New Phytologist

***New Phytologist* Supporting Information**

Article title: **Evolution and diversification of reproductive phased small interfering RNAs in *Oryza* species**

Authors: **Peng Tian, Xuemei Zhang, Rui Xia, Yang Liu, Meijiao Wang, Bo Li, Tieyan Liu, Jinfeng Shi, Rod A. Wing, Blake C. Meyers, Mingsheng Chen**

Article acceptance date: 12 October 2020

The following Supporting Information is available for this article:

Fig. S1 *PHAS* loci were mostly in the intergenic regions and 21-*PHAS* have high GC composition than 24-*PHAS*.

Fig. S2 miR2118 could target *NLRs* and produce no or sparse phasiRNAs in *O. sativa*.

Fig. S3 Most 21-*PHAS* and 24-*PHAS* were unique or with low copies, and examples showing 21-*PHAS* loci expanded by segment duplication.

Fig. S4 The syntenic distribution property of *PHAS* on the 11 chromosomes of five *Oryza* genomes.

Fig. S5 Both *pms1* and *pms3* are 21-*PHAS* loci in *O. sativa*.

Fig. S6 21-nt phasiRNA-targets were less stable than miRNA-targets.

Fig. S7 21-nt phasiRNAs with 5′-terminal U tend to induce *cis*-cleavage in the four *Oryza* species.

Fig. S8 The sequences of two 21-*PHAS* loci with *cis*-cleavage sites in *Oryza*.

Fig. S9 The expression level of two potential miR2118 NATs in different tissues.

Fig. S10 DNA methylation level of *PHAS* loci was higher in reproductive yong panicles than vegetative leaves.

Table S1 Summary of reads quantities and genome mapping qualities of sRNA and degradome data for five *Oryza* species.

Table S2 The lists of *PHAS* loci found in the five *Oryza* genomes (see separate file).

Table S3 miR2118 and miR2275 members in the five *Oryza* genomes (see separate file).

Table S4 The lists of 21-nt phasiRNAs targets identified by degradome analysis in the five *Oryza* species (see separate file).

Table S5 The 553 functional 21-nt phasiRNA-targets found in *O. sativa*, and their orthologous 21-nt phasiRNA and targets in the other *Oryza* species (see separate file).

Table S6 The number of total, or subset of the 540 functional 21-nt phasiRNAs, with homology between *O. sativa* and the other four *Oryza* species.

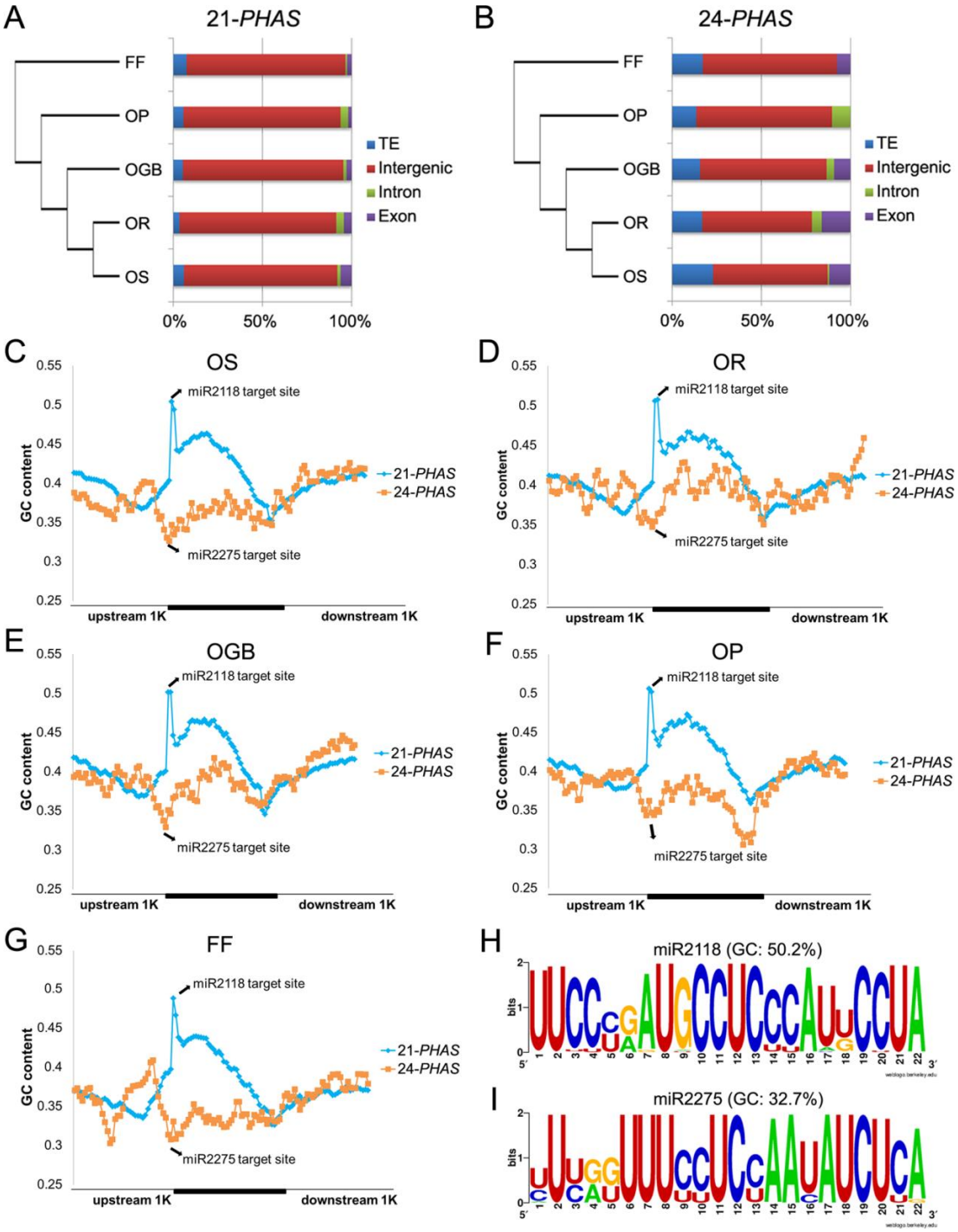


Fig. S1 *PHAS* loci were mostly in the intergenic regions and 21-*PHAS* have high GC composition than 24-*PHAS*. (A-B) The horizontal bars show the ratio of (A) 21-*PHAS* and (B) 24-*PHAS* in different genomic regions. The topology trees on the left show the phylogenetic relationship of the five *Oryza* species. (C-G) The average GC composition at 21-*PHAS*, 24-*PHAS* and their flanking regions in the five *Oryza* genomes. Estimated by sliding-window approach as that in Fig. 4C in the main text. (H-I) Weblogos showing the nucleotide composition of mature (H) miR2118 and (I) miR2275 family.

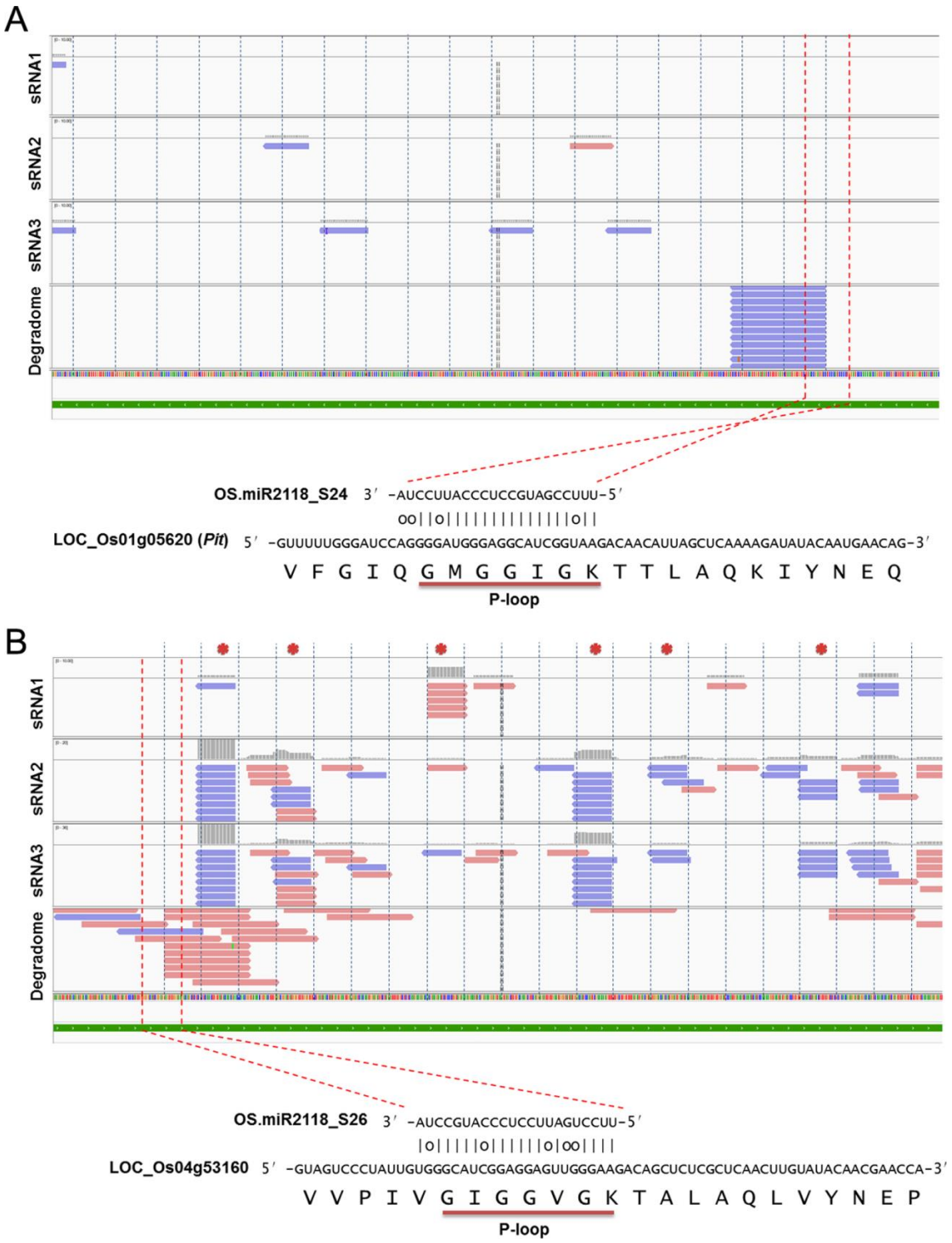


Fig. S2 miR2118 could target *NLRs* and produce no or sparse phasiRNAs in *O. sativa*. (A) Cleavage of miR2118 at LOC_Os01g05620 did not trigger significant phasiRNA synthesis. (B) Cleavage of miR2118 at LOC_Os04g53160 initiated weak phasiRNA synthesis. The red asterisks indicate the signals of 21-nt siRNAs in phase cycles after the miR2118 cleavage site. The sRNA1 to 3 correspond to the three sRNA-seq datasets of *O. sativa* used in this study. The bottom panels show the target sites that are inside the P-loop encoding region.

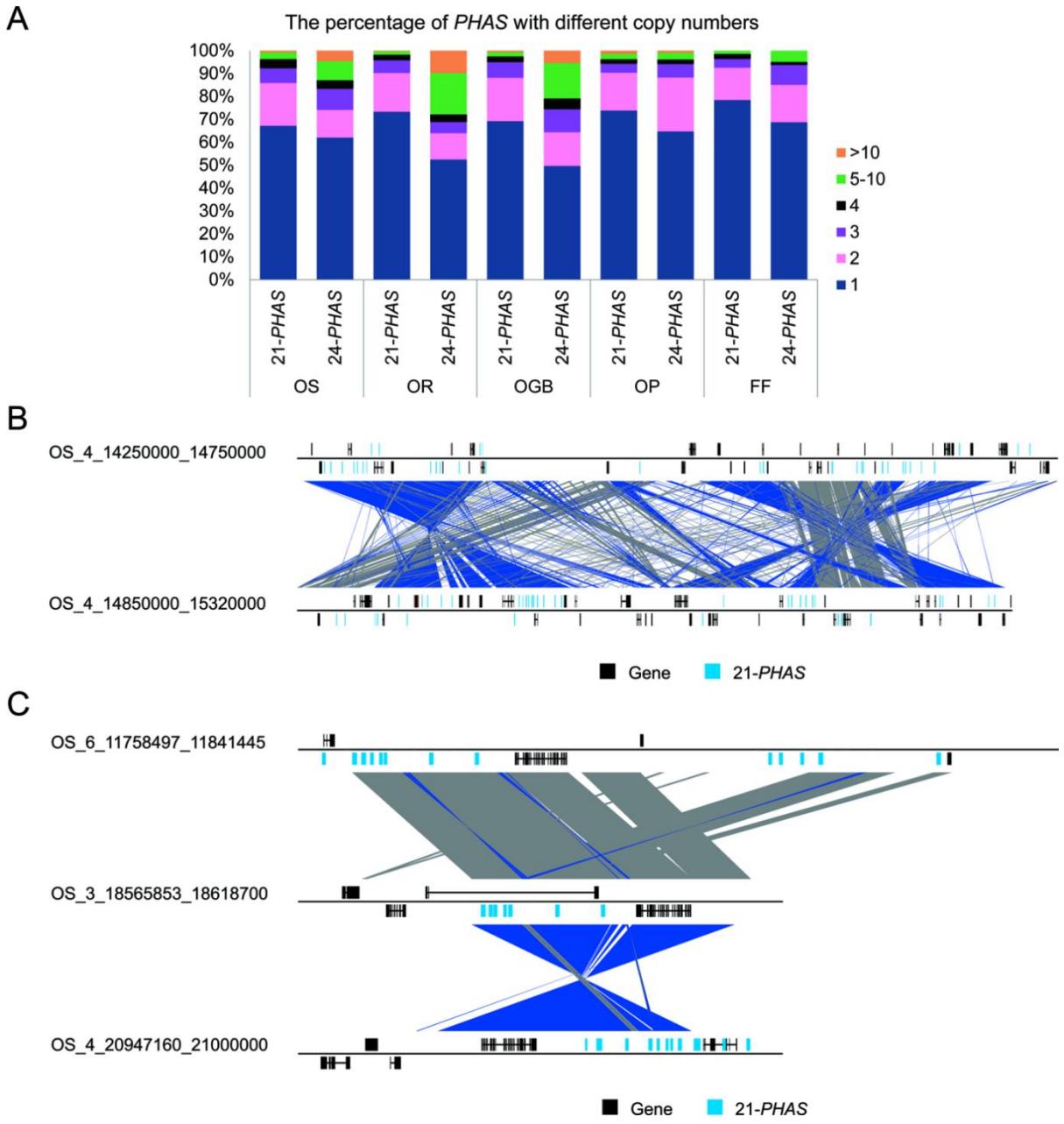


Fig. S3 Most 21-*PHAS* and 24-*PHAS* were unique or with low copies, and examples showing 21-*PHAS* loci expanded by segment duplication. (A) The ratio of 21-*PHAS* and 24-*PHAS* with different number of copies in the five *Orzya* genomes. (B) Example of intrachromosomal segmental duplication of *PHAS* loci. Over thirty 21-*PHAS* loci were expanded in chromosome 4 by duplications between a 500-kb (chromosome 4 (Chr 4), ranged from 14,250,000 bp to 14,750,000 bp) and a 470-kb (Chr 4, ranged from 14,850,000 bp to 15,320,000 bp) segment with distance about 100kb. (C) Example of interchromosomal duplications of *PHAS* loci. The expansion on Chr 6, Chr 3 and Chr 4 were induced by the segment duplication including both protein-coding genes and 21-*PHAS* loci.

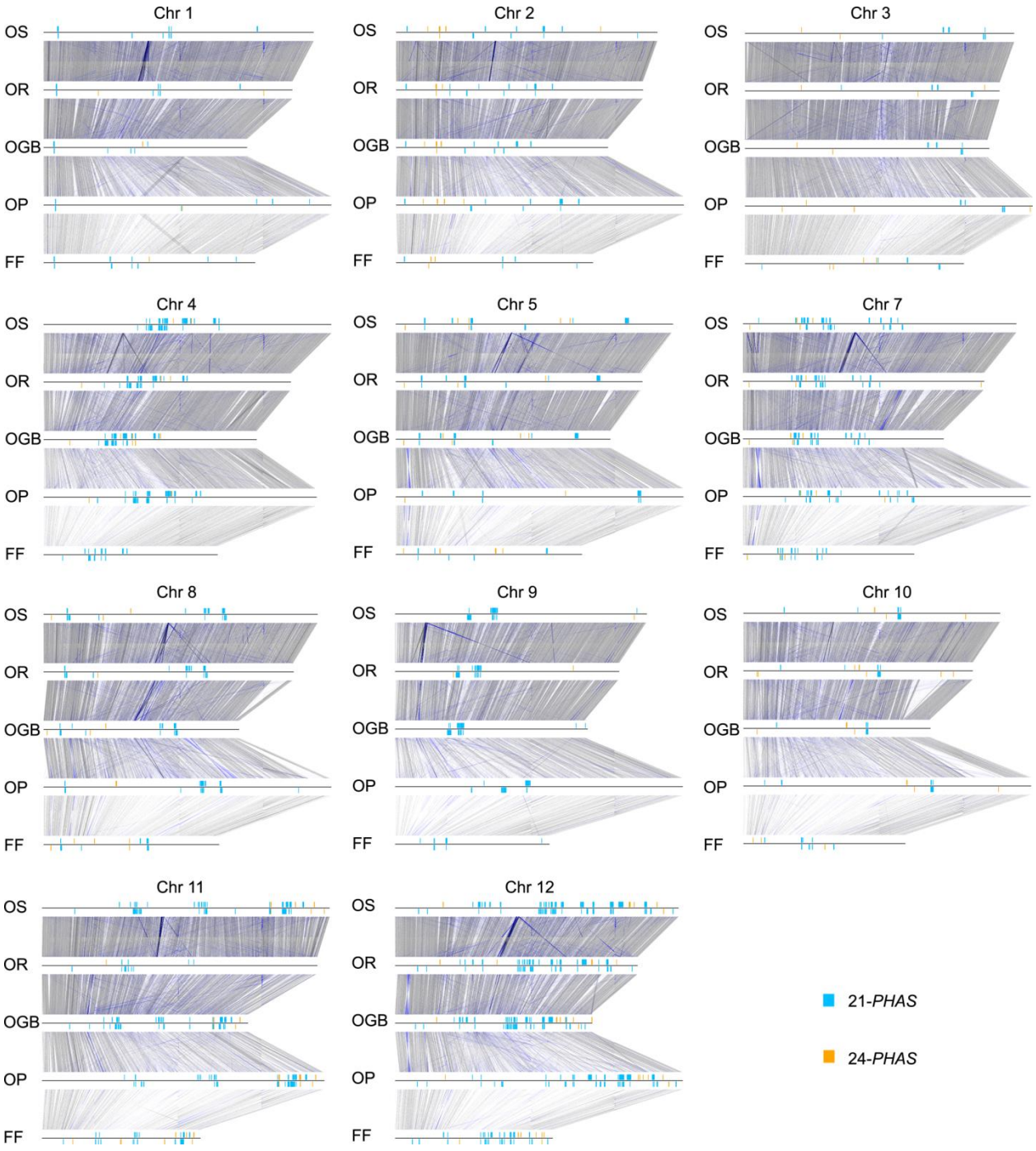
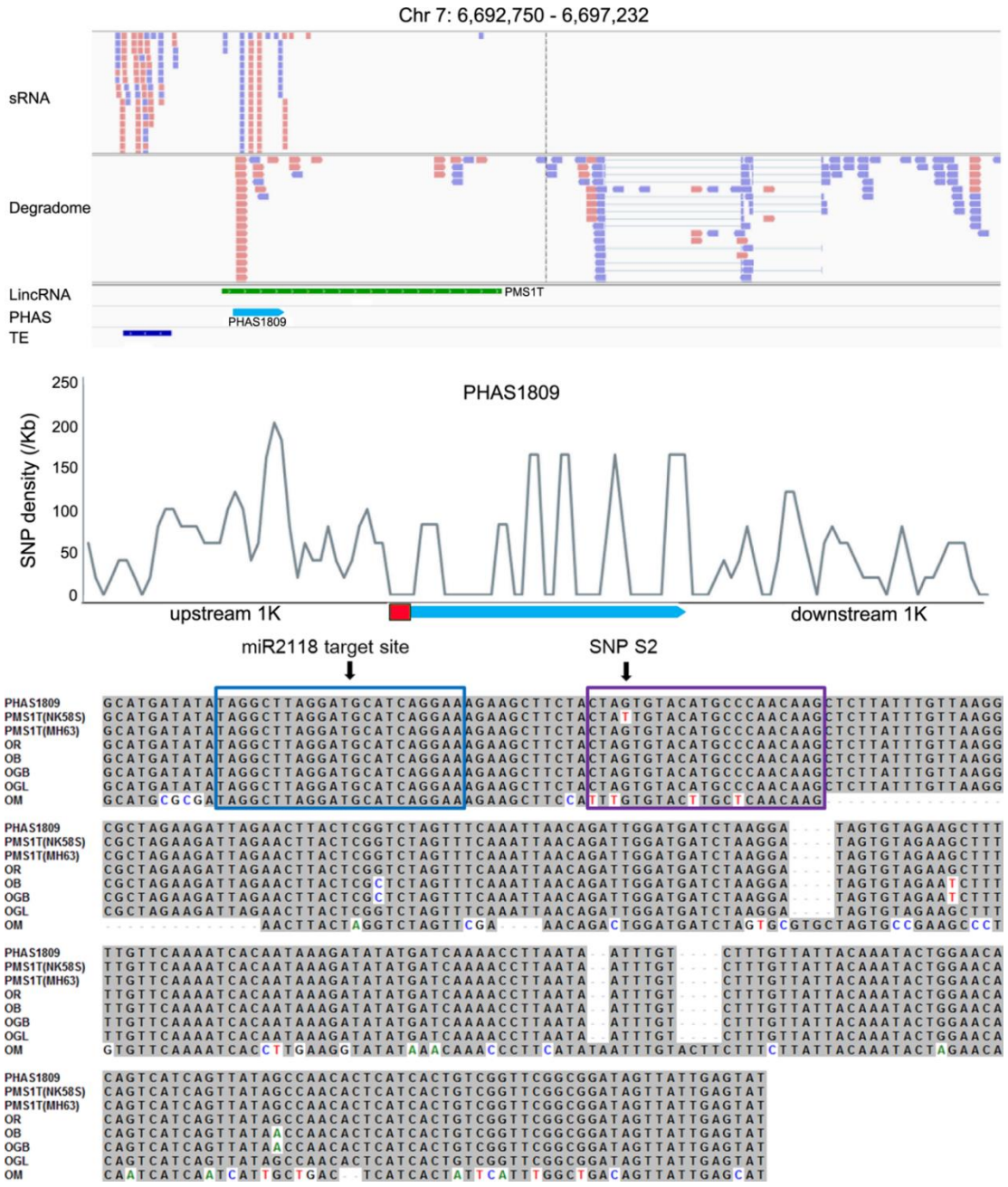
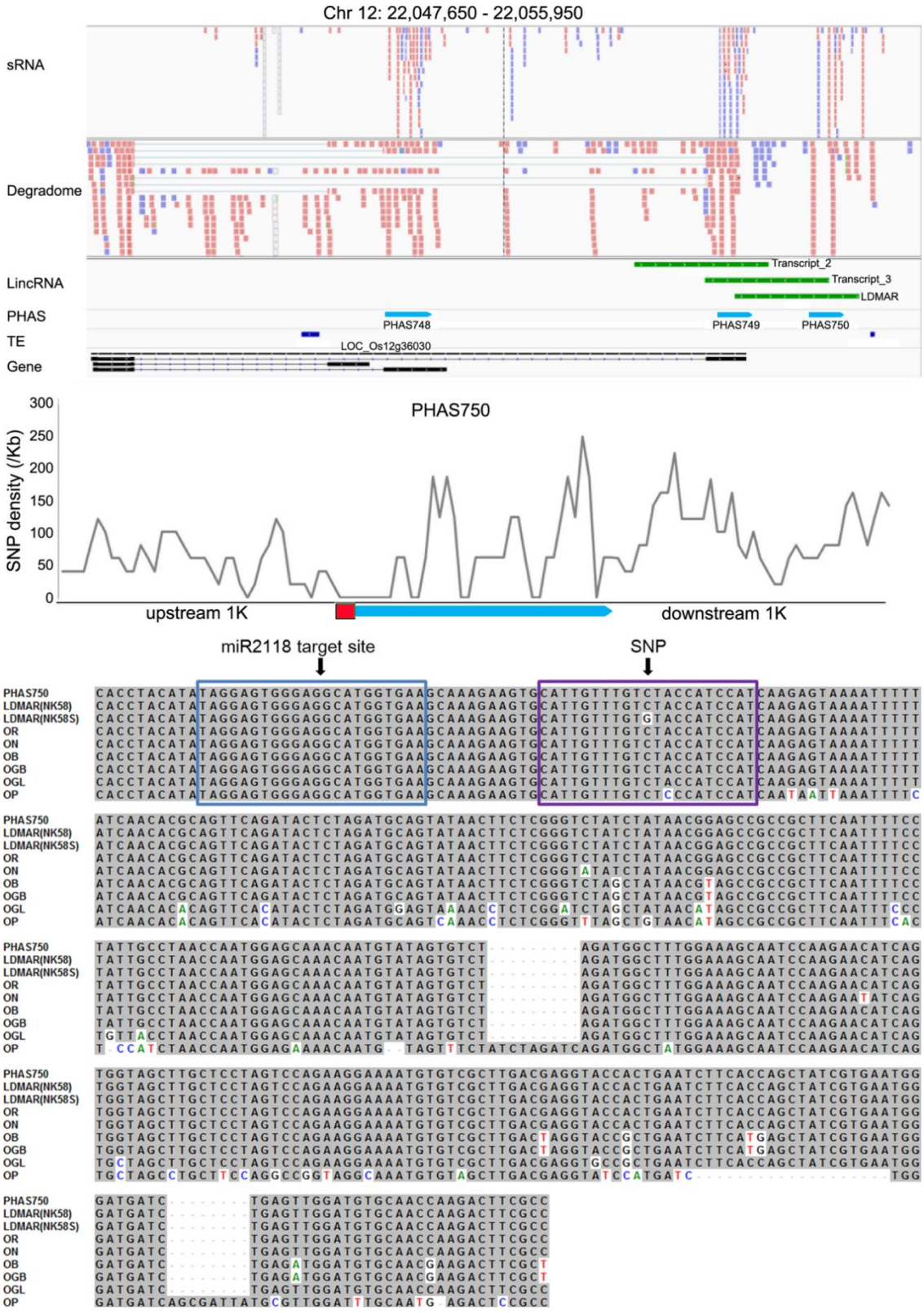


Fig. S4 The syntenic distribution property of *PHAS* on the 11 chromosomes of five *Oryza* genomes.



(Fig. S5A)

Fig. S5 Both *pms1* and *pms3* are 21-*PHAS* loci in *O. sativa*. (A) The *pms1* locus in *O. sativa* and its orthologs in *Oryza* species. (B, see next page) The *pms3* locus in *O. sativa* and its orthologs in *Oryza* species. The lncRNA positions are based on past research on *pms1* (Fan et al., 2016) and *pms3* (Ding et al., 2012). The upper panel is the sRNA and degradome signals at the *pms* region in *O. sativa*. The middle panel shows the SNP density at the pms-*PHAS* and surrounding regions, the red boxes point to the miR2118 target sites. The bottom panel is the sequence alignment of pms-*PHAS* and their orthologs in *Oryza* species. The blue rectangles indicate miR2118 target sites. The purple boxes show the causal SNPs in the second phase cycle. The sequence of *PMS1T* and *LDMAR* are according to the published paper cited above.


Chr 12: 22,047,650 - 22,055,950
sRNA
Degradome
LincRNA
PHAS
TE
Gene
Transcript_2
Transcript_3
LDMAR
PHAS748
PHAS749
PHAS750
LOC_Os12g36030
PHAS750
SNP density (/Kb)
300
250
200
150
100
50
0
upstream 1K
downstream 1K
miR2118 target site
SNP
PHAS750
LDMAR(NK58)
LDMAR(NK58S)
OR
ON
OB
OGB
OGL
OP

(Fig. S5B)

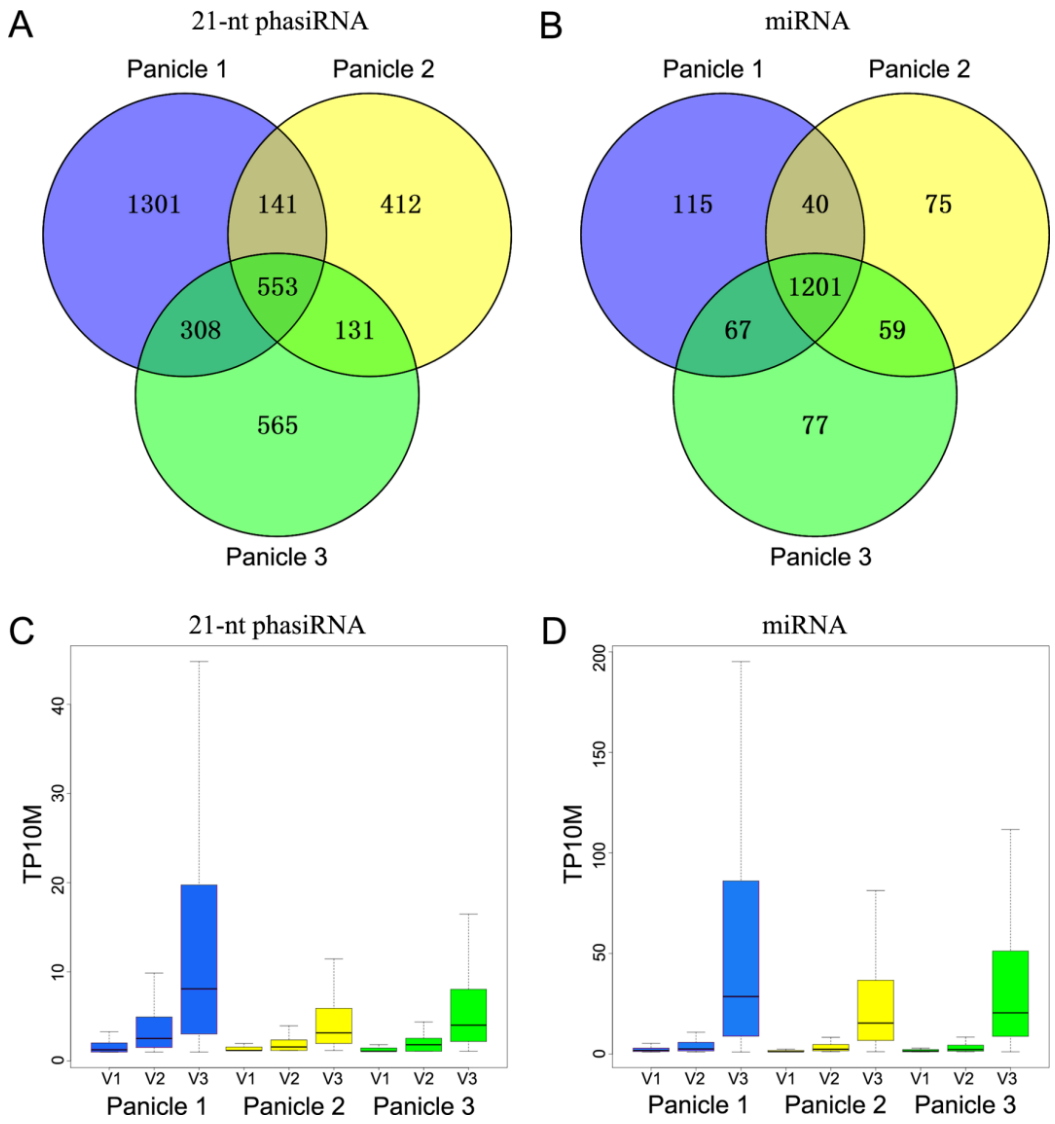


Fig. S6 21-nt phasiRNA-targets were less stable than miRNA-targets. (A-B) Venn diagrams show the number of (A) 21-nt phasiRNA-targets or (B) miRNA-targets found in the three degradome libraries in *O. sativa*. The results showed that many 21-nt phasiRNA-targets were only detected in one library, while miRNA-targets were mostly consistent in the three libraries. (C-D) Boxplots showing the degradome reads abundance at (C) 21-nt phasiRNA or (D) miRNA target sites with different degrees of support. The horizontal lines in the boxes are the median value, and the two bars were the 25th and 75th percentiles, respectively. V1 was supported in one library, V2 in two, and V3 in all three. The results show that sample-specific targets (V1) had less degradome reads than those could be replicated in two (V2) or three samples (V3).

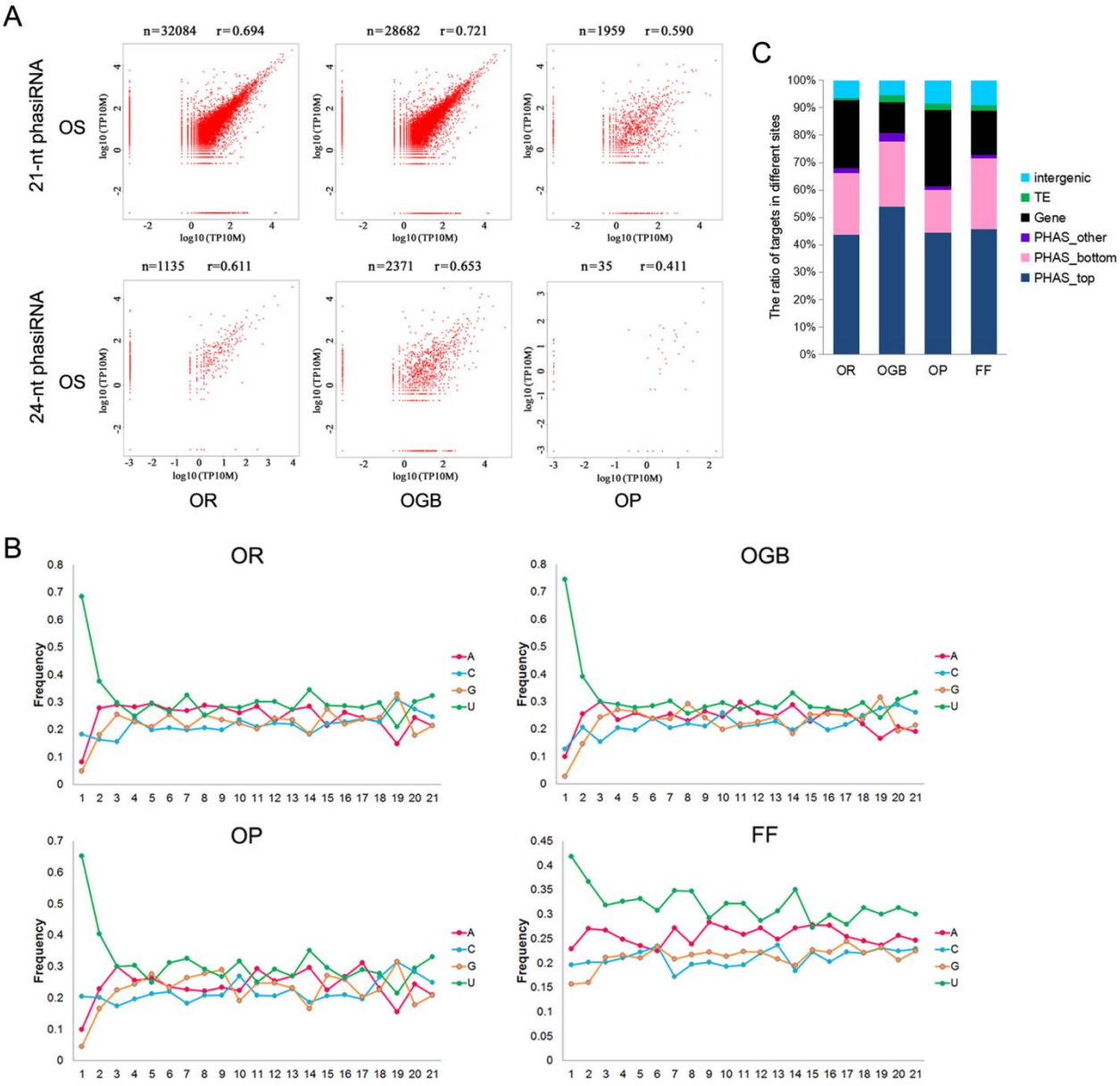


Fig. S7 21-nt phasiRNAs with 5′-terminal U tend to induce *cis*-cleavage in the four *Oryza* species. (A) The expression level of homologous phasiRNAs was in positive correlation between *O. sativa* and the other three *Oryza* species. ‘‘n’’ is the number of homologous phasiRNAs between *O. sativa* and another *Oryza* species. ‘‘r’’ is the Pearson’s correlation coefficient. (B) The Nucleotide composition of 21-nt phasiRNAs with potential cleavage-directing function in the four *Oryza* species. The results show that the 5′-terminal nucleotide was also biased to uridine (U). (C) The ratio of 21-nt phasiRNAs targets in different categories. The results show that most targets are inside *PHAS* regions, suggesting *cis*-activity of 21-nt phasiRNAs.

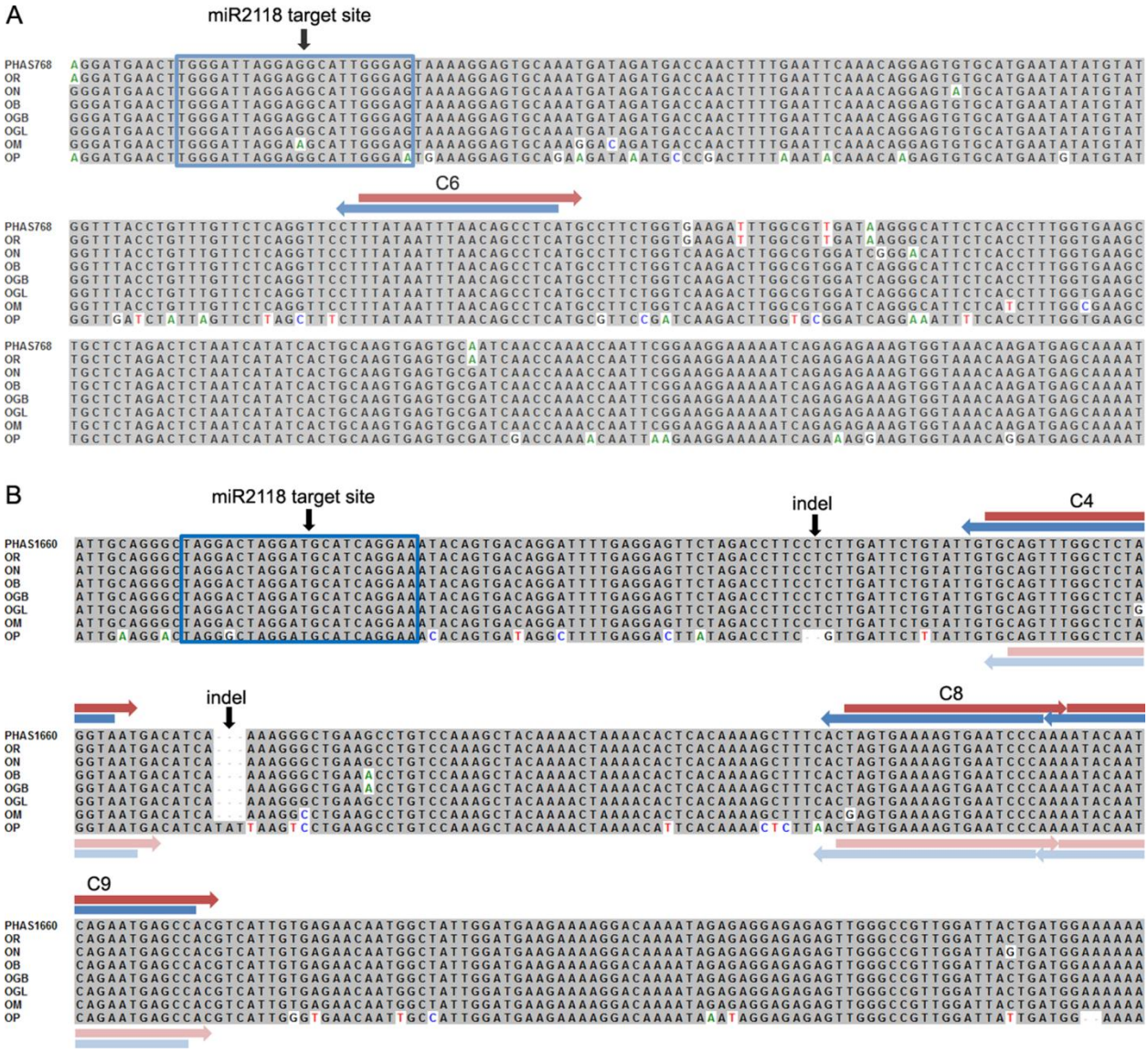


Fig. S8 The sequences of two 21-*PHAS* loci with *cis*-cleavage sites. (A) The sequence alignment of 21-PHAS768 and its orthologs in *Oryza.* The sequences at C6, in which *cis*-cleavage is detected, are conserved in *Oryza* AA group and *O. punctata*. (B) The sequence alignment of 21-PHAS1660 and its orthologs in *Oryza*. *Cis*-cleavage is found in C4, C8 and C9 in the AA group, but not in *O. punctata*. The blue rectangles indicate the miR2118 recognition sites. The red arrows show the phasiRNAs from top-strand, and the blue arrows show the *cis*-cleavage phasiRNAs from bottom-strand, which are complementary to the top-strand sequences showed in this figure. In (B), the light-red or light-blue arrows show the phasiRNAs in *O. punctata* at the three cycles, which are different with that in AA group, because of the shift of phase register induced by the two indels.

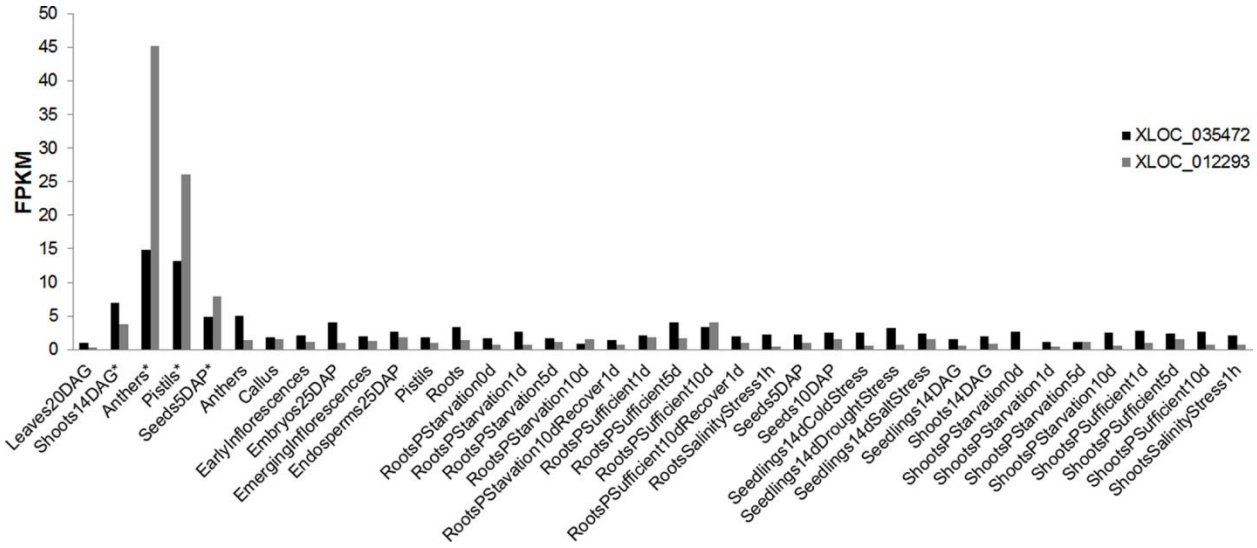


Fig. S9 The expression level of two potential miR2118 NATs in different tissues. The expression data and tissues list are according to a previous study (Zhang *et al.*, 2014).

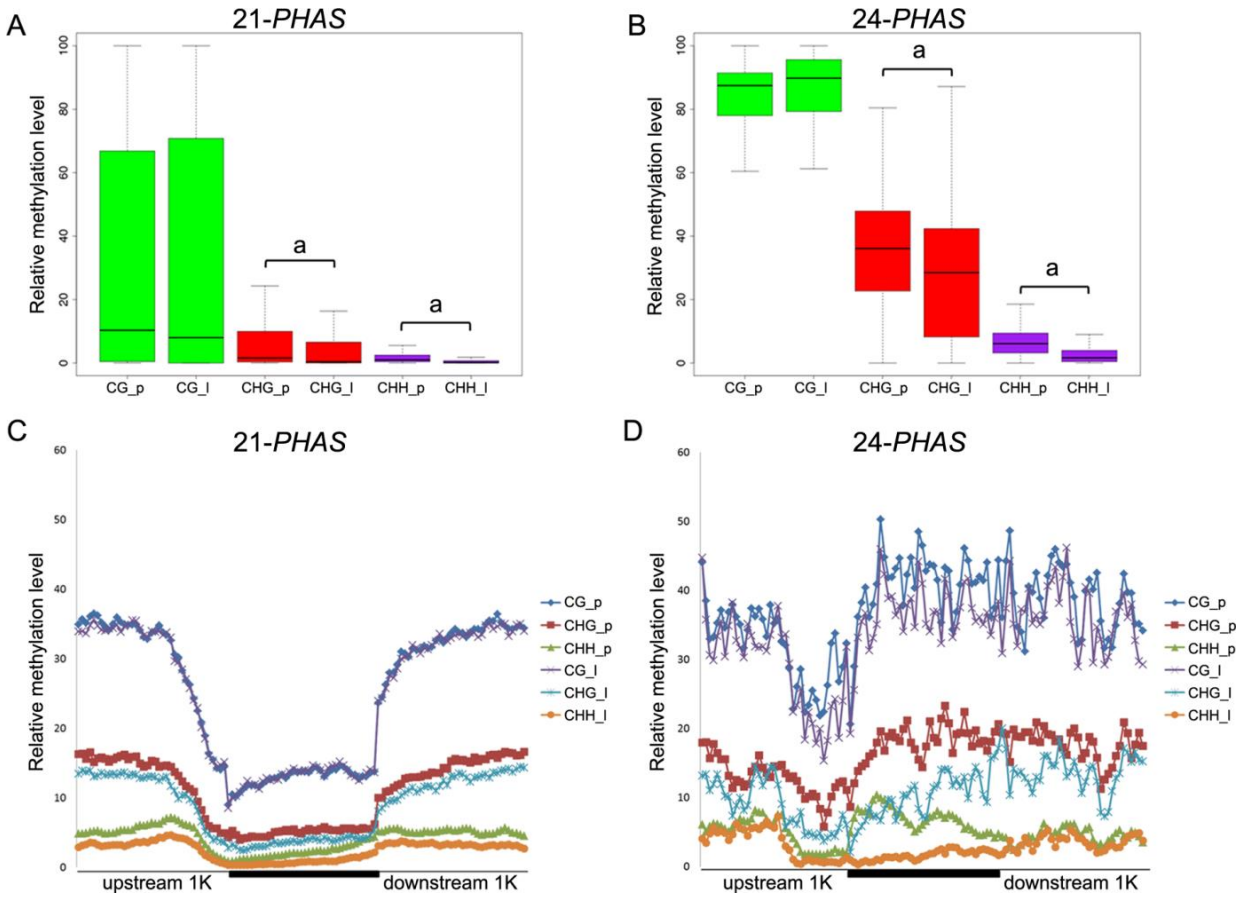


Fig. S10 DNA methylation level of *PHAS* loci was higher in reproductive young panicles than vegetative leaves. (A-B) Boxplots showing the relative DNA methylation level for the three contexts at (A) 21-*PHAS* and (B) 24-*PHAS* regions between young panicle (p) and mature leaf (l) tissues. The horizontal lines in the boxes are the median value, and the two bars were the 25th and 75th percentiles, respectively. "a" above the bars indicated DNA methylation level was significantly higher ($P < 0.001$) in young panicles than that in leaves. (C) DNA methylation pattern at 21-*PHAS* and flanking regions. (D) DNA methylation pattern at 24-*PHAS* and flanking regions. The public MethylC-seq data from mature leaves (Zemach *et al.*, 2010) and young panicles (Li *et al.*, 2012) were analyzed by Bismark (V0.16.3) (Krueger & Andrews, 2011). The methylation levels, which are expressed as percentage (%) per site for each of the three contexts of cytosines, CG, CHG and CHH (H = A, C or T), were calculated by dividing the number of non-converted (methylated) cytosines by the total number of cytosines within the assay. For the methylation level in a region, we used relative methylation (total methylation level of a methylation type divided by total number of that type in the calculated region). These distribution patterns were also estimated by sliding-window approach as that in Fig. 4C in the main text.

Table S1 Summary of reads quantities and genome mapping qualities of sRNA and degradome data for five *Oryza* species.

(a) sRNA data

Species	Total reads	Unique reads	Mapped reads	Mapped Unique	Mapped rate (%)
O. sativa (OS_1)	53,592,035	8,607,903	45,843,767	6,095,773	85.54
O. sativa(OS_2)	34,761,897	12,944,030	30,232,877	10,655,419	86.97
O. sativa (OS_3)	43,789,243	15,561,372	37,372,408	12,829,056	85.35
O. rufipogon	37,116,354	5,374,164	26,875,448	3,470,147	72.41
O. glaberrima	49,124,855	8,312,321	35,880,673	4,926,713	73.04
O. punctata	49,266,270	6,707,649	39,585,162	4,163,999	80.35
O. brachyantha	63,413,565	10,240,267	52,115,087	7,957,886	82.18

(b) Degradome data

Species	Total reads	Unique reads	Mapped reads	Mapped Unique	Mapped rate (%)
O. sativa (OS_1)	48,257,487	10,915,380	39,506,975	8,742,647	81.87
O. sativa (OS_2)	28,875,049	6,372,637	25,347,476	4,982,673	87.78
O. sativa (OS_3)	31,355,274	7,357,302	27,332,647	5,744,287	87.17
O. rufipogon	50,646,433	10,069,972	39,471,470	7,395,169	77.94
O. glaberrima	51,089,304	11,671,252	33,982,894	8,054,908	66.52
O. punctata	47,150,960	10,653,362	37,944,234	8,093,196	80.47
O. brachyantha	25,956,929	8,164,563	17,907,420	6,001,480	68.99

Table S2 The lists of *PHAS* loci found in the five *Oryza* genomes (see separate excel-file).

Table S3 miR2118 and miR2275 members in the five *Oryza* genomes (see separate excel-file).

Table S4 The lists of 21-nt phasiRNAs targets identified by degradome analysis in the five *Oryza* species (see separate excel-file).

Table S5 The 553 functional 21-nt phasiRNA-targets found in *O. sativa*, and their orthologous 21-nt phasiRNAs and targets in the other *Oryza* species (see separate excel-file).

Table S6 The number of total, or subset of the 540 functional 21-nt phasiRNAs, with homology between *O. sativa* and the other four *Oryza* species.

	Total 21-nt phasiRNAs with homology [a]	Functional 21-nt phasiRNAs with homology	Target site consistency [b]	Fisher's exact test P-value [c]
O. rufipogon	32,084 / 41,440	309	202	2.86E-11
O. glaberrima	28,682 / 55,134	295	238	2.44E-14
O. punctata	1959 / 44,854	21	5	0.052
O. brachyantha	2 / 36,296	0	NA	NA

a: The first number is the shared homologous 21-nt phasiRNAs between this species and *O. sativa*, and the second number is the total 21-nt phasiRNAs counted in this species. b: The number of target sites consistent between this species and *O. sativa*, for the subset of the 540 functional 21-nt phasiRNAs. c: The Fisher's exact test was used to estimate whether the 540 functional 21-nt phasiRNAs in *O. sativa* tend to be found as homologous in this species. "NA", not available.

References

Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q. 2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci USA* **109**: 2654-2659.

Fan Y, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, et al. 2016. *PMS1T*, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proc Natl Acad Sci USA* **113**: 15144-15149.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, et al. 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* **13**: 300.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916-919.

Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, Chen YQ. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* **15**: 512.