Divergence of regulatory networks governed by the orthologous transcription factors FLC and PEP1 in Brassicaceae species

Julieta L. Mateos[a,1,2], Vicky Tilmes[a,1], Pedro Madrigal[b,3], Edouard Severing[a], René Richter[a], Colin W. M. Rijkenberg[a], Paweł Krajewski[b], and George Coupland[a,4]

[a]Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, D-50829 Cologne, Germany; and [b]Department of Biometry and Bioinformatics, Institute of Plant Genetics, Polish Academy of Sciences, 60-479 Poznań, Poland

Contributed by George Coupland, November 6, 2017 (sent for review November 7, 2016; reviewed by Kerstin Kaufmann and Hao Yu)

Genome-wide landscapes of transcription factor (TF) binding sites (BSs) diverge during evolution, conferring species-specific transcriptional patterns. The rate of divergence varies in different metazoan lineages but has not been widely studied in plants. We identified the BSs and assessed the effects on transcription of FLOWERING LOCUS C (FLC) and PERPETUAL FLOWERING 1 (PEP1), two orthologous MADS-box TFs that repress flowering and confer vernalization requirement in the Brassicaceae species *Arabidopsis thaliana* and *Arabis alpina*, respectively. We found that only 14% of their BSs were conserved in both species and that these contained a CArG-box that is recognized by MADS-box TFs. The CArG-box consensus at conserved BSs was extended compared with the core motif. By contrast, species-specific BSs usually lacked the CArG-box in the other species. Flowering-time genes were highly overrepresented among conserved targets, and their CArG-boxes were widely conserved among Brassicaceae species. Cold-regulated (COR) genes were also overrepresented among targets, but the cognate BSs and the identity of the regulated genes were usually different in each species. In cold, COR gene transcript levels were increased in *flc* and *pep1-1* mutants compared with WT, and this correlated with reduced growth in *pep1-1*. Therefore, FLC orthologs regulate a set of conserved target genes mainly involved in reproductive development and were later independently recruited to modulate stress responses in different Brassicaceae lineages. Analysis of TF BSs in these lineages thus distinguishes widely conserved targets representing the core function of the TF from those that were recruited later in evolution.

vernalization | flowering | *Arabidopsis* | *Arabis alpina* | binding-site evolution

Variation in gene transcription is a major source of phenotypic diversity contributing to adaptation and speciation (1). Understanding how transcriptional patterns arise and evolve is an important question in biology. Transcription is regulated by transcription factors (TFs) that bind, often in combinations, to specific DNA sequences within genes to modulate the activity of RNA polymerase. Variation in transcription of a gene can arise through *cis*-acting differences that alter TF binding or through *trans*-acting differences in TF activity. Recently, these issues have been addressed at the genome-wide level by utilizing ChIP followed by next-generation sequencing (ChIP-seq), allowing determination of binding sites (BSs) for individual TFs across the whole genome and comparison of the repertoire of BSs for orthologous TFs in different species. This approach has been widely applied in yeast and metazoans to study evolution of TF BSs (2–4). However, diversification of TF BSs has seldom been studied at the genome-wide level in plant lineages (5). We determined the BSs of the *Arabidopsis thaliana* TF FLOWERING LOCUS C (FLC) and its ortholog PERPETUAL FLOWERING 1 (PEP1) of *Arabis alpina*, which are critical repressors of flowering (6–8).

Variation in TF binding between species has been extensively studied in yeast and metazoans. Surprisingly, in vertebrates, the BSs of TFs that contribute to developmental processes showed low conservation even among closely related species, and the extent of conservation decreased exponentially with increasing evolutionary distance (9). However, in *Drosophila* species, TF BSs appear to be more conserved (3, 10), and the extent of conservation decreased at only a linear rate with evolutionary distance (9). In plants, only one comparative study of TF binding has been performed in two sister species (5). BSs of the MADS-box TF SEPALLATA3 (SEP3) were compared between *A. thaliana* and *Arabidopsis lyrata*, and approximately 21% of BSs were conserved. This rate of divergence resembles that described in vertebrates rather than in insects. However, the extreme variation in genome size observed in plant lineages (11) and the differences in transposon content found between members of the same family suggested that constraints on genome-wide patterns of TF BSs might vary significantly in different parts of the phylogeny (9). Therefore, we focused on PEP1 and FLC, orthologous MADS-box TFs in *A. alpina* and *A. thaliana*, respectively. In both species, these TFs confer a flowering response to low winter

Significance

Developmental programs of higher plants show plasticity to environmental signals. In the Brassicaceae, the transcription factor (TF) FLOWERING LOCUS C (FLC) represses reproduction until plants are exposed to winter cold. Here we define the target genes of FLC in two species in different lineages of the Brassicaceae and compare the target sequences across the family. Fewer than 20% of target genes were conserved between the species examined, and genes involved in flowering were overrepresented among these. By contrast, many of the nonconserved target genes were involved in stress responses. We propose that, for TFs like FLC, which control environmental responses of plants, core sets of targets are conserved between species, but the majority change rapidly during evolution.

Author contributions: J.L.M., V.T., and G.C. designed research; J.L.M., V.T., R.R., and C.W.M.R. performed research; J.L.M., V.T., P.M., E.S., P.K., and G.C. analyzed data; and J.L.M., V.T., and G.C. wrote the paper.

Reviewers: K.K., Humboldt University Berlin; and H.Y., National University of Singapore.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE89889).

[1]J.L.M. and V.T. contributed equally to this work.

[2]Present address: Fundación Instituto Leloir, Instituto de Investigaciones Bioquímicas de Buenos Aires–National Scientific and Technical Research Council (CONICET), C1405BWE Buenos Aires, Argentina.

[3]Present address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, United Kingdom.

[4]To whom correspondence should be addressed. Email: coupland@mpipz.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618075114/-/DCSupplemental.

temperatures (i.e., vernalization), and their inactivation by mutation causes early flowering (6–8). Transcript levels of *FLC* and *PEP1* decrease progressively in prolonged cold, so that, after several weeks, they reach trough levels. In the annual species *A. thaliana*, *FLC* mRNA level remains low after vernalization, allowing plants to flower continuously until senescence, and *FLC* mRNA level is reset in the progeny. By contrast, in perennial *A. alpina*, *PEP1* transcript level is reset after vernalization, allowing the same individual to respond again to cold the following year (8). Allelic variation at *PEP1* and *FLC* is also important in natural populations contributing extensively to variation in flowering behavior (12, 13). Similar to other MADS-box TFs, FLC binds to CArG-box motifs with the consensus sequence CC[A/T]$_6$GG (14). Hundreds of BSs for FLC have been described in the genome of *A. thaliana* (15, 16), and these suggest that, in addition to vernalization response, FLC is associated with vegetative phase transition, flower development, gibberellin (GA) synthesis and signaling, as well as other environmental responses (15–20).

We compared the BSs and effects on gene expression of FLC and PEP1 in *A. thaliana* and *A. alpina*, respectively, species with markedly different genome sizes and life histories that are in different lineages of the Brassicaceae (21, 22). Our study suggests that determining evolutionarily conserved TF BSs in different Brassicaceae lineages can be a powerful approach to identify target genes contributing to the core function of a TF, as for FLC/PEP1 in the control of flowering. In addition, it allows investigation of how the BSs of orthologous TFs have independently evolved to regulate other processes.

Results

Definition of the Genome-Wide Occupancy and Transcriptional Network Controlled by PEP1 in *A. alpina*. To identify the BSs of PEP1 in *A. alpina* Pajares, ChIP-seq was performed on long-day (LD) grown seedlings of *A. alpina* Pajares and *pep1-1* mutant, which produces no WT PEP1 protein (*SI Appendix*, Fig. S1). The assay was performed in three biological replicates for each genotype by using antiserum raised against PEP1 (12), and high reproducibility was found between biological replicates (*SI Appendix*, Fig. S2). This approach identified 204 high-confidence peaks (false discovery rate < 0.01, present in at least two of three replicates) at unique genomic regions that represent PEP1 BSs (Fig. 1*A* and Dataset S1). Among these BSs, 180 were assigned to 324 annotated neighboring genes, where the center of a peak resided within 5 kb upstream of the transcriptional start site (TSS) and no further than 3 kb downstream of the transcriptional end site (TES) of these genes, which were therefore referred to as PEP1 direct target genes (Fig. 1*A*). In *A. thaliana*, target genes are usually defined as those in which the center of the ChIP-seq peak resides within 3 kb upstream of the TSS and 1 kb downstream of the TES (23, 24), but these distances were extended for *A. alpina* because of the longer intergenic regions in this species (22). Nevertheless, the PEP1 targets obtained by using the parameters employed in *A. thaliana* are also listed (Dataset S1). MADS-box TFs bind mostly in the 5′ promoter region or 5′UTR of target genes (15, 16, 23–27). Consistent with these previous reports, 70% of the detected PEP1 peaks were in promoter regions (Fig. 1*B*). Also, several BSs were validated independently by ChIP-PCR (Fig. 1*C*).

DNA sequences of PEP1 BSs were examined to identify positional weight matrices that can be assigned as consensus sequences recognized by the TF. De novo motif discovery identified the canonical CArG-box [CC(A/T)$_6$GG], the recognition sequence of MADS-box TFs, in 79% of the BSs, and most of these were located close to the center of the peak (Fig. 1*D*). In addition, TCP-binding motifs and G-boxes, which are bound by bHLH and bZIP TFs, were significantly enriched, although the positions of these elements within the peak were more variable (Fig. 1*D*). All three motifs were previously identified in the binding regions of FLC in *A. thaliana* (15, 16).

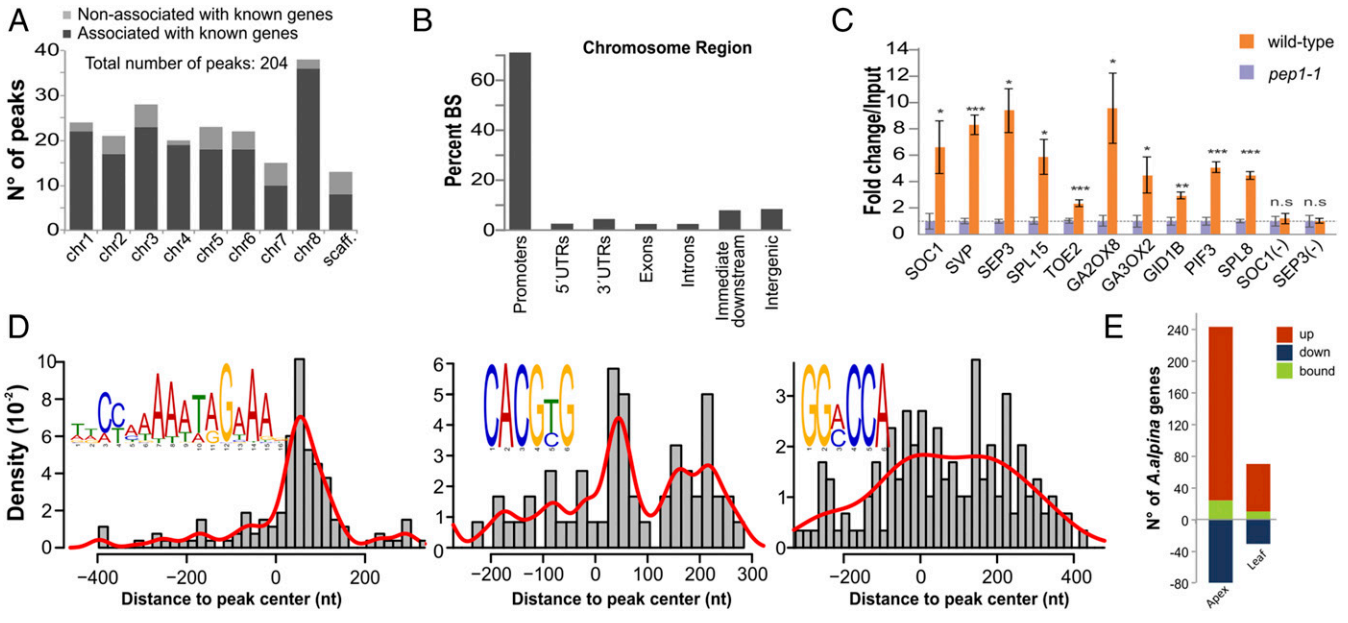


Fig. 1. Characterization of PEP1 targets in *A. alpina*. (*A*) Number of significant ChIP-seq peak calls and their associated genes for PEP1 binding across each chromosome. The proportion of peaks associated with genes is shown in black, and others are shown in gray. (*B*) PEP1 peak distribution over different genomic features in the *A. alpina* genome. (*C*) Validation of ChIP-seq for selected PEP1 BSs using ChIP-PCR. For each target, fold enrichment relative to its input is shown. Minus signs indicate primers not flanking predicted BSs used as negative controls. Data are shown as mean ± SEM ($n = 3$). Asterisks indicate significant enrichment in WT compared with *pep1-1* (n.s., not significant; $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$, Student's *t* test). (*D*) Density plot of distance of CArG-box, G-box, or TCP-binding motif consensus sequences to the center of PEP1 peaks. (*Inset*) Logo of the motifs found by MEME motif analysis. CArG-box, 161 sites (E-value, 7e^{-106}); G-box, 84 sites (E-value, 4e^{-5}); TCP-binding motif, 128 sites (E-value, e^{-7}). (*E*) Proportion of PEP1 direct targets among genes that are up- or down- regulated in *pep1-1* in leaves or apices. For RNA-seq experiment, plants were grown for 2 wk in LD conditions, and leaves and apices were collected at ZT8.

To analyze the influence of PEP1 on gene expression, RNA sequencing (RNA-seq) was performed on *A. alpina* Pajares and the *pep1-1* mutant. In these experiments, leaves and apices of 2-wk-old seedlings were used. In leaves and apices, 96 and 325 transcripts, respectively, were differentially expressed [$\log_2$ (fold change) > 1.5 and adjusted $P < 0.05$] in *pep1-1* compared with Pajares (Dataset S2). The gene set regulated by PEP1 differed between the two tissues (*SI Appendix*, Fig. S3), but approximately 54% of genes differentially expressed in *pep1-1* leaves were also differentially expressed in apices. Overall, more genes were differentially expressed in apices of *pep1-1* mutants than in leaves, which contrasts with what was previously observed in comparisons of *FLC* and *flc-3 A. thaliana* plants (16). PEP1 therefore appears to have a broader role in *A. alpina* apices than FLC does in *A. thaliana* apices, in which the related protein SHORT VEGETATIVE PHASE (SVP) plays a greater role (16).

To assess the effect of PEP1 on transcription of its direct target genes, differentially expressed genes (DEGs) in *pep1-1* mutants were compared with those identified by PEP1 ChIP-seq. A total of 27 genes representing ∼8% of PEP1 direct targets were detected as DEG in either tissue analyzed (Fig. 1*E*). In addition, PEP1 acted almost exclusively as a repressor of transcription (Fig. 1*E*), because nearly all differentially expressed direct target genes were increased in expression in the *pep1-1* mutant compared with Pajares. A similar result was previously obtained for FLC in *A. thaliana* (15, 16). PEP1 direct target genes that are differentially expressed in the *pep1-1* mutant include the orthologs of genes that are associated with different aspects of flowering in *A. thaliana*, including *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*) (28, 29), *CONSTANS LIKE1* (*COL1*) (30), *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 15* (*SPL15*) (31, 32), *SEPALLATA 3* (*SEP3*) (33), and *PHYTOCHROME INTERACTING FACTOR 3* (*PIF3*) (34). PEP1 therefore directly controls genes that are likely to be involved in several flowering-related processes in *A. alpina*.

A Small Proportion of PEP1 BSs in *A. alpina* Are Orthologous to *A. thaliana* FLC BSs. PEP1 BSs in *A. alpina* and FLC BSs in *A. thaliana* were compared to determine their conservation and the extent to which they regulate similar biological processes. To provide an FLC dataset directly comparable to that described here earlier for PEP1, *A. thaliana* plants were grown under similar conditions to *A. alpina* and ChIP-seq was performed. The sequences obtained after ChIP-seq of both TFs generated similar genome coverage (Dataset S3). The ChIP-seq of FLC was performed by using an antiserum that was previously described (15) (*SI Appendix*, Fig. S4, and Dataset S1). Again, high reproducibility was found between biological replicates (*SI Appendix*, Fig. S2). A total of 297 FLC BSs were identified. The number of FLC BSs in *A. thaliana* appears therefore to be higher than the number of PEP1 BSs in *A. alpina*. Genes were defined as direct targets of FLC if the center of a ChIP-seq peak was within 3 kb upstream of the TSS or 1 kb downstream of the TES. By using these criteria, 487 direct target genes were identified under the LD conditions used here, and these showed 50% overlap with those previously found in plants grown under short-day (SD) conditions (16) (*SI Appendix*, Fig. S4*B*). Some of the differences may represent photoperiod-dependent variation in activity of FLC. Furthermore, 61% of the targets detected here overlap with those identified by Deng et al. (15) (*SI Appendix*, Fig. S4*B*). As expected, FLC BSs were mostly located in promoter regions (*SI Appendix*, Fig. S4*C*).

The conservation of FLC BSs and PEP1 BSs was examined by direct sequence alignment comparisons using BLAST (Fig. 2*A*). Only 28 PEP1 BSs (representing 14% of all PEP1 BSs) were found to align to FLC BSs and are referred to as conserved BSs (Fig. 2 *A* and *B*, blue lines). Thus, the binding landscapes of PEP1 and FLC are very different in *A. alpina* and *A. thaliana*, respectively. Target gene conservation was approached independently of BS conservation by comparing directly the target genes identified in each species (Dataset S1). A total of 39 genes were identified as bound by FLC and PEP1, hereafter referred to as common target genes (Fig. 2 *A* and *B* and Dataset S1). In addition, PEP1 targets were compared with FLC targets that were detected by Deng et al. (15) but not found in this study. Twenty genes were identified (Dataset S4), which represented 4% of the FLC targets unique to the list of Deng et al., further supporting the low proportion of target genes conserved in both species. Genes associated with the 28 conserved BSs were then compared with the 39 common target genes detected in our data. Among the common target genes, 28 contained conserved BSs and were referred to as conserved target genes, whereas the remaining 11 contained BSs that differed between the two species (visualized in *SI Appendix*, Fig. S5).

Whether PEP1 and FLC BSs are located in conserved syntenic regions was then examined. Each BS in one species was aligned

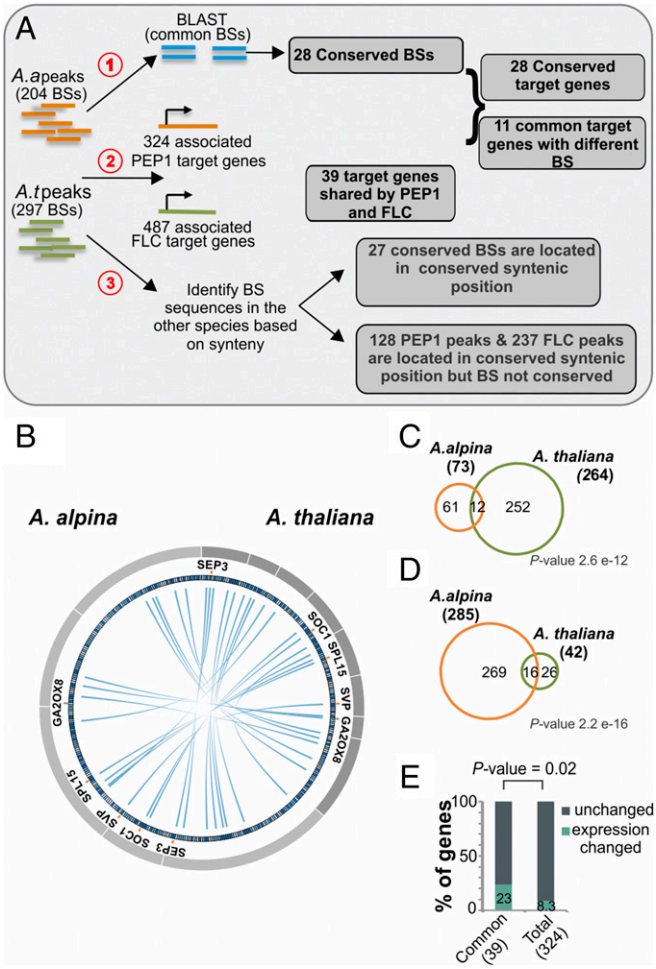


Fig. 2. Comparison of direct targets of FLC and PEP1 as well as of the DEGs in the respective mutants. (*A*) Flow diagram illustrating the rationale of the method to assess BS conservation and target gene conservation for FLC and PEP1 in *A. thaliana* and *A. alpina*, respectively. (*B*) Summary of the analysis described in *A* shown as a Circos plot. Heat map showing the percentage of identity of sequence alignments of BSs to the orthologous locus in the other genome (percentage identity: <50, gray; 50–60%, very light blue; 60–70%, blue; >70%, dark blue). Connecting lines represent alignments between FLC and PEP1 common BSs. Selected flowering related genes are labeled. (*C* and *D*) Venn diagram comparing DEG in leaves and apices of the *A. thaliana flc-3* mutant and the *A. alpina pep1-1* mutant. FLC transcriptome data were obtained from ref. 16. (*E*) Proportion of DEG in *pep1-1* among common target genes and all *A. alpina* PEP1 target genes. *P* value indicates a statistically significant difference (hypergeometric test).

to the orthologous locus in the other genome. This approach demonstrated that all conserved BSs that are associated with genes (24 BSs in *A. alpina* and 27 BSs in *A. thaliana*) are located in orthologous genomic positions. Furthermore, the majority of species-specific BSs (72% of PEP1 and 88% of FLC BSs; Fig. 2 *A* and *B*) are present in the other genome, with a sequence identity of 50–80% (*SI Appendix*, Fig. S6, and Dataset S5). Therefore, the sequences containing species-specific BSs are mostly present in the genome of the other species but are not recognized by the orthologous TF.

The effects of PEP1 and FLC on the transcriptome of *A. alpina* and *A. thaliana*, respectively, also differed greatly (Fig. 2 *C* and *D*), consistent with the low conservation of BSs. Among PEP1 direct targets, a higher proportion of conserved target genes showed altered expression in the mutants compared with all target genes (Fig. 2*E*; validation of differences for a subset of genes is shown in *SI Appendix*, Fig. S7). Thus, under standard growth conditions, conserved targets are more likely to show a change in expression than species-specific targets.

In summary, these results indicate that PEP1 and FLC targets include a small core set of conserved genes, but that, overall, their genome-wide binding landscapes are very different.

Sequence Variation at CArG-Box Motifs Contributes to Divergence in PEP1 and FLC Binding. FLC and PEP1 proteins show high sequence identity, which reaches 100% in the DNA-binding MADS-domain (*SI Appendix*, Fig. S8), yet their genome-wide BSs are highly divergent (Fig. 2). Comparison of the overall sequence conservation of *A. alpina* BSs in *A. thaliana* demonstrated that BSs that are conserved across species showed similar levels of sequence identity to those that are specific to *A. alpina* (*SI Appendix*, Fig. S6). This result differs from that described for SEP3 in *A. thaliana* and *A. lyrata*, in which genomic regions that were bound by SEP3 in both species were significantly more conserved than regions that were bound specifically in either species (5). As overall sequence across the FLC and PEP1 BSs is similarly conserved at specific and conserved BSs, the configuration of *cis*-elements, particularly of the CArG-boxes recognized by the TFs, was compared in both sets of BSs. This motif was significantly enriched in species-specific BSs as well as conserved BSs, and the motifs showed high similarity between species (Fig. 3*A*), suggesting that both TFs recognize the same *cis*-elements. However, the consensus sequences of CArG motifs in conserved and species-specific BSs differed slightly (Fig. 3*A*): only the consensus sequences of conserved BSs showed significant (Z-score >3) enrichment of a TTT extension at the 5′ end of the core motif (Fig. 3*A*). This suggests that the 5′ TTT is a functional part of the CArG-box at the conserved BSs, similar to the 3′ AAA extension that was found in all BSs and that was previously found in FLC BSs (15). Interestingly, ChIP-seq peaks present at conserved BSs are also generally more significant than those present at specific BSs (Mann–Whitney U test for PEP1, $P = 0.00611$; Mann–Whitney U test for FLC, $P = 0.0329$; Fig. 3*A*), suggesting that in vivo FLC/PEP1 might bind more strongly to those sites containing both the 5′ TTT and 3′ AAA sequences.

These differences between conserved and species-specific CArG-box motif sequences are found in both species, and therefore do not explain the divergence in the BSs of PEP1 and FLC between *A. alpina* and *A. thaliana*, respectively. To analyze these sequences in more detail, position-specific PhastCons scores, which represent the probability that a given nucleotide is part of a conserved region, were calculated for the CArG-box motifs and flanking nucleotide positions. Analysis of position-specific PhastCons scores within CArG-box motifs and 10 flanking nucleotides on each side in the conserved and specific BSs revealed that the CArG motifs within conserved BSs tend to have higher PhastCons scores (Fig. 3*B*). These results indicate that CArG-boxes in the conserved BSs are located within more conserved regions than CArG-boxes of *A. alpina*-specific BSs. Furthermore, a strong association was found between BS conservation across species and the presence of a CArG-box in those sites in both species (Fig. 3*B*, *Inset*). Approximately 80% of conserved PEP1 BSs contain a CArG-box in the *A. thaliana* orthologous sequence that is bound by FLC. In contrast, only 45% of the *A. alpina* specific PEP1 BSs contain this motif in their *A. thaliana* counterpart (Fig. 3*B*, *Inset*). This dramatic reduction in CArG-box conservation at the *A. thaliana* orthologs of specific PEP1 BSs probably largely explains the lack of FLC binding at these sites. Absence of CArG-boxes in these regions of *A. thaliana* was generally not caused by large-scale rearrangements, because overall homology of the BSs sequence between species is mostly maintained (*SI Appendix*, Fig. S6). Examples of modifications that change the CArG-box motif are shown in Fig. 3*C*. In each of these cases, the CArG-box is absent in *Aethionema arabicum* and in the *Arabidopsis* species examined, suggesting that they were gained in a lineage leading to the *Arabis* genus. Similarly, for *A. thaliana*-specific BSs, CArG-boxes were not present in most sequence counterparts in the *A. alpina* genome (Fig. 3*B*, *Inset*) and regions surrounding the *cis*-elements are more conserved in common BSs than in specific ones (Fig. 3*B*). In contrast to CArG-boxes, the conservation of G-box elements and TCP-BSs at PEP1 BSs did not strongly differ between conserved and specific BSs and thus did not correlate with binding conservation (*SI Appendix*, Fig. S9). In summary, conservation of binding at orthologous sequences across species is strongly correlated with presence of the CArG-box motif in both species (Fig. 3 *A* and *B*), whereas species-specific PEP1 binding correlated with absence of the CArG-box in *A. thaliana* but not with large sequence rearrangements.

Although the absence of a CArG-box motif is strongly correlated with loss of binding, ∼40% of specific BSs retained a CArG-box in the species in which no binding was detected (Fig. 3*B*, *Inset*). Lack of binding at these sites was also not correlated with absence of other motifs, such as the G-box or the TCP-binding motif (*SI Appendix*, Fig. S9). CArG-box motifs are recognized by many MADS-box TFs in addition to FLC, some of which were shown to form a protein complex with FLC (16, 35), and the genome-wide BSs of several of these have been identified in *A. thaliana* (15, 16, 23–26, 36–38). Conservation of CArG-boxes at species-specific BSs might be necessary to allow binding of other MADS-box TFs to these sites in the species in which FLC/PEP1 does not bind. Binding of different MADS-box TFs could allow repurposing of a *cis*-element, as was described in human and mouse (39). Comparison of the BSs detected for FLC/PEP1 with those previously described for other MADS-box TFs in *A. thaliana* showed that ∼80% of FLC-specific BSs are bound by at least one other MADS-box TF, whereas, for 40% of PEP1-specific BSs, the orthologous *A. thaliana* sequence has been found to bind another MADS-box TF (Fig. 3*D*). *A. thaliana*-specific BSs containing a conserved CArG-box are slightly more frequently bound by other MADS-box TFs compared with all species-specific BSs [Fig. 3*D*, row 1 (red) vs. row 2 (black)], suggesting that these CArG-boxes might be selected for because other TFs recognize these regions. Interestingly, BSs that are conserved between *A. thaliana* and *A. alpina* show a higher frequency of overlap with BSs of other MADS-box TFs than the species-specific BSs (Fig. 3*D*, row 3). The conserved function of PEP1 and FLC might involve interaction with some of these other MADS-box TFs, or these regions might be recognized by multiple MADS-box complexes.

In summary, sequence analysis of BSs showed that PEP1/FLC binding conservation is associated with the presence of a CArG-box at the BSs in both species, but CArG-box conservation is not sufficient for binding to occur. The binding of other MADS-box TFs to the same BS as PEP1/FLC might explain why, in some cases, the CArG-box is maintained even though conserved binding is not observed.

Target Genes of PEP1 and FLC Show Related Biological Functions. Gene Ontology (GO) enrichment analysis of direct target genes was performed to address whether differences in the binding landscapes of PEP1 and FLC cause divergence in the

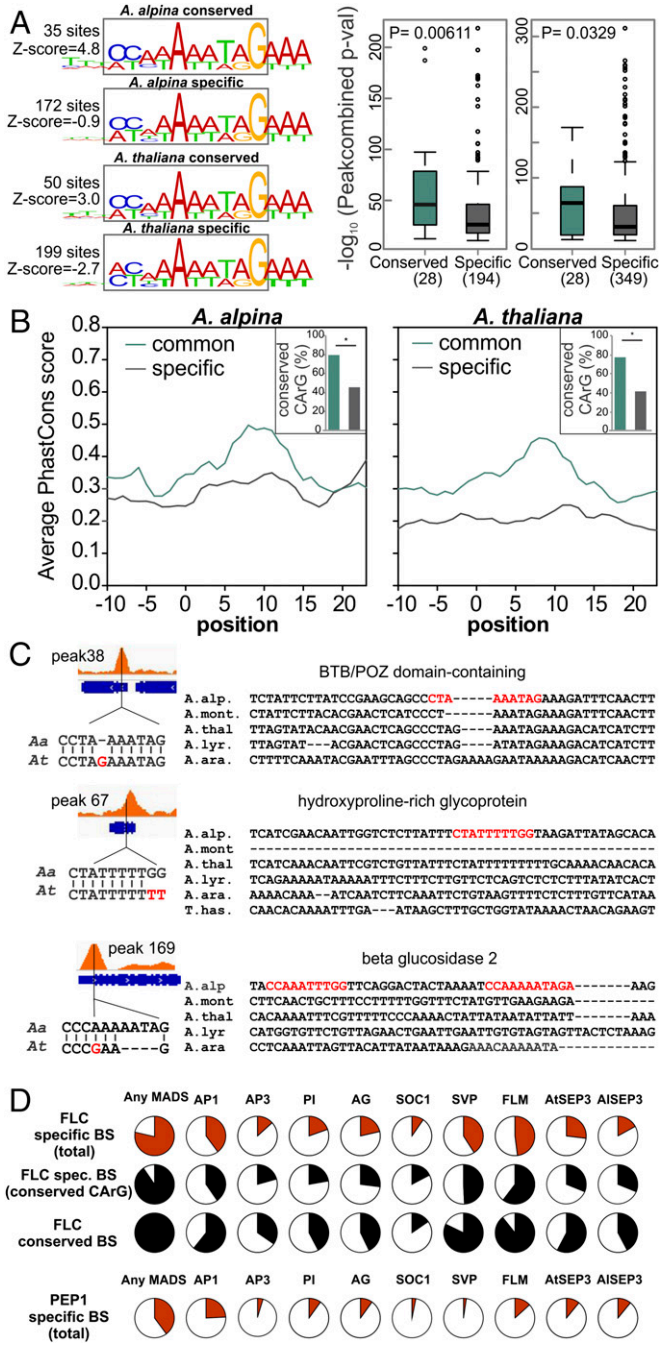


Fig. 3. PEP1 binding conservation depends on *cis*-element conservation. (*A*, *Left*) CArG-boxes in conserved and species-specific BSs in *A. alpina* and *A. thaliana* (consensus sequences are boxed). The number of CArG-boxes identified in each subset is indicated. *Z*-scores indicate significance of TTT enrichment at positions 1–3. (*A, Right*) Binding significance for conserved (green) and species-specific (gray) BSs in either species (Mann–Whitney *U* test *P* values are shown). (*B*) Average PhastCons scores of CArG-boxes in conserved and species-specific BSs for *A. alpina* (*Left*) or *A. thaliana* (*Right*). The CArG-box itself is located from positions 0 to 13. (*Insets*) Percentages of orthologous regions in the other species that contain a CArG-box. Asterisks indicate significant enrichment (*Z*-score > 3) and significant difference between groups (test based on hypergeometric distribution). Binding strength is higher among conserved BSs compared with specific BSs. (*C*) Three examples of *A. alpina*-specific BSs that contain a CArG-box in *A. alpina* that is lost in other Brassicaceae. (*Left*) G-browse capture of PEP1 binding and CArG-box sequence. Changes relative to the consensus motif in *A. thaliana* are highlighted in red. (*Right*) Alignment of *A. alpina* sequence around CArG-box with orthologous sequences in several Brassicaceae species. CArG-box from *A. alpina* is highlighted in red. (*D*) Percentages of different subsets of FLC or PEP1 BSs that overlap with BSs described for other MADS-box TFs in *A. thaliana*. For comparison of *A. alpina* PEP1 BSs, the orthologous sequence of BSs in *A. thaliana* was used. Subsets of BSs for each species are specific (red), specific with conserved CArG-box (black), and conserved BSs (black).

biological processes they regulate. In contrast to the high divergence in the identity of FLC and PEP1 direct target genes (Fig. 2), the biological processes associated with the targets of each TF were highly similar (Fig. 4*A*, *Left*). Consistent with the early-flowering phenotypes of *flc* and *pep1* mutants, the most enriched gene set regulated directly by FLC and PEP1 contained genes involved in flowering-time control with representation factors of 8.6 and 5.0 for PEP1 and FLC, respectively (Fig. 4*A*, *Left*, and Dataset S1). In both species, direct targets include genes that control flowering in distinct pathways and tissues of the plant, particularly in the photoperiod/light sensing/circadian clock pathway, in GA metabolism or signaling, and in meristem response and development (Dataset S1). Among genes bound by PEP1 whose orthologs are associated with the photoperiodic pathway were *CONSTANS LIKE 5* (COL5), *FT*, *NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT B-2* (*NFY B-2*), *PIF3*, *TARGET OF EAT 1* (TOE1), and *TOE2*. *COL5* is under circadian clock and diurnal regulation, and reduced or increased activity of this gene alter flowering time (40, 41). Furthermore, FLC binds to eight genes associated with photoperiodic flowering response, but only *FT* and *PIF3* are conserved targets with PEP1. Specific targets of FLC in *A. thaliana* are *TEMPRANILLO 1* (*TEM1*), *NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT C-9* (*NFY C-9*), *SPA1-RELATED 2* (*SPA2*), *CIRCADIAN 1* (*CIR1*), *AGAMOUS LIKE-15* (*AGL15*), and the photoreceptor-encoding gene *PHYTOCHROME B* (*PHYB*).

Genes encoding GA metabolism enzymes were also enriched among targets of both TFs (Fig. 4*A*), which might be related to the role of GA in promoting flowering. FLC was previously shown to control expression of genes involved in GA regulation, and the FLC-interacting MADS-box protein SVP has a role in reducing GA biosynthesis (42–46). Here, PEP1 was found to directly bind genes encoding enzymes involved in GA catabolism or biosynthesis such as GIBBERELLIN 2-OXIDASE 8 (GA2ox8) and GIBBERELLIN 3-OXIDASE 2 (GA3ox2), as well as genes in the signaling pathway including GA INSENSITIVE DWARF1B (GID1B). By contrast, FLC bound directly to genes involved in GA signaling, such as *GA INSENSITIVE DWARF1C* (*GID1C*) and *RGA-like 2* (*RGL2*). FLC also bound to *GA2ox8*, but at a different BS than PEP1 (*SI Appendix*, Fig. S5), and expression of *GA2ox8* in *A. thaliana* was up-regulated in *flc-3* (16) rather than down-regulated as in *pep1-1* (*SI Appendix*, Fig. S7). These data indicate that FLC and PEP1 both regulate GA-related processes but that they do so mostly by binding to different genes. Other GO categories enriched for both TFs included flower development, abiotic stimuli, response to cold, and response to temperature stimulus.

The 39 common target genes (Fig. 2*A* and Dataset S1) were then analyzed separately. The GO terms for this group showed overrepresentation of flowering-time genes and of genes involved in reproductive development (Fig. 4*A*, *Right*). Seven common target genes were involved in flowering-time control, and represent almost 20% of the common target genes. Four of these correspond to meristem response and development, suggesting that control of these processes might be more strongly conserved during evolution (Dataset S1). Shared flowering target genes included *SOC1*, *SVP*, *SPL15*, *SEP3, GA2ox8, FT*, and *PIF3*, some of which have genetically confirmed functions downstream of FLC (28, 35). Except for *GA2ox8* and *FT*, all of these common target genes encode TFs with major regulatory functions, which indicates that at least some of the mechanisms by which FLC and PEP1 repress flowering are conserved.

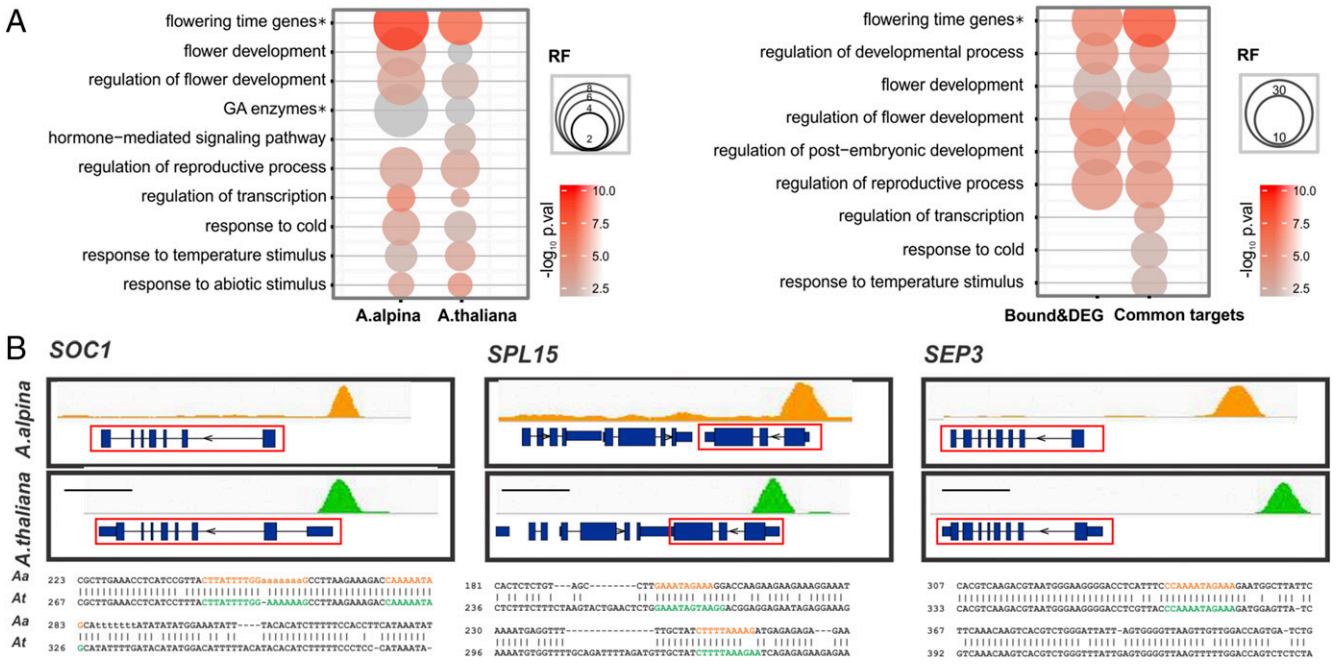


Fig. 4. Conservation of PEP1 targets and their biological functions between *A. alpina* and *A. thaliana*. (*A*) Bubble plot showing GO terms enriched among high-confidence targets of PEP1 and FLC (*Left*) and of a subset of PEP1 targets, namely DEG (27 genes) and shared with FLC (39 genes) (*Right*). Genes involved in flowering-time control or synthesis and degradation of GA reported in the literature (i.e., not standard GO categories) are marked with asterisks. Color represents *P* value, and size of the bubble represents the representation factor (*SI Appendix*). (*B*) GBrowse captures showing local enrichment of PEP1 (orange) and FLC (green) binding in three different genes (red box) associated with flowering or flower development to illustrate conserved binding. The sequence alignments of BSs around CArG-boxes are shown. CArG-box motifs in orange for *A. alpina* or green for *A. thaliana*. Numbers correspond to position within the BS. Bar represents a 1-kb window.

Analysis of the genes that were directly bound by PEP1 and differentially expressed in *pep1-1* mutants again identified categories related to flowering (Fig. 4*A*, *Right*). Thus, conserved binding of PEP1 to flowering-time genes is often associated with repression of their transcription. On the contrary, GO categories related to temperature responses were not enriched among those differentially expressed in *pep1-1* (Fig. 4*A*), indicating that direct binding to temperature related genes is not related with controlling transcription under normal growth temperatures. However, as is shown later, some of these genes were differentially expressed in *pep1-1* compared with WT at lower temperatures.

The sequences of three conserved targets with key regulatory roles in flowering or flower development were analyzed (Fig. 4*B*). In these cases, the predicted CArG-box BSs and surrounding sequences were highly conserved between *A. alpina* and *A. thaliana* (Fig. 4*B*). Moreover, the predicted BSs were highly conserved in other Brassicaceae species, and two were conserved in *A. arabicum*, the most basal lineage within the Brassicaceae family, and *Tarenaya hassleriana*, a member of the sister family (Fig. 4*C* and *SI Appendix*, Fig. S10). Among the whole set of conserved target genes, the CArG-boxes are widely conserved in the Brassicaceae, but only a small subset is conserved in all species (*SI Appendix*, Fig. S11). These findings suggest that the regulation of these genes by PEP1/FLC is broadly conserved in the Brassicaceae, although this remains to be directly tested.

In summary, to control flowering, FLC and PEP1 regulate transcription of several orthologous genes through binding to highly conserved CArG motifs, whereas other biological processes regulated by both TFs mainly involve their binding to different target genes, suggesting that they arose by convergent evolution.

PEP1 and FLC Modulate Responses to Short-Term Cold Exposure by Regulating Different Sets of Genes. Several direct target genes of FLC and PEP1 were found to be associated with the functional categories "response to cold" or "response to temperature" (Figs. 1*C*, 4*A*, and 5*A*). Moreover, PEP1 bound to orthologs of 33 of the robust list of 1,279 *A. thaliana* cold-regulated (COR) genes (47), and orthologs of 62 COR genes were differentially regulated in *pep1* (17% of *pep1-1* differentially regulated genes; Fig. 5*B* and Dataset S6). Comparison of the expression levels of these 62 genes in the *pep1-1* mutant with published expression levels of their *A. thaliana* orthologs after 1 h of cold treatment (47) revealed highly correlated patterns (*SI Appendix*, Fig. S12). These results suggest that mutation of *PEP1* in *A. alpina* has a similar effect on the transcription of these genes as cold treatment in *A. thaliana*, and therefore PEP1 might negatively regulate the cold response. Similarly, FLC target genes include those belonging to the response to cold or response to temperature GO categories and 44 COR genes (Fig. 5*B* and Dataset S6). Furthermore, 70 COR genes are differentially regulated in *flc-3* (26% of *flc-3* differentially regulated genes) compared with *FLC* (Dataset S6), and expression patterns of these genes in WT background after 1 h of cold treatment resemble those in the *flc-3* mutant (*SI Appendix*, Fig. S12). Together, these findings suggest that FLC and PEP1 are negative regulators of the cold response. However, comparison of COR genes regulated by FLC and those regulated by PEP1 revealed only a small overlap (Fig. 5*B*), indicating that most of the cold-network genes controlled by PEP1 and FLC are different.

PEP1 was found to bind to many COR genes (Figs. 4*A* and 5 *A* and *B*), but analysis of direct targets that are also differentially regulated in *pep1-1* under ambient temperature did not identify cold-related terms (Fig. 4*A*). Therefore, to test whether PEP1 directly regulates the cold response, the expression of these target genes was analyzed at low temperature in WT and *pep1-1* plants. Even though *PEP1* expression was slightly increased in response to short-term cold treatment (*SI Appendix*, Fig. S13), similar to what was described for *FLC* in *A. thaliana* (48), many PEP1 direct target genes were expressed at higher levels in the *pep1-1* mutant compared with WT (Fig. 5*C*), suggesting that PEP1 functions in cold temperature to modulate the cold response. Expression of COR genes was also analyzed

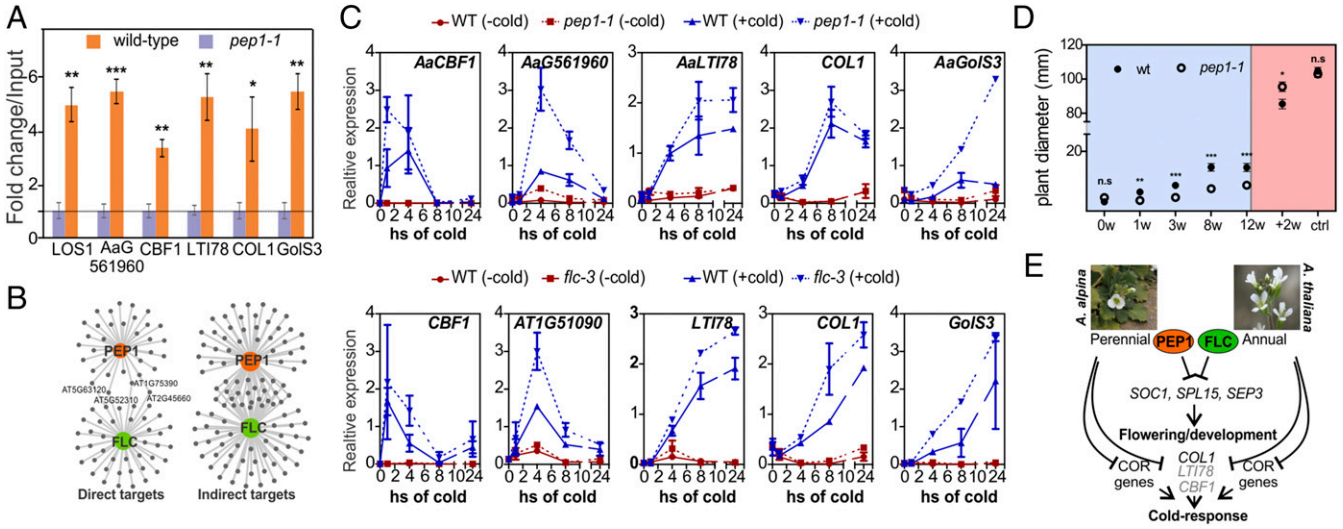


Fig. 5. Analysis of the roles of PEP1 and FLC in the regulation of cold response. (*A*) ChIP-PCR of PEP1 to sites associated with selected cold-related target genes. Data are shown as mean ± SEM ($n = 3$). Student's *t* test was performed to compare WT and *pep1-1* (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$). (*B*) Network showing direct targets and indirectly regulated genes (DEG in the respective mutant) of PEP1 and FLC among the COR genes defined in ref. 47. Genes are listed in Dataset S6. (*C*) Effect of mutations in *PEP1* and *FLC* on expression of COR genes after transferring plants to 4 °C for 24 h compared with control conditions of maintaining plants at 22 °C. Data are shown as mean ± SEM ($n = 2$). Each experiment was normalized to the average expression of the mutant in cold. (*D*) Growth phenotype of *pep1-1* mutant in cold. Plant diameter during cold treatment is shown as mean ± SEM ($n = 2$). Student's *t* test between WT and *pep1-1* was performed (n.s., not significant; *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$). Controls represent plants that were grown at 22 °C for 3 wk. (*E*) Model of PEP1 and FLC regulation of cold responses and flowering. Direct targets with nonconserved BS are shown in gray.

during vernalization, when *PEP1* is eventually silenced, and the repressive effect of PEP1 on cold-induced gene induction of some of these genes, such as *AaCOL1* and *AaGolS3*,was still detectable in the first weeks of vernalization (*SI Appendix*, Fig. S14), correlating with *PEP1* expression (8). The expression patterns of orthologs of the COR genes that were direct targets of PEP1 were examined in response to short-term cold treatment in *A. thaliana* (Fig. 5*C*). The *flc-3* mutation had a similar effect on the amplitude of induction of these cold-responsive genes as *pep1-1* did in *A. alpina* (Fig. 5*C*). However, only three of the five tested are direct targets of FLC, and, of those, only *COL1* is bound at a conserved BS, but this BS does not contain a conserved CArG-box identified in the analysis of conserved target genes (*SI Appendix*, Figs. S5 and S11, and Dataset S1). This further supports our suggestion that PEP1 and FLC negatively regulate the cold response but do so by directly binding to different genes or BSs.

During cold exposure, plant growth is reduced (49), and, in *A. thaliana*, *CBF1* overexpression caused severe growth reduction (50), suggesting that active repression of growth is part of the response to cold. Stronger expression of cold-responsive genes, including *CBF1* (Fig. 5*C*), in *pep1-1* prompted us to test whether the *pep1-1* mutation also affects plant growth in response to cold. Compared with WT plants, growth of the *pep1-1* mutant was significantly reduced during cold treatment (Fig. 5*D*). By contrast, before or after cold exposure or under control conditions in which plants were not exposed to cold, this effect was not detected or was even reversed. A stronger growth retardation in response to cold in *pep1-1* suggests that, by modulating expression of cold-responsive genes, PEP1 might promote growth in cold, a function not previously assigned to this gene to our awareness.

Taken together, PEP1 and FLC negatively regulate cold induction of cold-responsive genes, but, in contrast to the regulation of reproductive development (Fig. 5*E*), the COR genes controlled by FLC and PEP1 are highly divergent.

Discussion

Phenotypic differences between closely related species are often caused by changes in gene transcription (51, 52). Variation in the activity of regulatory TFs or in the sequence of their BSs can cause alterations in the transcription of target genes (53). The rate with which genome-wide patterns of TF binding change during evolution seems to vary in different lineages (9), but has not been extensively studied in plants. Here we show that the BSs of the orthologous MADS-box TFs FLC and PEP1 vary greatly between *A. thaliana* and *A. alpina*, two species in the Brassicaceae family, but a core set of targets with roles in reproductive development is highly conserved. We argue that the functions of PEP1 and FLC in reproductive development are highly conserved but that they have independently acquired functions in abiotic stress responses.

Conservation and Diversification of TF BSs. The genome-wide landscapes of BSs for orthologous TFs differ markedly among related species of yeast and mammals. For example, analysis of BSs of the liver-specific TF CEBPα in five different vertebrates and six rodent species showed an exponential decrease in the proportion of BSs conserved with evolutionary distance (4, 54, 55). Analysis of genome-wide BSs of several TFs in different tissues of human and mouse also showed low conservation (39). However, genome-wide BSs for TFs among *Drosophila* species seem to diverge more slowly than in mammals. For instance, more than 60% of the BSs of the TF Twist, which is involved in embryo development, were conserved among *Drosophila* species that diverged more than 30 Mya (3). Similar results were obtained for six TFs that regulate segmentation (10). This apparent lower rate of TF BS divergence among *Drosophila* species might result from their more compact genome size and higher effective population size than those of mammals (9, 56). In plants, in which genome size and structure vary greatly even among closely related species (57, 58), little is known of the rate of divergence of genome-wide patterns of TF binding. The one example available so far showed that, between the sister plant species *A. lyrata* and *A. thaliana*, only ∼20% of BSs of the SEP3 TF were conserved, suggesting a fast rate of divergence more similar to that observed in mammals. However, analysis of more TFs in a wider range of species is required to provide a broader picture of the divergence of TF binding in plants. We analyzed the genome-wide binding patterns of the orthologous MADS-box TFs FLC and PEP1 of *A. thaliana* and *A. alpina*,

respectively, two species in different lineages of the Brassicaceae, and found that fewer than 20% were shared between species. Although the CArG-box motifs recognized by different MADS-box TFs can differ subtly in sequence (14, 59–61), this is unlikely to be the cause of the divergent binding patterns of FLC and PEP1 because de novo motif discovery of the ChIP-seq peaks indicated that these TFs recognize the same consensus sequences. Similar observations were made for orthologous metazoan TFs with divergent genome-wide binding patterns (3, 4, 10). Nevertheless, for FLC and PEP1, the conserved BSs appear to differ qualitatively from the species-specific BSs. De novo motif discovery performed specifically on their common BSs identified an extended palindromic CArG-box consensus. MADS-box TFs bind DNA as tetramers, and each dimer binds one CArG-box motif (62). Variability in the CArG-box sequence recognized in vivo by SEP3 was previously proposed to result from it acting in different homo- and heterodimers (23). Flowering-time genes were enriched among those targets containing common BSs for FLC and PEP1 and were overrepresented among genes bound by FLC and its partner MADS-box TF SVP (16, 35). Therefore, common targets might be predominantly recognized by the SVP/FLC complex, whereas species-specific BSs might be recognized more frequently by FLC acting together with other MADS-box partners. Defining such combinatorial activities of TFs can allow their BSs to be predicted more accurately by using computational approaches (63).

Much of the variation in TF binding patterns among species has been explained by species-specific sequence variation at DNA motifs recognized by the TFs (3, 4, 10) or by insertion of transposable elements (4, 5, 64). The divergent landscapes of PEP1 and FLC binding could partially be explained by different distributions of CArG-box motifs. The majority of *A. thaliana* genomic regions orthologous to *A. alpina*-specific BSs did not contain a consensus CArG-box. Similarly, in comparisons of SEP3 binding between *A. thaliana* and *A. lyrata*, species-specific binding was often associated with modification or absence of the CArG-box in the species in which no binding was detected (5). Also, in vertebrates, the liver TFs HNF1α, HNF4 α, and HNF6 of mouse recognized human-specific BSs in mouse cells containing a human chromosome, emphasizing the importance of *cis*-acting variation in generating species-specific binding patterns (65). No evidence was obtained for the direct involvement of transposons in the generation of the specific BSs for PEP1 and FLC between *A. alpina* and *A. thaliana*, in contrast to what was observed for SEP3 in *A. lyrata* (5) and despite the high transposon content of the *A. alpina* genome (22).

Although lack of binding often correlated with absence of a CArG-box, 45% of PEP1-specific BSs retained the CArG-box at the orthologous site in *A. thaliana*. Thus, at these sites, the presence of a CArG-box is not sufficient for FLC binding. In other systems, loss of TF binding despite the conservation of the binding motif has been explained in different ways. Variation in regions flanking the core motif can impair binding of a TF because it prevents combinatorial interactions with a second TF that binds to an adjacent site (16, 54, 55) or it affects the DNA structure (66) or chromatin accessibility (67). Furthermore, changes in methylation patterns could also account for differences in TF binding (68). These observations emphasize the importance of testing binding directly and not relying on predictions based on conservation of the binding motif. Other features of these sites might preclude FLC binding, such as the absence of binding of another TF complex at a neighboring site that acts in a combinatorial fashion with FLC or a more general feature of the chromatin structure. However, de novo motif discovery performed on the species-specific sites of FLC/PEP1 binding in the species in which binding occurred did not identify any sequence motif apart from the CArG-box, suggesting that the additional features present at these BSs are not sufficiently conserved in different genes to be identified by this approach or that they are not recognizable in the primary DNA sequence.

The genome of *A. alpina* is threefold larger than that of *A. thaliana*, and larger genome sizes have been proposed to correlate with rapid evolutionary turnover of TF BSs, as described here earlier in the comparison of *Drosophila* species and mammals. Variation in intergenic distances in the genomes of different plant species might also contribute to divergence in the rate at which TF BSs exhibit turnover, although the difference in genome size between *A. alpina* and *A. thaliana* is much smaller than that between flies and vertebrates. Despite the larger genome of *A. alpina*, we found fewer PEP1 BSs than FLC BSs in *A. thaliana*. However, to determine if this correlates with a lower BS turnover in *A. alpina*, it will be necessary to consider rates at which BSs were gained and lost in both species. Nevertheless, the rapid rate of turnover of FLC/PEP1 BSs described here, and that of SEP3 BSs in the only other study on genome-wide turnover of plant TF BSs of which we are aware (5), suggest that, in plants, as in vertebrates, BSs of TFs involved in developmental processes evolve rapidly.

Conservation of the Core Developmental Function of TFs. In metazoans, despite rather low conservation of BSs of TFs involved in development, their regulatory functions are often maintained. This apparent paradox seems to be explained by high conservation of binding to key genes involved in the core developmental function of the TF (3, 4, 54, 69). Sometimes, even though binding to the same gene occurs in different species, the BSs in the orthologs are at different positions (4, 39, 64, 69). Similarly, FLC and PEP1 each bound to hundreds of genes, but only 39 (Dataset S1) were identified as common target genes. Thus, despite the similar early-flowering phenotypes of *flc* and *pep1* mutants, the *cis*-regulatory networks regulated by these orthologous TFs are strongly divergent. Nevertheless, as described here earlier for metazoan TFs involved in development, genes involved in their common developmental function of flowering-time control were highly overrepresented among the common targets of FLC and PEP1. Early flowering is the most evident phenotype of the *flc* and *pep1* mutants, suggesting that the major evolutionarily conserved function of these TFs is to repress flowering. Similarly, many SEP3 target genes involved in flower development were conserved in two sister species (5). Thus, the conservation of core functions accompanied by rapid turnover of other BSs appears to be found widely in diverse organisms from vertebrates to plants.

Common targets in flowering-time control indicate that FLC and PEP1 repress the earliest stages of floral induction in the shoot meristem. *SOC1* and *SPL15* are both direct targets and were recently shown to cooperate to activate target genes in the meristem under noninductive conditions (31). *SOC1* is also an early-acting gene in photoperiodic response (29). Therefore, by repressing both of these genes, activity of the SPL15-SOC1 complex would be strongly reduced, and floral induction effectively repressed under different environmental conditions. Similarly, SEP3 is a member of many MADS-box complexes involved in reproductive development, and its promoter was bound by both PEP1 and FLC (70). The CArG-boxes recognized by PEP1 and FLC in these three genes are conserved in all Brassicaceae genomes tested except *A. arabicum*, in which only the CArG-box present in *SOC1* was conserved. Thus, these genes were probably regulated by FLC orthologs early during the divergence of the Brassicaceae, before separation of the *Arabis* and *Arabidopsis* lineages, and these BSs must be under strong selective pressure; however, binding of FLC orthologs to these sites remains to be tested experimentally in a broader range of species. Previously, *FT* and *FD*, which encode photoperiodic flowering pathway components, were identified as targets of FLC by ChIP-PCR (71). Neither here nor in other genome-wide analyses of FLC targets was *FD* identified as a target of FLC/PEP1 (15, 16). Therefore, despite the early ChIP-PCR result, *FD* is probably not directly regulated by FLC. By contrast *FT* is bound by FLC and PEP1 in *A. thaliana* and *A. alpina*, respectively (15, 16). Finally, the list of highly conserved targets includes genes that have not previously been shown to have roles in flowering, but are now candidates for testing, including several encoding TFs, a transporter, and a helicase (*SI Appendix*, Fig. S11).

Nonconserved Targets of FLC and PEP1 Support Convergent Evolution on Stress Responses. Analysis of all BSs of FLC or PEP1 identified many of the same enriched GO categories, particularly hormone stimuli or the response to cold, suggesting that these TFs regulate abiotic responses as well as reproductive development. Similarly, the LEAFY TF, which has developmental roles in conferring floral meristem identity, was shown to contribute to biotic stress responses in *A. thaliana* (72). Among yeast species, divergence in TF-binding patterns was proposed to facilitate adaptation to different environments (2). FLC and PEP1 both bind many COR genes, which are regulated by CBF TFs and are implicated in adaptation to cold. Also, overexpression of CBFs was previously shown to increase FLC transcription (48), suggesting a complex interaction between abiotic stress responses to cold and FLC activity. We detected increased transcription of COR genes in *pep1-1* and *flc-3* compared with WT when plants were grown in cold, suggesting that PEP1 and FLC function specifically in cold to repress COR gene induction. As cold stress tolerance comes along with a retardation of growth (49), PEP1 and FLC might modify this trade-off to maintain growth under cold but nonfreezing temperatures. Interestingly, most of the COR genes bound by FLC or PEP1 are specific for each TF or the COR target genes are orthologous but the BSs are different, as shown for *CBF1*. This dramatic difference in the set of COR genes regulated by FLC and PEP1 suggests that these TFs were independently recruited after divergence of the lineages containing *A. alpina* and *A. thaliana* to modulate responses to cold, and this might represent convergent evolution to adapt to colder climates after the Brassicaceae split from the Cleomaceae (21).

Perspectives. As observed in metazoans, analyzing TF BSs in plant species from different lineages of the Brassicaceae is informative in understanding TF function and evolution. Performing similar studies with more TFs will demonstrate how generally applicable these observations are, and analyzing species in other families will indicate whether conservation of BSs extends beyond the Brassicaceae. FLC orthologs and their regulation by vernalization have been described even in monocotyledonous species (73). Consistent with this, the CArG-boxes to which FLC and PEP1 bind in *SOC1* and several other flowering-time genes were present in the orthologs of *T. hassleriana*, a member of the Cleomaceae, a sister family to the Brassicaceae. However, the Cleomaceae are found in semitropical regions and do not show a vernalization response, suggesting that this CArG-box is bound by other MADS-box TFs in *T. hassleriana* and has been repurposed in the Brassicaceae to bind FLC orthologs that contribute to vernalization response or that the FLC orthologs present in *T. hassleriana* (*SI Appendix*, Fig. S15) bind to this site but do not confer vernalization response. Deepening our knowledge of the function and topology of these networks in different species will help in the understanding of the evolution of reproductive development and of plant transcriptional networks more generally.

Methods

Growth Conditions, Plant Materials, and Phenotypic Analysis. *A. alpina* plants from Pajares accession and *pep1-1* mutant (8) were grown under LD conditions (16 h light/8 h dark at 20 °C). The *A. thaliana FRI* introgression line (Col FRI) (74) was used as WT. The *flc* allele used was *flc-3* (6). Plants were grown on soil under controlled conditions. For phenotypic analysis of growth under cold conditions, plants were grown for 2 wk under LD conditions at 21 °C and then transferred to SD conditions at 4 °C or 21 °C. Further details of expression analysis are provided in the *SI Appendix*.

Genome-Wide Transcriptome Studies. RNA for RNA-seq experiments was obtained from three biological replicates, and total RNA was isolated by using an RNeasy Plant Mini Kit (Qiagen) and subsequently digested with RNase-free DNase (Ambion) according to the manufacturer's protocol. A total of 4 μg of RNA was used for library preparation using TruSeq RNA Sample Preparation (Illumina). Libraries were gel-purified from 200 to 350 bp and pair-end sequencing (100 bases read) was performed at the Cologne Center for Genomics, University of Cologne.

ChIP Experiments. For ChIP experiments, plants were grown in LD conditions for 2 wk and above-ground tissue was collected at Zeitgeber time (ZT) 8. Three independent biological replicates were performed for PEP1 ChIP-seq assays, and two were performed for FLC ChIP-seq. For ChIP on PEP1, the Pajares and *pep1-1* genotypes were used with 1 μL of PEP1 antiserum raised against the C-terminal domain of PEP1 (12). For ChIP on FLC, the *A. thaliana* genotypes Col *FRI* and *flc-3 FRI* were analyzed and 2 μL of FLC antiserum was used (15). After cross-linking, the ChIP and ChIP-seq were performed as in ref. 16 for both species. Further details are given in the *SI Appendix*.

ChIP-Seq and RNA-Seq Data Analysis. After quality checking, the Illumina sequence reads were mapped to the *A. alpina* Pajares reference genome sequence (22). The numbers of reads obtained for each replicate are listed in Dataset S3. ChIP-seq peak calling was performed for the replicates of Pajares against those of the *pep1* mutant negative control. Recommended guidelines were followed for the analysis of ChIP-seq data for quality control, read mapping, normalization, peak-calling, and assessment of reproducibility among biological replicates (75). The tools "ranger" and "wig" were used in the software PeakRanger (76) to identify read-enriched genomic regions ($P = 1 \times 10^{-6}$, q-value = 0.01, remainder of parameters set to default settings) and to generate variable-step wiggle files of read coverage. Further details of these methods and analysis of the RNA-seq data are given in the *SI Appendix*. Other bioinformatics analysis were carried out following standard procedures (77).

Bioinformatic Analysis on *cis*-Elements and Identification of Ortholog Sequences. All bioinformatics approaches to identify enriched *cis*-elements, distribution of motifs across peaks, and identification of ortholog peak sequences are given in the *SI Appendix*.

ACKNOWLEDGMENTS. We thank Uciel Chorostecki from the Institute of Molecular and Cell Biology of Rosario, Argentina, for bioinformatics assistance and Wen-Biao Jiao and Korbinian Schneeberger for providing the *Arabis montbretiana* genome assembly and annotation. This study was supported by the Alexander von Humboldt-Stiftung and Argentinean National Council of Sciences (J.L.M.), Deutsche Forschungsgemeinschaft via the Cluster of Excellence on Plant Science Grant EXC 1028 (to V.T.), EU Marie Curie Innovative Training Network SYSFLO 237909 (to P.M. and P.K.), the ERC through HyLife (G.C.), and a core grant of the Max Planck Society (to G.C.). Some computations were performed at the Poznań Supercomputing and Networking Center.

1. Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* 13:505–516.
2. Borneman AR, et al. (2007) Divergence of transcription factor binding sites across related yeast species. *Science* 317:815–819.
3. He Q, et al. (2011) High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. *Nat Genet* 43:414–420.
4. Schmidt D, et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328:1036–1040.
5. Muiño JM, et al. (2016) Evolution of DNA-binding sites of a floral master regulatory transcription factor. *Mol Biol Evol* 33:185–200.
6. Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–956.
7. Sheldon CC, et al. (1999) The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation. *Plant Cell* 11:445–458.
8. Wang R, et al. (2009) PEP1 regulates perennial flowering in Arabis alpina. *Nature* 459: 423–427.
9. Villar D, Flicek P, Odom DT (2014) Evolution of transcription factor binding in metazoans–Mechanisms and functional implications. *Nat Rev Genet* 15:221–233.
10. Bradley RK, et al. (2010) Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. *PLoS Biol* 8:e1000343.
11. Michael TP (2014) Plant genome size variation: Bloating and purging DNA. *Brief Funct Genomics* 13:308–317.
12. Albani MC, et al. (2012) PEP1 of Arabis alpina is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet* 8: e1003130.
13. Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proc Natl Acad Sci USA* 100:10102–10107.
14. de Folter S, Angenent GC (2006) trans meets cis in MADS science. *Trends Plant Sci* 11: 224–231.
15. Deng W, et al. (2011) FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. *Proc Natl Acad Sci USA* 108:6680–6685.
16. Mateos JL, et al. (2015) Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. *Genome Biol* 16:31.

17. Chiang GC, Barua D, Kramer EM, Amasino RM, Donohue K (2009) Major flowering time gene, flowering locus C, regulates seed germination in Arabidopsis thaliana. *Proc Natl Acad Sci USA* 106:11661–11666.
18. Edwards KD, et al. (2006) FLOWERING LOCUS C mediates natural variation in the high-temperature response of the Arabidopsis circadian clock. *Plant Cell* 18:639–650.
19. Mentzer L, Yee T, Wang TY, Himelblau E (2010) FLOWERING LOCUS C influences the timing of shoot maturation in Arabidopsis thaliana. *Genesis* 48:680–683.
20. Willmann MR, Poethig RS (2011) The effect of the floral repressor FLC on the timing and progression of vegetative phase change in Arabidopsis. *Development* 138: 677–685.
21. Couvreur TL, et al. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol Biol Evol* 27:55–71.
22. Willing E-M, et al. (2015) Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nat Plants* 1:14023.
23. Kaufmann K, et al. (2009) Target genes of the MADS transcription factor SEPALLATA3: Integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol* 7:e1000090.
24. Kaufmann K, et al. (2010) Orchestration of floral initiation by APETALA1. *Science* 328: 85–89.
25. Gregis V, et al. (2013) Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. *Genome Biol* 14:R56.
26. Immink RG, et al. (2012) Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. *Plant Physiol* 160:433–449.
27. Tao Z, et al. (2012) Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. *Plant J* 70:549–561.
28. Lee H, et al. (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev* 14:2366–2376.
29. Samach A, et al. (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science* 288:1613–1616.
30. Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. *Plant J* 26:15–22.
31. Hyun Y, et al. (2016) Multi-layered regulation of SPL15 and cooperation with SOC1 integrate endogenous flowering pathways at the Arabidopsis shoot meristem. *Dev Cell* 37:254–266.
32. Schwarz S, Grande AV, Bujdoso N, Saedler H, Huijser P (2008) The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. *Plant Mol Biol* 67:183–195.
33. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405:200–203.
34. Oda A, Fujiwara S, Kamada H, Coupland G, Mizoguchi T (2004) Antisense suppression of the Arabidopsis PIF3 gene does not affect circadian rhythms but causes early flowering and increases FT expression. *FEBS Lett* 557:259–264.
35. Li D, et al. (2008) A repressor complex governs the integration of flowering signals in Arabidopsis. *Dev Cell* 15:110–120.
36. Immink RG, et al. (2009) SEPALLATA3: The 'glue' for MADS box transcription factor complex formation. *Genome Biol* 10:R24.
37. Pajoro A, et al. (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol* 15:R41.
38. Wuest SE, et al. (2012) Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc Natl Acad Sci USA* 109:13452–13457.
39. Denas O, et al. (2015) Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. *BMC Genomics* 16:87.
40. Ali GS, Golovkin M, Reddy AS (2003) Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein. *Plant J* 36:883–893.
41. Hassidim M, Harir Y, Yakir E, Kron I, Green RM (2009) Over-expression of CONSTANS-LIKE 5 can induce flowering in short-day grown Arabidopsis. *Planta* 230:481–491.
42. Andrés F, et al. (2014) SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the Arabidopsis shoot apex to regulate the floral transition. *Proc Natl Acad Sci USA* 111:E2760–E2769.
43. Castillejo C, Pelaz S (2008) The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. *Curr Biol* 18:1338–1343.
44. Osnato M, Castillejo C, Matías-Hernández L, Pelaz S (2012) TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in Arabidopsis. *Nat Commun* 3:808.
45. Porri A, Torti S, Romera-Branchat M, Coupland G (2012) Spatially distinct regulatory roles for gibberellins in the promotion of flowering of Arabidopsis under long photoperiods. *Development* 139:2198–2209.
46. Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JA, Amasino RM (2003) Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15:151–163.
47. Park S, et al. (2015) Regulation of the Arabidopsis CBF regulon by a complex low-temperature regulatory network. *Plant J* 82:193–207.
48. Seo E, et al. (2009) Crosstalk between cold response and flowering in Arabidopsis is mediated through the flowering-time gene SOC1 and its upstream negative regulator FLC. *Plant Cell* 21:3185–3197.
49. Atkin OK, Loveys BR, Atkinson LJ, Pons TL (2006) Phenotypic plasticity and growth temperature: Understanding interspecific variability. *J Exp Bot* 57:267–281.
50. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* 17:287–291.
51. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
52. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134:25–36.
53. Zinzen RP, Furlong EE (2008) Divergence in cis-regulatory networks: Taking the 'species' out of cross-species analysis. *Genome Biol* 9:240.
54. Ballester B, et al. (2014) Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife* 3:e02626.
55. Stefflova K, et al. (2013) Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 154:530–540.
56. Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10:195–205.
57. Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol* 65: 505–530.
58. Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101:12404–12410.
59. Riechmann JL, Wang M, Meyerowitz EM (1996) DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res* 24:3134–3141.
60. Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. *Eur J Biochem* 229:1–13.
61. West AG, Causier BE, Davies B, Sharrocks AD (1998) DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. *Nucleic Acids Res* 26:5277–5287.
62. Melzer R, Theissen G (2009) Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Res* 37:2723–2736.
63. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462:65–70.
64. Kunarso G, et al. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42:631–634.
65. Wilson MD, et al. (2008) Species-specific transcription in mice carrying human chromosome 21. *Science* 322:434–438.
66. Muiño JM, Smaczniak C, Angenent GC, Kaufmann K, van Dijk AD (2014) Structural determinants of DNA recognition by plant MADS-domain transcription factors. *Nucleic Acids Res* 42:2138–2146.
67. Degner JF, et al. (2012) DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* 482:390–394.
68. Domcke S, et al. (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* 528:575–579.
69. Odom DT, et al. (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 39:730–732.
70. Liu C, Xi W, Shen L, Tan C, Yu H (2009) Regulation of floral patterning by flowering time genes. *Dev Cell* 16:711–722.
71. Searle I, et al. (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes Dev* 20:898–912.
72. Winter CM, et al. (2011) LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev Cell* 20:430–443.
73. Ruelens P, et al. (2013) FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat Commun* 4:2280.
74. Lee I, Amasino RM (1995) Effect of vernalization, photoperiod, and light quality on the flowering phenotype of Arabidopsis plants containing the FRIGIDA gene. *Plant Physiol* 108:157–162.
75. Bailey T, et al. (2013) Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLOS Comput Biol* 9:e1003326.
76. Feng X, Grossman R, Stein L (2011) PeakRanger: A cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* 12:139.
77. Conesa A, et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13.

Supporting Information

SI Materials and Methods

Growth conditions for transcriptomic studies

For RNA-seq of *A. alpina*, seedlings from Pajares and *pep1-1* mutant were grown for 2 weeks under LD and apices and leaves were harvested at ZT8. Three biological replicates were used for each tissue. Similar conditions were used for RNA-seq validations by real-time PCR.

For expression analysis upon cold exposure, plants of *A. thaliana* or *A. alpina* genotypes were grown for 2 weeks under SD at 21°C and then shifted to cold (4°C) for 1 day (SD) at ZT 4. Controls were kept under SD at 21°C. Samples were collected at 0, 1, 4, 8 and 24 hours after shift. Above ground tissue of 15 seedlings was harvested and used for expression analysis as described below. For vernalization of *A. alpina*, plants were grown for 5.5 weeks under LD at 21°C and then shifted to vernalization (4°C, SD) for 12 weeks. After vernalization plants were shifted back to LD at 21°C. All samples of the vernalization experiment were taken at ZT8.

RNA extraction and quantitative real-time PCR

Total RNA was isolated from leaves or apices tissues in four independent biological replicates using RNAeasy extraction kit (Qiagen) and treated with RNAse-free DNase (Ambion) to remove residual genomic DNA. 5µg of total RNA was used for reverse transcription (Superscript II, Invitrogen). Transcript levels were quantified by quantitative PCR in a CFX384 Touch Real-Time PCR System (Biorad) using the *PP2A* as house-keeping gene in both species. Primers used to quantify the expression are listed in Dataset S7. For single time point expression analysis, mean and SEM of 4 independent biological replicates were calculated. For expression time course under cold, mean and SEM was calculated for 2 independent biological replicates that were normalized to the average expression of the mutant in cold.

ChIP experiments

For chromatin immunoprecipitation (ChIP) experiments plants were grown in LD for 2 weeks and above-ground tissue was collected at ZT8. Three independent biological replicates were performed for PEP1 ChIP assays and two for FLC ChIP-seq. For ChIP on PEP1 we used Pajares and *pep1-1* mutant with 1 µl of PEP1 antiserum raised against the C-terminal domain of PEP1 protein (1). For ChIP on FLC we analyzed the Arabidopsis genotypes Col FRI and *flc-3* FRI and 2 µl of FLC antiserum was used (2). After crosslinking the tissue, the chromatin immunoprecipitation and the ChIP-seq was done as in (3) for both species. Primers for ChIP-qPCR analysis are described in Dataset S7.

ChIP-seq data analysis

For both species, low quality and duplicated reads in the raw sequencing data were filtered out using Parallel-QC v1.0 (4). Low quality reads were considered those not having Phred quality scores ≥ 13 (probability of the base called to be incorrect ≤ 0.05) in at least 90% of the bases called. Duplicated reads were also removed to achieve a better specificity (fewer false positive peaks) during the peak-calling step at each individual replicate as recommended in Bailey et al. (2013) (5). Reads kept were then mapped either to the *Arabis alpina* reference assembly V3 (6) (8 chromosomes, and > 15 kbp scaffolds) or to the *Arabidopsis thaliana* TAIR10 using Bowtie v2.0.2 (7), reporting alignments under default mode parameters (i.e., reporting the 'best' alignment if multiple mapping locations were found for a read). We used the tools 'ranger' and 'wig' in the software PeakRanger (8), to identify read-enriched genomic regions (*P*-value < 1e-6, q-value (FDR) < 0.01, rest of parameters default), and to generate variable step wiggle files of read coverage. Reads were extended in the appropriate strand to an approximated average fragment size of 300 bp. Using MULTOVL v1.2 with parameters '-L 1 -u -m 2 -M 0 -s multovl -f BED' (9), a total of 222 peak regions present in at least two replicates of PEP1 ChIP-seq or 377 peak regions present in the two replicates of FLC ChIP-seq were retained for further processing (the union of overlapping regions was reported, minimum overlap required 1bp). After filtering by peak length (peaks shorter than 1.5kb were kept), we obtained the final peak set for both species, namely 204 peaks in *Arabis alpina* and 297 peaks in *Arabidopsis thaliana*. Final sets of peaks were annotated to V3 of the *Arabis alpina* annotation, or to TAIR10 (*Arabidopsis*) using the Bioconductor package CSAR (10). Target genes of the final set of ChIP-seq peaks were defined as those having the start or end of a peak region in a distance spanning from 5 kb upstream (i.e., from start of the gene) to 3 kb downstream (i.e., from end of the gene) for *A. alpina* or spanning from 3 kb upstream to 1 kb downstream for *A. thaliana*. Visualization of peaks was done with IGV (11).

Analysis of peak distribution over exon, intron, enhancer, proximal promoter, 5´ UTR and 3´ UTR was done in R using ChIPpeakAnno (12). Proximal promoter and immediate downstream sequence were considered as respectively 3kb upstream of TSS and 1kb downstream the TTS.

Genome-wide transcriptome analysis.

Low-quality reads in the raw data (FASTQ files) were filtered out and reads kept were then mapped to the *Arabis alpina* reference assembly V3 (6) using TopHat with the GTF file used in (6) (--min-anchor-length 10; --max-multihits 5; library-type fr-unstranded) (13). Raw counts of mapped reads were used for differential gene expression analysis in the R/Bioconductor package DESeq2 1.10.0 (14), and genes with adjusted *P*-value < 0.05 and $\log_2$ (fold-change) > 1.5 were considered differentially expressed. Principal Component Analysis (PCA) of the data uncovered high reproducibility among replications, while the tissues were separated by principal component 1 and the genotypes by principal component 2 (Fig. S3). For *A. thaliana* FLC transcriptome we used the data from LD condition from (3). Venn diagrams in Figures 1 and 2 were generated using the online tool Venny 2.0 (15).

Analysis of growth in *A. alpina* in cold

To determine growth of *A. alpina* seedlings, plants were photographed from above and the diameter was measured using the Image-J 1.43u software (16). Plants used as controls were grown for 3 weeks and diameter was measured using a ruler. Data of two independent biological replicates with n ≤ 13 were pooled and mean and SEM were calculated, significance was tested by Student's t-test.

Identification of enriched cis-elements

De novo motif enrichment analysis was performed using MEME (17) with the 'zoops' (motif occurrence zero or one per sequence) or 'anr' (any number of motif occurrence per sequence) model and revealed similar results. Parameters were set to identify the 10 most significantly enriched motifs of a length between 5 and 20 nucleotides. In addition, DREME (18) with the standard settings was used to screen for shorter motifs. CArG-boxes shown here were identified by MEME with the 'zoops' model and the G-box and TCP motif were identified by DREME.

Permutation test for enriched cis-elements

To test significance of enrichment of *de novo* identified motifs or candidate motifs in different subsets of BSs, we performed a permutation test. Therefore, we counted the number of BSs in the subset of interest that contain at least one motif (for CArG we allowed variation and searched for MYHWAWWWRGWWW, for the TCP-binding motif, we searched for GGMCCA and for the G-box (CACGTG), we did not allow sequence variation). For the permutation test, we generated 1000 samples of random genomic sequences of the same size as the BS subset of interest. For these samples, the mean number of sequences that contain at least one motif as well as the standard deviation was calculated. Finally we calculated the Z-score, representing the number of standard deviations by which the number of BSs with motif in the subset of interest differed from the mean of random sequences. Z-scores above 3 were considered significant. To test for significant enrichment of TTT at position 1-3 of CArG-boxes, we counted the number of TTTMYHWAWWWRGWWW in the different subsets of BSs. Z-score was calculated using 1000 random samples of all CArG-boxes (NNNMYHWAWWWRGWWW) identified in the total BSs of each species.

Distance distribution of cis-elements to peak center

Distance of the peak center to the different cis-elements was calculated. For CArG-boxes, all motifs matching MYHWAWWWRGWWW were considered. For G-boxes, all motifs matching the canonical CACGTG were considered. For the TCP-binding element, all motifs matching GGMCCA were considered. Density graph was plotted with R statistical environment.

Identification of orthologous sequences in other species

BSs that were bound in *A. alpina* and *A. thaliana* were identified using BLAST. *A. alpina* BSs were aligned against *A. thaliana* BSs with BLAST (Word size for wordfinder algorithm 11, Penalty for a nucleotide mismatch -3, Penalty for a nucleotide match 2, Cost to open a gap 5, Cost to extend a gap 2, Discontiguous MegaBLAST template length 18, window size 40). To test if BSs identified by the BLAST search are bound in regions with conserved synteny and to identify orthologous sequences of BSs in other species, BSs were aligned to orthologous regions. Therefore, orthologs of genes associated to the BSs were identified by reciprocal blast and these genes, including 5 kb upstream of the transcriptional start site and 3 kb downstream of the transcriptional end site were extracted as orthologous loci. Finally, the best alignment of the BS in the orthologous locus was obtained by local Smith and Waterman alignments (19). Data were represented using Circos (20). Genome assemblies and annotation files of *A. lyrata* (A. lyrata v1.0) (21) were downloaded from Phytozome v11.0. Assemblies and annotations of the genomes of *A. arabicum* (22) and *T. hassleriana* (23), were downloaded from https://genomevolution.org/coge/. The *A. montbretiana* genome assembly and annotation was kindly provided by Wen-Biao Jiao and Korbinian Schneeberger prior

to publication (personal communication). Alignments were done using mVISTA (24) and GATA 1.0 using standard settings (25).

Analysis of overlapping binding sites of different TFs

Genome-wide BSs of *AP1, AP3*, *PI*, *AG*, *SOC1* and *SVP* were publically available as supplementary files (3, 26-29). *FLM* BSs (30) were downloaded from GEO (GSE48082) and peaks with FDR<0.1 were selected. *SEP3* BSs and *A. lyrata SEP3* BSs in the *A. thaliana* genome (31) were defined as 100 bp up- and downstream of the position with maximum ChIP-seq score using peaks with FDR < 0.01. Overlap of these BSs with the 200 central bp of *FLC* BSs or *PEP1* BSs in the *A. thaliana* genome were identified using the BEDtools intersect v2.16.2 (32).

Functional category enrichment analysis

Significant GO terms were identified using the BioMaps tool from the Virtual Plant (33) software (http://virtualplant.bio.nyu.edu/cgi-bin/vpweb webcite). The category named "flowering-time genes" was based on the gene list obtained from http://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes). To study *Arabis alpina* genes, GO terms were identified using the orthologous gene from *Arabidopsis*. We selected a subset of significant (*P*-value < 0.05) GO terms of our interest and for each term we calculated the representation factor. The representation factor (RF) is the proportion of genes that fall in a certain category in each dataset divided by the proportion of genes in that category in the genome. A representation factor > 1 indicates more overlap than expected by chance, a representation factor < 1 indicates less overlap than expected. The probability of each overlapping was determined using the hypergeometric probability distribution.

Data Visualization

RF for GO term analysis in Figure 4 was plotted as a bubble plot in R using ggplot2 package (34). ChIP-seq data visualization in Figure 3 and 4 was done using IGV (11). Concentric plot on Figure 2 was done with Circos (20). Results of mVISTA analysis was visualized with VISTA visualization tool (24). Expression data was visualized with Java TreeView 1.1.6r2. Venn diagrams were represented with Venny 2.0 (15).

Phylogenetic tree construction

Orthofinder (35) was used for generating clusters of orthologous genes from the following set of species: *Arabis alpina*, *Arabis montbretiana*; *Arabidopsis lyrata* (annotation obtained from (36); *Brassica rapa*, *Capsella grandiflora*, *Capsella rubella* and *Eutrema salsugineum* (annotations obtained from https://phytozome.jgi.doe.gov); *Aethionema arabicum* and *Tarenaya hassleriana* (annotation obtained from: https://genomevolution.org). Only those proteins corresponding to the representative transcripts were used for orthologous group construction.

A multiple sequence alignment of the protein sequences in the cluster containing the *A. thaliana FLC* gene was using muscle (37) and trimmed using trimAl (38) The trimmed alignment was subsequently used as guide to align the corresponding coding sequences.

jModelTest (39) was used to determine the most suitable substitution model for maximum likelihood phylogenic tree reconstruction based on the AIC criterion. The maximum likelihood tree was constructed using PhyML (40).

Synteny reconstruction

Synteny blocks between *A. thaliana* and *T. hassleriana* were determined using i-ADHoRe (41) using the orthology group assignments obtained using orthofinder.

1. Albani MC, Castaings L, Wotzel S, Mateos JL, Wunder J, Wang R, Reymond M, Coupland G (2012) **PLoS Genet** 8, e1003130.
2. Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES (2011) **Proc Natl Acad Sci U S A** 108, 6680-6685.
3. Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, Fornara F, Schneeberger K, Krajewski P, Coupland G (2015) **Genome Biol** 16, 31.
4. Zhou Q, Su X, Wang A, Xu J, Ning K (2013) **PLoS One** 8, e60234.
5. Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, Madrigal P, Taslim C, Zhang J (2013) **PLoS Comput Biol** 9, e1003326.
6. Willing E-M, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, Becker C, Warthmann N, Chica C, Szarzynska B, Zytnicki M, Albani MC, Kiefer C, Bergonzi S, Castaings L, Mateos JL, Berns MC, Bujdoso N, Piofczyk T, de Lorenzo L, Barrero-Sicilia C, Mateos I, Piednoël M, Hagmann J, Chen-Min-Tao R, Iglesias-Fernández R, Schuster SC, Alonso-Blanco C, Roudier F, Carbonero P, Paz-Ares J, Davis SJ, Pecinka A, Quesneville H, Colot V, Lysak MA, Weigel D, Coupland G, Schneeberger K (2015) **Nature Plants** 1.
7. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) **Genome Biol** 10, R25.
8. Feng X, Grossman R, Stein L (2011) **BMC Bioinformatics** 12, 139.
9. Aszodi A (2012) **Bioinformatics** 28, 3318-3319.
10. Muino JM, Kaufmann K, van Ham RC, Angenent GC, Krajewski P (2011) **Plant Methods** 7, 11.
11. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) **Brief Bioinform** 14, 178-192.
12. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR (2010) **BMC Bioinformatics** 11, 237.
13. Trapnell C, Pachter L, Salzberg SL (2009) **Bioinformatics** 25, 1105-1111.
14. Love MI, Huber W, Anders S (2014) **Genome Biol** 15, 550.
15. Oliveros JC. An interactive tool for comparing lists with Venn's diagrams. 2007 2015 [cited; Available from: http://bioinfogp.cnb.csic.es/tools/venny/index.html
16. Schneider CA, Rasband WS, Eliceiri KW (2012) **Nat Methods** 9, 671-675.
17. Bailey TL, Elkan C (1994) **Proc Int Conf Intell Syst Mol Biol** 2, 28-36.
18. Bailey TL (2011) **Bioinformatics** 27, 1653-1659.
19. Smith TF, Waterman MS (1981) **J Mol Biol** 147, 195-197.
20. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) **Genome Res** 19, 1639-1645.
21. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, Van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL (2011) **Nat Genet** 43, 476-481.
22. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, Dewar K, Stinchcombe JR, Schoen DJ, Wang X, Schmutz J, Town CD, Edger PP, Pires JC, Schumaker KS, Jarvis DE, Mandakova T, Lysak MA, van den Bergh E, Schranz ME, Harrison PM, Moses AM, Bureau TE, Wright SI, Blanchette M (2013) **Nat Genet** 45, 891-898.
23. Cheng S, van den Bergh E, Zeng P, Zhong X, Xu J, Liu X, Hofberger J, de Bruijn S, Bhide AS, Kuelahoglu C, Bian C, Chen J, Fan G, Kaufmann K, Hall JC, Becker A, Brautigam A, Weber AP, Shi C, Zheng Z, Li W, Lv M, Tao Y, Wang J, Zou H, Quan Z,

Hibberd JM, Zhang G, Zhu XG, Xu X, Schranz ME (2013) **Plant Cell** 25, 2813-2830.
24. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) **Bioinformatics** 16, 1046-1047.
25. Nix DA, Eisen MB (2005) **BMC Bioinformatics** 6, 9.
26. DS OM, Wuest SE, Rae L, Raganelli A, Ryan PT, Kwasniewska K, Das P, Lohan AJ, Loftus B, Graciet E, Wellmer F (2013) **Plant Cell** 25, 2482-2503.
27. Kaufmann K, Wellmer F, Muino JM, Ferrier T, Wuest SE, Kumar V, Serrano-Mislata A, Madueno F, Krajewski P, Meyerowitz EM, Angenent GC, Riechmann JL (2010) **Science** 328, 85-89.
28. Tao Z, Shen L, Liu C, Liu L, Yan Y, Yu H (2012) **Plant J** 70, 549-561.
29. Wuest SE, O'Maoileidigh DS, Rae L, Kwasniewska K, Raganelli A, Hanczaryk K, Lohan AJ, Loftus B, Graciet E, Wellmer F (2012) **Proc Natl Acad Sci U S A** 109, 13452-13457.
30. Pose D, Verhage L, Ott F, Yant L, Mathieu J, Angenent GC, Immink RG, Schmid M (2013) **Nature** 503, 414-417.
31. Muino JM, de Bruijn S, Pajoro A, Geuten K, Vingron M, Angenent GC, Kaufmann K (2016) **Mol Biol Evol** 33, 185-200.
32. Quinlan AR, Hall IM (2010) **Bioinformatics** 26, 841-842.
33. Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE, Coruzzi GM, Gutierrez RA (2010) **Plant Physiol** 152, 500-515.
34. Wickham H. Elegant Graphics for Data Analysis: Springer-Verlag New York; 2009.
35. Emms DM, Kelly S (2015) **Genome Biol** 16, 157.
36. Rawat V, Abdelsamad A, Pietzenuk B, Seymour DK, Koenig D, Weigel D, Pecinka A, Schneeberger K (2015) **PLoS One** 10, e0137391.
37. Edgar RC (2004) **Nucleic Acids Res** 32, 1792-1797.
38. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) **Bioinformatics** 25, 1972-1973.
39. Posada D, Crandall KA (1998) **Bioinformatics** 14, 817-818.
40. Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) **Methods Mol Biol** 537, 113-137.
41. Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, Vandepoele K (2012) **Nucleic Acids Res** 40, e11.

Supplementary figures

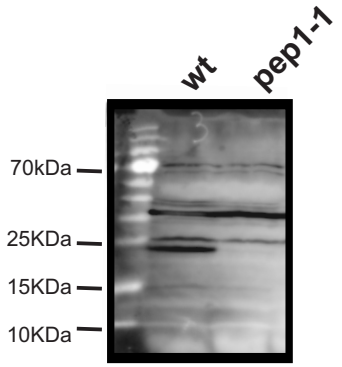


Figure S1. Antiserum used for PEP1 ChIP experiments
Western blot of PEP1 antiserum tested against wild-type and *pep1-1* mutant plants grown for 2 weeks in LD. Equal loading can be estimated by comparing unspecific signals. Total protein extract was used.

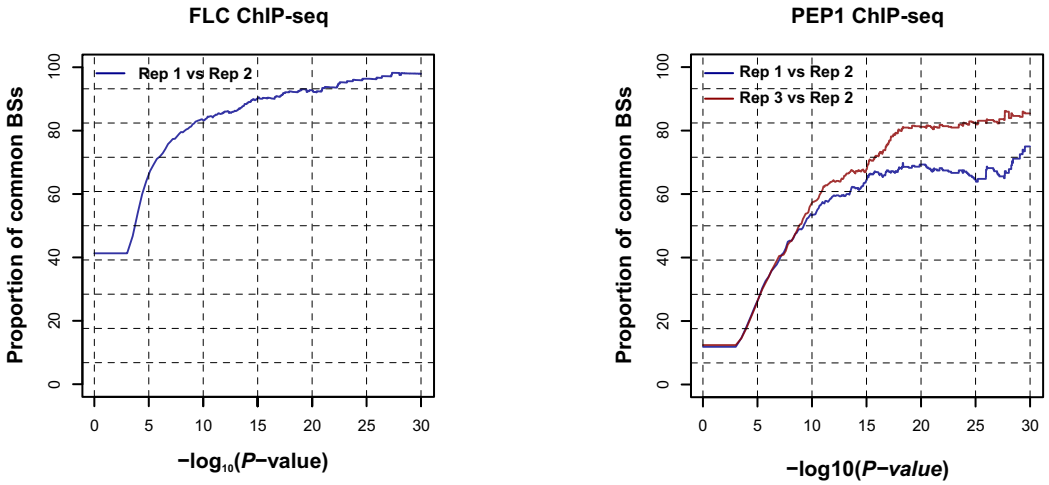


Figure S2. Proportion of conserved PEP1 and FLC BSs between replicates.
Left: Proportion of common BSs between replicate 1 and replicate 2 of FLC at different *P*-values as cut-off for ChIP-seq data sets. Right: Proportion of common BSs between replicate 2 and replicate 1 or replicate 3 of PEP1 at different *P*-values as cut-off for ChIP-seq data sets.

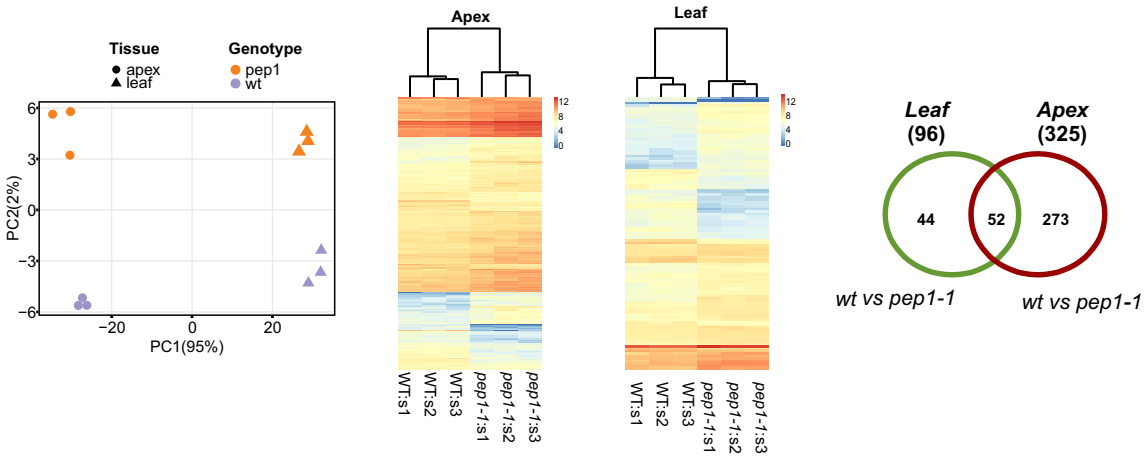


Figure S3. Genome-wide transcriptome analysis.
Transcriptional profile of *pep1-1* in leaves and apices. Left: PCA plot of the 12 samples for their first two principal components. Middle: Heat maps showing the expression data of the DEG from leaves and apices of *pep-1* vs wild-type. The data are normalized counts of each replicate (s1-s3). Hierarchical clustering was done and distances were computed by complete linkage method and visualized in the tree above the heatmap. Right: Venn diagram of DEG in *pep1-1* compared to wild-type in leaves and apices. Plants were grown for 2 weeks in LD conditions, and leaves and apices were collected at ZT8. Only genes with $\log_2$ (fold-change) > 1.5 and adjusted *P*-value < 0.05 were selected as DEG.

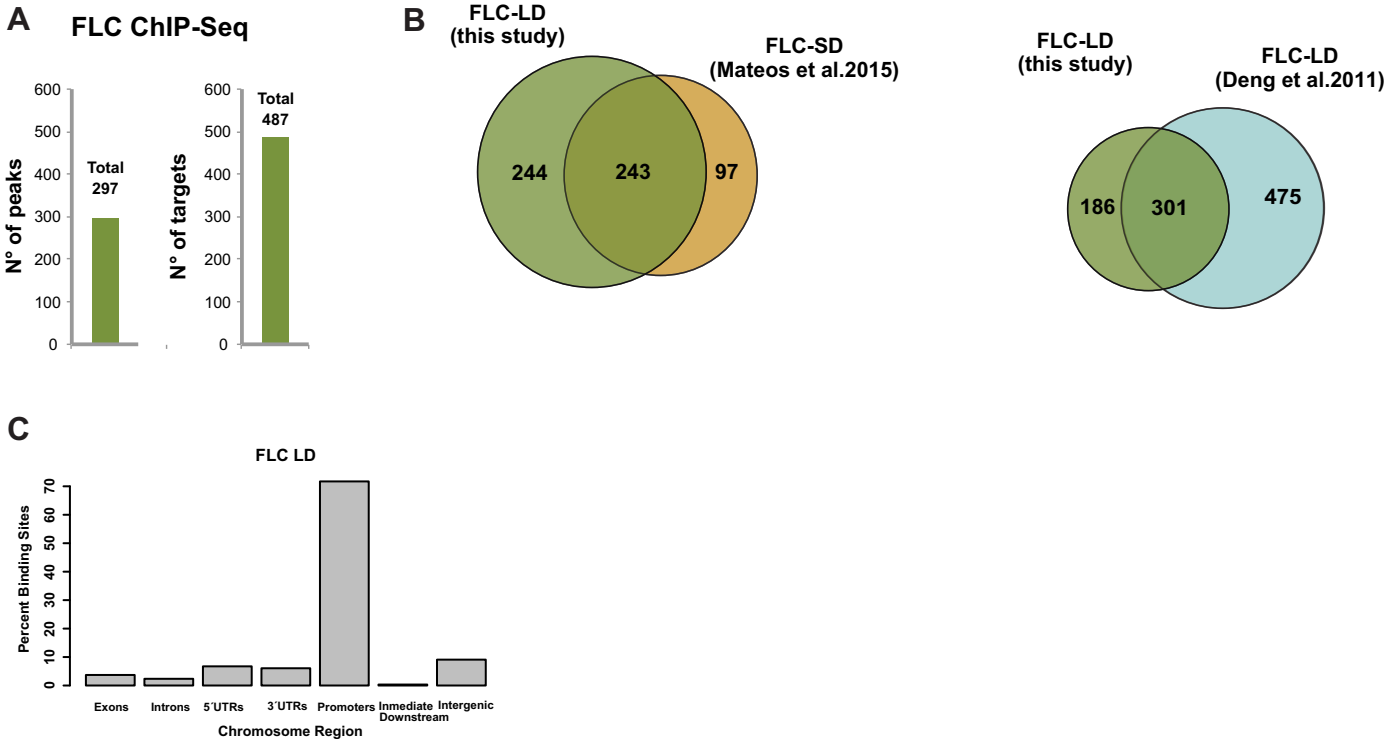
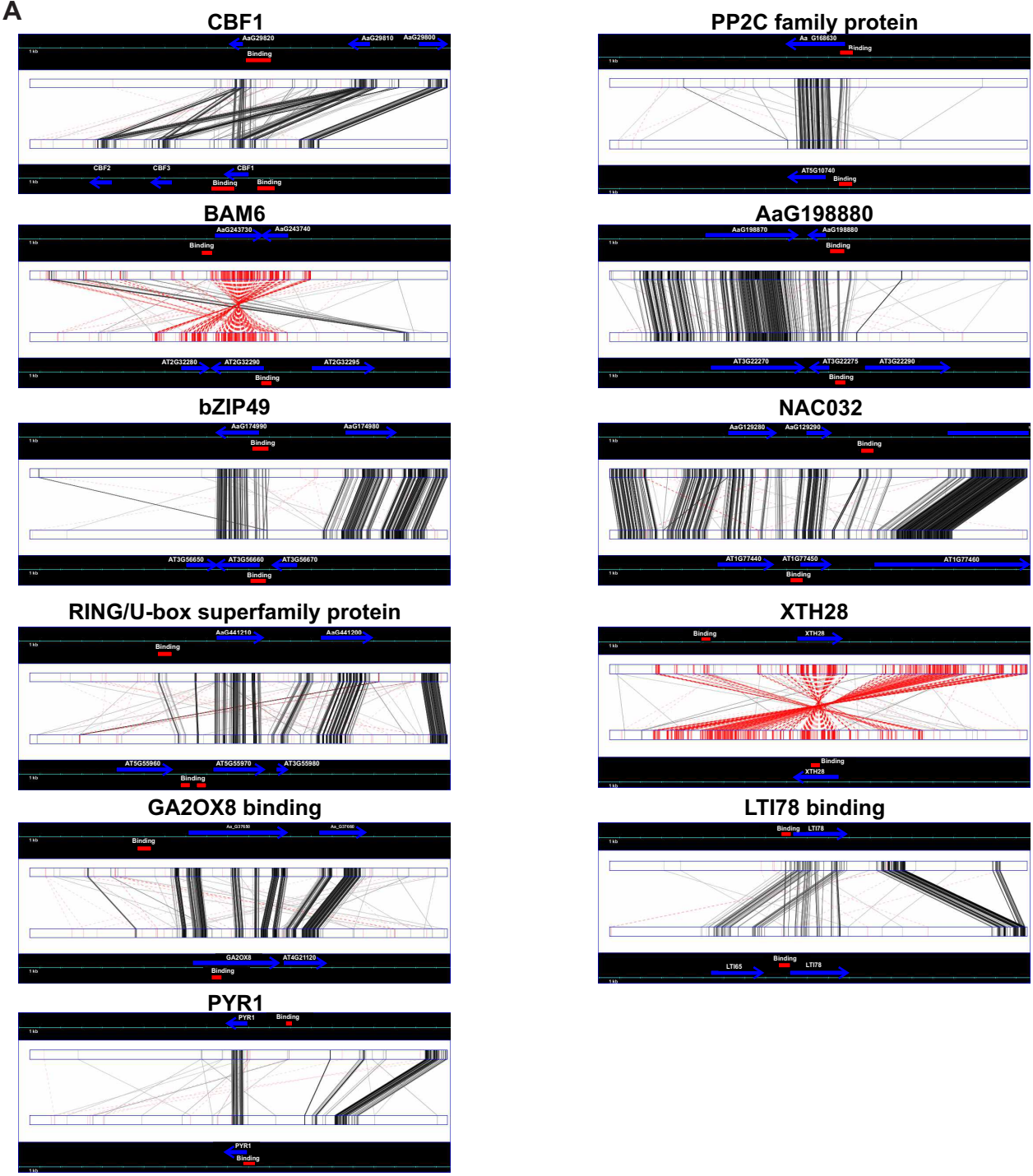


Figure S4. FLC targets in *A. thaliana* in LD.
(A) Number of significant FLC ChIP-seq peak calls and their associated genes in LD. Numbers above the bars indicate total numbers. *(B)* Venn diagrams showing common sets of target genes for FLC between this and previous studies. *(C)* FLC peak distribution over different genomic features in the *A. thaliana* genome.

A



B

Gene ID Aa	Gene ID At	Description	*A. alpina*	*A. thaliana*	comment
AaG242480	AT1G14720	ATXTH28	promoter	intron	
AaG152250	AT4G17870	PYR1	3´UTR	3´Downstream	
AaG37650	AT1G21200	GA2OX8	promoter	intron	
AaG29820	AT4G25490	CBF1	promoter	promoter+3´ UTR	a,b
AaG168630	AT5G10740	PP2C protein	promoter	promoter +5´UTR	b
AaG396870	AT5G52310	LTI78	promoter	promoter	b
AaG1292290	AT1G77450	NAC032	3´Downstream	promoter +5´UTR	
AaG243730	AT2G32290	BAM6	promoter	promoter +5´UTR+exon	b
AaG198980	AT3G22275	unkonwn	promoter	promoter	b
AaG174990	AT3G56660	bZIP49	promoter +5´UTR+exon	promoter +5´UTR+exon	b
AaG441210	AT5G55970	RING/U-box	promoter	promoter	a,b

*a= extra peak in Arabidopsis
*b= low conservation of sequence around PEP1 binding

Figure S5. Common target genes among PEP1 and FLC that are bound at different BSs. (A) Sequence alignment of PEP1 (upper sequence) and FLC (lower sequence) around binding regions created using GATA 1.0. Red bars represent BSs, blue bars represent genes. Black lines indicate homology, red lines are inversions. *(B)* Table listing features of BSs described in (A).

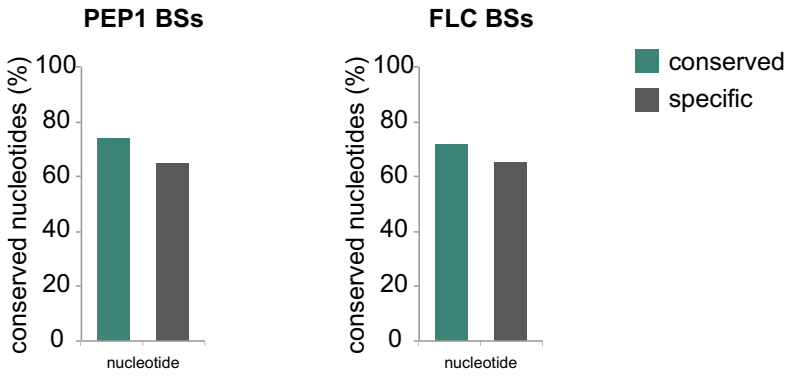


Figure S6. Nucleotide conservation of conserved and species-specific BSs.
(Left) Average nucleotide conservation of alignments of conserved and *A. alpina* specific PEP1 BSs in orthologous *A. thaliana* regions. (Right) Average nucleotide conservation of alignments of conserved and *A. thaliana* specific FLC BSs in orthologous *A. alpina* regions.

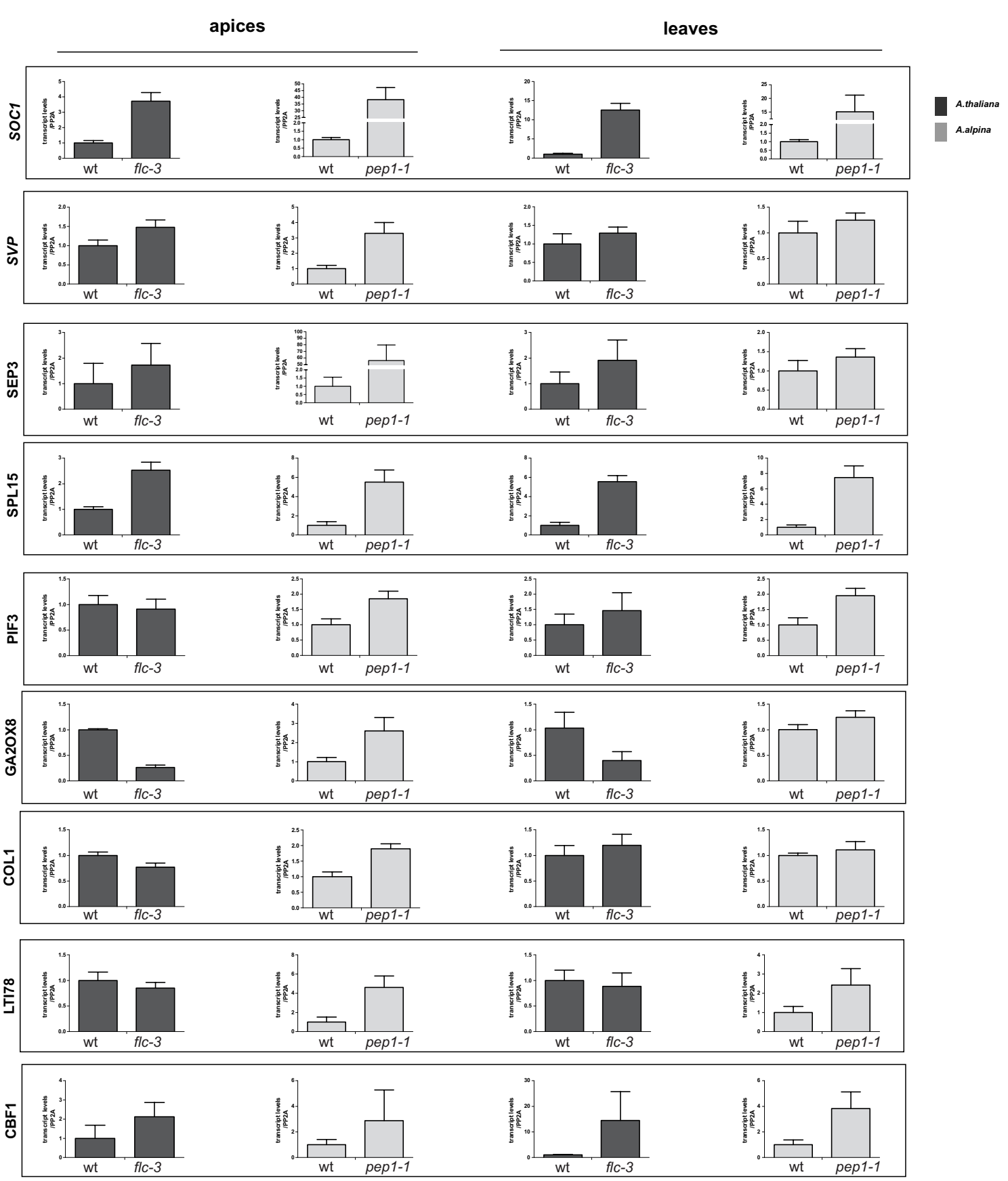


Figure S7. Transcript levels of common targets between PEP1 and FLC in both species. Real-time PCR was performed using leaves and apices of 2 weeks-old seedlings grown in LD. Expression was normalized to *PP2A* and shown relative to expression level in wild-type. Data are shown as mean ± SEM (n=4).

```
PEP1      MGRKKLEIKRIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSSSGKLYSFSSG
FLC       MGRKKLEIKRIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
          **************************************************:**********

PEP1      DNLVKILDRYGKRHADDLKALDLQSLLHGQDLQSKALNYGSHHELLEIVESKLVAPNVNN
FLC       DNLVKILDRYGKQHADDLKALDH---------QSKALNYGSHYELLELVDSKLVGSNVKN
          ************:*********          **********:****:*:****. **:*

PEP1      VSFDTLVQLEKHLETALAVVRAKKTELMLKLLESLKEKEKLLKEENQVLASQMEKKTLVG
FLC       VSIDALVQLEEHLETALSVTRAKKTELMLKLVENLKEKEKMLKEENQVLASQMENNHHVG
          **:*:*****:******:*.***********:*.******:*************::  **

PEP1      AEADDNMEISPGEISDINLPVTLPLLN
FLC       AEAE--MEMSPAGQISDNLPVTLPLLN
          ***:  **:**.   . **********

Score          Expect  Method                          Identities     Positives      Gaps
296 bits(758)  1e-106  Compositional matrix adjust.  172/208(83%)   185/208(88%)   13/208(6%)
```

Figure S8. Alignment of PEP1 and FLC proteins using ClustalW.
MADS-domain is labeled in red.

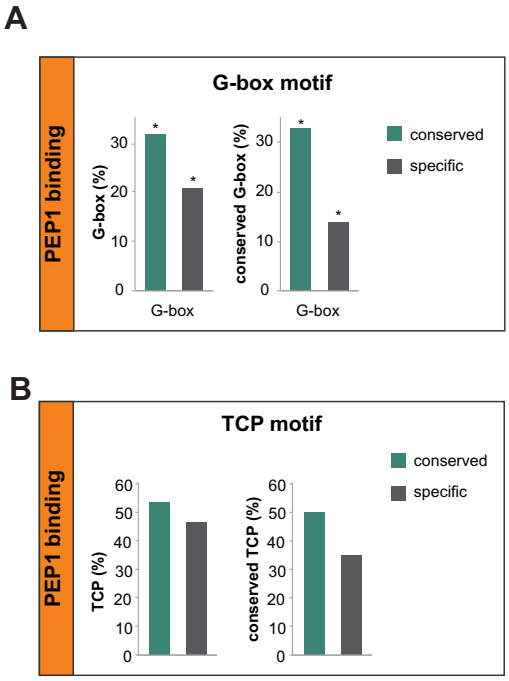


Figure S9. Percentage of *A. alpina* BSs that contain a G-box or TCP-binding motif.
(A) Percentage of *A. alpina* BSs that contain a G-box. Right, Percentage of orthologous regions in *A. thaliana* that contain a G-box. (B) Percentage of *A. alpina* BSs that contain a TCP-binding motif. Right, Percentage of orthologous regions in *A. thaliana* that contain a TCP-binding motif. Asterisks indicate significant enrichment (Z-score > 3).

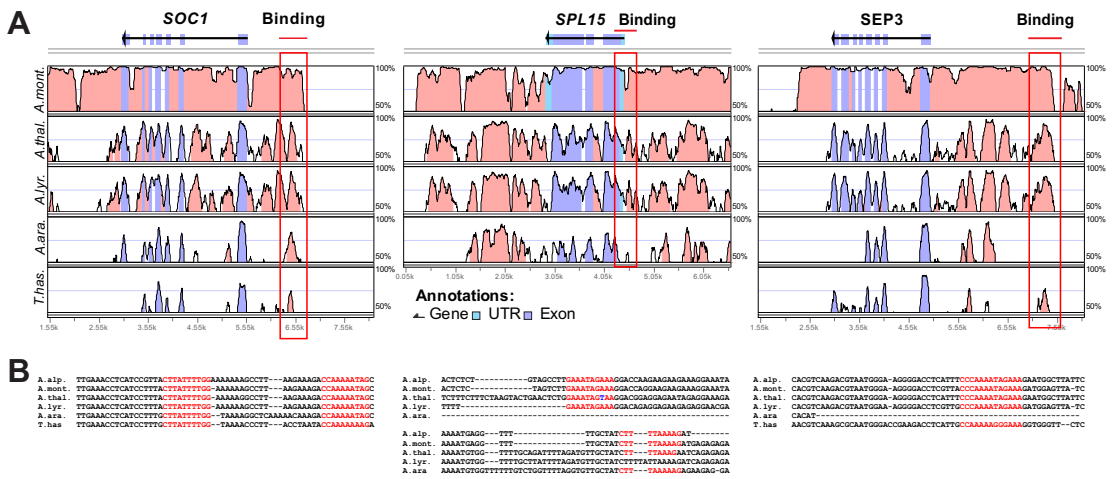


Figure S10. Conservation of CArG-box at different PEP1 BS across species.
(A) Comparative sequence analysis of PEP1 BSs at the 3 loci *SOC1*, *SPL15* and *SEP3* across six species. Approximately 6-kb regions of *A. montbretiana*, *A. thaliana*, *A. lyrata*, *A. arabicum* and *T. hassleriana* were aligned to the genomic region of *A. alpina*. All pairwise alignments are presented as VISTA plots. Pink color indicates regions with at least 70% conservation on a 100-bp sliding window. Red box indicates PEP1 BS. *(B)* Multiple sequence alignment of corresponding region around the CArG-box at PEP1 BSs of *SOC1*, *SPL15* and *SEP3* showing conservation of the motif across species. CArG-box in red in all sequences where it is present. Variations at the *cis*-element that do not disrupt the consensus sequence are shown in blue.

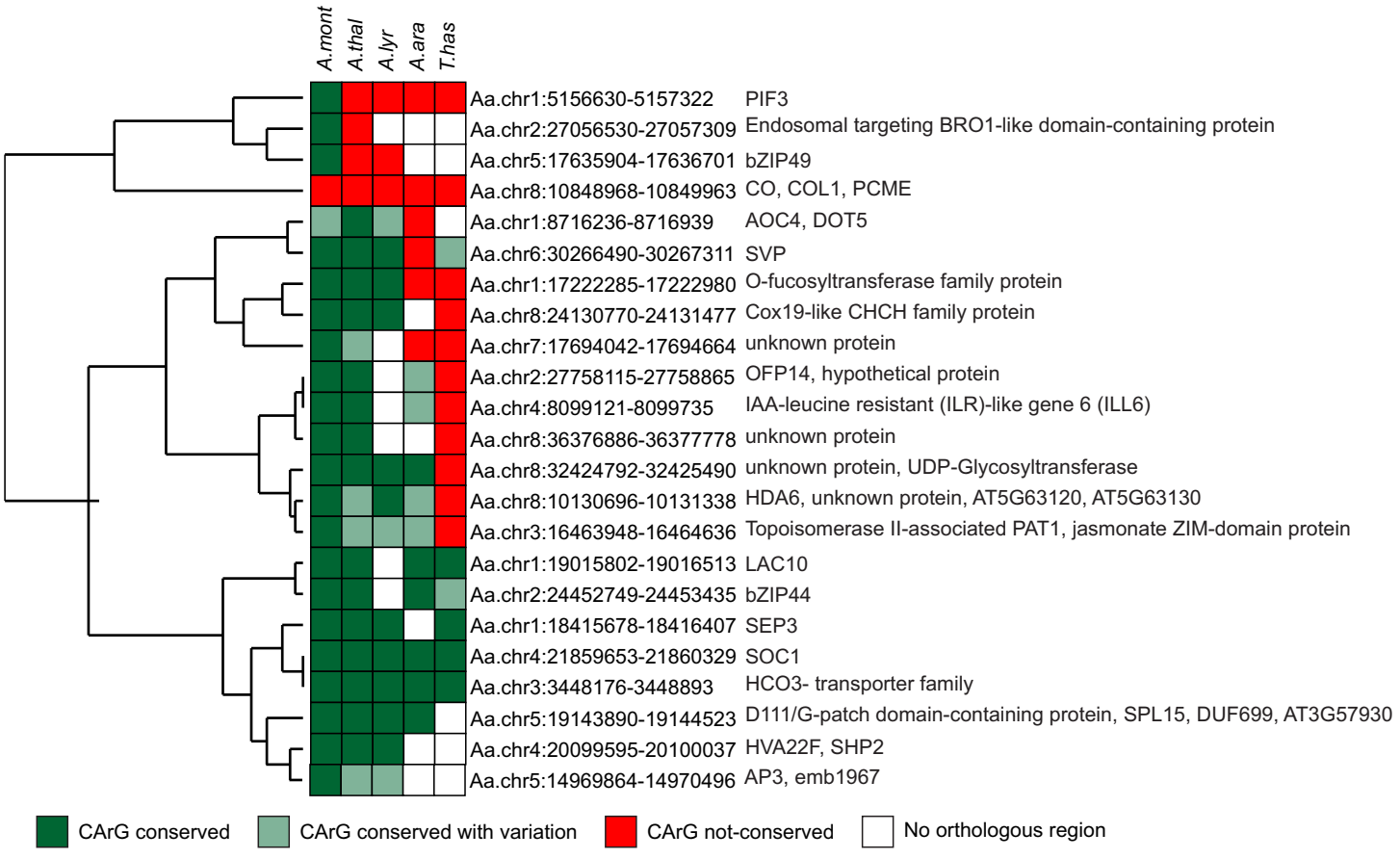


Figure S11. Heatmap showing conservation of CArG-box in different species.
Heatmap includes orthologous regions for all conserved *A. alpina* PEP1 BSs. Hierarchical clustering method organized regions in a tree structure, based on their similarity for conservation across species. Associated genes are listed

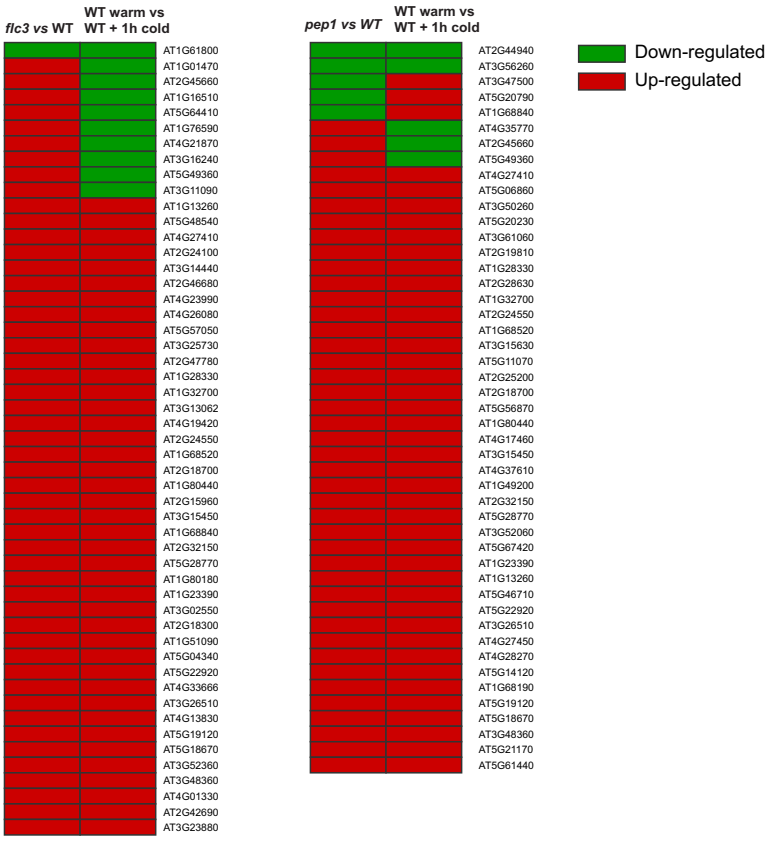


Figure S12. Heatmap showing effect of 1 h of cold-treatment compared to effect of *PEP1/FLC* mutation for a subset of COR genes.
(Left) Heatmap representing direction of expression (up- or down-regulated) of genes in flc-3 mutant compared to WT (as in (17)) compared to 1h cold treatment in WT as in (49). (Right) Heatmap representing direction of expression (up- or down-regulated) of genes in *pep1-1* mutant compared to WT (this study) compared to 1h cold treatment in WT as in (49). Up-regulation in the mutants or cold-treated plants is shown in red and down-regulation in green. For most of the genes, the effect of the cold-treatment was similar to the effect of the mutation. Data for cold-treated plants obtained from (49)

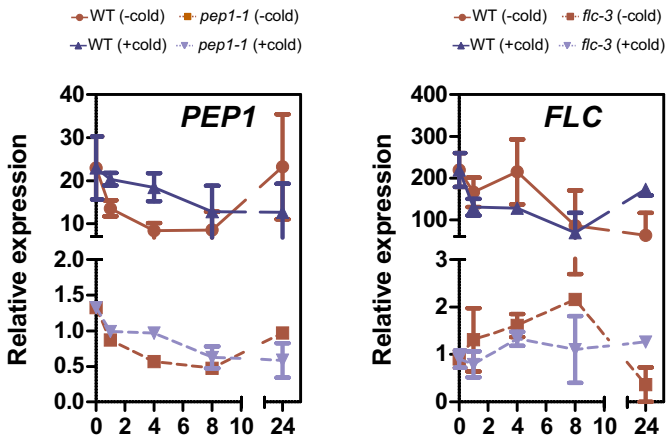


Figure S13. Expression of *PEP1* and *FLC* in response to short-term cold.
PEP1 expression in *A. alpina and FLC* expression in *A. thaliana* after transferring plants to 4°C for 24 hours compared to control conditions of maintaining plants at 22°C. Data are shown as mean ± SEM (n=2). Each experiment was normalized to the average expression of the mutant in cold.

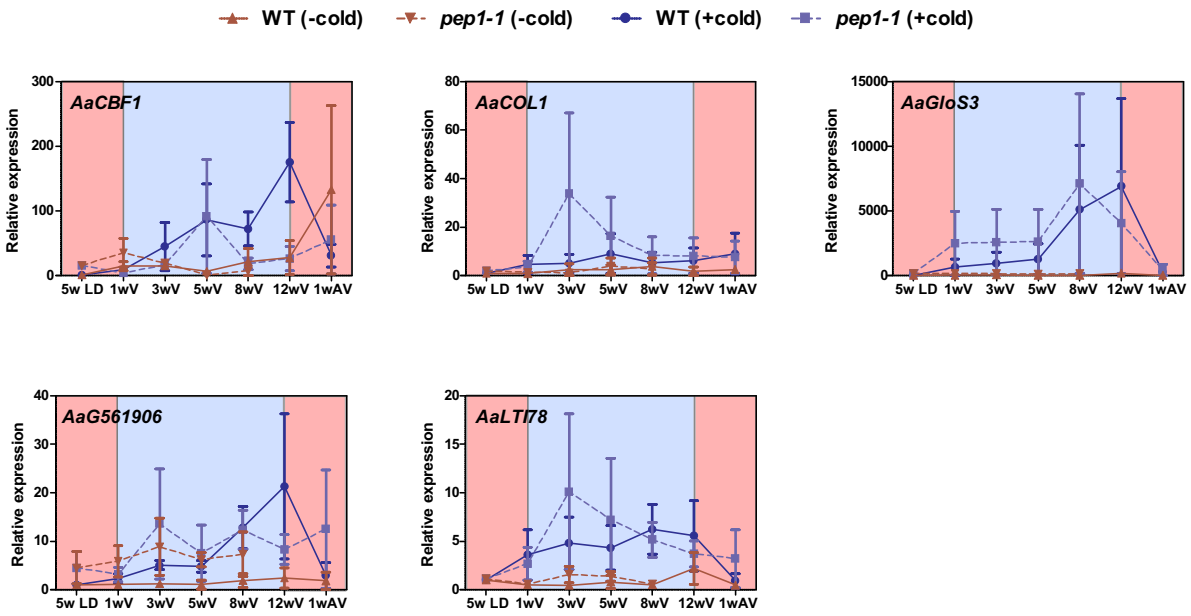


Figure S14. Effect of mutation in *PEP1* on expression of cold-regulated genes during vernalization.
Plants were grown for 5.5 weeks at 21°D under LD, then transferred to 4°C, SDs for 12 weeks and the shifted back to 21°C, LDs. Data are shown as mean ± SEM (n=2). Each experiment was normalized to expression in the wild-type at the start of the experiment

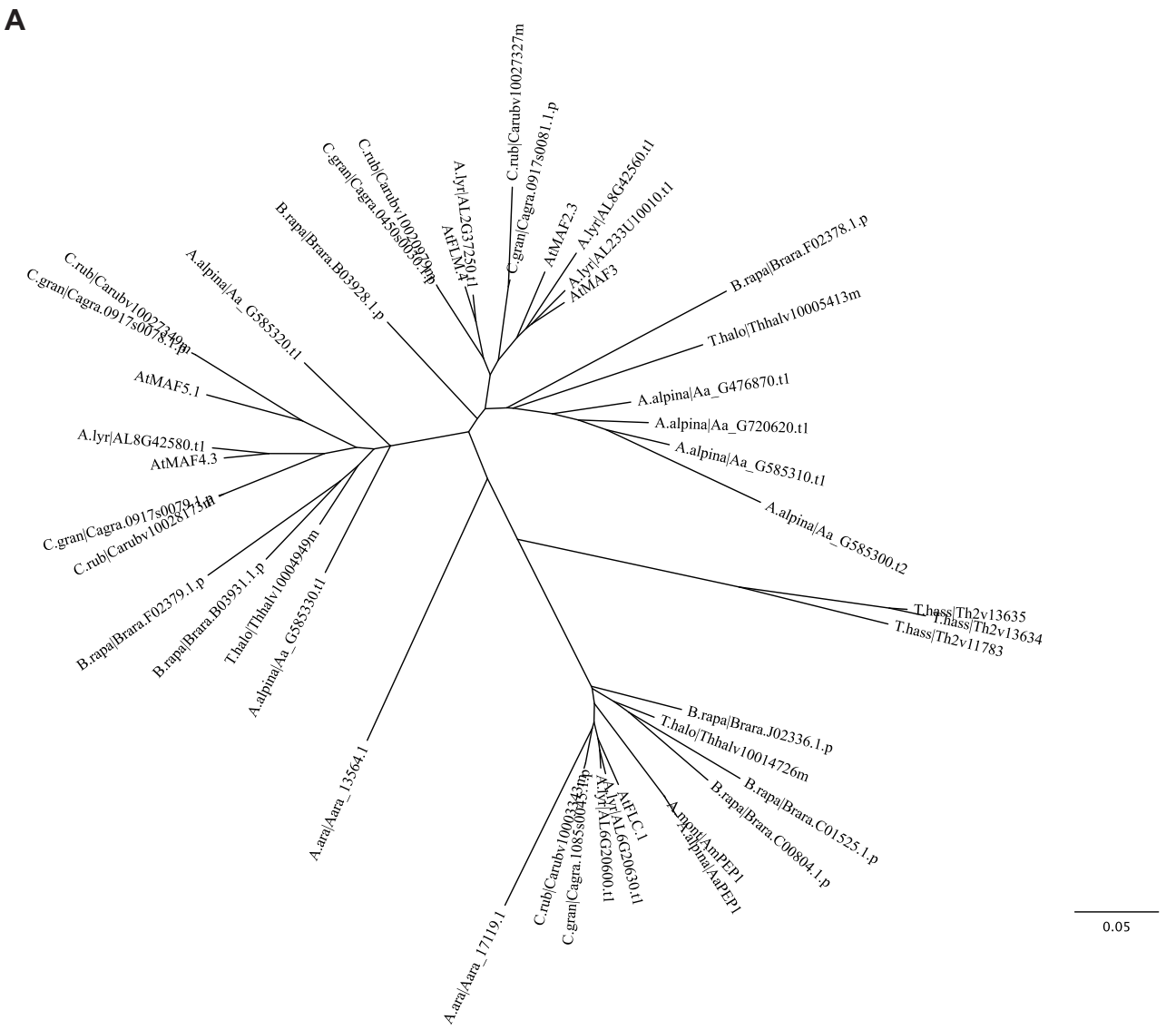


Figure S15. Phylogenetic analysis of *FLC* orthologs in Cleomaceae (A) Unrooted maximum likelihood tree of the gene sequences in the orthologous group containing the *A. thaliana FLC* gene. (B) Synteny conservation around *A. thaliana* and *T. hassleriana FLC* genes. The *A. thaliana FLC* and *T. hassleriana FLC*-like genes are indicated with bold lines. Note, for clarity, the orginal i-ADHoRe output was modified such that genes belonging to different ortholog groups were not aligned. In addition, genes masked out by i-ADHoRE as being tandem duplications were reinserted into the synteny alignment.

Dataset S1. High confidence TF BSs and associated target genes of PEP1 and FLC in LD. Common targets between FLC and PEP1, and flowering related targets are listed. Peaks in each replicate are shown. PEP1 targets obtained with shorter genomic distances is added.

Dataset S2. DEG in *pep1-1* vs wild-type *A. alpina* in leaves and apices.

Dataset S3. Metrics of the Sequencing.

Dataset S4. FLC target genes identified by Deng et al. 2010 but not in this study which are bound by PEP1 in this study.

Dataset S5. Percent of identity of BSs of PEP1 and FLC against the corresponding genome of the other species.

Dataset S6. List of COR genes among PEP1 and FLC direct targets and among DEG in *pep1-1* and *flc-3*.

Dataset S7 Primers used for ChIP-PCR and qRT-PCR analysis