Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evolution of PHAS Loci in the Young Spike of Allohexaploid Wheat

Rongzhi Zhang
Shandong Academy of Agricultural Sciences
Siyuan Huang
BGI
Shiming Li
BGI
Guoqi Song
Shandong Academy of Agricultural Sciences
Yulian Li
Shandong Academy of Agricultural Sciences
Wei Li
Shandong Academy of Agricultural Sciences
Jihu Li
Shandong Academy of Agricultural Sciences
Jie Gao
Shandong Academy of Agricultural Sciences
Tiantian Gu
Shandong Academy of Agricultural Sciences
Dandan Li
Shandong Academy of Agricultural Sciences
Shujuan Zhang
Shandong Academy of Agricultural Sciences
Genying Li (lgy111@126.com)
Shandong Academy of Agricultural Sciences https://orcid.org/0000-0003-3240-1100

Research article

Keywords: PhasiRNAs, PHAS, MicroRNAs, Evolution, Young spike, Wheatv

Posted Date: October 31st, 2019

DOI: https://doi.org/10.21203/rs.2.16650/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published on March 4th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6582-4.

Abstract

Background PhasiRNAs (phased secondary siRNAs) play important regulatory roles in the development processes and biotic and abiotic stresses in plants. A class of phasiRNAs involve in the reproductive development in grasses. Reproductive-associated phasiRNAs include two categories, 21-nt (nucleotide) and 24-nt phasiRNAs, which are triggered by miR2118 and miR2275 in premeiotic and meiotic anthers, respectively, which had been reported in rice, maize and other grass species. However, there were still absence in Triticum The allohexaploid wheat (Triticum aestivum) genome consists of three closely related subgenomes (subA, subB and subD), which is a model of allopolyploid in plants. And the evolution of PHAS loci in polyploid plants is still unavailable.

Results Here, using 261 small RNA expression datasets from various tissues, a batch of PHAS (phasiRNA precursors) loci were identified in the young spike of wheat, most of which were triggered by miR2118 and miR2275 in their target site regions. Dissection of PHAS and their trigger miRNAs among the diploid (AA and DD), tetraploid (AABB) and hexaploid (AABBDD) genomes of Triticum indicated that distribution of PHAS loci were dominant randomly in local chromosomes, while their trigger miR2118 was dominant only in the subB genome. The diversity of PHAS loci in the three subgenomes of wheat and their progenitor genomes (AA, DD and AABB) suggested that they originated or diverged before the occurrence of AABB. The positive relationship between the PHAS loci or the trigger miRNAs and the ploidy of genome indicated the expansion of genome was the major drive force for the increase of PHAS loci and their trigger miRNAs in Triticum . In addition, the PHAS transcripts responded to abiotic stresses such as cold stress in wheat.

Conclusions Altogether, non-coding phasiRNAs are conserved transcriptional regulators that display quick plasticity in genome dominance and sequence diversity and are involved in reproductive development and abiotic stress in wheat. It could be referred to molecular research on male reproductive development in Triticum .

Background

There is a particular class of small RNAs generated in 21- or 24-nt (nucleotide) intervals with a ‘head-to-tail’ pattern from their precursor transcripts, which are called phased, secondary, small interfering RNAs (phasiRNAs) [1–3]. PhasiRNAs have been found both in animals and plants. The genes encoding PIWI-interacting RNAs (piRNAs), which are a type of phasiRNA, are required for producing mature sperm in animals [4]. In plants, phased siRNAs play a series of roles in abiotic and biotic stresses [5, 6], seed germination [7] and reproductive development [3, 8–10]. Trans-acting siRNAs (ta-siRNAs) are a special class of phasiRNAs generated from *TAS* precursors (noncoding or coding transcripts), which silence targets in *trans.* Until now, ten *TAS* genes were characterized in plants. Both *TAS1* and *TAS2* are initiated by miR173 with 22 nt in eudicots [11, 12], and *TAS1* is involved in heat tolerance [12]. *TAS3* triggered by miR390 involved in auxin signalling in Land plants. And for other *TAS* genes, most of them are species-specfic and involved in reproductive process and biotic or abiotic stresses [2]. The regulation of disease

resistance genes with the *NB-LRR (*nucleotide-binding sites and leucine-rich repeat)domain by miRNAs has mostly been characterized in the ETI (effector-triggered immunity) pathway of plants [13]. miRNAs such as miR1507 and miR2109 in *medicago* [14], miR472 and miR482 in soybean [15] and citrus [16], and miR9863 in barley [6] can trigger 21-nt phasiRNA generation from *NB-LRR* transcripts. MiR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling in wheat [7].

In reproductive tissues, phasiRNAs are active in anther development of both eudicots and monocots. In grasses, there are two pathways that yield abundant phasiRNAs [3, 8]. MiR2118 triggers one class of 21-nt phasiRNAs in premeiotic anther development, while miR2275 triggers another class of 24-nt phasiRNAs associated with meiosis [8]. In eudicots, 24-nt phasiRNAs are also present in the anther or pollen development triggered by either miR2275 or not [17]. This finding indicated the ancient origin of phasiRNAs and their regulatory mechanism. In general, 5'-capped and polyadenylated noncoding *PHAS* RNA transcripts by RNA polymerase II are cleaved by 22-nt miR2118 or 22-nt miR2275, which mediates the generation of 21-nt or 24-nt phasiRNAs, respectively. Then, the 3' mRNA fragments are converted into double-strand RNAs by RNA-dependent RNA polymerase 6 (RDR6), which are processed by DCL4 (dice like 4) or DCL5 (also named DCL3b) to yield 21-nt or 24-nt phasiRNAs [18]. Mutations in *DCL4, DCL5* and *RDR6* in rice affect the generation of 21-nt or 24-nt phasiRNAs [18, 19]. These phasiRNAs are subsequently loaded into AGOs to function their regulatory roles. In rice, MEL1 (also named OsAGO5c) preferentially binds to 21-nt phasiRNAs [20], and ZmAGO18b is enriched in the tapetum and germinal cells, similar to the expression pattern of 24-nt phasiRNAs [8].

PhasiRNAs have been identified in a number of flowering plants in eudicots and monocots such as rice, maize, *Brachypodium,* sorghum, foxtail millet [21, 22] and litchi [17]. In the development of rice inflorescence, 828 21-*PHAS* that could produce 21-nt phasiRNAs and 35 24-*PHAS* that could produce 24-nt phasiRNAs were identified by Johnson *et al.* [3]. In addition, 1136 21-*PHAS* were identified in 93–11 and 1540 21-*PHAS* were identified in *Nipponbare* and 70 24-*PHAS* were identified in 93–11 and 34 24-*PHAS* were identified in *Nipponbare* by Song *et al.* [18]. In maize, 463 21-*PHAS* and 176 24-*PHAS* loci were identified [8]. In the flower of litchi, more than 178 21-*PHAS* loci were identified. These generation and regulation mechanisms are very conserved in grasses.

Numerous phasiRNAs are involved in the anther development process, which indicates that they may play a key role in normal anther development [8]. For example, *PMST1,* producing 21-*PHAS* triggered by miR2118, regulates photoperiod-sensitive male sterility in rice [23]. However, *PMST1* only exists in rice, and there is no homolog in other grasses. Thus, *PHAS* and phasiRNAs may be highly species-specific. Until now, few phasiRNAs involved in developing inflorescence have been characterized in *Triticeae.* Only miR9863 and miR9678 were characterized to mediate the generation of phasiRNAs in biotic stress [6] and seed germination [7], respectively. Knowledge about the role of phasiRNAs in the inflorescence development of wheat is absent, and the evolutionary path of phasiRNAs among their diploid progenitors, tetraploid progenitors and hexaploid wheat is still unclear. The large and complex genomes, including wheat and their progenitors in the AA, DD and AABB genomes, are all available at the present.

Furthermore, the number of small RNA datasets from high-throughput sequencing deposited in public databases, such as the GEO database, is increasing, and the databases contain various developmental and stress samples in wheat. These efforts make it possible to systematically identify the phasiRNAs in wheat.

Here, using small RNA datasets in wheat, we investigated the *PHAS* loci in various tissues, such as leaves, roots, flag leaves, young spikes, and grain. The evolution of these *PHAS* loci in *Triticum* species indicated that their independent dynamic evolution was accompanied by their regulators miR2118 and miR2275. This work will be referred for the research of anther development in wheat.

Results

Identification of 21- and 24-*PHAS* in Wheat

To identify the *PHAS* loci in wheat, we downloaded 261 small RNA datasets from the GEO database (*Table 1)*,, which included 12 seedling samples, 128 leaf samples, 12 root samples, one stem sample, one shoot sample, 29 young spike samples, two anther samples, one embryo sample, 17 spikelet samples, 12 rachis samples, 19 grain samples, 20 seed samples, 6 callus samples, and one mixed tissue sample. By comparing these small RNAs to the wheat genome using the Shortstack package [24] following the flowchart shown in *Supplementary Figure 1,* a batch of *PHAS* (phasiRNA precursors) loci were identified with phased scores greater than 15, 20, 25 or 30 (*Supplementary Table 1)*.. In the leaf, stem, root, spikelet, seed, callus, *etc.,* few *PHAS* were identified (*Supplementary Table 1 && Supplementary Figure 2)*.. However, abundant 21- and 24-*PHAS* were identified in the reproductive tissues, such as young spikes and anthers (*Figure 1A-B && Supplementary Table 1)*.. The abundance of phasiRNAs was not associated with the total reads of the small RNA samples, but it was highly related to the types of tissues (*Supplementary Figure 2)*.. In young spikes and anther tissues, numerous phasiRNAs were expressed.

In the downloaded small RNA samples of wheat, there were a group of datasets, which included the important development stage of young spike [25], *i.e.,* DR stage (double-ridge stage) at the initiation of spike formation and spikelet development; FM stage, the stage of appearance of the floret meristems (FMs) glume primordia, and lemma primordia; AM stage (anther primordia stage), stamen and pistil primordia emerged from FMs with visible anther primordia for some florets; TS stage (tetrads stage), young florets began to differentiate with immature anthers and unelongated pistils, and the pollen mother cells completed meiosis to form the tetrads at this stage. In the DR and FM stages of young spike tissues, there were few *PHAS* loci, while in the AM and TS stages, there were abundant 21- and 24-*PHAS.* For example, ~3600 and ~4000 of the 21-*PHAS* loci with phased scores greater than 30 and 1200 and 1000 of the 24-*PHAS* loci with phased scores greater than 30 were identified in AM (SRR3690677 and SRR3690678) and TS (SRR3690679 and SRR3690680) samples, respectively (*Supplementary Table 1)*.. There were more 21-*PHAS* loci in anthers with lengths of 1.0 mm than in those with lengths of 1.5, 2.2 and 3.0 mm, while for the 24-*PHAS,* the number of *PHAS* loci were very similar to that of 21-*PHAS* (*Figure 1A-B)*.. According to morphological development and stage determination of young spike and anther in

wheat [26], five small RNA datasets were selected for further study that represented the early and later anther development stages. The early anther development stage included the floret meristem (FM, SRR5460941 and SRR5460949), anther primordia (AM, SRR5460967 and SRR5460972) and tetrad stages (TS, SRR5461176 and SRR5461177), and the later anther development stage included free haploid microspores (FHM, SRR449365 with 1.5 mm anther), mitosis (MIT, SRR449366 with 2.2 mm anther), and mature pollen (MP, SRR449367 with 3.0 mm anther).

For the 21-nt phasiRNAs (21-phasiRNAs) in *PHAS* loci, the proportion of all 21-nt small RNAs was different. In the FM stage, there were few 21-phasiRNAs (300 TPM (transcripts per million)) with 0.6% of the total 21-nt small RNAs, while in AM, a sharp increase (36054 TPM) was observed, comprising 23.96% of the total. In the TS stage, the expression level continued to increase with 11410 TPM 21-phasiRNAs (44.07% of totals) (*Figure 1C & E).*. For the 24-phasiRNAs, the tendency of the proportion was similar to the 21-phasiRNAs. The proportions were 0%, 13.21% (101277 TPM), and 30.76% (156595 TPM) for the FM, AM, and TS stages, respectively (*Figure 1C & E).*. This indicated that the 21- or 24-phasiRNAs and 21- or 24-*PHAS* loci were present in the AM stage and peaked in the TS stage. For the later anther development stage, the 21- and 24-phasiRNAs occurred in the FHM stage with proportions of 8.99% (13701 TPM) and 8.23% (52030 TPM) peaked in the MIT stage with proportions of 15.21% (28633 TPM) and 10.72% (60758 TPM), and then decreased with proportions of 12.9% (24562 TPM) and 8.74% (50908 TPM), respectively (*Figure 1C & E).*.

The synthesis of phasiRNAs in monocot reproductive tissues requires both *PHAS* precursors and their initiated miRNAs, such as miR2118 and miR2275 [18, 19]. The expression level of miRNAs was concertedly related to the synthesis of the phasiRNAs and *PHAS* loci. For 21- and 24-*PHAS,* the *PHAS* loci peaked both in AM and TS and then rapidly decreased in the later anther development stage (*Figure 1A-B).*. In contrast, both 21- and 24-phasiRNAs were expressed in AM, peaked in TS and rapidly descended in the later anther development stage (*Figure 1C & E),*, which peaked after the development stage compared to 21- and 24-*PHAS.* We investigated the concert of the three elements at the transcriptome level. The expression of miR2118 peaked in the AM stage, which was similar to that of 21-*PHAS,* but occurred before the burst in expression of 21-nt phasiRNAs. Then, the expression of miR2118 decreased in the TS stage and disappeared in the later anther development stage. The expression of miR2275 peaked in the TS stage, which was before that of 24-phasiRNAs but the same as that of 24-*PHAS.* Then, the expression of miR2275, 24-phasiRNAs and 24-*PHAS* rapidly decreased in the later anther development stage (*Figure 1A-E).*. The expression of miR2275 was higher than that of miR2118 in the later anther development stage, which may be associate with the higher expression of 24-nt phasiRNAs than 21-nt phasiRNAs in these stages (*Figure 1A-E).*. Together, the expression of the 21- and 24-nt phasiRNAs and their trigger miRNAs burst in the early anther development stage.

The distribution of *PHAS* loci in the wheat genome

To investigate the genome distribution of *PHAS* in wheat, we used the Circos package to show the number of *PHAS* loci sliding the chromosome in window sizes of 500 kb. Here, we selected these small RNA datasets with abundant phasiRNAs in AM, TS, FHM, MIT and MP for further study. In the five developmental stages, the distribution of genome elements such as repeat sequences, gene body and intergenic regions were very similar in both 21- and 24-*PHAS (Supplementary Figure 3)*.. Few *PHAS* were located in the gene body or repeat sequence regions of the genome. Only 1~2% of these genes were distributed in gene body regions, and 12~21% of them were distributed in repeat sequence regions. In contrast, most of them (78~87%) were located in the intergenic regions. This result indicates that most *PHAS* loci may have independent transcript units that are not juxtaposed with the repeat sequences or coding genes.

According to the genome locations, we respectively merged all of the 21- and 24-*PHAS* in these samples derived from five florescence development stages. In total, there were 4850 and 3600 unique 21- and 24-*PHAS* in these samples, respectively (*Supplementary Figure 4A-B)*,, most of which were common. In total, 94.93% (2042 out of 2151), 53.14% (2385 out of 4488), 98.61% (637 out of 646), 94.85% (1263 out 1198) and 95.91% (1056 out of 1101) of 21-*PHAS* (*Supplementary Figure 4A)*,, and 49.87% (993 out of 1991), 69.09% (1571 out of 2274), 98.26% (1019 out of 1037), 98.38% (1157 out of 1176) and 98.79% (1057 out of 1070) of 24-*PHAS* (*Supplementary Figure 4B)*,, overlapped each other in at least two samples in AM, TS, FHM, MIT, and MP, respectively. The low number of common 21-*PHAS* in TS and 24-*PHAS* in AM indicated that there were more tissue-specific *PHAS* loci in these two development stages, which may be associated with the transition of the development stage from floret meristem to meiosis.

These merged unique *PHAS* were plotted on the chromosome rainbow with black lines (*Figure 2)*.. Most of the *PHAS* loci were located at the end of the chromosomes, *i.e.* telomere regions. In most regions, the peaks of the 21-*PHAS* (red lines in *Figure 2)* were higher than those of the 24-*PHAS* (blue lines in *Figure 2)* in the representative tissues. Most of the peaks in both 21- and 24-*PHAS* were similar among the subA, subB and subD genomes in each sample. However, some peaks of 21- and 24-*PHAS* in local chromosomes were preferred among the three subgenomes in each sample (*red and blue arrows in Figure 2 and Figure 3)*.

The genome plasticity of *PHAS* in allohexaploid wheat

Polyploidization is followed by genome partitioning or fractionation processes, *i.e.* a genome-wide diploidization, in which one or the other gene duplicate is lost. During this process, different functional protein-coding genes have been shown to behave differently. Transcription factors or regulators are often retained as duplicated copies following whole genome duplications (WGDs), whereas others are progressively deleted back to a single copy (singleton) state [27–29]. For allohexaploid wheat (AABBDD, *Triticum aestivum,* 2n = 6x = 42), high sequence similarity and structural conservation are retained with limited gene loss after polyploidization [30]. However, at the transcript level, cell type- and stage-dependent genome dominance was observed in the local chromosome regions [31]. Studies of the roles

of *PHAS* loci as a category of noncoding genes in the partitioning process of allohexaploid wheat have still not been performed. To investigate the location of *PHAS* in the subA, subB and subD of allohexaploid wheat, we performed a blast search against all genomes to identify the relationship of *PHAS* among the three subgenomes. With 80% identity and 80% matched sequence length, only 2.27%~6.40% of *PHAS* with scores of greater than 30 retained the triplet copies in the three subgenomes in the five samples, 11.11%~17.51% of *PHAS* retained the duplet copies in any two subgenomes, and 76.09%~86.22% of *PHAS* retained only singleton in any one subgenome (*Supplementary Figure 5).*. The homoeologous relationship is shown in *Figure 3* with the same color link lines in the homoeologous chromosomes. There were also some translocated homoeologs. For example, some *PHAS* in chr4A were homologous to those in chr5B/D and chr7A/D. These data showed that only partial *PHAS* retained the triple or duplet homoeologs, and most *PHAS* only possessed the singleton copy.

To investigate the subgenome distribution of *PHAS* clusters in allohexaploid wheat, the total number of *PHAS* loci in each sub-chromosome was calculated, and there were no biased in each subA, subB and subD chromosomes. However, for the local chromosome, there were some bias distribution in the local subA, subB and subD genomes. For 21-*PHAS* loci, in the bottom chromosomes of chr1, chr2, and chr3 and in the top and bottom chromosomes of chr4 and chr7, the *PHAS* loci were biased located in the chromosomes, but the tendency of preference was different, in either the top or bottom of one chromosome (*Figure 2 (the red arrow regions) and Figure 4A).*. In chr1-b (b, bottom of the chromosome), significantly less 21-*PHAS* were located in the subA genome than in the subB and subD genomes (Fisher's exact test, *P-value* < 0.05). In chr2-b, the number of 21-*PHAS* was significantly less in subB than in subA and subD (*P-value* < 0.05). In chr3-b and chr5-b, subB became the dominant genome with significantly more 21-*PHAS* than the subA and subD genomes (*P-value* < 0.05). In chr4-t (top of the chromosome), the number of 21-*PHAS* was less in subA than in subB and subD (*P-value* < 1.0e−5), while in chr4-b, the subA genome possessed many 21-*PHAS,* significantly more than the other subgenome (*P-value* < 2.2e−16). In chr7-t, subB possessed less of the phased loci, next to the subA genome, and the subD possessed many more 21-*PHAS* than subA and subB (*P-value* < 0.001).

For 24-*PHAS,* only chr3-t, chr4-t/b and chr7-t in all five samples exhibited a biased distribution (*Figure 2 (blue arrow regions) and Figure 4B).*. In chr4-t/b, the preference of 24-*PHAS* in chromosomes was very similar to that of 21-*PHAS.* In chr4-t, less *PHAS* loci were located in subA than in subB and subD (*P-value* < 1.0e−6), while in chr4-b, more *PHAS* loci were located in subA than in subB and subD (*P-value* < 2.2e−6). In chr3-t, more 24-*PHAS* were distributed in the subB genome than in the subA and subD genomes (*P-value* < 0.05). However, in chr7-t, far fewer 24-*PHAS* were located in the subB genome than in the subA and subD genomes (*P-value* < 1.0e−6). These data suggested that *PHAS* loci exhibited local chromosome preferences during the genome plasticity process.

Homoeologous *PHAS* loci in diploid, tetraploid, and allohexaploid wheat

The progenitors of allohexaploid wheat contain the diploid genomes AA, BB and DD and the tetraploid genome AABB. The genomes of AA (*Triticum urartu,* 2n = 2x = 14), DD (*Aegilops tauschii,* 2n = 2x = 14), and AABB (*Triticum turgidum,* 2n = 4x = 28) were sequenced, and their genome assembly was nearly perfect at the chromosome level. This made it possible to investigate the evolution of 21- and 24-*PHAS* in the different ploidy *Triticum* species. Thus, we downloaded the genome sequences of AA, DD and AABB, and then mapped the 21- and 24-*PHAS* with scores of greater than 30, which were identified in these samples from the five developmental stages of florescence, to these genome sequences with the BLAST program with 80% identity and 80% matched sequence length. Approximately 22.91%~25.98% of the 21-*PHAS* could be mapped to the AA genome, and 37.77%~44.12% of the 21-*PHAS* could be mapped to the DD genome, which was slightly higher than the 21-*PHAS* that were mapped to the AA genome. However, a greater proportion of *PHAS* (60.83%~62.72%) could be mapped to the AABB genome. For 24-*PHAS,* the mapped proportions to AA, DD and AABB were very similar to the mapped 21-*PHAS.* Approximately 22.52%~26.13%, 29.18%~37.42%, and 53.96%~58.05% of 24-*PHAS* could be aligned to the AA, DD and AABB genomes, respectively (*Supplementary Table 2).*. In the tetraploid AABB, the mapping rates of 21- and 24-*PHAS* were much higher than the diploid species AA and DD. The scatter diagram showed that the mapping rate was positively correlated with the ploidy times in 21-*PHAS* with r^2 = 0.8019, as determined by Pearson correlation test (*P-value* = 6.41e−6, *Supplementary Figure 6A),*, and 24-*PHAS* withr^2 = 0.8578*P-value* = 6.41e−6, *Supplementary Figure 6B).*. This indicated that the expansion of the whole genome level led to an increase in the *PHAS* loci.

For the mapped 21- and 24-*PHAS* to the AA genome, 67.55% and 75.73% ofthem were located in the subA genome of wheat, respectively. For the mapped 21- and 24-*PHAS* to DD genome, there were 72.78% and 77.53% of them located in subgenome D of wheat, respectively. For the mapped 21-*PHAS* to the AABB genome, 43.41% and 42.40% of them were from the subA and subB genome of hexaploid wheat, respectively, and for 24-*PHAS,* a similar proportion (*i.e.* 44.38% and 44.51% for subA and subB, respectively) was observed. The detailed genome distribution in the AABBDD genome for the mapped *PHAS* was clearly shown in each chromosome of *Figure 5.* For the *PHAS* mapped to the AA genome, there were more peaks (purple lines in *Figure 5)* from each chromosome of subA than from the other subgenomes. For the mapped *PHAS* to the DD genome, the peaks (green lines in *Figure 5)* were mostly distributed in the subD genome. For the mapped *PHAS* to the AABB genome, each chromosome of both subA and subB had more peaks (blue lines in *Figure 5)* than the subD genome. This finding indicated the orthologous relationship between subA and the AA genome, subB and BB genome, subD and DD genome. The evolution independence of *PHAS* sequences among the three diploid species, AA, BB and DD, may suggest that the *PHAS* sequences in *Triticum* may diverge before tetraploid synthesis and may diverge after divergence of the AA, BB and DD species.

Genome distribution of miR2118 and miR2275 in *Triticum*

The production of phasiRNAs was initiated by miR2118 and miR2275 *via* cleavage of their *PHAS* precursors. MiR2118 and miR2275 possessed many copies in grasses. There were 18 members of

miR2118 and four members of miR2275 in the rice genome and seven members of miR2118 and four members of miR2275 in the maize genome, based on the miRbase database [32]. To detect these miRNAs in *Triticum,* we mapped the mature sequences of miR2118 and miR2275 from the miRbase database to the *Triticum* genomes with perfect matches and found that there were 25, 30, 88, and 140 members of miR2118 in the DD, AA, AABB and AABBDD genomes, respectively, and 6, 5, 17 and 24 members of miR2275 in the four genomes, respectively. Most of these miRNAs were clustered on the chromosomes of the four species, as shown in *Figure 6A and Supplementary Figure 7–8.* The increase tendency of miR2118 and miR2275 was significantly correlated with ploidy, as determined by Pearson correlation test (*P-value* = 0.0015 and 0.0084, r^2 = 0.9971 and 0.9833, respectively; *Figure 6B-C).*. For miR2118, there was also a biased distribution in the three subgenomes. However, unlike the 21-*PHAS,* the tendency of miR2118 on the subgenome in each group of chromosomes was consistent (*Figure 6A).*. Except for chr2, subB was significantly dominant, with many more members of miR2118 than the subA and subD genomes in chr1 (Fisher exactly test, *P-value* = 1.87e−5), chr4 (*P-value* = 3.90e−14) and chr5 (*P-value* = 3.65e−14). On chr2, there were also more members of miR2118 in the subB genome but not significantly with *P-value* > 0.05. In the tetraploid AABB genome (*Supplementary Figure 7),,* miR2118 in subB was also dominant than the subA genome on chr1 (*P-value* = 2.06e−8), chr2 (*P-value = 1.43e−2),,* chr4 (*P-value = 3.72e−10),,* and chr5 (*P-value = 1.43e−2).*. The similar dominant subgenome distribution of miR2118 in the AABB and AABBDD genomes indicated that the dominance of subB may have occurred before the synthesized hexaploidy of wheat. In the AA and DD genomes, there were also fewer miR2118 in chr1, chr2, chr4 and chr5 than in subB of the AABB and AABBDD genomes (*Supplementary Figure 8A-B),,* which indicated that the expansion of miR2118 in the subB genome may have occurred before the synthesized tetraploidy of the AABB genome. This indicated the dynamic expansion of the trigger miRNAs following genome expansion or polyploidization.

PHAS loci as the targets of miR2118 and miR2275 in wheat

The production of the *PHAS* cluster could be initiated by miR2118 and miR2275 and then generate the 21-nt and 24-nt phasiRNAs, respectively. To identify whether miR2118 and miR2275 also target these *PHAS* transcripts, using the Targetfinder program [33], we aligned the two miRNA families to the *PHAS* sequences with a score less than 4. Taking the five development stages as examples, 35.43% (767 out of 2159), 34.06% (220 out of 646), 35.31% (446 out of 1263), 38.31% (824 out of 2151), and 35.09% (1575 out of 4488) of 21-*PHAS* sequences were predicted to be targeted by miR2118 in AM, ST, FHM, MIT, and MP, respectively. For 24-*PHAS,* 43.36% (575 out of 1323), 50.43% (523 out of 1037), 50.26% (591 out of 1176), 0.45% (9 out of 1991), and 48.86% (1111 out of 2274) were identified as the targets of miR2275, respectively. The alignment information between the miRNAs and *PHAS* loci are listed in *Supplementary Table 3–7.*

To validate whether these *PHAS* transcripts can be indeed cleaved by miR2118 and miR2275, we downloaded the degradome sequences from the GEO datasets for the young spike tissue under cold stress by Song *et al* [34], which corresponds to the small RNA datasets of the AM stage of young spike

(control samples, SRR3680677 and SRR3680678; and cold stress samples at 0°C after 48 hours, SRR3680679 and SRR3690680). According to the abundance of reads along the whole transcripts, using the Cleaveland program [35] (*P-value* < 0.05 and the category < = 2), miR2118 and miR2275 were considered to be interacted between *PHAS* transcripts and miRNAs. The target plots of miR2118 and miR2275 characterized in the degradome datasets are shown in *Figure 7A-B.* A total of 13.01% (520 out of 3996, SRR3680679) and 12.80% (525 out of 3996 in SRR3680680) of 21-*PHAS* were confirmed to be cleaved by miR2118 for the cold stressed samples. In the control samples, slightly less 21-*PHAS* were validated as cleaved targets, *i.e.* 9.86% (355 out of 3601 in SRR3680677) and 9.65% (353 out of 3659 in SRR3680678). For miR2275, few 24-*PHAS* were detected to be cleaved in the control and cold stressed samples. Only 0.18% (two out of 1122 in SRR3690680) and 0.38% (four out of 1052 in SRR3680679) of the 24-*PHAS* in the cold stressed samples were validated to be cleaved by miR2275, while in the control samples, there were slightly more 24-*PHAS* than in the cold stress samples, as confirmed by cleavage of miR2275, *i.e.* 3.16% (45 out of 1421 in SRR3680677) and 9.65% (38 out of 1203 in SRR3680678). This finding provided evidence that miR2118 and miR2275 could mediate the cleavage of *PHAS* transcripts in wheat. The cleavage information of the degradome for the *PHAS* loci is listed in *Supplementary Table 8–10.*

The expression level of *PHAS* loci

The generation of phasiRNAs may depended on the expression of their precursors. To detect the expression of the *PHAS* precursors, we downloaded the RNA-seq datasets from the GEO database, including the reproductive tissues at DR, FM, AM and TS. The *PHAS* precursors were mostly expressed in the AM and TS stages, and only a few of them were expressed in the DR and FM stages (*Figure 8A).*. Furthermore, most of their expression levels were very low. Only 434 and 289 *PHAS* showed RPKM > 1 in the AM and TS stages of young spikes, respectively, and 175 of them overlapped between the two stages (*Figure 8C).*. This finding indicated the specific expression of *PHAS* loci in different developmental stages of young spikes.

PHAS transcripts were expressed specifically in the reproductive tissues, but whether they responded to abiotic stress was still unknown until now. Taking cold stress as an example, using the transcriptomes with polyA in the DR and AM stages of young spike samples including 0 (control), 6, 12, 24 and 48 h (hours) after cold stress at 0°C, we identified the expression of these *PHAS* loci. In the cold stressed samples, most of the *PHAS* loci were also mainly expressed in the AM stage, and in the AM stage, the expression level also showed a difference (*Figure 8B).*. Using the DEseq program [36] with the fold change (stressed vs control) more than two times and adjusting the *P-value* to less than 0.01, 102, 104, 133 and 144 *PHAS* were characterized as differentially expressed *PHAS* transcripts at 6, 12, 24 and 48 hours after cold stress at 0°C, respectively. For the differentially expressed *PHAS,* a total of 66.67% (68 out of 102), 73.08% (76 out of 104), 74.43% (99 out of 133) and 68.06% (98 out of 144) were common at the four stressed time points (*Figure 8D).*. These results indicated that the *PHAS* transcripts responded to abiotic stress, such as cold stress.

Discussion

High conservation of thebiogenesis mechanism and low conservation of *PHAS* sequences in plants

PhasiRNAs are broadly present and play very important roles in the anther development of grasses, such as rice [3, 9, 18, 19], maize [8], *Brachypodium [*22], *sorghum [*1],and *foxtail millet [*1], and litchi [17]. Their biogenesis mechanisms were very conserved. The generation of phasiRNAs required miRNAs to trigger the cleavage of their RNA precursors, which are subsequently converted into double-stranded RNAs (dsRNAs) by RNA DEPENDENT RNA POLYMERASE6 (RDR6) [19]. These dsRNAs were then processed into 21-nt and 24-nt phasiRNAs by DCL4 or DCL5 proteins [18]. The resulting phasiRNAs are then loaded into an ARGONAUTE (AGO) family protein to direct the activity of their corresponding targets, typically by silencing [8, 37]. These trigger miRNAs and key genes in the biogenesis pathway of phasiRNAs are very conserved in grasses. Although the biogenesis is very conserved, the *PHAS* transcripts have very high diversity among the grass species. By comparing these *PHAS* loci in wheat to the genomes of their closely related species, including *Brachypodium,* rice, sorghum and maize, only 17, 7, 7 and 5 *PHAS* sequences in wheat could be matched with the 50% identity and 50% matched length, respectively. This result indicated that *PHAS* is less conserved across grass species. *PMST1,* as a typical example, regulates photoperiod-sensitive male sterility in rice [23], whereas *PMST1* exists only in rice without a homolog in other grasses. Thus, *PHAS* and phasiRNAs are highly species-specific in grasses.

There is a special class of *PHAS, i.e. TAS* genes. Ten TAS genes have been reported in plants, and only some of them have high conservation. *TAS3,* which is triggered by miR390, is conserved in land plants [38]. *TAS1* and *TAS2,* which are initiated by miR173, are conserved in eudicots [11, 12]. Most of the *TAS* genes are family-specific. *TAS5, TAS9* and *TAS10* are Solanaceae-specific [39, 40]. The other *TAS* genes are species-specific. The *TAS4* gene only exists in *Arabidopsis* [41], *TAS6* only exists in moss [42], and *TAS7–10* only exists in grapevine [43]. These *TAS* genes are involved in various biological processes, such as Auxin signaling, heat and chilling response [2].

Another category of *PHAS* is involved in regulating plant natural immunity. The regulation mechanism is also conserved both in eudicots and monocots. However, the trigger miRNAs are not conserved between eudicots and monocots. In eudicots and gymnosperm, miR472, miR482/2118, mir6024, and miR1507 can target *NB-LRR* transcripts and initiate the generation of phasiRNAs [44]. Most of these *NB-LRR* genes contain the *TIR* domain [45]. In monocots, very few miRNAs were identified to target the *NB-LRR* genes [44]. Recently, in *Triticum,* several miRNA families, such as miR9863, miR3117, miR3084, miR5071 and miR7757, were characterized to target *NB-LRR* transcripts and trigger the production of phasiRNAs in wheat [44]. Most of them are absent or lost function in other grass species. In *Triticum,* there is also variation in the expression level or copy number [44]. The relationship between miRNAs and *PHAS* may frequently become absent and present during the evolution of plants.

There are no common miRNAs in eudicots or monocots that regulate the *NB-LRR* genes. Intriguingly, miR2118, a common miRNA in eudicots and monocots induces neofunctionalization. In eudicots, miR2118/482 was found to mostly target *NB-LRR* genes and initiate the generation of 21-nt phasiRNAs [15]. In contrast, miR2118 in monocots mostly targets numerous noncoding sequences and triggers the generation of 21-nt phasiRNAs [3, 8, 18]. MiR2118 in eudicots is involved in plant immunity response [15], while in monocots, miR2118 plays an important role in anther development [8, 9]. In addition, miR2118 in grasses such as rice [3] and wheat undergoes a special tandem repeat expansion. The function variation of miR2118 in the plant evolution process may be due to the target sequence variation. The sequences of the *NB-LRR* genes rapidly diversified between eudicots and monocots [45]. For example, the *TIR* domain is present in most eudicots but absent in monocots. The domain loss or sequence variation of the *NB-LRR* genes may lead to the functional variation of their regulator.

Dynamic evolution of *PHAS* loci and their trigger miRNAs in *Triticum*

In grasses, numerous *PHAS* loci have been identified in maize [8], rice [3, 9] and wheat. Most of them were derived from the non-coding regions or intergenic regions (*Supplementary Figure 3).*. These *PHAS* loci have very high species specificity, as described in section 3.1. The *Triticum* species had a small time-scale in the divergence history. In *Triticum,* the progenitor of wheat AA, DD diverged approximately four million years ago (mya). AABB and AA diverged approximately 0.5 mya, and AABBDD were nascent approximately 10000 years ago. In the small time-scale of evolution, the *PHAS* loci still have high divergence. Most of the *PHAS* loci (76.09~86.22%) were singletons with only one homologue among the three subgenomes, and others possessed triplet and duplet homoeologs in wheat (*Supplementary Figure 5).*. This study demonstrated the lower conservation of *PHAS* among the three subgenomes. Approximately 22–25%, 61–62% and 37–44% of *PHAS* in AABBDD could be mapped to the AA, AABB and DD genomes, respectively. The lower identity of *PHAS* loci among subA, subB and subD indicated high heterogenization among the three subgenomes. Intriguingly, most of the *PHAS* mapped to AA, AABB and DD were located in the subA, subA & B and subD genomes of wheat, respectively. This indicated that the variation of *PHAS* among subA, subB and subD may occur after the divergence of the AA, BB and DD species and may occur before the synthesized AABB and AABBDD (the left panel of *Figure 9).*. The subgenome plasticity may contribute less to the divergence of *PHAS* among the three subgenomes (the left panel of *Figure 9).*. The *PHAS* that were not mapped to the AA, AABB or DD genomes may be the nascent *PHAS* after the yield of wheat. Overall, the non-coding *PHAS* loci may have a much higher evolution rate than the protein-coding genes.

The trigger miRNAs of *PHAS* were also expanded following genome expansion or polyploidization. The members of miR2118 and miR2275 in diploid, tetraploid and hexaploid were progressively increased and positively related to the ploidy. This result indicated the lower subgenome plasticity of the miRNA trigger among the three sets of genomes in wheat. More miR2118 was distributed in most subB chromosomes in AABB and AABBDD than in subA and subD (*Figure 6A and Supplementary Figure 7).*. However, there

was no evidence to deduce the occurrence times of miR2118 expansion in the subB genome. Nevertheless, in chr5B of the AABBDD genome, there were 35 members of miR2118 (*Figure 6A)*,, which was more than the 17 members of miR2118 in chr5B of the AABB genome (*Supplementary Figure 7)*.. This indicated the special expansion of miR2118 in chr5B of AABBDD genomes, and the expansion may occur after the synthesis of hexaploid wheat (right panel of *Figure 9)*..

The expansion of miR2118 in *Triticum* and the other grass species may be associated with the numerous *PHAS.* Increasing the number of copies of miR2118 may increase to regulate *PHAS* and yield phasiRNAs (the middle pannel of *Figure 9)*.. After initiating the cleavage of the mRNA precursor of phasiRNAs, the secondary 21-nt phasiRNAs can function in *cis* to target their own precursors, which was observed in both rice and maize [9]. The simultaneously expanding and subgenome plasticity in both miRNAs and *PHAS* of the polyploid *Triticum* species may indicate their co-evolution.

PHAS loci involved in male sterility

Numerous *PHAS* loci were observed in male reproductive organs, such as young spikes and anthers, in grasses. The *PHAS* loci were closely related to the fertility of the anther. Direct evidence is that the presence of *PMST1* showed that *PHAS* was involved in male sterility in rice [23]. Indirect evidence is the mutants of genes in the biogenesis pathway of phasiRNAs. The *DCL4* mutants in rice resulted in severe developmental defects in the spikelet, such as slight opening between the lemma and the palea. The lemma was partially or completely degenerated to the awn, and the strong loss-of-function transgenic plants were sterile [46]. A marked reduction of 21-nt phasiRNAs was observed in the *osdcl4–1* panicles [18]. The *OsDCL3b* RNAi lines led to reduced pollen fertility, seed setting rate and decreased grain yield in rice [47]. In addition, *osdcl3b* mutants specifically affected the generation of 24-nt phasiRNAs [18]. The *osrdr6–1* mutant was temperature sensitive and exhibited spikelet defects. The *osrdr6–1* mutants had a strong impact on the accumulation of both 21-nt and 24-nt phasiRNAs [19]. MEIOSIS ARRESTED AT LEPTOTENE1 (MEL1), a rice AGO, was shown to function in the development of pre-meiotic and meiotic procession, which preferred to load 24-nt phasiRNAs that bear a 5'-terminal cytosine [20]. In addition, *mel1* mutants also behaved sterile phenotype [48]. However, how *PHAS* leads to male sterility remains unclear. Whether all of the *PHAS* loci or phasiRNAs are associated with anther development or whether only several key *PHAS* loci or phasiRNAs play important roles is also unclear. The resolution of mysteries of *PHAS* in anther development remains a challenge for future studies.

PHAS or phasiRNAs may be photo-thermosensitive

The *PHAS* loci were preferentially identified in reproductive organs in grass species, which indicated their expression preference. Similar to phasiRNAs, few *PHAS* transcripts were expressed in the DR and FM of young spikes (*Figure 8A & B)*,, which were also found in maize [8] and rice [9]. Most of the *PHAS* in reproductive organs in wheat had low expression levels, and only 10.49% and 4.27% ofthe *PHAS* in the AM and TS stages had expression levels with RPKM > = 1 (*Figure 8C)*.. The expression level between AM

and TS was also different (*Figure 8C)*.. In addition, 274 *PHAS* in AM were differentially expressed at 6, 12, 24, and 48 hours after cold stress at 0°C (*Figure 8D)*.. This result indicated that *PHAS* not only involved in anther development but also could respond to cold stress. In addition to cold stress, the *PHAS* may also be sensitive to high temperatures in the development of anthers or flowers. The phenotype of *osrdr6* mutants was temperature dependent in rice. The phenotypic severity of *osrdr6–1* was enhanced with increased growth temperature under the fixed photoperiod. At lower temperatures of 28°C or 30°C, no obvious developmental defect or only 10% of the mutant spikelets showing a slight opening between lemma and palea was observed. At higher temperatures, the phenotypes of the mutant spikelets were very severe. At 32°C, the lemma was lost or became a completely or partially radial, awn-like structure or both the lemma and the pelea degenerated into awn-like structures. At 34°C, most of the lemma and pelea were defective, with shriveled or filament-like anthers. Both 21-nt and 24-nt phasiRNAs decreased in *osrdr6* at higher temperatures [19]. Thus, phasiRNAs or *PHAS* transcripts may be temperature dependent or thermo-sensitive. In addition to the thermo-response, some *PHAS* were photoperiod-sensitive, which affected the transition of male fertility and sterility under different photoperiod conditions. Under long-day conditions, *PMS1T* was targeted by miR2118 to produce 21-nt phasiRNAs, which affected rice fertility. Under short-day conditions or under long-day conditions with the mutated target site of miR2118, *PMS1T* could not produce 21-nt phasiRNAs because the transcript of *PMS1T* could not be recognized by miR2118 [23]. Overall, according to these features of *PHAS and* their associated key genes, the photothermosensitive genic male sterile lines may be a good model to study the further function of phasiRNAs or *PHAS,* and the detailed molecular mechanism of male sterility will be uncovered.

Methods

The identification of phasiRNAs and *PHAS* loci

The small RNA datasets used in this study were downloaded from the NCBI GEO database. In total, 261 small RNA datasets were collected, including 12 seedling samples, 128 leaf samples, 12 root samples, one stem sample, one shoot sample, 29 young spike samples, two anther samples, one embryo sample, 17 spikelet samples, 12 rachis samples, 19 grain samples, 20 seed samples, 6 callus samples, and one mixed tissue sample. (*Supplementary Table 1)*.. The reference genomes of AA (*Triticum urartu* (v1), 2n = 2x = 14), DD (*Aegilops tauschii* (v1), 2n = 2x = 14), AABB (*Triticum turgidum* (v1), 2n = 4x = 28) and AABBDD (*Triticum aestivum* (refv1.0), 2n = 6x = 42) were obtained from the URGI website (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). Next, the Trimmomatic program (version 0.38) was used to screen the raw small RNA datasets and remove the adaptor sequences and contaminated reads. Subsequently, the ShortStack program (version 3.8.5) [24] was used to align the cleaned data to the wheat reference genome (AABBDD). Then, the Shortstack package was used to map the small RNAs to the genome sequence. The distribution of small RNA on the reference genome was analyzed, and *PHAS* loci were identified with phased scores of greater than 15, 20, 25 or 30. According to the genome location of *PHAS,* the *PHAS* loci were annotated. A detailed flowchart is shown in *Supplementary Figure 1.*

The genome distribution of *PHAS* loci and their trigger miR2118 and miR2275

To show the distribution of *PHAS* loci on the genome, the total number of *PHAS* loci in each 500 kb window sliding each chromosome was calculated. Then, we used the Circos program (version 0.69–6) [49] to show the number of 21-nt (red lines in *Figure 2, 3 and 5)* and 24-nt *PHAS* (blue lines in *Figure 2, 3 and 5)* clusters on each chromosome in the wheat genome. The homoeologous chromosomes were filled with the same colors as the rainbow chromosomes (*Figure 2, 3 and 5)*.. The black lines in the rainbow chromosomes indicate the *PHAS* loci. The homoeologous *PHAS* loci among the homoeologous chromosomes were linked with the same color lines, and others were linked with different color lines (*Figure 3)*..

To identify miR2118 and miR2275, the mature sequences of the two families of miRNAs were downloaded from the miRBase website (http://www.mirbase.org/). Then, we mapped these sequences to the AA, DD, AABB, and AABBDD genomes using the BLAST program [50] with perfect matches. The distribution of miR2118 and miR2275 on each chromosome was drawn with the Mapchart program of the R package [51].

Target analysis of *PHAS* transcripts

To understand whether the *PHAS* clusters were regulated by miR2118 and miR2275 in wheat, we used the Targetfinder program [33] to predict the miRNA target sites in the *PHAS* sequences with scores less than four. To validate whether the miRNAs can indeed cleave the predicted targets, we first downloaded the degradome (SRP076763) datasets from the GEO database, including the control and cold treatment datasets [34]. The degradome datasets with the control and cold stress samples corresponded to the small RNA datasets for the control (SRR3680677 and SRR3680678) and cold stress samples (SRR3680679 and SRR3690680). Then, we used the two degradome libraries to confirm the target cleavage sites in the identified *PHAS* sequences in their corresponding control and cold stressed small RNA datasets with the CleaveLand program [35]. The cleavage sites were classified into 0, 1, 2, 3, and 4 categories according to the abundance of reads in the cleaved sites along the whole transcripts with a *P-value* less than 0.05. At the cleaved sites, the categories represented the following categories: category 4, only one read located at that position; category 3, >1 read but below or equal to the average depth of coverage on the transcript; category 2, >1 read above the average depth but not the maximum on the transcript; category 1, >1 read, equal to the maximum on the transcript when there was >1 position at the maximum value; and category 0, >1 read equal to the maximum on the transcript when there was just 1 position at the maximum value. Here, only these cleavages with the category < = 2 and *P-value* < 0.05 were selected for further study.

The expression level of *PHAS* transcripts

To understand the expression level of *PHAS* transcripts, we downloaded the transcriptome data of DR (SRR5464507 and SRR5464508), FM (SRR5464515 and SRR5464518), AM (SRR5464519 and SRR5464520) and TS (SRR5464523 and SRR5464524) stages of young spikes, which corresponded to the following small RNA datasets: DR (SRR5460930 and SRR5460939), FM (SRR5460941 and SRR5460949), AM (SRR5460967 and SRR5460972) and TS (SRR5461176 and SRR5461177), respectively. Then, the RSEM program [52] was used to calculate the gene expression value (RPKM) of the *PHAS* transcripts. To investigate whether these *PHAS* loci responded to abiotic stress, we compared the transcriptomes of control and cold stressed samples in the DR and AM samples of young spikes. Then, we used the DEseq program [36] to calculate the differentially expressed transcripts. The heatmaps of the *PHAS* transcripts were drawn using the Pheatmap program in the R package.

Conclusions

In summary, our results provided the first *PHAS* profiles in the young spike of wheat. They were also triggered by miR2118 and miR2275, which provided another evidence for the conservation of *PHAS* loci in generation and regulation mechanism of grasses. The increase of genome ploidy was the major drive force for the expansion of *PHAS* loci and their trigger miRNAs in *Triticum.* The sequence variations and biased distribution of *PHAS* in these genomes of AA, DD, AABB and AABBDD suggested their origination and diversity with rapid evolution in a small time-scale. Further study of the moleclular mechanism for these *PHAS* loci should improve our understanding of the function of phasiRNAs in the male reproductive developments of modern polyploid wheat.

Table 1

Table 1. The tissues of small RNA datasets used to identify the PHAS loci in wheat.

Tissues	Number of samples
Seedling	12
Leaf	128
Root	12
Stem	1
Shoot	1
Young Spike	29
Anther	2
Embryo	1
Spikelet	17
Rachis	12
Grain	19
Seed	20
Callus	6
Mixed tissues	1

Figures

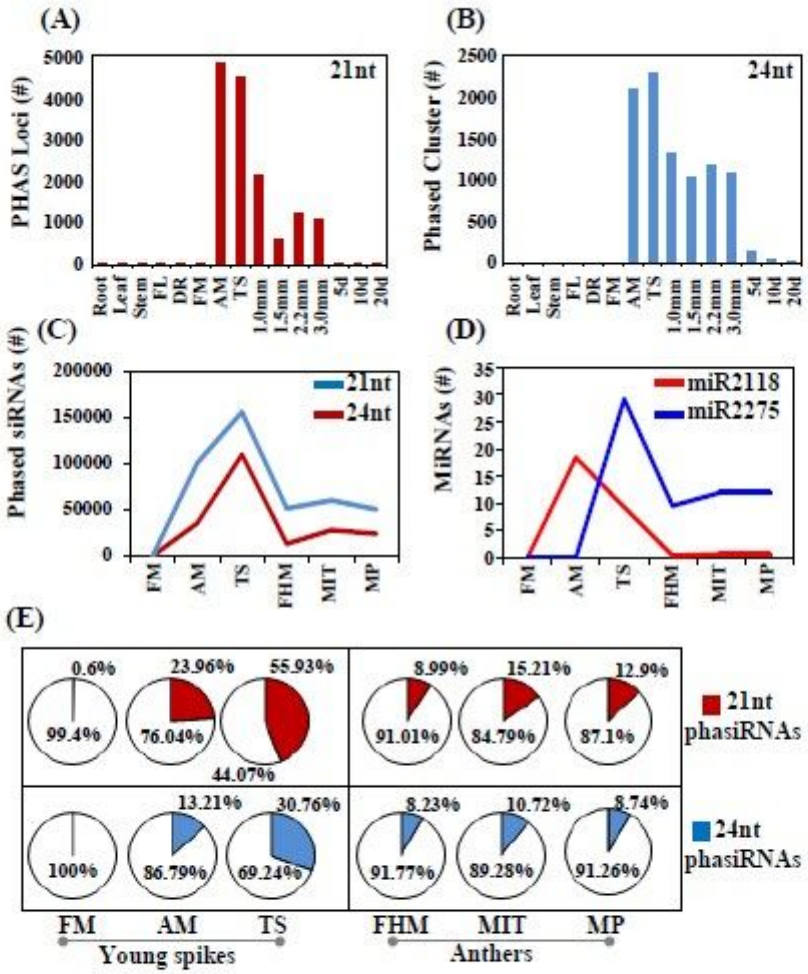

Figure 1

The number of 21- (A) or 24-PHAS loci (B) in roots, leaves, stems, flag leaves (FLs) young spikes (DR, FM, AM, TS), anthers with lengths of 1.0, 1.5, 2.2 and 3.0 mm, and grain 5, 10 and 20 days after flowering. (C) The expression level (TPM) of 21-nt phasiRNAs (blue lines) and 24-nt phasiRNAs (red lines). (D) The expression level (TPM) of miR2118 and miR2275 in the FM, AM, TS, FHM, MIT, and MP stages. (E) The proportion of 21-nt phasiRNAs (red color) and 24-nt phasiRNAs (blue color) in the total 21-nt siRNAs and 24-nt siRNAs, respectively.

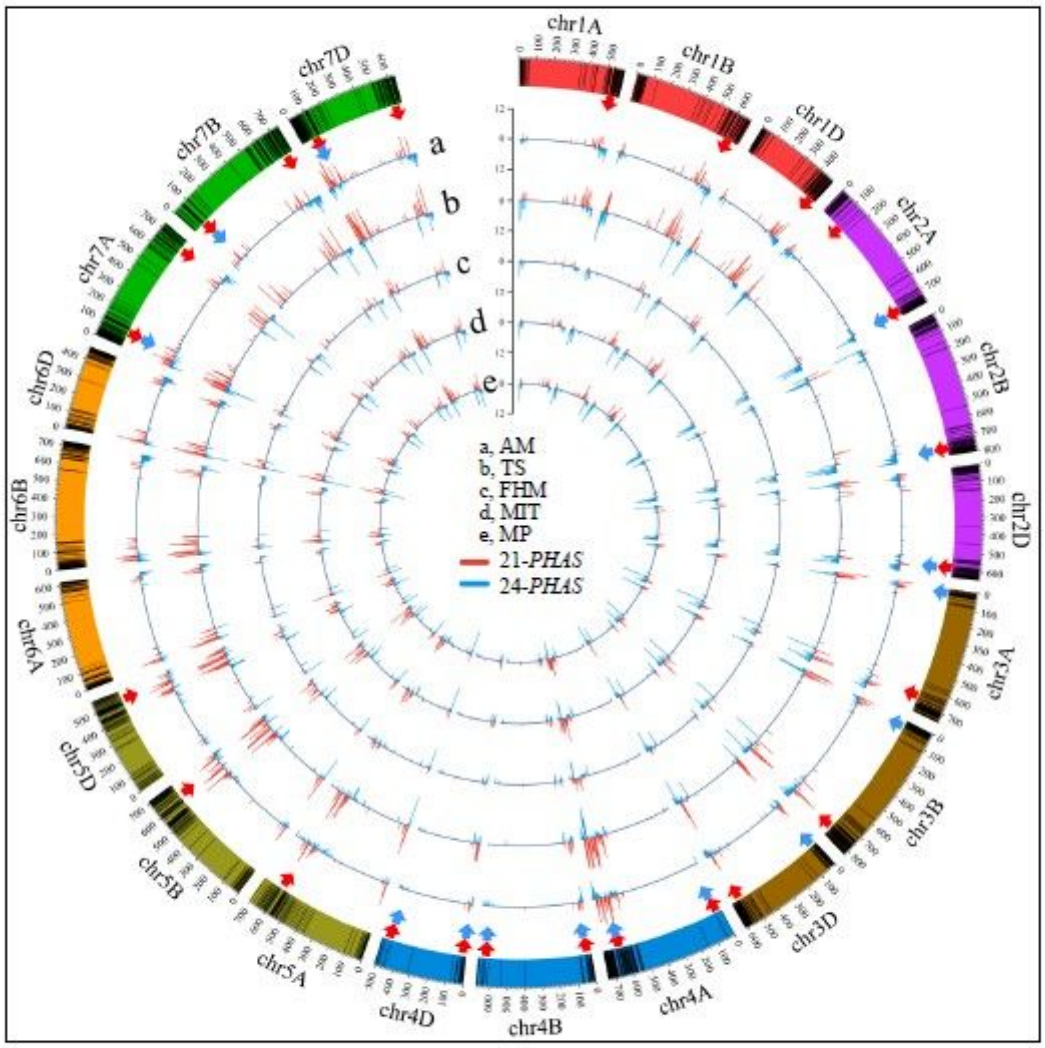

Figure 2

Density distribution of 21- (red lines) and 24-PHAS loci (blue lines) in the young spike samples. The black lines in the chromosome represent the PHAS loci. The peaks in circles a-e indicate the number of PHAS loci in each 500 kb region across each chromosome in AM, TS, FHM, MIT and MP, respectively. The red and blue arrows represent the biased distribution of 21- and 24-PHAS among the three subgenomes, respectively.

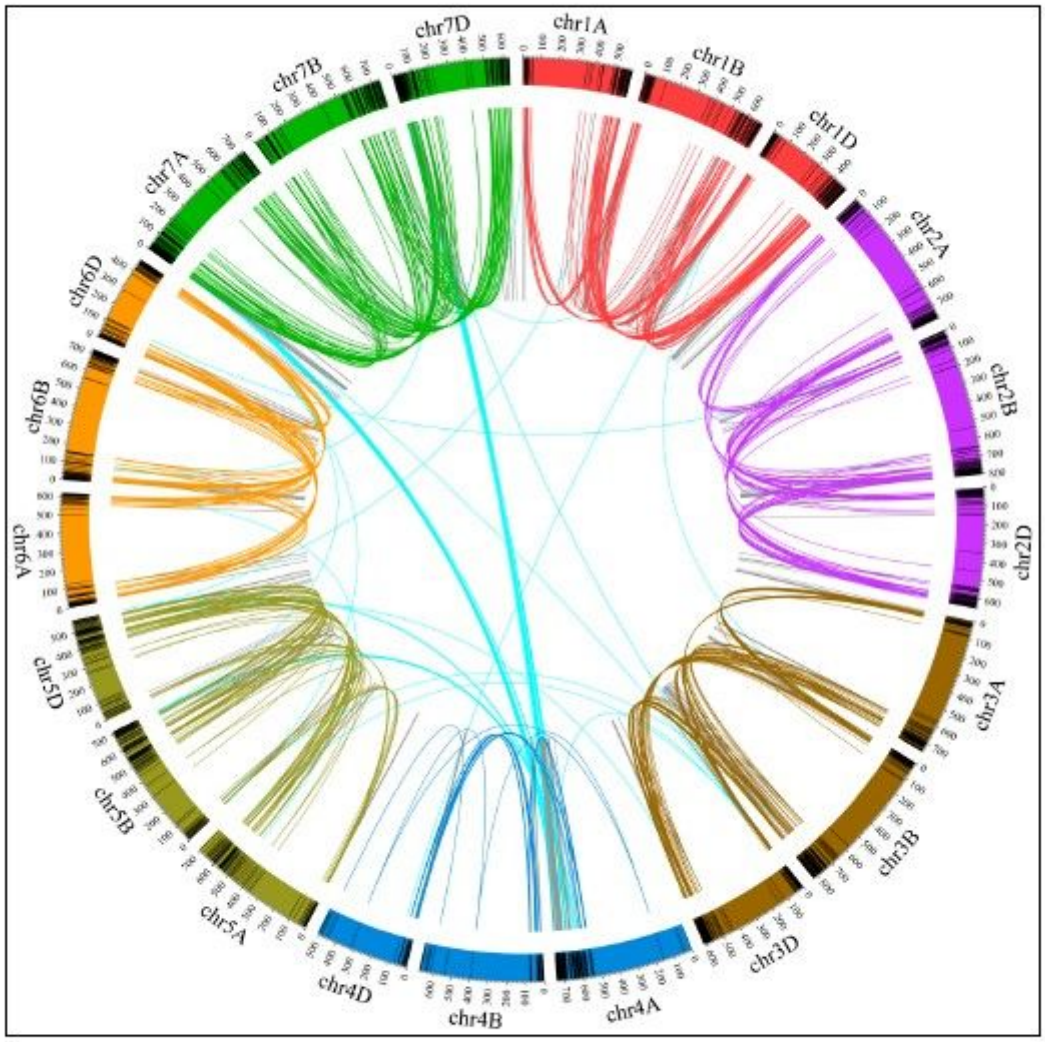

Figure 3

The duplication relationship among the three homologous chromosomes. The homoeologous PHAS loci are linked by lines with the same color. The black bar represents the PHAS loci in each chromosome.

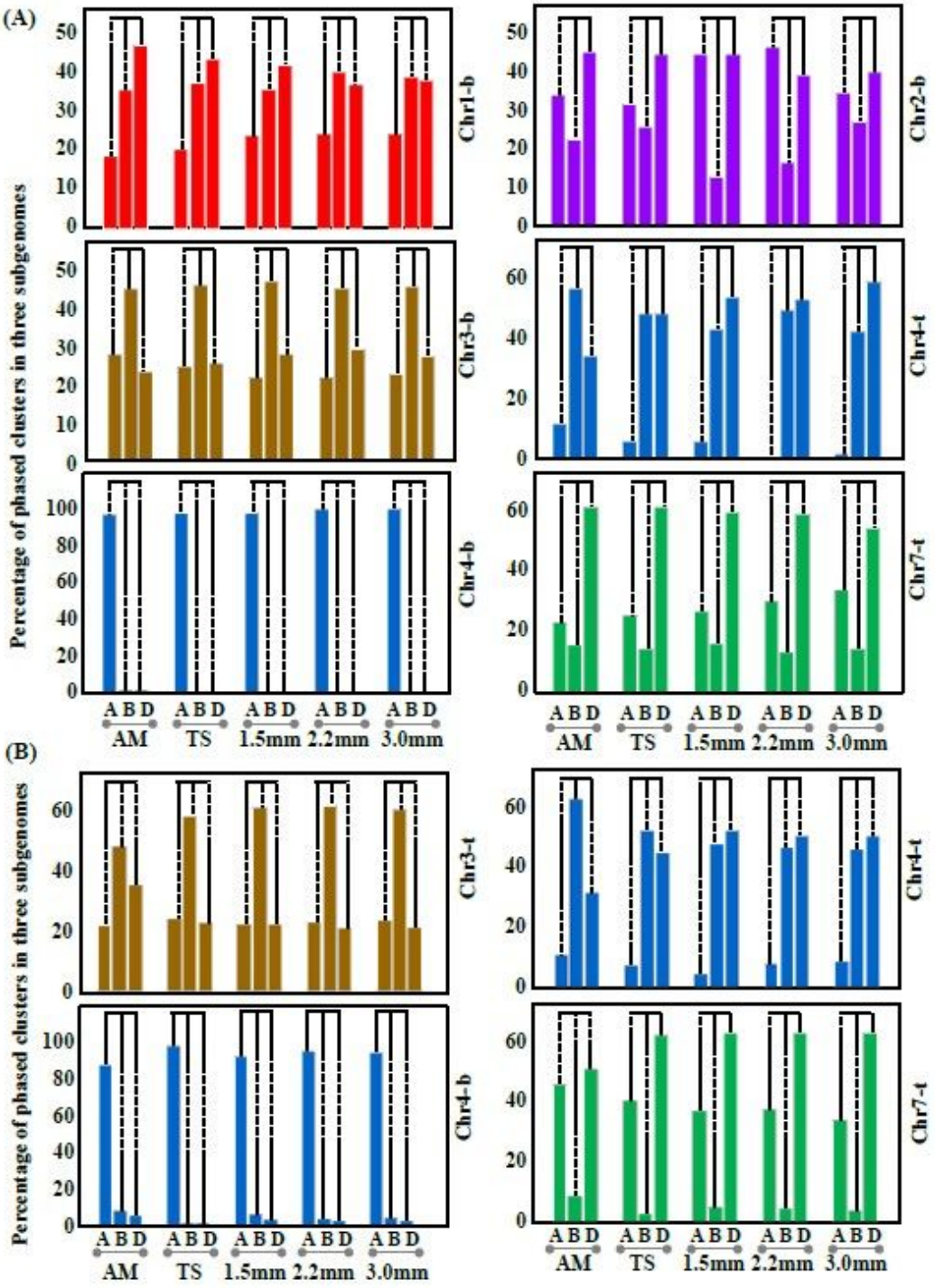

Figure 4

The biased distribution of 21- (A) and 24-PHAS loci (B) in each subgenome in AM, TS, FHM, MIT and MP. A, B and D indicate subA, subB and subD, respectively. “-t” and “-b” indictae the top of chromosome and bottom of chromosome from the centramere regions.

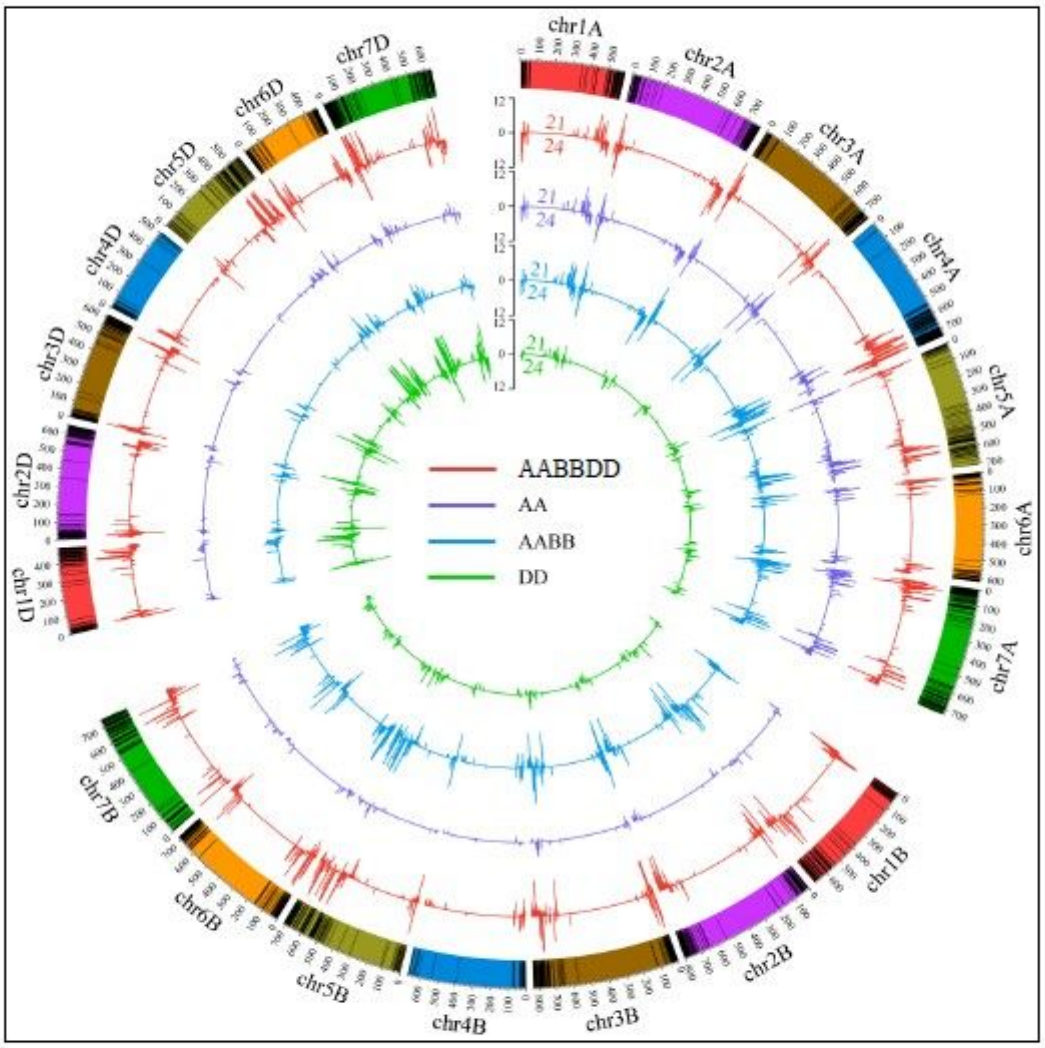

Figure 5

The density of 21- (top lines) and 24- (bottom lines) PHAS loci that can be mapped to the AABBDD (red lines), AA (purple lines), AABB (blue lines) and DD (green lines) genomes. The black bar represents the PHAS loci in each chromosome.

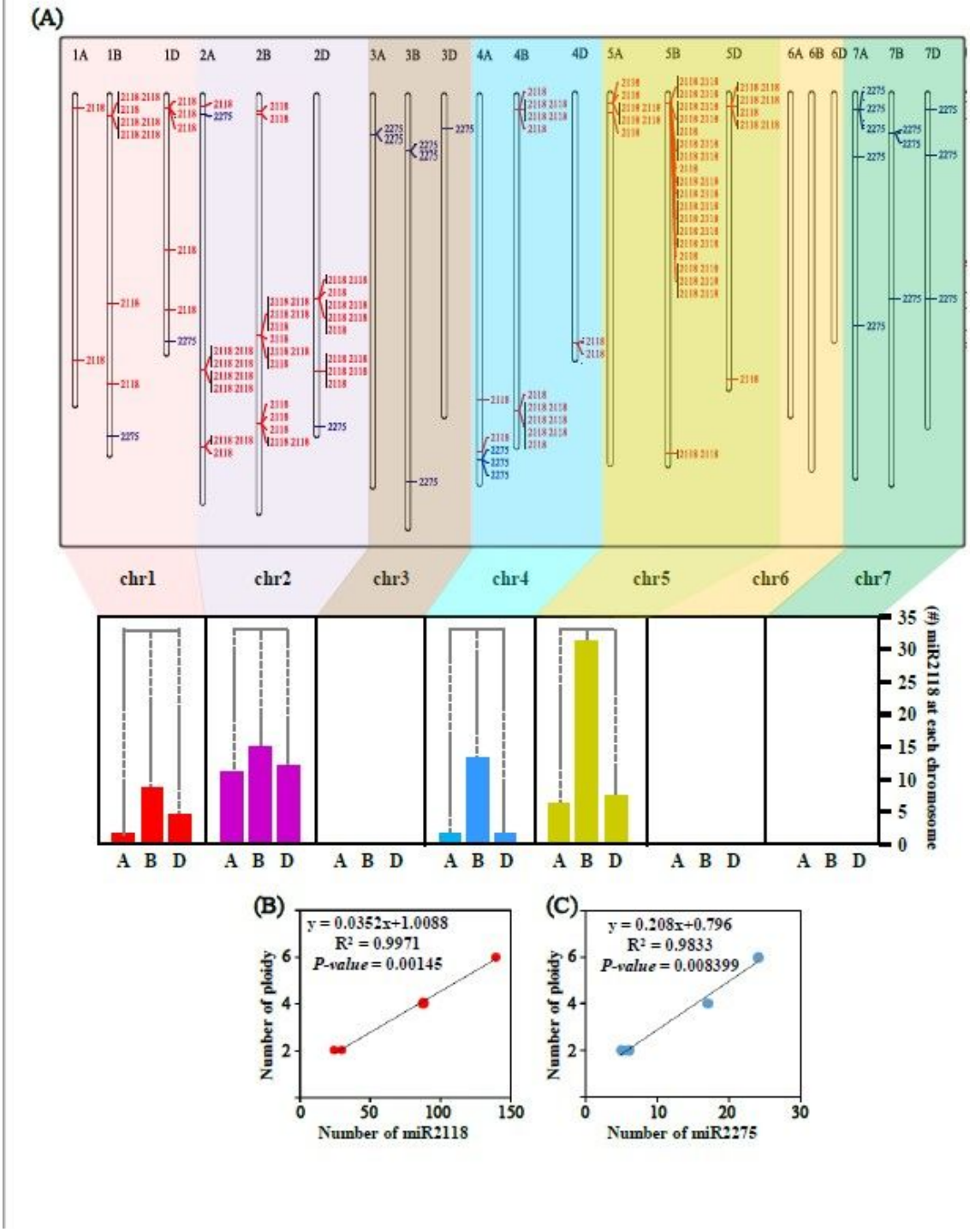

Figure 6

(A) The distribution of miR2118 (red lines) and miR2275 (blue lines) in each chromosome of wheat. (B-C) The correlation between the members of miR2118 (B) or miR2275 (C) and the times of ploidy. A, B and D represent the subA, subB and subD genomes, respectively.

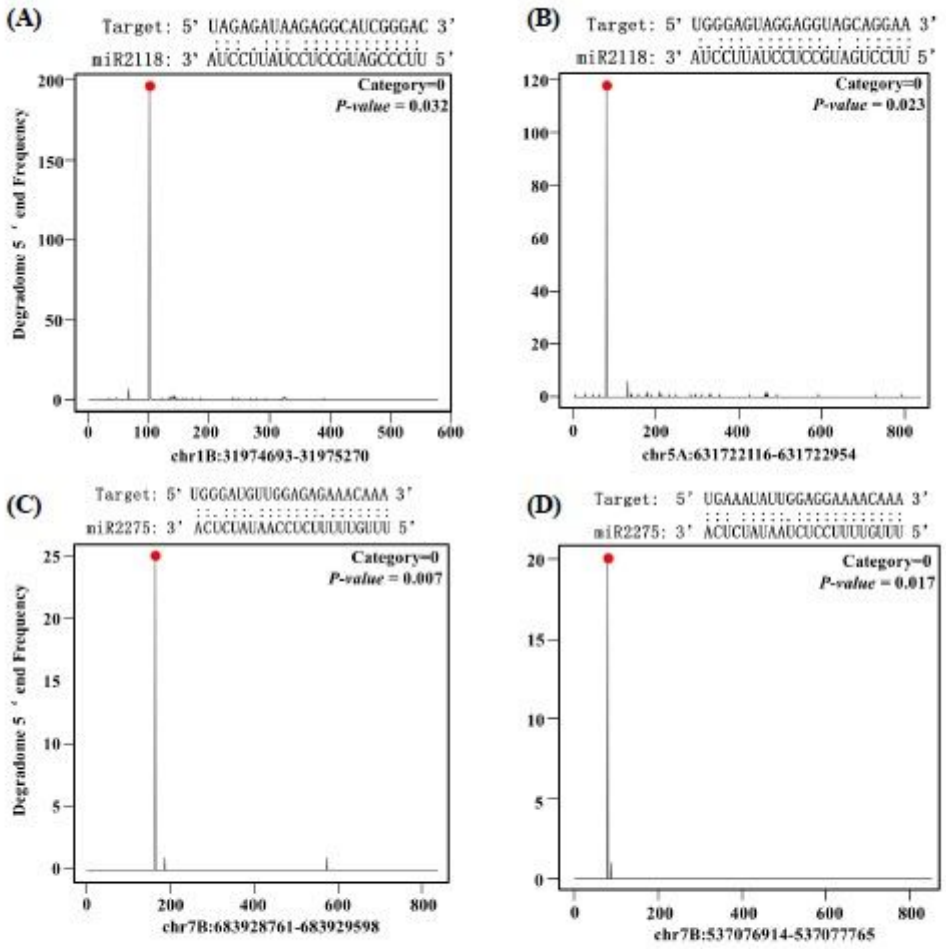

Figure 7

Target plots (T-plots) of miRNAs characterized in the degradome datasets. The abundance of signature tags was plotted along the indicated transcripts. The red dots indicate the predicted cleavage sites on the x-axis, and the black lines indicate the signatures produced by miRNA-directed cleavage.

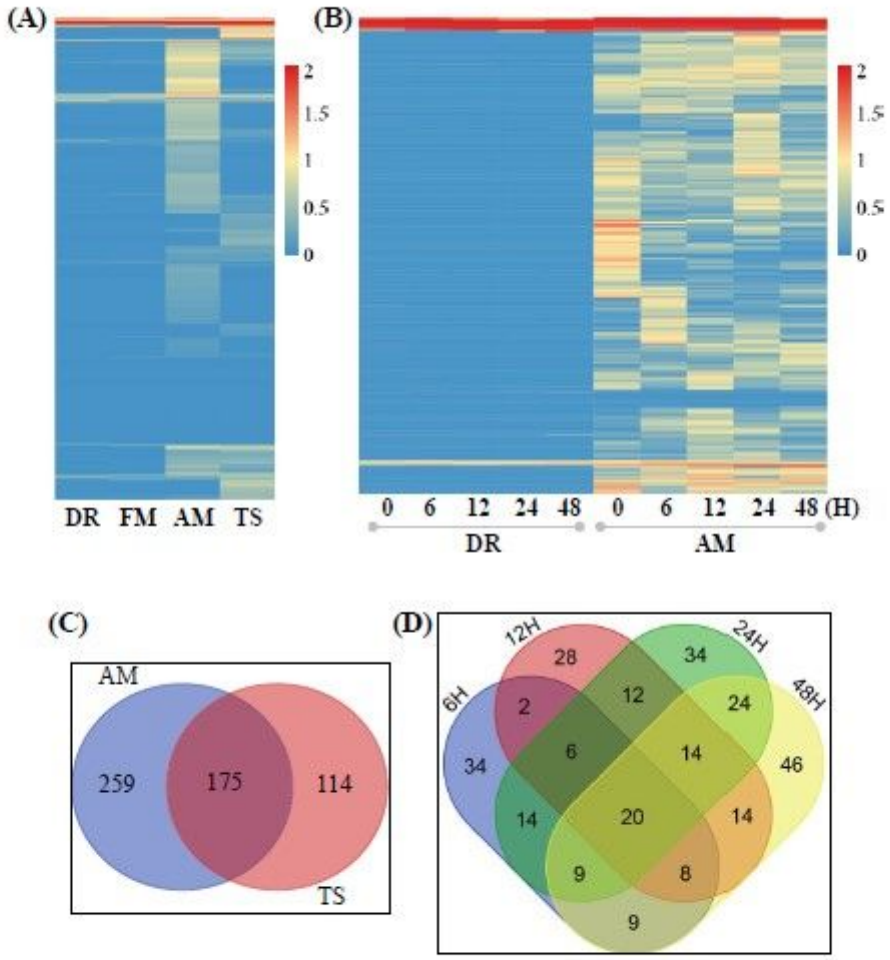

Figure 8

Expression heatmap of PHAS transcripts in the young spike of DR, FM, AM and TS and in the cold-stressed young spike transcriptome with polyA after 0, 6, 12, 24 and 48 hours in DR and AM.

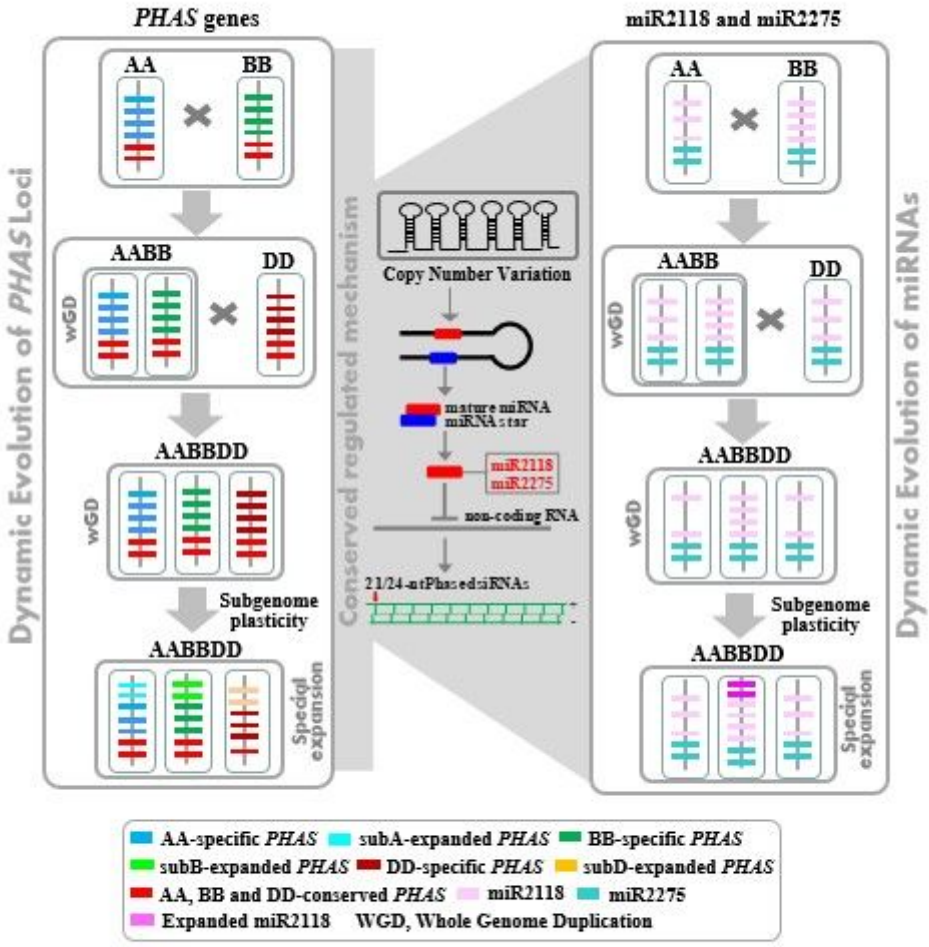

Figure 9

Dynamically evolved PHAS (the left panel) and their regulators miR2118 and miR2275 (the right panel) in the diploid, tetraploid and hexaploid Triticum species.

Supplementary Files

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

- Zhangetal20190929ST8.xlsx
- Zhangetal20190929ST9.xlsx
- Zhangetal20190929ST2.xlsx

- Zhangetal20190929ST1.xlsx
- Zhangetal20190929ST11.xlsx
- Zhangetal20190929ST3.xlsx
- Zhangetal20190929ST5.xlsx
- Zhangetal20190929ST4.xlsx
- Zhangetal20190929ST6.xlsx
- Zhangetal2019BMCSFig6.pdf
- Zhangetal2019BMCSFig4.pdf
- Zhangetal2019BMCSFig1.pdf
- Zhangetal2019BMCSFig9.pdf
- Zhangetal2019BMCSFig3.pdf
- Zhangetal2019BMCSFig7.pdf
- Zhangetal2019BMCSFig5.pdf
- Zhangetal20190929ST7.xlsx
- Zhangetal2019BMCSFig2.pdf
- Zhangetal20190929ST10.xlsx
- Zhangetal2019BMCSFig8.pdf