PROF. MEI   ZHANG (Orcid ID : 0000-0001-9803-0205)

PROF. BLAKE   MEYERS (Orcid ID : 0000-0003-3436-6097)

PROF. NATHAN   SPRINGER (Orcid ID : 0000-0002-7301-4759)

CHH DNA methylation increases at 24-*PHAS* loci depend on 24-nt phasiRNAs in maize meiotic anthers

Mei Zhang[1,†], Xuxu Ma[1,†], Chunyu Wang[1,2], Qing Li[3], Blake C. Meyers[4,5], Nathan M. Springer[3], Virginia Walbot[6,†]

[1] Key Laboratory of Plant Molecular Physiology, -Institute of Botany, Chinese Academy of Sciences, Nanxincun 20, Fragrant Hill, Beijing 100093, China.

[2] University of Chinese Academy of Sciences, Beijing 100049, China.

[3] Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108, USA.

[4] Donald Danforth Plant Science Center, St Louis, MO 63132, USA.

[5] University of Missouri-Columbia, Division of Plant Sciences, 52 Agriculture Lab, Columbia, MO 65211, USA.

[6] Department of Biology, Stanford University, Stanford, CA 94305 USA.

[†]These authors contributed equally to this work.

Author for correspondence:

Mei Zhang

Tel: +86(10) 6283-6251

Email: mei.zhang@ibcas.ac.cn

ORCID

Mei Zhang	http://orcid.org/0000-0001-9803-0205

Xuxu Ma	https://orcid.org/0000-0002-3206-8187

Chunyu Wang	https://orcid.org/0000-0001-7104-3874

Blake C. Meyers	http://orcid.org/0000-0003-3436-6097

Nathan M. Springer	http://orcid.org/0000-0002-7301-4759

Summary

- Plant phasiRNAs contribute to robust male fertility, however, specific functions remain undefined. In maize (*Zea mays*), *male sterile23* (*ms23*), necessary for both 24-nt phasiRNA precursor (24-*PHAS*) loci and *Dicer-like5* (*Dcl5*) expression, and *dcl5-1* mutants unable to slice *PHAS* transcripts, lack nearly all 24-nt phasiRNAs.

- Based on sequence capture bisulfite-sequencing, we find that CHH DNA methylation of most 24-*PHAS* loci is increased in meiotic anthers of control plants but not in the *ms23* and *dcl5* mutants.

- Because *dcl5-1* anthers express *PHAS* precursors, we conclude that the 24-nt phasiRNAs, rather than just activation of *PHAS* transcription, are required for targeting increased CHH methylation at these loci.

- Although *PHAS* precursors are processed into multiple 24-nt phasiRNA products, there is substantial differential product accumulation. Abundant 24-nt phasiRNA positions corresponded to high CHH methylation within individual loci, reinforcing the conclusion that 24-nt phasiRNAs contribute to increased CHH methylation in *cis*.

Key words: phasiRNA, DNA methylation, *Dcl5*, *Ms23*, *Zea mays* (maize)

Introduction

In flowering plants, small RNAs (sRNAs) are key regulators during reproduction in both the sporophyte and gametophyte (Borges & Martienssen, 2015). Reproductive phasiRNAs were first described in male reproductive organs of maize and rice (Johnson *et al.*, 2009; Song *et al.*, 2012a; Zhai *et al.*, 2015; Fei *et al.*, 2016) and are now known in diverse monocots and eudicots (Kakrana *et al.*, 2018; Xia *et al.*, 2019). Reproductive phasiRNAs exist in two size classes with discrete spatiotemporal regulation in anthers: the "premeiotic" 21-nt phasiRNAs are preferentially expressed during the cell fate specification and proliferation period, while the "meiotic" 24-nt phasiRNAs begin to be synthesized at the onset of meiosis and accumulate to high levels during meiosis I (Zhai *et al.*, 2015). The 24-nt reproductive phasiRNAs are broadly present in angiosperms, and they may have originated with the evolutionary emergence of anthers (Xia *et al.*, 2015; Shuai *et al.*, 2016; Kakrana *et al.*, 2018; Liu *et al.*, 2018; Yu *et al.*, 2018; Xia *et al.*, 2019).

The exact functions of phasiRNAs in anthers are unknown, but they play critical roles in male reproduction. For example, a mutation in long non-coding RNA *PMS1T*, a *PHAS* precursor yielding 21-nt phasiRNAs, is responsible for the agronomically important, photoperiod-sensitive male sterility used in hybrid rice production (Fan *et al.*, 2016). In maize, *dcl5-1* mutants, which are deficient in 24-nt phasiRNAs, exhibit conditional male fertility, with the main defect in tapetal cell differentiation during meiosis (Teng *et al.*, 2020). Quantification of *PHAS* precursor transcripts and phasiRNA abundance in several maize male sterile mutants, demonstrated that production of 21-nt phasiRNAs requires normal epidermal development, while 24-nt phasiRNAs require normally developed tapetal cells (Zhai *et al.*, 2015). In rice, tapetal cell differentiation during meiosis is also required for 24-nt phasiRNA biogenesis (Ono *et al.*, 2018). A fascinating aspect of male reproductive phasiRNAs is the parallel with mammalian PIWI-associated RNAs (piRNAs). Like phasiRNAs, there are two size classes, the shorter class accumulating prior to meiosis and the longer class occurring during meiosis (Girard *et al*., 2006; Grivna *et al*., 2006; Vagin *et al*., 2006; Fu & Wang, 2014). Furthermore, like phasiRNAs, piRNAs are required for spermatogenesis (Fu & Wang, 2014), and are partly generated in a phased

49 pattern (Han *et al.*, 2015; Mohn *et al.*, 2015).

50 The current understanding of phasiRNA biogenesis is that the precursors are long,

51 non-coding RNAs transcribed by DNA-DEPENDENT RNA POLYMERASE II (Pol II)

52 from several hundred *PHAS* loci. The production of 21- and 24-nt phasiRNAs is initiated

53 by 22-nt miR2118 or miR2275, directing the cleavage of 21- or 24-*PHAS* primary

54 transcripts, respectively (Johnson *et al.*, 2009; Arikit *et al.*, 2013; Zhai *et al.*, 2015).

55 RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) converts the resulting 3' cleaved

56 RNA fragments into double-stranded RNA (Song *et al.*, 2012b). This dsRNA is the

57 substrate for phased cleavage by a Dicer enzyme; DCL4 is proposed for 21-nt phasiRNAs

58 and DCL5 for the 24-nt class based on expression timing (Song *et al.*, 2012a; Arikit *et al.*,

59 2013). *Dcl5* is a duplication of *Dcl3* found in many monocots (Xia *et al.*, 2019). Proof that

60 DCL5 is required for 24-nt phasiRNA biogenesis, but not generation of other small RNAs,

61 comes from analysis of multiple *dcl5* alleles in maize, which lack nearly all 24-nt

62 phasiRNAs (Teng *et al.*, 2020). The entire 24-nt phasiRNA biogenesis pathway appears to

63 be regulated by MS23, a basic helix-loop-helix (bHLH) transcription factor that controls

64 the initial steps in tapetal differentiation. In maize *ms23* mutants with defective tapetum

65 development, *PHAS* precursors, the normal suite of miR2275 types, *Dcl5*, and 24-nt

66 phasiRNAs are missing (Zhai *et al.*, 2015; Nan *et al.*, 2017); additionally, several

67 thousand protein-coding genes are mis-regulated (Nan *et al.*, 2017).

68 Biogenesis of phasiRNAs in grasses parallels that of trans-acting small interfering

69 RNAs (tasiRNAs) (Arikit *et al.*, 2013; Deng *et al.*, 2018). The tasiRNAs interact with

70 mRNAs and result in mRNA degradation or block translation by virtue of

71 complementarity. *Arabidopsis thaliana* has four *TAS* gene families -- *TAS1*, *TAS2*, *TAS3*,

72 and *TAS4* -- and they are present at eight genetic loci (Allen *et al.*, 2005; Rajagopalan *et*

73 *al.*, 2006). In addition to targeting mRNAs, 21-nt tasiRNAs from *TAS3c* and *TAS1a* were

74 reported to direct DNA methylation at their cognate *TAS*-encoding loci (Wu *et al.*, 2012;

75 Wu, 2013), a type of regulation defined as "in *cis*" in the plant small RNA community.

76 Elevated DNA methylation of the *TAS3c* locus requires DCL1, while multiple DCLs

77 contribute to DNA methylation of *TAS1* loci (Wu *et al.*, 2012; Wu, 2013). Although 24-nt

78   maize phasiRNAs are proposed to be synthesized in the tapetum (Teng *et al.*, 2020), they

79   also accumulate in meiocytes (Zhai *et al.*, 2015). Dukowic-Schulze *et al.* (2016) proposed

80   that phasiRNA could be involved in *cis* DNA methylation at their own loci, because they

81   detected elevated CHH DNA methylation at *PHAS* loci in zygotene stage maize meiocytes

82   and in whole anthers compared to levels in seedlings. In mammalian testes, a small

83   fraction of piRNAs maintain genome integrity by epigenetically silencing transposons via

84   DNA methylation, especially in germline stem cells (Aravin *et al.*, 2007; Le Thomas *et al.*,

85   2013; Luteijn & Ketting, 2013; Fu & Wang, 2014; Zhai *et al.*, 2015); this possible

86   function has not been explored for 24-nt plant phasiRNAs, however, the plant phasiRNAs

87   are not complementary to known transposons.

88   Building on the foundation established by Dukowic-Schulze *et al*. (2016), we wished to

89   address two questions. First, do the 24-*PHAS* loci of anthers acquire elevated CHH DNA

90   methylation after synthesis of the 24-nt phasiRNAs or do anthers have a constitutively

91   elevated level? Second, if there is developmental regulation of CHH DNA methylation,

92   does it occur after transcriptional activation of 24-*PHAS* loci or does it require successful

93   generation of the 24-nt phasiRNA products? To answer these questions and explore the

94   potential role of 24-nt phasiRNAs in maize DNA methylation, we generated sequence

95   capture bisulfite-sequencing data from two male sterile mutants, *dcl5-1* and *ms23*, where

96   24-nt phasiRNAs are highly reduced or eliminated, and their control plants.

MATERIALS AND METHODS

Plant materials

Maize was grown in a greenhouse at Stanford, CA with a 14-h day (28 ℃)/ 10-h night (22 ℃) light cycle. Dissected anthers were staged using a micrometer.

For the *dcl5* mutant, we utilized the *dcl5-1* allele with a one-base pair deletion in the coding region of the *Dcl5* gene (Teng *et al.*, 2020). We started with the T_0 generation heterozygous *dcl5-1//+* plants in the HiII background. After several generations of self-crossing or crossing among sibling plants, we obtained T_3 and T_4 generation 1:2:1 populations for dissecting the anthers of the homozygous *dcl5-1* (*dcl5-1//dcl5-1*) mutant and its wildtype fertile control (*+//+*). For *ms23*, a 1:2:1 population of *ms23-ref* in the ND101 background introgressed several times into the W23 inbred line was used; anthers from the homozygous *ms23* (*ms23//ms23*) mutant and its fertile control (*ms23//+* or *+//+*) were collected. To simplify figure labeling, the fully fertile sibling plants compared to mutants in the same family are referred to as "control plants". Table **S1** provides detailed information about the biological samples.

WGBS library preparation and data analysis

Whole genome bisulfite sequencing (WGBS) was performed using 0.5 mm, 2.0 mm, and 4.0 mm W23 anthers. DNA was extracted from each sample using CTAB, and libraries were prepared for WGBS as previously described (Urich *et al.*, 2015). Sequencing was performed on a HiSeq 2000 platform and generated 101-bp paired reads, followed by quality control using FASTQC and adaptor trimming using Trim Galore. For each sample, reads were aligned to the maize B73 reference genome B73_ AGPv4 (Jiao *et al.*, 2017) by BSMAP (Xi & Li, 2009). We calculated the methylation level for each cytosine using methratio.py in BSMAP in the uniquely mapped, duplication-removed and properly paired reads.

ViewBS was utilized to construct the plots based on WGBS data (Huang *et al.*, 2018). Briefly, the 2 kb up- and down-stream regions of 21- and 24-*PHAS* loci, protein-coding genes, and transposons (TEs), and their body regions were divided into 100, 60, and 100

125　bins, separately. Then, the methylation level of each bin was summarized by dividing

126　detected methylated cytosines (#C) by total cytosines (#C and # delta T) (Schultz *et al.*,

127　2012).

128　**Sequence capture bisulfite-sequencing library preparation and data mapping**

129　To gain high-resolution methylation profiles for 24-*PHAS* loci, we chose ~20,000 specific

130　regions from the B73 genome, including 176 24-*PHAS* and 463 21-*PHAS* loci (Zhai *et al.*,

131　2015). We found that 122 of the 176 24-*PHAS* loci and 416 of 463 21-*PHAS* loci

132　identified in the B73 reference genome v4, contained a predicted miR2275 or miR2118

133　cleavage site (CS), and were used as analyzable *PHAS* loci in our study. The capture

134　probes of those regions were designed by Roche (Han *et al.*, 2018). Briefly, we

135　constructed a standard genome bisulfite library and extracted capture target regions by

136　hybridization to probes as described previously (Li *et al.*, 2014). Then the capture library

137　was sequenced on an Illumina MiSeq instrument using 150 cycles from both ends after

138　PCR amplification. The computational methods described for WGBS data were used to

139　analyze sequence capture bisulfite-sequencing data.

140　　　To plot the methylation profiles at 21- and 24-*PHAS* loci based on sequence capture

141　bisulfite-sequencing, the 200 bp upstream region, transcribed regions prior to the CS, the

142　transcribed region after the CS, and the 200 bp downstream region were divided into 5, 15,

143　45, and 5 bins, respectively. The same strategy was utilized to calculate the DNA

144　methylation of each bin as for WGBS data.

145　　　Handling replicate samples: 1) The replicates of 2.0 mm anther of *dcl5-1* mutant and its

146　controls were from the T_3 and T_4 generations; data were analyzed separately for

147　methylation level change among 21-*PHAS* and 24-*PHAS* loci. Subsequently, only the T_3

148　generation data were used, because of the relatively low coverage in the T_4 generation

149　dataset. 2) Regarding the replicates of 1.0 mm anthers, we used the average for the

150　analysis of methylation levels at 24-*PHAS* loci and 21-*PHAS* loci in the *dcl5-1* mutant and

151　its controls in pre-meiotic (0.5 and 1.0 mm) and meiotic 2.0 mm anthers. Other biological

152　samples did not have replicates.

sRNA-seq and RNA-seq library construction and sequencing

sRNA-seq and RNA-seq libraries were prepared as previously described (Mathioni *et al.*, 2017; Teng *et al.*, 2020). In brief, total RNA was isolated and libraries were constructed for *dcl5-1* and its wild-type control, with two biological replicates, and sequenced using the Illumina HiSeq2500 platform.

For sRNA-seq data, adapters were trimmed by Trim Galore (version 0.6.0), and Bowtie (version 1.2.2) (Langmead *et al.*, 2009) was performed for the alignment with the parameter "-q -e 1 -l 31 -n 0 -p 5 -m 20 -a -S --nomaqround --chunkmbs 524" to the maize B73 reference genome v4. Only uniquely mapped reads were used in this study. The abundance of sRNA is normalized on the basis of the total count of mapped reads in that sample. The abundance of 24-nt phasiRNAs in the *ms23* mutant and its control was calculated based on the published sRNA-seq data from Zhai *et al*. (2015).

For RNA-seq data, read quality was checked using Fastp (Chen *et al.*, 2018) and then mapped to the maize B73 reference genome B73_ AGPv4 (Jiao *et al.*, 2017) by Hisat2 (Kim *et al.*, 2015). Only uniquely aligned reads, as input for StringTie (Pertea *et al.*, 2015), were used to calculate 24-*PHAS* transcript expression levels.

Updating the list of 24-*PHAS* loci in maize anthers

sRNA-seq data from 2.0 mm anthers of the control of the *dcl5-1* mutant were used for 24-*PHAS* detection. Only 24-nt intervals were used for genome-wide phasing analysis, and the phasing scores were calculated as previously described (De Paoli *et al.*, 2009). With the criterion of a phasing score higher than 25 plus a manual check, we had an updated list of 151 24-*PHAS* loci, 23 of which were newly identified.

Identification of 24-*PHAS* in the 2.0 mm *dcl5-1* mutant anthers

Considering the relatively low abundance of 24-nt small RNAs mapped to 24-*PHAS* in *dcl5-1* 2.0 mm anthers, we required the largest phasing score of one 24-nt siRNA per locus to be greater than or about 20. For some examples with an obviously good phasing

179   pattern -- at least 3 continuous and high phasing score elements with a 24-nt spacing

180   interval by manual checking -- the largest phasing score can be low as 15.

181   **CHH DMR detection**

182   To detect differentially methylated regions, we summarized CHH methylation levels for

183   each 50-bp non-overlapping sliding window with CHHs and at least 10 reads, dividing

184   detected methylated cytosines (#C) by total cytosines (#C and # delta T). Then, the

185   p-value of Fisher's exact test was calculated for each window, to weigh the significance of

186   allele bias. The criteria were as follows: 1) Q-values (converted by p-value) < 0.05

187   between the DNA methylation level of the mutant and its control; 2) absolute methylation

188   difference between wild-type and mutant> 5%; 3) and the change relative to wild-type

189   ((wild-type - mutant)/wild-type) greater than 40%. Finally, we combined the adjoining

190   windows detected as DMRs within 200 bp, and then recalculated the percentage using the

191   same strategy.

192   **Abundant 24-nt siRNA position detection**

193   Because we observed that 24-nt phasiRNAs are unevenly distributed across individual

194   24-*PHAS* loci, we defined the 24-nt siRNA with the highest number of normalized reads

195   as the most abundant 24-nt siRNA position. Further, if other 24-nt siRNA positions

196   possessed a representation level at least half as high as the most abundant 24-nt siRNA

197   position, we also considered these 24-nt siRNA positions as abundant positions.

198

RESULTS

200 **24-*PHAS* exhibited increased CHH DNA methylation only at the meiotic stage of**

201 **maize anther development**

202 Meiosis is an essential stage in sexual reproduction. To explore if any DNA methylation

203 changes occur before, during, or after meiosis, we performed low-coverage whole genome

204 bisulfite sequencing at three stages of W23 inbred anther development: 0.5 mm anther

205 length (cell proliferation and differentiation), 2.0 mm (prophase I of meiosis), and 4.0 mm

206 (microspore development) anthers (Kelliher & Walbot, 2011) (Fig. **1a-c**; Table **S1**). First,

207 the average levels of CG, CHG, and CHH methylation were determined for each stage

208 (Fig. **1d-f**). The W23 anthers had DNA methylation levels for the CG (75%), CHG (60%),

209 and CHH (2%) contexts that are similar to levels in other inbred lines and tissues

210 examined in previous studies (Li *et al.*, 2014; Zhang *et al.*, 2014). Second, we compared

211 DNA methylation of protein-coding genes, transposons (TEs), and the *PHAS* loci among

212 these three stages. The up- and down-stream regions of protein-coding genes display

213 elevated CHH context DNA methylation at the 2.0 mm stage: about 2-3 times higher than

214 either the 0.5 or 4.0 mm stage (Fig. **1i**). In contrast, there is no major DNA methylation

215 change in the CG and CHG contexts in the gene body regions of protein-coding genes or

216 any contexts for the TEs across the three stages (Fig. **1j**, Fig. **S1c, d, g, h**). Prior work has

217 shown that the 21-nt phasiRNAs peak in abundance by 0.5 mm while the 24-nt

218 phasiRNAs peak at 2.0 mm (Zhai *et al.*, 2015), therefore, we determined if the

219 corresponding *PHAS* loci are differentially methylated at these stages or show any

220 differences relative to protein-coding genes (Fig. **1g, h,** Fig. **S1a, b, e, f**). As with

221 protein-coding genes, no major DNA methylation differences were found in the CG or

222 CHG context across the three stages for either the 21- or 24-*PHAS* loci (Fig. **S1a, b, e, f**).

223 Strikingly, in the CHH context, both 21- and 24-*PHAS* transcribed regions exhibit the

224 highest DNA methylation at 2.0 mm, followed by 0.5 mm; the lowest level occurred at the

225 4.0 mm stage (Fig. **1g, h**).

226 To quantify the stage-specific effects of CHH DNA methylation at various genomic

227 features, we separately analyzed their relative DNA methylation changes compared to the

228 0.5 mm stage (Fig. **1k**). All regions analyzed exhibit increased CHH methylation at the 2.0

229 mm stage and reduced CHH methylation at 4.0 mm suggesting dynamic levels of CHH

230 methylation during anther development. Further, we observed that the CHH methylation

231 levels of the transcribed and the up-stream regions of 24-*PHAS* loci, as well as the up- and

232 down-stream regions of protein-coding genes, are more than 50% higher than that of the

233 0.5 mm stage (Fig. **1k**). Taken together, these results suggest that 2.0 mm meiotic anthers

234 exhibit distinct CHH DNA methylation patterns and establish that there is developmental

235 regulation of this epigenetic feature.

236 **Elevated CHH DNA methylation of 24-*PHAS* at the meiotic stage is related to the**

237 **presence of 24-nt phasiRNAs**

238 The period of peak DNA methylation at 2.0 mm was investigated further with sequence

239 capture bisulfite-sequencing to enrich for *PHAS* loci. Our main question was whether the

240 elevated CHH DNA methylation on 24-*PHAS* at the 2.0 mm stage depends on the

241 activation of transcription of the loci or on 24-nt phasiRNA abundance. Two male-sterile

242 maize mutants with a profound loss of 24-nt phasiRNAs were utilized. In the *ms23* mutant

243 -- *MS23* encodes a master transcriptional regulator of tapetal development -- few 24-*PHAS*

244 transcripts exist, there are almost no 24-nt phasiRNAs, and *Dcl5* is not expressed (Nan *et*

245 *al.*, 2017). In the downstream *dcl5-1* mutant, 24-*PHAS* transcripts are as high or higher

246 than in the control plants, however, in the absence of DCL5 protein, very few 24-nt

247 phasiRNAs are present (Teng *et al.*, 2020). Analysis of these two mutants can distinguish

248 whether transcriptional activation of the 24-*PHAS* loci or the production of 24-nt

249 phasiRNAs are required for the change in CHH context DNA methylation at 24-*PHAS*

250 loci in 2.0 mm anthers.

251 To pursue the analysis, we generated in-depth, sequence capture bisulfite-sequencing

252 data, as well as sRNA-seq data for 2.0 mm anthers of fully sterile homozygous *dcl5-1*

253 mutants, and the control plants in the T_3 generation (see the Materials and Methods

254 section). In the *dcl5-1* anthers, CG and CHG DNA methylation levels are

255 indistinguishable from control plants (Fig. **2a, b**). There was obviously reduced DNA

256 methylation in the CHH context within the transcribed region of 24-*PHAS* loci,

257   particularly 3' of the miRNA cleavage site (CS), in the *dcl5-1* sterile individuals,

258   compared to its control (Fig. **2c**). In contrast, 21-*PHAS* loci CHH methylation levels are

259   similar in *dcl5-1* and its control anthers at 2.0 mm (Fig. **S2**), about 10 days after peak

260   transcript and 21-nt phasiRNA abundances (Zhai *et al.*, 2015). To confirm these results,

261   we generated sequence capture bisulfite-sequencing data from 2.0 mm anthers in T_4

262   generation plants. The substantial CHH methylation difference was present in the

263   transcribed region of 24-*PHAS* loci, but not in the 21-*PHAS* loci in the *dcl5-1* mutant (Fig.

264   **S3**). Because the sequence capture bisulfite-sequencing data from 2.0 mm anthers in the

265   T_4 generation plants were limited, we did further analysis with data from the T_3 generation

266   plants.

267   In analyzing the sequence capture bisulfite-sequencing data for the 2.0 mm anthers of

268   the *ms23* mutant, we found slightly elevated CG and CHG DNA methylation in the

269   24-*PHAS* transcribed regions in the mutant compared to its fertile control (Fig. **2d, e**).

270   Paralleling the results with *dcl5-1,* CHH DNA methylation levels were dramatically

271   reduced 3' of the miR2275-mediated CS across the transcribed region compared to its

272   fertile control (Fig. **2f**). In contrast, no major CHH DNA methylation difference was

273   observed at the 21-*PHAS* loci, and CG and CHG DNA methylation levels at these loci

274   were elevated in the mutant as was observed for the 24-*PHAS* loci (Fig. **S4**). We

275   concluded that these elevated CG and CHG DNA methylation levels in the *ms23* mutant

276   do not compensate for reduced CHH methylation here, because it also occurred in

277   21-*PHAS* loci (without reduced CHH DNA methylation) as well. We propose that absence

278   of the MS23 protein (and failure of tapetal cells to specify and differentiate) affected DNA

279   demethylases required to maintain the CG and CHG methylation in the genome. Because

280   the tapetal cells are such a large fraction (approximately one-third, Kelliher & Walbot,

281   2011) of all anther cells at 2.0 mm, significant changes in this one cell type will show up

282   in whole anther data.

283   Collectively, these data establish that defects in genes required for 24-nt phasiRNA

284   biogenesis result in lower CHH DNA methylation in the corresponding transcribed region

285   of 24-*PHAS* loci, with the most profound impact in the region immediately distal to the

286   miR2275 CS. *ms23* mutants lack both *PHAS* precursors and 24-nt phasiRNAs while

287   *dcl5-1* mutant anthers contain abundant precursors (Teng *et al.*, 2020), but few phasiRNA

288   products. Therefore, we conclude that the elevated CHH DNA methylation of 24-*PHAS*

289   loci at the 2.0 mm stage requires 24-nt phasiRNAs.

290   **Differentially methylated regions in the CHH context at 24-*PHAS* loci**

291   To explore methylation patterns, we identified the CHH-context differentially methylated

292   regions (DMRs) for the *dcl5-1* or *ms23* mutant and their control plants (Table **S2**; see the

293   Materials and Methods section). For the *dcl5-1* mutant and its control, among 120

294   analyzable 24-*PHAS* loci, we identified *dcl5*-DMRs (DMRs between the *dcl5-1* mutant

295   and its control in the CHH context, abbreviated as *dcl5*-DMRs) from 49 (49/120, 40.8%)

296   24-*PHAS* loci. There were a total of 70 *dcl5*-DMRs at these 49 loci; 19 loci had two or

297   three detected DMRs (Fig. **S5)**. Among 70 *dcl5*-DMRs, 69 of the 70 DMRs display

298   decreased methylation in the *dcl5-1* mutant. In contrast, among 407 analyzable 21-*PHAS*

299   loci, we only identified 34 *dcl5*-DMRs present in 32 (32/407, 6.9%) 21-*PHAS* loci and

300   only two loci had two DMRs; 30 of these DMRs display decreased methylation level in

301   the *dcl5-1* mutant (Fig. **S5**). Therefore, DMRs involving lower CHH methylation levels,

302   comparing the *dcl5-1* mutant to its controls, are more abundant in the 24-*PHAS* loci than

303   in the 21-*PHAS* loci in 2.0 mm anthers.

304   For the *ms23* mutant and its control there are 96 DMRs located in 68 (68/118, 57.6%)

305   24-*PHAS* loci from 118 analyzable loci; again nearly all are hypomethylated (95/96

306   DMRs) (Table **S2**). Thirty DMRs were detected (30/398, 7.5%) among 398 analyzable

307   21-*PHAS* loci. There were 22 24-*PHAS* with two or three detected DMRs, and 46 with one

308   DMR, whereas only one DMR was detected in each 21-*PHAS* locus (Fig. **S5**). In summary,

309   DMRs at 24-*PHAS* loci are more abundant than those found at 21-*PHAS* loci in both the

310   *dcl5-1* and *ms23* mutants.

311   Next, we compared the representation of CHH DMRs in the two mutants. We found 28

312   shared DMRs, 41 *dcl5* specific DMRs, and 67 *ms23* specific DMRs (Fig. **2g**, Table **S2**). If

313   we utilized a lower DMR identification criterion of an absolute methylation change of 2%,

314   then >72% of the DMRs are shared by these mutants. As summarized in the heatmap in

315   Fig. **2g**, almost all analyzable DMRs exhibit decreased CHH DNA methylation in both

316   mutants. Further analysis shows, for the shared 28 DMRs, that both the controls of the

317   *dcl5-1* and *ms23* mutant display relatively consistent CHH DNA methylation with the

318   median value of around 10%. In contrast, DMRs restricted to either *dcl5-1* or *ms23* have

319   relatively lower DNA methylation levels with the median value of around 5% in their

320   corresponding fertile controls (Fig. **S6**). We conclude that the elevated CHH DNA

321   methylation in 2.0 mm control anthers is focused on specific sites in 24-*PHAS* loci and

322   that the extent of CHH methylation depends on the presence of 24-nt phasiRNAs.

323   **CHH DNA methylation at 24-*PHAS* loci normally increases at the meiotic stage**

324   Having established that 24-*PHAS* loci in both *dcl5-1* and *ms23* mutant anthers exhibit low

325   levels of CHH DNA methylation after the CS and across the transcribed gene body at the

326   2.0 mm stage, we sought to distinguish whether this pattern reflected either (i) a failure to

327   maintain higher DNA methylation established at an earlier stage, or (ii) a failure to

328   increase DNA methylation at the 2.0 mm stage. The initial whole genome data presented

329   in Figure 1 established that overall CHH methylation peaked at 2.0 mm. The abundances

330   of 24-nt phasiRNAs from each loci peak at 2.0 mm, with low to moderate level at 1.0 mm,

331   and very low at 0.5 mm (Zhai *et al.*, 2015): transitioning from 0.5 and 1.0 mm there is a

332   slightly increased expression level of 24-nt phasiRNAs, while at the 2.0 mm anther stage

333   there is copious 24-nt phasiRNAs.

334   We generated sequence capture bisulfite-sequencing data for 0.5 and 1.0 mm anthers of

335   the *dcl5-1* mutant and its control (see the Materials and Methods section). First, we

336   compared the methylation patterns for both 21- and 24-*PHAS* in the three developmental

337   stages (0.5, 1.0, and 2.0 mm) from the control plants. No major DNA methylation

338   difference was observed in CG and CHG contexts for either the 21- or 24-*PHAS* across the

339   approximately 10 days of development (Fig. **3a**, **b**, Fig. **S7a, b**). There was, however, an

340   obvious CHH DNA methylation difference in the transcribed regions of 24-*PHAS* loci 3'

341   of the microRNA CS with relatively low levels in pre-meiotic (0.5 and 1.0 mm) and

342   higher levels in meiotic (2.0 mm) anthers (Fig. **3c**); no such difference was observed for

343   21-*PHAS* loci (Fig. **S7c**). These data clarify that the CHH methylation observed in

344   24-*PHAS* loci in 2.0 mm anthers represents an increase from earlier stages paralleling both

345   normal development and production of 24-nt phasiRNAs. In the *dcl5-1* mutant, CHH

346   methylation levels in the CS to the transcription termination site (TTS) region of 24-*PHAS*

347   loci are nearly constant across the three developmental stages and similar to control

348   anthers at the 0.5 and 1.0 mm stages (Fig. **3c, f**). The developmental analysis supports the

349   conclusion that elevated CHH methylation at 24-*PHAS* loci requires the 24-nt phasiRNAs.

350   A parallel analysis of 21-*PHAS* loci identified no obvious DNA methylation differences in

351   any context (Fig. **S7d-f**).

352   We analyzed CHH DNA methylation patterns of the *dcl5*-DMRs (detected between the

353   *dcl5-1* mutant and its control at the 2.0 mm stage) in the 0.5 and 1.0 mm anthers. Almost

354   all *dcl5*-DMRs showed consistent low levels of CHH DNA methylation across these two

355   stages (Fig. **3g**). Collectively, we conclude that the elevated CHH DNA methylation on

356   24-*PHAS* (but not 21-*PHAS*) in the 2.0 mm control anthers is highly related to the burst of

357   24-nt phasiRNAs. Therefore, we conclude that the elevated CHH DNA methylation of

358   24-*PHAS* loci is dependent on the presence of 24-nt phasiRNAs during the period when

359   both transcription of the loci and production of the small RNA products are normally very

360   high in the control anthers.

361   **Abundant phasiRNAs are necessary but not specifically predictive of increased CHH**

362   **DNA methylation at individual 24-*PHAS* loci**

363   There is a low to moderate level of 24-nt phasiRNAs present in 1.0 mm control anthers,

364   with a peak 6 days later at 2.0 mm (Zhai *et al.*, 2015). Despite this, we did not observe an

365   increased CHH DNA methylation level in the 1.0 mm control anthers (Fig. **3c**), suggesting

366   that the methylation change requires abundant 24-nt phasiRNAs. We sought to answer the

367   question of whether increased CHH DNA methylation is correlated to the level of 24-nt

368   small RNAs derived from that location. First, we evaluated the CHH DNA methylation

369   level for five 24-*PHAS* subgroups, ranking loci by the levels of 24-nt phasiRNA

370   abundance in the controls of the *dcl5-1* mutant. As shown in Fig. **4a**, the median value of

371   the average CHH DNA methylation lowest abundance subgroup is around 3%, while the

372   subsequent 4 subgroups steadily increase from 5% to 9% (Fig. **4a**). Moreover, we found

373   that 24-*PHAS* with *dcl5*-DMRs tend to have a higher level of 24-nt phasiRNAs in control

374   plants (Fig. **S8**).

375   Second, we calculated the CHH DNA methylation levels and small RNA abundance for

376   each 24-*PHAS* locus. Obviously, almost all analyzable 24-*PHAS* loci display relatively

377   elevated CHH DNA methylation in the control plants, and much lower in both *dcl5-1* and

378   *ms23* mutants (Fig. **4b**). Importantly, we also observed that 24-*PHAS* loci with the highest

379   CHH methylation are not the ones with the most abundant 24-nt phasiRNAs (Table **S3**).

380   In other words, the level of CHH methylation at individual 24-*PHAS* loci is not highly

381   correlated with the cumulative abundance of the 24-nt phasiRNA products of that locus.

382   When we investigated the distributions of 24-nt phasiRNA, as well as the *dcl5*- and

383   *ms23*-DMRs, we found that neither the 24-nt phasiRNAs nor the DMRs are evenly

384   distributed. Both show a higher abundance 3' of the microRNA CS (Fig. **S9**). Additionally,

385   a higher proportion of *ms23*-DMRs was seen than of *dcl5*-DMRs across analyzable *PHAS*

386   (Fig. **S9b**), which is consistent with the differential abundance of 24-nt siRNAs on

387   24-*PHAS* in the *ms23* (lower) and *dcl5-1* (higher) mutants (Fig. **4b**).

388   Finally, we examined small RNA and the DNA methylation patterns for individual

389   24-*PHAS* loci in the *dcl5-1* and *ms23* mutants and their controls. Fig. **4c** and Fig. **S10**

390   display two examples: 24-*PHAS*_NO291 (Fig. **4c**) and 24-*PHAS*_NO132 (Fig. **S10**). For

391   24-*PHAS*_NO291, based on the sRNA-seq data, there are two abundant 24-nt phasiRNA

392   positions after the miRNA CS in the controls of both the *dcl5-1* and *ms23* mutants; only a

393   few small RNA reads are detected in the *dcl5-1* mutant, and almost no 24-nt phasiRNAs

394   were present in the *ms23* mutant (Fig. **4c**). Based on the sequence capture

395   bisulfite-sequencing data, it is obvious that the controls possess higher CHH DNA

396   methylation levels across the whole 24-*PHAS* transcript compared to the *dcl5-1* and *ms23*

397   mutants. Three DMRs after the CS were identified in the *dcl5-1* mutant and its control,

398   while another three DMRs were detected in the *ms23* mutant and its control. Two

399   *ms23*-DMRs partially overlap or are adjacent to the three DMRs detected in the *dcl5-1*

400   mutant. And for non-shared *ms23*-DMR regions, the methylation difference was also

401 observed in *dcl5-1* and its control plants. Similar patterns were observed in the

402 24-*PHAS*_NO132 (Fig. **S10**).

403   Taken together, the high CHH DNA methylation at 24-*PHAS* requires the presence of

404 24-nt phasiRNAs, but the abundance of 24-nt phasiRNAs is not highly predictive of the

405 CHH methylation level at a *PHAS* locus, although a general low to moderate correlation

406 was observed (Fig. **4a** and Fig. **S11**).

407 **Only one or a few abundant 24-nt phasiRNA(s) accumulate from individual**

408 **24-*PHAS* and likely influence increased CHH DNA methylation**

409 Surprisingly, despite being derived from the same precursor, the abundance of individual

410 24-nt phasiRNAs is highly variable across a locus, similar to the report of

411 non-stoichiometric yield of 21-nt reproductive phasiRNAs from their precursor transcripts

412 (Tamim *et al.*, 2018). The reason for this is unclear: equivalent stoichiometry would be

413 predicted for each 24-nt phasiRNA from the same precursor, but this pattern is not true for

414 any of the expression abundance classes of 24-*PHAS* loci. Therefore, we can investigate

415 within individual loci, whether DMRs are correlated with the abundant 24-nt phasiRNA

416 positions at that locus.

417   To explore as many 24-*PHAS* loci as possible, we re-identified 24-nt phasiRNAs using

418 our newly generated sRNA-seq data of HiII line fertile anthers at the 2.0 mm stage, using

419 the same strategy as previously described (Zhai *et al.*, 2015) (see the Materials and

420 Methods section). As a result, we detected 151 24-*PHAS* loci, 23 of which were newly

421 identified, compared with the list of 24-*PHAS* in the previous study of W23 anthers.

422 Among 176 published 24-*PHAS* in RefB73_v2 in Zhai *et al.* (2015), only 163 of them

423 were evident in RefB73_v4. Among these, only 128 of them were detectable in our

424 sRNA-seq data. Altogether, 186 24-*PHAS* have been detected in different datasets and

425 maize genome assemblies (Zhai *et al.*, 2015 in B73 v4) (Table **S4**).

426   By looking at the 24-nt small RNA distribution across these 24-*PHAS* individually, we

427 found that 24-*PHAS*_NO132 and 24-*PHAS*_NO291 each had only one or a few highly

428 accumulated 24-nt phasiRNA(s) in the fertile control of the *dcl5-1* mutant (Fig. **4c**, Fig.

429   **S10**). These are not exceptional cases as we observed that all *PHAS* loci tend to possess

430   one or a few highly accumulated 24-nt small RNA type(s), violating the expected

431   stoichiometry from a single precursor. Among 151 expressed 24-*PHAS* loci in our data, 96

432   (63.6%) have only one abundant 24-nt phasiRNA with a super-high accumulation level

433   (35.4 to 1.2 million-fold higher than the lowest accumulated 24-nt type). This is visualized

434   for 24-*PHAS*_NO210 and 24-*PHAS*_NO296 in the fertile control of the *dcl5-1* mutant (Fig.

435   **5a, b**). Another 35 (23.2%) 24-*PHAS* loci have two abundant positions, exemplified by

436   24-*PHAS*_NO173 (Fig. **S12a**), and 20 (13.2%) have three to six abundant positions, as

437   found in 24-*PHAS*_NO203 (Fig. **S12b**).

438   To address if the abundant 24-nt phasiRNA position is well correlated with elevated

439   CHH DNA methylation in 2.0 mm control anthers, we analyzed the methylation

440   distribution across 24-*PHAS* loci. As a result, the average value of the CHH DNA

441   methylation levels of the abundant 24-nt phasiRNA position was 4.6%, while other

442   regions with the lowest expressed 24-nt siRNAs were about 2.3%, and this two-fold

443   difference is highly significant (*P* value < 0.001 in the Wilcoxon signed-rank test) (Fig.

444   **5e**). The average abundance of small RNAs from abundant position(s) is around 3,200

445   reads per million, more than three orders of magnitude higher than the value of 2 in other

446   regions (Fig. **5f**). Overall, these results indicate that within an individual locus and then

447   collectively in all highly expressed 24-*PHAS* loci, one or a few 24-nt phasiRNAs with

448   high abundance products are most likely resulting in the elevated DNA methylation levels.

449   **Some 24-nt phasiRNAs accumulate in the *dcl5-1* mutant, but not in *ms23* anthers**

450   The *dcl5-1* mutants retain accumulation of some 24-nt small RNAs generated from

451   24-*PHAS* loci (Fig. **S13**). We were curious to know whether these small RNAs are actual

452   phasiRNAs or another class of small RNA. Interestingly, we observed fifty 24-*PHAS* loci

453   with 24-nt siRNAs generated in an obvious phasing pattern by manual check (phasing

454   score greater than 15, and passing our manual check, see the Materials and Methods

455   section) (Table **S5**). Compared to fertile anthers, however, the accumulation of these

456   phasiRNAs was 73-fold lower in *dcl5-1* mutants at 2.0 mm. Therefore, other DCLs,

457   perhaps DCL3 given its propensity to process heterochromatic siRNAs into 24 nt lengths,

458 substitute by slicing the abundant precursor transcripts, albeit inefficiently, when DCL5 is

459 absent (Teng *et al.*, 2020). In the *ms23* mutant, none of the detected 24-nt small RNAs

460 exhibited a phasing pattern, as expected because there is no detectable expression of

461 24-*PHAS* loci.

462   Two representative examples of the distribution of 24-nt small RNAs in the 2.0 mm

463 *dcl5-1* mutants are displayed in Fig. **5**. Notably, only one 24-nt small RNA with highly

464 abundant accumulation is present in 24-*PHAS*_NO296 and another in 24-*PHAS*_NO210.

465 In both cases, these map 3' of the miR2275 CS (Fig. **5a, b**). Other 24-nt small RNAs

466 (minor peaks) from these two 24-*PHAS* loci have relatively low accumulation levels.

467 Importantly, the 24-nt siRNAs from 24-*PHAS*_NO296 exhibited a clear phased pattern in

468 the *dcl5-1* mutant (phasing score greater than 25), whereas 24-*PHAS*_NO210 24-nt RNAs

469 were not in a phased pattern (Fig. **5c, d**). Consequently, some 24-*PHAS* loci generate

470 24-nt siRNAs independently of the usual biogenesis of phasiRNAs.

Discussion

472

473   This study sought to further explore the hypothesis that 24-nt phasiRNAs can impact

474 DNA methylation at their cognate loci, a type of *cis* regulation (Dukowic-Schulze *et al.*,

475 2016). We first demonstrated that CHH DNA methylation levels are developmentally

476 regulated, being highest in 2.0 mm, mid-meiosis anthers compared to earlier or later stages.

477 In fertile anthers, CHH DNA methylation levels increase significantly at 2.0 mm, the stage

478 of peak 24-nt phasiRNA abundance. To distinguish whether transcriptional activation of

479 24-*PHAS* loci or the 24-nt phasiRNA products contribute to the observed CHH DNA

480 methylation increase, we analyzed two mutants: *ms23*, in which transcriptional activation

481 of the 24-*PHAS* loci fails to occur, and *dcl5-1*, in which transcript abundance is nearly

482 normal (Fig. **S14**), but far fewer 24-nt phasiRNA products exist. Both *ms23* and *dcl5-1*

483 mutants failed to exhibit the elevated CHH methylation at the 24-*PHAS* loci found at 2.0

484 mm in normal anthers, suggesting that 24-nt phasiRNAs are necessary to direct increased

485 CHH DNA methylation at this developmental stage. Further support for a role of 24-nt

486 phasiRNAs in modulating CHH DNA methylation is provided by the observation that

487 elevated CHH DNA methylation of each 24-*PHAS* locus requires the abundant 24-nt small

488 RNA derived from that 24-*PHAS*. Additionally, within a locus, there is enormous

489 variation in the abundance of individual 24-nt phasiRNAs despite derivation from a

490 common precursor; this matches observations from 21-*PHAS* loci (Tamim *et al.*, 2018).

491 The CHH DNA methylation level of the abundant 24-nt phasiRNA position is obviously

492 higher than that at the genomic positions corresponding to low abundance 24-nt siRNAs

493 across the locus (Fig. **5e, f**). Collectively, these three lines of evidence support the idea

494 that 24-nt phasiRNAs can direct increased CHH DNA methylation in *cis*.

495   In whole anthers, the detectable changes in DNA methylation will be from the abundant

496 somatic cells, not the meiocytes. The tapetum synthesizes 24-nt phasiRNAs (Zhai *et al.*,

497 2015; Teng *et al.*, 2020) and both the *ms23* and *dcl5-1* mutants impact tapetal

498 development. Therefore, we propose that the CHH DNA methylation changes we

499 quantified reflect events in the tapetum. As CHH DNA methylation flanking genes

500 exhibits a positive relationship to gene expression (Gent *et al.*, 2013), we speculate that

501 the presumably tapetal CHH DNA methylation might regulate expression of these *PHAS*

502 loci (up or down) or help with chromatin compaction or binucleation of tapetal cells.

503   A reasonable question is whether *cis*-directed enhancement of methylation is the actual

504 or only function of 24-nt reproductive phasiRNAs. It seems somewhat unlikely, as it

505 would represent a closed circuit: highly expressed 24-nt phasiRNAs -> high CHH

506 methylation --> increase of PHAS transcription. At a minimum, our results do

507 demonstrate that 24-nt phasiRNAs are competent to direct CHH DNA methylation. This

508 could result from mis-loading, in an abundance-dependent manner into AGO4, the

509 Argonaute typically loaded with 24-nt heterochromatic siRNAs to direct DNA

510 methylation (Zilberman *et al.*, 2003). Like the observations described here, 21-nt

511 reproductive phasiRNAs in rice and maize have been shown and validated to function in

512 *cis* to direct slicing of their own precursors (Tamim *et al.*, 2018). But it is unclear also in

513 that case if this *cis* activity is the primary function of the phasiRNAs or an unintended

514 consequence of highly abundant small RNAs misloaded into the 'wrong' AGO protein.

515 Until we know the precise functions of 24-nt reproductive phasiRNAs, the AGO proteins

516 into which they are loaded, and their predominant targets if any exist, the importance of

517 the *cis*-directed DNA methylation reported here may remain unclear.

Author contributions

M.Z. and V.W. conceived of the project; M.Z. performed experiments of the project as part of M.Z.'s postdoctoral training at Stanford University; X.M., C.W., and Q.L. analyzed sequencing data and built graphs; M.Z. and V.W. wrote the manuscript with editing by B.C.M., N.M.S., and with input from all co-authors. MZ, XM and VW contributed equally to this work.

References

537

538 **Allen E, Xie Z, Gustafson AM, Carrington JC**. **2005**. MicroRNA-directed phasing during

539 *trans-acting* siRNA biogenesis in plants. *Cell* **121**: 207–221.

540 **Aravin AA, Hannon GJ, Brennecke J**. **2007**. The piwi-piRNA pathway provides an adaptive

541 defense in the transposon arms race. *Science* **318**: 761–764.

542 **Arikit S, Zhai J, Meyers BC**. **2013**. Biogenesis and function of rice small RNAs from non-coding

543 RNA precursors. *Current Opinion in Plant Biology* **16**: 170–179.

544 **Borges F, Martienssen RA**. **2015**. The expanding world of small RNAs in plants. *Nature Reviews*

545 *Molecular Cell Biology* **16**: 727–741.

546 **Chen S, Zhou Y, Chen Y, Gu J. 2018.** fastp: an ultra-fast all-in-one FASTQ preprocessor.

547 *Bioinformatics* **34**: i884–i890.

548 **De Paoli E, Dorantes-Acosta A, Zhai J, Accerbi M, Jeong DH, Park S, Meyers BC, Jorgensen**

549 **RA, Green PJ**. **2009**. Distinct extremely abundant siRNAs associated with cosuppression in petunia.

550 *RNA* **15**: 1965–1970.

551 **Deng P, Muhammad S, Cao M, Wu L**. **2018**. Biogenesis and regulatory hierarchy of phased small

552 interfering RNAs in plants. *Plant Biotechnology Journal* **16**: 965–975.

553 **Dukowic-Schulze S, Sundararajan A, Ramaraj T, Kianian S, Pawlowski WP, Mudge J, Chen C**.

554 **2016**. Novel meiotic miRNAs and indications for a role of phasiRNAs in meiosis. *Frontiers in Plant*

555 *Science* **7**: 762.

556 **Fan Y, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, *et al.* 2016**.

557 *PMS1T*, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in

558 rice. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 15144–

559 15149.

560 **Fei Q, Yang L, Liang W, Zhang D, Meyers BC**. **2016**. Dynamic changes of small RNAs in rice

561 spikelet development reveal specialized reproductive phasiRNA pathways. *Journal of Experimental*

562 *Botany* **67**: 6037–6049.

563 **Fu Q, Wang PJ**. **2014**. Mammalian piRNAs: Biogenesis, function, and mysteries. *Spermatogenesis* **4**:
564 e27889.

565 **Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK**. **2013**. CHH islands: De novo
566 DNA methylation in near-gene chromatin regulation in maize. *Genome Research* **23**: 628–637.

567 **Girard A, Sachidanandam R, Hannon GJ, Carmell MA**. **2006**. A germline-specific class of small
568 RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.

569 **Grivna ST, Beyret E, Wang Z, Lin H**. **2006**. A novel class of small RNAs in mouse spermatogenic
570 cells. *Genes and Development* **20**: 1709–1714.

571 **Han Z, Crisp PA, Stelpflug S, Kaeppler SM, Li Q, Springer NM**. **2018**. Heritable epigenomic
572 changes to the Maize methylome resulting from tissue culture. *Genetics* **209**: 983–995.

573 **Han BW, Wang W, Li C, Weng Z, Zamore PD**. **2015**. PiRNA-guided transposon cleavage initiates
574 Zucchini-dependent, phased piRNA production. *Science* **348**: 817–821.

575 **Huang X, Zhang S, Li K, Thimmapuram J, Xie S. 2018.** ViewBS: A powerful toolkit for
576 visualization of high-throughput bisulfite sequencing data. *Bioinformatics* **34:** 708–709**.**

577 **Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS**
578 *et al.* **2017**. Improved maize reference genome with single-molecule technologies. *Nature* **546**: 524–
579 527.

580 **Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V,**
581 **Vance V, Bowman LH**. **2009**. Clusters and superclusters of phased small RNAs in the developing
582 inflorescence of rice. *Genome Research* **19**: 1429–1440.

583 **Kakrana A, Mathioni SM, Huang K, Hammond R, Vandivier L, Patel P, Arikit S, Shevchenko**
584 **O, Harkess AE, Kingham B** *et al.* **2018**. Plant 24-nt reproductive phasiRNAs from intramolecular
585 duplex mRNAs in diverse monocots. *Genome Research* **28**: 1333–1344.

586 **Kelliher T, Walbot V. 2011.** Emergence and patterning of the five cell types of the *Zea mays* anther
587 locule. *Developmental Biology* **350**: 32–49.

588 **Kim D, Langmead B, Salzberg SL. 2015.** HISAT: a fast spliced aligner with low memory
589 requirements. *Nature Methods* **12**: 357–360.

590 **Langmead B, Trapnell C, Pop M, Salzberg SL**. **2009**. Ultrafast and memory-efficient alignment of
591 short DNA sequences to the human genome. *Genome Biology* **10**:R25.

592 **Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA,**
593 **Tóth KF**. **2013**. Piwi induces piRNA-guided transcriptional silencing and establishment of a
594 repressive chromatin state. *Genes and Development* **27**: 390–399.

595 **Li Q, Eichten SR, Hermanson PJ, Zaunbrecher VM, Song J, Wendt J, Rosenbaum H, Madzima**
596 **TF, Sloan AE, Huang J *et al.* 2014**. Genetic perturbation of the maize methylome. *Plant Cell* **26**:
597 4602–4616.

598 **Liu L, Ren S, Guo J, Wang Q, Zhang X, Liao P, Li S, Sunkar R, Zheng Y**. **2018**. Genome-wide
599 identification and comprehensive analysis of microRNAs and phased small interfering RNAs in
600 watermelon. *BMC Genomics* **19**(Suppl 2):111.

601 **Luteijn MJ, Ketting RF**. **2013**. PIWI-interacting RNAs: From generation to transgenerational
602 epigenetics. *Nature Reviews Genetics* **14**: 523–534.

603 **Mathioni SM, Kakrana A, Meyers BC**. **2017**. Characterization of plant small RNAs by next
604 generation sequencing. *Current Protocols in Plant Biology* **2**: 39–63.

605 **Mohn F, Handler D, Brennecke J**. **2015**. PiRNA-guided slicing specifies transcripts for
606 Zucchini-dependent, phased piRNA biogenesis. *Science* **348**: 812–817.

607 **Nan GL, Zhai J, Arikit S, Morrow D, Fernandes J, Mai L, Nguyen N, Meyers BC, Walbot V**.
608 **2017**. MS23, a master basic helix-loop-helix factor, regulates the specification and development of the
609 tapetum in maize. *Development* **144**: 163–172.

610 **Ono S, Liu H, Tsuda K, Fukai E, Tanaka K, Sasaki T, Nonomura K**. **2018**. EAT1 transcription
611 factor, a noncell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis
612 in rice anther tapetum. *PLoS Genetics* **14**: e1007238.

613 **Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015.** StringTie
614 enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**:
615 290–295.

616 **Rajagopalan R, Vaucheret H, Trejo J, Bartel DP**. **2006**. A diverse and evolutionarily fluid set of
617 microRNAs in *Arabidopsis thaliana*. *Genes & Development* **20**: 3407–3425.

618 **Schultz MD, Schmitz RJ, Ecker JR**. **2012**. 'Leveling' the playing field for analyses of single-base
619 resolution DNA methylomes. *Trends in Genetics* **28**: 583–585.

620 **Shuai P, Su Y, Liang D, Zhang Z, Xia X, Yin W**. **2016**. Identification of phasiRNAs and their
621 drought- responsiveness in *Populus trichocarpa*. *FEBS Letters* **590**: 3616–3627.

622 **Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C *et al.* 2012a**.
623 Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant Journal* **69**: 462–474.

624 **Song X, Wang D, Ma L, Chen Z, Li P, Cui X, Liu C, Cao S, Chu C, Tao Y *et al.* 2012b**. Rice
625 RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. *Plant
626 Journal* **71**: 378–389.

627 **Tamim S, Cai Z, Mathioni SM, Zhai J, Teng C, Zhang Q, Meyers BC**. **2018**. *Cis*-directed
628 cleavage and nonstoichiometric abundances of 21-nucleotide reproductive phased small interfering
629 RNAs in grasses. *New Phytologist* **220**: 865–877.

630 **Teng C, Zhang H, Hammond R, Huang K, Meyers BC, Walbot V**. **2020**. *Dicer-like 5* deficiency
631 confers temperature-sensitive male sterility in maize. *Nature Communications* **11**: 2912.

632 **Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR**. **2015**. MethylC-seq library preparation for
633 base-resolution whole-genome bisulfite sequencing. *Nature Protocols* **10**: 475–483.

634 **Vagin V V, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD**. **2006**. A distinct small RNA pathway
635 silences selfish genetic elements in the germline. *Science* **313**: 320 – 324.

636 **Wu L**. **2013**. DICER-LIKE1 processed *trans-acting* siRNAs mediate DNA methylation: Case study
637 of complex small RNA biogenesis and action pathways in plants. *Plant Signaling & Behavior* **8**:
638 e22476.

639 **Wu L, Mao L, Qi Y**. **2012**. Roles of DICER-LIKE and ARGONAUTE proteins in *TAS*-derived small
640 interfering RNA-triggered DNA methylation. *Plant Physiology* **160**: 990–999.

641 **Xi Y, Li W**. **2009**. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC*
642 *Bioinformatics* **10**: 232.

643 **Xia R, Chen C, Pokhrel S, Ma W, Huang K, Patel P, Wang F, Xu J, Liu Z, Li J *et al.* 2019**. 24-nt
644 reproductive phasiRNAs are broadly present in angiosperms. *Nature Communications* **10**: 627.

645 **Xia R, Xu J, Arikit S, Meyers BC**. **2015**. Extensive families of miRNAs and *PHAS* loci in Norway
646 spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Molecular Biology and*
647 *Evolution* **32**: 2905–2918.

648 **Yu Y, Zhou Y, Zhang Y, Chen Y**. **2018**. Grass phasiRNAs and male fertility. *Science China Life*
649 *Sciences* **61**: 148–154.

650 **Zhai J, Zhang H, Arikit S, Huang K, Nan GL, Walbot V, Meyers BC**. **2015**. Spatiotemporally
651 dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proceedings of*
652 *the National Academy of Sciences of the United States of America* **112**: 3146–3151.

653 **Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, Li H, Yang W, Zhao H, Wang G *et al.* 2014**.
654 Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns
655 of imprinting regulation in maize. *Genome Research* **24**: 167–176.

656 **Zilberman D, Cao X, Jacobsen SE**. **2003**. ARGONAUTE4 control of locus-specific siRNA
657 accumulation and DNA and histone methylation. *Science* **299**: 716–719.

658

Figure legends

Figure 1. 24-*PHAS* loci display stage-specific DNA methylation patterns during maize anther development.

(**a**) to (**c**) Transverse sections of maize anthers at three length stages: 0.5 mm, cell proliferation and differentiation (**a**); 2.0 mm, prophase I of meiosis (**b**); 4.0 mm, binucleate gametophyte (**c**). Abbreviations: AR, archesporial cell(s); Mei, meiocytes; Msp, microspore; SPC, secondary parietal cell(s); T, tapetum; ML, middle layer; EN, endothecium; E, Epidermis. Bars = 30 μm. (**d**) to (**f**) Genome-wide DNA methylation levels in the CG (**a**), CHG (**b**), and CHH (**c**) contexts at three stages. Y axis shows the average CG DNA methylation level (**d**), CHG methylation level (**e**), and CHH methylation level (**f**). (**g**) to (**j**) CHH DNA methylation profiles in the transcribed regions and 2 kb before and 2 kb after the 21-*PHAS* loci (**g**), 24-*PHAS* loci (**h**), protein-coding genes (**i**), and transposons (TEs) (**j**). Y axis is the CHH methylation level (**g-j**). From (**g**) to (**j**), TSS indicates the transcription start site, and TTS indicates the transcription termination site. (**k**) Percentage change in CHH methylation levels of 2.0 and 4.0 mm anthers compared to the 0.5 mm stage at the 21-*PHAS*, 24-*PHAS*, protein-coding genes, and TEs, calculated as level in percent at 2.0 mm - 0.5 mm and 4.0 mm - 0.5 mm. Y axis displays the resulting % change relative to the 0.5 mm stage. From (**d**) to (**k**), fertile anthers were dissected from the W23 inbred, and low-coverage whole genome bisulfite sequencing data were used.

679   **Figure 2. DNA methylation levels at 24-*PHAS* loci in the *dcl5-1* and *ms23* mutants compared to**

680   **their control plants at the 2.0 mm stage of maize anthers.**

681   These deep coverage data were generated using a sequence capture bisulfite-sequencing

682   strategy. (**a**) to (**c**) Plots of the three context methylation profiles at 24-*PHAS* in the *dcl5-1*

683   mutant (blue line) and the control plants (red line). (**d**) to (**f**) Alterations in the methylation

684   profiles at 24-*PHAS* in the *ms23* mutant (purple line) and its control plants (gold line).

685   From (**a**) to (**f**), TSS indicates the transcription start site, CS indicates the microRNA

686   cleavage site in the precursor RNA transcript, and TTS indicates the transcription

687   termination site. MS (*dcl5*) and MS (*ms23*) indicates the male sterile *dcl5-1* or *ms23*

688   mutant, while F (*dcl5*) and F (*ms23*) represents the controls for *dcl5-1* and *ms23* mutants.

689   Under the p-value < 0.001 as threshold, the significant bins were highlighted in grey

690   background in (**c**) and (**f**). Only *PHAS* loci (N = 117 ~ 120, different numbers for each plot)

691   with >5 mapped reads were used for calculation, and the methylation level of each bin was

692   calculated using the weighted DNA methylation computing method (#C/(#C+#T))

693   (Schultz *et al.*, 2012). (**g**) Heatmap plot of CHH DNA methylation levels for the

694   *dcl5*-Differentially Methylated Regions (DMRs identified between the *dcl5-1* mutant and

695   its control plants) and *ms23*-DMRs (DMRs identified between the *ms23* mutant and its

696   control plants). All DMRs identified in 24-*PHAS* loci were included in this analysis. The

697   label changing from white to dark blue color indicates increasing CHH methylation (%).

698   MS refers to male-sterile mutants, and F refers to the control plants.

700 **Figure 3. DNA methylation levels at 24-*PHAS* loci in the *dcl5-1* mutant and its**

701 **control plants in maize pre-meiotic (0.5 and 1.0 mm), and meiotic 2.0 mm anthers.**

702 Deep coverage data were generated using the sequence capture bisulfite-sequencing

703 strategy. (**a**) to (**c**) Plot of the three context methylation profiles at 24-*PHAS* loci in the

704 control plants of the *dcl5-1* mutant in 0.5 (orange line), 1.0 (purple line), and 2.0 mm (red

705 line) stage anthers. (**d**) to (**f**) Plot of the three context methylation profiles at 24-*PHAS* loci

706 in the *dcl5-1* mutant in the 0.5 (light green line), 1.0 (pink line), and 2.0 mm (light blue

707 line) stage anthers. From (**a**) to (**f**), plots were compiled using the same strategy as

708 described for Fig. **2a-f**. TSS, transcription start site; CS, the miR2275 cleavage site in the

709 precursor RNA transcript; and TTS, transcription termination site. Under the p-value <

710 0.001 as threshold, the significant bins were highlighted in grey background (between 1.0

711 and 2.0 mm) and orange background (between 0.5 and 2.0 mm) (**g**) Heatmap plot of CHH

712 DNA methylation levels for the *dcl5*-DMRs identified between the *dcl5-1* mutant and its

713 control plants at the 2.0 mm stage at all three stages. The label changing from white to

714 dark blue color indicates increasing CHH methylation (%). MS refers to male-sterile, and

715 F refers to control plants.

717   **Figure 4. High CHH DNA methylation at 24-*PHAS* loci requires abundant 24-nt**

718   **small RNAs in maize.**

719   (**a**) CHH DNA methylation levels of 5 subgroups of 24-*PHAS* with different levels of

720   small RNA abundance in the control plants of the *dcl5-1* mutant. All detected 151

721   24-*PHAS* loci were divided into 5 subgroups, with their phasiRNA abundance ranging

722   from 10^2 to 10^3, 10^3 to 10^4, 10^4 to 10^5, 10^5 to 10^6, and 10^6 to 10^7 reads. Correspondingly,

723   the number of analyzable 24-*PHAS* loci is 8, 24, 51, 48, and 10, while the numbers of

724   un-analyzable examples are 2, 4, 3, 1, and 0. The CHH DNA methylation level was

725   calculated by weighted DNA methylation for two adjacent 50 bp windows with the

726   highest CHH DNA methylation level at the *PHAS*. MS indicates the *dcl5-1* or *ms23*

727   mutant, while F (*dcl5*) and F (*ms23*) represent the control plants for *dcl5-1* and *ms23*

728   mutants. Boxplot encloses the 25 - 75th percentiles (interquartile range); bold black line is

729   the median; the whiskers refer to values greater than interquartile range excluding outliers;

730   black dots refer to outliers. (**b**) Heatmap of 24-nt small RNA abundance calculated by

731   reads per million (RPM, blue scale) and CHH DNA methylation levels (red) for each

732   analyzable 24-*PHAS* (N = 120). The expression level of 24-nt phasiRNAs in the *ms23*

733   mutant and its control was calculated based on the published sRNA-seq data from Zhai et

734   al., 2015. The control for *dcl5-1* shows the loci ranked by DNA methylation level, and the

735   other seven rows correspond to this order. (**c**) A view of 24-nt small RNA and CHH DNA

736   methylation patterns in control plants and two mutants across 24-*PHAS*_NO291 locus and

737   flanking 200 bp regions. The top four graphs display the 24-nt siRNA abundance with

738   discontinuous line by reads per million (RPM), while the bottom four graphs show the

739   CHH DNA methylation percent in analyzable windows at *PHAS* loci. Each gray box

740   indicates a *dcl5*- or *ms23*-DMR. Dotted lines indicate the microRNA cleavage site.

741   Vertical gray bars represent 24 nt spacing.

742

744 **Figure 5. PhasiRNA abundances, CHH DNA methylation, and phased patterns**

745 **across the 24-*PHAS*_NO210 and 24-*PHAS*_NO296 loci in four genotypes at the 2.0**

746 **mm stage of maize anthers.**

747 (**a-b**) phasiRNA abundance across the loci 24-*PHAS*_NO210 (**a**) and 24-*PHAS*_NO296 (**b**)

748 in the *dcl5-1* and *ms23* mutant, and their controls. (**c-d**) Phased patterns at

749 24-*PHAS*_NO210 (**c**) and 24-*PHAS*_NO296 (**d**) in the *dcl5-1* and *ms23* mutant, and their

750 controls. Dotted lines indicate the microRNA cleavage site. Vertical gray bars represent

751 24 nt spacing. MS indicates the *dcl5-1* or *ms23* mutant, while F (*dcl5*) and F (*ms23*)

752 represent the control plants for *dcl5-1* and *ms23* mutants. (**e**) Comparison of CHH DNA

753 methylation of the abundant 24-nt phasiRNA positions (this corresponds to the peak

754 region with super highly expressed 24-nt phasiRNA(s)) and other regions (with no or

755 lowly expressed 24-nt phasiRNAs) at the 24-*PHAS*. One hundred and forty 24-*PHAS* loci

756 were used here. (**f**) Comparison of 24-nt small abundance from the abundant 24-nt

757 phasiRNA positions and other regions on 24-*PHAS* loci. Boxplot encloses the 25 - 75th

758 percentiles (interquartile range); bold black line is the median; the whiskers refer to values

759 greater than interquartile range excluding outliers; black dots refer to outliers. One

760 hundred and fifty-one 24-*PHAS* loci were used here.

761

Supporting Information

Fig S1. CG and CHG DNA methylation profiles in the transcribed regions and 2 kb before and 2 kb after the 21-*PHAS*, 24-*PHAS*, protein-coding genes, and transposons (TEs).

Fig S2. DNA methylation levels across the 21-*PHAS* transcribed regions and flanking 200 bp regions in the CG, CHG, and CHG contexts at the 2.0 mm stage of the *dcl5-1* mutants (MS) and its controls (F) in a family segregating 3:1 for sterile individuals.

Fig S3. DNA methylation levels at 24-*PHAS* and 21-*PHAS* transcribed regions and flanking 200 bp regions in 2.0 mm anthers.

Fig S4. DNA methylation levels within the transcribed regions and flanking 200 bp regions of 381 to 400 21-*PHAS* loci (analyzed based on presence in a dataset) in the CG, CHG, and CHG contexts in 2.0 mm anthers.

Fig S5. Differentially methylated region (DMR) counts at 24-*PHAS* and 21-*PHAS* loci with *dcl5-* or *ms23*-DMRs at the 2.0 mm stage.

Fig S6. CHH DNA methylation distribution of DMRs specific to *dcl5-1* mutant, shared by *dcl5-1* and *ms23* samples, and *ms23*-specific DMRs.

Fig S7. DNA methylation levels at 21-*PHAS* loci in the *dcl5-1* mutant and its control in pre-meiotic (0.5 and 1.0 mm), and meiotic 2.0 mm anthers.

Fig S8. Comparison of 24-nt phasiRNA abundance for 49 24-*PHAS* with *dcl5*-DMR(s) and 71 24-*PHAS* without detectable DMR(s).

Fig S9. Distribution of 24-nt small RNAs and 24-*PHAS* with *dcl5-* or *ms23*-DMRs across 24-*PHAS* loci plus the 200 bp flanking regions in the *dcl5-1* and *ms23* mutants and their fertile controls.

Fig S10. A view of 24-nt small RNA and CHH DNA methylation patterns in two mutants, *dcl5-1* and *ms23*, and their fertile controls across locus 24-P*HAS*_NO132 and its 200 bp flanking regions.

Fig S11. Correlation between the abundance of 24-nt phasiRNAs and CHH methylation level at 24-*PHAS*.

Fig S12. 24-nt small RNA distribution plotted across the transcribed regions of the 24-*PHAS*_NO173 and 24-*PHAS*_NO203 loci in the control (F) of the *dcl5-1* mutant at the 2.0 mm stage.

Fig S13. Abundance of 24-nt phasiRNAs from 151 24-*PHAS* loci in the 2.0 mm anthers.

Fig S14. The expression levels of all 24-*PHAS* loci in the two replicates of the *dcl5-1* mutants and its control plants.

Table S1. Summary of whole genome bisulfite sequencing, sequence capture bisulfite-sequencing, sRNA-seq, and RNA-seq libraries newly generated in this study.

Table S2. CHH DNA methylation levels of each *dcl5-* and *ms23*-DMR in the 2.0 mm *dcl5-1* and *ms23* mutants and their fertile controls.

Table S3. 24-nt sRNA abundance and CHH methylation level of analyzable 24-*PHAS* loci.

Table S4. Updated list of 186 24-*PHAS* in maize (RefB73_v4) with the 24-nt small RNA abundance calculated by reads per million (RPM) on each locus in the four genotypes.

Table S5. The largest phasing score of certain 24-nt phasiRNA on each 24-*PHAS* in the *dcl5-1* and *ms23* mutants, as well as their controls.

(a) 0.5 mm
(b) 2.0 mm
(c) 4.0 mm

AR/Mei/Msp
SPC
T
ML
EN
E

(d) CG context (%) — 0.5 mm, 2.0 mm, 4.0 mm
(e) CHG context (%) — 0.5 mm, 2.0 mm, 4.0 mm
(f) CHH context (%) — 0.5 mm, 2.0 mm, 4.0 mm

(g) 21-*PHAS*
(h) 24-*PHAS*
(i) Genes
(j) TE

CHH context (%)

-2 kb TSS TTS 2 kb

—— 0.5 mm   —— 2.0 mm   —— 4.0 mm

(k)
% changes relative to 0.5

■ 2.0 mm ■ 4.0 mm

upstream 21-*PHAS* downstream | upstream 24-*PHAS* downstream | promoter exon intron downstream | upstream TE downstream

21-*PHAS* | 24-*PHAS* | Genes | TE
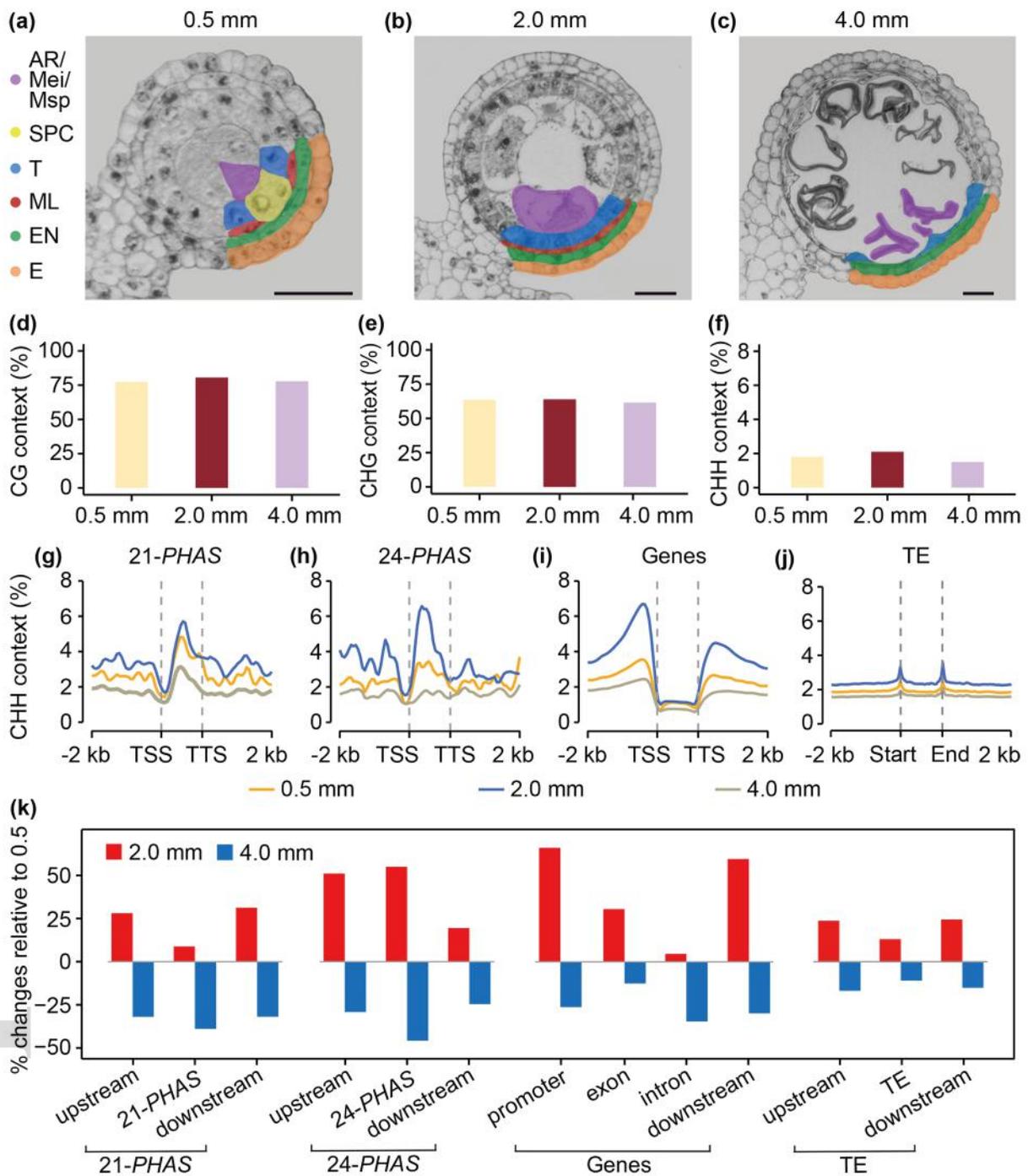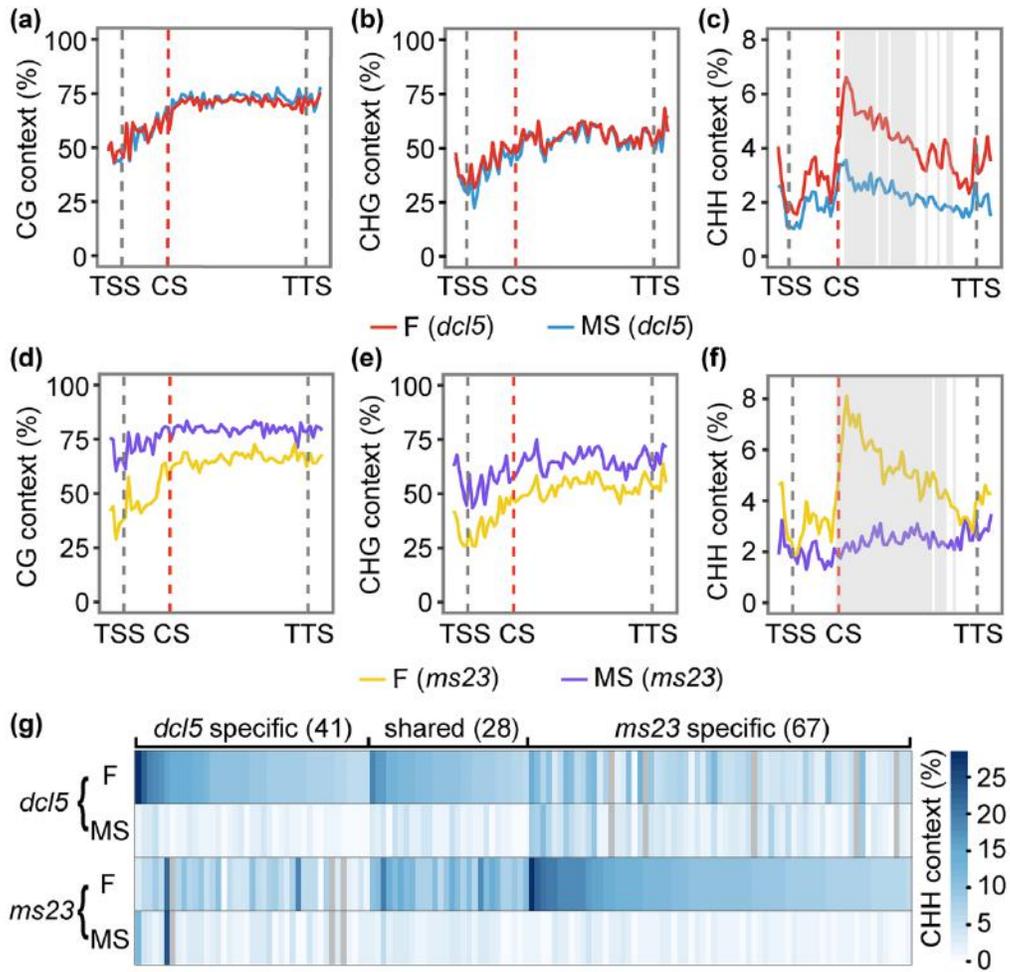
nph_17060_f1.tiff

nph_17060_f2.tiff

nph_17060_f3.tiff

nph_17060_f4.tiff

(a) 24-*PHAS*_NO210

(b) 24-*PHAS*_NO296

nph_17060_f5.tiff