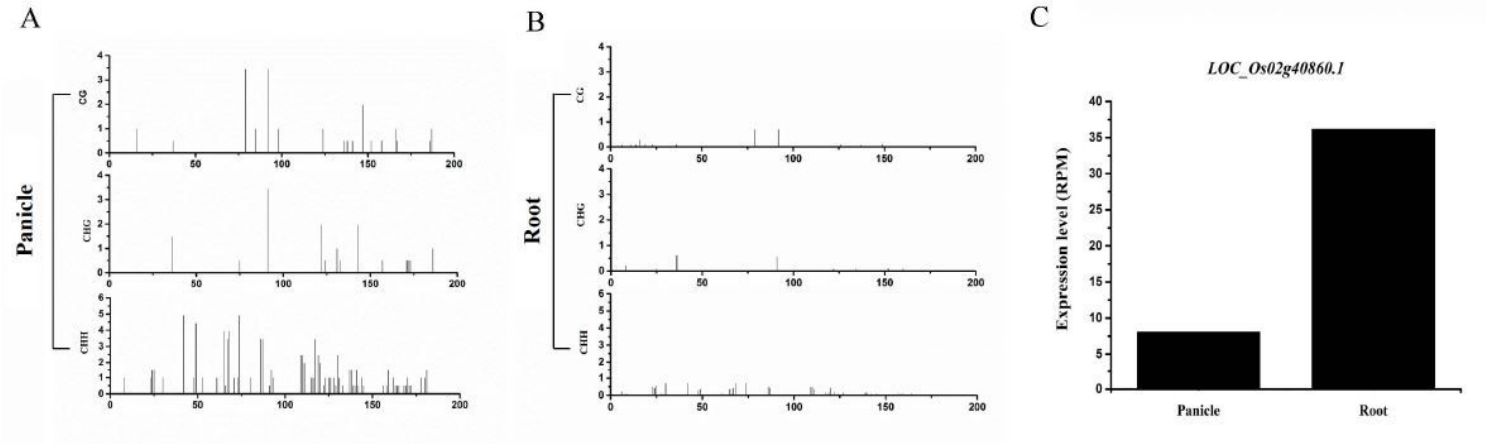

Figure 5

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).

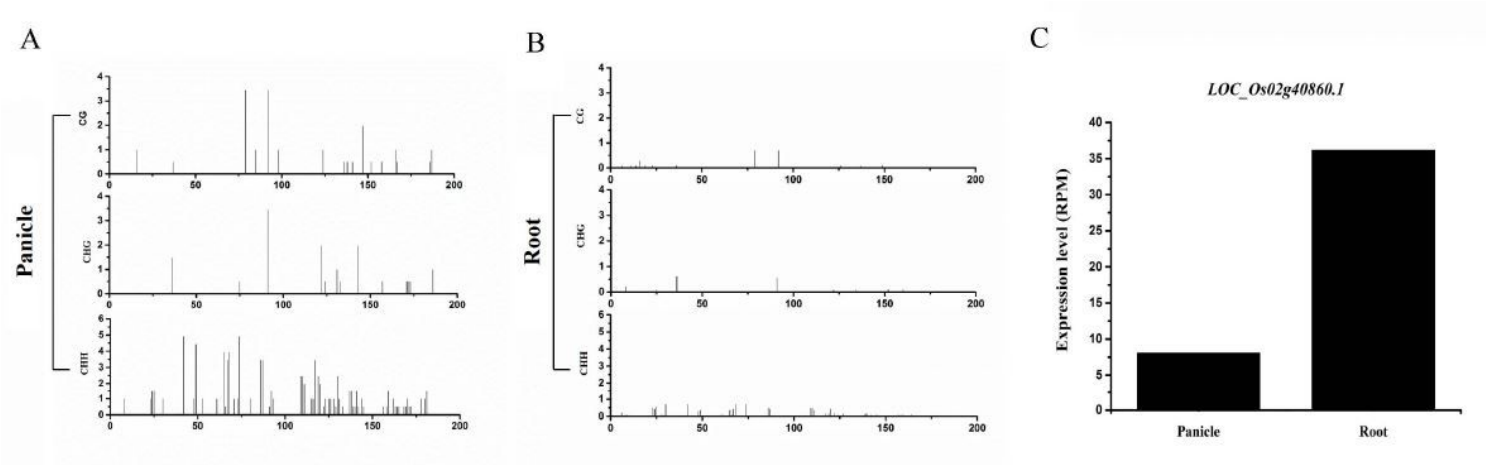

Figure 5

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
- figureS5.pdf
- figureS4.pdf
- figureS4.pdf
- TableS3.xlsx
- TableS3.xlsx
- TableS1revised.xlsx
- TableS1revised.xlsx
- TableS2revised.xlsx
- TableS2revised.xlsx
- figureS1.pdf
- figureS1.pdf
- figureS2.pdf
- figureS2.pdf
- figureS3.pdf
- figureS3.pdf