# CRISPR/Cas9-targeted mutagenesis of *TaDCL4*, *TaDCL5* and *TaRDR6* induces male sterility in common wheat

Rongzhi Zhang[1,2,3,†], Shujuan Zhang[1,2,3,†] (iD), Jihu Li[1,2,3], Jie Gao[1,2,3], Guoqi Song[1,2,3], Wei Li[1,2,3], Shuaifeng Geng[4], Cheng Liu[1,2,3], Yanxiang Lin[1,2,3], Yulian Li[1,2,3,*] and Genying Li[1,2,3,*] (iD)

[1]*Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China*

[2]*Ministry of Agriculture, Key Laboratory of Wheat Biology and Genetic Improvement on North Yellow and Huai River Valley, Jinan, China*

[3]*National Engineering Research Center for Wheat and Maize, Jinan, China*

[4]*National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China*

## Summary

Phased, small interfering RNAs (phasiRNAs) are important for plant anther development, especially for male sterility. PhasiRNA biogenesis is dependent on genes like RNA polymerase 6 (RDR6), DICER-LIKE 4 (DCL4), or DCL5 to produce 21- or 24 nucleotide (nt) double-strand small RNAs. Here, we generated mutants of *DCL4*, *DCL5* and *RDR6* using CRISPR/Cas9 system and studied their effects on plant reproductive development and phasiRNA production in wheat. We found that *RDR6* mutation caused sever consequence throughout plant development starting from seed germination and the *dcl4* mutants grew weaker with thorough male sterility, while *dcl5* plants developed normally but exhibited male sterility. Correspondingly, DCL4 and DCL5, respectively, specified 21- and 24-nt phasiRNA biogenesis, while RDR6 contributed to both. Also, the three key genes evolved differently in wheat, with *TaDCL5-A/B* becoming non-functioning and *TaRDR6-A* being lost after polyploidization. Furthermore, we found that *PHAS* genes (phasiRNA precursors) identified *via* phasiRNAs diverged rapidly among sub-genomes of polyploid wheat. Despite no similarity being found among phasiRNAs of grasses, their targets were enriched for similar biological functions. In light of the important roles of phasiRNA pathways in gametophyte development, genetic dissection of the function of key genes may help generate male sterile lines suitable for hybrid wheat breeding.

## Introduction

Wheat yield and sustainable production are important issues for global food safety. Hybrid vigour is an important consideration in increasing crop yield. Male sterile (MS) mutants are appropriate materials for understanding the molecular mechanisms underlying fertility regulation in wheat and are potential germplasms for hybrid seeds production. Depending on the fertility gene sources, MS includes cytoplasmic male sterility (CMS) and genic male sterility (GMS). The three line-system, composed of the CMS line, the maintainer line and the restorer line, has been successfully used for producing hybrid wheat. However, the establishment of the three lines is a time-consuming and cumbersome process, along with unstable or incomplete male sterility, which limited the use of the CMS technology in commercial wheat seed industry (Geyer *et al.*, 2018; Liu *et al.*, 1997, 2002). In contrast, genic male sterile (GMS) mutants can be restored by any wild-type germplasms and thus have a broader choice of paternal lines to produce hybrid wheat of superior heterosis. In photoperiod and/or thermosensitive GMS (PTGMS), the same line serves as both the MS and maintainer lines under different environmental conditions. Despite of the lower cost of seed production in PTGMS, the system is easily affected by unpredictable environmental conditions, which will affect the propagation of sterile lines and maintainer lines and hybrid seed production (Chen *et al.*, 2005). The propagation of recessive GMS lines and their maintainers could be solved through seed production technology (SPT) (Chang *et al.*, 2016; Wu *et al.*, 2016; Zhang *et al.*, 2018a).

But it requires multiple laborious and complicated steps. Qi *et al.* (2020) designed a simple next-generation hybrid seed production strategy that takes advantage of the CRISPR/Cas9 technology to create a manipulated GMS maintainer (MGM). This strategy can generate a GMS line and its maintainer line simultaneously. It paves the way for efficient breeding schemes using the next-generation hybrid crop SPT.

In common wheat, the complex allohexaploid genome have limited our understanding of the genetic and molecular mechanisms of fertility. Until now, only six GMS genes in wheat were demonstrated to confer male sterility. *TaMS2* is a dominant GMS gene, encoding an orphan protein. Its activation confers male sterility in grass species (Ni *et al.*, 2017; Xia *et al.*, 2017). The other five GMS genes are recessive, including *TaMS1*, *TaMS5*, *TaMS26*, *TaMS45* and *TaNP1*. *TaMS1* encodes a glycosylphosphatidylinositol-anchored lipid transfer protein, which is necessary for pollen exine development (Tucker *et al.*, 2017; Wang *et al.*, 2017). *TaMs5* as a glycosylphosphatidylinositol-anchored lipid transfer protein is required for normal pollen exine development (Pallotta *et al.*, 2019). *TaMs26* gene plays an important role in pollen development in sporopollenin biosynthesis for pollen exine formation (Singh *et al.*, 2017). *TaMs45* gene encodes a strictosidine synthase-like enzyme and is required for microspore cell wall development (Singh *et al.*, 2018). *TaNP1* encodes a putative glucose-methanol-choline oxidoreductase, and functions in tapetum degeneration and pollen exine formation (Li *et al.*, 2020). Mutating any of the five genes, that is, *TaMS1* in chromosome

(chr) 4B, *TaMS5* in chr3D, the three copies of *TaMS26*, *TaMS45* and *TaNP1*, could confer male sterility. In addition to these coding genes, numerous 21-nt and 24-nt phased interfering small RNAs (phasiRNAs), deriving from hundreds of genomic loci, were identified in pre-meiotic, meiotic and post-meiotic anthers of wheat that may be relevant to male fertility (Bélanger *et al*., 2020; Zhang *et al*., 2020a).

PhasiRNAs are produced from the long noncoding RNAs at *PHAS* loci. The phase is usually determined by 22-nt miRNAs that cut *PHAS* RNA at recognition sites, which trigger conversion of the RNA into double-strand RNAs by RNA-dependent RNA polymerase 6 (RDR6). The subsequent cleavage steps are carried out by DICER-LIKE 4 (DCL4) and DCL5 (also named DCL3b) producing 21-nt and 24-nt phasiRNAs, respectively. The trigger miRNAs were miR2118 or miR11308 for 21-nt phasiRNAs and miR2275, tae_cand1 or miR14051 for 24-nt phasiRNAs (Bélanger *et al*., 2020; Pokhrel *et al*., 2021a, 2021b; Song *et al*., 2012a, 2012b). The phasiRNAs were first reported in grasses (Johnson *et al*., 2009) and then discovered in eudicots (Pokhrel *et al*., 2021a, 2021b). These non-coding RNAs played important roles in male gametophyte development. For example, the mutated target site of *PMS1T* escaping the regulation of miR2118 produces no phasiRNA, which led to male sterility (MS) in rice (Fan *et al*., 2016). In addition, mutating key genes in the phasiRNA biogenesis pathways also induce anther defect and MS. In rice, for example, loss of function of *OsDCL4* led to male sterility (Liu *et al*., 2007; Zhang *et al*., 2020b), and generated few or no 21-nt phasiRNAs (Song *et al*., 2012a). The maize *dcl5* mutant had short anthers with defective tapetal cells and produced nearly no 24-nt phasiRNAs (Song *et al*., 2012a; Teng *et al*., 2020). The *Osrdr6* mutant was also male sterile and had a strong impact on the accumulation of 21-nt and 24-nt phasiRNAs (Song *et al*., 2012b). Therefore, 21-nt or 24-nt phasiRNAs are deeply involved in grass reproductive processes.

The rice reproductive 21-nt phasiRNAs are essential for the elimination of a specific set of RNAs during the stamen development. Recent studies suggested that 21-nt phasiRNAs, by direct cleavage of target mRNAs, may have similar regulatory modes to that of miRNAs (Jiang *et al*., 2020; Zhang *et al*., 2020b). These target mRNAs appeared to be enriched for carbohydrate biosynthetic and metabolic pathways (Jiang *et al*., 2020). The 24-nt phasiRNAs contributed to increase CHH methylation in *cis* at 24-*PHAS* loci in maize anther at meiosis (Zhang *et al*., 2021). Thus, reproductive phasiRNAs play a role in fine-tuning gene expression in the post- or pre-transcription level.

Common wheat (*Triticum aestivum*) is a hexaploid plant (2n = 6x = 42, BBAADD) that contains three closely related sub-genomes, A, B and D, which were reunited by two polyploidization events of three diploid species. *T. turgidum* (2n = 4x = 28, AABB) was synthesized by *T. urartu* (2n = 2x = 14, AA) and an unknown species similar to current *Aegilops speltoides* (2n = 2x = 14, BB). *T. turgidum* was then united with *Aegilops tauschii* (2n = 2x = 14, DD) *via* a more recent allohexaploid speciation event about 8000–10 000 years ago. Estimation using whole-genome single-nucleotide polymorphism (SNP) showed that 0.68–4.9 million years ago (mya) divergence time among the three sub-genomes and their donors (A, 1.15 mya; D, 0.68 mya; B, <4.49 mya) (Li *et al*., 2022). The genome availability of these polyploid wheat and their progenitors allows systematical study of the evolution of phasiRNAs in *Triticum* species.

Despite that phasiRNAs were abundant in developing wheat anthers (Bélanger *et al*., 2020; Zhang *et al*., 2020a), their

biogenesis and function remain unclear. Here, we generated CRISPR/Cas9 mutants of the three key genes for phasiRNA production: *DCL4*, *DCL5* and *RDR6*, and studied for the first time their functions of phasiRNA biogenesis in wheat. We also investigated the pattern of evolution of *PHAS* loci during wheat polyploidization. Our work provided useful information regarding using the phasiRNA pathway components for male sterile line generation and hybrid breeding in wheat.

## Results

### Key genes involved in phasiRNA biogenesis in wheat

The key genes that are known in other grasses for 21-nt and 24-nt phasiRNA productions are DCL4 and DCL5, respectively, which cleave the double stranded *PHAS* RNA produced by RDR6 (Song *et al*., 2012a, 2012b). As a hexaploid, wheat has three copies (homologues) of *DCL4* and *DCL5*, while only two homologues (*TaRDR6-B* and *TaRDR6-D*) were present for *RDR6* (Figure 1a and Figure S1). Despite this, the A genome donor conferred a *RDR6* gene, indicating the A homologue in hexaploid may be lost upon hexaploidization. Protein sequences of these genes were conserved among the sub-genomes of the polyploids and their progenitors (Figures S1 and S2). TaRDR6 showed similar expression patterns to that of TaDCL4, which can be detected in both vegetative and reproductive stages, with much higher level in the latter stage (Figure S3a,b). TaDCL5, however, was detected only in tissues at reproductive stage, such as young spikes and anther, with the D homologue having the highest expression level among the three sub-genomes (Figure S3c). qRT-PCR assay showed that TaDCL4 and TaRDR6 were most highly expressed in 0.2-mm anther, whereas the highest expression was found between 0.8- and 1.0-mm anthers for TaDCL5 (Figure 1b), indicating their different roles in wheat anther development.

### Creation of the *dcl4*, *dcl5* and *rdr6* mutants by genome editing

Three single guide RNAs (sgRNAs) targeting the conserved domains of their homologues were designed at the DEAD/DEAH box helicase of TaDCL4, the PAZ domain for TaDCL5, and the RNA dependent RNA polymerase domain of TaRDR6 (Figure 1a). The CRISPR/Cas9 constructs were transformed into wheat using the *Agrobacterium tumefaciens*-mediated approach and positive plants were confirmed by Hi-TOM sequencing. Totally, 19, 5 and 9 $T_0$ plants were mutated in the genome level for *TaDCL4*, *TaDCL5* and *TaRDR6*, respectively (Figure 1c and Tables S1–S3). However, the mutated plants contained homozygous genotype such as aabbdd of *dcl4-17* or bbdd of *rdr6-13* that were fertile. The 3-bp deletion or its integral number of times (Tables S2 and S3) of these plants that may have relatively little effect on the protein structure. Thus, in this study, we defined the mutations that only consisted of early stop codons or frameshift mutations. Then, we selected the $T_0$ plants (*dcl4-17*, *dcl4-31*, *dcl5-4*, and *rdr6-13*) with higher proportion of mutations to propagate their progenies (Tables S1–S3). At $T_1$ or $T_2$ generation, mutants of the three genes showed male sterility, and their gene expression levels were much lower than WT (Figure 1d and Tables S1–S3). All the mutants used in the further study had the MS phenotype. The details for creating these mutants were shown in the Supporting Materials and Methods.

The $T_0$ plant of *dcl5-4* was completely male sterile and had to cross with the wild type Fielder to maintain heterozygous seeds. (Table S1 and Figure 1c). The MS phenotypes in *dcl5* happened
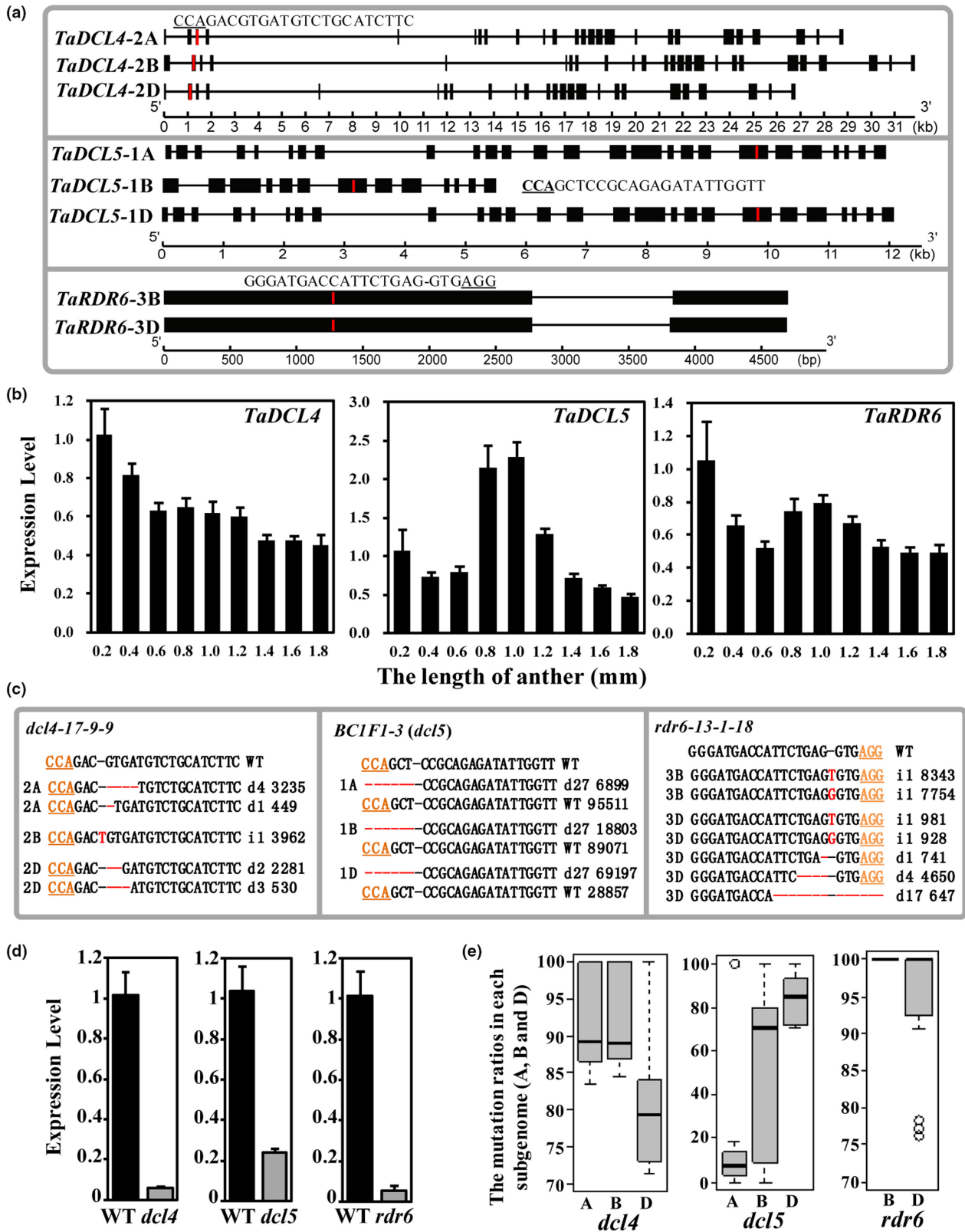
**Figure 1** (a) The gene structure of the *TaDCL4*, *TaDCL5* and *TaRDR6* genes and their sgRNAs targeting all homologous chromosomes. (b) The expression level of the three genes in the young spike with different anther lengths. (c) The mutagenesis of the three genes induced by CRISPR/Cas9. PAM sites are underlined and indicated in orange; numbers indicated the reads of sequences; d, deletion; i, insertion. (d) The expression level of the three genes in the young spike of the WT and their mutants. (e) The mutation ratios in each sub-genome (A, B and D) in *dcl5*, *dcl4* and *rdr6* mutants.

only when the mutation ratio to at least 70% in chr1D (Figure 1e). It suggested that the chr1A and chr1D may lost function after hexaploidization. In contrast, mutated phenotypes such as MS only occurred when each homologue of *TaDCL4* and *TaRDR6* were edited with mutations to at least 70% or 75% in each homologue (Figure 1e and Tables S2, S3). Otherwise, no obvious phenotype was observed, indicating that the pathways mediated by these two genes were to some extent redundant. The male sterile mutants of *dcl4* were inheritable by self-crossing. For some *rdr6* mutants, about half of the tillers were male sterile and thus could be propagated by self-crossing too.

## Phenotypes of the *dcl4*, *dcl5* and *rdr6* mutants

For severe mutants of the three genes, anthers were the most affected organs that were shorter, 2.1, 2.4, and 2.3 mm for *dcl4*, *dcl5* and *rdr6* plants respectively relative to those of the wild type (3.2 mm) (Figure 2a,c and Figure S4). The spike and spikelet of the *dcl4* mutant was shorter or smaller too (Figure 2b), and moreover a few of spikelets in the *dcl4* mutant contained only a pair of empty glumes without the three little anther flowers (Figure S5a). The plants' tiller, spike, and height in *rdr6* were like the WT, and only half of the tillers behaved as MS (Figure 2d and Figure S5c). For most of the *rdr6* mutants, the spike and spikelet behaved similarly to the WT (Figure 2b), whereas several spikes had abnormal spikelets without anther and the lemma became thin and aristiform. In addition to anther morphology, *rdr6* mutants showed wider effects and severe phenotypes such as more tillers, dwarfism and plants with difficulty to germination or to enter reproductive phase (Figure S5b,c and Table S4), indicating that the functions of *TaRDR6* were essential for multiple developmental process in wheat. The *dcl5* mutants behaved to be thorough male sterile in all the tillers, while these plants in tiller, spike, and height were like the WT (Figure 2a–d and Figure S5d).

## Fertility of the *dcl4*, *dcl5* and *rdr6* mutants

Further study showed that mutants had no fertile pollen grains when stained with I₂-KI and were shrunken in morphology. Scanning electron microscopy (SEM) also revealed mutant pollen walls to be wrinkled, irregular or shrunken (Figure 2e). Cytological observation showed that defective meiocytes of the *dcl4* and *rdr6* mutants gradually began to disintegrate during meiotic to tetrad development, while only incomplete tapetal layer was observed in the *dcl5* mutant (Figure S6). Correspondingly, fewer pollen grains were observed in *dcl4* and *rdr6* anthers, but the pollen number in *dcl5* anther was similar to the WT (Figure 2e and Figure S6). These data indicated different functional domains of these genes during pollen development.

## Meiosis and microspore development of *dcl4*, *dcl5* and *rdr6* MS mutants

We then observed the development of pollen mother cells (PMCs) of *dcl4* and *rdr6* and found that they could complete the meiosis process (Figure 3). However, *dcl4* plants showed numerous of PMCs with lagged chromosome, which thus produce malformed dyad, triad, and tetrad causing different chromosome ploidy and hence deformed microspores (Figure S7). In *rdr6* mutants, only a few PMCs showed lagging chromosomes with multiple nuclei (Figure S8). Some larger pollen grains were observed due to the presence of conglutinated microspores. The microspores of *dcl5* became deformed when germinal aperture appeared (Figure 3) and furthermore ~4%~6% of microspores produced two

germinal apertures (Figure 3 and Table S5). These observations demonstrated that *TaDCL4* or *TaRDR6* were involved in meiosis, while *TaDCL5* functioned in microspore development.

## Biogenesis of phasiRNAs in the reproductive stage was dependent on TaDCL4, TaDCL5, and TaRDR6

To study the roles of TaDCL4, TaDCL5 and TaRDR6 in phasiRNAs biogenesis, we sequenced 18 small RNA libraries derived from tissues of anther primordia (AM) and anthers at tetrad stage (TS) of the wild type and three gene editing mutants with two or three biological replicates. A total of 630 21-*PHAS* loci were identified in AM libraries and 3376 in TS libraries, where comparable 24-*PHAS* loci were found for both stages (AM vs. TS, 1072 vs. 1364) (Figure 4a and Data S1, S2). Despite this, only 17% (185) 24-*PHAS* at AM was overlapped with those at TS, while nearly 40% (246) of the 21-*PHAS* reappeared at TS (Figure 4b). In *dcl4* mutants, no 21-*PHAS* loci was identified, either at the AM or TS, whereas numbers of 24-*PHAS* remained in a similar ratio at both stages as in the wild type (AM vs. TS, 1016 vs. 1573) (Data S3, S4), indicating that 21-nt phasiRNAs biogenesis was largely dependent on TaDCL4 in both the AM and TS stage. At the TS stage of *dcl5* mutants, 24-*PHAS* was completely diminished, while 21-nt ones remained similar as in the wild type (Figure 4c and Data S5), indicating that the biogenesis of 24-nt phasiRNAs was totally dependent on TaDCL5. The mutation of *RDR6* caused significant reduction of both 21- and 24-*PHAS* (Figure 4a and Data S6, S7), consistent with its role in double strand RNA production during phasiRNA biogenesis as reported in other species (Song *et al.*, 2012b). Interestingly, few *PHAS* loci were overlapping between developmental stages in the wild type or between the wild type and the three mutants, indicating temporal and spatial expression specificity of phasiRNAs (Figure 4c). Consistent with previous reports (Jiang *et al.*, 2020; Zhang *et al.*, 2020b), we found that more than 90% of 21-*PHAS* were targeted by miR2118 with TargetFinder score ≤ 8, whereas 27.21%–71.83% of 24-*PHAS* were targeted by miR2275 (Figure 4d,e). In addition, 30.66%–91.97% 24-*PHASs* were also targeted by sRNA tae_cand1 (Table S6). Degradome data confirmed that 21- and 24-*PHAS* could be cleaved by miR2118, miR2275, and tae_cand1 (Figure 4e), respectively.

## PhasiRNAs targeted both coding and noncoding RNAs

Despite the presence of *PHAS* loci, phasiRNAs with reads >50 and TargetFinder score ≤4 were only abundantly identified in wild type (both AM and TS, or AM$^{wt}$ and TS$^{wt}$) and the TS of the *dcl5* mutant (TS$^{dcl5}$), with 480, 619, and 2226 21-nt phasiRNAs targeting 2485, 13 024, and 10 703 coding gene transcripts respectively. Totally, 509, 675, 311 and 309 high-confident 24-nt phasiRNAs were identified to target 575, 310, 619 and 675 coding gene transcripts in AM$^{wt}$, AM$^{dcl4}$, TS$^{wt}$, and TS$^{dcl4}$, respectively. Moreover, a total of 2 450 002 long non-coding RNAs (lncRNAs) were identified from the TS-stage anthers of WT and *dcl4* plants by the strand-specific transcriptome sequencing. On average, 559 21-nt phasiRNAs targeted 16 981 lncRNAs, whereas 1250 24-nt phasiRNAs targeted 5535 lncRNAs in AM$^{wt}$. In TS$^{wt}$, an average of 1214 21-nt phasiRNAs targeted 38 890 lncRNAs. In other words, one 21-nt phasiRNAs targeted ~30 lncRNAs in both AM and TS, whereas one 24-nt phasiRNA targeted only 3–4 lncRNAs (Table S7). It had to be noted that a significant portion of phasiRNAs targeted transposon elements (TEs), ~70% for 21-nt phasiRNA targets from lncRNAs and ~60% for 24-nt phasiRNAs (Table S7 and Figure S9), indicating that

**Figure 2** The phenotype of anthers (a and c), spikes (b) and plant phenotypes (d) in the WT, and *dcl4*, *dcl5* and *rdr6* mutants. The white arrows indicated the sterile tillers. (e) I$_2$-KI staining of mature pollen grains (the left panel) for the WT, *dcl4*, *dcl5* and *rdr6*. SEM examination of the pollen grains in the middle and right panels, respectively.
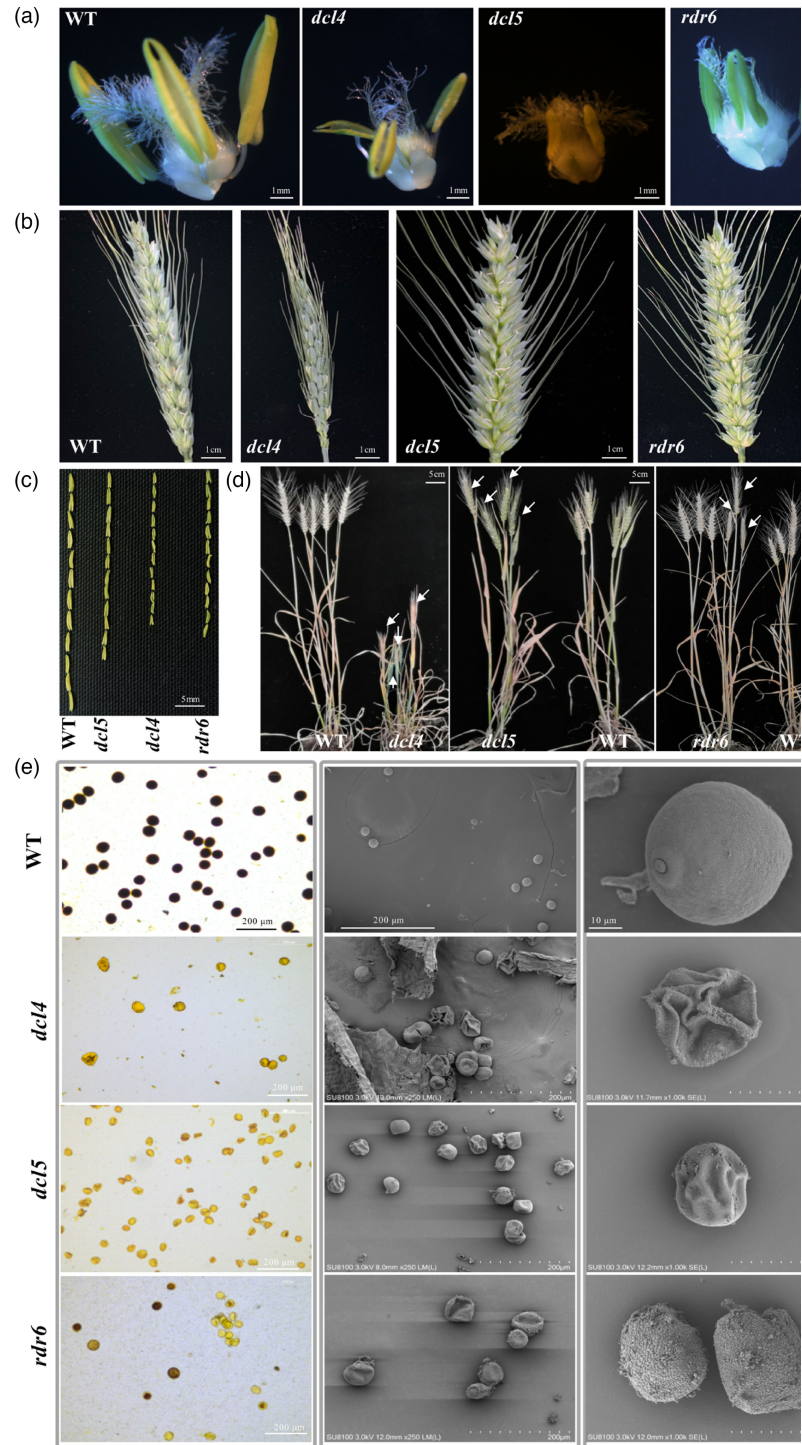
phaisRNAs may also involve in genome stability by affecting chromatin structure.

To study the action mode of phasiRNAs, we performed degradome sequencing and target sites prediction using the CleaveLand program with a category ≤1 and a *P-value* < 0.05. In the wild type, a total of 343 and 439 21-nt phasiRNAs respectively in TS$^{wt}$ and TS$^{dcl5}$ were found to align to the breaking points located at 384 and 502 potential target gene mRNA (Data S8–S11). In contrast, a total of 105 and 38 24-nt phasiRNAs respectively in TS$^{wt}$ and TS$^{dcl4}$ were aligned to 117 and 43 coding target transcripts (Figure 4f and Data S8–S11). Furthermore, there were 141 21-nt phasiRNAs and 32 24-nt phasiRNAs were verified to target 162 and 34 lncRNAs, respectively. Then, we verified such cleavage by rapid amplification of 5′ cDNA ends (5′-RACE) for the lncRNA TCONS_02748198 that was cleaved by 21-nt phasiRNA (Figure 4f). These data indicated that the cleavage is one of the main modes of actions of phasiRNAs on target transcripts.
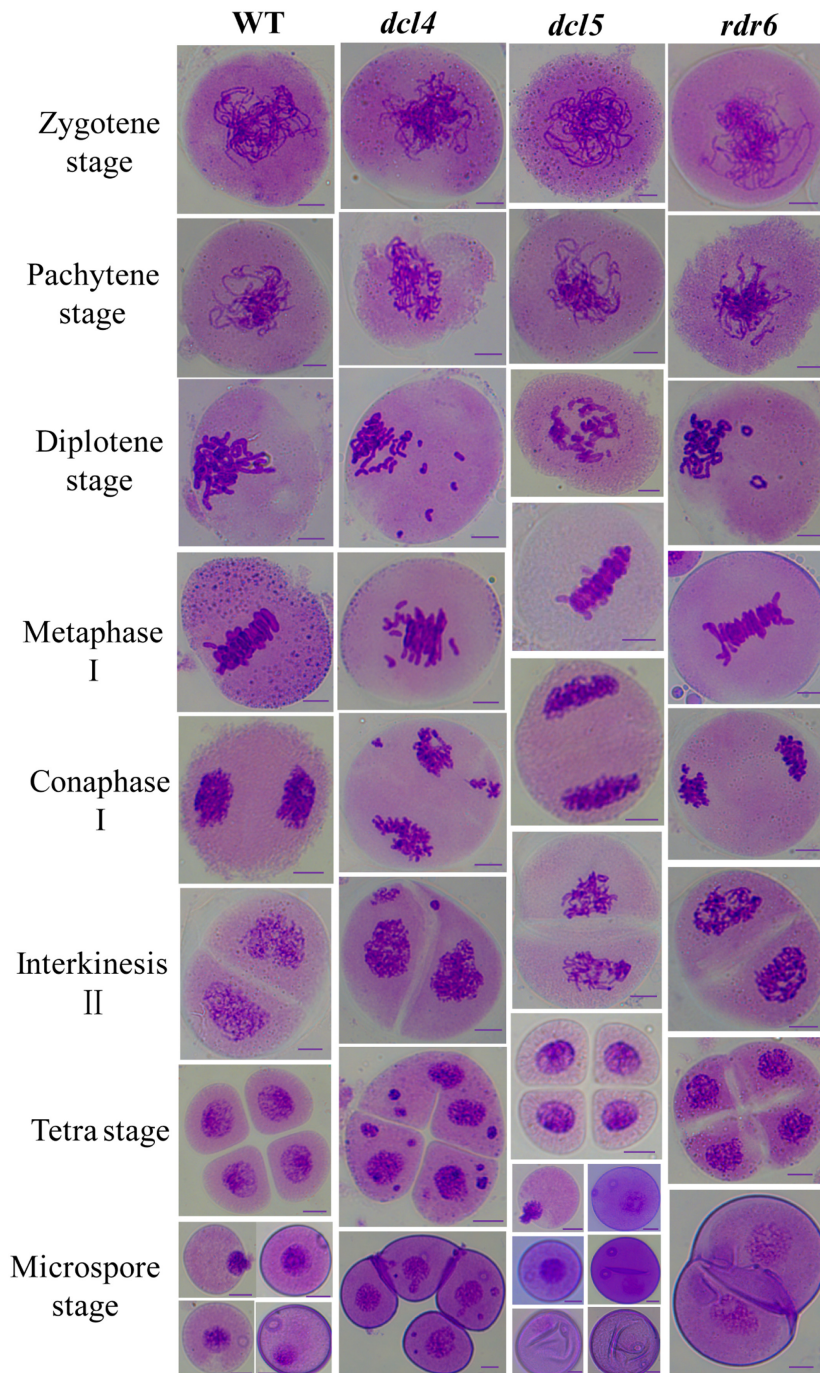
**Figure 3** The meiosis processes by staining with modified card magenta of the WT and MS mutants of *dcl4*, *dcl5*, and *rdr6* in PMCs. Scale bars = 10 μm.

## PhasiRNAs were extensively involved in meiosis, tapetum and microspore development

Gene Ontology (GO) enrichment analysis showed that targets of 21-nt phasiRNAs were enriched for functions of basic biological processes such as protein modification, carbohydrate biosynthesis and ncRNA metabolism, as well as regulation of the reproductive process (Figure 5a and Data S12, S13). Particularly, GO terms related to meiosis progress, such as microtubule movement and nucleation and chromatin modification such as histone modification, were enriched (Figure 5a and Figure S10a,b, Data S12, S13). For 24-nt phasiRNA targets, enriched GO-terms included cell wall organization and modification, telomere organization and

maintenance, phosphatidylserine biosynthetic and metabolic process, and sterol metabolic process (Figure 5b and Figure S11, Data S14). The functions for cation homeostasis, regulation of cell communication and signalling pathway were also enriched (Figure 5b and Data S14), suggesting their potential roles in anther development.

To give a more precise picture about the function of the phasiRNAs during microsporegenesis, we aligned them to wheat genes that were homologous to rice genes for pollen wall development (PWD) and male sterility (Shi *et al.*, 2015; Wan *et al.*, 2021). As shown in Figure 5c,d and Table S8, phasiRNAs targeted the transcripts of genes involved in nearly all steps of pollen intine, primexine and exine development, tapetum
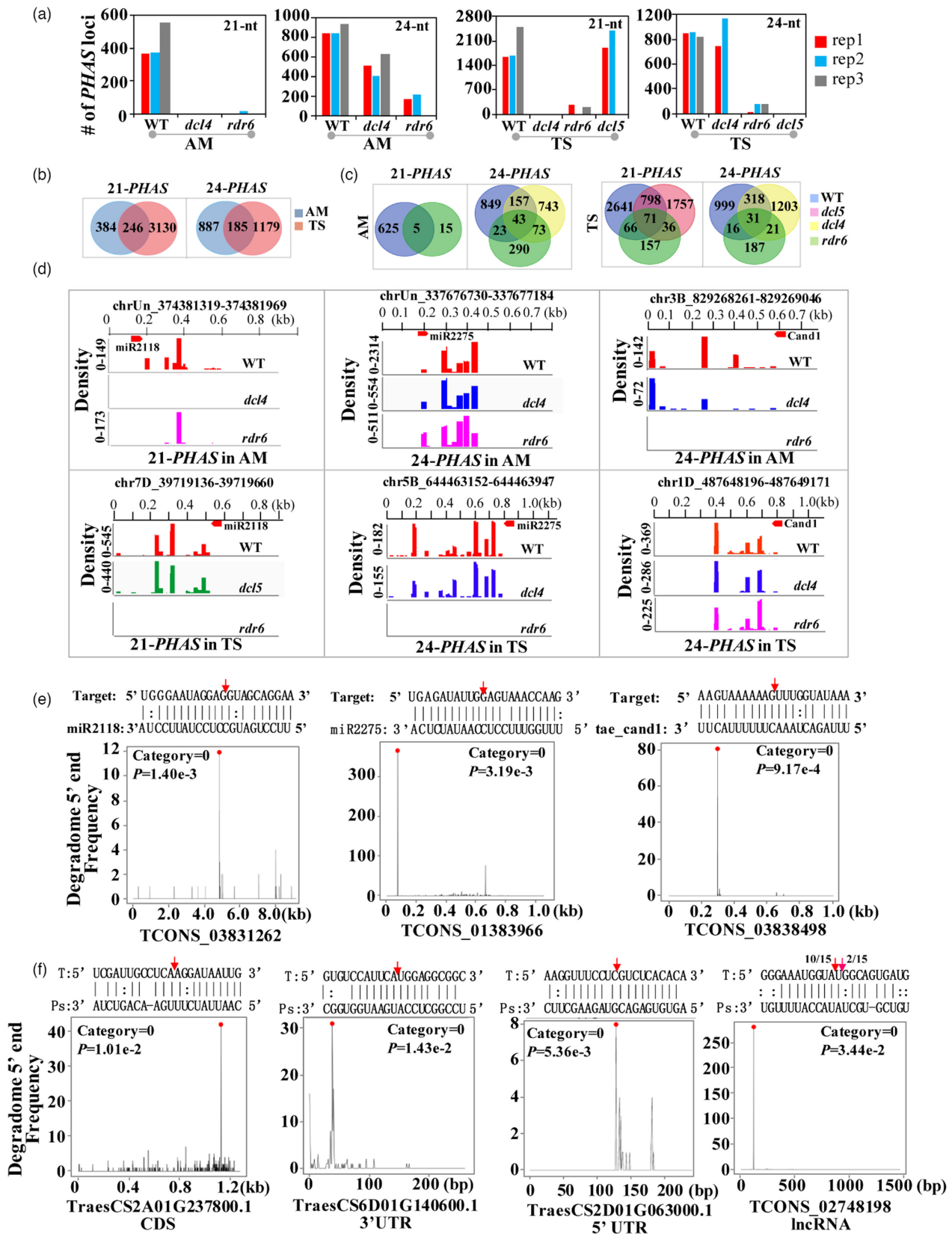
**Figure 4** (a) The predicted 21- and 24-*PHAS* loci in the AM/TS^WT, AM/TS^*dcl4*, AM/TS^*rdr6*, and TS^*dcl5*. (b) The overlapped 21- and 24-*PHAS* loci between the AM^WT and TS^WT. (c) The overlapped 21- and 24-*PHAS* loci in the AM/TS^*dcl4*, AM/TS^*rdr6*, and TS^*dcl5*. (d) The density of 21-nt and 24-nt phasiRNAs in 21- and 24-*PHAS* loci targeted by miR2118 and miR2275 or tae_cand1 in the WT and *dcl4*, *dcl5* and *rdr6* mutant. (e) The verified cleavage in *PHAS* loci by miR2118, miR2275 and tae_cand1 by degradome data. (f) The validated cleavage target site guided by 21-nt phasiRNAs in the gene body including the coding regions, 3′ UTR, and 5′ UTR regions, and the lncRNA by degradome data and 5′-RACE.
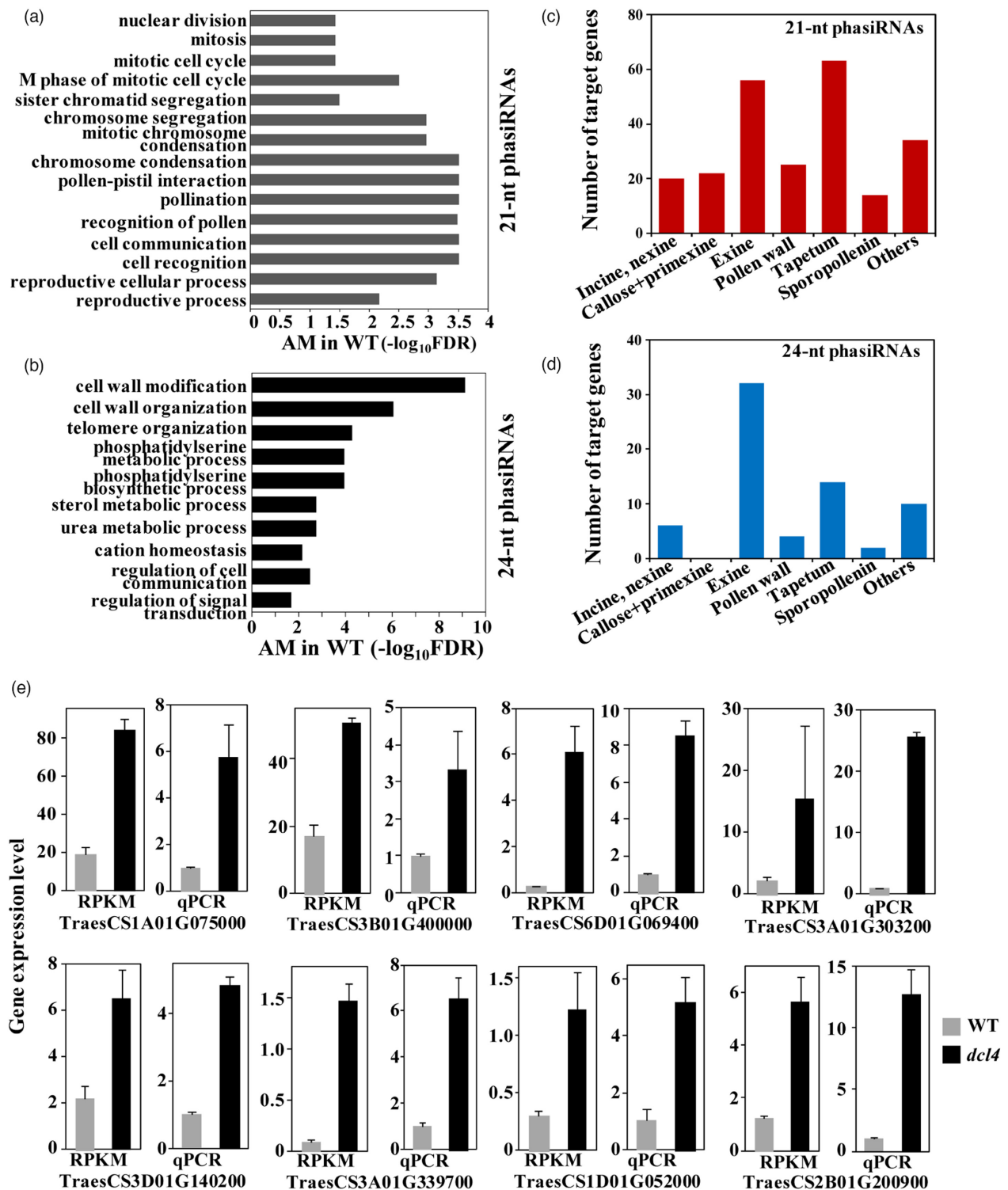
**Figure 5** The enriched GO-terms for the targets of 21-nt (a) and 24-nt phasiRNAs (b) in the AM^WT. [(c) and (d)] The distribution targets genes of 21-nt (c) and 24-nt (d) phasiRNAs involved in fertility regulation. (e) Expression levels of anther development genes targeted by phasiRNAs between WT and *dcl4*, or *dcl5* mutant. RPKM, gene expression level in transcriptome data.

differentiation, and sporopollenin metabolism development, and PWD. To confirm the regulation of phasiRNAs on their targets, we performed transcriptome sequencing to verify expression level variation of anther development genes targeted by phasiRNAs in

mutants and WT. In 245 genes as listed in Table S8, we found that 68 were differentially expressed genes (DEGs) in *dcl4* mutants (Table S9). Among them, 62 were upregulated, whereas six were downregulated. Thus, *dcl4* mutant caused more genes

upregulated than those in WT. To further verify these DEGs, we selected eight up-regulated DEGs for qRT-PCR and found that all of them were consistently significantly up-regulated (Figure 5e). Thus, both the transcriptome data and the result of qRT-PCR confirmed that the phasiRNAs indeed regulated their targets during anther development.

## Evolution of phasiRNAs and their targets in grasses

The biogenesis mechanism of phasiRNAs and their action mode were conserved among grasses as shown above. To further understand whether phasiRNAs were also conserved, we compared phasiRNA and their targets in wheat and rice by BLASTN program, with an e-value cutoff <1 e-20. Approximately 1.47% (43/2930) 21-*PHAS* and 4.99% (88/1736) 24-*PHAS* loci of wheat could be matched to the rice genome, demonstrating lower homology in *PHAS* loci between species. Despite this, target genes of phasiRNAs appeared to be conserved to some extent. Totally 484, 2474 and 2048 target genes in AM$^{WT}$, TS$^{WT}$, and TS$^{dcl5}$ could be matched to 234, 698 and 605 target genes in rice respectively (Zhang *et al*., 2020b) (Data S15–S17). The overall similarity of these 21-nt phasiRNAs that target the homologous genes was lower than 50%. Thus, although the similarity was low for 21-nt phasiRNAs, there were common targets between wheat and rice. This could be due to that phasiRNAs in wheat and rice had their individually specific target sites (Figure S12). For example, phasiRNA targeted the UTR region of rice, whereas the phasiRNA in wheat targeted the cds region of the homologous genes. These data indicated the high diversity of phasiRNAs and/or their targets.

## Divergence of *PHAS* genes in polyploid wheat

To investigate the divergence of *PHAS* genes among three subgenomes of wheat, we merged all *PHAS* loci into one dataset and yielded 2924 21-nt (21-*PHAS*) and 1736 24-nt (24-*PHAS*) nonredundant *PHAS* loci. Aligning these *PHAS* loci to the genomes with stringent criterion of 80% identity and 80% length, we found that only a small fraction of *PHAS* loci, 2.6% for 21-*PHAS* and 1.7% for 24-*PHAS*, had additional homologues in other subgenome(s), significantly less than protein-coding genes, of which 56% genes were triads with three homologues in all the subgenomes (Evans *et al*., 2022). In other words, 81.22% 21-*PHAS* and 86.52% 24-*PHAS* were singleton, significantly more than that of the coding genes, that is, 25.97% (Figure 6a,b). Plotting of wheat *PHAS* loci across each chromosomes showed their preferential distribution at telomere regions, which was similar to the protein coding genes (Figure 6c,d and Figures S13, S14). It was interesting that most *PHAS* loci on chromosomes 4A and 4D were concentrated located on the long arm telomere, indicaing that some sort of deletions may have occurred on these two chromosomes.

To study the evolution of wheat *PHAS* genes, we mapped them to the genomes of *T. urartu* (AA), *T. dicoccoides* (AABB), and *Aegilops tauschii* (DD). With 80% identity and 80% matched length, 63.03%, 60.42% and 91.26% of 21-*PHAS* genes from subA, subB and subD of hexaploid wheat were able to find orthologs in progenitors with AA, ancestor of BB, or DD genomes, respectively. The percentages of *PHAS* genes from tetraploid wheat were even higher, with 87.82% and 85.89% 21-*PHAS*, respectively, from subA and subB, matching to those in the AABB genome. Similar patterns were found for 24-*PHAS* (Figure 6e). Moreover, the proportion of *PHAS* in each subgenome matched to its corresponding progenitor was negatively to their divergence time with $R^2$ = 0.684 and *P*-

value = 3.16 e-3 (Figure 6f), indicating dynamic evolution of *PHAS* loci among *Triticium* species.

To check whether *PHAS* homologues were targeted by the same phasiRNA simultaneously, we investigated the cleavage sites of 21-nt phasiRNAs. Genes fell in GO terms in Figure 5b for reproductive, meiosis, microtube-based and chromatin process, were classified and 63 triads, 26 duplet, and 14 singletons were identified (Figure 6g). For triad genes, 51.67% maintained cleavage sites in all three homologues, while a similar percentage (56.52%) of duplets had both homologue carrying cleavage sites. Therefore, in hexaploid wheat, about half (50%) of *PHAS* genes had diverged and lost their target sites, indicating rapid evolution of *PHAS* genes in polyploid wheat.

# Discussion

## Functional conservation of phasiRNA biogenesis pathways

PhasiRNAs are interesting regulators with an intriguing phasing mechanism by miRNAs and a series of small RNAs of the same length but may target different transcripts with different modes of actions (Jiang *et al*., 2020; Zhang *et al*., 2021). By making mutations in major components, TaDCL4, TaDCL5 and TaRDR6, of the phasiRNA biogenesis pathway using CRISPR/Cas9 techniques, we showed that the functions of these genes were largely conserved even in polyploid wheat (Figure 7). Editing of *DCL4* produced dwarf and weak wheat plants, with shorter spikes and fewer pollen grains. In anthers, many PMCs were arrested at prophase I and could enter the zygotene stage. Despite that some pollen grains did enter meiosis, chromosomes behaved abnormally in the dyads and tetrads leading to many sterile pollen grains. Wheat *dcl4* mutants were similar to that of rice and caused significant reduction in 21-nt phasiRNA production (Liu *et al*., 2007; Zhang *et al*., 2020b).

In contrast, the *dcl5* mutants showed clear tapetum defectiveness. But PMCs could normally develop to tetraploid and then to microspores, thus, the amount of pollen grains was similar to that of the wild type. *Dcl5* mutants grew and developed normally (Figure S5d). Male sterility was evident, even at lower temperature 20 °C/16 °C in a photoperiod of 16 h/8 h (day/night). This may be different to that in maize *dcl5* mutants which showed temperature-sensitive male sterility (Teng *et al*., 2020), which may indicate the difference of the DCL5 gene between *panicoideae* and *pooideae*. And whether *dcl5* mutants were sensitive to the lower temperature, we will perform the further study. Despite this, in grasses, *DCL5* may be optimized to create a male sterile line for hybrid breeding in grasses.

*RDR6* is known to play key roles in small RNA production. Indeed, 40%–60% of wheat *rdr6* mutant seeds even could not germinate or could not transition into the reproductive stage. Accordingly, no phasiRNAs were produced in *rdr6* mutants. The mutants demonstrated a wider role of RDR6 and small RNAs throughout plant development, which is conserved among grass plants (Liu *et al*., 2020; Song *et al*., 2012b).

## Divergence of *DCL4*, *DCL5*, and *RDR6* genes in hexaploid wheat

Common wheat is a hexaploid with three sets of genomes that have diverged for ~5 million years. The hexaploidization occurred about 8000 years ago when the tetraploid cultivated wheat with AABB genomes (*T. dicoccum*) combined with the DD g the current wheat reference genome. CRISPR/Cas9 editing showed
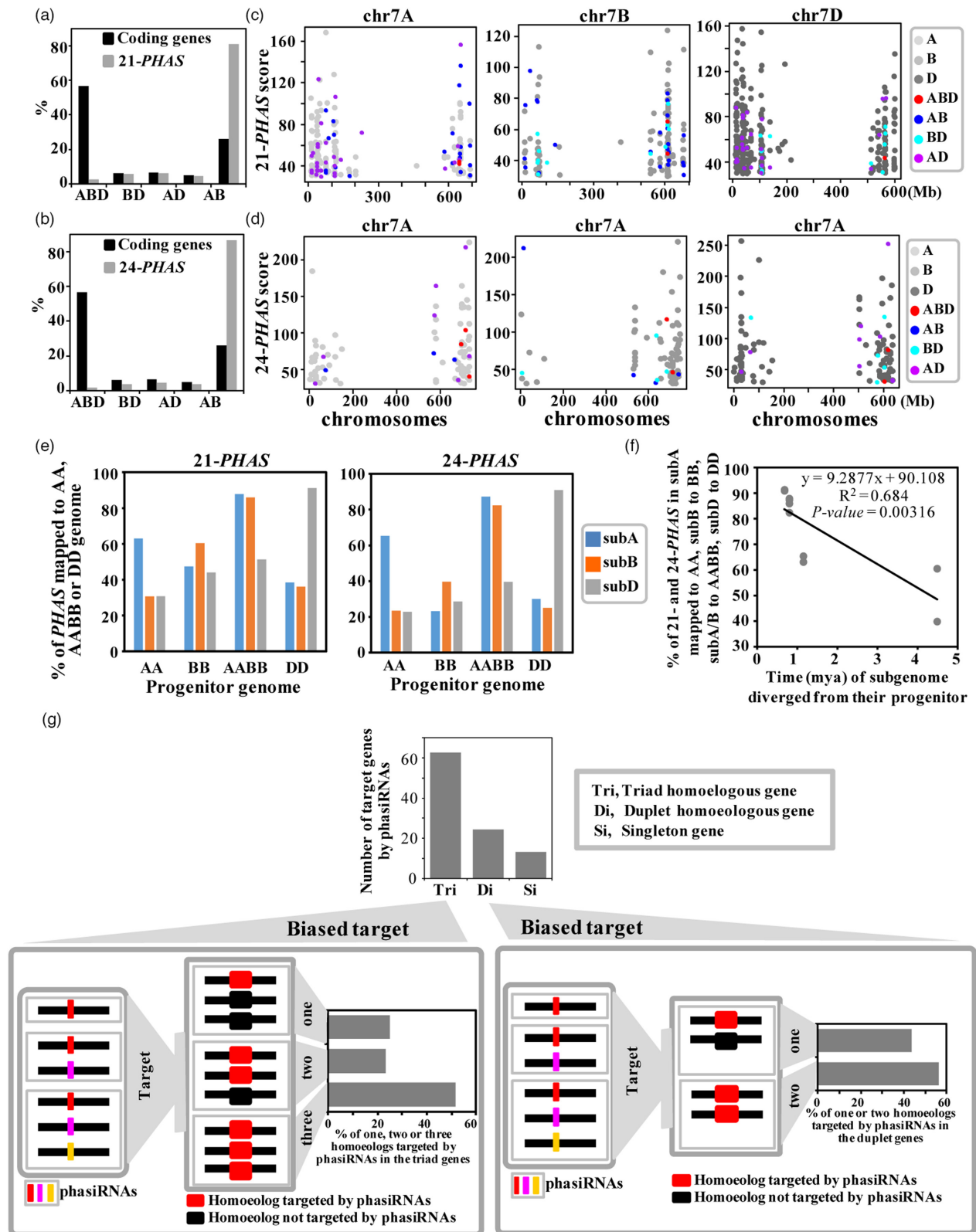
**Figure 6** Comparison of the percentage of triads, diads and singleton between 21- (a) or 24-*PHAS* (b) and the coding genes. The distribution of 21- (c) and 24-*PHAS* loci (d) across the chromosomes. (e) Percentage of 21-*PHAS* and 24-*PHAS* in each sub-genome mapped to their progenitor genome. (f) The correlation between the time of each sub-genome diverged from its progenitor, and the percentage of *PHAS* in subA mapped to AA, subB to B, subA/B to AB, subD to D. (g) The gene number of triads (Tri), diads (Di) and singleton (Si) genes targeted by 21-nt phasiRNAs, and the percentage of one, two or three homologs targeted by phasiRNAs in Tri or Di genes.
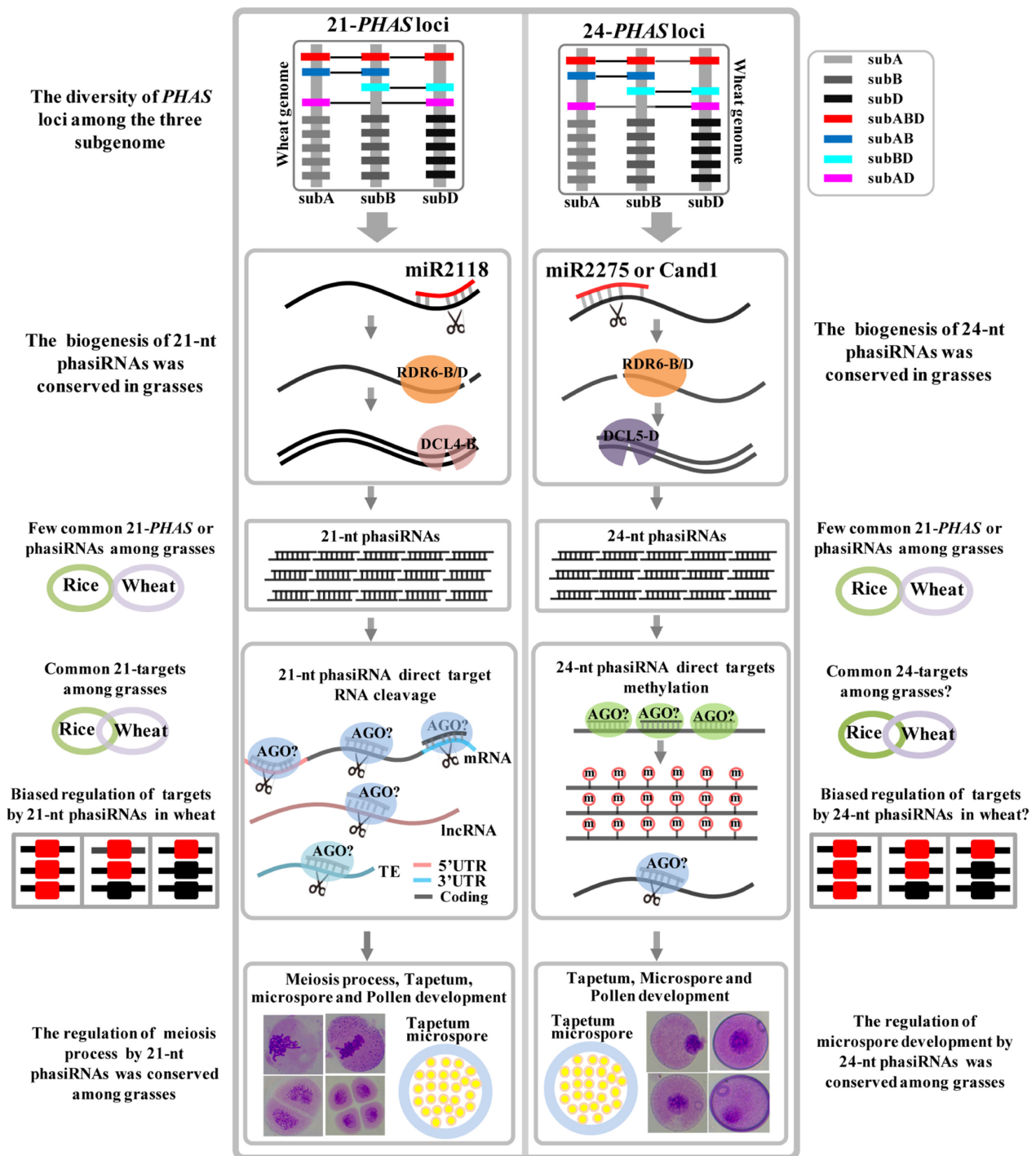
**Figure 7** The diversity of *PHAS* among three subgenomes (the top of panel); the production process of 21-nt or 24-nt phasiRNAs by *TaDCL4* or *TaDCL5*, and *TaRDR6* (the middle of panel); their potential regulatory mechanism of pollen development in hexaploid wheat (the bottom panel); and their evolution among grasses (the left and right in the outside panel).

that both 3B and 3D homologues had to be knocked out for mutated phenotype to be shown. Accordingly, no 21-nt and 24-nt phasiRNA can be found.

## Diverged phasiRNAs targeted genes of conserved functions

Despite the conserved biogenesis mechanism of phasiRNAs among grasses plants (Song *et al.*, 2012a, 2012b; Xia *et al.*, 2019; Zhai

*et al.*, 2015), the similarity of phasiRNAs between species was low. Nevertheless, we found that some target genes in wheat could be matched to rice target genes (Figure S12). These genes in rice were enriched in carbohydrate biosynthetic process, which is consistent with our observation in some of target gene GO enrichment analysis. In addition, those for recognition of pollen, nucleoside bisphosphate metabolic process, and those with functions for gene expression reprogramming during meiotic progression were also

found in both rice and wheat. Thus, despite the quick diversification between phasiRNAs, their targets are large conserved. The target sites of phasiRNAs perhaps located in the unconstrained regions where the nucleotide diversity would not affect the gene function. Also, the diversity of the target sites probably drives the rapid variation of phasiRNAs in grasses. It may indicate the co-evolution between phasiRNAs and their targets in the regulation network.

### Rapid diversification of *PHAS* genes in wheat

The phasiRNAs have been reported in many flowering plants, demonstrating their ancient origins. These non-coding RNAs, however, were by no means homologous among grass species (Figure 7). And in *Oryza* species, *PHAS* sequences also diverged rapidly, which were under strong selection for conservation (Tian *et al.*, 2021). For the *Triticum* family, members were diverged for ~5 mya (Li *et al.*, 2022). Within such a time frame, there was still a trace number of *PHAS* genes that were recognizable as homologues, with 18.78% 21-*PHAS* and 13.48% 24-*PHAS* existing as triads in common wheat and many can still be recognized in the genomes of their progenitors. Despite this, we did not identify homologues for 233 21-*PHAS* and 172 24-*PHAS* in their progenitors. Besides the possible problem of genome quality, they may be nascent *PHAS* genes, arising after wheat polyploidization. However, in a larger time scale, such as the ones between cereal species that separated for more than 50 mya, no homology can be found between *PHAS*s of rice and wheat. Thus, *PHAS* loci showed a much higher frequency of birth and death than the coding genes in wheat.

## Conclusions

PhasiRNAs are extensively involved male gametophyte development and strictly dependent on *TaDCL4*, *TaDCL5* and *TaRDR6*. Our work revealed function of genes for phasiRNA biogenesis as well as their precursors and target genes. These findings provide important insight into wheat gametophyte development and may aid in selecting optimal genes to produce male sterile lines and establishment of novel hybrid wheat breeding strategy in this staple food plant.

## Experimental procedures

### Plant materials

The plants were grown in greenhouses in a photoperiod of 16 h/8 h (day/night) at 20 °C/16 °C. Transgenic lines were selected based on BASTA resistance. Transgenic wheat plants were identified with the Enviologix QuickStix kit for bar protein (ENVIROLOGIX, Portland, ME). The $T_1$ and $T_2$ generation progenies were collected from independent $T_0$ lines.

### Vector construction and genetic transformation

Using CRISPRdirect (http://crispr.dbcls.jp/), the sgRNAs were designed in their target sequences. The Cas9-sgRNA expression vectors were constructed as previously described (Zhang *et al.*, 2018b), which was introduced into the *Agrobacterium tumefaciens* strain EHA105. Immature embryos of wheat (Fielder cultivar) were isolated and *Agrobacterium*-mediated transformation was performed according to Zhang *et al.* (2018b).

### Detection of the targeted genome editing events

Genomic DNA from the leaves of transgenic wheat plants was extracted using the Tiangen DNAquick Plant System (Tiangen, Beijing, China). The construction of the next-generation sequencing libraries were performed using the Hi-TOM Gene Editing Detection Kit (Qingxue, Xi'an, China) (Liu *et al.*, 2019). PCR amplification, library construction and Illumina sequencing and the analysis of nucleotide sequence variation in the targeted sites were performed according to Zhang et al. (2019). The primers used were listed in Table S10.

### SEM observation

Anthers were collected directly into 2.5% glutaraldehyde overnight at 4 °C. The fixed materials were dehydrated through a series of ethanol solutions. The dehydrated samples were critical-point dried and coated with platinum (Hitachi s-3400N). The SEM images of pollen were observed using a scanning electron microscope.

### Expression analysis by real-time qRT-PCR

Total RNA was isolated from Fielder and their mutant young spike using TRIzol reagent (SparkJade, China) and cDNA was synthesized using RT SuperMix for qRT-PCR (Vazyme, Nanjing, China). qRT-PCR was performed with SYBR qRT-PCR Master Mix (Vazyme) on a LightCycler® 480 Real-Time PCR system (Roche, Rotkreuz, Switzerland). The expression of *ubiquitin* was used as the endogenous control to normalize the gene expression levels during different samples. Each sample was performed with three independent biological replicates and three technical repetitions. The gene expression level was calculated using the $2^{-\Delta\Delta ct}$ method (Livak and Schmittgen, 2001). Primers used were listed in Table S10.

### 5′-RACE

5′-RLM-RACE was performed using the 5′/3′ RACE kit (Vazyme). Following the protocol, total RNA was directly ligated to the RNA adapter oligonucleotide with the T4 RNA ligase, which were then reversely transcribed using 5′-TS oligos. To confirm the cleavage site of mRNA, nested PCR was done using the reverse gene-specific primers and the universal primer mix that was homologous to the 5′-TS oligos. The PCR products were purified and cloned using the TA/Blunt-Zero Cloning kit (Vazyme), which were then sequenced. The nested primers used were listed in Table S10.

### Strand-specific RNA-seq, small RNA, and degradome library construction

Young spikes were pooled from the WT and the MS mutants with two or three replications and then were stored in liquid nitrogen. Total RNA was isolated using the TRIzol reagent according to the manufacturer's instructions (Invitrogen, Carlsbad, CA). NEBNext® UltraTM Directional RNA Library Prep Kit, and Multiplex Small RNA Library Prep Set were used to construct the libraries for strand-specific RNA-seq and sRNA-seq, respectively. The degradome libraries were constructed following the method by German *et al.* (2009). Their corresponding deep-sequencing using Illumina NovaSeq 6000 with PE150, SE50, and SE50 was performed in Novogene company, respectively.

### Bioinformatics analysis of strand-specific RNA-seq, small RNA sequencing data and degradome data

The raw data were cleaned by removing reads containing poly-N, low-quality reads, and reads containing adapter through in-house perl scripts. Then, the cleaned reads were aligned to the Chinese Spring genome (IWGSC version 1) from the URGI website (https://wheaturgi.versailles.inra.fr/Seq-Repository/Assemblies) by HISAT2 (V2.0.4; Kim *et al.*, 2015).

*Analysis of stand-specific RNA-seq*

The cleaned reads were assembled to transcripts by Stringtie (V1.3.1) (Pertea *et al.*, 2015). The assembled transcripts were filtered out as coding potential by the following four tools, that is, Coding-Non-Coding-Index V2 (Sun *et al.*, 2013), Coding Potential Calculator 0.9-r2 (Kong *et al.*, 2007), Pfam Scan v1.3 (Bateman *et al.*, 2002) and phylogenetic codon substitution frequency v20121028 (Lin *et al.*, 2011). The left transcripts were the candidate set of lncRNAs. Cuffdiff (v2.1.1) was used to calculate RPKMs of genes (Trapnell *et al.*, 2010). The analysis of differential gene expression was performed by edgeR (v3.38.4) (McCarthy *et al.*, 2012).

*PhasiRNA prediction*

The ShortStack program V3.8.5 (Michael, 2013) was used to align the cleaned data to the wheat reference genome (AABBDD). Then, the distribution of small RNAs on the reference genome was analysed by the perl scripts, and *PHAS* loci were identified with phased scores ≥30. Using the TargetFinder program, we predicted the target sites of phasiRNAs, miR2118, miR2275 and miRCand1 (Bo and Wang, 2005). The distribution of phasiRNA in the genome was shown by the IGV program (Robinson *et al.*, 2011).

*Analysis of degradome-seq*

In the degradome libraries, the target cleavage sites of phasiRNAs were confirmed in coding and lncRNAs by the CleaveLand program (German *et al.*, 2009). The cleavage sites were classified into the categories of 0, 1, 2, 3 and 4 according to the abundance of reads in the cleaved sites along the whole transcripts with a *P*-value of less than 0.05.

## Acknowledgements

## Conflicts of interest

The authors declare no conflicts of interest.

## Author contributions

R.Z., S.Z. and G.L. designed the research, analysed data and wrote the manuscript. J.G. and Y.L. performed wheat transformation. J.L., G.S, W.L., S.G. and C.L. performed the experiment. All authors read and approved the final manuscripts.

## Data availability statement

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (Degradome data: CRA007480, small RNA data: CRA007491, lncRNA data: CRA007492) and are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

## References

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, R.S., Griffiths-Jones, S. *et al.* (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280.

Bélanger, S., Pokhrel, S., Czymmek, K. and Meyers, C.B. (2020) Premeiotic, 24-nucleotide reproductive PhasiRNAs are abundant in anthers of wheat and barley but not rice and maize. *Plant Physiol.* **184**, 1407–1423.

Bo, X.B. and Wang, S. (2005) TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics*, **21**, 1401–1402.

Chang, Z., Chen, Z., Wang, N., Gang, X., Lu, J., Yan, W., Zhou, J. *et al.* (2016) Construction of a male sterility system for hybrid rice breeding and seed production using a nuclear male sterility gene. *Proc. Natl Acad. Sci. USA*, **113**, 14145–14150.

Chen, G., Gong, D., Guo, X., Qiu, W. and He, Q. (2005) Problems of the hybrid with Chongqing thermo-photo-sensitive male sterility wheat C49S in the plain of Jiang Han. *Mailei Zuowu Xuebao*, **25**, 147–148.

Evans, E.C., Arunkumar, R. and Borrill, P. (2022) *Transcription factor retention through multiple polyploidisation steps in wheat.* bioRxiv. https://doi.org/10.1101/2022.02.15.480382

Fan, Y., Yang, J., Mathioni, S.M., Yu, J., Shen, J., Yang, X., Wang, L. *et al.* (2016) PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proc. Natl Acad. Sci. USA*, **113**, 15144–15149.

German, A.M., Luo, S., Schroth, G., Meyers, C.B. and Green, J.P. (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat. Protoc.* **4**, 356–362.

Geyer, M., Albrecht, T., Hartl, L. and Mohler, V. (2018) Exploring the genetics of fertility restoration controlled by Rf1 in common wheat (*Triticum aestivum* L.) using high-density linkage maps. *Mol. Genet. Genomics*, **293**, 451–462.

Jiang, P., Lian, B., Liu, C., Fu, Z., Shen, Y., Cheng, Z. and Qi, Y. (2020) 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *Nat. Commun.* **11**, 5191.

Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.-L., Walbot, V., Sundaresan, V. *et al.* (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res.* **19**, 1429–1440.

Kim, D., Langmead, B. and Salzberg, L.S. (2015) HISAT: a fast spliced aligner with low memory requirements Daehwan HHS Public Access. *Nat. Methods*, **12**, 357–360.

Kong, L., Zhang, Y., Ye, Z., Liu, X., Zhao, S.-Q., Wei, L. and Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349.

Li, J., Wang, Z., He, G., Ma, L. and Deng, X. (2020) CRISPR/Cas9-mediated disruption of TaNP1 genes results in complete male sterility in bread wheat. *J. Genet. Genomics*, **47**, 263–272.

Li, L., Zhang, Z., Wang, Z., Li, N., Sha, Y., Wang, X., Ding, N. *et al.* (2022) Genome sequences of five Sitopsis species of Aegilops and the origin of polyploid wheat B subgenome. *Mol. Plant*, **15**, 488–503.

Lin, F.M.F., Jungreis, I. and Kellis, M. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, **27**, i275–i282.

Liu, C., Wu, Y., Zhang, C., Ren, S. and Zhang, Y. (1997) A preliminary study on the effects of *Aegilops crassa* 6x cytoplasm on the characters of common wheat. *J. Genet. Genomics*, **24**, 241–247.

Liu, C., Hou, N., Liu, G., Wu, Y., Zhang, C. and Zhang, Y. (2002) Studies on fertility genes and its genetic characters in D2-type CMS lines of common wheat. *J. Genet. Genomics*, **29**, 638–645.

Liu, B., Chen, Z., Song, X., Liu, C., Cui, X., Zhao, X., Fang, J. *et al.* (2007) Oryza sativa dicer-like4 reveals a key role for small interfering RNA silencing in plant development. *Plant Cell*, **19**, 2705–2718.

Liu, Q., Wang, C., Jiao, X., Zhang, H., Song, L., Li, Y., Gao, C. *et al.* (2019) Hi-TOM: a platform for high-throughput tracking of mutations induced by CRISPR/Cas systems. *Sci. China Life Sci.* **62**, 1–7.

Liu, C., Shen, Y., Qin, B., Wen, H., Jiawen, C., Mao, F., Shi, W. *et al.* (2020) Oryza sativa RNA-dependent RNA polymerase 6 contributes to double-strand break formation in meiosis. *Plant Cell*, **32**, 3273–3289.

Livak, K.J.L. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, **25**, 402–408.

McCarthy, D., Chen, Y. and Smyth, G. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297.

Michael, A.J. (2013) ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA*, **19**, 740–751.

Ni, F., Qi, J., Hao, Q., Lyu, B., Luo, M.C., Wang, Y., Chen, F. *et al.* (2017) Wheat Ms2 encodes for an orphan protein that confers male sterility in grass species. *Nat. Commun.* **8**, 15121.

Pallotta, M.A., Warner, P., Kouidri, A., Tucker, E.J., Baes, M., Suchecki, R., Watson-Haigh, N. *et al.* (2019) Wheat ms5 male-sterility is induced by recessive homoeologous A and D genome non-specific lipid transfer proteins. *Plant J.* **99**, 673–685.

Pertea, M., Pertea, M.G., Antonescu, M.C., Chang, T.-C., Mendell, J.T. and Salzberg, L.S. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295.

Pokhrel, S., Huang, K., Bélanger, S., Zhan, J., Caplan, L.J.L., Elena, K.M. and Blake, M.C. (2021a) Pre-meiotic 21-nucleotide reproductive phasiRNAs emerged in seed plants and diversified in flowering plants. *Nat. Commun.* **12**, 4941.

Pokhrel, S., Huang, K. and Meyers, C.B. (2021b) Conserved and non-conserved triggers of 24-nucleotide reproductive phasiRNAs in eudicots. *Plant J.* **107**, 1332–1345.

Qi, X., Zhang, C., Zhu, J., Liu, C., Huang, C., Li, X. and Xie, C. (2020) Genome editing enables next-generation hybrid seed production technology. *Mol. Plant*, **13**, 1262–1269.

Robinson, T.J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, S.E., Getz, G. and Mesirov, P.J. (2011) Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26.

Shi, J.S., Cui, M., Yang, L., Kim, Y.-J. and Zhang, D. (2015) Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci.* **20**, 741–753.

Singh, M., Kumar, M., Thilges, K., Cho, M.J. and Cigan, A.M. (2017) MS26/CYP704B is required for anther and pollen wall development in bread wheat (*Triticum aestivum* L.) and combining mutations in all three homeologs causes male sterility. *PLoS One*, **12**, e0177632.

Singh, M., Kumar, M., Albertsen, M.C., Young, J.K. and Cigan, A.M. (2018) Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* **97**, 371–383.

Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B., Jeong, D.H. *et al.* (2012a) Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J.* **69**, 462–474.

Song, X., Wang, D., Ma, L., Chen, Z., Li, P., Cui, X., Liu, C. *et al.* (2012b) Rice RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. *Plant J.* **71**, 378–389.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y. *et al.* (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **41**, e166.

Teng, C., Zhang, H., Hammond, R., Huang, K., Meyers, B.C. and Walbot, V. (2020) Dicer-like 5 deficiency confers temperature-sensitive male sterility in maize. *Nat. Commun.* **11**, 2912.

Tian, P., Zhang, X., Xia, R., Liu, Y., Wang, M., Li, B., Liu, T. *et al.* (2021) Evolution and diversification of reproductive phased small interfering RNAs in Oryza species. *New Phytol.*, **229**(5), 2970–2983.

Trapnell, C., Williams, B., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M., Salzberg, S. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.

Tucker, E.J., Baumann, U., Kouidri, A., Suchecki, R., Baes, M., Garcia, M., Okada, T. *et al.* (2017) Molecular identification of the wheat male fertility gene Ms1 and its prospects for hybrid breeding. *Nat. Commun.* **8**, 869.

Wan, X., Wu, S. and Li, X. (2021) Breeding with dominant genic male-sterility genes to boost crop grain yield in the post-heterosis utilization era. *Mol. Plant*, **14**, 531–534.

Wang, Z., Li, J., Chen, S., Heng, Y., Chen, Z., Yang, J., Zhou, K. *et al.* (2017) Poaceae-specific MS1 encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl Acad. Sci. USA*, **114**, 12614–12619.

Wu, Y., Fox, W.T., Trimnell, R.M., Wang, L., Xu, R., Cigan, A.M., Huffman, G.A. *et al.* (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. *Plant Biotechnol. J.* **14**, 1046–1054.

Xia, C., Zhang, L., Zou, C., Gu, Y., Duan, J., Zhao, G., Wu, J. *et al.* (2017) A TRIM insertion in the promoter of Ms2 causes male sterility in wheat. *Nat. Commun.* **8**, 15407.

Xia, R., Chen, C., Pokhrel, S., Ma, W., Huang, K., Patel, P., Wang, F. *et al.* (2019) 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nat. Commun.* **10**, 627.

Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.L., Walbot, V. and Meyers, B.C. (2015) Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc. Natl Acad. Sci. USA*, **112**, 3146–3151.

Zhang, D., Wu, S., An, X., Xie, K., Dong, Z., Zhou, Y., Xu, L. *et al.* (2018a) Construction of a multicontrol sterility system for a maize male-sterile line and hybrid seed production based on the ZmMs7 gene encoding a PHD-finger transcription factor. *Plant Biotechnol. J.* **16**, 459–471.

Zhang, S., Zhang, R., Song, G., Gao, J., Li, W., Han, X., Chen, M. *et al.* (2018b) Targeted mutagenesis using the Agrobacterium tumefaciens-mediated CRISPR-Cas9 system in common wheat. *BMC Plant Biol.* **18**, 302.

Zhang, S., Zhang, R., Gao, J., Gu, T., Song, G., Li, W., Li, D. *et al.* (2019) Highly efficient and heritable targeted mutagenesis in wheat via the *Agrobacterium tumefaciens*-mediated CRISPR/Cas9 system. *Int. J. Mol. Sci.* **20**, 4257.

Zhang, R., Huang, S., Li, S., Song, G., Li, Y., Li, W., Li, J. *et al.* (2020a) Evolution of PHAS loci in the young spike of Allohexaploid wheat. *BMC Genomics*, **21**, 200.

Zhang, Y., Lei, M., Zhou, Y.Z., Yang, Y., Lian, J., Wu, Y., Feng, Y. *et al.* (2020b) Reproductive phasiRNAs regulate reprogramming of gene expression and meiotic progression in rice. *Nat. Commun.* **11**, 6031.

Zhang, M., Ma, X., Wang, C., Li, Q., Meyers, C.B., Springer, M.N. and Walbot, V. (2021) CHH DNA methylation increases at 24-PHAS loci depend on 24-nt phased small interfering RNAs in maize meiotic anthers. *New Phytol.* **229**, 2984–2997.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix S1** Creation of the *dcl5* male sterile lines; Creation of the *dcl4* male sterile lines; Creation of the *rdr6* male sterile lines.

**Data S1** The 21-nt and 24-*PHAS* loci in AM$^{WT}$

**Data S2** The 21-nt and 24-*PHAS* loci in TS$^{WT}$

**Data S3** The 21-nt and 24-*PHAS* loci in AM$^{dcl4}$

**Data S4** The 21-nt and 24-*PHAS* loci in TS$^{dcl4}$

**Data S5** The 21-nt and 24-*PHAS* loci in AM$^{rdr6}$

**Data S6** The 21-nt and 24-*PHAS* loci in TS$^{rdr6}$

**Data S7** The 21- and 24-*PHAS* loci in TS$^{dcl5}$

**Data S8** The cleavage information for targets in CDS region by 21-nt or 24-nt phasiRNAs in WT and *dcl5*

**Data S9** The cleavage information for targets in 5′ UTR region by 21- or 24-nt phasiRNAs in WT and mutants of *dcl4* and *dcl5*

**Data S10** The cleavage information for targets in 3′ UTR region by 21- or 24-nt phasiRNAs in WT and mutants of *dcl4* and *dcl5*

**Data S11** The cleavage information for targets in lncRNAs regions by 21- or 24-nt phasiRNAs in WT and mutants of *dcl4* and *dcl5*

**Data S12** The enriched GO-terms in targets of coding sequences targeted by 21-nt phasiRNAs in AM$^{WT}$ and TS$^{WT}$

**Data S13** The enriched GO-terms in targets of coding sequences by 21-nt phasiRNAs in TS$^{dcl5}$

**Data S14** The enriched GO-terms in targets of coding sequences targeted by 24-nt phasiRNAs in AM$^{WT}$ and TS$^{WT}$

**Data S15** The homologous genes between wheat and rice targeted by phasiRNAs in AM$^{WT}$

**Data S16** The homologous genes between wheat and rice targeted by phasiRNAs in TS$^{WT}$

**Data S17** The homologous genes between wheat and rice targeted by phasiRNAs in TS$^{dcl5}$

**Figure S1** The phylogenetic of DCL4 (a), DCL5 (b) and RDR6 gene (c) in *Arabidopsis*, sorghum, maize, rice, *Brachypodium*, barley, wheat (AABBDD) and their progenitor including AA, DD and AABB species.

**Figure S2** The domains of TaDCL4 (a), TaDCL5 (b) and TaRDR6 (c) in wheat.

**Figure S3** The expression pattern of *TaDCL4* (a), *TaRDR6* (b) and *TaDCL5* (c) in wheat different development stages. Flb, flag leaf blade; 30% Sp, 30% spike; Fboot, full boot.

**Figure S4** The length of anther in WT and *dcl4*, *dcl5* and *rdr6* mutants.

**Figure S5** The worse phenotype of plant architecture and spike in *dcl4* (a), *rdr6* (b-c) and *dcl5* (d) mutants compared with WT.

**Figure S6** Transverse section analysis of anthers from the WT, *dcl4*, *rdr6* and *dcl5*. The images are cross sections of a single locule. Bars = 10 mm. The white arrow indicates the tapetal layer.

**Figure S7** The types of PMCs during the meiosis processes by staining with modified card magenta of *dcl4*. Scale bars = 10 μm.

**Figure S8** The types of PMCs during the meiosis processes by staining with modified card magenta of *rdr6*. Scale bars = 10 μm.

**Figure S9** The percentage of lncRNAs targeted by phasiRNAs deriving from TEs.

**Figure S10** The enriched GO-terms in the target coding genes of 21-nt phasiRNAs in TS$^{WT}$ (a) and 24-nt phasiRNA in TS$^{dcl5}$ (b).

**Figure S11** The enriched GO-terms for the targets of 24-nt phasiRNAs in the TS$^{WT}$.

**Figure S12** Common homologous targets between wheat and rice targeted by 21-nt phasiRNAs. Bars indicate target sites; the coloured phasiRNAs corresponded to the bars with the same colours in the target sites.

**Figure S13** The distribution of 21-*PHAS* loci across the chromosome.

**Figure S14** The distribution of 24-*PHAS* loci across the chromosome.

**Table S1** The mutation types and mutant ratios for *TaDCL5* gene in T$_0$, hybrid F$_1$ and BC1F1 generation

**Table S2** The mutation types and mutation ratios for *TaDCL4* gene in T$_0$, T$_1$ and T$_2$ generation

**Table S3** The mutation types and mutation ratios for *TaRDR6* gene in T$_0$, T$_1$ and T$_2$ generation

**Table S4** The worse phenotypes of *rdr6* mutants in *rdr6-13* line

**Table S5** The percentage of two germinal apertures in *dcl5* mutants

**Table S6** *PHAS* LOCI targeted by miR2118 and miR2275 or tae_cand1

**Table S7** Predicted targets deriving from lncRNAs by phasiRNAs

**Table S8** Pollen wall and MS genes targeted by 21-nt and 24-nt phasiRNAs

**Table S9** DEGs for anther development that were targeted by phasiRNAs between *dcl4* mutant and WT

**Table S10** Primers for mutation detection by Hi-TOM sequencing in transgenic wheat plants, primer for qRT-PCR and 5′ RACE