TE nearby a gene introduces regulatory sequences that alter the expression of the gene in a way that is advantageous. For example, many traits that have been selected during crop domestication involve regulatory changes caused by TE insertions, including anthocyanin production in

immunoprecipitation experiments indicate that the core PRC2 is associated with several plant-specific PcG proteins that share common mutant phenotypes with the PRC2 members. Thus in Arabidopsis, various proteomic studies show that the chromodomain-containing protein LIKE HETEROCHROMATIN PROTEIN 1 (LHP1)/TERMINAL FLOWER 2 (TFL2), is a PRC2 component [18, 26, 29]. LHP1 is homologous to HETEROCHROMATIN PROTEIN 1 (HP1) of animals, but its function has diverged as it binds H3K27me3 unlike HP1 which binds the heterochromatic mark H3K9me2. EMBRYONIC FLOWER1 (EMF1) is a plant-specific PcG protein whose function is less clear, but biochemical studies suggest it may play a role in chromatin compaction and

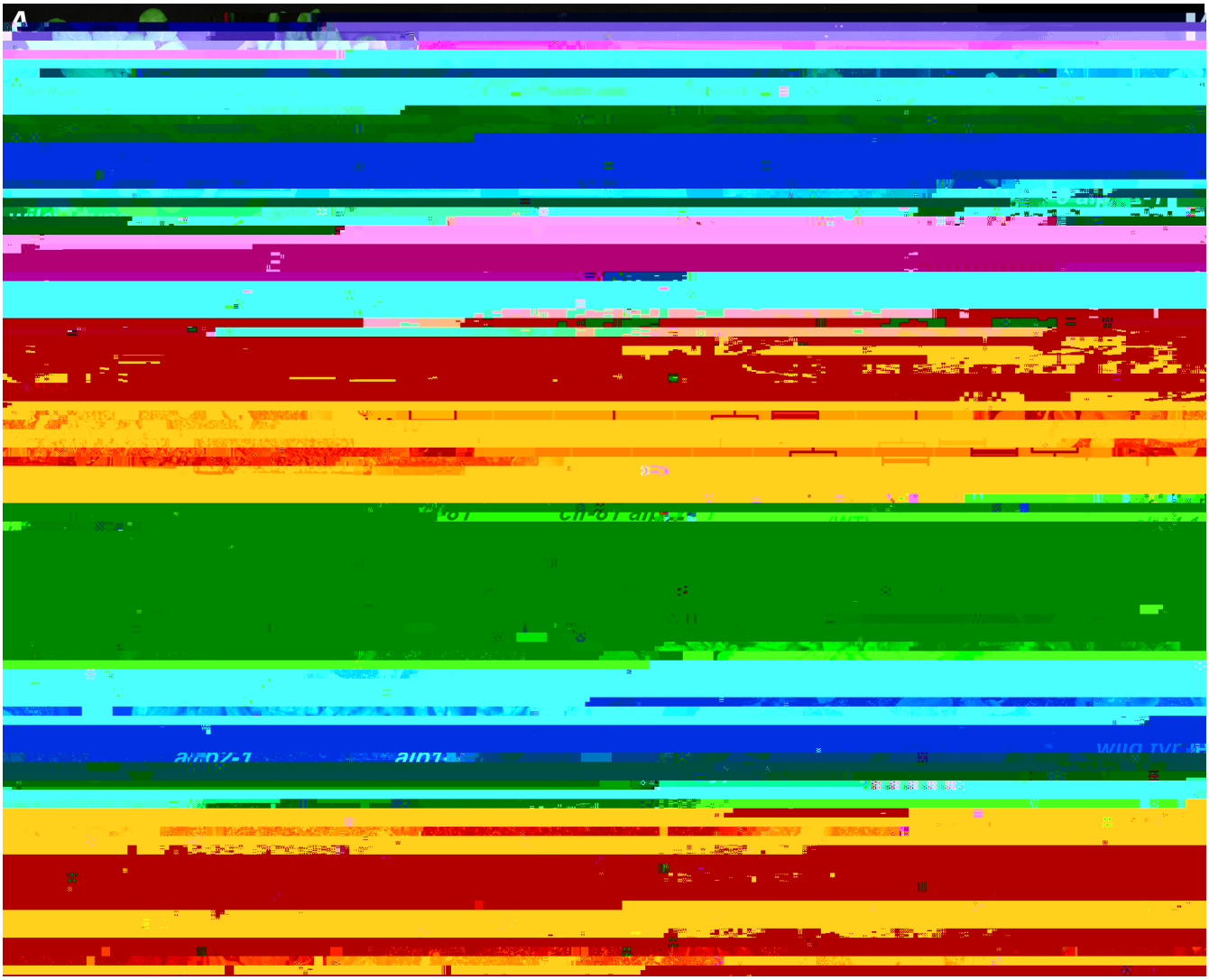


Fig 1. ***ALP2*** acts in the same pathway as ***ALP1***. A. Rosette phenotypes of 9 week old plants grown in short days. B. Box plots showing median and inter

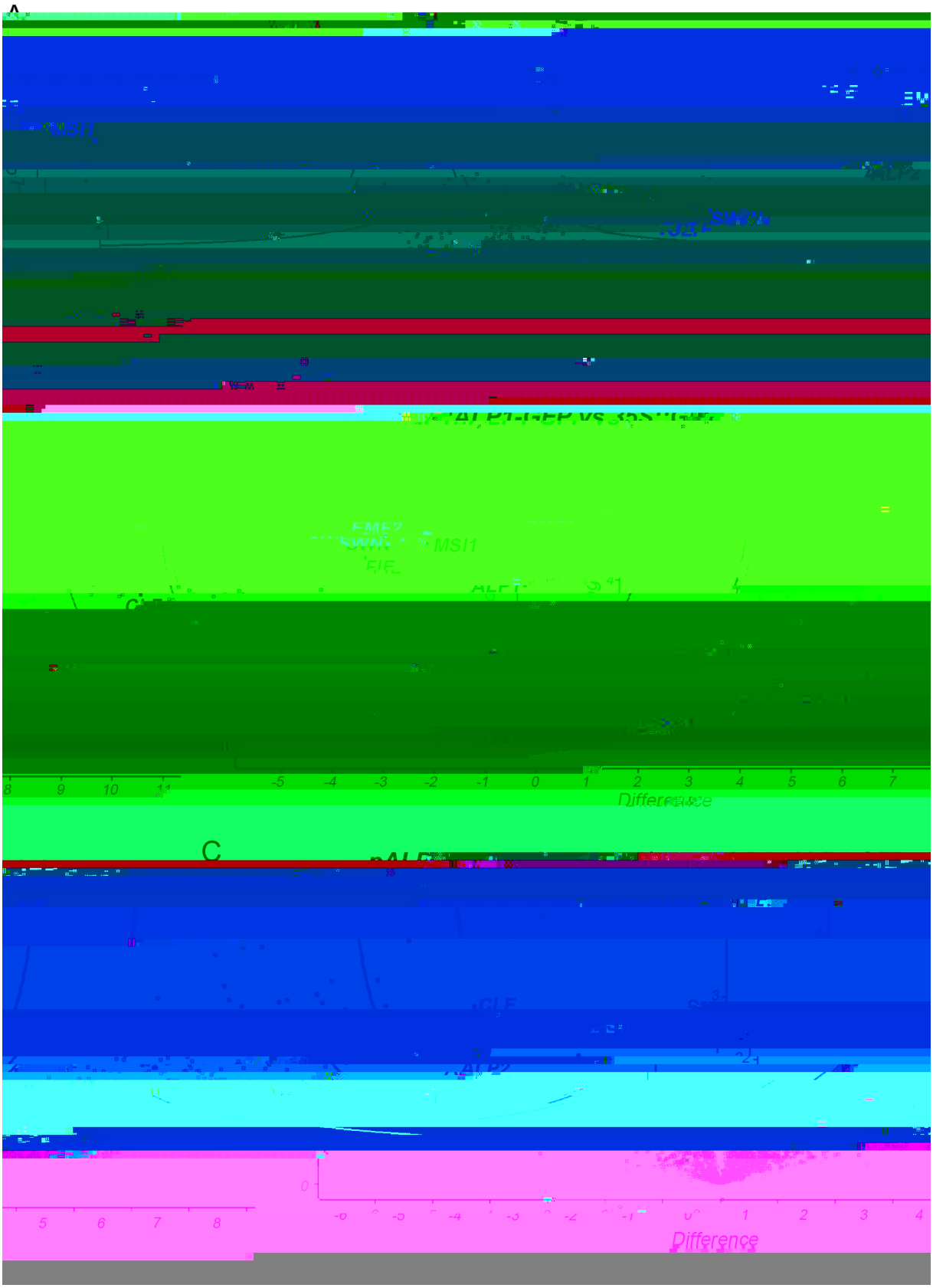
mutant plants there was no further reduction in *SEP3* or *FT* expression relative to the *lhp1 alp* double mutants. Together these results showed that *alp1* and *alp2* have similar mutant phenotypes and that in the absence of *ALP1* activity there was no further effect of withdrawing *ALP2* activity, so *ALP1* and *ALP2* act in a common genetic pathway. Consistent with this, *ALP2* was widely expressed in plant tissues similar to *ALP1* and *CLF* (Fig 2B).

Although the *alp* mutations had a strong effect in PcG mutant backgrounds, their phenotypes were weaker in wild-type (*LHP1*$^{+}$) backgrounds. However, both *alp1-1* and *alp2-1* mutations were slightly late flowering relative to their wild-type (Col-0) progenitors (Fig 1B and 1D) and had mild vegetative phenotypes in which older leaves were curled slightly downward

(Fig 1E). To further characterize the effects of the *ALP* genes on floral induction, we grew plants in non-inductive

GFP-ALP2 fusion (S4 Fig). We immunoprecipitated protein extracts from inflorescence tissue of *35S*::*GFP-ALP2*, *pALP1*::*ALP1-GFP* and *35S*::*GFP* plants using an antibody to the GFP tag and identified the proteins isolated by mass spectrometry (IP-MS). Consistent with ALP2 interacting with ALP1, we identified numerous ALP1 peptides when ALP2 was immunoprecipitated, although a few (0±3) ALP1 peptides were also found in the *35S*::*GFP* control; in the ALP1 IP, abundant ALP2 peptides were detected and none were found in the *35S*::*GFP* control

(Table 1 and S2 Dataset). Strikingly, the ALP2 IP clearly contains all four core members of the Arabidopsis PRC2 (i.e., FIE, MSI1,


A
MSI1
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7
Difference
C
5 6 7 8
-5 -4 -3 -2 -1 0 1 2 3 4
Difference

similar domains and if so whether they can also interact with ALPs. To address this we performed yeast two hybrid assays using the Myb DNA binding protein and the nuclease protein (also termed ORF1 and TPase, respectively) encoded by the rice *Ping* TE. Consistent with the fact that both Ping proteins are needed to catalyse transposition, we found that their full length proteins interacted in yeast (S8 Fig). To delimit the interacting regions of the Ping proteins further, we made truncations and found that a N-terminal half of the

Arafew12 11 0.6 0 Td (opeptidoT12 1ts55860 Td (tfroT

Although full-length ALP1 and ALP2 proteins did not interact with any PRC2 member in yeast (S6 Fig), an N-terminal portion of ALP2 interacted with MSI1 as a bait but not a prey fusion (Fig 5C). As an additional confirmation of the ALP2 interaction with MSI1, we co-expressed epitope-tagged versions of ALP2 and MSI1 in insect cell cultures using a baculoviral expression system and performed pulldown assays. This showed that ALP2 could pull down MSI1, whereas in control experiments co-expressing ALP1 and MSI1 no interaction of ALP1 with MSI1 was found (Fig 7). Collectively, these experiments indicate that ALP2 but not ALP1 contacts the PRC2 via the MSI1 core member.

The above results suggested that the association of ALP1 with the PRC2 might be indirect and rely on ALP2 as a bridge. To test this, we first asked whether expressing ALP2 in tobacco could facilitate an interaction of ALP1 with MSI1 in BiFC assays. When tobacco leaves were co-infiltrated with a mixture of Agrobacterium cells harbouring *35S*::*ALP1-nYFP* and *35S*::*cYFP-MSI1* constructs, no fluorescence was observed. By contrast, when we included a third Agrobacterium strain harbouring *35S*::*ALP2*, many cells showed fluorescence indicating that ALP2 could cause ALP1 and MSI1 to associate closely with one another (Fig 6 and S7 Fig). To confirm that these interactions were relevant in Arabidopsis, we introduced an *pALP1*::*ALP1-GFP* transgene into the *alp2-1* mut iTd (pistidown)Tj 7.2925 0 Td (abidopsis)...

results indicate that ALP2 recruits ALP1 to PRC2 via direct interaction with the MSI1 component.

Analysis of H3K27me3 enrichment in ***alp*** mutants

Since *ALP1* and *ALP2* acts antagonistically to the PRC2 by de-repressing PcG target genes (Fig 2), we analysed the effects on H3K27me3 deposition at target gene chromatin using chromatin immunoprecipitation (ChIP) from seedling extracts. We found that in *clf-28* mutants, consistent with previous studies [29, 39], H3K27me3 levels at selected PcG targets were decreased relative to the wild type, whereas the *alp* mutants showed no consistent change (Fig 8A and 8B). We next investigated whether the partial suppression of the *clf* phenotype in *clf alp* double mutants (Fig 1) might reflect a restoration of H3K27me3 at PcG targets, which may be caused, for example, by an increase in SWN-PRC2 activity when *ALP* function is removed. To this

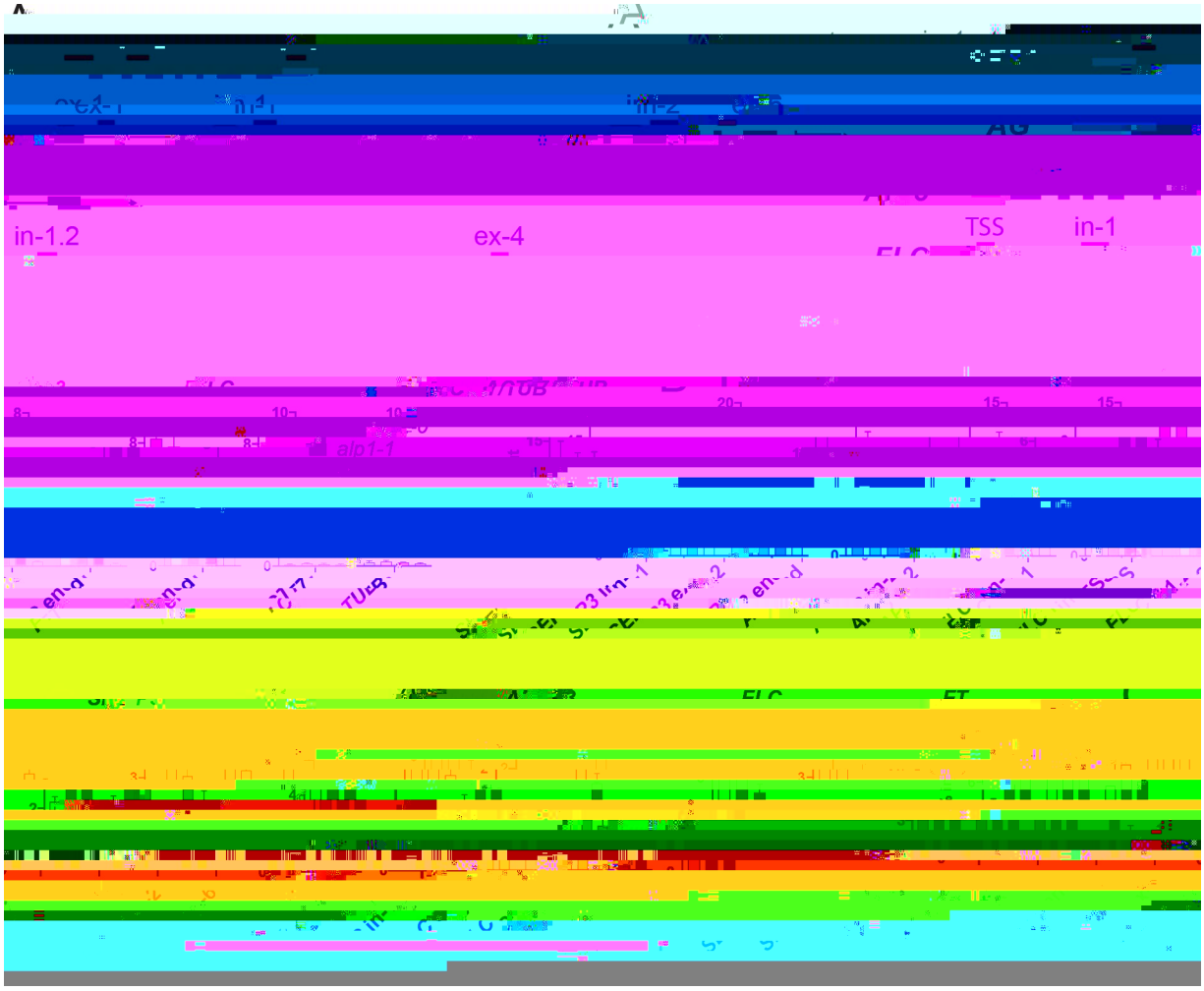


Fig 8. ChIP analysis of H3K27me3 enrichment at loci known to be affected by PRC2-CLF activity. A. Schematics of genomic regions examined during ChIP experiments. Approximate regions amplified via real-time PCR are denoted by red bars. Blue bars: untranslated regions; black bars: exons; black lines: introns; arrow heads: direction of transcription. B. Immunoprecipitation of chromatin prepared from 10-day-old Col-0, *alp1-1*, *alp2-1* and *clf-28* seedlings. An H3K27me3 antibody was used to precipitate chromatin, and enrichments are displayed as percentage input. C. Immunoprecipitation of chromatin from *clf-28*, *alp1-1 clf-28* and *alp2-1 clf-28* 10 day old seedlings. Three biological replicates were used for each genotype and error bars indicate the standard deviation from the mean. Primers that amplify regions in *ACT7* (AT5G[illegible]Tj 2.9375 We)Tj 1.60l4 /F12 1 TTd (bar & tes)Tj 3.9614 0 t 4.45.

end, we compared H3K27me3 levels in *clf-28*, *clf-28 alp1-1* and *clf-28 alp2-1* mutants by initially testing the same genomic regions used for Fig 8B and then additional regions for genes that appeared to show differences. We found that H3K27me3 levels at *SEP3*, *AG* and *FLC* were generally higher in the double mutants than in *clf-28* (Fig 8C), whereas little change was observed for *FT* and *AP3*. Thus, the incomplete suppression of the *clf* phenotype by *alp* mutations8B ( )$ 40148 0 Td (the)T

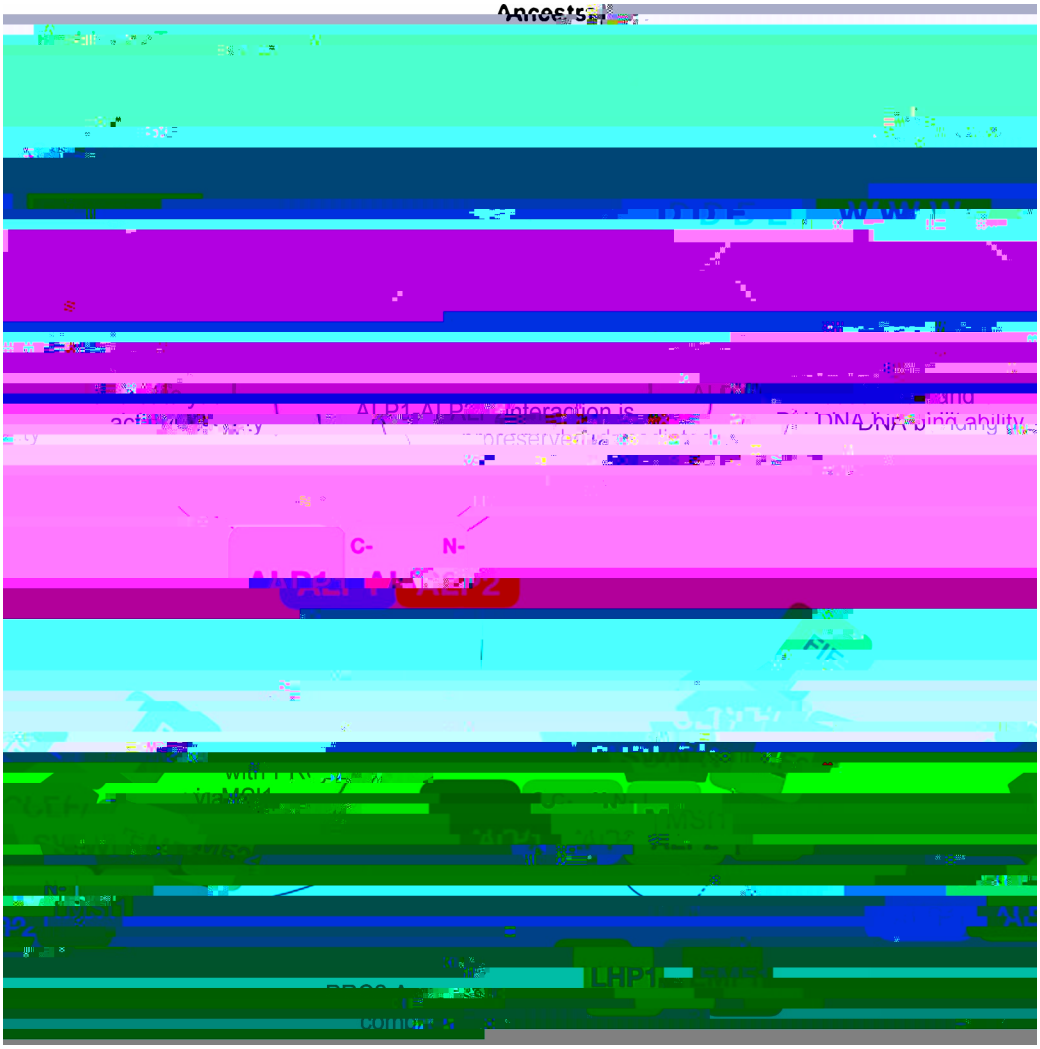


Fig 9. Model for evolution of ALP-PRC2. The ancestral transposases accumulated mutations that disabled their ability to catalyse transposition but maintained their interaction. The interaction of ALP2 with MSI1 displaced EMF1 and LHP1, which also interact with MSI1, leading to mutually exclusive PRC2 subcomplexes.

https://doi.org/10.1371/journal.pgen.1008681.g009

ALP2 can interact with both MSI1 and ALP1, via distinct N- and C-terminal regions respectively, it potentially recruits ALP1 to PRC2. Consistent with this, we found that only when ALP2 was co-expressed with ALP1-nYFP and cYFP-MSI1 in tobacco leaves was an ALP1-MSI1 interaction observed in BiFC assays, presumably because ALP2 brought the two proteins into close proximity. Additionally, immunoprecipitation experiments showed that *ALP2* activity was needed in order for ALP1 to associate with PRC2 components in Arabidopsis inflorescences, supporting that the interactions observed in the heterologous systems (tobacco and yeast) are biologically relevant.

ALP1 and ALP2 form a distinct PRC2 complex

A distinctive feature of the ALP-PRC2 complex is that it lacks the accessory components that are commonly observed in pull down experiments involving core members of the PRC2. For example, IP-MS experiments variously using CLF, EMF2, MSI1 or FIE as baits have repeatedly identified EMF1, LHP1, VRN5 and VEL1 [18, 26, 29]. By contrast, IP-MS with ALP1 or ALP2

fluorescent strip lights on a compost mixture containing Levington's F2, Perlite, horticultural sand (in a ratio of 15:6:4). Plants for the IP-MS experiments were grown in controlled environment cabinets at 16ÊC in continuous light. All mutants were in Col-0 genetic background. The *alp1-1* and *alp2-1* mutants were identified in the genetic screen described in [35], the *alp2-2* mutation is a T-DNA insertion (Salk_150532) in the first exon of *ALP2*. The *pALP1*::*ALP1-GFP* and *35S*::*ALP1-GFP* transgenic lines were described previously [18].

Total RNA extraction, cDNA synthesis and Gene Expression Analysis

Total RNA was extracted from 0.1 g of plant material, ground to fine powder in N_2(l), using RNeasy Plant Mini Kit (Qiagen # 74904) according to the manufacturer's instructions, including the steps of on-column DNA digestion with RNase-Free DNase (Qiagen # 79254). First strand cDNA was synthesised using 1 µg total RNA primed with an anchored oligo dT primer and Superscript III reverse transcriptase (Thermo Fisher # C18080044-X) as described in product's documentation. Quantification of gene expression by real time PCR was performed using a LightCycler480 (Roche) and PCR reactions containing 1X SYBR master premix (Roche) as described in [18]. Crossing point (Ct) values were determined using the 2nd derivative max method in the LightCycler480 (Roche) software. For each primer set, the amplification efficiency was checked by analysis of a calibration curve made using a ten-fold dilution series (10^0–10^{-3}) of cDNA again using the supplied software. The crossing point (Ct) values were converted to relative expression or enrichment levels manually in Excel using the 2^{-Ct} method. Gene expression was normalised relative to the *EiF4A1* reference gene. The primers used to amplify target genes are listed in

with 40% or more gaps were removed). The phylogenetic trees were produced using the Mr Bayes program implemented on the CIPRES server (https://www.phylo.org/) with a mixed amino acid model, substitution rates modelled using invgamma ngamma = 4. The trees were run for up to 10 million generations

energy of 27 (Olsen *et al*, 2007). The isolation window in the quadrupole was set at 1.4 Thomson. Only ions with charge between 2 and 7 were selected for MS2. The MaxQuant software platform [54] version 1.6.6.0 was used to process raw files and search was conducted against the complete *Arabidopsis thaliana* database (Uniprot, released October 2016), using the Andromeda search engine [54]. The first search peptide tolerance was set to 20 ppm while the main search peptide tolerance was set to 4.5 pm. Isotope mass tolerance was 2 ppm and maximum charge to 7. Maximum of two missed cleavages were allowed. Carbamidomethylation of cysteine was set as fixed modification. Oxidation of methionine and acetylation of the N-terminal as well as the Gly-Gly (diglycil) on lysine were set as variable modifications. For peptide and protein identifications FDR was set to 1%. For the volcano plot statistical analysis, the spectra obtained from the mass spectrometric analysis of the Trypsin-digested IP-MS experimental samples were searched against the *Arabidopsis thaliana* proteome database (UniprotKB_aratha downloaded on 15.03.2015) and label free quantification performed using the MaxQuant programme [55] and the following search parameters: Digestion mode = Trypsin; Variable modifications = Acetyl (protein N terminus), Oxidation (M); Maximum missed cleavages = 4. The output of this analysis was analysed by a two sample T-test using Perseus software [56] to find proteins whose increased abundance in a given sample relative to the negative control (e.g. in sample expressing GFP-tagged protein of interest relative to the sample expressing GFP alone) was statistically significant.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018911

Tandem Affinity Purification (TAP) and MS of GS^{rhino}-ALP2

The constructs expressing *35S::GSrhino-ALP2* were generated by Gateway recombination and PSB-D suspension cultures were transformed as previously described [38]. The tandem affinity purification followed the protocol described in [38]

the cells in a bench-top microcentrifuge for 30 sec at maximum speed. The supernatant was removed and the cells were washed by resuspending in 1 ml of sterile dH_2O. Cells were then pelleted again (30 sec max speed) and resuspended in adequate volume of 1x TE buffer, calculated from the OD_{600} value of the original culture, so that for the resulting yeast suspension $OD_{600} = 1$.

Two 1/10 serial dilutions of every yeast suspension were prepared and 8 μl were immediately spotted on the different plates along with 8 μl of the undiluted sample. Non selective YMM + CSM-L-W plates were used as a control for the viability of the spotted yeast. Low stringency YMM + CSM-L-W-H (Anachem # 4530±112) plates were used to detect weak interactions and high stringency YMM + CSM-L-W-H-A (Anachem # 4540±412) plates, with or without Aureobasidin (Clontech # 630466), were employed for detection of stronger interactions.

BiFC assay

The constructs for expression of *CLF* and *SWN* in BiFC assays were described previously[28]. For *FIE* the coding sequenc 4.42245 (sequenc 3rm)Tj e

the pFASTIN vector backbone. Primers were designed using the online assembly tool http://nebuilder.neb.com/#!/ and the assembly itself was performed with the NEBuilder HiFi DNA Assembly Cloning Kit (NEB # E5520S) following the suggested protocols. The recombinant pFASTIN constructs bearing the ALPs-SSF fusions were verified by sequencing and then introduced into DH10Bac cells by heat shock transformation. Putative transformants (white colonies) were selected (blue/white selection) on Bluo-Gal (Thermo Fisher Scientific) containing LB agar plates supplemented with appropriate antibiotics. Recombinant bacmids were isolated from putative transformants via standard miniprep procedure and successful integration of the ALPs-SSF fusions was confirmed via PCR.

Spodoptera frugiperda Sf9 cells were grown in suspension cultures in HyClone CCM3 medium (GE # SH30065.02) at 27ÊC and 125 rpm (1º orbit). Cell density was maintained between 0.5±6x10^6 cells/ml. To generate baculovirus stocks, 1.8x10^6 Sf9 cells were plated in a 35-mm tissue culture dish and transfected with 1µg of recombinant bacmid using X-treme-GENE HP DNA Transfection Reagent (Roche # 06 366 236 001) following the manufacturer's instructions. After 4±5 days, cell culture supernatants containing recombinant baculovirus particles were clarified by centrifugation at 1000 x g for 5 min at room temperature and used for further titer amplification in suspension cultures at a density of 2x10^6 cells/ml, using initially a 1:20 dilution of the clarified supernatant, followed by 2±3 sequential rounds of amplification with 1:100 dilutions of the previous virus stock.

For protein expression and Pull-Down assays, Sf9 cells at a density of 4x10^6 cells/ml were inoculated with the highest titer baculovirus stocks at a dilution of 1:60 each of either 6xHIS-MSI1 and ALP1-SSF, or 6xHIS-MSI1 and SSF-ALP2. After 65 h, cells were harvested via centrifugation at 500 x g for 5 min at room temperature. The pellet was either frozen in LN_2 and stored at -80 ^oC or used immediately for protein extraction. Total protein was extracted by resuspending the cell pellet in 8 ml of pre-chilled lysis buffer (150 mM NaCl, 10% v/v Glycerol, 50 mM Tris HCl pH 8, 0.1% v/v Triton-X-100, 1 mM DTT, Protease inhibitors cocktail tablets, 1 mM EDTA). The cell suspension was subjected to three 30-s sonication cycles, 35% amplitude, 1 s Pulse ON and 0.1 s Pulse OFF, and subsequently centrifuged at 40.000 x g for 30 min at 4 ^{o}

Supporting information

S1 Fig. Induction of flowering genes in day-shifted plants. RT-qPCR of *SEP3*, *AG*, and *PI* in SAM-enriched tissue of Col-0, *alp1-1*, and *alp2-1* plants that were grown under short day conditions (8 hr light, 16 hr dark) for 14 days, and then grown under long day conditions (16 hr light, 8 hr dark). Tissue was collected on the days specified after the photoperiod shift 1 hr prior to the commencement of the dark period (ZT 15). Values are the mean of three biological replicates and are presented relative to the reference gene *PP2AA3* (*AT1G13320*). Error bars indicate the standard deviation of the mean
(PDF)

S2 Fig. Phylogeny of ALP2, HDP2 and Harbinger transposases. An unrooted Bayesian phylogenetic tree was made based on alignment of the amino acid sequences. Numbers at the branches indicate probabilities (percent), the scale bar indicates the average number of substitutions per site. ALP2 sequences are indicated in red, *Harbinger* transposases in blue and HDP2 sequences in green.
(PDF)

S3 Fig. Synteny between ***A***. ***thaliana*** and ***A***. ***lyrata*** in ***ALP2*** region. In *A*. *thaliana*, *ALP2* is neighboured by *At5G24490* and *At5G24510* encoding 30S and 60S ribosomal proteins, respectively. The *A*. *lyrata ALP2* orthologue (*Al ALP2*) is located in a syntenous region, i.e. is flanked by orthologous genes with the same arrangement and orientation as in *A*. *thaliana*.
(PDF)

S4 Fig. ALP2-GFP protein expression. Immunoblot analysis of total protein extracts using -GFP antibodies. A protein with the predicted size (65.2 Kda) for the ALP2-GFP fusion is specifically detected in extracts from *35S*::*GFP-ALP2* plants. *35S*::*LHP1-GFP* is included as a positive control.
(PDF)

S5 Fig. The core PRC2 co-purifies with GS^{rhino}-ALP2. Volcano plot analysis with the x axis showing relative abundance of proteins identified in tandem affintity purified *35S*:*GSrhino-ALP2* samples relative to non transgenic control. The y axis shows probability values. Analysis is based on three biological replicates. ALP proteins highlighted in red, PRC2 components in blue.
(PDF)

S6 Fig. Yeast two hybrid assays for ALP protein interactions. A. Truncated forms of ALP1 did not interact with ALP2. The interaction of a C-terminal region of ALP2 (residues 171± 261) is included as a control. B. Full length ALP1 and ALP2 proteins did not interact with core PRC2 components.
(PDF)

S7 Fig. ALP protein interactions in BiFC assays. Low magnification images showing epidermal cells from *N*. *benthamiana* leaves transformed by infiltration with Agrobacterium. Images in left of panels are YFP channel, on right is merge of light field and fluorescence channels. ALP proteins are tested with each other and with the core PRC2 components. Images

dilutions of five pooled transformants were spotted onto selective media.
(PDF)

S9 Fig. Interaction of truncated Ping and ALP proteins in yeast two hybrid assays. A N-terminal fragment (amino acids 1±223) of the Ping nuclease interacts with the full length Ping Myb DNA binding protein reciprocally. A C-terminal fragment (amino acids 292±465) of the Ping Myb DNA binding protein interacts

8. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene *TB1*. Nature Genetics. 2011; 43(11):1160±3. https://doi.org/10.1038/ng.942 PMID: 21946354.

9. Jangam D, Feschotte C, Betran E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. 2017; 33(11):817±31. https://doi.org/10.1016/j

27. Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, et al. *Cis* and *trans* determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. Nature Genetics. 2017. https://doi.org/10.1038/ng.3937 PMID: 28825728.

28. Zhou Y, Wang Y, Krause K, Yang T, Dongus JA, Zhang Y, et al. Telobox Yang

46. Jiang N,