The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons

Rita A. Batista[1], Jordi Moreno-Romero[1,2], Yichun Qiu[1], Joram van Boven[1], Juan Santos-González[1], Duarte D. Figueiredo[1,3], Claudia Köhler[1*]

1. Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Centre for Plant Biology, Uppsala, Sweden
2. Present address: Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain
3. Present address: Institute for Biochemistry and Biology, University of Potsdam, Germany

* Corresponding author (claudia.kohler@slu.se)

Abstract

MADS-box transcription factors are ubiquitous in eukaryotic organisms and play major roles during plant development. Nevertheless, their function in seed development remains largely unknown. Here we show that the imprinted *Arabidopsis thaliana* MADS-box TF PHERES1 (PHE1) is a master regulator of paternally expressed imprinted genes, as well as of non-imprinted key regulators of endosperm development. PHE1 binding sites show distinct epigenetic modifications on maternal and paternal alleles, correlating with parental-specific transcriptional activity. Importantly, we show that the CArG-box-like DNA-binding motifs bound by PHE1 have been distributed by RC/Helitron transposable elements. Our data provide an example of molecular domestication of these elements, which by distributing PHE1 binding sites throughout the genome, have facilitated the recruitment of crucial endosperm regulators into a single transcriptional network.

Introduction

MADS-box transcription factors (TFs) are present in most eukaryotes, and classified into two groups: type I or SRF (Serum Response Factor)-like, and type II or MEF2 (Myocyte Enhancing Factor2)-like (Gramzow and Theissen, 2010). In flowering plants, type I MADS-box TFs are associated with reproductive development and many are active in the endosperm, a nutritive seed tissue that supports embryo growth (Bemer et al., 2010). Deregulation of type I MADS-box TF genes has been frequently linked to failure of endosperm development and seed inviability, both in the Brassicaceae (Erilova et al., 2009; Lu et al., 2012; Rebernig et al., 2015; Tiwari et al., 2010), and in crop species like tomato and rice (Ishikawa et al., 2011; Roth et al., 2019). In *Arabidopsis thaliana* and other species, the activity of type I MADS-box TFs has been shown to be associated with the timing of endosperm cellularisation – an important developmental transition, which is fundamental for seed viability (Erilova et al., 2009; Kang et al., 2008; Lu et al., 2012; Tiwari et al., 2010). Nevertheless, a mechanistic explanation for these observations, clarifying the role of MADS-box TFs in endosperm development, remains to be established.

In this work we characterise the function of the type I MADS-box TF PHE1. *PHE1* is active in the endosperm and is a paternally expressed imprinted gene (Köhler et al., 2005). Imprinting is defined as an epigenetic phenomenon causing a gene to be preferentially expressed from the maternal or the paternal allele. In plants, imprinting is mainly manifested in the endosperm, and similarly to type I MADS-box TFs, imprinted genes have been previously implicated in endosperm development (Erilova et al., 2009; Figueiredo et al., 2015; Jullien and Berger, 2010; Pignatta et al., 2018; Wolff et al., 2015). Imprinting relies on the asymmetric distribution of epigenetic marks in maternal and paternal alleles, which is thought to distinctly impact their transcription (Gehring, 2013; Rodrigues and Zilberman, 2015). However, transcriptional regulators of imprinted genes have not been identified until now. Thus, the interaction between epigenetic modifications of imprinted genes and their transcriptional regulation remains unknown. Our results identify PHE1 as a key transcriptional regulator of imprinted genes, as well as of other genes important for endosperm development. Furthermore, we explore the

cross-talk between epigenetic regulation and PHE1 transcriptional activity at imprinted gene loci, and show that differential parental epigenetic modifications of PHE1 binding sites correlates with the transcriptional status of the parental alleles. Finally, we reveal that RC/Helitron transposable elements (TEs) have served as distributors of PHE1 DNA-binding sites, providing an example of molecular domestication of TEs.

Results and Discussion

To identify the genes regulated by PHE1, we performed a ChIP-seq experiment using siliques from a PHE1::PHE1-GFP line, which has been previously shown to expressed PHE1-GFP exclusively in the endosperm (Weinhofer et al., 2010). We identified a total of 1942 PHE1 target genes, which were enriched for Gene Ontology (GO)-terms associated with development, metabolic processes, and transcriptional regulation, showing that many PHE1 targets are themselves transcriptional regulators (**Fig. 1 – Fig. Supplement 1a**). Among PHE1 targets are several known regulators of endosperm development like *AGL62, YUC10, IKU2, MINI3*, and *ZHOUPI*, revealing a central role for PHE1 in regulating this developmental process (**Fig. 1 – Fig. Supplement 2**). Importantly, type I MADS-box family genes were over-represented among PHE1 targets (**Fig. 1 – Fig. Supplement 1a**), pointing to a high degree of cross-regulation among members of this family.

Our data revealed that PHE1 uses two distinct DNA-binding motifs: motif A was present in about 53% of PHE1 binding sites, while motif B could be found in 43% of those (**Fig. 1a**). PHE1 motifs closely resemble type II CArG-boxes, the signature motif of MADS-box TFs (Folter and Angenent, 2006) (**Fig. 1a, Fig. 1 – Fig. Supplement 3**), suggesting that DNA-binding properties between type I and type II members are conserved.

Strikingly, PHE1 binding sites significantly overlapped with TEs, preferentially with those of the RC/Helitron superfamily (29% (723/2494) in PHE1 binding sites, versus 9% (224/2494) in random binding sites) (**Fig. 1b**), and in particular with some RC/Helitron subfamilies (**Fig. 1 – Fig. Supplement 4**). We thus addressed the question whether RC/Helitrons contain sequence properties that promote

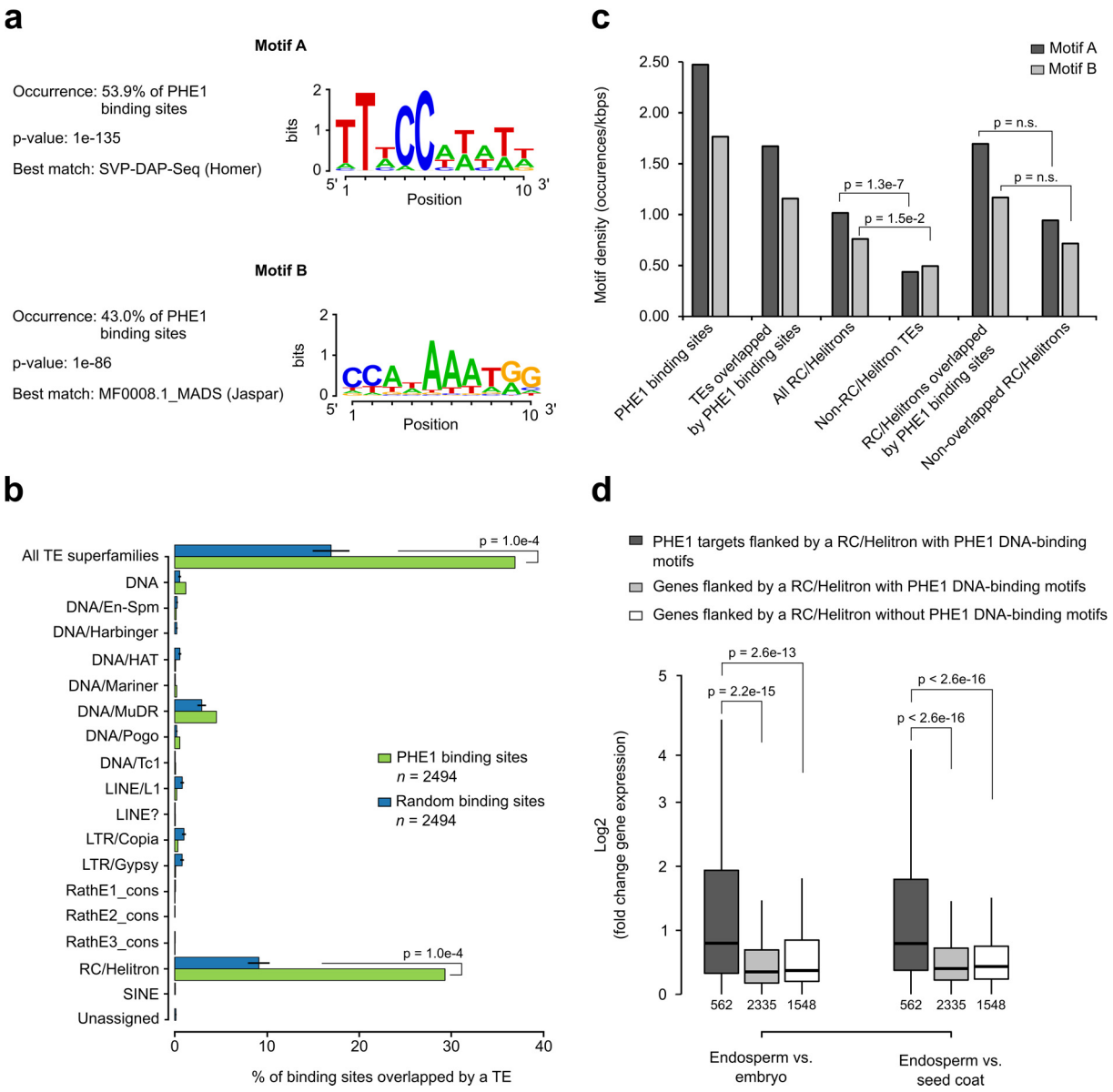


Figure 1 - RC/Helitrons carry PHE1 DNA-binding motifs. a, CArG-box like DNA-binding motifs identified from PHE1 ChIP-seq data. **b,** Fraction of PHE1 binding sites (green) overlapping transposable elements (TEs.) Overlap is expressed as the percentage of total binding sites where spatial intersection with features on the y-axis is observed. A set of random binding sites is used as control (blue). This control set was obtained by randomly shuffling the PHE1 binding sites in *A. thaliana* gene promoters (see Methods). P-values were determined using Monte Carlo permutation tests (see Methods). Bars represent ± sd, (n = 2494, PHE1 binding sites, Random binding sites). **c,** Density of PHE1 DNA-binding motifs in different genomic regions of interest. P-values were determined using χ2 tests. **d,** Fold change gene expression of genes flanked by RC/Helitron TEs. Fold change was determined between endosperm and embryo, or between endosperm and seed coat. Genes were divided in three categories depending on their PHE1 target status, and the presence of RC/Helitrons with and without PHE1 DNA-binding motifs. Gene expression data was retrieved from Belmonte et al., (2013). Pre-globular seed stage was used in this analysis. P-values were determined using two-tailed Mann-Whitney tests (*n* = number represented below boxplots).

PHE1 binding. Indeed, a screen of genomic regions for the presence of PHE1 DNA-binding motifs, revealed significantly higher motif densities within all RC/Helitrons, compared to other TE superfamilies (**Fig. 1c**). We found that most RC/Helitron sequences associated with both PHE1 DNA-binding motifs share sequence homology, and the majority of them could be grouped into one large cluster, based on sequence identity (**Fig. 1 – Figure Supplement 5)**. While the presence of additional smaller clusters points to a few instances of independent gains of PHE1 DNA-binding motifs through *de novo* mutation or sequence capture, the grouping of most sequences within one cluster suggests that a single ancestral RC/Helitron likely acquired a perfect or nearly perfect motif sequence. Moreover, we could detect the presence of perfect or nearly perfect PHE1 DNA-binding motifs within the consensus sequences of several RC/Helitron families (**Fig. 1 – Figure Supplement 6**), suggesting that radiation of RC/Helitrons occurred after the acquisition of the binding motif. Together, these data suggest that an ancestral RC/Helitron likely played a role in acquiring PHE1 DNA-binding motifs, which were subsequently amplified in the genome through transposition.

Interestingly, even though motif densities were higher in RC/Helitrons overlapped by PHE1 binding sites than in non-overlapped RC/Helitrons, this difference was not significant (**Fig. 1c**). As the enrichment of PHE1 DNA-binding motifs is a specific feature of RC/Helitrons, the domestication of these TEs as *cis*-regulatory regions might facilitate TF binding and modulation of gene expression. In line with this, we detected that genes flanked by RC/Helitrons carrying bound PHE1 DNA-binding motifs were more highly expressed in the endosperm than in other seed tissues (**Fig. 1d**). This shows that for a subset of genes, these TEs can be effectively used as sites for PHE1 binding, thus triggering endosperm-specific expression.

Among the PHE1 target genes we detected a significant enrichment of imprinted genes, with 12% of all maternally expressed genes (MEGs) and 31% of all paternally expressed genes (PEGs) being targeted (**Fig. 2a**). Imprinted gene expression relies on parental-specific epigenetic modifications, which are asymmetrically established during male and female gametogenesis, and inherited in the endosperm (Gehring, 2013; Rodrigues and Zilberman, 2015). Demethylation of repeat sequences and

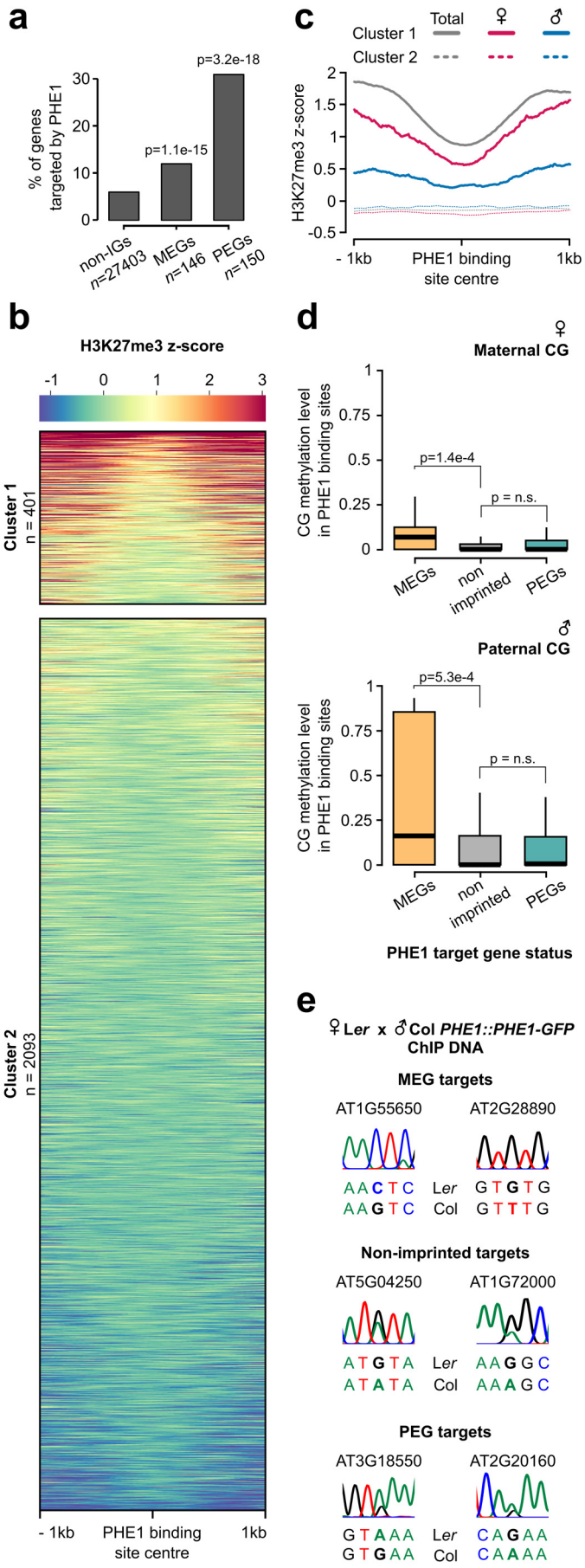


Figure 2 - Parental asymmetry of epigenetic marks in imprinted gene promoters conditions PHE1 binding. **a,** Fraction of non-imprinted and published imprinted genes targeted by PHE1. P-values were determined using the hypergeometric test. The list of published imprinted genes used for this analysis is detailed in **Fig 1. - source data 1**. **b,** Heatmap of endosperm H3K27me3 distribution along PHE1 binding sites. Each horizontal line represents one binding site. Clusters were defined based on the pattern of H3K27me3 distribution (see Methods) **c,** Metagene plot of average maternal (♀, pink), paternal (♂, blue) and total (grey) endosperm H3K27me3 along PHE1 binding sites. **d,** CG methylation levels in maternal (♀, upper panel) and paternal (♂, lower panel) alleles of PHE1 binding sites associated with MEGs (yellow), PEGs (green) and non-imprinted (grey) PHE1 targets. P-values were determined using two-tailed Mann-Whitney tests. **e,** Sanger sequencing of imprinted and non-imprinted gene promoters bound by PHE1. SNPs for maternal (L*er*) and paternal (Col) alleles are shown, (*n* = 1 biological replicate). Maternal:total read ratios for imprinted genes are as follows: AT1G55650 – 1.0; AT2G28890 – 0.97; AT3G18550 – 0.44; AT2G20160 – 0.32.

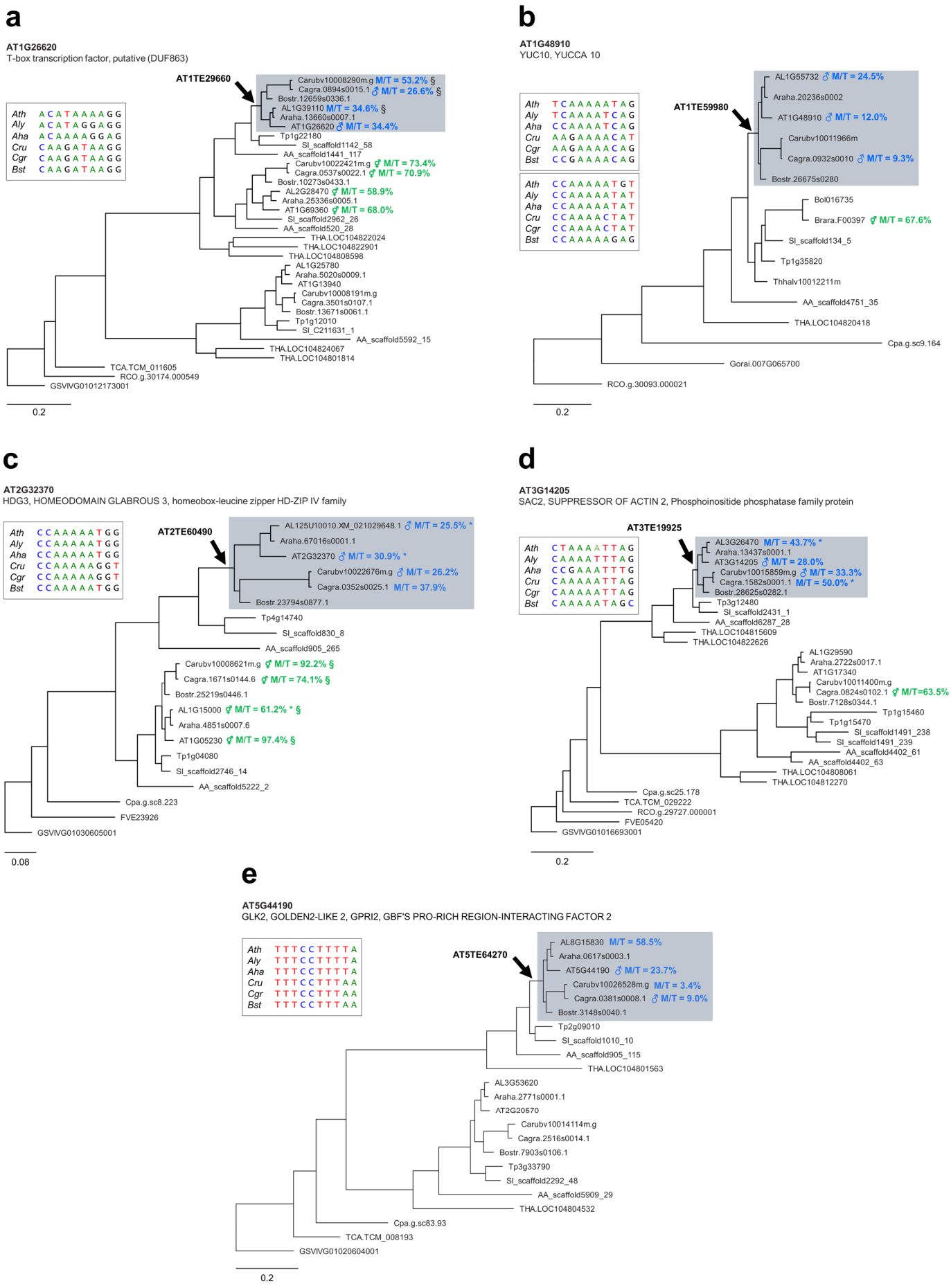
TEs – occurring in the central cell, but not in sperm – is a major driver for imprinted gene expression (Gehring, 2013; Rodrigues and Zilberman, 2015). In MEGs, DNA hypomethylation of maternal alleles leads to their expression, while DNA methylation represses the paternal allele (Gehring, 2013; Rodrigues and Zilberman, 2015). In PEGs, the hypomethylated maternal allele undergoes trimethylation of lysine 27 of histone H3 (H3K27me3), a repressive histone modification established by the Fertilization Independent Seed-Polycomb Repressive Complex2 (FIS-PRC2). This renders the maternal alleles inactive, while the paternal allele is expressed (Gehring, 2013; Moreno-Romero et al., 2016; Rodrigues and Zilberman, 2015).

Given the significant overrepresentation of imprinted genes among PHE1 targets, we assessed how the epigenetic landscape at those loci correlates with DNA-binding by PHE1. We surveyed levels of endosperm H3K27me3 within PHE1 binding sites and identified two distinct clusters (**Fig. 2b**). Cluster 1 was characterized by an accumulation of H3K27me3 in regions flanking the centre of the binding site, while the centre itself was devoid of this mark. Cluster 2 contained binding sites largely devoid of H3K27me3. The distribution of H3K27me3 in cluster 1 was mostly attributed to the deposition of H3K27me3 on the maternal alleles, while the paternal alleles were devoid of this mark (**Fig. 2c**) – a pattern usually associated with PEGs (Moreno-Romero et al., 2016). Consistently, genes associated with cluster 1 binding sites had a more paternally-biased expression in the endosperm when compared to genes associated with cluster 2 (**Fig. 2 – Fig. Supplement 1a**). This is reflected by more PEGs and putative PEGs being associated with cluster 1 (**Fig. 2 – Fig. Supplement 1b**). We also identified parental-specific differences in DNA methylation, specifically in the CG context: PHE1 binding sites associated with MEGs had significantly higher methylation levels in paternal alleles than in maternal alleles (**Fig. 2d**).

We hypothesized that the differential parental deposition of epigenetic marks in parental alleles of PHE1 binding sites can result in differential accessibility of each allele. This might impact the binding of PHE1, and therefore also impact transcription in a parent-of-origin-specific manner. To test this, we performed ChIP using a L*er* maternal plant and a Col *PHE1::PHE1-GFP* pollen donor, taking advantage

129 of SNPs between these two accessions to discern parental preferences of PHE1 binding. Using Sanger
130 sequencing, we determined the parental origin of enriched ChIP-DNA in MEG, non-imprinted, and PEG
131 targets (**Fig. 2e, Fig. 2 – Fig. Supplement 2**). While binding of PHE1 was biallelic in non-imprinted
132 targets (**Fig 2e**), only maternal binding was detected in the tested MEG targets, supporting the idea
133 that CG hypermethylation of paternal alleles prevents their binding by PHE1 (**Fig. 2e, Fig. 2 – Fig.**
134 **Supplement 3a**). Interestingly, we observed biallelic binding in PEG targets (**Fig. 2e**). Even though the
135 maternal PHE1 binding sites in PEGs were flanked by H3K27me3 (**Fig. 2b-c**), correlating with
136 transcriptional repression of maternal alleles, the absence of this mark within the binding site centres
137 seems to be permissive for maternal PHE1 binding. We speculate that the accessibility of this site might
138 be important to mediate recruitment of H3K27me3 in the central cell and/or for maintenance of this
139 mark during endosperm development (**Fig. 2 – Fig. Supplement 3a**).
140 Previous studies have shown that PEGs are often flanked by RC/Helitrons (Pignatta et al., 2014;
141 Wolff et al., 2011), a phenomenon suggested to lead to the parental asymmetry of epigenetic marks

Figure 3 - Ancestral RC/Helitron insertions are associated with gain of imprinting in the Brassicaceae. a-e, Phylogenetic analyses of PHE1-targeted PEGs and their homologs. Each panel represents a distinct target gene, and its corresponding homologs in different species. The genes in grey background have homologous RC/Helitron sequences in their promoter region. The arrow indicates the putative insertion of an ancestral RC/Helitron. The identity of the RC/Helitron identified in *A. thaliana* is indicated. These *A. thaliana* RC/Helitrons contain a PHE1-DNA binding motif and are associated with a PHE1 binding site. The inset boxes represent the alignment between the *A. thaliana* PHE1 DNA-binding motif and similar DNA motifs contained in RC/Helitrons present at the promoter regions of orthologous genes. When available, the imprinting status of a given gene is indicated by the presence of ♂ (PEG), or ☿ (non-imprinted), and reflects the original imprinting analyses done in the source publications (see Methods). The maternal:total read ratio (M/T) for each gene is also indicated. §: potential contamination from maternal tissue. *: accession-biased expression. Scale bar represents substitution per site for the ML tree. The tree is unrooted. Gene identifier nomenclatures: AT, *Arabidopsis thaliana*; AL, *Arabidopsis lyrata*; Araha, *Arabidopsis halleri*; Bostr, *Boechera stricta*; Carubv, *Capsella rubella*; Cagra, *Capsella grandiflora*; Tp, *Schrenkiella parvula*; SI, *Sisymbrium irio*; Bol, *Brassica oleracea*; Brapa, *Brassica rapa*; Thhalv, *Eutrema salsugineum*; AA, *Aethionema arabicum*; THA, *Tarenaya hassleriana*; Cpa, *Carica papaya*; TCA, *Theobroma cacao*; Gorai, *Gossypium raimondii*; RCO, *Ricinus communis*; FVE, *Fragaria vesca*; GSVIVG, *Vitis vinifera*.



in these genes (Moreno-Romero et al., 2016; Pignatta et al., 2014). Consistent with our finding that PHE1 binding sites overlapped with RC/Helitrons (**Fig. 1b**), we found that PHE1 DNA-binding motifs were contained within these TEs significantly more frequently in PEGs than in non-imprinted genes (**Fig. 2 – Fig. Supplement 3b**). Furthermore, we detected the presence of homologous RC/Helitrons containing PHE1 binding motifs in the promoter regions of several PHE1-targeted PEG orthologs (**Fig. 3 a-e, Fig. 3 – Figure Supplement 1**), indicating ancestral insertion events. The presence of these RC/Helitrons correlated with paternally-biased expression of the associated orthologs, providing further support to the hypothesis these TEs contribute to the gain of imprinting, especially of PEGs. Thus, besides facilitating the asymmetry of epigenetic marks, these TEs can contribute to the generation of novel gene promoters that ensure the timely endosperm expression of PEGs, under the control of PHE1, and possibly other type I MADS-box TFs (**Fig. 2 – Fig. Supplement 3a**).

Among the PEGs targeted by PHE1 were *ADM*, *SUVH7*, *PEG2,* and *NRPD1a* (**Fig. 1 – Fig. Supplement 2**). Mutants in all four PEGs suppress the abortion of triploid (3x) seeds generated by paternal excess interploidy crosses (Martinez et al., 2018; Satyaki and Gehring, 2019; Wolff et al., 2015). Furthermore, we found that close to 50% of highly upregulated genes in 3x seeds are targeted by PHE1 (**Fig. 4a**), suggesting this TF might play a central role in mediating the strong gene deregulation observed in these seeds. If this is true, removal of PHE1 in 3x seeds is expected to suppress 3x seed inviability. To test this hypothesis, we generated a *phe1* CRISPR/Cas9 mutant in the *phe2* background, since both genes are likely redundant (Villar et al., 2009) (**Fig. 4 – Fig. Supplement 1a**). We introduced *phe1 phe2* into the *omission of second division 1 (osd1)* mutant background, which produces diploid gametes at high frequency (d'Erfurth et al., 2009). Wild-type (wt) and *phe2* maternal plants pollinated with *osd1* pollen form 3x seeds that abort at high frequency (Kradolfer et al., 2013) (**Fig. 4b, Fig. 4 – Fig. Supplement 1b**). In contrast, *phe1 phe2 osd1* pollen strongly suppressed 3x seed inviability, reflected by the increased germination of 3x *phe1 phe2* seeds (**Fig. 4c, Fig. 4 – Fig. Supplement 1c**). This phenotype could be reverted by introducing the *PHE1::PHE1-GFP* transgene paternally (**Fig. 4b-c, Fig. 4 – Fig. Supplement 1b-c**). Notably, 3x seed rescue was mostly mediated by *phe1*, as the presence of a wt *PHE2*

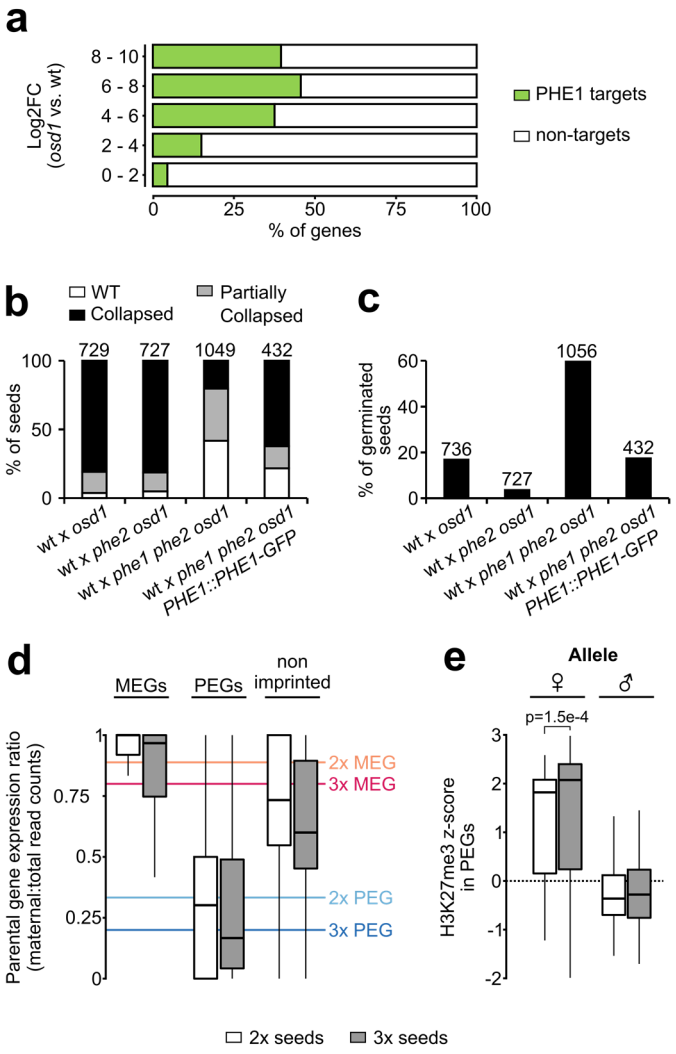


Figure 4 - PHE1 establishes 3x seed inviability of paternal excess crosses. a, Target status of upregulated genes in paternal excess crosses. Highly upregulated genes in 3x seeds are more often targeted by PHE1 (p<2.2e-16, χ^2 test). **b-c,** Seed inviability phenotype (b) of paternal excess crosses in wild-type (wt), *phe1 phe2*, and *phe1* complementation lines, with respective seed germination rates (c). The maternal parent is always indicated first. Remaining control crosses are shown in Fig. 4 – Fig. Supplement 1b-c. *n* = numbers on top of bars (seeds). **d,** Parental expression ratio of imprinted genes in the endosperm of 2x (white) and 3x seeds (grey). Solid lines indicate the ratio thresholds for definition of MEGs and PEGs, in 2x and 3x seeds. **e,** Accumulation of H3K27me3 across maternal (♀) and paternal (♂) gene bodies of PEGs in the endosperm of 2x and 3x seeds (white and grey, respectively). H3K27me3 accumulation in MEGs and non-imprinted genes is shown in Fig. 4 – Fig. Supplement 2. P-value was determined using a two-tailed Mann-Whitney test

allele in 3x seeds (wt x *phe1 phe2 osd1*) led to comparable rescue levels than when having no wt *PHE2* allele present (*phe2* x *phe1 phe2 osd1*) (**Fig. 4b-c, Fig. 4 – Fig. Supplement 1b-c**). Importantly, 3x seed rescue was accompanied by reestablishment of endosperm cellularisation (**Fig. 4 – Fig. Supplement 1d-e**), and reduced expression of PHE1 target genes (**Fig. 4 – Fig. Supplement 1f**).

Loss of FIS-PRC2 function causes a similar phenotype to that of paternal excess 3x seeds, correlating with largely overlapping sets of deregulated genes (Erilova et al., 2009; Tiwari et al., 2010). Since FIS-RC2 is a major regulator of PEGs in *Arabidopsis* endosperm (Moreno-Romero et al., 2016), we addressed the question whether imprinting is disrupted in 3x seeds. To assess this, we analysed the

177 parental expression ratio of imprinted genes in the endosperm of 2x and 3x seeds. Surprisingly,
178 imprinting was not disrupted in 3x seeds (**Fig. 4d**), consistent with the presence of similar
179 H3K27me3levels on the maternal alleles of PEGs in the endosperm of 2x and 3x seeds (**Fig. 4e, Fig. 4 –**
180 **Fig. Supplement 2**). Collectively, these data show that the major upregulation of imprinted gene
181 expression in 3x seeds is due to increased transcription of the active allele, likely mediated by PHE1
182 and other MADS-box TFs, with maintenance of the imprinting status.

183 In summary, this work reveals that the MADS-box TF PHE1 is a major regulator of imprinted genes
184 in the *Arabidopsis* endosperm, and that this TF establishes a reproductive barrier in response to
185 interploidy hybridizations. We furthermore show that deregulated PEGs in 3x seeds remain imprinted,
186 but that the active allele becomes strongly overexpressed, correlating with increased PHE1 activity in
187 these seeds (Erilova et al., 2009). Importantly, we reveal a novel role for RC/Helitrons in the regulation
188 of imprinted genes by showing that they contain PHE1 DNA-binding sites. Our data favour a scenario
189 where these elements have been domesticated to function as providers of *cis*-regulatory sequences
190 that facilitate transcription. Similarly, binding sites for type II MADS-box TFs, as well as for E2F TFs,
191 have been shown to be carried by TEs (Hénaff et al., 2014; Muiño et al., 2016). These results, together
192 with our study, provide examples of TE-mediated distribution of TF binding sites throughout flowering
193 plant genomes, adding support to the long-standing idea that transposition facilitates the formation
194 of *cis*-regulatory architectures required to control complex biological processes (Britten and Davidson,
195 1971; Feschotte, 2008; Hirsch and Springer, 2017). We speculate that this process may have
196 contributed to endosperm evolution by allowing the recruitment of crucial developmental genes into
197 a single transcriptional network, regulated by type I MADS-box TFs. The diversification of the
198 mammalian placenta has been connected with the dispersal of hundreds of placenta-specific
199 enhancers by endogenous retroviruses (Chuong et al., 2013; Dunn-Fletcher et al., 2018), suggesting
200 that the convergent evolution of the endosperm in flowering plants and the mammalian placenta have
201 been promoted by TE transpositions.

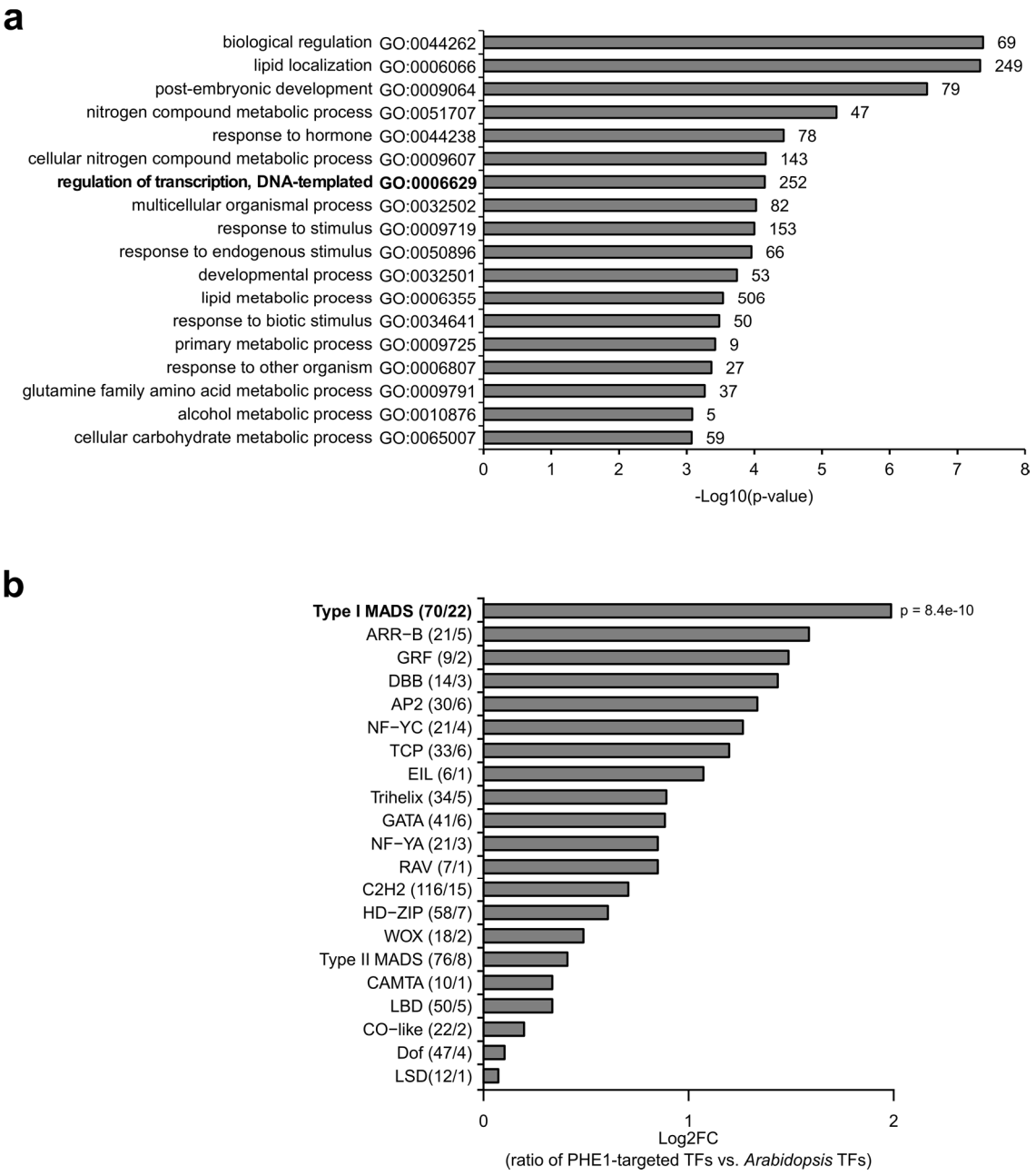


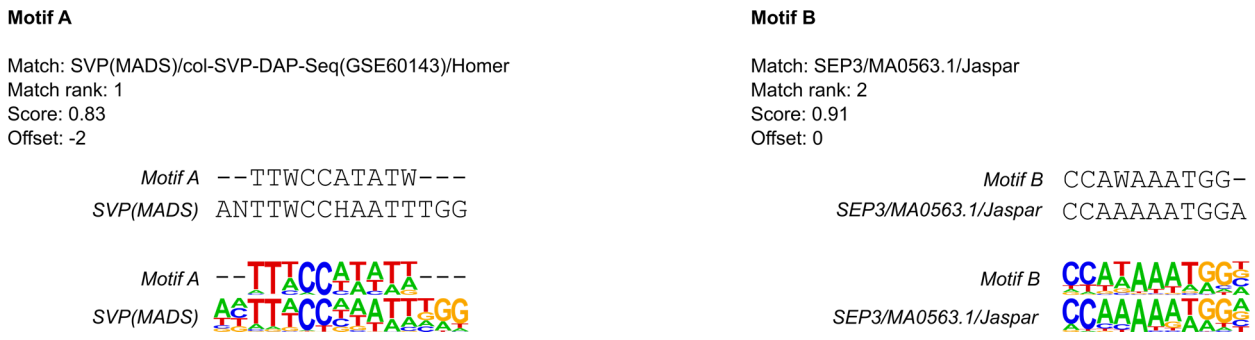
Figure 1 – Figure Supplement 1. Transcription factor genes are enriched among PHE1 targets.

a, Enriched biological processes within PHE1 target genes. Numbers on bars indicate number of PHE1 target genes within each GO-term. **b,** Enrichment of transcription factor (TF) families among PHE1 target genes (see Methods). Numbers in parenthesis indicate the total number of *Arabidopsis* genes belonging to a certain transcription factor family, and the total number of genes in that family targeted by PHE1, respectively. P-value was determined using the hypergeometric test.

208 **Figure 1 – Figure Supplement 2. PHE1 target genes previously implicated in endosperm**
209 **development.**

Gene ID	Imprinting status	Description of function
AGL62	Non imprinted	Type I MADS-box TF involved in endosperm proliferation and seed coat development (Figueiredo et al., 2016, 2015; Kang et al., 2008; Roszak and Köhler, 2011)
YUC10	PEG	Flavin monooxygenase that catalyses the last step of the Trp-dependent auxin biosynthetic pathway (Zhao, 2012). Involved in endosperm proliferation and cellularisation (Batista et al., 2019; Figueiredo et al., 2015)
IKU2	Non imprinted	Encodes a leucine-rich repeat receptor kinase protein that, together with *MINI3*, is part of the IKU pathway controlling seed size. *iku2* mutants show reduced endosperm growth and early endosperm cellularisation (Garcia et al., 2003; Luo et al., 2005).
MINI3	Non imprinted	WRKY TF that, together with *IKU2*, is part of the IKU pathway controlling seed size. *mini3* mutants show reduced endosperm growth and early endosperm cellularisation (Luo et al., 2005).
ZHOUPI	Non imprinted	Encodes a bHLH TF expressed in the embryo surrounding region of the endosperm. It is essential for embryo cuticle formation and endosperm breakdown after its cellularisation (Xing et al., 2013; Yang et al., 2008).
MEA	MEG	Subunit of the FIS-PRC2 complex, responsible for depositing H3K27me3 at target loci, including PEGs (Moreno-Romero et al., 2019; Moreno-Romero et al., 2016). Loss of MEA and, consequently, paternally biased expression of PEGs, leads to a 3x seed like phenotype (Grossniklaus et al., 1998; Kiyosue et al., 1999).
ADM	PEG	Interacts with SUVH9 and AHL10 to promote H3K9me2 deposition in TEs, influencing neighbouring gene expression. Mutations in *ADM* lead to rescue of the 3x seed abortion phenotype (Jiang et al., 2017; Kradolfer et al., 2013; Wolff et al., 2015).
SUVH7	PEG	Encodes a putative histone-lysine N-methyltransferase. Mutations in *SUVH7* lead to rescue of the 3x seed abortion phenotype (Wolff et al., 2015).
PEG2	PEG	Encodes a putative SNARE-like superfamily protein, which is not translated in endosperm. *PEG2* transcripts act as a sponge for siRNA854, thus regulating UBP1 abundance (Wang et al., 2018). Mutations in *PEG2* lead to rescue of the 3x seed abortion phenotype (Wang et al., 2018; Wolff et al., 2015).
NRPD1a	PEG	Encodes the largest subunit of RNA POLYMERASE IV, involved in the RNA-directed DNA methylation pathway. Mutations in *NRPD1a* lead to rescue of the 3x seed abortion phenotype (Martinez et al., 2018).

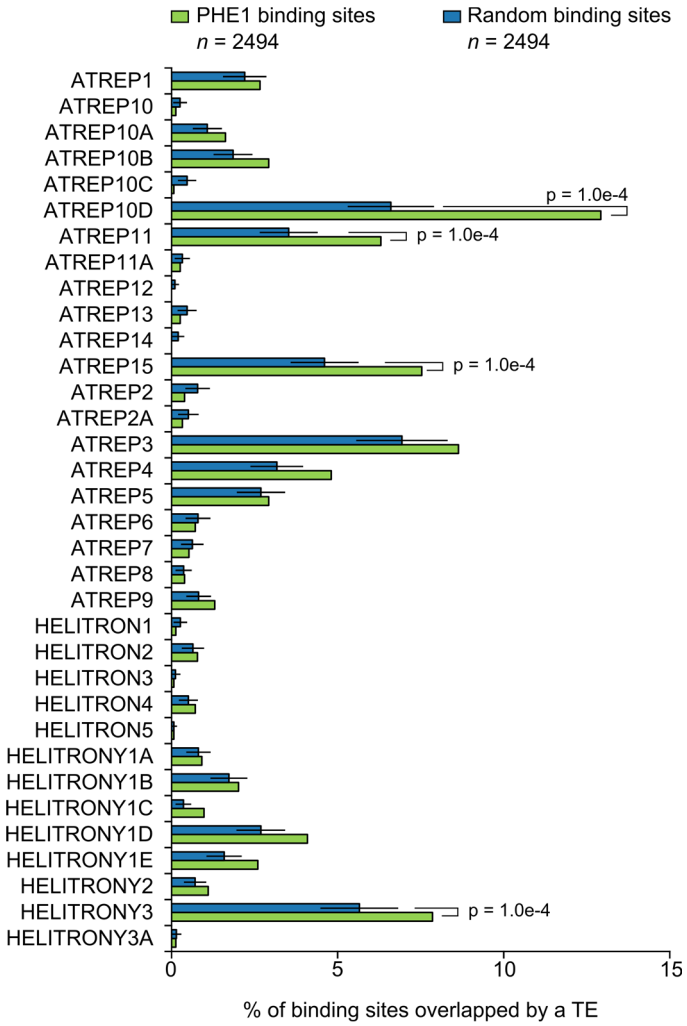
210



211 **Figure 1 – Figure Supplement 3. PHE1 DNA-binding motifs show similarity to type II MADS-box CArG-**

212 **boxes.**

213 Alignment between PHE1 DNA-binding motifs and known motif matches.



214 **Figure 1 – Figure Supplement 4. RC/Helitron family overlap with PHE1 binding sites.**

215 Fraction of PHE1 binding sites (green) overlapping with different RC/Helitron TE families. Overlap is
216 expressed as the percentage of binding sites where spatial intersection with the families specified on
217 the y-axis is observed. A set of random binding sites is used as control (blue). This control set was
218 obtained by randomly shuffling the PHE1 binding sites in *A. thaliana* gene promoters (see Methods).
219 P-values were determined using Monte Carlo permutation tests (see Methods). Bars represent ± sd, (*n*
220 = 2494, PHE1 binding sites, Random binding sites).

221

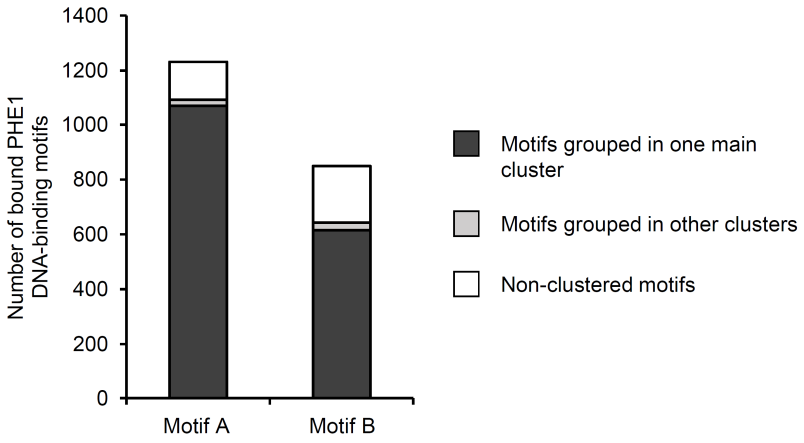


Figure 1 – Figure Supplement 5. Clustering of bound PHE1 DNA-binding motifs.

PHE1-DNA binding motifs and their flanking sequences were paired based on sequence homology. The pairwise homologous sequences were then merged into higher order clusters, based on shared elements in the homologous pairs (see Methods). Motifs are those identified in **Fig. 1a**.

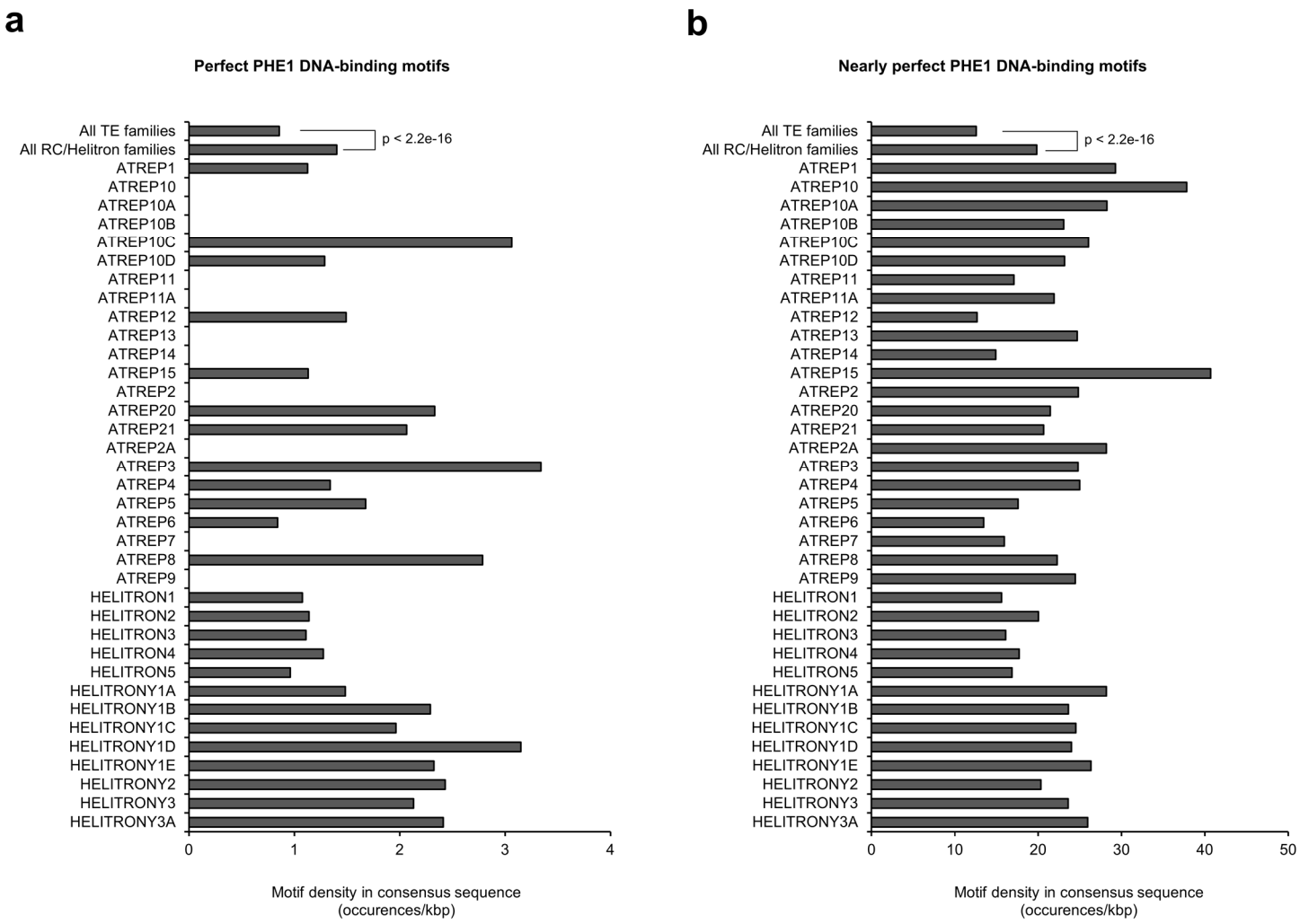
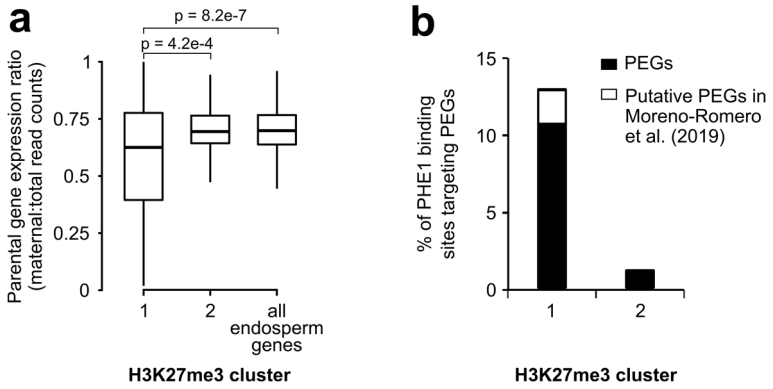


Figure 1 – Figure Supplement 6. PHE1 DNA-binding motif densities in RC/Helitron consensus sequences.

a, Density of perfect PHE1-DNA binding motifs in consensus sequences of different RC/Helitron families. Perfect motifs are those described in **Fig. 1a**. As a reference, the density of perfect PHE1 DNA-binding motifs in consensus sequences of all TE families is shown. **b,** Density of nearly perfect PHE1-DNA binding motifs in the consensus sequences of different RC/Helitron families. Nearly perfect motifs are those sequences where only one nucleotide substitution is required to generate a perfect PHE1 DNA-binding motif. As a reference, the density of nearly perfect PHE1 DNA-binding motifs in the consensus sequences of all TE families is shown. P-values were determined using χ^2 tests.



236

237 **Figure 2 – Figure Supplement 1. Characterization of H3K27me3 clusters.**

238 **a,** Parental gene expression ratio of PHE1 targets associated with Cluster 1 and Cluster 2 binding sites,

239 and all endosperm expressed genes. P-values were determined using two-tailed Mann-Whitney tests.

240 **b,** Fraction of Cluster 1 and Cluster 2 binding sites that target PEGs (**Fig. 1 – source data 1**) and putative

241 PEGs (Moreno-Romero et al., 2019). (a-b) H3K27me3 clusters are those defined in **Fig. 2b**.

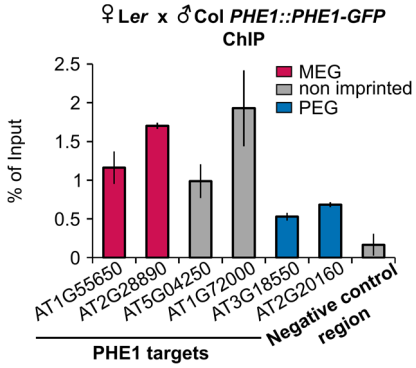


Figure 2 – Figure Supplement 2. Parental-specific PHE1 ChIP.

qPCR of purified ChIP-DNA. Enrichment is shown as % of input DNA, in regions associated with MEGs (pink), PEGs (blue), and non-imprinted (grey) PHE1 target genes. Bars represent ± sd. Data from one representative biological replicate is shown (*n* = 2 biological replicates).

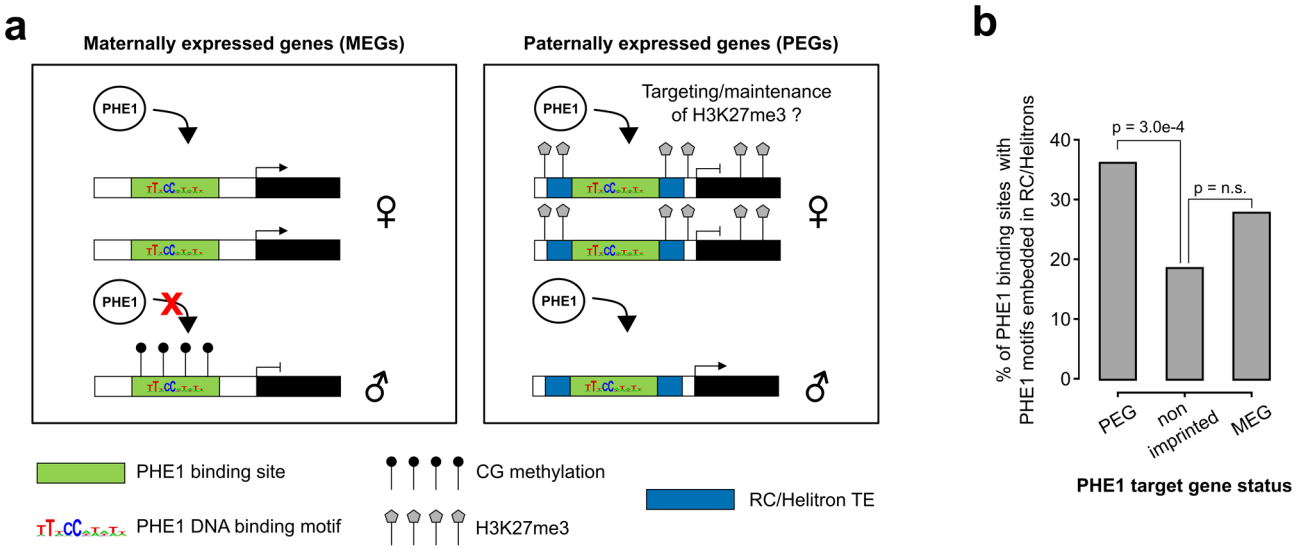


Figure 2 – Figure Supplement 3. Control of imprinted gene expression by PHE1.

a, Schematic model of imprinted gene control by PHE1. Maternally expressed genes (left panel) show DNA hypermethylation of paternal (♂) PHE1 binding sites. This precludes PHE1 accessibility to paternal alleles, leading to predominant binding and transcription from maternal (♀) alleles. In paternally expressed genes (right panel), RC/Helitrons found in flanking regions carry PHE1 DNA-binding motifs, allowing PHE1 binding. The paternal PHE1 binding site is devoid of repressive H3K27me3, facilitating binding of PHE1 and transcription of this allele. H3K27me3 accumulates at the flanks of maternal PHE1 binding sites, while the binding sites remain devoid of this repressive mark. PHE1 is able to bind maternal alleles, but fails to induce transcription. We hypothesize that the accessibility to maternal PHE1 binding sites might be important for deposition of H3K27me3 during central cell development (possibly by another type I MADS-box transcription factor). It may furthermore be required for maintenance of H3K27me3 during endosperm division. **b,** Fraction of PHE1 binding sites targeting MEGs, PEGs, or non-imprinted genes where a spatial overlap between a RC/Helitron and a PHE1 DNA-binding motif is observed (as illustrated in a, right panel). P-values were determined using the hypergeometric test.

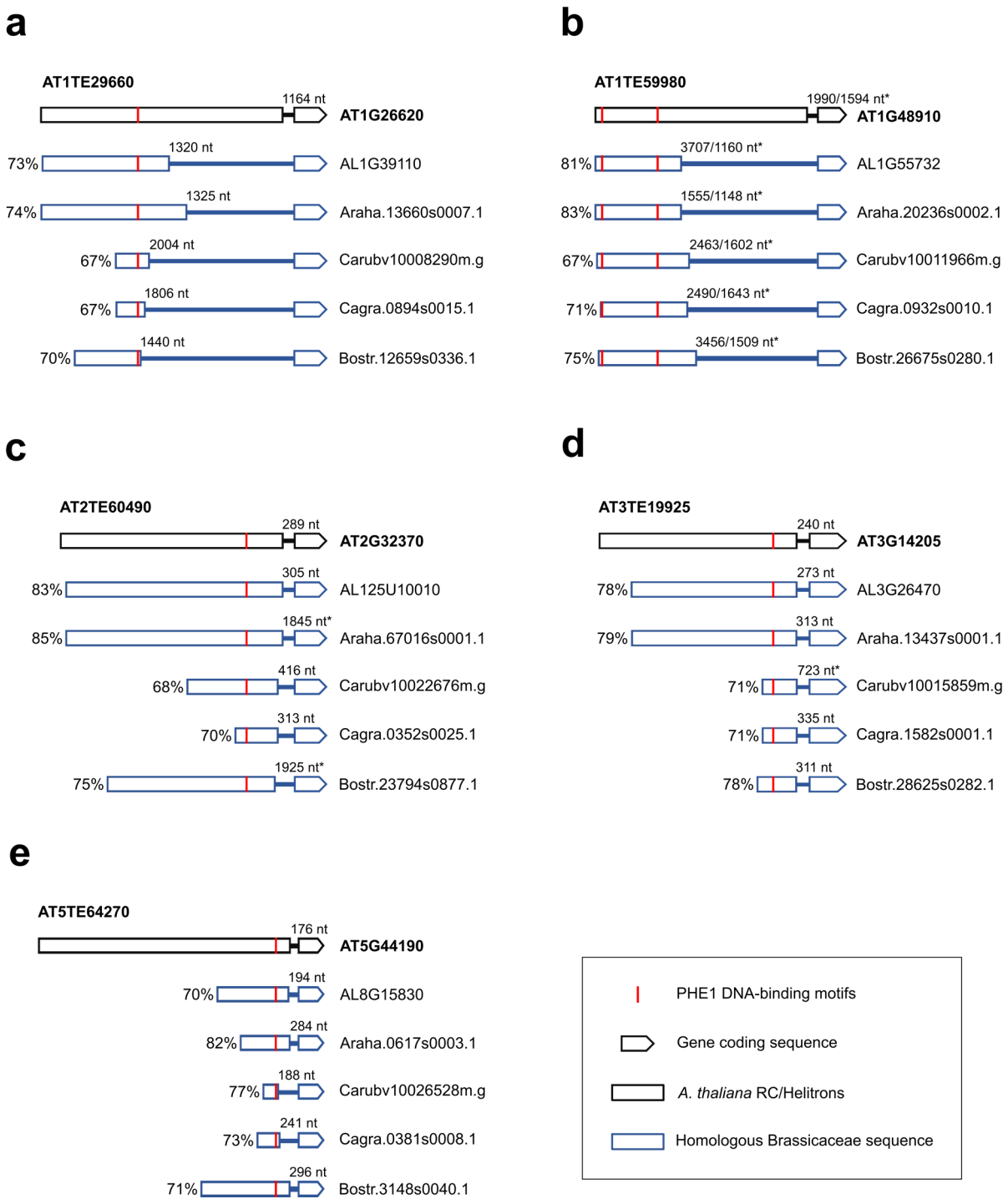
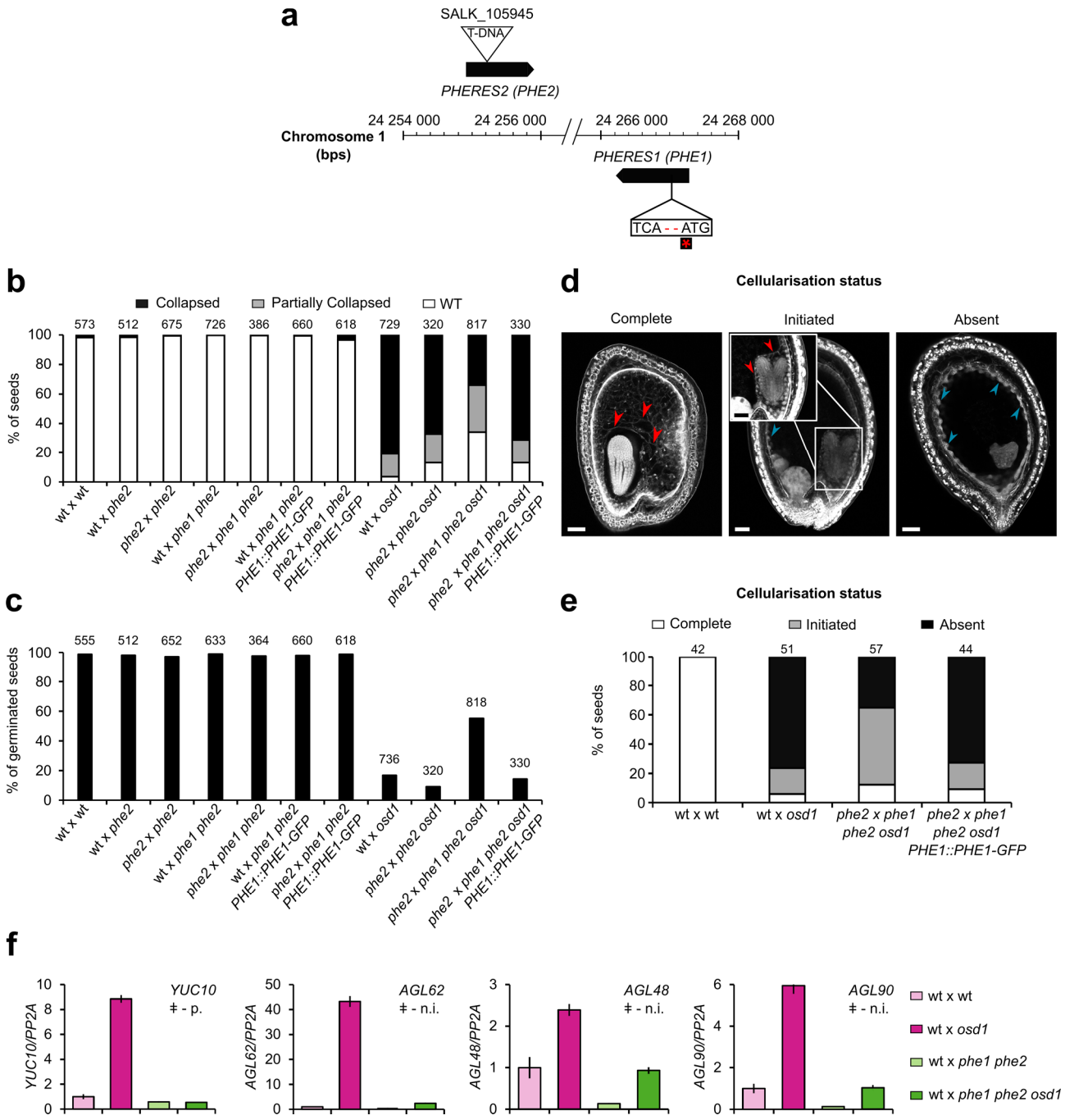


Figure 3 – Figure Supplement 1. Homology analysis of PEGs and upstream RC/Helitron sequences within Brassicaceae.

Alignments between *Arabidopsis thaliana* RC/Helitrons associated with PEGs (**Fig. 3a-e**) and homologous sequences in other Brassicaceae, where the putative homologous PHE1-DNA binding motifs are present across all six species. For each aligned fragment, the percent sequence identity is indicated. The distances between the DNA-binding motifs and the TSS of PEGs are labelled. *

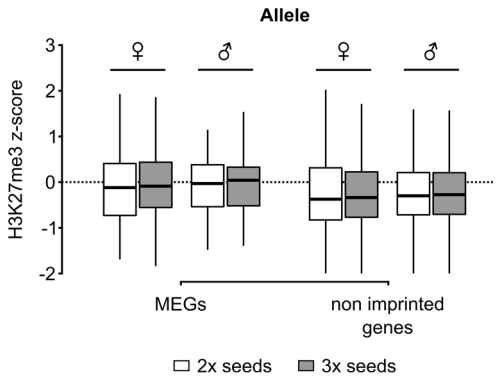
corresponds to the distance to the translation start site, in cases where the annotation for the 5'-UTR of the gene is not available.



273 **Figure 4 – Figure Supplement 1. Rescue of 3x seed inviability in *phe1 phe2*.**

274 **a,** Schematic representation of the *phe1 phe2* mutant. CRISPR/Cas9 was used to generate a premature
275 stop codon in *PHE1*, represented by the red asterisk. *PHE1* mutagenesis was done in the *phe2*
276 background, due to the proximity and likely redundancy of both genes (see Methods). **b-c,** Seed
277 inviability phenotype (b) of wild-type (wt) and paternal excess crosses, in wt, *phe1 phe2*, and *phe1*
278 complementation lines, with respective seed germination rates (c). **d-e**, Status of endosperm
279 cellularisation in wt and paternal excess seeds. (b-c,e) *n* = numbers on top of bars (seeds). **f,** Expression

of PHE1 target genes in rescued 3x seeds. ‡ - gene neighbours an RC/Helitron containing a PHE1 DNA-binding motif; p. – PEG; n.i. – non-imprinted gene. Bars represent ± sd. Data from one representative biological replicate is shown (*n* = 2 biological replicates).



283

284 **Figure 4 – Figure Supplement 2. Distribution of H3K27me3 in 2x and 3x seeds.**

285 Accumulation of H3K27me3 in maternal (♀) and paternal (♂) gene bodies of MEGs and non-imprinted

286 genes in the endosperm of 2x and 3x seeds (white and grey, respectively).

Methods

Plant material and growth conditions

Arabidopsis thaliana seeds were sterilized in a closed vessel containing chlorine gas, for 3 hours. Chlorine gas was produced by mixing 3 mL HCl 37% and 100 mL of 100% commercial bleach. Sterile seeds were plated in ½ MS-medium (0.43% MS salts, 0.8% Bacto Agar, 0.19% MES hydrate) supplemented with 1% Sucrose. When required, appropriate antibiotics were supplemented to the medium. Seeds were stratified for 48 h, at 4°C, in darkness. Plates containing stratified seeds were transferred to a long-day growth chamber (16h light / 8h dark; 110 µmol $s^{-1}m^{-2}$; 21ºC; 70% humidity), where seedlings grew for 10 days. After this period, the seedlings were transferred to soil and placed in a long-day growth chamber.

Several mutant lines used in this study have been previously described: *osd1-1* (d'Erfurth et al., 2009), *osd1-3* (Heyman et al., 2011) and *pi-1* (Goto and Meyerowitz, 1994). The phe2 allele corresponds to a T-DNA insertion mutant (SALK_105945). Phenotypical analysis of this mutant revealed no deviant phenotype relative to Col wt plants (data not shown). Genotyping of phe2 was done using the following primers (PHE2 fw 5'-AAATGTCTGGTTTTATGCCCC-3', PHE2 rv 5'-GTAGCGAGACAATCGATTTCG-3', T-DNA 5'-ATTTTGCCGATTTCGGAAC-3').

Generation of phe1 phe2

The *phe1 phe2* double mutant was generated using the CRISPR/Cas9 technique. A 20-nt sgRNA targeting *PHE1* was designed using the CRISPR Design Tool (Ran et al., 2013). A single-stranded DNA oligonucleotide corresponding to the sequence of the sgRNA, as well as its complementary oligonucleotide, were synthesized. BsaI restriction sites were added at the 5' and 3' ends, as represented by the underlined sequences (sgRNA fw 5'- ATTGCTCCTGGATCGAGTTGTAC-3'; sgRNA rv 5'-AAACGTACAACTCGATCCAGGAG-3'). These two oligonucleotides were then annealed to produce a double stranded DNA molecule.

The double-stranded oligonucleotide was ligated into the egg-cell specific pHEE401E CRISPR/Cas9 vector (Z.-P. Wang et al., 2015) through the BsaI restriction sites. This vector was transformed into the *Agrobacterium tumefaciens* strain GV3101, and *phe2/-* plants were subsequently transformed using the floral-dip method (Clough and Bent, 1998).

To screen for T1 mutant plants, we performed Sanger sequencing of *PHE1* amplicons derived from these plants, and obtained with the following primers (fw 5'-AGTGAGGAAAACAACATTCACCA-3'; rv 5'-GCATCCACAACAGTAGGAGC-3'). The selected mutant contained a homozygous two base pair deletion that leads to a premature stop codon, and therefore a truncated $PHE1_{1\text{-}50aa}$ protein. In the T2 generation, the segregation of pHEE401E allowed to select plants that did not contain this vector and that were double homozygous *phe1 phe2* mutants. Genotyping of the *phe1* allele was done using primers fw 5'- AAGGAAGAAAGGGATGCTGA-3' and rv 5'-TCTGTTTCTTTGGCGATCCT-3', followed by RsaI digestion.

Seed imaging

Analysis of endosperm cellularisation status was done following the Feulgen staining protocol described previously (Batista et al., 2019). Imaging of Feulgen-stained seeds was done in a Zeiss LSM780 NLO multiphoton microscope with excitation wavelength of 800 nm, and acquisition between 520 nm – 695 nm.

RT-qPCR

RNA extraction, cDNA synthesis and qPCR analyses for *AGL* and *PEG* expression were performed as described previously (Batista et al., 2019). Two biological replicates per cross were used. Primer sequences for *YUC10*, *AGL62*, and *PP2A* were described previously (Batista et al., 2019). For the remaining genes, primer sequences were as follows: *AGL48* (fw 5'- TTCGCGATCCACCAGTGTTT-3', rv 5'-GACCGCCTCCTACAAAACCA-3'), *AGL90* (fw 5'-TTGGTGATGAGTCGTTTTCCGA-3', rv 5'-TCATATTCGCATTTGCGTCCG-3').

338

339 *Chromatin immunoprecipitation (ChIP)*

340 To find targets of PHE1, we performed a ChIP experiment using a reporter line containing the PHE1
341 protein tagged with GFP. This reporter line, which we denote as *PHE1::PHE1-GFP,* contains the *PHE1*
342 promoter, its coding sequence and its 3' regulatory sequence, and is present in a Col background
343 (Weinhofer et al., 2010). Given the presence of the 3' regulatory regions (Makarevich et al., 2008), this
344 reporter behaves as a paternally expressed gene, similarly to the endogenous *PHE1* gene. Crosslinking
345 of plant material was done by collecting 600 mg of 2 days after pollination (DAP) *PHE1::PHE1-GFP*
346 siliques, and vacuum infiltrating them with a 1% formaldehyde solution in PBS. The vacuum infiltration
347 was done for two periods of 15 min, with a vacuum release between each period. The crosslinking was
348 then stopped by adding 0.125 mM glycine in PBS and performing a vacuum infiltration for a total of 15
349 min, with a vacuum release each 5 min. The material was then ground in liquid nitrogen, resuspended
350 in 5mL Honda buffer (Moreno-Romero et al., 2017), and incubated for 15 min with gentle rotation.
351 This mixture was filtered twice through Miracloth and one time through a CellTrics filter (30 µm), after
352 which a centrifugation for 5 min, at 4°C and 1,500 g was performed. The nuclei pellet was then
353 resuspended in 100 µL of nuclei lysis buffer (Moreno-Romero et al., 2017), and the ChIP protocol was
354 continued as described before (Moreno-Romero et al., 2017). ChIP DNA was isolated using the Pure
355 Kit v2 (Diagenode), following the manufacturer's instructions.

356 For the parental-specific PHE1 ChIP the starting material consisted of 600 mg of 2DAP siliques from
357 crosses between a *pi-1* mother (L*er* ecotype) and a *PHE1::PHE1-GFP* father (Col ecotype). The male
358 sterile *pi-1* mutant was used to avoid emasculation of maternal plants. Crosslinking of plant material,
359 nuclei isolation, ChIP protocol, and ChIP-DNA purification were the same as described before.

360 To assess parental-specific H3K27me3 profiles in 3x seeds, the INTACT system was used to isolate
361 4DAP endosperm of seeds derived from L*er pi-1* x Col *INTACT osd1-1* crosses, as described previously
362 (Jiang et al., 2017; Moreno-Romero et al., 2017). *osd1* mutants were used for their ability to generate
363 unreduced male gametes, leading to the formation of 3x seeds. ChIPs against H3 and H3K27me3 were

364 then performed on the isolated endosperm nuclei, following the previously described protocol
365 (Moreno-Romero et al., 2017).
366 The antibodies used for these ChIP experiments were as follows: GFP Tag Antibody (A-11120,
367 Thermo Fisher Scientific), anti-H3 (Sigma, H9289), and anti-H3K27me3 (Millipore, cat. no. 07-449). All
368 experiments were performed with two biological replicates.
369
370 *Library preparation and sequencing*
371 *PHE1::PHE1-GFP* ChIP libraries were prepared using the Ovation Ultralow v2 Library System
372 (NuGEN), with a starting material of 1 ng, following the manufacturer's instructions. These libraries
373 were sequenced at the SciLife Laboratory (Uppsala, Sweden), on an Illumina HiSeq2500 platform, using
374 50 bp single-end reads.
375 Library preparation and sequencing of H3K27me3 ChIPs in 3x seeds was done as described
376 previously (Jiang et al., 2017).
377 Both datasets were deposited at NCBI's Gene Expression Omnibus database
378 (https://www.ncbi.nlm.nih.gov/geo/), under the accession number GSE129744.
379
380 *qPCR and Sanger sequencing of parental-specific PHE1 ChIP*
381 Purified ChIP DNA and its respective input DNA obtained from the parental-specific PHE1 ChIP were
382 used to perform qPCR. Positive and negative genomic regions for PHE1 binding were amplified using
383 the following primers: AT1G55650 (fw 5'-CGAAGCGAAAAAGCACTCAC-3'; rv 5'-
384 CCTTTTACATAATCCGCGTTAAA-3'), AT2G28890 (fw 5'-TTTGTGGTTGGAGGTTGTGA-3'; rv 5'-
385 GTTGTTCGTGCCCATTTCTT-3'), AT5G04250 (fw 5'-AATTGACAAATGGTGTAATGGT-3'; rv 5'-
386 CCAAAGAATTTGTTTTTCTATTCC-3'), AT1G72000 (fw 5'-AACAAATATGCACAAGAAGTGC-3'; rv 5'-
387 ACCTAGCAAGCTGGCAAAAC-3'), AT3G18550 (fw 5'-TCCTTTTCCAAATAAAGGCATAA-3'; rv 5'-
388 AAATGAAAGAAATAAAAGGTAATGAGA-3'), AT2G20160 (fw 5'-TCCTAAATAAGGGAAGAGAAAGCA-3'; rv
389 5'-TGTTAGGTGAAACTGAATCCAA-3'), negative region (fw 5'-TGGTTTTGCTGGTGATGATG-3', rv 5'-

CCATGACACCAGTGTGCCTA-3'). HOT FIREPol EvaGreen qPCR Mix Plus (ROX) (Solis Biodyne) was used as a master mix for qPCR amplification in a iQ5 qPCR system (Bio-Rad).

For Sanger sequencing, positive genomic regions for PHE1 binding containing SNPs that allowed distinction between parents were amplified by PCR using the Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific), in combination with the primers described above. Amplified DNA was purified using the GeneJET PCR Purification kit (Thermo Fisher Scientific) and used for Sanger sequencing. The chromatograms obtained from Sanger sequencing were then analysed for the presence of SNPs. The maternal:total read ratios were retrieved from the publications where these genes were identified as imprinted.

Bioinformatic analysis of ChIP-seq data

For the *PHE1::PHE1-GFP* ChIP, reads were aligned to the *Arabidopsis* (TAIR10) genome using Bowtie version 1.2.2 (Langmead, 2010), allowing 2 mismatches (-v 2). Only uniquely mapped reads were kept. ChIP-seq peaks were called using MACS2 version 2.1.1, with its default settings (--gsize 1.119e8, --bw 250) (Zhang et al., 2008). Input samples served as control for their corresponding GFP ChIP sample. Each biological replicate was handled individually using the same peak calling settings, and only the overlapping peak regions between the two replicates were considered for further analysis. These regions are referred throughout the text as PHE1 binding sites (**Supplementary File 1**). Peak overlap was determined with BEDtools version 2.26.0 (Quinlan and Hall, 2010). Each PHE1 binding site was annotated to a genomic feature and matched with a target gene using the peak annotation feature (annotatePeaks.pl) provided in HOMER version 4.9 (Heinz et al., 2010) (**Supplementary File 1**). Only binding sites located less than 3 kb away from the nearest transcription start site were considered.

PHE1 DNA-binding motifs were identified from *PHE1::PHE1-GFP* ChIP-seq peak regions with HOMER's findMotifsGenome.pl function, using the default settings. P-values of motif enrichment, as well as alignments between PHE1 motifs and known motifs were generated by HOMER.

Read mapping, coverage analysis, purity calculations, normalisation of data, and determination of parental origin of reads derived from H3K27me3 ChIPs in 3x seeds was done following previously published methods (Moreno-Romero et al., 2016) (**Supplementary File 1**).

Analysis of PHE1 target genes

Significantly enriched Gene Ontology terms within target genes of PHE1 were identified using AtCOECIS (Vandepoele et al., 2009), and further summarized using REVIGO (Supek et al., 2011).

Enrichment of specific transcription factor families within PHE1 targets was calculated by first normalizing the number of PHE1-targeted TFs in each family, to the total number of TFs targeted by PHE1. As a control, the number of TFs belonging to a certain family was normalised to the total number of TFs in the *Arabidopsis* genome. The Log_2 fold change between these ratios was then calculated for each family. Significance of the enrichment was assessed using the hypergeometric test. Annotation of transcription factor families was done following the Plant Transcription Factor Database version 4.0 (Jin et al., 2014). Only TF families containing more than 5 members were considered in this analysis.

To determine which imprinted genes are targeted by PHE1, a custom list consisting of the sum of imprinted genes identified in different studies was used (**Fig. 1 – source data 1**) (Gehring et al., 2011; Hsieh et al., 2011; Pignatta et al., 2014; Schon and Nodine, 2017; Wolff et al., 2011).

To determine the proportion of genes overexpressed in paternal excess crosses that are targeted by PHE1, a previously published transcriptome dataset of 3x seeds was used (Schatlowski et al., 2014).

Spatial overlap of TEs and PHE1 binding sites

Spatial overlap between PHE1 ChIP-seq peak regions (binding sites) and TEs was determined using the regioneR package version 1.8.1 (Gel et al., 2015), implemented in R version 3.4.1 (Core Team R, 2017). As a control, a mock set of binding sites was created, to which we refer to as random binding sites. This random binding site set had the same total number of binding sites and the same size distribution as the PHE1 binding site set. Using regioneR, a Monte Carlo permutation test with 10000

iterations was performed. In each iteration the random binding sites were arbitrarily shuffled in the 3 kb promoter region of all *Arabidopsis thaliana* genes. From this shuffling, the average overlap and standard deviation of the random binding site set was determined, as well as the statistical significance of the association between PHE1 binding sites and TE superfamilies/families.

BedTools version 2.26.0 (Quinlan and Hall, 2010) was used to determine the fraction of PHE1 binding sites targeting MEGs, PEGs, or non-imprinted genes where a spatial overlap between binding sites, RC/Helitrons and PHE1 DNA-binding motifs is simultaneously observed. The hypergeometric test was used to assess the significance of the enrichment of PHE1 binding sites where this overlap is observed, across different target types.

Expression analyses of genes flanked by RC/Helitrons

We investigated the expression level of genes containing RC/Helitrons in their promoter regions (defined as 3kb upstream of the TSS), in embryo, endosperm, and seed coat of pre-globular stage seeds. Affymetrix GeneChip ATH1 Arabidopsis Genome Array data were extracted from Belmonte et al., (2013). The expression values in micropylar endosperm, peripheral endosperm, and chalazal endosperm were averaged to represent the endosperm expression level. The endosperm expression levels of a given gene were then compared to the expression levels in the embryo and seed coat.

RC/Helitron-associated genes were classified into three groups: (i) genes with RC/Helitrons and PHE1 DNA-binding motifs within the 3kb promoter, thus representing genes with domesticated RC/Helitrons; (ii) genes having RC/Helitrons and PHE1 DNA-binding motifs within the 3kb promoter, but not bound by PHE1; and (iii) genes having RC/Helitrons located within the 3kb promoter, and no PHE1 DNA-binding motifs. A two-tailed Mann-Whitney test with continuity correction was used to assess statistical significance of differences in expression levels between gene groups.

Calculation of PHE1 DNA-binding motif densities

466 To measure the density of PHE1 DNA-binding motifs within different genomic regions of interest,
467 the fasta sequences of these regions were first obtained using BEDtools. HOMER's
468 scanMotifGenomeWide.pl function was then used to screen these sequences for the presence of PHE1
469 DNA-binding motifs, and to count the number of occurrences of each motif. Motif density was then
470 calculated as the number of occurrences of each motif, normalized to the size of the genomic region
471 of interest. Motif densities in RC/Helitron consensus sequences were calculated as above. Perfect PHE1
472 DNA-binding motif sequences were defined as those represented in **Fig. 1a**. Nearly perfect motif
473 sequences were defined as sequences where only one nucleotide substitution could give rise to a
474 perfect PHE1 DNA-binding motif. Consensus sequences were obtained from Repbase (Bao et al., 2015).
475 Chi-square tests of independence were used to test if there were any associations between specific
476 genomic regions and PHE1 DNA-binding motifs. This was done by comparing the proportion of DNA
477 bases corresponding to PHE1 DNA-binding motifs in each genomic region.
478
479 *Identification of homologous PHE1 DNA-binding motifs carried by RC/Helitrons*
480 To assess the homology of PHE1 DNA-binding motifs and associated RC/Helitron sequences,
481 pairwise comparisons were made among all sequences, using the BLASTN program. The following
482 parameters were followed: word size = 7, match/mismatch scores = 2/-3, gap penalties, existence = 5,
483 extension = 2. The RC/Helitron sequences were considered to be homologous if the alignment covered
484 at least 9 out of the 10 bp PHE1 DNA-binding motif sites, extended longer than 30 bp, and had more
485 than 70% identity. Because the mean length of intragenomic conserved non-coding sequences is
486 around 30 bp in *A. thaliana* (Thomas et al., 2007), we considered this as the minimal length of
487 alignments to define a pair of related motif-carrying TE sequences. The pairwise homologous
488 sequences were then merged in higher order clusters, based on shared elements in the homologous
489 pairs.
490
491 *Phylogenetic analyses of PHE1-targeted PEG orthologs in the Brassicaceae*

Amino acid sequences and nucleotide sequences of PHE1-targeted PEGs were obtained from TAIR10. The sequences of homologous genes in the Brassicaceae and several other rosids were obtained in PLAZA 4.0 (https://bioinformatics.psb.ugent.be/plaza/) (Van Bel et al., 2018), BRAD database (http://brassicadb.org/brad/) (X. Wang et al., 2015), and Phytozome v.12 (https://phytozome.jgi.doe.gov/) (Goodstein et al., 2012).

For each PEG of interest, the amino acid sequences of the gene family were used to generate a guided codon alignment by MUSCLE with default settings (Edgar, 2004). A maximum likelihood tree was then generated by IQ-TREE 1.6.7 with codon alignment as the input (Nguyen et al., 2015). The implemented ModelFinder was executed to determine the best substitution model (Kalyaanamoorthy et al., 2017), and 1000 replicates of ultrafast bootstrap were applied to evaluate the branch support (Hoang et al., 2018). The tree topology and branch supports were reciprocally compared with, and supported by another maximum likelihood tree generated using RAxML v. 8.1.2 (Stamatakis, 2014).

We selected PEGs that had well supported gene family phylogeny with no lineage-specific duplication in the *Arabidopsis* and *Capsella* clades, and where imprinting data were available for all *Capsella grandiflora* (Cgr), *C. rubella* (Cru), and *Arabidopsis lyrata* (Aly) orthologs of interest (Klosinska et al., 2016; Lafon-Placette et al., 2018; Pignatta et al., 2014). We then obtained the promoter region, defined as 3 kb upstream of the TSS, of the orthologs and paralogs in Brassicaceae and rosids species. These promoter sequences were searched for the presence of homologous RC/Helitron sequences, as well as for putative PHE1 DNA-binding sites contained in these TEs.

Homology between *A. thaliana* RC/Helitron sequences and Brassicaceae sequences was detected by aligning the *A. thaliana* sequence to the promoter region of the orthologous PEG, using the BLASTN program, with the following parameters: word size = 11, match/mismatch scores = 2/-3, gap penalties, existence = 5, extension = 2. Aligned sequences are considered homologous if spanning more than 100 nt of the *Arabidopsis thaliana* RC/Helitron, with over 60% identity.

Epigenetic profiling of PHE1 binding sites

Parental-specific H3K27me3 profiles (Moreno-Romero et al., 2016) and DNA methylation profiles (Schatlowski et al., 2014) generated from endosperm of 2x seeds were used for this analysis. Levels of H3K27me3 and CG DNA methylation were quantified in each 50 bp bin across the 2 kb region surrounding PHE1 binding site centres using deepTools version 2.0 (Ramírez et al., 2016). These values were then used to generate H3K27me3 heatmaps and metagene plots, as well as boxplots of CG methylation in PHE1 binding sites. Clustering analysis of H3K27me3 distribution in PHE1 binding sites was done following the k-means algorithm as implemented by deepTools. A two-tailed Mann-Whitney test with continuity correction was used to assess statistical significance of differences in CG methylation levels.

Parental gene expression ratios in 2x and 3x seeds

To determine parental gene expression ratios in 2x and 3x seeds, we used previously generated endosperm gene expression data (Martinez et al., 2018). In this dataset, L*er* plants were used as maternal plants pollinated with wt Col or *osd1* Col plants, allowing to determine the parental origin of sequenced reads following the method described before (Moreno-Romero et al., 2019).

Parental gene expression ratios were calculated as the number of maternally-derived reads divided by the sum of maternally- and paternally-derived reads available for any given gene. Ratios were calculated separately for the two biological replicates of each cross (L*er* x wt Col and L*er* x *osd1* Col), and the average of both replicates was considered for further analysis. The MEG and PEG ratio thresholds for 2x and 3x seeds indicated in Figure 3d were defined as a four-fold deviation of the expected read ratios, towards more maternal or paternal read accumulation, respectively. The expected read ratio for a biallelically expressed gene in 2x seeds is 2 maternal reads : 3 total reads, while for 3x paternal excess seeds this ratio is 2 maternal reads : 4 total reads. Deviations from these expected ratios were used to classify the expression of published imprinted genes (**Fig. 1 – source data 1**) as maternally or paternally biased in 3x seeds, according to the direction of the deviation. As a control, the parental bias of these imprinted genes was also assessed in 2x seeds.

544

545 *Parental expression ratios of genes associated with H3K27me3 clusters*

546 Previously published endosperm gene expression data, generated with the INTACT system, was
547 used for this analysis (Del Toro-De León and Köhler, 2018). Parental gene expression ratios were
548 determined as the mean between ratios observed in the L*er* x Col cross and its reciprocal cross. As a
549 reference, the parental gene expression ratio for all endosperm expressed genes was also determined.
550 A two-tailed Mann-Whitney test with continuity correction was used to assess statistical significance
551 of differences between parental gene expression ratios.

552

553 *H3K27me3 accumulation in imprinted genes*

554 Parental-specific accumulation of H3K27me3 across imprinted gene bodies in the endosperm of 2x
555 (Moreno-Romero et al., 2016) and 3x seeds (this study) was estimated by calculating the mean values
556 of the H3K27me3 z-score across the gene length. Imprinted genes were considered as those genes
557 previously identified in different studies (Gehring et al., 2011; Hsieh et al., 2011; Pignatta et al., 2014;
558 Schon and Nodine, 2017; Wolff et al., 2011). A two-tailed Mann-Whitney test with continuity
559 correction was used to assess statistical significance of differences in H3K27me3 z-score levels.

560

561 *Statistics*

562 Sample size, statistical tests used, and respective p-values are indicated in each figure or figure
563 legend, and further specified in the corresponding Methods sub-section.

564

565 *Data availability*

566 ChIP-seq data generated in this study is available at NCBI's Gene Expression Omnibus database
567 (https://www.ncbi.nlm.nih.gov/geo/), under the accession number GSE129744. Additional data used
568 to support the findings of this study are available at NCBI's Gene Expression Omnibus, under the
569 following accession numbers: H3K27me3 ChIP-seq data from 2x endosperm (Moreno-Romero et al.,

2016) - GSE66585; Gene expression data in 2x and 3x endosperm (Martinez et al., 2018) - GSE84122; Gene expression data in 2x and 3x seeds and parental-specific DNA methylation from 2x endosperm (Schatlowski et al., 2014) - GSE53642; Parental-specific gene expression data of 2x INTACT-isolated endosperm nuclei (Del Toro-De León and Köhler, 2018) - GSE119915. Gene expression profile of different seed compartments (Belmonte et al., 2013) – GSE12404.

Acknowledgments

We thank Qi-Jun Chen for providing the pHEE401E CRISPR/Cas9 vector. We are grateful to Cecilia Wärdig for technical assistance. Sequencing was performed by the SNP&SEQ Technology Platform, Science for Life Laboratory at Uppsala University, a national infrastructure supported by the Swedish Research Council (VRRFI) and the Knut and Alice Wallenberg Foundation. This research was supported by a grant from the Swedish Research Council (to C.K.), a grant from the Knut and Alice Wallenberg Foundation (to C.K.), and support from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine.

Author contributions

R.A.B., J.M-R., Y.Q, D.D.F and C.K. performed the experimental design. R.A.B., J.M-R., J.V.B. and Y.Q. performed experiments. R.A.B., J.M-R., Y.Q., J.S-G., and C.K. analysed the data. R.A.B and C.K. wrote the manuscript. All authors read and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Materials and correspondence

The materials generated in this study are available upon request to C.K. (claudia.kohler@slu.se).

591 **References**

592 Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in
593 eukaryotic genomes. Mobile DNA. https://doi.org/10.1186/s13100-015-0041-9

594 Batista, R.A., Figueiredo, D.D., Santos-González, J., Köhler, C., 2019. Auxin regulates endosperm
595 cellularization in Arabidopsis. Genes & Development. https://doi.org/10.1101/gad.316554.118

596 Belmonte, M.F., Kirkbride, R.C., Stone, S.L., Pelletier, J.M., Bui, A.Q., Yeung, E.C., Hashimoto, M., Fei, J.,
597 Harada, C.M., Munoz, M.D., Le, B.H., Drews, G.N., Brady, S.M., Goldberg, R.B., Harada, J.J., 2013.
598 Comprehensive developmental profiles of gene activity in regions and subregions of the
599 Arabidopsis seed. Proceedings of the National Academy of Sciences 110, E435–E444.
600 https://doi.org/10.1073/pnas.1222061110

601 Bemer, M., Heijmans, K., Airoldi, C., Davies, B., Angenent, G.C., 2010. An atlas of type I MADS box gene
602 expression during female gametophyte and seed development in Arabidopsis. Plant Physiology
603 154, 287–300. https://doi.org/10.1104/pp.110.160770

604 Britten, R.J., Davidson, E.H., 1971. Repetitive and Non-Repetitive DNA Sequences and a Speculation on
605 the Origins of Evolutionary Novelty. The Quarterly Review of Biology 46, 111–138.
606 https://doi.org/10.1086/406830

607 Chuong, E.B., Rumi, M.A.K., Soares, M.J., Baker, J.C., 2013. Endogenous retroviruses function as
608 species-specific enhancer elements in the placenta. Nature Genetics 45, 325–329.
609 https://doi.org/10.1038/ng.2553

610 Clough, S.J., Bent, A.F., 1998. Floral dip: A simplified method for Agrobacterium-mediated
611 transformation of Arabidopsis thaliana. Plant Journal 16, 735–743.
612 https://doi.org/10.1046/j.1365-313X.1998.00343.x

613 Core Team R, 2017. R: A language and environment for statistical computing., R Foundation for
614 Statistical Computing. Vienna, Austria. https://doi.org/ISBN 3-900051-07-0

615 d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., Mercier, R., 2009. Turning Meiosis into
616 Mitosis. PLoS Biology 7, e1000124. https://doi.org/10.1371/journal.pbio.1000124

617 Del Toro-De León, G., Köhler, C., 2018. Endosperm-specific transcriptome analysis by applying the
618 INTACT system. Plant Reproduction 32, 55–61. https://doi.org/10.1007/s00497-018-00356-3
619 Dunn-Fletcher, C.E., Muglia, L.M., Pavlicev, M., Wolf, G., Sun, M.-A., Hu, Y.-C., Huffman, E.,
620 Tumukuntala, S., Thiele, K., Mukherjee, A., Zoubovsky, S., Zhang, X., Swaggart, K.A., Lamm, K.Y.B.,
621 Jones, H., Macfarlan, T.S., Muglia, L.J., 2018. Anthropoid primate–specific retroviral element
622 THE1B controls expression of CRH in placenta and alters gestation length. PLOS Biology 16,
623 e2006337. https://doi.org/10.1371/journal.pbio.2006337
624 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
625 Nucleic Acids Research 32, 1792–1797. https://doi.org/10.1093/nar/gkh340
626 Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Scheid, O.M., Hennig, L., Köhler, C., 2009.
627 Imprinting of the Polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. PLoS
628 Genetics 5, e1000663. https://doi.org/10.1371/journal.pgen.1000663
629 Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nature Reviews
630 Genetics 9, 397–405. https://doi.org/10.1038/nrg2337
631 Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L., Köhler, C., 2016. Auxin production in the
632 endosperm drives seed coat development in Arabidopsis. eLife 5, 1–23.
633 https://doi.org/10.7554/eLife.20542
634 Figueiredo, D.D., Batista, R.A., Roszak, P.J., Köhler, C., 2015. Auxin production couples endosperm
635 development to fertilization. Nature Plants 1, 15184. https://doi.org/10.1038/nplants.2015.184
636 Folter, S. de, Angenent, G.C., 2006. trans meets cis in MADS science. Trends in Plant Science 11, 224–
637 231. https://doi.org/10.1016/J.TPLANTS.2006.03.008
638 Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G., Berger, F., 2003. Arabidopsis haiku
639 mutants reveal new controls of seed size by endosperm. Plant Physiology 131, 1661–70.
640 https://doi.org/10.1104/pp.102.018762
641 Gehring, M., 2013. Genomic Imprinting: Insights From Plants. Annual Review of Genetics 47, 187–208.
642 https://doi.org/10.1146/annurev-genet-110711-155527

643 Gehring, M., Missirian, V., Henikoff, S., 2011. Genomic analysis of parent-of-origin allelic expression in
644 Arabidopsis thaliana seeds. PLoS ONE 6, e23687. https://doi.org/10.1371/journal.pone.0023687
645 Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., Malinverni, R., 2015. RegioneR: An
646 R/Bioconductor package for the association analysis of genomic regions based on permutation
647 tests. Bioinformatics 32, 289–291. https://doi.org/10.1093/bioinformatics/btv562
648 Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten,
649 U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant
650 genomics. Nucleic Acids Research 40, D1178–D1186. https://doi.org/10.1093/nar/gkr944
651 Goto, K., Meyerowitz, E.M., 1994. Function and regulation of the Arabidopsis floral homeotic gene
652 PISTILLATA. Genes & Development 8, 1548–1560. https://doi.org/10.1101/gad.8.13.1548
653 Gramzow, L., Theissen, G., 2010. A hitchhiker's guide to the MADS world of plants. Genome Biology.
654 https://doi.org/10.1186/gb-2010-11-6-214
655 Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M. a, Gagliano, W.B., 1998. Maternal control of
656 embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280, 446–450.
657 https://doi.org/10.1126/science.280.5362.446
658 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass,
659 C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-
660 Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell 38, 576–589.
661 https://doi.org/10.1016/j.molcel.2010.05.004
662 Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C., Casacuberta, J.M., 2014.
663 Extensive amplification of the E2F transcription factor binding sites by transposons during
664 evolution of Brassica species. The Plant Journal 77, 852–862. https://doi.org/10.1111/tpj.12434
665 Heyman, J., Van den Daele, H., De Wit, K., Boudolf, V., Berckmans, B., Verkest, A., Kamei, C.L.A., De
666 Jaeger, G., Koncz, C., De Veylder, L., 2011. Arabidopsis ULTRAVIOLET-B-INSENSITIVE4 Maintains
667 Cell Division Activity by Temporal Inhibition of the Anaphase-Promoting Complex/Cyclosome. The
668 Plant Cell 23, 4394–4410. https://doi.org/10.1105/tpc.111.091793

669 Hirsch, C.D., Springer, N.M., 2017. Transposable element influences on gene expression in plants.
670 Biochimica et Biophysica Acta 1860, 157–165. https://doi.org/10.1016/j.bbagrm.2016.05.010
671 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the
672 Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518–522.
673 https://doi.org/10.1093/molbev/msx281
674 Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada,
675 J.J., Zilberman, D., Fischer, R.L., 2011. Regulation of imprinted gene expression in Arabidopsis
676 endosperm. PNAS 108, 1755–62. https://doi.org/10.1073/pnas.1019273108
677 Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., Kinoshita, T., 2011. Rice interspecies
678 hybrids show precocious or delayed developmental transitions in the endosperm without change
679 to the rate of syncytial nuclear division. Plant Journal 65, 798–806.
680 https://doi.org/10.1111/j.1365-313X.2010.04466.x
681 Jiang, H., Moreno-Romero, J., Santos-González, J., De Jaeger, G., Gevaert, K., Van De Slijke, E., Köhler,
682 C., 2017. Ectopic application of the repressive histone modification H3K9me2 establishes post-
683 zygotic reproductive isolation in Arabidopsis thaliana. Genes & Development 31, 1272–1287.
684 https://doi.org/10.1101/gad.299347.117
685 Jin, J., Zhang, H., Kong, L., Gao, G., Luo, J., 2014. PlantTFDB 3.0: a portal for the functional and
686 evolutionary study of plant transcription factors. Nucleic Acids Research 42, D1182–D1187.
687 https://doi.org/10.1093/nar/gkt1016
688 Jullien, P.E., Berger, F., 2010. Parental genome dosage imbalance deregulates imprinting in
689 Arabidopsis. PLoS Genetics 6, e1000885. https://doi.org/10.1371/journal.pgen.1000885
690 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast
691 model selection for accurate phylogenetic estimates. Nature Methods 14, 587–589.
692 https://doi.org/10.1038/nmeth.4285
693 Kang, I.-H., Steffen, J.G., Portereiko, M.F., Lloyd, A., Drews, G.N., 2008. The AGL62 MADS domain
694 protein regulates cellularization during endosperm development in Arabidopsis. The Plant Cell

20, 635–47. https://doi.org/10.1105/tpc.107.055137

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., Fischer, R.L., 1999. Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. PNAS 96, 4186–4191. https://doi.org/10.1073/pnas.96.7.4186

Klosinska, M., Picard, C.L., Gehring, M., 2016. Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nature Plants 2. https://doi.org/10.1038/nplants.2016.145

Köhler, C., Page, D.R., Gagliardini, V., Grossniklaus, U., 2005. The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nature Genetics 37, 28–30. https://doi.org/10.1038/ng1495

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., Köhler, C., 2013. An Imprinted Gene Underlies Postzygotic Reproductive Isolation in Arabidopsis thaliana. Developmental Cell 26, 525–535. https://doi.org/10.1016/J.DEVCEL.2013.08.006

Lafon-Placette, C., Hatorangan, M.R., Steige, K.A., Cornille, A., Lascoux, M., Slotte, T., Köhler, C., 2018. Paternally expressed imprinted genes associate with hybridization barriers in Capsella. Nature Plants 4, 352–357. https://doi.org/10.1038/s41477-018-0161-6

Langmead, B., 2010. Aligning short sequencing reads with Bowtie. Current Protocols in Bioinformatics 11.7.1-11.7.14. https://doi.org/10.1002/0471250953.bi1107s32

Lu, J., Zhang, C., Baulcombe, D.C., Chen, Z.J., 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. PNAS 109, 5529–5534. https://doi.org/10.1073/pnas.1203094109

Luo, M., Dennis, E.S., Chaudhury, A., Peacock, W.J., Berger, F., 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. PNAS 102, 17531–17536. https://doi.org/10.1073/pnas.0508418102

Makarevich, G., Villar, C.B.R., Erilova, A., Köhler, C., 2008. Mechanism of PHERES1 imprinting in Arabidopsis. Journal of Cell Science 121, 906–912. https://doi.org/10.1242/jcs.023077

721 Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L.L., DeFraia, C.,
722 Slotkin, R.K., Köhler, C., 2018. Paternal easiRNAs regulate parental genome dosage in Arabidopsis.
723 Nature Genetics 50, 193–198. https://doi.org/10.1038/s41588-017-0033-4
724 Moreno-Romero, J., Del Toro-De León, G., Yadav, V.K., Santos-González, J., Köhler, C., 2019. Epigenetic
725 signatures associated with imprinted paternally expressed genes in the Arabidopsis endosperm.
726 Genome Biology 20, 41. https://doi.org/10.1186/s13059-019-1652-0
727 Moreno-Romero, J., Santos-González, J., Hennig, L., Köhler, C., 2017. Applying the INTACT method to
728 purify endosperm nuclei and to generate parental-specific epigenome profiles. Nature Protocols
729 12, 238–254. https://doi.org/10.1038/nprot.2016.167
730 Moreno-Romero, J., Jiang, H., Santos-González, J., Köhler, C., 2016. Parental epigenetic asymmetry of
731 PRC2-mediated histone modifications in the Arabidopsis endosperm. The EMBO Journal 35,
732 1298–1311. https://doi.org/10.15252/embj.201593534
733 Muiño, J.M., De Bruijn, S., Pajoro, A., Geuten, K., Vingron, M., Angenent, G.C., Kaufmann, K., 2016.
734 Evolution of DNA-binding sites of a floral master regulatory transcription factor. Molecular
735 Biology and Evolution 33, 185–200. https://doi.org/10.1093/molbev/msv210
736 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective
737 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and
738 Evolution 32, 268–274. https://doi.org/10.1093/molbev/msu300
739 Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., Gehring, M., 2014. Natural epigenetic
740 polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife 3.
741 https://doi.org/10.7554/eLife.03198
742 Pignatta, D., Novitzky, K., Satyaki, P.R. V., Gehring, M., 2018. A variably imprinted epiallele impacts
743 seed development. PLOS Genetics 14, e1007469. https://doi.org/10.1371/journal.pgen.1007469
744 Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features.
745 Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033
746 Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke,

747 T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic
748 Acids Research 44, W160–W165. https://doi.org/10.1093/nar/gkw257
749 Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D. a, Zhang, F., 2013. Genome engineering using the
750 CRISPR-Cas9 system. Nature Protocols 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143
751 Rebernig, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T., Köhler, C., 2015. Non-reciprocal
752 Interspecies Hybridization Barriers in the Capsella Genus Are Established in the Endosperm. PLoS
753 Genetics 11, e1005295. https://doi.org/10.1371/journal.pgen.1005295
754 Rodrigues, J.A., Zilberman, D., 2015. Evolution and function of genomic imprinting in plants. Genes and
755 Development 29, 2517–2531. https://doi.org/10.1101/gad.269902.115
756 Roszak, P., Köhler, C., 2011. Polycomb group proteins are required to couple seed coat initiation to
757 fertilization. PNAS 108, 20826–31. https://doi.org/10.1073/pnas.1117111108
758 Roth, M., Florez-Rueda, A.M., Städler, T., 2019. Differences in Effective Ploidy Drive Genome-Wide
759 Endosperm Expression Polarization and Seed Failure in Wild Tomato Hybrids. Genetics
760 genetics.302056.2019. https://doi.org/10.1534/genetics.119.302056
761 Satyaki, P.R.V., Gehring, M., 2019. Paternally Acting Canonical RNA-Directed DNA Methylation Pathway
762 Genes Sensitize Arabidopsis Endosperm to Paternal Genome Dosage. The Plant Cell 31, 1563–
763 1578. https://doi.org/10.1105/TPC.19.00047
764 Schatlowski, N., Wolff, P., Santos-Gonzalez, J., Schoft, V., Siretskiy, A., Scott, R., Tamaru, H., Kohler, C.,
765 2014. Hypomethylated Pollen Bypasses the Interploidy Hybridization Barrier in Arabidopsis. The
766 Plant Cell 26, 3556–3568. https://doi.org/10.1105/tpc.114.130120
767 Schon, M.A., Nodine, M.D., 2017. Widespread Contamination of Arabidopsis Embryo and Endosperm
768 Transcriptome Data Sets. The Plant Cell 29, 608–617. https://doi.org/10.1105/tpc.16.00845
769 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
770 phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
771 Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. Revigo summarizes and visualizes long lists of gene
772 ontology terms. PLoS ONE 6, e21800. https://doi.org/10.1371/journal.pone.0021800

773 Thomas, B.C., Rapaka, L., Lyons, E., Pedersen, B., Freeling, M., 2007. Arabidopsis intragenomic
774 conserved noncoding sequence. PNAS 104, 3348–53. https://doi.org/10.1073/pnas.0611574104
775 Tiwari, S., Spielman, M., Schulz, R., Oakey, R.J., Kelsey, G., Salazar, A., Zhang, K., Pennell, R., Scott, R.J.,
776 2010. Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana.
777 BMC Plant Biology 10, 72. https://doi.org/10.1186/1471-2229-10-72
778 Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., Vandepoele, K.,
779 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant
780 genomics. Nucleic Acids Research 46, D1190–D1196. https://doi.org/10.1093/nar/gkx1002
781 Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., Van de Peer, Y., 2009. Unraveling
782 Transcriptional Control in Arabidopsis Using cis-Regulatory Elements and Coexpression Networks.
783 Plant Physiology 150, 535–546. https://doi.org/10.1104/pp.109.136028
784 Villar, C.B.R., Erilova, A., Makarevich, G., Trösch, R., Köhler, C., 2009. Control of PHERES1 Imprinting in
785 Arabidopsis by Direct Tandem Repeats. Molecular Plant 2, 654–660.
786 https://doi.org/10.1093/MP/SSP014
787 Wang, G., Jiang, H., Del Toro de León, G., Martinez, G., Köhler, C., 2018. Sequestration of a Transposon-
788 Derived siRNA by a Target Mimic Imprinted Gene Induces Postzygotic Reproductive Isolation in
789 Arabidopsis. Developmental Cell 46, 696-705.e4. https://doi.org/10.1016/J.DEVCEL.2018.07.014
790 Wang, X., Wu, J., Liang, J., Cheng, F., Wang, Xiaowu, 2015. Brassica database (BRAD) version 2.0:
791 integrating and mining Brassicaceae species genomic resources. Database 2015, bav093.
792 https://doi.org/10.1093/database/bav093
793 Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., Chen, Q.-J., 2015. Egg cell-specific
794 promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target
795 genes in Arabidopsis in a single generation. Genome Biology 16, 144.
796 https://doi.org/10.1186/s13059-015-0715-0
797 Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., Köhler, C., 2010. H3K27me3 profiling of the
798 endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS

799 Genetics 6, 1–14. https://doi.org/10.1371/journal.pgen.1001152

800 Wolff, P., Jiang, H., Wang, G., Santos-González, J., Köhler, C., 2015. Paternally expressed imprinted
801 genes establish postzygotic hybridization barriers in Arabidopsis thaliana. eLife 4.
802 https://doi.org/10.7554/eLife.10074

803 Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M.,
804 Rehmsmeier, M., Köhler, C., 2011. High-Resolution Analysis of Parent-of-Origin Allelic Expression
805 in the Arabidopsis Endosperm. PLoS Genetics 7, e1002126.
806 https://doi.org/10.1371/journal.pgen.1002126

807 Xing, Q., Creff, A., Waters, A., Tanaka, H., Goodrich, J., Ingram, G.C., Machida, Y., 2013. ZHOUPI controls
808 embryonic cuticle formation via a signalling pathway involving the subtilisin protease ABNORMAL
809 LEAF-SHAPE1 and the receptor kinases GASSHO1 and GASSHO2. Development 140, 770–9.
810 https://doi.org/10.1242/dev.088898

811 Yang, S., Johnston, N., Talideh, E., Mitchell, S., Jeffree, C., Goodrich, J., Ingram, G., 2008. The
812 endosperm-specific ZHOUPI gene of Arabidopsis thaliana regulates endosperm breakdown and
813 embryonic epidermal development. Development 135, 3501–9.
814 https://doi.org/10.1242/dev.026708

815 Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M.,
816 Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9,
817 R137. https://doi.org/10.1186/gb-2008-9-9-r137

818 Zhao, Y., 2012. Auxin Biosynthesis: A Simple Two-Step Pathway Converts Tryptophan to Indole-3-Acetic
819 Acid in Plants. Molecular Plant 5, 334–338. https://doi.org/10.1093/MP/SSR104

820