The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*.

Short title: The PMD MAIN/MAIL1 and PP7L complex regulates gene expression and TE silencing.

Melody Nicolau[1,2], Nathalie Picault[1,2], Julie Descombin[1,2], Yasaman Jami-Alahmadi[3], Suhua Feng[4], Etienne Bucher[5], Steven E. Jacobsen[4,6], Jean-Marc Deragon[1,2], James Wohlschlegel[3] and Guillaume Moissiard[1,2]*.

1 LGDP-UMR5096, CNRS, Perpignan, France.

2 LGDP-UMR5096, Université de Perpignan Via Domitia, France.

3 Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, CA, USA.

4 Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA, USA.

5 Plant Breeding and Genetic Resources, Agroscope, Nyon, Switzerland.

6 Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA, USA.

* Corresponding author

E-mail: guillaume.moissiard@univ-perp.fr

ABSTRACT

Transposable elements (TEs) are DNA repeats that must remain silenced to ensure cell integrity. Several epigenetic pathways including DNA methylation and histone modifications are involved in the silencing of TEs, and in the regulation of gene expression. In *Arabidopsis thaliana*, the TE-derived plant mobile domain (PMD) proteins have been involved in TE silencing, genome stability, and control of developmental processes. Using a forward genetic screen, we found that the PMD protein MAINTENANCE OF MERISTEMS (MAIN) acts synergistically and redundantly with DNA methylation to silence TEs. We found that MAIN and its close homolog MAIN-LIKE 1 (MAIL1) interact together, as well as with the phosphoprotein phosphatase (PPP) PP7-like (PP7L). Remarkably, *main*, *mail1, pp7l* single and *mail1 pp7l* double mutants display similar developmental phenotypes, and share common subsets of upregulated TEs and misregulated genes. Finally, phylogenetic analyses of PMD and PP7-type PPP domains among the Eudicot lineage suggest neo-association processes between the two protein domains to potentially generate new protein function. We propose that, through this interaction, the PMD and PPP domains may constitute a functional protein module required for the proper expression of a common set of genes, and for silencing of TEs.

AUTHOR SUMMARY

The plant mobile domain (PMD) is a protein domain of unknown function that is widely spread in the angiosperm plants. Although most PMDs are associated with repeated DNA sequences called transposable elements (TEs), plants have domesticated the PMD to produce genic versions that play important roles within the cell. In *Arabidopsis thaliana*, MAINTENANCE OF MERISTEMS (MAIN) and MAIN-LIKE 1 (MAIL1) are genic PMDs that are involved in genome stability, developmental processes, and silencing of TEs. The mechanisms involving MAIN and MAIL1 in these cellular processes remain elusive. Here, we show that MAIN, MAIL1 and the phosphoprotein phosphatase (PPP) named PP7-like (PP7L) interact to form a protein complex that is required for the proper expression of genes, and the

47 silencing of TEs. Phylogenetic analyses revealed that PMD and PP7-type PPP domains are evolutionary
48 connected, and several plant species express proteins carrying both PMD and PPP domains. We
49 propose that interaction of PMD and PPP domains would create a functional protein module involved
50 in mechanisms regulating gene expression and repressing TEs.
51

52 ## INTRODUCTION

53 In eukaryotes, DNA methylation and post-translational modifications of histones are
54 epigenetic marks involved in chromatin organization, regulation of gene expression and silencing of
55 DNA repeats such as transposable elements (TEs) [1-3]. Constitutive heterochromatin is highly
56 condensed and enriched in silenced TEs that are targeted by DNA methylation and histone H3 lysine
57 9 dimethylation (H3K9me2). Euchromatin is more relaxed and composed of genes that are more
58 permissive to transcription, depending on the recruitment of transcription factors (TFs), cofactors and
59 RNA polymerases [1, 4]. In plants, DNA methylation occurs in three different cytosine contexts: CG,
60 CHG and CHH (where H = A, T or C), involving specialized DNA methyltransferases [5]. In *Arabidopsis*
61 *thaliana,* DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and DRM1 mediate de novo DNA
62 methylation in all sequence contexts through the RNA-directed DNA methylation (RdDM) pathway,
63 which involves among other components, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and DICER-
64 LIKE 3 (DCL3) for the production of short interfering (si)RNAs [6, 7]. The maintenance of CG
65 methylation is specifically performed by METHYLTRANSFERASE 1 (MET1), while CHROMOMETHYLASE
66 2 (CMT2) and CMT3 are involved in the maintenance at CHG sites [8, 9]. CMT2 can also be involved in
67 the deposition of CHH methylation at specific genomic location [10, 11]. Finally, DRM2 is mostly
68 required for the maintenance of CHH methylation through the RdDM pathway [6, 7, 9]. Together with
69 DNA methylation, additional pathways play important roles in TE silencing. The MICRORCHIDIA 1
70 (MORC1) and MORC6 ATPases interact together, and are required for heterochromatin condensation
71 and repression of TEs, acting mostly downstream of DNA methylation and RdDM pathway [12-14].
72 More recently, the *A. thaliana* plant mobile domain (PMD) proteins MAINTENANCE OF

MERISTEM (MAIN) and MAIN-LIKE 1 (MAIL1) were identified as new factors required for TE silencing [15]. In addition, these two proteins have been involved in genome stability, and regulation of developmental processes such as cell division and differentiation [16, 17]. The PMD is a large protein domain of unknown function that is widely represented among the angiosperms, predominantly associated with TEs [15, 18]. It has been proposed that genic PMD versions, such as the MAIN and MAIL1 proteins derived from TEs after gene domestication [15, 18, 19]. Previous studies suggested that genic PMDs could act as cellular factors related to transcription, possibly acting as transcription factor (TF)-like, co-factor or repressor proteins regulating this cellular process [16, 18]. Nevertheless, the role of PMD proteins in the regulation of transcription remains elusive. Most of genic PMD proteins are standalone versions, however, in some cases, the PMD is fused to another protein domain, such as protease, kinase or metallo-phosphatase (MPP) domains. For instance in *A. thaliana*, the MAIL3 protein carries a PMD, which is fused to a putative serine/threonine-specific phosphoprotein phosphatase (PPP) domain phylogenetically related to the plant-specific protein phosphatase 7 (PP7) [20]. PP7 is a calmodulin-binding PPP that has been related to cryptochrome (CRY)-mediated blue-light signaling, and to the control of stomatal aperture [20-22]. PP7 is also involved in the perception of red/far red light by controlling the phytochrome pathway [23, 24]. In addition to PP7 and MAIL3 (also known as "long PP7"), the protein PP7-like (PP7L) belongs to the same phylogenetic clade [20]. PP7L was recently identified as a nuclear protein involved in chloroplast development and abiotic stress tolerance [25]. The *pp7l* mutant plants showed photosynthetic defects and strong developmental phenotype associated with misregulation of several genes [25].

In this study, we described a forward genetic screen based on a GFP reporter gene that allowed us to identify a mutant population in which *MAIN* is mutated, leading to GFP overexpression. We then deciphered the genetic interaction between the DRM2, CMT3 and MAIN, showing that these proteins are part of different epigenetic pathways that act redundantly or synergistically to repress TEs. Biochemical analyses indicated that MAIN and MAIL1 physically interact together. These analyses also identified PP7L as a robust interactor of the two PMD proteins. In addition, the characterization

of developmental and molecular phenotypes of *pmd* and *pp7l* single and double mutant plants strongly suggest that these proteins interact together to silence TEs, and regulate the expression of a common set of genes. Finally, phylogenetic analyses allowed us to determine the distribution of PMD and PP7/PP7L domains among the Eudicots. Based on these analyses, we have evidences of co-evolution linked to the neo-association of the PMD and PP7-type PPP domains on single proteins in several Eudicot species, suggesting a convergent evolution between these two protein domains.

RESULTS

Mutation in *MAIN* is responsible for TE silencing defects.

The *ATCOPIA28* retrotransposon *AT3TE51900* (hereafter called *ATCOPIA28*) is targeted by several epigenetic pathways such as DNA methylation and the MORC1/6 complex, which altogether contribute to its repression. We engineered a construct in which the 5' long terminal repeat (LTR) promoter region of *ATCOPIA28* controls GFP transcription (Fig 1A). While the *ATCOPIA28::GFP* transgene is fully silenced in wild type (WT) plants, it is weakly expressed in the DNA methylation-deficient *drm1 drm2 cmt3 (ddc)* triple mutant background (Fig 1B) [26]. We performed an ethyl methane sulfonate (EMS) mutagenesis using the *ATCOPIA28::GFP ddc* plants as sensitized genetic material, and screened for mutant populations showing GFP overexpression. Among, the selected populations, we retrieved two new mutant alleles of *MORC6* carrying missense mutations in either the GHKL or S5 domains of the protein (S1A-C Fig). We also identified the population *ddc #16* showing strong overexpression of GFP and misregulation of several endogenous TEs, including *ATCOPIA28* (Fig 1B-D). Mapping experiments based on whole genome resequencing and bulk segregant analysis indicated that *ddc #16* carries a missense point mutation (C230Y) in the gene *AT1G17930*, previously named *MAIN* (S1D and S1E Fig). Genetic complementation analyses by crossing the *ddc #16* EMS mutant with the knock-out (KO) transferred DNA (T-DNA) insertion line *main*-2 generated F1 *ddc #16* x *main-2* plants that did not express the GFP (S1F Fig). Transcriptional profiling analyses showed,

however, that endogenous TEs, including *ATCOPIA28*, were upregulated in F1 *ddc #16* x *main-2* plants, but not in F1 control plants generated from the backcross of *ddc #16* with WT Columbia (Col) plants (S1G Fig). Self-fertilization of F1 *ddc #16* x *main-2* plants allowed us to retrieve several F2 *ddc #16* x *main-2* plants overexpressing the GFP (S1F Fig). Among these GFP positive F2 plants, we identified individuals that were either homozygote for the EMS mutation in the *MAIN* gene, or plants carrying both the EMS and T-DNA *main*-2 mutant alleles (S1F Fig). Moreover, while all these plants were homozygote for the *drm2* mutation, half of them segregated the *cmt3* mutation. Thus, altogether, these analyses suggested that *ATCOPIA28::GFP* silencing is more DRM2- than CMT3-dependent. More importantly, they confirmed that *MAIN* was the mutated gene causing the upregulation of *ATCOPIA28::GFP* and several endogenous TEs. Therefore, *ddc #16* was renamed *ddc main-3*.

MAIN, DRM2 and CMT3 act redundantly or synergistically to repress TEs and DNA-methylated genes.

To determine the genetic interaction of *ddc* and *main-3* mutations on TE silencing, we carried out two independent RNA sequencing (RNA-seq) analyses in the hypomorphic *main-3* single, *ddc* triple and *ddc main-3* quadruple mutant plants (Fig 2A). We observed some variations among the misregulated loci between the two RNA-seq analyses (S2A-B Fig and S1 Table). Therefore, we decided to focus our analyses on loci that were reproducibly misregulated in each mutant background (S2C Fig and S2 Table). As previously described, the *ddc* mutant showed upregulation of several TEs spread over the five chromosomes (Fig 2B-D and S2D Fig, and S2 Table) [11]. Loss of TE silencing was also observed to a milder degree in the *main-3* mutant, with the significant enrichment of pericentromeric TEs among the upregulates TEs (Fig 2B-D and S2D Fig, and S2 Table). The *ddc main-3* mutant showed an exacerbation of TE silencing defects, with a large number of pericentromeric TEs being specifically upregulated in this mutant background (Fig 2B-D and S2 Table). Comparative analyses revealed that upregulated TEs cluster into four distinct classes (Fig 2E and S2E Fig). Class I TEs are upregulated in both *ddc* and *main-3* mutants, and combining these mutations in *ddc main*-3 lead to synergistic effects

(Fig 2E and S2E-F Fig). Class II and class III TEs are targeted by the MAIN and DRM2/CMT3 pathways, respectively (Fig 2E and S2E-F Fig). However, the upregulation of class II and class III TEs is further enhanced in *ddc main-3*, which suggests that the MAIN and DRM2/CMT3 pathways can partially compensate each other at these genomic locations (S2F Fig). Finally, the most abundant class IV TEs are only misregulated in *ddc main-3*, which implies that the MAIN and DRM2/CMT3 pathways act redundantly to silence these TEs (Fig 2E and S2E-F Fig).

Several genes were reproducibly misregulated in the three mutant backgrounds (S2 Table). Among these genes, a subset was commonly upregulated in *ddc*, *main-3* and *ddc main-3* (S2G Fig). Remarkably, genes that were upregulated in the three mutants were significantly enriched in pericentromeric regions of chromosomes, where constitutive heterochromatin resides (S2G Fig). This is consistent with the fact that, among the upregulated genes, we identified a large proportion of DNA-methylated genes (S2G Fig). Conversely, we could not identify a subset of genes commonly downregulated in *ddc*, *main-3* and *ddc main-3* (S2G Fig). Furthermore, downregulated genes were not enriched in pericentromeric regions, and most of them were not targeted by DNA methylation (S2G Fig).

To further dissect the genetic interaction between the DRM2, CMT3 and MAIN pathways, we generated the *drm1 drm2 main-3 (dd main-3)* and *cmt3 main-3* mutants (S2H Fig). We then analyzed the expression level of several TEs previously identified as misregulated in *ddc*, *main-3* and/or *ddc main-3*. The endogenous *ATCOPIA28* was the most expressed in *ddc main-3* and *dd main-3*, and to a lesser extent, in *cmt3 main-3* (Fig 2F). This is consistent with the fact that all the F2 *ddc #16* x *main-2* plants overexpressing *ATCOPIA28::GFP* were *drm2* homozygote, although they segregated the *cmt3* mutation (S1F Fig). Further analyses showed that most of the tested TEs tend to be more expressed in *cmt3 main-3* than in *dd main-3*, with the exception of *ATIS112A* that was more upregulated in *dd main-3* than in *cmt3 main-3* (Fig 2G). In conclusion, these analyses showed complex genetic interactions between the DRM2, CMT3 and MAIN pathways. Depending on the locus, the three pathways act redundantly or synergistically to repress TEs and DNA-methylated genes.

MAIN and MAIL1 are required for the proper expression of a common set of genes and TEs.

To study the role of MAIN and MAIL1 in the regulation of gene expression and TE silencing, we performed two independent RNA-seq experiments in the *main-2* and *mail1-1* null mutants (Exp1 and Exp3, S3 Table), and reanalyzed previously published RNA-seq datasets (Exp2, S3 Table) [15]. Although all the RNA-seq experiments were performed using plants at the same developmental stage (3-week-old seedlings), we observed important variations among the number of genes and TEs that were misregulated in the two *pmd* null mutants in each RNA-seq experiment (S3A Fig and S3 Table). Variations among WT replicates from the three independent RNA-seq experiments were not as pronounced, and principal component analysis (PCA) showed that all the WT samples tend to cluster together (S3B Fig). Conversely, the *pmd* mutant samples did not cluster together based on their genetic background, but rather based on the RNA-seq experiment (S3B Fig). This suggests that *main-2* and *mail1-1* perceive their environment in a similar way, leading to the misregulation of similar sets of loci in each independent RNA-seq experiment. Furthermore, this is consistent with the hypothesis that MAIN and MAIL1 act in the same pathway [15]. Comparative analyses between the three independent RNA-seq experiments allowed us to identify genes and TEs that were reproducibly misregulated in *main-2* and *mail1-1* (S3C Fig). For stringency purpose, we decided to focus on these lists of reproducibly misregulated loci to perform further analyses (Fig 3A-B and S4 Table).

We then compared the transcriptomes of *main-2* and *mail1-1* null mutants, together with the hypomorphic *main-3* mutant allele (S2 and S4 Tables). We identified sets of genes and TEs that were specifically misregulated in each mutant background (Fig 3C and D). These analyses also revealed some loci that were commonly misregulated in the three mutants (Fig 3C-D, S3D-F Fig and S5 Table). The biggest overlaps were among the downregulated genes, as most of the downregulated genes in *main-2* were also downregulated in *mail1-1* and, to a smaller proportion, in *main-3* (Fig 3D). As observed in *main-3* (S2G Fig), the genes that were downregulated in *main-2* and *mail1-1* were mostly located in the chromosome arms, and they were not DNA-methylated genes (Fig 3E). However, unlike

in *main-3*, the upregulated genes in *main-2* and *mail1-1* were not enriched in pericentromeric regions, and only one quarter of them were DNA-methylated genes (Fig 3E). This discrepancy suggests that MAIN and MAIL1 are not only required for the silencing of pericentromeric and DNA-methylated loci, but also for the repression of genes that are not targeted by DNA methylation. In conclusion, these comparative analyses allowed us to precisely define the sets of genes and TEs that are commonly misregulated in *main-2, mail1-1* and *main-3* mutants. Moreover, these results revealed important overlaps between the misregulated loci in *main-2* and *mail1-1,* which strongly suggests that the two proteins act in the same pathway to regulate the expression of common sets of loci.

Slight increase in non-CG methylation in the *main-2* mutant does not correlate with changes in gene expression and TE silencing defect.

Whole genome bisulfite sequencing (BS-seq) analyses showed that, at the chromosome scale, DNA methylation level is mostly unchanged in *main-2* in comparison to WT, with the exception of a slight increase in CHG methylation in pericentromeric regions (Fig 4A). Subtle but statistically significant CHG hypermethylation was further confirmed in pericentromeric TEs and genes, which are mostly TE genes (Fig 4B and C). Slight CHG and CHH hypermethylation was also detected in TEs located in chromosome arms (Fig 4D). Conversely, genes located in chromosome arms did not show significant changes in DNA methylation level in *main-2* (Fig 4E). Identical results were obtained by analyzing the DNA methylation level at TEs and genes that were specifically misregulated in *main-2* (Fig 4F-H). We then analyzed the DNA methylation level at genomic locations previously defined as differentially hypomethylated regions (hypo DMRs) at CHG and CHH sites in *cmt3* and *drm1drm2 (dd)* mutants, respectively [26]. The *cmt3* and *dd* hypo DMRs are mostly located in TEs. As observed with pericentromeric genes and all TEs (Fig 4B-D), we found slight increases in CHG and CHH methylation at *cmt3* and *dd* hypo DMRs, respectively, in *main-2* (S4A and S4B Fig). Finally, DMR calling in *main-2* using stringent parameters only identified a few DMRs (S4C Fig). Thus, DNA methylation is mostly unaffected in *main*-2, with the exception of a slight increase in non-CG methylation at pericentromeric

genes and all TEs. Moreover, this subtle non-CG hypermethylation does not correlated with changes in gene and TE expression observed in *main-2*.

MAIN, MAIL1 and the metallo-phosphatase PP7L physically interact together.

The *main-2* and *mail1-1* null mutants display similar molecular and developmental phenotypes (Fig 3 and Fig 5A). Thus, we hypothesized that the two PMD proteins may act in the same pathway, possibly by interacting together. To test this hypothesis, we generated transgenic lines expressing FLAG- and MYC-tagged genomic PMD versions driven by their endogenous promoters. We confirmed that epitope-tagged MAIN and MAIL1 proteins were produced at the expected sizes, and they could complement the respective developmental phenotypes of null mutant plants (Fig 5A and B). Importantly, they could also efficiently rescue the TE silencing and gene expression defects observed in the two *pmd* mutants, implying that epitope-tagged MAIN and MAIL1 are functional proteins (Fig 5C-E). Using FLAG-tagged MAIN and MAIL1 expressing plants, immunoprecipitation followed by mass spectrometry (IP-MS) analyses were carried out to determine potential protein interactors. Mass spectrometry (MS) analyses indicated that MAIL1 was strongly immunoprecipitated with MAIN-FLAG and *vice versa* (Fig 5F). To validate IP-MS results, we crossed the MAIN-FLAG and MAIL1-MYC lines together. We then performed co-immunoprecipitation (co-IP) experiments using F1 hybrid plants co-expressing the two transgenes, and confirmed that MAIN and MAIL1 interact together (Fig 5G). MS analyses of MAIN-FLAG and MAIL1-FLAG IP also identified the metallo-phosphatase PP7L as putative interactor (Fig 5F). MAIN, MAIL1 and PP7L were the only three proteins reproducibly enriched across multiple replicates (Fig 5F). Co-IP experiments using plants co-expressing either PP7L-FLAG together with MAIN-MYC or MAIL1-MYC constructs confirmed the interaction between PP7L and each PMD protein (Fig 5H and I). Thus, the three proteins MAIN, MAIL1 and PP7L physically interact together.

The *main, mail1* and *pp7l* mutants display similar developmental and molecular phenotypes.

PP7L is a putative metallo-phosphatase that was recently identified as a nuclear protein required for photosynthesis [20, 25]. The *pp7l-2* null mutant displays abnormal developmental phenotype reminiscent of the two *pmd* null mutant plants, and 3-week-old *mail1-1 pp7l-2* double mutant plants do not show exacerbation of this phenotype (Fig 6A). To determine the genetic interaction between PMD and PP7L, we compared the transcriptomes of *main-2*, *mail1-1, pp7l-2* single and *mail1-1 pp7l-2* double mutants (RNA-seq Exp3, S3 and S6 Tables). As previously described for *main-2* and *mail1-1* (S3B Fig), PCA analyses showed that all the mutant samples tend to cluster together, and away from WT controls (S5A Fig). We identified a large number of misregulated loci in each mutant background (S5B Fig). Comparative analyses revealed that significant proportions of these loci were commonly misregulated in all the mutant backgrounds, which is consistent with the fact that MAIN, MAIL1 and PP7L interact together to possibly regulate gene expression (Fig 6B and C). These analyses also identified loci that were specifically misregulated in *main-2*, *mail1-1* or *pp7l-2*, which suggests that each protein is independently required for the proper expression of subsets of loci (Fig 6C). Besides, these analyses revealed loci that were exclusively misregulated in the *mail1-1 pp7l-2* double mutant, which implies that PP7L and MAIL1 may act redundantly to ensure the proper expression of these loci (Fig 6B-C). Further analyses showed that, among the loci that were misregulated in *mail1-1 pp7l-2*, upregulated genes were significantly more expressed in the double mutant than in each single mutant, and upregulated TEs were significantly differentially expressed only between *mail1-1 pp7l-2* and *pp7l-2* mutants (Fig 6D-E). Conversely, there was no significant difference of expression between the double mutant and single mutants for the downregulated genes (Fig 6F). Thus, these analyses suggest that combining the *pp7l-2* and *mail1-1* mutations may lead to synergistic defects mostly at genes that are upregulated in the double mutant. Considering the variations observed among the misregulated loci in *main-2* and *mail1-1* between independent RNA-seq experiments, we then decided to focus our analyses on loci previously defined as reproducibly misregulated in the two *pmd* mutants (Fig 3, S3 Fig and S4 Table). Among these loci, we identified several genes and TEs that were also misregulated in *pp7l-2* and *mail1-1 pp7l-2*, with the most

significant overlap among downregulated genes (Fig 6G-J and S7 Table). These commonly misregulated loci did not show any significant difference of expression between the double and single mutants (Fig6K-M). This suggests that MAIN, MAIL1 and PP7L are equally required for the proper expression of these subsets of loci (S7 Table). Finally, we performed in silico analyses to identify enriched DNA motif within a 1kb promoter region upstream of start codon of genes that were up- or downregulated in the different mutant backgrounds. We could not detect any enrichment of a DNA motif among any lists of upregulated genes. Likewise, we could not identify a DNA motif significantly enriched in the lists of downregulated genes from the RNA-seq Exp3. However, we identified a discrete DNA motif (hereafter called 'DOWN' motif) that was enriched in the promoters of genes that were reproducibly downregulated in *main-3*, *ddc main-3*, *main-2* and *mail1-1* mutants (S5C Fig). The 'DOWN' motif was further enriched and refined in the promoters of genes commonly downregulated in all the mutant backgrounds analyzed in this study: eighteen genes out of nineteen, 95% of enrichment (S5C Fig, S8 and S9 Tables). We analyzed the DNA methylation level of the 'DOWN' motif in the promoters of these eighteen genes in WT and *main-2*, and found that this DNA motif was not targeted by DNA methylation. Besides, further analyses showed that only a small fraction of all *Arabidopsis* genes carried the 'DOWN' motif in their promoter (11%, S5C Fig), which is in the range of genomic distribution of several DNA motifs recognized by bZIP TFs (S10 Table). Finally, random test analyses based on nineteen randomly picked genes strongly suggested that the enrichment of the 'DOWN' motif in the promoter of downregulated genes was substantial (S9 Table). Thus, altogether, these analyses showed that MAIN, MAIL1 and PP7L are equally required for the proper expression of a common set of genes that are downregulated in each single mutant as well as in *mail1-1 pp7l-2* double mutant, and these genes carry the 'DOWN' DNA motif in their promoter.

PP7L is not required for heterochromatin condensation.

WT *Arabidopsis* nuclei at interphase exhibit condensed DNA foci called chromocenters that are composed of constitutive heterochromatin, and are enriched in H3K9me2 [27]. In several

306 epigenetic mutants, decondensation of constitutive heterochromatin correlates with disruption of
307 chromocenters, and loss or diffusion of H3K9me2 in the nucleoplasm [27]. Thus, analyzing H3K9me2
308 subnuclear distribution by immunofluorescence (IF) experiments has been reproducibly used as a
309 cytological approach to assay for heterochromatin decondensation [12, 27, 28]. A previous study
310 showed that subnuclear distributions of chromocenters and H3K9me2 were unchanged in *main-2* and
311 *mail1-1* mutants [15]. However, fluorescent in situ hybridization (FISH) experiments using a DNA
312 probe for the 106B pericentromeric repeats suggested that heterochromatin was decondensed in the
313 two *pmd* mutants in comparison to WT plants [15]. We performed IF experiments to analyze the
314 subnuclear distribution of H3K9me2 in the *pp7l-2* mutant. These analyses did not show any change in
315 the condensation level of chromocenters in *pp7l-2* nuclei in comparison to WT (Fig 7). Instead, we
316 observed that *pp7l-2* nuclei were proportionally more condensed than WT nuclei (Fig 7). This is likely
317 due to the fact that *pp7l-2* mutant displays abnormal phenotype and growth delay in comparison to
318 WT plants that are entering the floral transition stage, a developmental stage where partial
319 chromocenter decondensation has been documented [29]. In conclusion, based on the H3K9me2 IF
320 experiments, we can conclude that *pp7l-2* is not impaired in chromocenter condensation.
321
322 **The PMD and PP7 domains have co-evolved among the Eudicots.**
323 Among the Angiosperms, most of the genic PMDs, like MAIN and MAIL1, are standalone
324 versions [18]. However, some genic PMDs can associate with other protein domains, such as for
325 instance a PPP domain. In *A. thaliana*, the protein MAIL3, which carries a PMD fused to a PPP domain,
326 is a close homolog of both MAIN/MAIL1 and PP7/PP7L through its PMD and PPP domains, respectively.
327 Considering that the PMD proteins MAIN and MAIL1 interact with PP7L, and are required for the
328 expression of similar set of loci, we decided to determine the distribution of related genic PMD and
329 PPP domains, and to retrace their evolutionary history among plant species. The *A. thaliana* MAIN,
330 MAIL1 and MAIL3 are all members of the PMD-C family that also includes MAIL2 [15]. Since our
331 objective is to retrace the evolution of genic (and not TE-containing) PMD-C, we have decided to

restraint our search to Eudicots. Indeed, Eudicot species contain mainly genic PMD-C, while other angiosperms may contain variable numbers of closely related genic and TE-associated PMD-C motifs that would be difficult to distinguish in our analysis. To retrace the evolution history of the genic PMD-C family, we used *A. thaliana* PMD-C genes to search and collect their relatives (paralogues and orthologues) in 30 genomes representative of the Eudicot diversity (see S11 Table for a list of species and their corresponding codes used in Fig 8, and S12 Table for motif sequences).

In our phylogenetic analysis, the genic PMD-C family can be clearly separated in two major clades. The first clade is composed of orthologues of *A. thaliana* MAIL2, MAIL1 and MAIN, while the second one includes orthologues of *A. thaliana* MAIL3 (Fig 8A). MAIL2 orthologues were found in all species tested, forming a closely related group, which suggests that they are under strong purifying selection (see the very short branch lengths linking most MAIL2 genes in Fig 8A). In several species, additional MAIL2 paralogues were also detected. They were either imbedded in the major MAIL2 group, or forming independent and more divergent subgroups, like in the case of MAIL1 and MAIN that are Brassicaceae-specific MAIL2 paralogues. By comparison, MAIL3 orthologues were not found in all Eudicot species tested, and, except in Brassicaceae, *MAIL3* genes appear to be under much weaker purifying selection compare to *MAIL2* and *MAIL2-like* genes (see the longer branch lengths in the tree of Fig 8A). Brassicaceae *MAIL3* genes contrast with other *MAIL3*, by forming a closely related group in the phylogenetic tree. This suggests a clear change in selection pressure, typical of a neofunctionalization event that could correlate with the acquisition of the PPP motif by these genes (Fig 8B and see below). Remarkably, another fusion event between PMD-C and PPP motifs occurred independently in grapevine, but this time involving a MAIL2 paralogue (VvMAIL2.2, Fig 8A).

We then used the PPP motif found in *A. thaliana* MAIL3, to collect orthologous genes and retrace the evolution history of this motif in the same Eudicot species used above. We confirmed that these genes can be clearly separated in two distinct clades: PP7 and PP7-like (PP7L) (Fig 8B). All tested species present one or several closely related PP7 paralogues. Although the Brassicaceae MAIL3 PPP motif belongs to the PP7 clade, it diverged significantly compared to other standalone PP7 paralogues

(Fig 8B). Same observation was made regarding the PP7 domain of VvMAIL2.2. Thus, as described for the PMD of Brassicaceae MAIL3 and grapevine VvMAIL2.2, this suggests a fast-evolving period and neofunctionalization of the PP7 domain in these species, subsequently to the PMD-C/PP7 fusion. Conversely, PP7L orthologues were not found in all species tested and, accordingly, these genes are under weaker purifying selection compare to genes belonging to the PP7 subfamily. In conclusion, phylogenetic analyses showed that, in at least Brassicaceae and grapevine, neo-association of PMD-C and PP7 domains have potentially create new protein functions that were maintained through evolution.

DISCUSSION

In *A. thaliana,* MAIN and MAIL1 are standalone PMD proteins that have been involved in genome integrity, regulation of cell division and differentiation, and silencing of TEs [15-17]. In this study, we show that TE silencing is widely impaired in the *ddc main-3* higher order mutant, which is both partially defective in DNA methylation and MAIN activity. We also identify the putative phosphatase protein PP7L as MAIN and MAIL1 protein interactor, and show that among the loci that are commonly misregulated in *pmd* and *pp7l* single and double mutants, most of the downregulated genes carry the 'DOWN' DNA motif in their promoter. Finally, phylogenetic analyses among Eudicots suggest a mechanism of neofunctionalization between the PMD and PP7-type PPP, to potentially acquire a functional module that requires the two protein domains.

The PMD MAIN protein acts independently of DRM2- and CMT3 pathways to silence TEs.

Previous analyses showed that some TEs were synergistically upregulated in the *mail1 rdr2* double mutant plants, suggesting that MAIL1 acts independently of RdDM pathway [15]. In our whole genome transcriptomic analyses, we show that several TEs are upregulated in both *main-3* and *ddc* mutants, and their upregulation is dramatically exacerbated in the *ddc main-3* quadruple mutant (Fig 2 and S2 Fig). We also identify TEs that are upregulated in either *ddc* or *main-3* mutants, but display

stronger misregulation in the *ddc main-3* higher order mutant (Fig 2 and S2 Fig). Finally, we identify a large class of TEs that are only upregulated in *ddc main-3* (Fig 2 and S2 Fig). Altogether, these analyses reveal complex genetic interaction between the MAIN, DRM2 and CMT3 proteins to silence TE. Depending on the locus, these proteins act redundantly or synergistically to efficiently silence TE. Previous work showed that DNA methylation is not impaired in *mail1-1* [15]. We found that DNA methylation is mostly unaffected in the *main-2* null mutant. However, we detected a mild but significant hypermethylation at non-CG sites in TEs and pericentromeric genes (Fig 4). One hypothesis is that CHG and CHH hypermethylation observed in *main-2* is a backup mechanism to compensate for MAIN loss of function, and to dampen TE silencing defects. Although further studies will be required to test this hypothesis, it is consistent with the fact that combining the *main-3* and *ddc* mutations leads to an exacerbation of TE silencing defects. Synergistic effects between different epigenetic pathways have already been described. For instance, it has been shown that MORPHEUS MOLECULE 1 (MOM1) and MORC1/MORC6 proteins, or MOM1 and the RdDM pathway act synergistically to efficiently silence TEs [13, 30]. Altogether, these observations contribute to the "mille-feuille" (i.e. "multiple layers") model, in which different epigenetic pathways converge towards the silencing of TEs [31].

The putative phosphatase PP7L interacts with the PMD MAIN and MAIL1 protein to regulate a similar set of genes and TEs.

Recently, the putative phosphoprotein phosphatase PP7L was involved in the biogenesis of chloroplasts and plant response upon abiotic stress [25]. Here, we show that PP7L interact with MAIN and MAIL1, and *main-2*, *pp7l-2, mail1-1* single and *mail1-1 pp7l-2* double mutant plants display similar developmental and molecular phenotypes (Fig 5 and 6). We also show that, as described for *main-2* and *mail1-1* [15], the subnuclear distribution of chromocenters and H3K9me2 are unaltered in *pp7l-2* (Fig 7). The 106B pericentromeric repeats appeared decondensed in the two *pmd* mutants [15], future work will determine if similar phenotype is observed in *pp7l-2*. Although MAIN, MAIL1 and PP7L

interact together, we cannot exclude that an additional protein is required for the interaction. In addition, PP7L may have additional partners independently of MAIN and MAIL1. Further biochemical studies such as IP-MS analyses using the FLAG-tagged PP7L line will contribute to addressing these points.

Transcriptomic analyses revealed complex genetic interaction between MAIN, MAIL1 and PP7L; the three proteins acting either independently or together to ensure the proper expression of genes, and to perform TE silencing. Moreover, transcriptome profiling of *mail1-1 pp7l-2* double mutant revealed that the two mutations may have synergistic effects, specifically at genes that are upregulated in the mutant. To further study the genetic interaction between the three proteins, it will be important to analyze the transcriptome of *main-2 mail1-1 pp7l-2* triple mutant. Altogether and considering that *i)* MAIN, DRM2 and CMT3 pathways genetically interact together, and *ii)* the *main-2* mutant show a slight increase in DNA methylation at CHG and CHH sites, we cannot rule out that MAIN is playing a dual role: regulating gene expression through its interaction with MAIL1 and PP7L, and involved in TE silencing through its genetic interaction with DNA methylation. In the future, it will be important to analyze DNA methylation in *pp7l-2*, but also in *pmd pp7l-2* higher order mutants. In parallel, studying the *ddc pp7l-2* mutant will allow to further decipher the genetic interaction between the PP7L and DNA methylation pathways.

Genes that are commonly downregulated in *main, mail1* and *pp7l* mutants carry the 'DOWN' motif in their promoters.

Genes that are commonly downregulated in the different *pmd* and *pp7l* mutant backgrounds show a bigger overlap than other misregulated loci (Fig 6, and S8 Table). Furthermore, eighteen out of nineteen genes commonly downregulated in the all the mutant backgrounds carry the 'DOWN' DNA motif in their promoter (S5 Fig). The 'DOWN' motif is also found in the promoter of genes reproducibly downregulated in *main-2*, *mail1-1*, *main-3* and *ddc main-3*. However, it is not significantly enriched in the promoter of genes identified as downregulated in the *pmd* and *pp7l* mutants from the RNA-seq

436 Exp3. One explanation for this discrepancy is that too many loci were identified as downregulated in
437 each dataset of the RNA-seq Exp3, which created a dilution of the loci carrying the 'DOWN' motif in
438 their promoter. In the future, further RNA-seq experiments in *pp7l-2* and *mail1-1 pp7l-2* will precise
439 the lists of reproducibly misregulated loci in these mutant backgrounds.

440 Based on our results, we hypothesize that the 'DOWN' motif may act as a putative cis-
441 regulatory element (CRE) recognized by an unidentified TF, which would be required for the
442 transcription of genes identified as downregulated in *pmd* and *pp7l* mutants. This unknown TF could
443 be recruited or activated by the MAIN/MAIL1/PP7L protein complex. Another hypothesis is that the
444 'DOWN' motif is directly recognized by the MAIN/MAIL1/PP7L protein complex. Further study will be
445 required to test if MAIN/MAIL1/PP7L protein complex interact with chromatin, and bind the 'DOWN'
446 motif. In parallel, further biochemical analyses may allow to identify an uncharacterized putative TF
447 as MAIN/MAIL1/PP7L protein interactor.

448

449 **The association of PMD-C and PP7/PP7L domains creates a functional protein module.**

450 In this study, we identified PP7L has a protein partner of the two standalone PMDs MAIN and
451 MAIL1, and showed that these proteins are required for the proper expression of a common set of
452 genes, and for TE silencing. Besides, we showed that the Brassicaceae MAIL3 and the grapevine
453 VvMAIL2.2 proteins carry a PMD fused to a PP7 domain. Based on these results, we hypothesize that
454 depending on the configuration, the association of PMD-C and PP7/PP7L domains would create a
455 functional protein module in trans or in cis. It is likely that the cis-association of PMD and PP7 found
456 in the Brassicaceae MAIL3 proteins occurred in the common ancestors of this Eudicot lineage, possibly
457 through the process of gene duplication. Since then, the MAIL3 PMD/PP7 fusion was maintained
458 under strong purifying selection, arguing for a neofunctionalization of the fusion protein. It is likely
459 that a similar process happened in grapevine, and possibly, in closely related Vitaceae species. To
460 some extent, the two distinct events that occurred in Brassicaceae and grapevine are reminiscent of
461 convergent evolution processes leading to the production of a functional PMD/PP7 module.

462 The occurrence of PMD and PP7/PP7L protein fusion in several Brassicaceae and grapevine is
463 reminiscent of the concept of Rosetta stone chimera proteins, which describes that two proteins
464 interacting together in one organism can be found fused together in another species to facilitate
465 enzymatic activity [32]. There are several examples of Rosetta stone proteins, described for instance
466 with different subunits of DNA topoisomerase or RNA polymerase [32]. Here, we show that, at least
467 in *A. thaliana,* the Rosetta stone chimera MAIL3 coexist with its close homologs MAIN/MAIL1 and PP7L
468 that interact together. The fact that the PMD and PP7 domains are fused together in MAIL3 may be a
469 strategy to optimize protein activity. Conversely, the enzymatic activity of the MAIN/MAIL1/PP7L
470 protein complex could be further regulated by allowing, or not, the three proteins to interact together.
471 Nevertheless, in both scenarios, it is likely that PMD and PP7/PP7L association creates a functional
472 protein module, which might be specialized in distinct biological processes depending on its
473 composition. Thus, we hypothesize that the MAIL3 and MAIN/MAIL1/PP7L protein complexes play
474 different role in the plant. This is consistent with the fact that, unlike *main-2*, *mail1-1* and *pp7l-2*
475 mutant, the *mail3-2* mutant does not show abnormal developmental phenotype [17]. Further studies
476 will be required to describe the role of MAIL3 in the plants.

477 In conclusion, we show here that the two *A. thaliana* PMD MAIN and MAIL1 proteins interact with
478 PP7L, and are involved in the regulation of a common set of genes and TEs. In addition, we show that
479 distinct events of PMD-C and PP7 fusions have occurred among the Eudicots (among several
480 Brassicaceae species and in grapevine), suggesting some convergent evolution processes and a
481 potential neofunctionalization of PMD/PP7 module in cis. The biological significance of PMD/PP7
482 fusion proteins will be investigated in the future by studying the role of MAIL3 in *A. thaliana*. In
483 addition, it will be important to determine whether the PMD proteins play important roles in other
484 plant species with agronomic value.

485

486 MATERIALS AND METHODS

487

488 **Plant material and growing conditions.** Wild-type (WT) and all mutant lines are in the Columbia (Col)
489 ecotype. The *drm1-2* (SALK_031705), *drm2-2* (SALK_150863), *cmt3-11* (SALK_148381), *ddc* triple,
490 *main-2* (GK-728H05), *mail1-1* (GK-840E05) and *pp7l-2* (SALK_003071) null mutant lines were
491 previously described [15-17, 25, 26], and obtained from The Nottingham Arabidopsis Stock Centre.
492 The *mail1-1 pp7l-2* double mutant was obtained by crossing the respective single mutants. T-DNA
493 insertions were confirmed by PCR-based genotyping and RT-qPCR analyses. The *main-3* mutation was
494 genotyped by derived cleaved amplified polymorphic sequences (dCAPS) using the restriction enzyme
495 FokI. Primer sequences are described in S13 Table. All the WT Col and T-DNA mutant plants were
496 grown on soil under a 16h-light/8h-dark cycle. When experiments required to screen for GFP
497 expression under UV light, plants carrying the *ATCOPIA28::GFP* transgene were first grown on
498 Murashige and Skoog (MS) plates under continuous light, 10-day old plants were then screened for
499 GFP expression under UV light, and subsequently transferred onto soil. For *in vitro* plant culture, seeds
500 were surface-sterilized and sowed on solid MS medium containing 0.5% sucrose (w/v).
501
502 **Cloning of ATCOPIA28::GFP.** The pCambia3300-NLS-GFP-T35S vector was previously described [12].
503 The 5'LTR promoter corresponding to a region of ~1 kb upstream of *ATCOPIA28 (AT3TE51900)* was
504 PCR amplified from WT genomic DNA, and cloned into pCR2.1 TOPO vector (Invitrogen). Quikchange
505 site-directed mutagenesis (Stratagene) was performed according to Manufacturer's instruction to
506 create a polymorphism site (MfeI→NdeI) within the 5'LTR promoter, which was subsequently
507 mobilized into pCambia3300 upstream of NLS-GFP-T35S sequence. *ddc* triple mutant plants were
508 transformed with the AT*COPIA28::GFP* construct using the *Agrobacterium*-mediated floral dip
509 method [33]. Transgenic plants showing GFP fluorescence were backcrossed with a WT plant to
510 promote the silencing of *ATCOPIA28::GFP* in the F1 generation. F1 plants were self-crossed and their
511 F2 progenies were screened for GFP fluorescence, and PCR-based genotyped to obtain
512 *ATCOPIA28::GFP* WT and *ATCOPIA28::GFP ddc* plants. Primer sequences used for *ATCOPIA28::GFP*
513 cloning and PCR genotyping are described in S13 Table.

514

515 **EMS mutagenesis, GFP screening and mapping analyses.** Five thousand seeds of *ATCOPIA28::GFP ddc*
516 were mutagenized in 0.26% EMS solution for 12 hours with rotation. Seeds were subsequently washed
517 with water and sown on soil. Fifteen hundred M2 populations were collected, and subsequently
518 screened for GFP fluorescence under UV light using a SMZ18 Nikon Fluorescence Stereomicroscope
519 coupled with the C-HGFI intensilight fluorescence filter. Pictures were taken using the DS Qi1MC digital
520 camera kit. Mapping and identification of the EMS mutation responsible for the phenotype were
521 performed by bulk segregant analysis coupled with deep genome re-sequencing as previously
522 described [12], with the following differences. Reads were mapped against the reference genome
523 (*Arabidopsis* TAIR10) and single nucleotide polymorphisms called in Geneious (Biomatters). Using R,
524 single nucleotide polymorphisms were filtered for EMS mutations (G:C→A:T) and zygosity called
525 based on the variant frequency provided by Geneious (≥80% homozygous mutation, ≥45%, and ≤55%
526 heterozygous mutation). Plots were then created by calculating the ratio of the number of
527 homozygous and heterozygous and mutations in a 500-kb window as previously described [34].

528

529 **Cloning of epitope-tagged versions of PMD and PP7L proteins.** *MAIN*, *MAIL1* and *PP7L* genomic
530 regions were PCR amplified and FLAG or Myc epitopes were added to the C-terminus of each protein
531 as previously described [12]. Each time, the amplified region includes a ~1Kb promoter sequence
532 upstream of the respective transcriptional start site. For the *MAIN* promoter, a MluI site was modified
533 to allow LR reaction without changing the sequence integrity of the gene. *main-2*, *mail1-1* and *pp7l-2*
534 mutant plants were transformed with the *MAIN-FLAG*, *MAIN-MYC*, *MAIL1-MYC* and *PP7L-*
535 *FLAG* constructs using the *Agrobacterium*-mediated floral dip method [33]. Primer sequences are
536 described in S13 Table.

537

538 **IP and MS analysis.** Ten grams of 3-week-old seedling tissue were ground in liquid nitrogen and
539 resuspended in 50mL ice-cold IP buffer [50mM Tris HCl pH 7.6, 150mM NaCl, 5mM $MgCl_2$, 0.1%

Nonidet P-40, 10% glycerol (v/v), 0.5mM DTT, 1x Protease Inhibitor Mixture (Roche)] and centrifuged 2 times for 15 min at 4°C at 15 350g. 400µL of M2 magnetic FLAG-beads (Sigma, M8823) were added to the supernatants, and incubated for 90min rotating at 4°C. M2 magnetic FLAG-beads were washed seven times in ice-cold IP buffer for 5 min rotating at 4°C, and immunoprecipitated proteins were eluted 3 times with 150µL 3x-FLAG peptides (Sigma, F4799) for 25 min each at 25°C. The eluted protein complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously described [13]. Peptide and protein-level false discovery rates were calculated by the DTASelect algorithm using the decoy database approach. Based on a peptide PSM level p-value filter of less than 0.01 and a requirement for at least two peptides per protein, the protein-level false discovery rate was less than 1% for all proteins detected.

Co-IP and immunoblotting. 0.5 g of 3-week-old seedling tissue were ground in liquid nitrogen, resuspended in 1.5mL ice-cold IP buffer [50mM Tris pH 7.6, 150mM NaCl, 5mM MgCl2, 0.1% Nonidet P-40, 10% glycerol, 0.5 mM DTT, 1x Protease Inhibitor Mixture (Roche)], and centrifuged 2 times for 15 min at 4°C, 16 000g. 50µL M2 magnetic FLAG-beads (Sigma, M8823) were added to the supernatants and incubated for 2 hour rotating at 4°C. Beads were washed 3 times in ice-cold IP buffer for 10 min rotating at 4°C. Immunoprecipitated proteins were denatured in Laemmli buffer for 5min at 95°C. 10µL of input and bead elution were run on 10% SDS-PAGE gels, and proteins were detected by western blotting using either Anti-FLAG M2 monoclonal antibody-peroxidase conjugate (Sigma, A8592) at a dilution of 1:10000, or c-Myc rat monoclonal antibody (Chromotek, 9E1-100) at a dilution of 1:1000 followed by goat anti-rat IgG horseradish peroxidase (Abcam, ab205720) used at a dilution of 1:20000 as secondary antibody. Western blots were developed using Substrat HRP Immobilon Western (Merck Millipore, WBKLS0500).

564 **RNA extraction.** Total RNA was extracted from aerial parts of 3-week-old seedlings grown on soil using
565 either RNeasy Plant Mini Kit (Qiagen, 74904) or Monarch Total RNA Miniprep Kit (NEB, T2010)
566 according to the manufacturer's protocols.
567
568 **RNA sequencing.** RNA-seq libraries were generated from 1µg of input RNA using NEBNext Ultra II
569 Directional RNA Library Prep Kit for Illumina (NEB, E7490) according to the manufacturer's protocols.
570 Libraries were sequenced on an Illumina HiSeq 4000 or NextSeq 550 machines. Reads were trimmed
571 using Trimmomatic [35], and mapped to the *A. thaliana* genome (*Arabidopsi*s TAIR10 genome) using
572 HISAT2 [36]. The sequence alignment files were sorted by name and indexed using SAMtools [37].
573 Files were converted to BAM files and number of reads mapped onto a gene calculated using HTSeq-
574 count [38]. Differentially expressed genes were obtained with DESeq2 [39], using a log2 fold-change
575 ≥ 2 (up-regulated genes) or ≤ -2 (down-regulated genes) with an adjusted *p*-value of 0,01. Re-analyses
576 of previously published RNA-seq datasets from *main-2* and *mail1-1* (PRJEB15202) [15] were
577 performed as described above.
578
579 **RT-qPCR.** 1 µg of input RNA was converted to cDNA using GoScript Reverse Transcriptase (Promega
580 A501C) according to the manufacturer's protocol. The final reaction was diluted 6 times with RNase
581 free water. RT-qPCR experiments were performed with 4µL of cDNA combined to the Takyon No Rox
582 SYBR MasterMix (Eurogentec, UF-NSMT-B0701), using a LightCycler 480 instrument (Roche).
583 Amplification conditions were as follows: 95°C 5 min; 45 cycles, 95°C 15s, 60°C 15s, 72°C 30s; melting
584 curves. RT-qPCR analyses used the $2^{-\Delta\Delta Ct}$ method. For each analysis, ΔCt was first calculated based on
585 the housekeeping *RHIP1* gene Ct value [40]. ΔΔCt were then obtained by subtracting the wt ΔCt from
586 the ΔCt of each sample. Values were represented on bar charts relative to WT. Three technical
587 replicates were performed per biological replicate, and 3 biological replicates were used in all
588 experiments, unless otherwise stated. Primer sequences are described in S13 Table.
589

590 **DNA motif detection.** The motifs for enhancer sequences (1kb upstream the TSS) were discovered
591 using MEME (Multiple Em for Motif Elicitation). MEME represents motifs as position-dependent
592 letter-probability matrices which describe the probability of each possible letter at each position in
593 the pattern [41].
594
595 **Bisulfite sequencing.** Genomic DNA was extracted from aerial parts of 3-week-old seedlings using
596 Quick-DNA Plant/Seed Miniprep Kit (Zymo research, D6020) according to the manufacturer's protocol.
597 Whole genome bisulfite sequencing (WGBS) library was prepared from 50 ng genomic DNA using
598 NuGen Ovation Ultralow Methyl-Seq kit. Bisulfite treatment was carried out by Qiagen Epitect bisulfite
599 kit. WGBS libraries were sequenced on an Illumina HiSeq 4000 machine. The raw reads (single end)
600 were trimmed using Trimmomatic in order to remove adapter sequences [35]. The remaining
601 sequences were aligned against the *A. thaliana* genome TAIR10 version using Bismark [42]. Duplicated
602 reads were collapsed into one read. For visualization of the data, we used ViewBS [43].
603
604 **Sequence selection, multiple sequences alignments and phylogenetic reconstruction.**
605 Blast searches (blastp) were performed starting from known *A. thaliana* PMD-C and PP7/PP7L motifs
606 on the thirty species representing the diversity of the Eudicot lineages. When necessary tblastn
607 searches were also used to obtain complete protein sequences. To build the phylogenetic trees, PMD-
608 C or PP7/PP7L motifs were aligned using the multiple sequence comparison by log-expectation
609 (MUSCLE v3.7) software [44]. Trees were reconstructed using the fast-maximum likelihood tree
610 estimation program PHYML [45] using the LG amino acids replacement matrix [46]. Statistical support
611 for the major clusters were obtained using the approximate likelihood-ratio test (aLRT) [47].
612
613 **Immunofluorescence and DAPI-staining.** Leaves from 3-week-old plants, were fixed for 20 min
614 rotating at 4°C in 2% formaldehyde in Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl),
615 washed two times for 10 min rotating at 4°C in cold Tris buffer and subsequently chopped in LB01

buffer (15 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20mM NaCl and 0.1% Triton- X-100). Nuclei were filtered through a 30 µm cell strainer cap (Sysmex, 04-0042-2316) and 5µl of the nuclei solution was diluted in 10 µl of sorting buffer (100mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM MgCl2, 0.05% Tween-20 and 5% sucrose). 20µl of the nuclei dilution were spread onto a polylysine slide and air-dried for 40 min. Slides were post-fixed in 2% formaldehyde in 1X PBS for 5 min and washed 2 times with water. Slides were incubated 15 min in 1X PBS, 0.5% Triton X-100 at RT and washed 3 times with 1X PBS for 5 min. For detection, slides were incubated over night with a mouse anti-H3K9me2 monoclonal antibody (Abcam, Ab 1220) at 1:500 in 3% BSA, 0.05% Tween in 1X PBS at 4°C in a moist chamber. After 3 washes in 1X PBS for 5 min, slides were incubated 2h with a goat anti-mouse antibody coupled to Alexa fluor 568 (Invitrogen, A11004) at 1:1000 in 3% BSA, 0.05% Tween in 1X PBS in a moist chamber. Slides were washed 1 time 5 min with 1X PBS, 1 time 10 min with 1X PBS, 1µg/mL DAPI, and 1 time 5 min with 1X PBS. DNA was counterstained with 1µg/mL DAPI in Vectashield mounting medium (Vector Laboratories). Observation and imaging were performed using a LSM 700 epifluorescence microscope (Zeiss).

Data availability. Nucleotide sequencing data generated in this study have been deposited in European Nucleotide Archive (ENA) under the accession number PRJEB33240 (http://www.ebi.ac.uk/ena/data/view/PRJEB33240). The proteomics data have been deposited to the MassIVE data repository (https://massive.ucsd.edu) with the dataset identifier MSV000084089. All other data and material are available within the manuscript and its supplementary files, or from the corresponding author upon request.

ACKNOWLEDGMENTS

The authors want to thank Thierry Lagrange, Frederic Pontvianne and other team members for fruitful discussions, and all the LGDP platform members for their outstanding technical assistance and plant care. This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-

642 10-LABX-41). S.E.J. is an investigator of the Howard Hughes Medical Institute.

643

644 ## REFERENCES

645

646 1. Grewal SI, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8(1):35-46.
647 2. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the
648 genome. Nat Rev Genet. 2007;8(4):272-85.
649 3. Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications.
650 Nat Rev Genet. 2019;20(7):417-31.
651 4. Dergai O, Hernandez N. How to Recruit the Correct RNA Polymerase? Lessons from snRNA
652 Genes. Trends Genet. 2019;35(6):457-69.
653 5. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in
654 plants and animals. Nat Rev Genet. 2010;11(3):204-20.
655 6. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of
656 increasing complexity. Nat Rev Genet. 2014;15(6):394-408.
657 7. Wendte JM, Pikaard CS. The RNAs of RNA-directed DNA methylation. Biochim Biophys Acta.
658 2016.
659 8. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with
660 histone methylation. Nat Rev Mol Cell Biol. 2015;16(9):519-32.
661 9. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. Nat Rev Mol
662 Cell Biol. 2018;19(8):489-506.
663 10. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The
664 Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing
665 heterochromatin. Cell. 2013;153(1):193-205.
666 11. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape
667 the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21(1):64-72.
668 12. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, et al. MORC family ATPases required
669 for heterochromatin condensation and gene silencing. Science. 2012;336(6087):1448-51.
670 13. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, Yen L, et al. Transcriptional gene
671 silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proc Natl
672 Acad Sci U S A. 2014;111(20):7474-9.
673 14. Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M. Involvement of a GHKL ATPase in RNA-
674 directed DNA methylation in Arabidopsis thaliana. Curr Biol. 2012;22(10):933-8.
675 15. Ikeda Y, Pelissier T, Bourguet P, Becker C, Pouch-Pelissier MN, Pogorelcnik R, et al.
676 Arabidopsis proteins with a transposon-related domain act in gene silencing. Nat Commun.
677 2017;8:15122.
678 16. Wenig U, Meyer S, Stadler R, Fischer S, Werner D, Lauter A, et al. Identification of MAIN, a
679 factor involved in genome stability in the meristems of Arabidopsis thaliana. Plant J. 2013;75(3):469-
680 83.
681 17. Uhlken C, Horvath B, Stadler R, Sauer N, Weingartner M. MAIN-LIKE1 is a crucial factor for
682 correct cell division and differentiation in Arabidopsis thaliana. Plant J. 2014;78(1):107-20.
683 18. Babu MM, Iyer LM, Balaji S, Aravind L. The natural history of the WRKY-GCM1 zinc fingers
684 and the relationship between transcription factors and transposons. Nucleic Acids Res.
685 2006;34(22):6505-20.
686 19. Steinbauerova V, Neumann P, Novak P, Macas J. A widespread occurrence of extra open
687 reading frames in plant Ty3/gypsy retrotransposons. Genetica. 2011;139(11-12):1543-55.
688 20. Farkas I, Dombradi V, Miskei M, Szabados L, Koncz C. Arabidopsis PPP family of
689 serine/threonine phosphatases. Trends Plant Sci. 2007;12(4):169-76.

21. Sun X, Kang X, Ni M. Hypersensitive to red and blue 1 and its modification by protein phosphatase 7 are implicated in the control of Arabidopsis stomatal aperture. PLoS Genet. 2012;8(5):e1002674.
22. Liu HT, Li GL, Chang H, Sun DY, Zhou RG, Li B. Calmodulin-binding protein phosphatase PP7 is involved in thermotolerance in Arabidopsis. Plant Cell Environ. 2007;30(2):156-64.
23. Genoud T, Santa Cruz MT, Kulisic T, Sparla F, Fankhauser C, Metraux JP. The protein phosphatase 7 regulates phytochrome signaling in Arabidopsis. PLoS One. 2008;3(7):e2699.
24. Uhrig RG, Labandera AM, Moorhead GB. Arabidopsis PPP family of serine/threonine protein phosphatases: many targets but few engines. Trends Plant Sci. 2013;18(9):505-13.
25. Xu D, Marino G, Klingl A, Enderle B, Monte E, Kurth J, et al. Extrachloroplastic PP7L Functions in Chloroplast Development and Abiotic Stress Tolerance. Plant Physiol. 2019.
26. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152(1-2):352-64.
27. Fransz P, de Jong H. From nucleosome to chromosome: a dynamic organization of genetic information. Plant J. 2011;66(1):4-17.
28. Zhao S, Cheng L, Gao Y, Zhang B, Zheng X, Wang L, et al. Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. Cell Res. 2019;29(1):54-66.
29. Tessadori F, Schulkes RK, van Driel R, Fransz P. Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis. Plant J. 2007;50(5):848-57.
30. Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, Mittelsten Scheid O, et al. MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. Embo J.29(2):340-51.
31. Rigal M, Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. Biochim Biophys Acta. 2011;1809(8):452-8.
32. Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. Nat Biotechnol. 2000;18(6):609-13.
33. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 1998;16(6):735-43.
34. Hristova E, Fal K, Klemme L, Windels D, Bucher E. HISTONE DEACETYLASE6 Controls Gene Expression Patterning and DNA Methylation-Independent Euchromatic Silencing. Plant Physiol. 2015;168(4):1298-308.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-60.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
38. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.
39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
40. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. 2005;139(1):5-17.
41. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-8.
42. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571-2.
43. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. Bioinformatics. 2018;34(4):708-9.

741 44. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
742 Nucleic Acids Res. 2004;32(5):1792-7.
743 45. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies
744 by maximum likelihood. Syst Biol. 2003;52(5):696-704.
745 46. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol.
746 2008;25(7):1307-20.
747 47. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate,
748 and powerful alternative. Syst Biol. 2006;55(4):539-52.
749 48. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of
750 histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE.
751 2008;3(9):e3156.

752

753 FIGURE CAPTIONS

754 **Fig 1. The *ddc #16* EMS population shows overexpression of *ATCOPIA28::GFP* and upregulation of**
755 **endogenous TEs.**

756 (A) Schematic representation of the *ATCOPIA28::GFP* transgene. The 5' long terminal repeat (LTR)
757 promoter region of an *ATCOPIA28* LTR-retrotransposon (*AT3TE51900*) is used to control the
758 expression of GFP. The construct carries a Nuclear Localization Signal (NLS) to target the GFP in the
759 nucleus. (B) WT and *drm1 drm2 cmt3* (*ddc*) triple mutant plants carrying the *ATCOPIA28::GFP*
760 transgene showed no and weak GFP fluorescence under UV light, respectively. By comparison, the *ddc*
761 *#16* EMS mutant showed strong GFP fluorescence. Insets show plants under white light. (C) Western
762 blot using anti-GFP antibody confirmed *ATCOPIA28::GFP* overexpression in *ddc #16*. Coomassie
763 staining of the large Rubisco subunit (rbcL) is used as a loading control. KDa: kilodalton. (D) Relative
764 expression analyses of *ATCOPIA28::GFP (GFP)* and three endogenous TEs in *ddc* and *ddc #16* assayed
765 by Real-Time quantitative PCR (RT-qPCR). RT-qPCR analyses were normalized using the housekeeping
766 *RHIP1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate
767 standard deviation based on three independent biological replicates. Screening of EMS mutant
768 populations was done on MS plates to allow for visualization of GFP-positive individuals under UV
769 light.

Fig 2. MAIN, DRM2 and CMT3 act synergistically to repress TEs.

(A) Representative pictures showing the developmental phenotype of 3-week-old *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plant. (B) Number of upregulated TEs in *ddc*, *main-3* and *ddc main-3*, and classified by TE superfamily. (C) Chromosomal distributions of misregulated loci in *ddc*, *main-3* and *ddc main-3* over WT. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as defined in [48]. Upregulated genes are represented in blue, downregulated genes in green and TEs are represented in red above the chromosomes (upregulated) or below (downregulated). (D) Fraction of upregulated TEs in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [48]. Asterisks indicate statistically significant enrichments of TEs in pericentromeric regions in comparison to the genomic distribution of all *A. thaliana* TEs (Chi-Square test, *: p-value≤ 0.05, n.s: not significant). (E) Heatmap showing normalized count reads of upregulated TEs in *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plants. (F-G) Relative expression analyses of *ATCOPIA28* (F) and several endogenous TEs (G) in *ddc*, *main-3, ddc main-3, cmt3 main-3* and *drm1 drm2 (dd) main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Analyses described in panels B-E are based on the overlaps of two independent RNA-seq experiments (EMS Exp1 and Exp2, four biological replicates, S1 and S2 Tables). RNA-seq threshold: log2≥2, or log2≤-2 ; p-adj< 0.01.

Fig 3. MAIN and MAIL1 are required for the proper expression of similar genes, and for TE silencing.

(A-B) Number of reproducibly misregulated genes (A) and upregulated TEs (B) in *main-2, mail1-1* and *main-3* mutants in comparison to WT plants. TEs are classified by superfamily. (C) Heatmap showing normalized count reads of misregulated loci in *main-2*, *mail1-1* and *main-3* in comparison to

795 respective WT controls. Asterisks represents loci that are commonly misregulated in the three mutant
796 backgrounds. (D) Venn diagrams analyses representing the overlaps between misregulated loci in
797 *main-2*, *mail1-1* and *main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram
798 overlaps (p-value <1,2.10e-14). (E) Fraction of misregulated loci in *main-2* and *mail1-1* located in
799 chromosome arms or in pericentromeric regions as defined in [48]. Asterisks indicate statistically
800 significant enrichments of downregulated genes and upregulated TEs in chromosome arms and
801 pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana*
802 genes and TEs (Chi-Square test, *: p-value≤ 0.05, **: p-value≤ 0.01, n.s: not significant). Percentages
803 of DNA-methylated genes were calculated based on whole genome bisulfite sequencing (BS-seq) of
804 WT plants described in [26]. n.d.: not determined. Analyses are based on the overlaps of three
805 independent RNA-seq experiments (Exp1 and Exp2, two biological replicates each, and Exp3, four
806 biological replicates, S3 and S4 Tables). RNA-seq threshold: log2≥2, or log2≤-2 ; p-adj< 0.01.

807

808 **Fig 4. The *main-2* mutation has a slight effect on non-CG DNA methylation levels.**

809 (A) Genome-wide DNA methylation levels along the five *Arabidopsis* chromosomes in *main-2* versus
810 WT plants. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as
811 defined in [48]. Mb: megabase. (B-H) Boxplot analyses in two *main-2* and WT biological replicates
812 showing the DNA methylation levels of all pericentromeric TEs (B) and genes (C), all chromosome arms
813 TEs (D) and genes (E), TEs that are upregulated in *main-2* (F), and genes that are upregulated (G) and
814 downregulated (H) in *main-2*. p-values were calculated using a Wilcoxon test. ***: p-value < 2.10e-16.

815

816 **Fig 5. MAIN, MAIL1 and PP7L physically interact together.**

817 (A) Representative pictures of 3-week-old *main-2* and *mail1-1* mutants, and epitope-tagged
818 complementing lines in comparison to WT Col plants. (B) Western blots using anti-FLAG and anti-Myc

819 antibodies showing the accumulation of epitope-tagged PMD proteins at the expected sizes in the
820 different complementing lines. Coomassie staining of the large Rubisco subunit (rbcL) is used as a
821 loading control. KDa: kilodalton.(C-E) Relative expression analyses of upregulated TEs (C), upregulated
822 genes (D) and downregulated genes (E) in the different complementing lines assayed by RT-qPCR. RT-
823 qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the
824 complementing lines and mutants are represented relative to WT Col. Error bars indicate standard
825 deviation based on three independent biological replicates. (F) FLAG-tagged MAIN and MAIL1 proteins
826 were immunoprecipitated and putative interacting proteins were identified by mass spectrometry.
827 Numbers of identified spectra, peptides and the normalized spectral abundance factor (NSAFe5) are
828 shown for two independent experiments, including three *main-2* and two *mail1-1* replicates. WT
829 replicates are used as a negative control. Only proteins reproducibly enriched in all the FLAG-MAIN
830 and FLAG-MAIL1 IP, and depleted in WT controls across multiple replicates are described in the table.
831 (G) MAIL1-MYC was co-immunoprecipitated with MAIN-FLAG in F1 plants obtained by crossing MAIL1-
832 MYC and MAIN-FLAG lines together. Parental MAIL1-MYC and MAIN-FLAG lines were used as negative
833 controls. (H) The MAIN-MYC line was supertransformed with the PP7L-FLAG construct, and MAIN-MYC
834 was co-immunoprecipitated with PP7L-FLAG. Plants expressing only MAIN-MYC or PP7L-FLAG were
835 used as negative controls. (I) Same as H but using MAIL1-MYC plants supertransformed with the PP7L-
836 FLAG construct. Epitope-tagged proteins were detected by Western blotting. Arrowheads indicates
837 expected bands. Asterisks indicates non-specific hybridization. Co-exp: plants co-expressing PP7L-
838 FLAG and MAIN-MYC (H) or PP7L-FLAG and MAIL1-MYC (I).

839

840 **Fig 6. *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants display similar**
841 **developmental and molecular phenotypes.**

842 (A) Representative pictures of 3-week-old *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double
843 mutants in comparison to WT Col plant. (B) Heatmap showing normalized count reads of misregulated

844 loci in *main-2*, *mail1-1, pp7l-2* and *mail1-1 pp7l-2* mutants in comparison to WT col plants using the
845 datasets of RNA-seq Exp3 (4 biological replicates, S3 and S6 Tables). One asterisk defines the loci that
846 are commonly misregulated in all mutant backgrounds. Two asterisks define the loci that are
847 misregulated in the *mail1-1 pp7l-2* double mutant. (C) Venn diagrams analyses representing the
848 overlaps between misregulated loci in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2*. Fisher's exact test
849 statistically confirmed the significance of Venn diagram overlaps (p-value <1,2.10e-14). (D-F) Boxplots
850 analyses showing average RPKM values of upregulated TEs (D), upregulated genes (E) and
851 downregulated genes (F) in *mail1-1 pp7l-2* in the indicated genotypes of RNA-seq Exp3. These analyses
852 are based on the misregulated loci datasets defined by ** in panel B. P-values were calculated using
853 a Wilcoxon test, and only significant p-values are shown. *: p-value< 1.10e-3; ***: p-value< 2.10e-16.
854 (G) Heatmap showing normalized count reads of reproducibly misregulated loci in *main-2* and *mail1-*
855 *1* in the indicated genotypes of RNA-seq Exp3. Lists of reproducibly misregulated loci in *main-2* and
856 *mail1-1* were as defined in Fig 3 and S4 Table. One asterisk defines the loci that are commonly
857 misregulated in *main-2* and *mail1-1* mutants (S5 Table). Two asterisks define subsets of loci that are
858 commonly misregulated in all the indicated genotypes (S7 Table). (H-J) Relative expression analyses
859 of upregulated TEs (H), upregulated genes (I) and downregulated genes (J) in the different genotypes
860 assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and
861 transcript levels in the different mutants are represented relative to WT Col. Error bars indicate
862 standard deviation based on three independent biological replicates. (K-M) Boxplots analyses showing
863 average RPKM values of commonly upregulated TEs (K), upregulated genes (L) and downregulated
864 genes (M) in the indicated genotypes of RNA-seq Exp3. These analyses are based on the commonly
865 misregulated loci datasets defined by ** in panel G. P-value was calculated using a Wilcoxon test. ***:
866 p-value < 2.10e-16.

867

868 **Fig 7. Constitutive heterochromatin appears unaltered in *pp7l-2* mutant.**

869 Proportion of nuclei showing condensed, partially decondensed (intermediate), or decondensed
870 chromocenters in the *pp7l-2* mutant in comparison to WT control (Col) based on H3K9me2
871 immunostaining of nuclei. Representative pictures of nuclei displaying condensed, partially
872 decondensed or decondensed chromocenters. DAPI: DNA stained with 4',6-diamidino-2-
873 phenylindole.

874 **Fig 8. Evolutionary history of PMD-C and PP7 proteins in plants.**

875 (A) An alignment of the PMD-C motifs from 30 representative Eudicot species was used to construct
876 a phylogenetic tree. The two major clades (MAIL2/MAIL2-like and MAIL3) are indicated. The species
877 codes are given in S11 Table, and corresponding protein sequences in S12 Table). In red are genes
878 presenting a fusion between a PMD-C and a PP7 motif. Statistical supports of key nodes calculated
879 with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The
880 tree was rooted using the *Amborella trichopoda* PMD-C motif (Atr1PMDC). (B) Phylogenetic tree
881 constructed using an alignment of the PP7 motif from the same species as in (A). The two major clades
882 (PP7 and PP7L) are indicated. In red are genes presenting a fusion between a PP7 and a PMD-C motif.
883 Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated.
884 Scale bar indicates one substitution/site. The tree was rooted using the *A. thaliana* PP5 motif (AtPP5).

885
886

SUPPORTING INFORMATION CAPTIONS

S1 Fig. *MAIN* is the mutated gene responsible for *ATCOPIA28::GFP* and TE overexpression in the *ddc #16* mutant.

(A) Representative pictures of *ddc #18* (*ddc morc6-8*) and *ddc #344* (*ddc morc6-9*) mutants in comparison to *ATCOPIA28::GFP* WT and *ddc* control plants under UV light. Insets show plants under white light. (B) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the populations *ddc #18* and *ddc #344*. Mb: megabase. (C) Location of the point mutations corresponding to the *morc6-8* and *morc6-9* alleles within the *MORC6* genomic sequence. Nucleotide and corresponding amino acid changes are indicated above the gene. Positions of the mutations are indicated relative to the transcription start site (+1). Grey boxes represent 5' and 3' UTR, blue boxes and lines represent exons and introns, respectively. (D) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the population *ddc #16*. (E) Location of the point mutation corresponding to the *main-3* mutant allele within the *MAIN* genomic sequence. (F) Genetic complementation analyses using the KO T-DNA insertion line *main-2*. *ddc #16* plants were crossed with *main-2* plants. F1 plants were self-crossed, and F2 plants were screened under UV light to select GFP-overexpressing plants. Western blotting using anti-GFP antibodies confirmed GFP overexpression in selected F2 plants. Coomassie staining of the large Rubisco subunit (rbcL) is used as a loading control. KDa: kilodalton. Among the selected F2 plants, the presence of *main-3* EMS and *main-2* T-DNA mutant alleles were determined by dCAPS-PCR and PCR analyses, respectively. *DRM2* and *CMT3* genotyping were determined by PCR analyses. WT: Wild type, Ho: Homozygote mutant. He: Heterozygote. (G) Relative expression analyses of several TEs in the indicated genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation

based on two independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

S2 Fig. Combining the *drm2*, *cmt3* and *main-3* mutations exacerbate TE silencing defects.

(A) Number of misregulated loci in *ddc*, *main-3* and *ddc main-3* in comparison to *ATCOPIA28::GFP* WT plants in two independent RNA-seq experiments (EMS Exp1 and Exp2, two biological replicates each, S1 Table). (B) Principal component analysis (PCA) performed on normalized count reads for first two components of the sixteen samples described in RNA-seq EMS Exp1 and Exp2. (C) Venn diagrams analyses defining the reproducibly misregulated loci in the different genotypes based on the overlaps of loci identified as misregulated in RNA-seq EMS Exp1 and Exp2 (S2 Table). (D) Relative expression analyses of *ATCOPIA28* and *HELITRONY1D* (*AT5TE35950*) in *ddc*, *main-3* and *ddc main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. (E) Venn diagrams analysis showing the overlaps between reproducibly upregulated TEs in *ddc*, *main-3* and *ddc main-3*. The numbers of reproducibly upregulated TEs for each genotype were defined by the overlaps of upregulated TEs shown in panel C. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <1,2.10e-14). (F) Same as panel D for TEs defined as class I-IV TEs. Frames of RT-qPCR graphs are using the same color code as shown in panel E. (G) *top*, Venn diagrams analyses defining the overlaps between reproducibly up- and downregulated genes in the different genotypes. Reproducibly misregulated genes were defined based on the overlaps described in panel C. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <1,2.10e-14). *bottom,* Fraction of misregulated genes in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [48]. Asterisks indicate statistically significant enrichments of misregulated genes in chromosome arms or pericentromeric regions in comparison to the genomic

937 distributions of all *A. thaliana* genes (Chi-Square test, *: p-value≤ 0.05, **: p-value≤ 0.01, n.s: not
938 significant). Percentages of DNA-methylated genes were calculated based on whole genome bisulfite
939 sequencing (BS-seq) of WT plants described in [26]. n.d.: not determined. (H) Relative expression
940 analyses of *DRM2* and *CMT3* in *ddc*, *main-3*, *ddc main-3, cmt3 main-3* and *dd main-3* assayed by RT-
941 qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in
942 the different genotypes are represented relative to WT. Error bars indicate standard deviation based
943 on three independent biological replicates. Screening of EMS mutant populations was done on MS
944 plates to allow for visualization of GFP-positive individuals under UV light.

945

946 **S3 Fig. Identification of reproducibly misregulated loci in *main-2*, *mail1-1* and *main-3*.**

947 (A) Number of misregulated loci in *main-2* and *mail1-1* in comparison to WT Col plants in three
948 independent RNA-seq experiments (Exp1, Exp2 [15], and Exp3; two, two and four biological replicates,
949 respectively, S3 Table) (B) Principal component analysis (PCA) performed on normalized count reads
950 for first two components of the twenty-four *main-2*, *mail1-1* and WT Col samples described in RNA-
951 seq Exp1, Exp2 and Exp3. (C) Venn diagrams analyses defining the reproducibly misregulated loci in
952 *main-2* and *mail1-1* based on the overlaps of loci identified as misregulated in RNA-seq Exp1, Exp2
953 and Exp3 (S4 Table). (D-F) Relative expression analyses of several upregulated TEs (D), upregulated
954 genes (E), and downregulated genes (F) in *main-2, mail1-1* and *main-3* assayed by RT-qPCR. RT-qPCR
955 analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different
956 genotypes are represented relative to respective WT controls. Error bars indicate standard deviation
957 based on three independent biological replicates.

958 **S4 Fig. DNA methylation analyses in the *main-2* mutant**

959 (A-B) Boxplot analyses in two *main-2* and WT biological replicates showing the DNA methylation levels
960 at genomic sites previously defined as hypo CHG differentially methylated regions (DMR) in *cmt3* (A)
961 and hypo CHH DMR in *drm1 drm2* (B) based on [26]. p-values were calculated using a Wilcoxon test.

962 *: p-value <5.10e-7, **: p-value <5.10e-10, ***: p-value < 2.10e-16.

963 **S5 Fig. MAIN, MAIL1 and PP7L are required for the proper expression of similar loci, and commonly**
964 **downregulated genes carry the 'DOWN' DNA motif in their promoter.**

965 (A) Principal component analysis (PCA) performed on normalized count reads for first two components
966 of the twenty samples described in RNA-seq Exp3. (B) Number of misregulated loci in the different
967 genotypes in comparison to WT Col plants from RNA-seq Exp3 (four biological replicates, S3 and S6
968 Tables). (C) Identification and proportions of the 'DOWN' DNA motif among the promoters of
969 downregulated genes and all *Arabidopsis* genes using the MEME software. Promoter regions are
970 defined as 1kb upstream of ATG. The list of all *Arabidopsis* genes used to determine genomic
971 distributions is based on the TAIR file: TAIR10_upstream_1000_translation_start_20101028.

972 **S6 Fig. Full size images of panels described in Fig 5G-I.**

973 **S1 Table. Lists of differentially regulated loci in *ddc*, *main-3* and *ddc main-3* in two independent**
974 **RNA-seq experiments.**
975
976 **S2 Table. Lists of reproducibly differentially regulated loci in *ddc*, *main-3* and *ddc main-3*.**
977
978 **S3 Table. Lists of differentially regulated loci in *main-2* and *mail1-1* in three independent RNA-seq**
979 **experiments.**
980
981 **S4 Table. Lists of reproducibly differentially regulated loci in *main-2* and *mail1-1*.**
982
983 **S5 Table. Lists of loci reproducibly and commonly misregulated in *main-2*, *mail1-1* and *main-3*.**
984
985 **S6 Table. Lists of differentially regulated loci in *pp7l-2* and *mail1-1pp7l-2* in RNA-seq Exp3.**
986

S7 Table. Lists of loci commonly misregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1pp7l-2*.

S8 Table. Lists of loci commonly misregulated in all mutant backgrounds analyzed in this study.

S9 Table. Lists of commonly downregulated genes displaying the "DOWN" motif in their promoter and random test analyses.

S10 Table. Distribution of several bZIP DNA motifs among the promoter regions of all *Arabidopsis* genes.

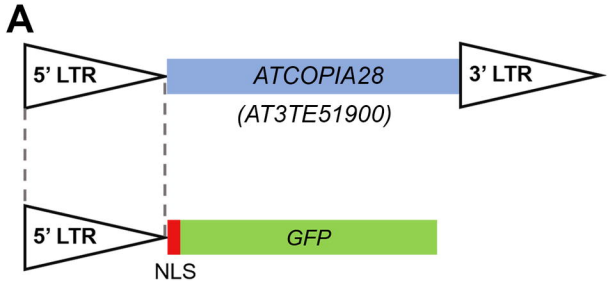
S11 Table. List of species used to construct the two trees of figure 8, their codes and the presence/absence of the different PMD-C and PP7 motifs.

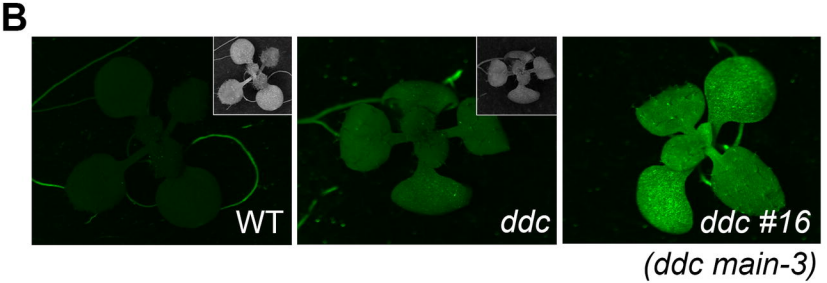
S12 Table. (A) PMD-C and (B) PP7/PP7L motifs used to construct the two phylogenetic trees of Figure 8.

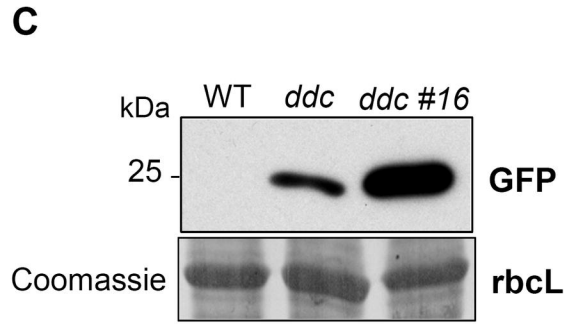
S13 Table. List of primers used in this study.

S14 Table. Next Generation Sequencing (NGS) mapping and coverage statistics.

Fig1


A
5' LTR
ATCOPIA28
(AT3TE51900)
3' LTR
5' LTR
GFP
NLS


B
WT
ddc
ddc #16
(ddc main-3)


C
kDa
WT
ddc
ddc #16
25
GFP
Coomassie
rbcL

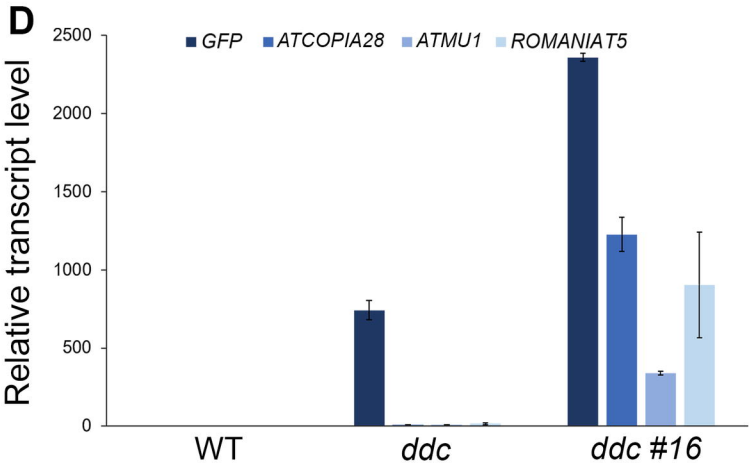
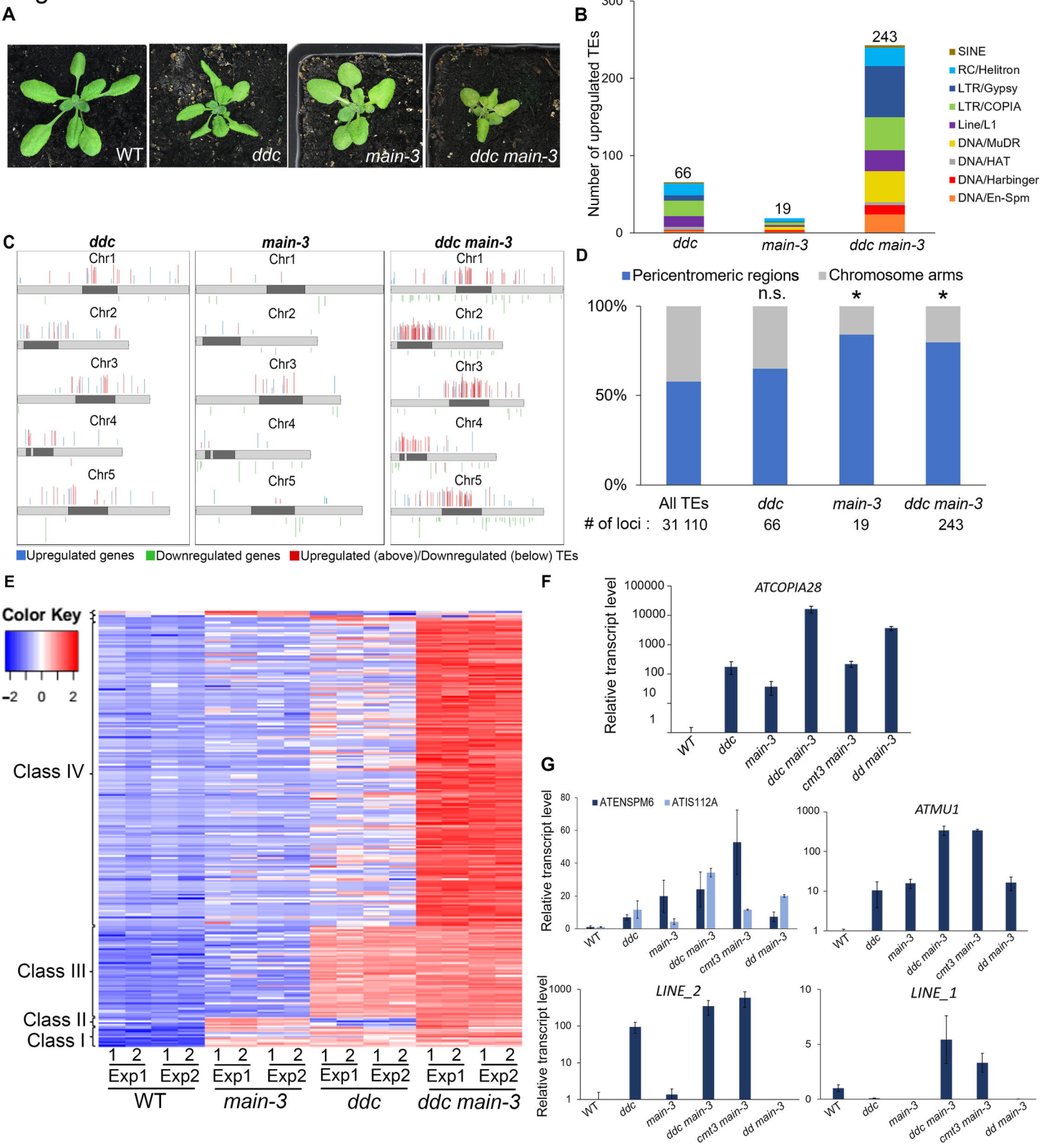

D
GFP
ATCOPIA28
ATMU1
ROMANIAT5
Relative transcript level
2500
2000
1500
1000
500
0
WT
ddc
ddc #16

Fig 2


A
WT
ddc
main-3
ddc main-3
B
Number of upregulated TEs
300
200
100
0
66
19
243
ddc
main-3
ddc main-3
SINE
RC/Helitron
LTR/Gypsy
LTR/COPIA
Line/L1
DNA/MuDR
DNA/HAT
DNA/Harbinger
DNA/En-Spm
C
ddc
main-3
ddc main-3
Chr1
Chr2
Chr3
Chr4
Chr5
Upregulated genes
Downregulated genes
Upregulated (above)/Downregulated (below) TEs
D
Pericentromeric regions
Chromosome arms
n.s.
*
*
100%
50%
0%
All TEs
ddc
main-3
ddc main-3
# of loci : 31 110
66
19
243
E
Color Key
−2 0 2
Class IV
Class III
Class II
Class I
1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
Exp1 Exp2 Exp1 Exp2 Exp1 Exp2 Exp1 Exp2
WT
main-3
ddc
ddc main-3
F
ATCOPIA28
Relative transcript level
100000
10000
1000
100
10
1
WT
ddc
main-3
ddc main-3
cmt3 main-3
dd main-3
G
ATENSPM6
ATIS112A
Relative transcript level
80
60
40
20
0
ATMU1
1000
100
10
1
LINE_2
LINE_1
10
5
0

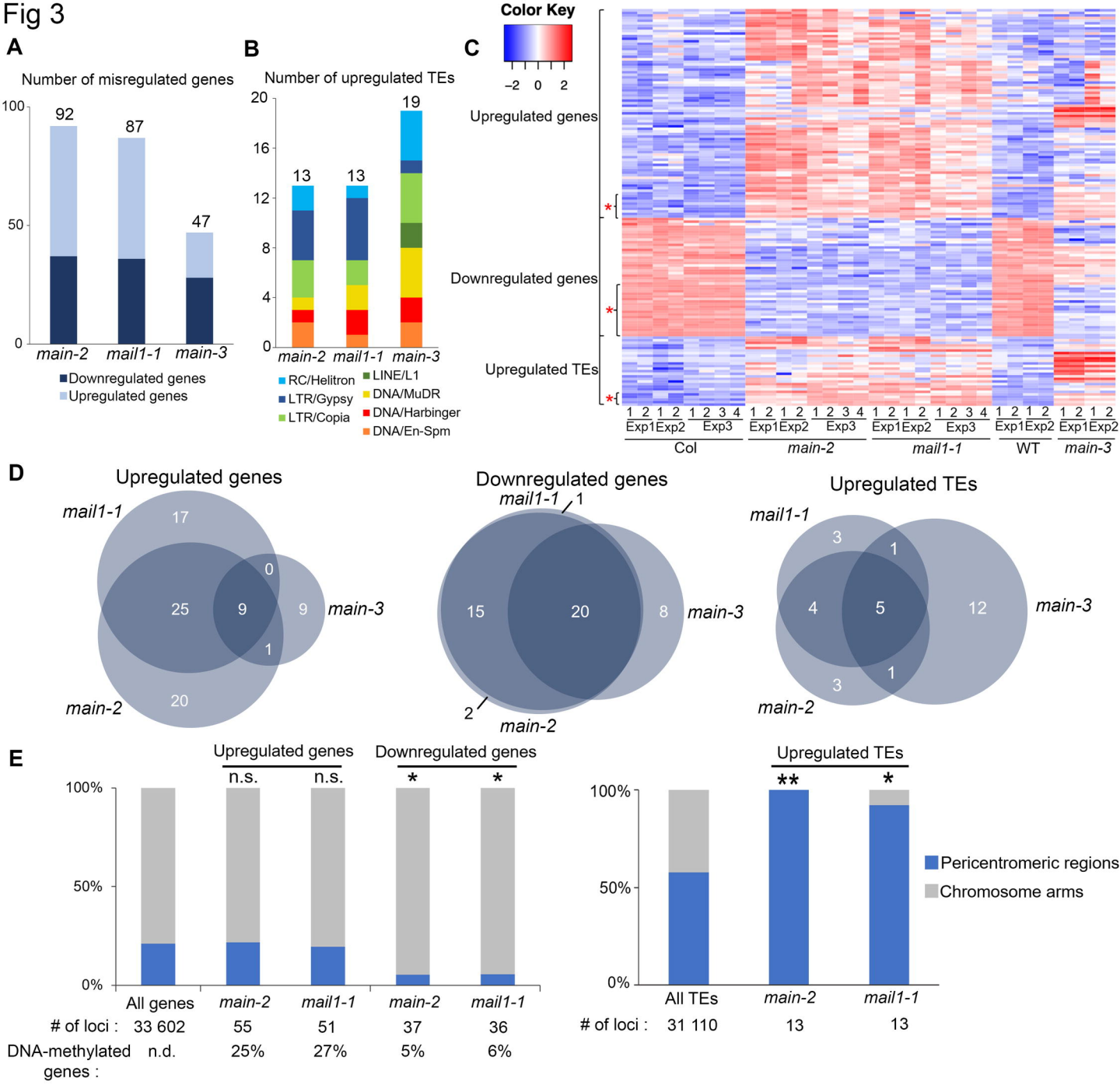


Fig 3
A
Number of misregulated genes
100
50
0
92
87
47
main-2
mail1-1
main-3
Downregulated genes
Upregulated genes
B
Number of upregulated TEs
20
16
12
8
4
0
13
13
19
main-2
mail1-1
main-3
RC/Helitron
LTR/Gypsy
LTR/Copia
LINE/L1
DNA/MuDR
DNA/Harbinger
DNA/En-Spm
C
Color Key
−2 0 2
Upregulated genes
Downregulated genes
Upregulated TEs
1 2 1 2 1 2 3 4 1 2 1 2 1 2 3 4 1 2 1 2 1 2 3 4 1 2 1 2 1 2 1 2
Exp1 Exp2 Exp3 Exp1 Exp2 Exp3 Exp1 Exp2 Exp3 Exp1 Exp2 Exp1 Exp2
Col
main-2
mail1-1
WT
main-3
D
Upregulated genes
mail1-1
17
0
25
9
9
main-3
1
main-2
20
Downregulated genes
mail1-1
1
15
20
8
main-3
2
main-2
Upregulated TEs
mail1-1
3
1
4
5
12
main-3
3
1
main-2
E
Upregulated genes
Downregulated genes
n.s.
n.s.
*
*
100%
50%
0%
All genes
main-2
mail1-1
main-2
mail1-1
# of loci :
33 602
55
51
37
36
DNA-methylated genes :
n.d.
25%
27%
5%
6%
Upregulated TEs
**
*
100%
50%
0%
All TEs
main-2
mail1-1
# of loci :
31 110
13
13
Pericentromeric regions
Chromosome arms

Fig 4

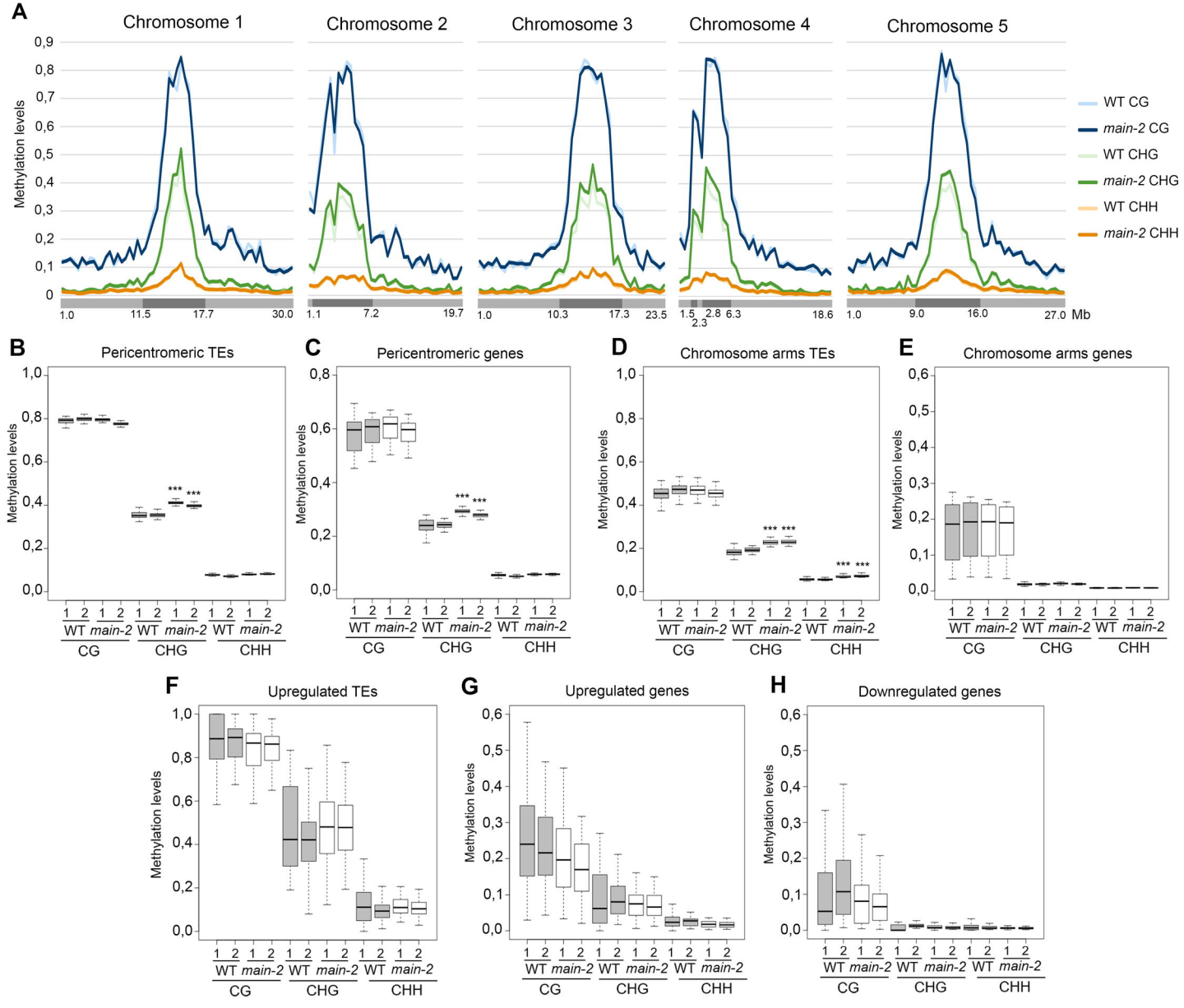
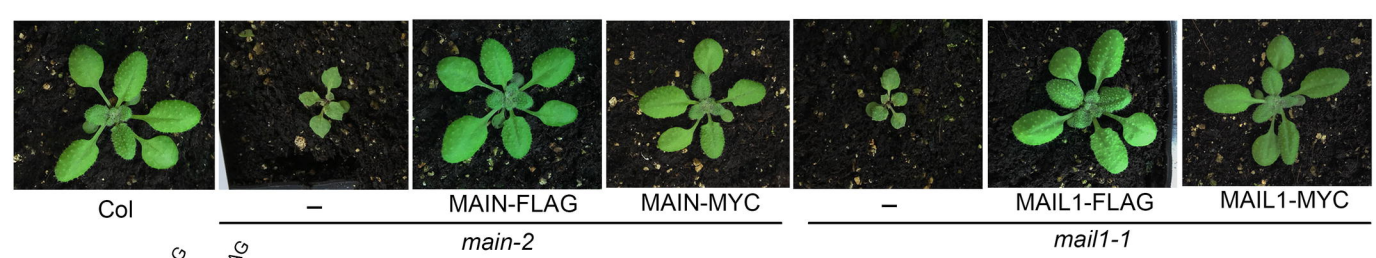
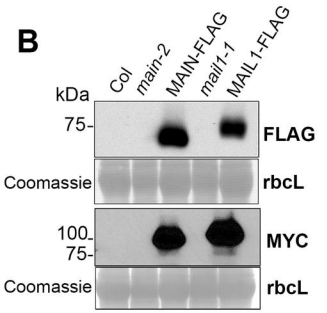
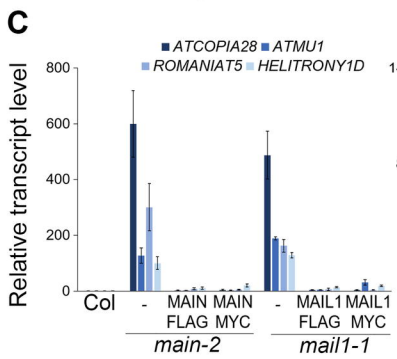
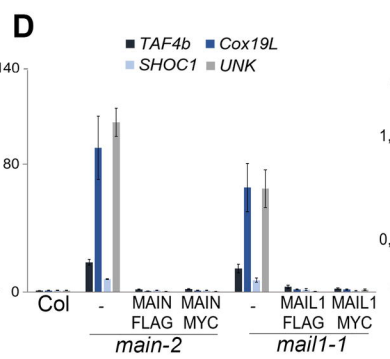

A
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Methylation levels
WT CG
main-2 CG
WT CHG
main-2 CHG
WT CHH
main-2 CHH
Mb
B
Pericentromeric TEs
C
Pericentromeric genes
D
Chromosome arms TEs
E
Chromosome arms genes
WT main-2
CG
CHG
CHH
F
Upregulated TEs
G
Upregulated genes
H
Downregulated genes

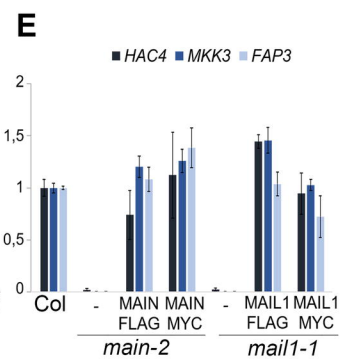
Fig 5





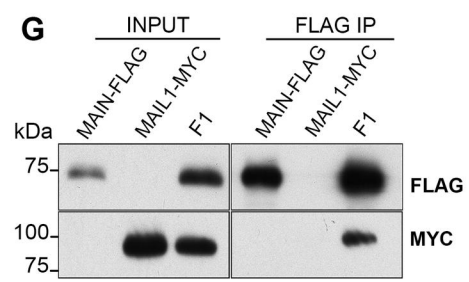


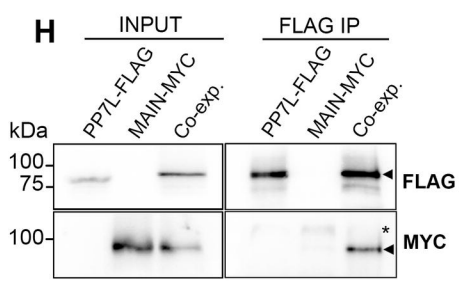




F

			Spectral count			Peptide count			NSAFe5		
			MAIN	MAIL1	PP7L	MAIN	MAIL1	PP7L	MAIN	MAIL1	PP7L
FLAG IP	Exp1	MAIN-FLAG	Rep. 1	33	39	35	19	20	22	794.0899	881.3135	670.9659
			Rep. 2	34	42	38	16	18	23	1298.8062	1506.6931	1156.4468
		MAIL1-FLAG	Rep. 1	20	25	14	13	14	12	905.1708	1062.5531	504.78354
		WT control	Rep. 1	0	0	0	0	0	0	0	0	0
			Rep. 2	0	0	0	0	0	0	0	0	0
	Exp2	MAIN-FLAG	Rep. 3	90	75	59	23	33	28	487.18513	381.26144	254.43695
		MAIL1-FLAG	Rep. 2	28	57	37	17	27	21	170.688	326.30966	179.68976
		WT control	Rep. 3	0	0	0	0	0	0	0	0	0





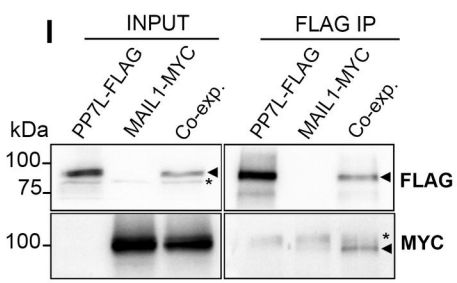


Fig 6

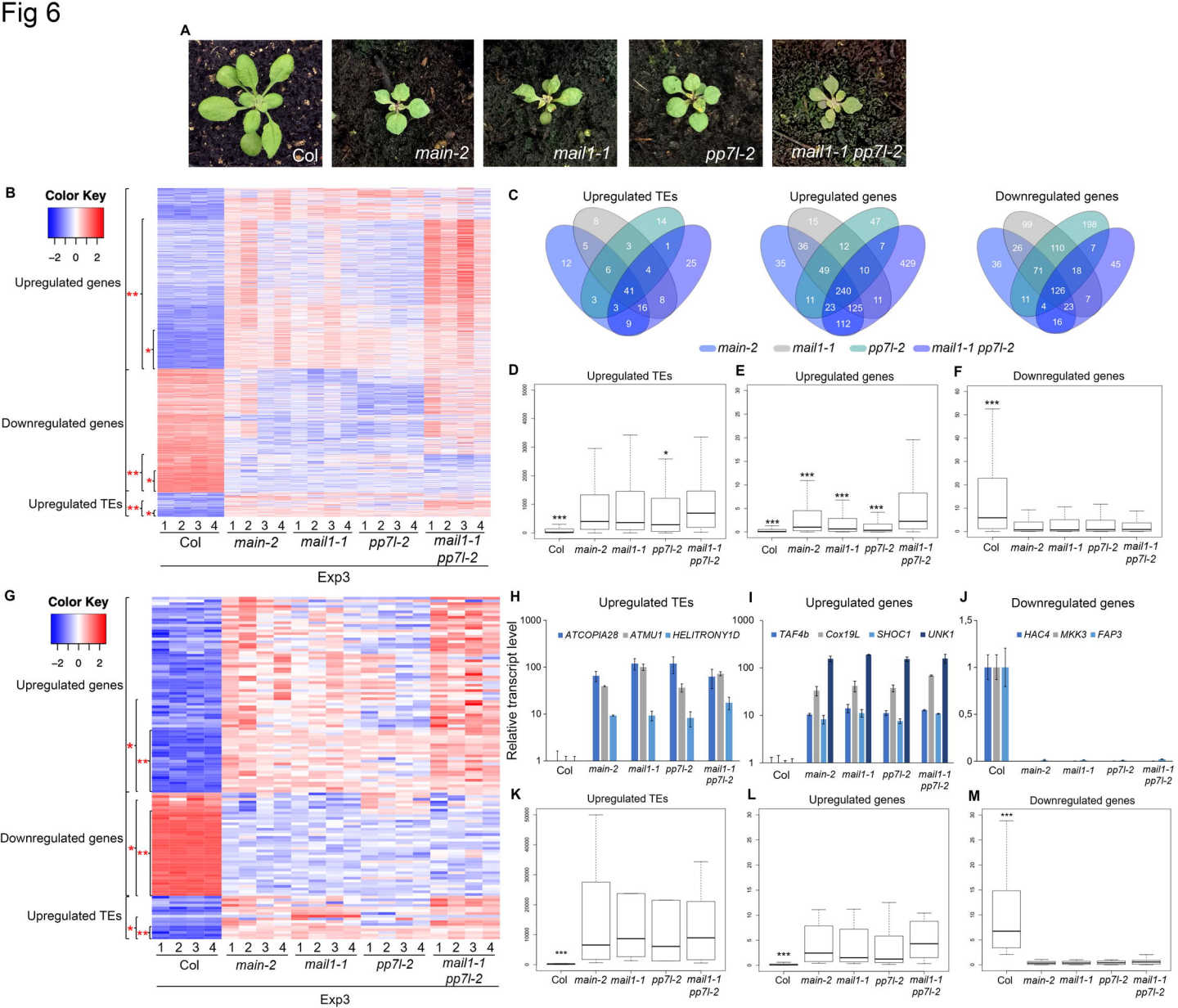

A
Col
main-2
mail1-1
pp7l-2
mail1-1 pp7l-2
B
Color Key
−2 0 2
Upregulated genes
Downregulated genes
Upregulated TEs
1 2 3 4
Col
main-2
mail1-1
pp7l-2
mail1-1 pp7l-2
Exp3
C
Upregulated TEs
Upregulated genes
Downregulated genes
main-2
mail1-1
pp7l-2
mail1-1 pp7l-2
D
Upregulated TEs
E
Upregulated genes
F
Downregulated genes
G
Color Key
−2 0 2
Upregulated genes
Downregulated genes
Upregulated TEs
1 2 3 4
Col
main-2
mail1-1
pp7l-2
mail1-1 pp7l-2
Exp3
H
Upregulated TEs
ATCOPIA28
ATMU1
HELITRONY1D
Relative transcript level
I
Upregulated genes
TAF4b
Cox19L
SHOC1
UNK1
J
Downregulated genes
HAC4
MKK3
FAP3
K
Upregulated TEs
L
Upregulated genes
M
Downregulated genes

Fig 7

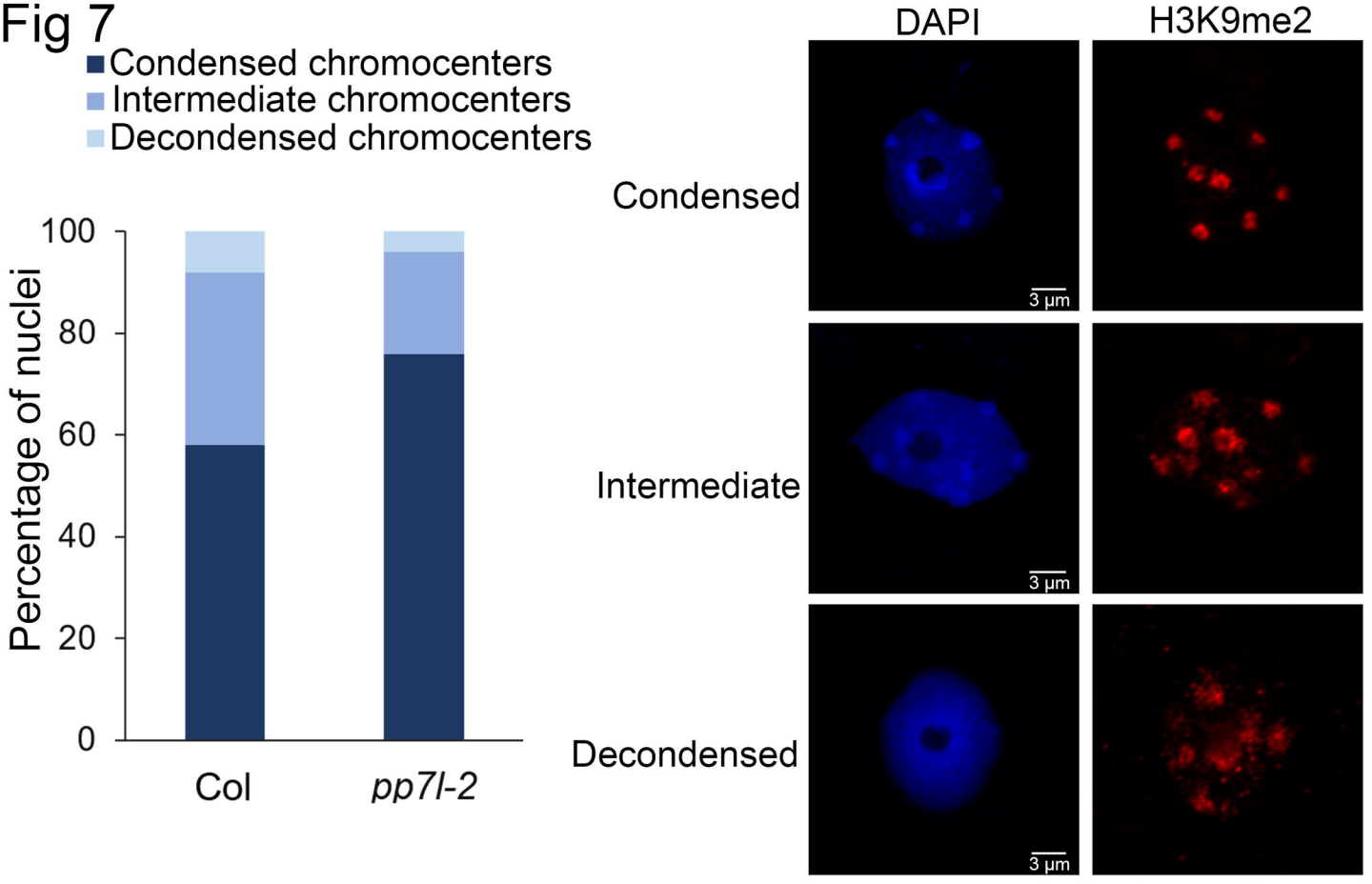
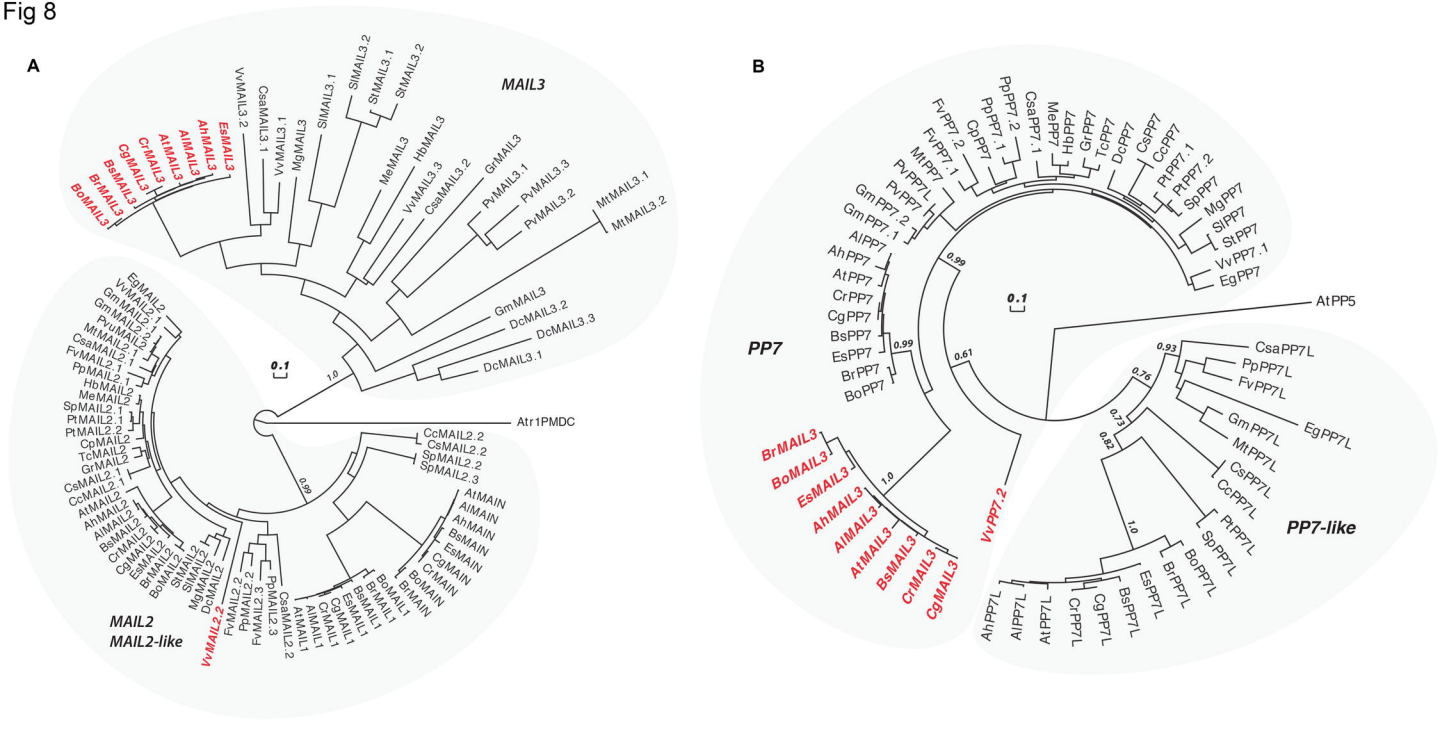

Condensed chromocenters
Intermediate chromocenters
Decondensed chromocenters
Percentage of nuclei
100
80
60
40
20
0
Col
pp7l-2
DAPI
H3K9me2
Condensed
Intermediate
Decondensed
3 µm
3 µm
3 µm

Fig 8


A
MAIL3
VvMAIL3.2
CsaMAIL3.1
VvMAIL3.1
MgMAIL3
SlMAIL3.1
SlMAIL3.2
SlMAIL3.3
MeMAIL3
HbMAIL3
VvMAIL3.3
CsaMAIL3.2
GrMAIL3
PvMAIL3.1
PvMAIL3.3
PvMAIL3.2
MtMAIL3.1
MtMAIL3.2
EsMAIL3
AhMAIL3
AlMAIL3
AtMAIL3
CrMAIL3
CgMAIL3
BsMAIL3
BrMAIL3
BoMAIL3
GmMAIL3
DcMAIL3.2
DcMAIL3.3
DcMAIL3.1
0.1
1.0
Atr1PMDC
MAIL2
MAIL2-like
EgMAIL2
VvMAIL2.1
GmMAIL2.1
PvMAIL2.1
MtMAIL2.1
CsaMAIL2.1
FvMAIL2.1
PpMAIL2
HbMAIL2
MeMAIL2
SpMAIL2.1
PtMAIL2.1
PtMAIL2.2
CpMAIL2
TcMAIL2
GrMAIL2
CsMAIL2.1
CcMAIL2.1
AtMAIL2
AhMAIL2
AlMAIL2
BsMAIL2
CrMAIL2
CgMAIL2
EsMAIL2
BrMAIL2
BoMAIL2
SlMAIL2.1
SlMAIL2.2
MgMAIL2
DcMAIL2
VvMAIL2.2
FvMAIL2.2
PpMAIL2.2
FvMAIL2.3
PpMAIL2.3
CsaMAIL2.2
CcMAIL2.2
CsMAIL2.2
SpMAIL2.2
SpMAIL2.3
0.99
AtMAIN
AlMAIN
AhMAIN
BsMAIN
EsMAIN
CgMAIN
CrMAIN
BrMAIN
BoMAIN
AtMAIL1
AlMAIL1
CrMAIL1
CgMAIL1
BsMAIL1
EsMAIL1
BrMAIL1
BoMAIL1
B
PP7
PpPP7.2
CsaPP7.1
MePP7
HbPP7
GrPP7
TcPP7
DcPP7
CsPP7
CcPP7
PtPP7.1
PtPP7.2
SpPP7
MgPP7
SlPP7
StPP7
VvPP7.1
EgPP7
FvPP7.2
CpPP7
PpPP7.1
MtPP7
PvPP7
PvPP7L
GmPP7.2
GmPP7.1
AlPP7
AhPP7
AtPP7
CrPP7
CgPP7
BsPP7
EsPP7
BrPP7
BoPP7
0.99
0.99
0.61
0.1
AtPP5
PP7-like
CsaPP7L
PpPP7L
FvPP7L
EgPP7L
GmPP7L
MtPP7L
CsPP7L
CcPP7L
PtPP7L
SpPP7L
BoPP7L
BrPP7L
EsPP7L
BsPP7L
CgPP7L
CrPP7L
AtPP7L
AlPP7L
AhPP7L
0.93
0.76
0.73
0.62
1.0
1.0
BrMAIL3
BoMAIL3
EsMAIL3
AhMAIL3
AlMAIL3
AtMAIL3
BsMAIL3
CrMAIL3
CgMAIL3
VvPP7.2