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* and *pp7l* 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 genes and 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 heavily DNA methylated, while euchromatin is more
57 relaxed and composed of genes that are more permissive to transcription, depending on the
58 recruitment of transcription factors (TFs), cofactors and RNA polymerases [1, 4]. In plants, DNA
59 methylation occurs in three different cytosine contexts: CG, CHG and CHH (where H = A, T or C),
60 involving specialized DNA methyltransferases [5]. In *Arabidopsis thaliana,* DOMAINS REARRANGED
61 METHYLTRANSFERASE 1 (DRM1) and DRM2 mediate de novo DNA methylation in all sequence
62 contexts through the RNA-directed DNA methylation (RdDM) pathway, which involves among other
63 components, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and DICER-LIKE 3 (DCL3) for the
64 production of short interfering (si)RNAs [6, 7]. The maintenance of CG methylation is specifically
65 performed by METHYLTRANSFERASE 1 (MET1), while CHROMOMETHYLASE 2 (CMT2) and CMT3 are
66 involved in the maintenance at CHG sites [8, 9]. CMT2 can also be involved in the deposition of CHH
67 methylation at specific genomic location [10, 11]. Finally, DRM1/2 are mostly required for the
68 maintenance of CHH methylation through the RdDM pathway [6, 7, 9]. Together with DNA
69 methylation, additional pathways play important roles in TE silencing. MORPHEUS MOLECULE 1
70 (MOM1) is required for TE silencing [12]. MICRORCHIDIA 1 (MORC1) and MORC6 interact together,
71 and are required for heterochromatin condensation and repression of TEs, acting mostly downstream
72 of DNA methylation and RdDM pathway [13-15]. It has been proposed that the MORC1/6 complex

73 would interact with the RdDM machinery to promote its own recruitment to heterochromatic loci that
74 must remain condensed and silenced [16].

75 More recently, the *A. thaliana* plant mobile domain (PMD) proteins MAINTENANCE OF
76 MERISTEM (MAIN) and MAIN-LIKE 1 (MAIL1) were identified as new factors required for TE silencing
77 [17]. In addition, these two proteins have been involved in genome stability, and regulation of
78 developmental processes such as cell division and differentiation [18, 19]. The PMD is a large protein
79 domain of unknown function that is widely represented among the angiosperms, predominantly
80 associated with TEs [17, 20]. It has been proposed that genic PMD versions, such as the MAIN and
81 MAIL1 proteins derived from TEs after gene domestication [17, 20, 21]. Previous studies suggested
82 that genic PMDs could act as cellular factors related to transcription, possibly acting as transcription
83 factor (TF)-like, co-factor or repressor proteins regulating this cellular process [18, 20]. Nevertheless,
84 the role of PMD proteins in the regulation of transcription remains elusive. Most of genic PMD
85 proteins are standalone versions, however, in some cases, the PMD is fused to another protein
86 domain, such as protease, kinase or metallo-phosphatase (MPP) domains. For instance in *A. thaliana*,
87 the MAIL3 protein carries a PMD, which is fused to a putative serine/threonine-specific
88 phosphoprotein phosphatase (PPP) domain phylogenetically related to the plant-specific protein
89 phosphatase 7 (PP7) [22]. PP7 is a calmodulin-binding PPP that has been related to cryptochrome
90 (CRY)-mediated blue-light signaling, and to the control of stomatal aperture [22-24]. PP7 is also
91 involved in the perception of red/far red light by controlling the phytochrome pathway [25, 26]. In
92 addition to PP7 and MAIL3 (also known as "long PP7"), the protein PP7-like (PP7L) belongs to the same
93 phylogenetic clade [22]. PP7L was recently identified as a nuclear protein involved in chloroplast
94 development and abiotic stress tolerance [27]. The *pp7l* mutant plants showed photosynthetic defects
95 and strong developmental phenotype associated with misregulation of several genes [27].

96 In this study, we describe a forward genetic screen based on a GFP reporter gene that allowed
97 us to identify a mutant population in which *MAIN* is mutated, leading to GFP overexpression. We then
98 decipher the genetic interaction between the DRM1/2, CMT3 and MAIN, showing that these proteins

99 act in different epigenetic pathways that act redundantly and synergistically to repress TEs.
100 Biochemical analyses indicate that MAIN and MAIL1 physically interact together. We have also
101 identified PP7L as a robust interactor of the two PMD proteins. In addition, the characterization of
102 developmental and molecular phenotypes of *pmd* and *pp7l* mutant plants strongly suggest that these
103 proteins interact together to silence TEs, and regulate the expression of a common set of genes.
104 Finally, phylogenetic analyses allowed us to determine the distribution of PMD, PP7 and PP7L domains
105 among the Eudicots. Based on these analyses, we have evidences of co-evolution linked to the neo-
106 association of the PMD and PP7-type PPP domains on single proteins in several Eudicot species,
107 suggesting a convergent evolution between these two protein domains.
108

109 ## RESULTS

110 **MAIN acts synergistically with DRM1/2 and CMT3 to repress TEs.**

111 The *ATCOPIA28* retrotransposon *AT3TE51900* (hereafter called *ATCOPIA28*) is targeted by
112 several epigenetic pathways such as DNA methylation and the MORC1/6 complex, which altogether
113 contribute to its repression. We engineered a construct in which the 5' long terminal repeat (LTR)
114 promoter region of *ATCOPIA28* controls GFP transcription (S1A Fig). While the *ATCOPIA28::GFP*
115 transgene is fully silenced in wild type (WT) plants, it is weakly expressed in the DNA methylation-
116 deficient *drm1 drm2 cmt3 (ddc)* triple mutant background (Fig 1A) [28]. We performed an ethyl
117 methane sulfonate (EMS) mutagenesis using the *ATCOPIA28::GFP ddc* plants as sensitized genetic
118 material, and screened for mutant populations showing GFP overexpression. Among, the selected
119 populations, we retrieved two new mutant alleles of *MORC6* carrying missense mutations in either
120 the GHKL or S5 domains of the protein (S1B-D Fig). We also identified the population *ddc #16* showing
121 strong overexpression of GFP and misregulation of several endogenous TEs, including *ATCOPIA28* (Fig
122 1A-C). Mapping experiments based on whole genome resequencing and bulk segregant analysis
123 indicated that *ddc #16* carries a missense point mutation (C_{230}→Y) in the gene *AT1G17930*, previously

124 named *MAIN* (S1C and S1E Fig). Genetic complementation analyses using the knock-out (KO)
125 transferred DNA (T-DNA) insertion line *main*-2 confirmed that *MAIN* was the mutated gene causing
126 the overexpression of GFP (S1F Fig). Therefore, *ddc #16* was renamed *ddc main-3.*

127 To determine the genetic interaction of *ddc* and *main-3* mutations on TE silencing, we carried
128 out RNA sequencing (RNA-seq) analyses in the hypomorphic *main-3* single, *ddc* triple and *ddc main-3*
129 quadruple mutant plants (S1G Fig). As previously described, the *ddc* mutant showed upregulation of
130 several TEs [11], which was also observed to a milder degree in the *main-3* mutant (Figs 1D and S1H,
131 and S1 Table). The *ddc main-3* mutant showed an exacerbation of TE silencing defects, with a large
132 number of pericentromeric TEs being specifically upregulated in this mutant background (Figs 1D-E
133 and S1H-I, and S1 Table). Reverse transcription quantitative PCR (RT-qPCR) analyses confirmed the
134 RNA-seq results, and showed that *ddc* and *main-3* mutations have synergistic effects on TE
135 upregulation (Fig 1F). These analyses reveal that DRM1/2, CMT3 and MAIN act redundantly and
136 synergistically in the silencing of several TEs, strongly suggesting that these proteins act in different
137 silencing pathways.

138

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

140 The *main-2* and *mail1-1* null mutants show similar developmental phenotypes, and
141 transcriptomic analyses have revealed that many TEs, including *ATCOPIA28*, are overexpressed in
142 these two mutants (S2A and B Fig, and S2 Table) [17]. In addition, several genes are misregulated in
143 the two *pmd* mutants (S2C and D Fig, and S2 Table). It has been suggested that MAIN and MAIL1 act
144 in the same pathway, possibly by interacting together [17]. To test this hypothesis, we generated
145 transgenic lines expressing FLAG- and MYC-tagged genomic PMD versions driven by their endogenous
146 promoters. We showed that epitope-tagged MAIN and MAIL1 proteins were produced at the expected
147 sizes, and they could complement the respective developmental phenotypes of null mutant plants
148 (S2A and S2E Fig). Importantly, the different epitope-tagged PMD proteins could also efficiently rescue
149 the TE silencing defects observed in the two *pmd* mutants (S2F Fig). Using FLAG-tagged MAIN and

150 MAIL1 expressing plants, immunoprecipitation followed by mass spectrometry (IP-MS) analyses were
151 carried out to determine potential protein interactors. Mass spectrometry (MS) analyses indicated
152 that MAIL1 was strongly immunoprecipitated with MAIN-FLAG and *vice versa* (Fig 2A). Co-
153 immunoprecipitation (co-IP) experiments using F1 transgenic line expressing MAIN-FLAG and MAIL1-
154 MYC transgenes validated IPMS results (Fig 2B). MS analyses of MAIN-FLAG and MAIL1-FLAG IP also
155 identified the metallo-phosphatase PP7L as putative interactor (Fig 2A). Co-IP experiments using
156 plants co-expressing either PP7L-FLAG together with MAIN-MYC, or PP7L-FLAG in combination with
157 MAIL1-MYC constructs confirmed the interaction between PP7L and the two PMD proteins (Fig 2C and
158 D). Thus, the three proteins MAIN, MAIL1 and PP7L physically interact together.
159
160 **The *main, mail1* and *pp7l* mutants display similar developmental and molecular phenotypes.**
161 PP7L is a putative metallo-phosphatase that is related to the phosphatase PP7, and it was
162 recently identified as a nuclear protein required for photosynthesis [22, 27]. Because PP7L interacts
163 with MAIN and MAIL1, and the *pp7l-2* null mutant displays abnormal developmental phenotype
164 reminiscent of the two *pmd* mutant plants (Figs 2 and S3A), we decided to reanalyze available *pp7l*
165 RNA-seq data [27]. Like in the two *pmd* mutants, several TEs were upregulated and hundreds of genes
166 were misregulated in *pp7l* (S3B and S3C Fig, and S2 Table). We then compared the transcriptomes of
167 *pmd* and *pp7l* mutants, and found that 56% and 60% of TEs upregulated in *main* and *mail1* mutants
168 were also upregulated in *pp7l* (Fig 3A). Similar analyses with misregulated genes revealed that a
169 fraction of genes that were down- or upregulated in *pp7l* were similarly misregulated in the two *pmd*
170 mutants (Fig 3B and C). Moreover, we found that 38% and 47% of genes downregulated in *main* and
171 *mail1* mutants were also repressed in *pp7l*, while only 6 % of genes upregulated in the *pmd* mutant
172 were similarly overexpressed in *pp7l* (Fig 3B and C). RT-qPCR analyses confirmed the upregulation of
173 several TEs in the *pp7l-2* null allele (Fig 3D). These analyses also validated that several genes were
174 similarly misregulated in *pp7l*, *main* and *mail1* mutants (Figs 3E-F and S2D). In addition, whole genome
175 bisulfite sequencing analyses performed in *main* showed that DNA methylation level is largely

176 unaffected in the *pmd* mutant in comparison to WT plants (S3D-E Fig). Finally, we performed in silico
177 analyses to identify any putative DNA cis regulatory elements (CREs) within a 1-kb promoter region
178 upstream of start codon of genes that were similarly down- or upregulated in the three mutant
179 backgrounds (Fig 3A-C). We found that a specific DNA decamer motif was significantly enriched in
180 twenty-one out of the twenty-four downregulated genes (Fig 3G and S3 Table). We also identified a
181 DNA motif in ten out of the eleven upregulated genes, although the E-value was not significant, most
182 likely because of the small sample size (S3F Fig and S3 Table). Remarkably, the DNA motif identified
183 among the dowregulated genes was also strongly enriched in the list of downregulated genes in the
184 hypomorphic *main-3* mutant (twenty-seven out of twenty-eight genes), and very similar motifs were
185 retrieved in the lists of downregulated genes in *ddc main-3*, *main-2* and *mail1-1* mutants (S3G Fig and
186 S3 Table). Thus, the MAIN/MAIL1 protein complex interacts with PP7L, and the three proteins are
187 required for the proper expression of a common set of loci. Moreover, most of the loci that are
188 commonly downregulated in the *main*, *mail1* and *pp7l* mutants display a specific DNA motif within
189 their promoters, suggesting that the transcriptional regulation of these loci could occur through the
190 recognition of a putative CRE.
191
192 **The PMD and PP7 domains have co-evolved among the Eudicots.**
193 Among the Angiosperms, most of the genic PMDs, like MAIN and MAIL1, are standalone
194 versions [20]. However, some genic PMDs can associate with other protein domains, such as for
195 instance a PPP domain. In *A. thaliana*, the protein MAIL3, which carries a PMD fused to a PPP domain,
196 is a close homolog of both MAIN/MAIL1 and PP7/PP7L through its PMD and PPP domains, respectively.
197 Considering that the PMD proteins MAIN and MAIL1 interact with PP7L, and are required for the
198 expression of similar set of loci, we decided to determine the distribution of related genic PMD and
199 PPP domains, and to retrace their evolutionary history among plant species. The *A. thaliana* MAIN,
200 MAIL1 and MAIL3 are all members of the PMD-C family that also includes MAIL2 [17]. Since our
201 objective is to retrace the evolution of genic (and not TE-containing) PMD-C, we have decided to

202 restraint our search to Eudicots. Indeed, Eudicot species contain mainly genic PMD-C, while other
203 angiosperms may contain variable numbers of closely related genic and TE-associated PMD-C motifs
204 that would be difficult to distinguish in our analysis. To retrace the evolution history of the genic PMD-
205 C family, we used *A. thaliana* PMD-C genes to search and collect their relatives (paralogues and
206 orthologues) in 30 genomes representative of the Eudicot diversity (see S4 Table for a list of species
207 and their corresponding codes used in Fig 4, and S5 Table for motif sequences).

208 In our phylogenetic analysis, the genic PMD-C family can be clearly separated in two major
209 clades. The first clade is composed of orthologues of *A. thaliana* MAIL2, MAIL1 and MAIN, while the
210 second one includes orthologues of *A. thaliana* MAIL3 (Fig 4A). Notably, a MAIL2 gene was found in
211 all species tested. MAIL2 genes form a closely related group suggesting that they are under strong
212 purifying selection (see the very short branch lengths linking most MAIL2 genes in Fig 4A). Additional
213 MAIL2 paralogues were also detected in several species. Some MAIL2 paralogues are imbedded in the
214 major MAIL2 group while others are more divergent and form independent subgroups. This is the case
215 of MAIL1 and MAIN that are clearly Brassicaceae-specific MAIL2 paralogues. MAIL1 and MAIN diverge
216 significantly from Brassicaceae MAIL2 following a gene duplication event that likely occurred in the
217 common ancestor of Brassicaceae, before falling under strong purifying selection, a situation often
218 encountered in neofunctionalization events. The MAIL3 distribution is irregular compared to MAIL2,
219 as a MAIL3 orthologue was not found in all Eudicot species tested. Accordingly, except in Brassicaceae,
220 *MAIL3* genes appear to be under much weaker purifying selection compare to *MAIL2* and *MAIL2-like*
221 genes (see the longer branch lengths in the tree of Fig 4A). Brassicaceae *MAIL3* genes contrast with
222 other *MAIL3*, by forming a closely related group in the phylogenetic tree. This suggests a clear change
223 in selection pressure, typical of a neofunctionalization event, this time correlated with the acquisition
224 of the PPP motif by these genes (Fig 4B and see below). Remarkably, another fusion event between a
225 PMD-C and a PPP motif occurred independently in grapevine, but this time involving a MAIL2
226 paralogue (VvMAIL2.2, Fig 4A).

227 We then used the PP7/PP7L-related motif found in *A. thaliana* MAIL3, to collect orthologous

228 genes and retrace the evolution history of PP7/PP7L in the same Eudicot species used above. We
229 confirmed by this analysis that these PPP genes can be clearly separated in two distinct clades: PP7
230 and PP7-like (PP7L) (Fig 4B). All tested species present one or several closely related PP7 paralogues.
231 Although Brassicaceae MAIL3 PPP motifs belong to the PP7 clade, it diverged significantly compared
232 to Brassicaceae or any other standalone PP7 paralogues. This suggests a fast-evolving period for
233 Brassicaceae MAIL3 PP7 motifs subsequently to their fusion with the PMD-C motif. This is in complete
234 agreement with the apparent change of selection regime that occurred for the MAIL3 PMD motif
235 following its fusion with PP7 (Fig 4A). It is interesting to note that in grapevine, the PPP motif found
236 on the same protein as the MAIL2 PMD-C is also of the PP7 type, and is quite divergent compared to
237 the other grapevine standalone PP7 and MAIL2 domains (VvMAIL2.2 and VvPP7.2 in Fig 4A and B,
238 respectively). Finally, PP7L orthologues were not found in all species tested and, accordingly, these
239 genes appear to be under weaker purifying selection compare to genes belonging to the PP7
240 subfamily. However, this was less obvious among the Brassicaceae clade, where the PP7L domain
241 appears highly conserved (see the branch lengths in the tree of Fig 4B). In conclusion, phylogenetic
242 analyses showed that, in at least two distinct events, neo-association of PMD-C and PP7 domains have
243 potentially create a new protein function that was maintained through evolution.
244

245 **DISCUSSION**

246 In *A. thaliana,* MAIN and MAIL1 are standalone PMD proteins that have been involved in
247 genome integrity, regulation of cell division and differentiation, and silencing of TEs [17-19]. In this
248 study, we showed that TE silencing is widely impaired in the *ddc main* higher order mutant, which is
249 both partially defective in DNA methylation and MAIN activity. We also identified the putative
250 phosphatase protein PP7L as a MAIN and MAIL1 protein interactor, and showed that most of the genes
251 that are commonly misregulated in *main*, *mail1* and *pp7l* mutants share identical CRE motifs. This
252 suggests that transcription at these loci might be regulated through similar mechanisms involving
253 specific CRE recognition. Finally, phylogenetic analyses among Eudicots suggest mechanism of

254 neofunctionalization between the PMD and PP7-type PPP, to potentially acquire a functional module
255 that requires the two protein domains.
256
257 **The PMD MAIN protein acts independently of DRM1/2- and CMT3-mediated DNA methylation**
258 **pathways to silence TEs.**
259 Previous analyses showed that some TEs were synergistically upregulated in the *mail1 rdr2*
260 double mutant plants, suggesting that MAIL1 acts independently of RdDM pathway [17]. In our whole
261 genome transcriptomic analyses, we showed that several TEs are upregulated in both *main-3* and *ddc*
262 mutants, and their upregulation is dramatically exacerbated in the *ddc main-3* quadruple mutant (Fig
263 1 E and F, red color code). We also identified TEs that are only upregulated in *ddc* or *main-3* mutants,
264 but display stronger misregulation in the *ddc main-3* higher order mutant (Fig 1 E and F, blue and
265 yellow color codes). Finally, we identified a large number of TEs that are only overexpressed in *ddc*
266 *main-3* (Fig 1 E and F, green color code). Thus, the PMD MAIN, DRM1/2 and CMT3 pathways act
267 synergistically at a subset of TEs (Fig 1 E, blue, red and yellow color codes), but also operate
268 redundantly to efficiently silence a much larger set of TEs that are predominantly pericentromeric (Fig
269 1 E, green color code and S1H Fig). Previous work showed that DNA methylation is not impaired in
270 *mail1* [17]. Likewise, we found that DNA methylation is largely unaffected in the *main* null mutant
271 (S3E Fig). Considering the strong synergistic effect of *main* and *ddc* mutations on TE derepression, we
272 can exclude the hypothesis that MAIN act downstream of DRM1/2- or CMT3-mediated DNA
273 methylation (no epistasis). Instead, it is very likely that MAIN, DRM1/2 and CMT3 pathways act
274 independently to efficiently silence TEs. Notably, synergistic effects between different epigenetic
275 pathways have already been observed. For instance, it has been shown that MORC1/6 and MOM1
276 pathways, or RdDM and MOM1 pathways act synergistically and redundantly to efficiently silence TEs
277 [14, 29]. Altogether, these observations contribute to the “mille-feuille” (i.e. “multiple layers”) model,
278 in which different epigenetic pathways converge towards the silencing of TEs [30].
279

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

282 Recently, PP7L was involved in the biogenesis of chloroplasts and plant response upon abiotic
283 stress [27]. The *pp7l* mutant shows dramatic developmental phenotype, which correlates with
284 misregulation of gene expression and upregulation of TEs (S3A-C Fig) [27]. Using IPMS and co-IP
285 experiments, we first showed that MAIN and MAIL1 interact together, which was already suggested
286 based on transcriptomic analyses of *main mail1* double mutant plants (Fig 2A and B) [17]. In addition,
287 we identified PP7L, which is a putative PPP, as a robust interactor of MAIN and MAIL1, suggesting that
288 the three proteins form heteromers (Fig 2A, C and D).

289 Comparison of upregulated TEs and misregulated genes in *main*, *mail1* and *pp7l* mutants revealed
290 that a significant proportion of loci that are misregulated in the two *pmd* mutants are also
291 misregulated in *pp7l* (Fig 3A-F, S2D and S2F). This suggests that the three proteins act, at least partially,
292 in the same pathway, which is consistent with the fact that they interact together.

293

294 **Genes that are commonly downregulated in *main, mail1* and *pp7l* carry an identical DNA motif in**
295 **their promoter regions.**

296 In silico analyses identified a specific DNA motif (5'-CATGCCGTTT-3') among the promoter of genes
297 that are commonly downregulated in *main*, *mail1* and *pp7l* mutants (Fig 3 G). this DNA motif was also
298 strongly enriched in most of the downregulated genes in the hypomorphic *main-3* mutant, and in
299 approximately half of the downregulated genes in *main-2* and *mail1-1* null mutant (S3G Fig). Based on
300 these results, we hypothesize that the MAIN/MAIL1/PP7L protein complex would be involved in
301 cellular processes that positively regulate transcription through the recognition of the putative CRE
302 motif 5'-CATGCCGTTT-3'. Consistently with this hypothesis, the highly similar DNA pentamer 5'-
303 TGACGT-3' was previously described as a G-box-like DNA element specifically recognized by
304 basic/leucine zipper (bZIP) TFs [31]. Therefore, it is likely that either a bZIP-like TF is recruited through
305 the MAIN/MAIL1/PP7L pathway, or the PMDs/PP7L complex directly recognizes the DNA motif to

306 promote transcription. In the future, it will be important to test if the MAIN/MAIL1/PP7L protein
307 complex interact directly with chromatin through the specific recognition of CRE motif. It will also be
308 essential to determine if this protein complex is required for the activation of specific TFs that bind
309 the CRE motif to regulate transcription.
310
311 **The association of PMD-C and PP7/PP7L domains creates a functional protein module.**
312 In this study, we identified PP7L has a protein partner of the two standalone PMDs MAIN and
313 MAIL1, and showed that *main*, *mail1* and *pp7l* display similar mutant phenotypes (Figs 2, 3 and S3). In
314 addition, phylogenetic analyses suggest that the association in trans of Brassicaceae PMD
315 MAIN/MAIL1 with PP7L would have created a selection pressure to optimize the MAIN/MAIL1/PP7L
316 interaction and function within the plant species. Indeed, both Brassicaceae MAIN/MAIL1 and PP7L
317 are clearly divergent compared to their closest orthologues and present much shorter branch lengths,
318 suggesting that a new correlated selection regime resulted from their cooperative functions in this
319 lineage (Fig 4). Besides, we showed that PMD and PP7 domains are fused in several Brassicaceae
320 MAIL3 and in the grapevine VvMAIL2.2 proteins (Fig 4). We propose a model in which the MAIL3-
321 specific PMD/PP7 fusion found in several Brassicaceae would have occurred in the common ancestors
322 of this Eudicot lineage, leading to a transient fast-evolving period for both motifs (possibly under a
323 positive selection regime). Since then, the MAIL3 PMD/PP7 fusion was maintained under strong
324 purifying selection, arguing for a neofunctionalization of the fusion protein. It is likely that similar
325 process happened in grapevine, and possibly, in closely related Vitaceae species. To some extent, the
326 two distinct events that occurred in Brassicaceae and grapevine are reminiscent of convergent
327 evolution processes leading to the production of a functional PMD/PP7 module, in which the two
328 protein domains would interact in cis. Although the Brassicaceae MAIL3 and grapevine VvMAIL2.2
329 proteins are probably involved in distinct biological processes, we hypothesize that, like the
330 MAIN/MAIL1/PP7L complex, they may interact with chromatin at specific genomic locations to
331 regulate gene expression.

332

333 In conclusion, we show here that the two *A. thaliana* PMD MAIN and MAIL1 proteins interact
334 with PP7L, and are involved in the regulation of a common set of genes and TEs. This association in
335 trans would have created a selection pressure on PP7L to optimize its interaction with MAIN and
336 MAIL1. In the future, it will be important to test if the MAIN/MAIL1/PP7L protein complex interact
337 with chromatin. In addition, we showed that distinct events of PMD-C and PP7 fusions occurred
338 among the Eudicots (among several Brassicaceae species and in grapevine), suggesting some
339 convergent evolution processes and a potential neofunctionalization of PMD/PP7 module in cis. In the
340 future, it will be essential to determine the role of these PMD/PP7 fusion proteins, such as the *A.*
341 *thaliana* MAIL3. In addition, it will be important to decipher the function of PMD proteins in plant
342 species with agronomic value.

343

344 MATERIALS AND METHODS

345

346 **Plant material and growing conditions.** Wild-type (WT) and all mutant lines are in the Columbia (Col)
347 ecotype. The *drm1-2* (SALK_031705), *drm2-2* (SALK_150863) *cmt3-11* (SALK_148381) *ddc* triple
348 mutant line was previously described [28]. The *main-2* (GK-728H05), *mail1-1* (GK-840E05) and *pp7l-2*
349 (SALK_003071) were previously described [17-19, 27], and obtained from The Nottingham Arabidopsis
350 Stock Centre. T-DNA insertions were confirmed by PCR-based genotyping. The *main-3* mutation was
351 genotyped by derived cleaved amplified polymorphic sequences (dCAPS) using the restriction enzyme
352 FokI. Primer sequences are described in S6 Table. Plants were grown on soil under a 16h-light/8h-dark
353 cycle or *in vitro* under continuous light and transferred on soil after 10 days. For *in vitro* plant culture,
354 seeds were surface-sterilized and sowed on solid Murashige and Skoog medium containing 0.5%
355 sucrose (w/v).

356

357 **Cloning of ATCOPIA28::GFP.** The pCambia3300-NLS-GFP-T35S vector was previously described [13].
358 The 5'LTR promoter corresponding to a region of ~1 kb upstream of *ATCOPIA28 (AT3TE51900)* was
359 PCR amplified from WT genomic DNA, and cloned into pCR2.1 TOPO vector (Invitrogen). Quikchange
360 site-directed mutagenesis (Stratagene) was performed according to Manufacturer's instruction to
361 create a polymorphism site (MfeI→NdeI) within the 5'LTR promoter, which was subsequently
362 mobilized into pCambia3300 upstream of NLS-GFP-T35S sequence. *ddc* triple mutant plants were
363 transformed with the AT*COPIA28::GFP* construct using the *Agrobacterium*-mediated floral dip
364 method [32]. Transgenic plants showing GFP fluorescence were backcrossed with a WT plant to
365 promote the silencing of *ATCOPIA28::GFP* in the F_1 generation. F_1 plants were self-crossed and their
366 F_2 progenies were screened for GFP fluorescence, and PCR-based genotyped to obtain
367 *ATCOPIA28::GFP* WT and *ATCOPIA28::GFP ddc* plants. Primer sequences used for *ATCOPIA28::GFP*
368 cloning are described in S6 Table.
369
370 **EMS mutagenesis, GFP screening and mapping analyses.** Five thousand seeds of *ATCOPIA28::GFP ddc*
371 were mutagenized in 0.26% EMS solution for 12 hours with rotation. Seeds were subsequently washed
372 with water and sown on soil. Fifteen hundred M2 populations were collected, and subsequently
373 screened for GFP fluorescence under UV light using a SMZ18 Nikon Fluorescence Stereomicroscope
374 coupled with the C-HGFI intensilight fluorescence filter. Pictures were taken using the DS Qi1MC digital
375 camera kit. Mapping and identification of the EMS mutation responsible for the phenotype were
376 performed by bulk segregant analysis coupled with deep genome re-sequencing as previously
377 described [13], with the following differences. Reads were mapped against the reference genome
378 (*Arabidopsis* TAIR10) and single nucleotide polymorphisms called in Geneious (Biomatters). Using R,
379 single nucleotide polymorphisms were filtered for EMS mutations (G:C→A:T) and zygosity called
380 based on the variant frequency provided by Geneious (≥80% homozygous mutation, ≥45%, and ≤55%
381 heterozygous mutation). Plots were then created by calculating the ratio of the number of
382 homozygous and heterozygous and mutations in a 500-kb window as previously described [33].

383

384 **Cloning of epitope-tagged versions of PMD proteins.** *MAIN* and *MAIL1* genomic regions were PCR
385 amplified and FLAG or Myc epitopes were added to the C-terminus of each protein as previously
386 described [13]. In both cases, the amplified region includes a ~1Kb promoter sequence upstream of
387 the respective transcriptional start site. For the *MAIN* promoter, a MluI site was modified to allow LR
388 reaction without changing the sequence integrity of the gene. Primer sequences are described in S6
389 Table.

390

391 **IP and MS analysis.** Ten grams of 3-week-old seedling tissue were ground in liquid nitrogen and
392 resuspended in 50mL ice-cold IP buffer [50mM Tris HCl pH 7.6, 150mM NaCl, 5mM $MgCl_2$, 0.1%
393 Nonidet P-40, 10% glycerol (v/v), 0.5mM DTT, 1x Protease Inhibitor Mixture (Roche)] and centrifuged
394 2 times for 15 min at 4°C at 15 350g. 400µL of M2 magnetic FLAG-beads (Sigma, M8823) were added
395 to the supernatants, and incubated for 90min rotating at 4°C. M2 magnetic FLAG-beads were washed
396 seven times in ice-cold IP buffer for 5 min rotating at 4°C, and immunoprecipitated proteins were
397 eluted 3 times with 150µL 3x-FLAG peptides (Sigma, F4799) for 25 min each at 25°C. The eluted protein
398 complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously
399 described [14].

400

401 **Co-IP and immunoblotting.** 0.5 g of 3-week-old seedling tissue were ground in liquid nitrogen,
402 resuspended in 1.5mL ice-cold IP buffer [50mM Tris pH 7.6, 150mM NaCl, 5mM MgCl2, 0.1% Nonidet
403 P-40, 10% glycerol, 0.5 mM DTT, 1x Protease Inhibitor Mixture (Roche)], and centrifuged 2 times for
404 15 min at 4°C, 16 000g. 50µL M2 magnetic FLAG-beads (Sigma, M8823) were added to the
405 supernatants and incubated for 2 hour rotating at 4°C. Beads were washed 3 times in ice-cold IP buffer
406 for 10 min rotating at 4°C. Immunoprecipitated proteins were denatured in Laemmli buffer for 5min
407 at 95°C. 10µL of input and bead elution were run on 10% SDS-PAGE gels, and proteins were detected
408 by western blotting using either Anti-FLAG M2 monoclonal antibody-peroxidase conjugate (Sigma,

409 A8592) at a dilution of 1 :10 000, or c-Myc rat monoclonal antibody (Chromotek, 9E1-100) at a dilution
410 of 1 : 1000 followed by goat anti-rat IgG horseradish peroxidase (Abcam, ab205720) used at a dilution
411 of 1 :20000 as secondary antibody. Western blots were developed using Substrat HRP Immobilon
412 Western (Merck Millipore, WBKLS0500).
413
414 **RNA extraction.** Total RNA was extracted from aerial parts of 3-week-old seedlings using either
415 RNeasy Plant Mini Kit (Qiagen, 74904) or Monarch Total RNA Miniprep Kit (NEB, T2010) according to
416 the manufacturer's protocols.
417
418 **RNA sequencing.** RNA-seq libraries were generated from 1µg of input RNA using NEBNext Ultra II
419 Directional RNA Library Prep Kit for Illumina (NEB, E7490) according to the manufacturer's protocols.
420 Libraries were sequenced on an Illumina HiSeq 4000 or NextSeq 550 machines. Reads were trimmed
421 using Trimmomatic [34], and mapped to the *A. thaliana* genome (*Arabidopsis* TAIR10 genome) using
422 HISAT2 [35]. The sequence alignment files were sorted by name and indexed using SAMtools [36].
423 Files were converted to BAM files and number of reads mapped onto a gene calculated using HTSeq-
424 count [37]. Differentially expressed genes were obtained with DESeq2 [38], using a log2 fold-change
425 > 2 (up-regulated genes) or < -2 (down-regulated genes) with an adjusted *p*-value of 0,01. Re-analyses
426 of previously published RNA-seq datasets from *main-2* and *mail1-1* (PRJEB15202) [17], and *pp7l*
427 mutants (GSE122495) [27] were performed as described above.
428
429 **RT-qPCR.** 1 µg of input RNA was converted to cDNA using GoScript Reverse Transcriptase (Promega
430 A501C) according to the manufacturer's protocol. The final reaction was diluted 6 times with RNase
431 free water. RT-qPCR experiments were performed with 4µL of cDNA combined to the Takyon No Rox
432 SYBR MasterMix (Eurogentec, UF-NSMT-B0701), using a LightCycler 480 instrument (Roche).
433 Amplification conditions were as follows: 95°C 5 min; 45 cycles, 95°C 15s, 60°C 15s, 72°C 30s; melting

434 curves. Three technical replicates were performed per biological replicate, and 3 biological replicates
435 were used in all experiments.
436
437 **Cis-elements detection.** The motifs for enhancer sequences (1kb upstream the TSS) were discovered
438 using MEME (Multiple Em for Motif Elicitation). MEME represents motifs as position-dependent
439 letter-probability matrices which describe the probability of each possible letter at each position in
440 the pattern [39].
441
442 **Bisulfite sequencing.** Genomic DNA was extracted from aerial parts of 3-week-old seedlings using
443 Quick-DNA Plant/Seed Miniprep Kit (Zymo research, D6020) according to the manufacturer's protocol.
444 Whole genome bisulfite sequencing (WGBS) library was prepared from 50 ng genomic DNA using
445 NuGen Ovation Ultralow Methyl-Seq kit. Bisulfite treatment was carried out by Qiagen Epitect bisulfite
446 kit. WGBS libraries were sequenced on an Illumina HiSeq 4000 machine. The raw reads (single end)
447 were trimmed using Trimmomatic in order to remove adapter sequences [34]. The remaining
448 sequences were aligned against the *A. thaliana* genome TAIR10 version using Bismark [40]. Duplicated
449 reads were collapsed into one read. For visualization of the data, we used ViewBS [41].
450
451 **Sequence selection, multiple sequences alignments and phylogenetic reconstruction.**
452 Blast searches (blastp) were performed starting from known *A. thaliana* PMD-C and PP7/PP7L motifs
453 on the thirty species representing the diversity of the Eudicot lineages. When necessary tblastn
454 searches were also used to obtain complete protein sequences. To build the phylogenetic trees, PMD-
455 C or PP7/PP7L motifs were aligned using the multiple sequence comparison by log-expectation
456 (MUSCLE v3.7) software [42]. Trees were reconstructed using the fast-maximum likelihood tree
457 estimation program PHYML [43] using the LG amino acids replacement matrix [44]. Statistical support
458 for the major clusters were obtained using the approximate likelihood-ratio test (aLRT) [45].
459

460 **Data availability.** Nucleotide sequencing data generated in this study have been deposited in
461 European Nucleotide Archive (ENA) under the accession number PRJEB33240
462 (http://www.ebi.ac.uk/ena/data/view/PRJEB33240). The proteomics data have been deposited to the
463 MassIVE data repository (https://massive.ucsd.edu) with the dataset identifier MSV000084089. All
464 other data and material are available within the manuscript and its supplementary files, or from the
465 corresponding author upon request.
466

467 ## ACKNOWLEDGMENTS

468 The authors want to thank Thierry Lagrange, Frederic Pontvianne and other team members for fruitful
469 discussions, and all the LGDP platform members for their outstanding technical assistance and plant
470 care. This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-
471 10-LABX-41). S.E.J. is an investigator of the Howard Hughes Medical Institute.
472

473 ## REFERENCES
474

475 1. Grewal SI, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8(1):35-46.
476 2. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the
477 genome. Nat Rev Genet. 2007;8(4):272-85.
478 3. Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications.
479 Nat Rev Genet. 2019;20(7):417-31.
480 4. Dergai O, Hernandez N. How to Recruit the Correct RNA Polymerase? Lessons from snRNA
481 Genes. Trends Genet. 2019;35(6):457-69.
482 5. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in
483 plants and animals. Nat Rev Genet. 2010;11(3):204-20.
484 6. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of
485 increasing complexity. Nat Rev Genet. 2014;15(6):394-408.
486 7. Wendte JM, Pikaard CS. The RNAs of RNA-directed DNA methylation. Biochim Biophys Acta.
487 2016.
488 8. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with
489 histone methylation. Nat Rev Mol Cell Biol. 2015;16(9):519-32.
490 9. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. Nat Rev Mol
491 Cell Biol. 2018;19(8):489-506.
492 10. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The
493 Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing
494 heterochromatin. Cell. 2013;153(1):193-205.
495 11. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape
496 the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21(1):64-72.

12. Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J. Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature. 2000;405(6783):203-6.
13. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. Science. 2012;336(6087):1448-51.
14. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, Yen L, et al. Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proc Natl Acad Sci U S A. 2014;111(20):7474-9.
15. Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. Curr Biol. 2012;22(10):933-8.
16. Gallego-Bartolome J, Liu W, Kuo PH, Feng S, Ghoshal B, Gardiner J, et al. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. Cell. 2019;176(5):1068-82 e19.
17. Ikeda Y, Pelissier T, Bourguet P, Becker C, Pouch-Pelissier MN, Pogorelcnik R, et al. Arabidopsis proteins with a transposon-related domain act in gene silencing. Nat Commun. 2017;8:15122.
18. Wenig U, Meyer S, Stadler R, Fischer S, Werner D, Lauter A, et al. Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana. Plant J. 2013;75(3):469-83.
19. Uhlken C, Horvath B, Stadler R, Sauer N, Weingartner M. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. Plant J. 2014;78(1):107-20.
20. Babu MM, Iyer LM, Balaji S, Aravind L. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. Nucleic Acids Res. 2006;34(22):6505-20.
21. Steinbauerova V, Neumann P, Novak P, Macas J. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. Genetica. 2011;139(11-12):1543-55.
22. Farkas I, Dombradi V, Miskei M, Szabados L, Koncz C. Arabidopsis PPP family of serine/threonine phosphatases. Trends Plant Sci. 2007;12(4):169-76.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. Rigal M, Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. Biochim Biophys Acta. 2011;1809(8):452-8.
31. Schindler U, Beckmann H, Cashmore AR. TGA1 and G-box binding factors: two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. Plant Cell. 1992;4(10):1309-19.
32. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 1998;16(6):735-43.

548 33. Hristova E, Fal K, Klemme L, Windels D, Bucher E. HISTONE DEACETYLASE6 Controls Gene
549 Expression Patterning and DNA Methylation-Independent Euchromatic Silencing. Plant Physiol.
550 2015;168(4):1298-308.
551 34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
552 Bioinformatics. 2014;30(15):2114-20.
553 35. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
554 requirements. Nat Methods. 2015;12(4):357-60.
555 36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
556 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
557 37. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
558 sequencing data. Bioinformatics. 2015;31(2):166-9.
559 38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
560 seq data with DESeq2. Genome Biol. 2014;15(12):550.
561 39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for
562 motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-8.
563 40. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
564 applications. Bioinformatics. 2011;27(11):1571-2.
565 41. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for
566 visualization of high-throughput bisulfite sequencing data. Bioinformatics. 2018;34(4):708-9.
567 42. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
568 Nucleic Acids Res. 2004;32(5):1792-7.
569 43. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies
570 by maximum likelihood. Syst Biol. 2003;52(5):696-704.
571 44. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol.
572 2008;25(7):1307-20.
573 45. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate,
574 and powerful alternative. Syst Biol. 2006;55(4):539-52.

575 ## FIGURE CAPTIONS

576 **Fig 1. MAIN is required for the silencing of TEs, acting redundantly and synergistically with the**
577 **DRM1/2 and CMT3 pathways.**

578 (A) WT and *drm1 drm2 cmt3* (*ddc*) triple mutant plants carrying the *ATCOPIA28::GFP* transgene
579 showed no and weak GFP fluorescence under UV light, respectively. By comparison, the *ddc #16* EMS
580 mutant showed strong GFP fluorescence. Insets show plants under white light. (B) Western blot using
581 anti-GFP antibody confirmed *ATCOPIA28::GFP* overexpression in *ddc #16*. Coomassie staining of the
582 large Rubisco subunit (rbcL) is used as a loading control. (C) Relative expression analyses of
583 *ATCOPIA28::GFP (GFP)* and three endogenous TEs in *ddc* and *ddc #16 over WT* assayed by Real-Time
584 quantitative PCR (RT-qPCR), and normalized to the house keeping gene *RHIP1*. Error bars indicate
585 standard deviation based on three independent biological replicates. (D) Number of upregulated TEs

586 in *ddc*, *main-3* and *ddc main-3*, and classified by TE superfamily. Threshold: log2>2, p-adj< 0,01.
587 Analyses are based on four biological replicates. (E) Venn diagram analyses showing the overlaps
588 between upregulated TEs in *ddc*, *main-3* and *ddc main-3*. (F) Relative expression analyses of several
589 TEs in *ddc*, *main-3* and *ddc main-3* over WT assayed by RT-qPCR and normalized to *RHIP1*. Error bars
590 indicate standard deviation based on three independent biological replicates. The same color code
591 was used for the Venn diagram (E) and the frames of RT-qPCR graphs (F).

592

593 **Fig 2. MAIN and MAIL1 physically interacts together and with PP7L.**

594 (A) FLAG-tagged MAIN and MAIL1 proteins were immunoprecipitated and putative interacting
595 proteins were identified by mass spectrometry. The total number of identified spectra is shown for
596 three MAIN and two MAIL1 biological replicates. WT replicates are used as a negative control. (B)
597 MAIL1-MYC was co-immunoprecipitated with MAIN-FLAG in F1 plants co-expressing the two
598 transgenes. Plants expressing only MAIL1-MYC or MAIN-FLAG were used as negative controls. (C)
599 Plants expressing the MAIN-MYC transgene were "supertransformed" with the PP7L-FLAG construct,
600 and MAIN-MYC was co-immunoprecipitated with PP7L-FLAG. Plants expressing only MAIN-MYC or
601 PP7L-FLAG were used as negative controls. (D) Same as C but using MAIL1-MYC plants
602 "supertransformed" with the PP7L-FLAG construct. Epitope-tagged proteins were detected by
603 Western blotting. Arrowheads indicates expected bands. Asterisks indicates non-specific
604 hybridization.

605

606 **Fig 3. The PMD proteins and PP7L regulates a same set of loci.**

607 (A-C) Venn diagrams showing the overlaps between upregulated TEs (A), downregulated genes (B) and
608 upregulated genes (C) in *main-2*, *mail1-1* and *pp7l-1* based on analyses described in S2 and S3 Figs.
609 (D-E) Relative expression analyses assayed by RT-qPCR of upregulated TEs (D), downregulated genes

610 and (E) upregulated genes in the *main-2* and *pp7l-2* mutants over WT. Transcript level is normalized
611 to *RHIP1*. Error bars indicate standard deviation based on three independent biological replicates. (G)
612 Overrepresentation of a specific DNA motif in the 1-kb promoter regions of genes that are commonly
613 downregulated in *main*, *mail1* and *pp7l* using the MEME software. Under the logo are represented
614 the ratio of genes carrying the motif among the list of downregulated loci, and the significant E-value
615 associated with this motif.

616

617 **Fig 4. Evolutionary history of PMD-C and PP7 proteins in plants.**

618 (A) An alignment of the PMD-C motifs from 30 representative Eudicot species was used to construct
619 a phylogenetic tree. The two major clades (MAIL2/MAIL2-like and MAIL3) are indicated. The species
620 codes are given in S4 Table, and corresponding protein sequences in S5 Table). In red are genes
621 presenting a fusion between a PMD-C and a PP7 motif. Statistical supports of key nodes calculated
622 with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The
623 tree was rooted using the *Amborella trichopoda* PMD-C motif (Atr1PMDC). (B) Phylogenetic tree
624 constructed using an alignment of the PP7 motif from the same species as in (A). The two major clades
625 (PP7 and PP7L) are indicated. In red are genes presenting a fusion between a PP7 and a PMD-C motif.
626 Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated.
627 Scale bar indicates one substitution/site. The tree was rooted using the *A. thaliana* PP5 motif (AtPP5).

628

629
630

631 # SUPPORTING INFORMATION CAPTIONS

632

633 **S1 Fig. *MAIN* is the mutated gene responsible for *ATCOPIA28::GFP* and TE overexpression in the *ddc***
634 ***#16* mutant.**

635 (A) Schematic representation of the *ATCOPIA28::GFP* construct. The 5' long terminal repeat (LTR)
636 promoter region of an ATCOPIA28 LTR-retrotransposon (*AT3TE51900*) is used to control the
637 expression of GFP. The construct also carries a Nuclear Localization Signal (NLS) to target the GFP in
638 the nucleus. (B) Representative pictures of WT, *ddc* and *ddc #18* (*ddc morc6-8)* and *ddc #344* (*ddc*
639 *morc6-9)* mutant plants under UV light. Insets show plants under white light. (C) Enrichment in
640 homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining
641 the linkage intervals for the populations *ddc #18*, *ddc #344* and *ddc #16* (*ddc main-3)*. (D) Location of
642 the point mutations corresponding to the *morc6-8 and morc6-9* alleles within the *MORC6* genomic
643 sequence. (E) Location of the point mutation corresponding to the *main-3* allele within the *MAIN*
644 genomic sequence. Nucleotide and corresponding amino acid changes are indicated above the gene.
645 Positions of the mutations are indicated relative to the transcription starting site (+1). Grey boxes
646 represent 5' and 3' UTR, blue boxes and lines represent exons and introns, respectively. (F) Genetic
647 complementation analysis using the KO T-DNA insertion line *main-2*. *ddc #16* plants were crossed with
648 *main-2* plants, F1 plants were self-crossed and F2 plants were screened under UV light to analyze GFP
649 fluorescence. Western blotting using anti-GFP antibodies confirmed GFP overexpression in selected
650 F2 plants. Coomassie staining of the large Rubisco subunit (rbcL) is used as a loading control. The
651 *DRM2* and *CMT3* genotype of selected F2 plants were subsequently analyzed by PCR using specific
652 primers. Ho: Homozygous. He: Heterozygous. All the F2 plants that were selected were homozygote
653 for the *drm2* mutation. (G) Representative pictures showing the developmental phenotype of 3-week-
654 old WT, *ddc*, *main-3* and *ddc main-3* plants. (H) Chromosomal distributions of misregulated loci in *ddc*,
655 *main-3* and *ddc main-3* over WT. Chromosome arms are represented in light grey, pericentromeric

656 regions in dark grey. Upregulated genes are represented in blue, downregulated genes in green and
657 TEs are represented in red above the chromosomes (upregulated) or below (downregulated). The size
658 of each bar is proportional to the misregulation of the loci. Threshold: log2>2, or log2<-2; p-adj< 0,01.
659 (I) Relative expression analyses of *ATCOPIA28* and *HELITRONY1D* (*AT5TE35950*) in *ddc*, *main-3* and *ddc*
660 *main-3* over WT assayed by RT-qPCR. Relative transcript level is normalized to *RHIP1*. Error bars
661 indicate standard deviation based on three independent biological replicates.

662

663 **S2 Fig. Epitope-tagged MAIN and MAIL1 constructs complement the developmental and molecular**
664 **phenotypes of *main-2* and *mail1-1* null mutants.**

665 (A) Representative pictures of 2-week-old WT, *main-2*, *mail1-1* and epitope-tagged complementing
666 mutant plants. (B) Number of upregulated TEs in *main-2* and *mail1-1*, classified by TE superfamily. (C)
667 Number of misregulated genes in *main-2* and *mail1-1* classified by upregulated and downregulated
668 genes. Threshold: log2>2, or log2<-2; p-adj< 0,01. These analyses are based on four biological
669 replicates, including the reanalysis of RNA-seq data published in reference [17]. (D) Relative
670 expression analyses of upregulated genes (above panel) and downregulated genes (below panel) in
671 *main-2* and *mail1-1* over WT assayed by RT-qPCR. Relative transcript level is normalized to *RHIP1*.
672 Error bars indicate standard deviation based on three independent biological replicates. (E) Western
673 blots using anti-FLAG and anti-MYC antibodies confirmed the accumulation of epitope-tagged PMD
674 proteins at the expected sizes in the different complementing lines. (F) Relative expression analyses
675 of several TEs in the different complementing lines over WT assayed by RT-qPCR and normalized to
676 *RHIP1*. The *main-2* and *mail1-1* mutants are used as negative controls. Error bars indicate standard
677 deviation based on three independent biological replicates.

678

679 **S3 Fig. Developmental and molecular phenotypes of *pp7l* mutant.**

680 (A) Representative pictures of 5-week-old *main-2*, *mail1-1* and *pp7l-2* mutant plants showing similar
681 abnormal developmental phenotypes in comparison to WT plant. (B) Number of upregulated TEs in
682 *pp7l-1* and classified by TE superfamily. (C) Number of misregulated genes in *pp7l-1* classified by
683 upregulated and downregulated genes. Threshold: log2>2, or log2<-2; p-adj< 0,01. These analyses are
684 based on reanalysis of RNA-seq data published in reference [27]. (D-E) Metaplot analyses showing the
685 average distribution of DNA methylation levels in upregulated TEs (D) and genes (E) in *main-2* versus
686 WT plants. Upstream and downstream regions are 2 kb long. (F) Overrepresentation of a specific DNA
687 motif in the 1-kb promoter regions of genes that are commonly upregulated in *main*, *mail1* and *pp7l*
688 using the MEME software. Under the logo are represented the ratio of genes carrying the motif among
689 the list of commonly upregulated loci. Note that the E-value associated with this motif is not
690 significant. (G) Same as (F) for the hypomorphic and null *pmd* mutants. Only DNA motifs with
691 significant E-values are shown.

692

693 **S1 Table. Lists of differentially regulated loci in *ddc*, *main-3* and *ddc main-3*.**

694

695 **S2 Table. Lists of differentially regulated loci in *main-2*, *mail1-1* and *pp7l-1*.**

696

697 **S3 Table. Lists of misregulated genes carrying a specific DNA motif within their promoter.**

698

699 **S4 Table. List of species used to construct the two trees of Fig 4, their codes and the**
700 **presence/absence of the different PMD-C and PP7motifs are indicated.**

701

702 **S5 Table. (A) PMD-C and (B) PP7/PP7L motifs used to construct the two phylogenetic trees of Fig 4.**
703 **Sequences were collected at the NCBI or JGI databases.**

704

705 **S6 Table. List of primers used in this study.**

706

707

708

Fig 1

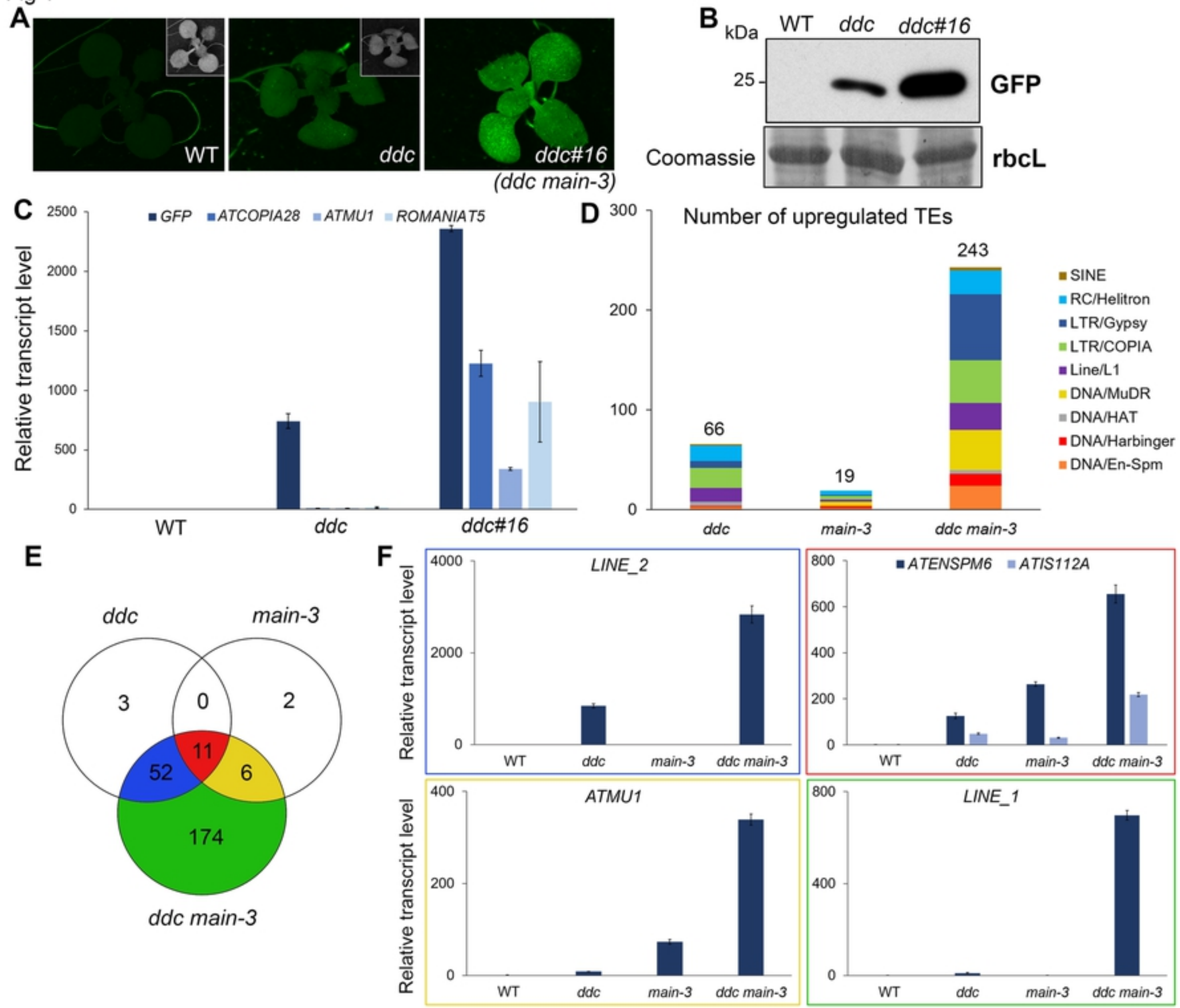

Figure 1

Fig 2

A

			Spectral Count		
			MAIN	MAIL1	PP7L
FLAG IP	MAIN-FLAG	rep.1	33	39	35
		rep.2	34	42	38
		rep.3	90	75	59
	MAIL1-FLAG	rep.1	20	25	14
		rep.2	28	57	37
	WT control	rep.1	0	0	0
		rep.2	0	0	0
		rep.3	0	0	0

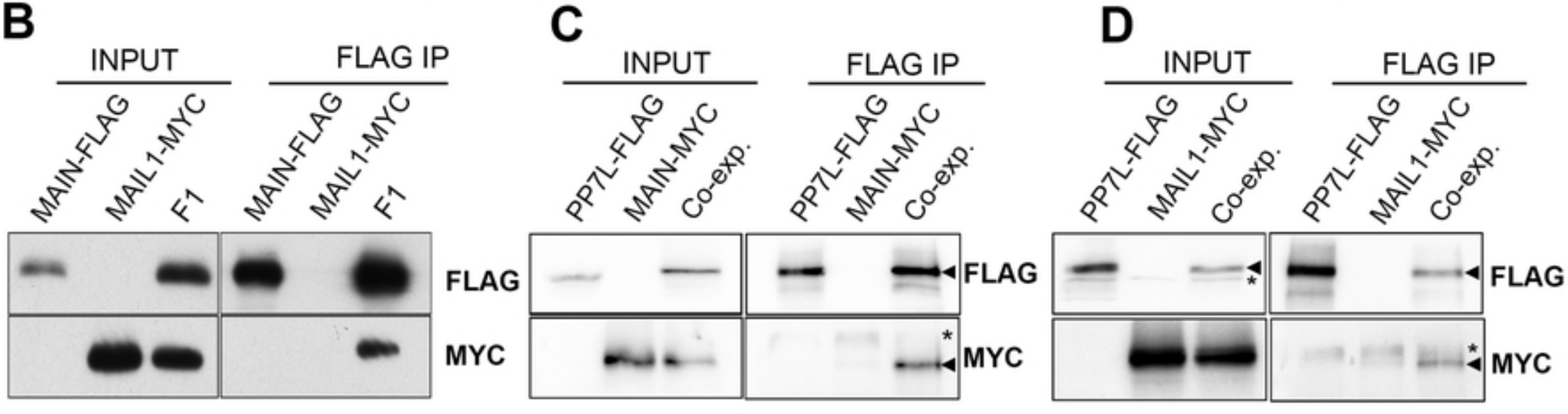

Figure 2

Fig 3

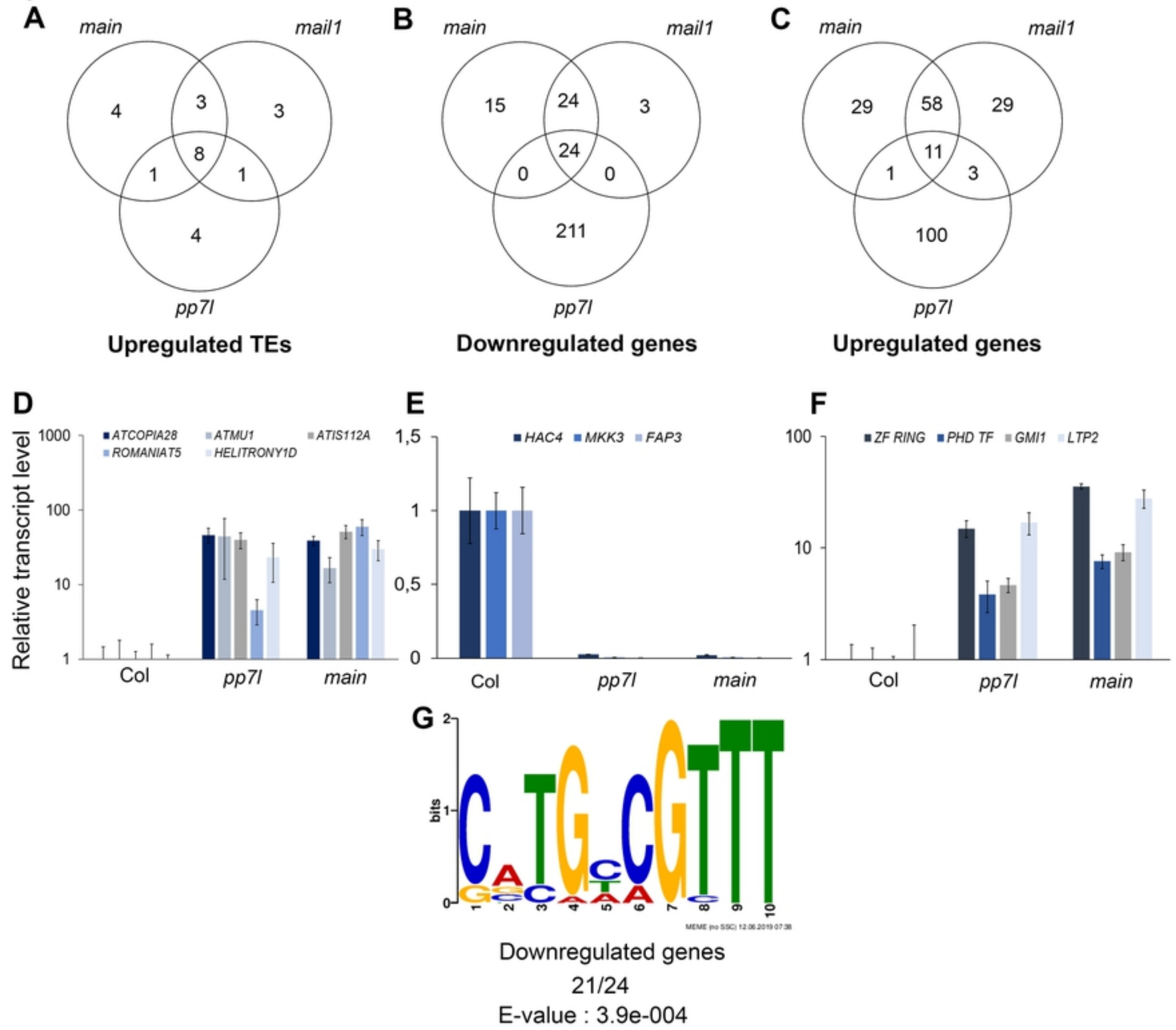

Figure 3

Fig 4

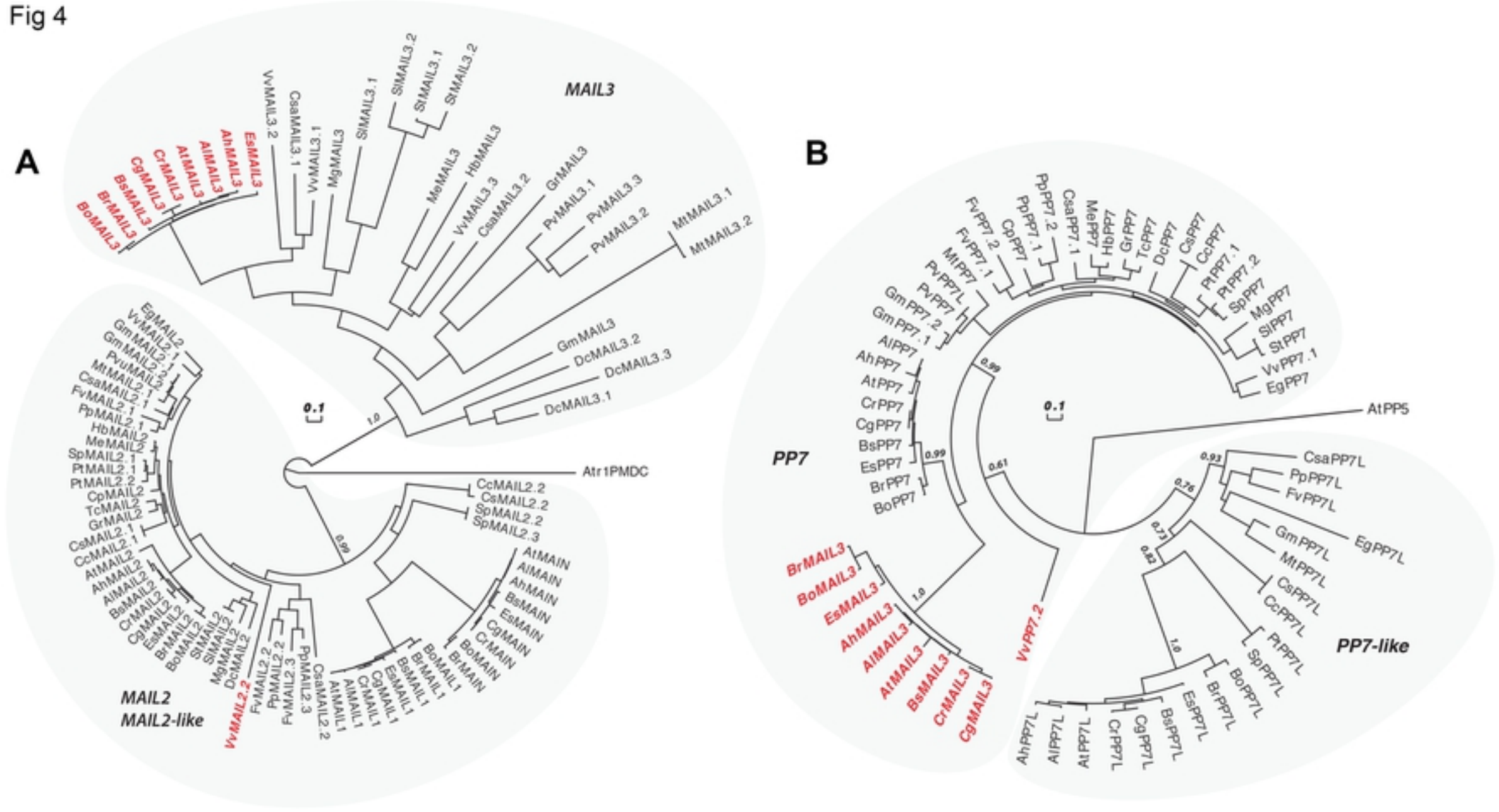

Figure 4