The evolutionary analysis of PEBP Gene Family among Rosaceae tree species

CURRENT STATUS: UNDER REVIEW

Man Zhang
Beijing Forestry University

Ping Li
Beijing Forestry University

Xiaolan Yan
Mei Flower Research Center

Jia Wang
Beijing Forestry University

Tangren Cheng
Beijing Forestry University

Qixiang Zhang  zqxbjfu@126.com
Corresponding Author

DOI:
10.21203/rs.2.13507/v1

SUBJECT AREAS
Evolutionary Biology  Evolutionary Developmental Biology

KEYWORDS
PEBP genes family; Rosaceae species; Gene family evolution; Florigen; Molecular evolution
Abstract

Phosphatidylethanolamine-binding proteins (PEBPs) are a common gene family found among animals, plants and microbes. Plant PEBP proteins play an important role in regulating flowering time, as well as seed and bud dormancy. PEBP proteins can be divided into three major clades: FLOWERING LOCUS T-like (FT-like), TERMINAL FLOWER1-like (TFL1-like), and MOTHER OF FT AND TFL1-like (MFT-like). Though PEBP family genes have been well studied in Arabidopsis and other model species, their functional role in perennial trees is not fully understood. To characterize the evolution of PEBP genes and their role in flowering control among Rosaceae species, we identified a total of 46 PEBP members in seven Rosaceae species. Sequence and gene structure analysis revealed highly conserved intron/exon distributions and featured motifs among Rosaceae PEBP proteins. Analysis of synonymous/nonsynonymous substitution rates showed purifying selection constraining divergence within most lineages, while positive selection appears to have driven divergence of FT-like and TFL-like genes from the MFT clade. The expression of PEBP genes varied among different tissues indicating their functional divergence during gene family evolution. Furthermore, by employing a weighted gene co-expression network approach, we inferred a putative FT regulatory module essential for dormancy release and floral induction in P. mume. Our study sheds new light on the evolution of PEBP genes and their functional roles in controlling flowering time among Rosaceae tree species.

Introduction

Phosphatidylethanolamine-binding proteins (PEBPs) form a superfamily of genes containing the PEBP domain, which is highly conserved across taxa, from bacteria to insects to mammals and plants [1–3]. Mammalian PEBPs are globular proteins composed of a functional binding site for acetate, phosphate groups and phosphorylethanolamine [4,
Plant PEBP homologs share similar conserved motifs except their C-terminal part is deleted [6, 7]. The animal PEBP proteins were reported to function as serine protease or Raf kinase inhibitors controlling cell growth and differentiation [8–11]. In plants, PEBP genes are key regulators in determining flowering time, plant architecture and seed germination [12–15]. In angiosperms, members of PEBP family fall into three clades of genes: *FLOWERING LOCUS T (FT)*, *TERMINAL FLOWER 1 (TFL1)* and *MOTHER OF FT AND TFL1 (MFT)* [16, 17]. It was reported that *MFT*-like genes exist in both basal land plants and seed plants, while *FT*-like and *TFL*-like genes were only found in gymnosperm and angiosperms, indicating *MFT* clade might be the evolutionary ancestor to *FT*-like and *TFL*-like genes [17, 18]. Despite extensive sequence similarity among PEBP members, their functions diverged from each other [19].

*FT* and *TFL1* are two major PEBP proteins well studied in Arabidopsis and in many other plant species [20–23]. In *Arabidopsis*, *FT* acts as a floral signal transducer moving from leaves to shoot meristem to promote flowering; while *TFL* maintains inflorescence meristem in the shoot apex by suppressing *FT* expression [24–26]. *FT* and *TFL1* share ~60% amino acid sequence identity, but only a few amino acid change can convert *FT* from a floral promoter to a *TFL1*-like floral repressor [22, 27]. In addition to *FT* and *TFL1*, *Arabidopsis* PEBP gene family includes *MOTHER OF FT AND TFL1 (MFT)*, *TWIN SISTER OF FT (TSF)*, *BROTHER OF FT AND TFL1 (BFT)*, and *CENTRORADIALIS (CEN)* [13]. *MFT* integrates abscisic acid (ABA) and gibberellic acid (GA) signaling pathways and acts in a PIF1-dependent manner to repress seed germination under far-red light [14, 28]. *TSF* encodes the closest homolog of *FT* and resembles *FT* as a floral inducer [29]. *BFT* and *CEN* are two floral repressors in *Arabidopsis*, the overexpression of either one resulted in late flowering phenotype similar to plants over-expressing *TFL1* [30–32].

The functions of *FT*- and *TFL1*-like genes in regulating floral transition are highly
conserved across crop plants [33–36] and woody perennials [37–40]. In rice, \textit{DATE} 3a (HD3a) is orthologous to \textit{Arabidopsis FT} and it promotes heading under short day conditions [41]. Over-expression of apple (\textit{Malus \times domestica} Borkh.) \textit{FT}-like genes conferred precocious flowering in both \textit{Arabidopsis} and in apple [39]. In addition to flowering promoting role, \textit{FT} can also regulate growth cycling and bud dormancy in poplar [38, 42] and in Norway spruce (\textit{Picea abies}) [43]. Two \textit{FT} homologs have been characterized in poplar: \textit{PtFT1} regulates reproductive growth in response to low temperature while \textit{PtFT2} promote vegetative growth and suppresses fall bud set [38]. On the other hand, \textit{TFL1}-like genes play a key role in maintaining the inflorescence meristem in many species. For example, the presence of dominant allele of tomato \textit{SELF-PRUNING} (\textit{SP}) gene, orthologous to \textit{Arabidopsis TFL1}, generated tomato plants with continuous formation of inflorescences and fruit [44, 45]. Similarly, plants carrying mutation in \textit{TFL1} in rose and strawberry also exhibited recurrent flowering habit [23, 46].

The family of \textit{Rosaceae} consists of over 2500 species from about 90 genera, most of which are native to temperate zones around the world [47–49]. \textit{Prunus} is the largest \textit{Rosaceae} genus belonging to tribe \textit{Amygdaleae}, containing about 430 species, many of which are important fruit crops, such as plums, cherries, apricots and peaches [50]. Additionally, \textit{Prunus} includes a large number of spring-blooming trees with high ornamental and economic value. Though many \textit{Prunus} trees flower in spring, their flowering initiation time differs from one another. \textit{Prunus mume} is one of the earliest flowering species, which blooms in late winter or early spring, followed by apricots, peaches, cherries and plums that flush during March to April. Apple and pear trees from tribe \textit{Maleae} bloom much later around April to May in Northern China [51]. With divergent flowering time among \textit{Rosaceae} tree species, it is of great interest to dissect the evolution and their roles in contributing to flowering time variation in \textit{Rosaceae} tree species. Although PEBP gene
family has been recognized as key floral regulators in model plant species, their molecular evolution and function remains less clear in Rosaceae perennials. Here we provide a systematic study on the molecular evolution and function of PEBP gene family in Rosaceae tree species. We identified PEBP gene family across seven Rosaceae species and evaluated the conservation of their sequence, gene structure, and protein motifs. We then used synonymous/nonsynonymous substitution ratios to test for selection signatures on PEBP family genes. We further examined the expression pattern of PEBP genes across different organs and inferred a co-expression network of FT during dormancy break and floral induction in P. mume.

Results

Characterization of PEBP genes in Rosaceae species

By combing HMM and BLAST searches, we identified 45 PEBP-like proteins across seven Rosaceae species (Table 1). Each putative gene was validated by blasting against SMART, Pfam and NCBI CDD to ensure they contain PEBP domains. We then assigned all Rosaceae PEBPs to their closest Arabidopsis homologs (Figure 1). In total, we retrieved 45 Rosaceae PEBPs including 9 FT/TSF-like, 9 TFL-like, 9 CEN-like, 8 MFT-like and 10 BFT-like genes (Table 1). FT and TFL related genes showed the highest identity 74.38–82.18% with their Arabidopsis orthologs, while BFT-like proteins showed the lowest identity 66.29 to 67.24% comparing with AtBFT. Five to six PEBPs were detected among Prunus species, while the average number of PEBP genes almost doubled in M. domestica and Pyrus communis (Table 1). Duplicated paralogous pairs, such as MdTFL1 and MdTFL2, PcTFL1 and PcTFL2 are likely generated and retained after whole genome duplication in M. domestica and Pyrus communis [52, 53]. However, only one copy of MFT retained in M. domestica and Pyrus communis (Table 1).

Phylogenetic analyses
Phylogenetic trees were constructed based on protein sequence alignment of *Rosaceae* PEBPs using three approaches: neighbor-joining, maximum likelihood, and Bayesian inference methods (Figure 2; Figure S1). All three phylogenetic trees shared similar topology (Figure 2; Figure S1). The phylogenetic tree inferred showed that the total 51 PEBP proteins can be clustered into three major clades, *FT*-clade, *TFL*-clade, and *MFT*-clade (Figure 2). The *FT*-clade could be further split into *FT/TSF-like* genes and *BFT*-like genes, *TFL*-clade into *TFL1*-like and *CEN*-like genes (Figure 2). Within each subfamily, *Prunus* genes grouped closely together, while genes from *Maleae* species formed a separate group (Figure 2). Among *Prunus* PEBPs, proteins of *P. dulcis* and *P. persica* from *Amygdalus* subgenus first grouped together, then with that of *P. mume*, and with proteins of *P. yedoensis* and *P. avium* from *Cerasus* subgenus (Figure 2).

Structural analysis of PEBP family genes

*Roseaceae* PEBP family genes displayed conserved genomic structure and high amino acid sequence similarity with each other (Figure S2; Figure 3). The length of coding regions of PEBPs ranged from 519 to 579 bps, with *FT*-like genes between 525 to 529 bps, *MFT*-like genes between 519 to 579 bps, *BFT*-like genes between 522 to 525 bps, *TFL1*-like genes between 519 to 534 bps, and *CEN*-like genes between 522 to 528 bps. Gene structure analysis revealed a rather loose gene structure among all PEBPs consisting of four exons and three introns (Figure S2). For example, BFT-like genes harbor shortest introns of total length larger than 200 bp (Figure S2). Most *Rosaceae* PEBP genes were located on different chromosomes or scaffolds except *MdTFL2* and *MdFT* collocated on chromosome 12, *PpCEN* and *PpFT* co-located on chromosome 6 (Figure S3).

Sequence alignment of *Rosaceae* PEBP proteins revealed a high degree of conservation across the entire protein and within functional domains (Figure 1). Five conserved motifs were identified with the MEME program among *Rosaceae* PEBP proteins covering 164
amino acids (Figure 3). Among these, Motif 1, 2, and 5 together spread over the whole PEBP domain (Figure 3). Motifs DPDXP (Asp-Pro-Asp-X-Pro) and GIHR (Gly-Ile-His-Arg), which are essential for the anion-binding activity of PEBPs, were detected in the exon 4 among PEBPs (Figure 1). We also observed conserved residues that distinguish FT-like genes from TFL-like genes (Figure S4). Previously reported key residues conferring flowering promoting role of FT including V73, Y88, E112, L131, Y138, G142, W143, Q145, N156 were found among all Rosaceae FT-like proteins (Figure S4) [6, 27, 54]. While their corresponding residues (I/T)73, H88, E112, (K/N/T)131, (F/N)138, (P/S)142, S143, D145, and D156 were found among all TFL-like proteins (Figure S4). Residues determining 14-3-3 receptor binding interface (R66, F105, R134) were also detected in both types of proteins (Figure S4).

Molecular evolutionary analysis of PEBPs

To investigate the evolution of PEBP genes in Rosaceae species, we performed selection scans using branch model, site model, and branch site model in CODEML program of PAML (Table 2; Table S1; Table S2). Branch models specifying different $\omega$ parameters for foreground lineages (i.e. FT-like, TFL-like, CEN-like, and BFT-like lineages, and (FT, BFT, TFL, CEN)) were compared with fixed ratio model (Table S1). The likelihood ratio tests (LRT) on models specifying individual lineages of FT, TFL, CEN, BFT genes as foreground branch showed no significant $\omega$ difference between foreground and background branches (P>0.05) (Table S1). However, the LRT test on branch model 5 suggested that there is significant divergence between genes in the FT/TFL clade and those in MFT clade (P<0.001) (Table S1). We then applied the site model LRT and found no significant differences in $\omega$ values among sites of the PEBP coding sequence (Table S2). However, the branch-site LRT test revealed selection at specific sites in FT/TFL lineages (Table 2). The Bayes Empirical Bayes model suggested modest selection at at position 28, and position
10, 108, 128 when *FT/TFL*-clade and *TFL* lineage were set as foreground branch (Table 2).

Cis-acting element analysis of *FT* promoter

We extracted the 2000 bp region of *FT* genes and scanned for putative cis-element by searching against PlanPan and the PlantCARE database (Table 3). We compared the type and copy number of cis-element for 10 *FT* genes from *A. thaliana, P. trichocarpa, M. domestica, Pyrus communis, P. mume, P. persica, P. dulcis*, and *P. yedoensis* (Table 3). Within the promoter region of investigated *FTs*, five to seven CCACA boxes (binding site for CO) were identified across 8 species, while none were found within 2kb promoter region of *PtFT2* (Table 3). CArG boxes, binding site for MADS transcription factor, were found among all *FT* promoters, with *AtFT* promoter containing the most (Table 3). In addition, binding sites for MYB, MYC transcription factor, and ethylene-responsive transcription factor were present in all *FT* promoters (Table 3). Gibberellin-responsive elements of different types were present in all *FT* promoters (Table 3). Additionally, some cis-element showed species-specific distribution pattern. For example, low-temperature responsiveness element was only detected within promoter of *AtFT, PcFT1, and PmFT* (Table 3), while *W-box*, binding sites for WRKY transcription factor, were detected exclusively among *Prunus FT* promoters and *PtFT2* (Table 3).

Tissue-specific expression pattern of PEBPs

To explore the functional role of PEBP genes, we examined their expression pattern in different tissues of three *Prunus* species *P. persica, P. mume*, and *P. yedoensis* (Figure 4). In general, all tissues had expression of at least one of the PEBPs (Figure 4). *FT* were found expressed in most organs but were highly expressed in leaf and leaf bud in all three *Prunus* species (Figure 4). *PpTFL* and *PmTFL* were mostly expressed in vegetative tissues
such as leaf and stem, while the transcription of their closest homologs \textit{PpCEN} and \textit{PmCEN} were detected in leaf and root tissues respectively (Figure 4). \textit{MFT} were found expressed in embryo and fruit tissues (Figure 4). The tissue-expression pattern of \textit{BFT} diverges in two \textit{Prunus} species where \textit{PmBFT} were mostly expressed in fruit, \textit{PyBFT1} and \textit{PyBFT2} were detected in leaf but barely in fruit (Figure 4). The somewhat inconsistent expression pattern of ortholog pairs across three \textit{Prunus} species is likely a result of non-uniform sampling time and tissue specificity across three independent studies. The divergent expression pattern observed among PEBPs indicate functional divergence among PEBP lineages.

Co-expression network analysis of \textit{FT} during floral induction in \textit{P. mume}

To explore the functional role of \textit{FT} in floral induction in \textit{Prunus} species, we re-analyzed transcriptome changes of \textit{P. mume} during dormancy release [55] and performed a weighted co-expression network analysis (WGCNA). We identified 25 modules with distinct expression patterns (Figure S4a). Module-trait relationship analysis revealed six modules ‘darkorange’, ‘blue’, ‘turquoise’, ‘paleturquoise’, ‘indianred4’ and ‘salmon’ associated with the progression of bud flushing ($R^2 > 0.8$). Among them, module ‘darkorange’ was showing the strongest correlation with the FPKM of \textit{PmFT} (Figure S4b). The ‘darkorange’ module genes were significantly enriched in biological processes including flower development (GO:0050793), glucan metabolic process (GO:0009251), auxin transport (GO:0060918), and developmental growth involved in morphogenesis (GO:0060560). We further identified top 50 genes most associated with \textit{PmFT} from ‘darkorange’ module and 11 flowering related genes such as \textit{PmLFY}, \textit{PmAP1}, and \textit{PmCOL} that were previously identified [56, 57]. Among genes in module ‘darkorange’, \textit{SVP} (SHORT VEGETATIVE PHASE), \textit{SOC1} (SUPPRESSOR OF OVEREXPRESSION OF CO 1), \textit{GI} (GIGANTEA),, and \textit{CIB1} (CRYPTOCHROME-INTERACTING BASIC-HELIX-LOOP-HELIX 1) were previously identified as
key players in the FT-dependent flowering control in *Arabidopsis* [58, 59] (Figure 5a). Four tandem duplicated *PmDAMs* (*PmDAM1, PmDAM4, PmDAM5, PmDAM6*) from module ‘darkorange’ also exhibited expression pattern highly correlated with *PmFT* (Figure 5a). The expression pattern of other known floral regulators such as *COL* (*CONSTANS-LIKE*), *LHY1* (*LATE ELONGATED HYPOCOTYL 1*), and *AP1* (*APETALA1*) from module ‘paleturquoise’ and ‘turquoise’ were not highly correlated with that of *PmFT* (*R^2* < 0.62) (Figure 5a). The *FT* co-expressed genes were annotated to other biological processes including phosphomonoester hydrolysis, membrane trafficking, and ethylene signaling pathways (Table S3). *PmFT* showed weak transcription level in endodormant floral buds (Figure 5b). As floral bud continued accumulating chilling units and exiting dormancy, *PmFT* expression increased and showed high expression in flushing buds (Figure 5b). *PmCIB1* and 36 other genes were showing similar expression pattern with *PmFT*, while *PmGI*, *PmCOL*, *PmSVP*, *PmSOC1*, and four *PmDAMs* displayed contrasting expression pattern with their expression decreased as floral buds exiting endodormancy (Figure 5b).

**Discussion**

**Evolution of PEBP gene family in Rosaceae species**

PEBP’s form an ancient gene family with important functions in floral induction and plant architecture among species [17, 60]. Though previous studies have cloned and explored the function of PEBP family genes in model plants, none have focused on a comparative analysis of PEBP among Rosaceae perennials. A systematic search of Rosaceae genomes identified 45 PEBP family genes, orthologous to six *Arabidopsis* genes *FT/TSF, TFL1, CEN, BFT*, and *MFT*. The number of PEBP family genes we identified in *Prunus* species (chromosome 2n = 2x = 16) were approximately the same as in *Arabidopsis* (five to six copies). However, the PEBP family expanded in *M. domestica* and *Pyrus communis*
(chromosome 2n = 2x = 34), which is likely attributable to a recent whole-genome duplication (WGD) event that occurred in apple and pear after their split from Prunus species [52]. A few genes such as MdFT, MdMFT, and PcMFT retained only one copy in apple and pear, indicating the other copy may be lost during the species evolution after the WGD. There exist two copies of PyFT, PyMFT, and PyBFT in the genome of P. yedoensis (Figure S3g). Given that they scattered on different scaffolds or chromosomes, these duplicated gene pairs may arise from segmental duplication (Figure S3g) [61].

Previous studies have shown that PEBP gene family experienced two ancient duplications, giving rise to three types: FT-like genes promoting flowering, TFL1-like genes maintaining inflorescence meristem identity, and MFT-like genes controlling seed germination [3, 13, 17]. Since MFT genes are ancestral to FT and TFL1 genes, the duplication of MFT-clade gene and the latter diversification give rise to FT-like and TFL-like clades [16–18]. Our phylogenetic analysis suggests that Rosaceae PEBPs can be clustered into three distinct clades (FT, TFL, and MFT) consistent with other species [3, 13, 17]. The FT-like clade can be further divided into FT and BFT lineage, and TFL1-like clade can be divided into TFL and CEN lineages. With branch model maximum likelihood test specifying each of the five lineages (FT, TFL, CEN, MFT, and BFT) as foreground branch, we detected no evidence of positive selection acting on any of them. However, we observed significant selection acting on FT/TFL clade genes with the MFT clade specified as background branch, which supports the theory that divergence of the FT/TFL clade occurred after splitting from the MFT clade following duplication [18]. We detected no signs of position selection on individual sites, but strong purifying selection on most of them. However, this does not rule out the possibility that positive selection acting on a few codons masked by purifying selection preserving most of other sites [3]. With the branch-site likelihood ratio test, we detected a few codons under selection within TFL lineage and on the branch harboring
FT/TFL clades as a whole, but not within FT lineage. These results indicate adaptive evolution driving the FT/TFL clade diversification from the MFT clade, while purifying selection constrains evolution within most PEBP lineages.

Functional conservation of FT/TFL among Rosaceae tree species

Our structural analysis of PEBP proteins revealed a highly conserved gene structure and amino acid sequence within PEBP featured motifs. All PEBP family genes had a common gene structure with exactly four exons of similar sizes. Among conserved protein motifs, anion-binding D-P-D-x-P and G-x-H-R motifs are important for the conformation of ligand binding site in PEBP proteins [62]. Mutations close to this region may affect the binding of FT protein with phosphate ions, and alter its interaction with FD (FLOWERING LOCUS D) [63]. Segment B on exon 4 encodes an external loop, and together with its adjacent segment C determine the opposite function of FT and TFL1 in Arabidopsis [20]. Another key protein motif is the 14–3–3 binding domain essential for FT/TFL interaction with 14–3–3 receptors to promote flowering [6]. Key residues within these motifs were critical in determining FT/TFL functions. For example, substitution of amino acid (replace His–88 in TFL with Tyr) can convert TFL1 into a floral promoter [22]. In another study, specific mutations at four residues Glu–109, Trp–138, Gln–140, and Asn–152 convert FT into a TFL1-like repressor [27]. The amino acid at each of these critical positions were highly conserved and specific to FT and TFL, which suggest their functional conservation in floral promoting and repressing functions in Rosaceae species.

Recent molecular analyses have characterized the function of FT and TFL-like genes in several Rosaceae perennials [18]. The over-expression of MdFT in Arabopsis and apple both lead to precocious flowering [39]. Similarly, Arabidopsis transformed with PmFT, and European plum (P. domestica) transformed with poplar FT1 also demonstrate early flowering phenotype, indicating the floral promoting role of FT is conserved [37, 64]. In
addition to early flowering, plum trees transformed with PtFT1 displayed a shrub-like growth habit, reduced chilling requirement, and insensitive to short-day signals [64]. On the other hand, prolonged vegetative growth and late-flowering phenotype were observed for transgenic Arabidopsis over-expressing PpTFL1, PmTFL1, MdTFL1-1, or MdTFL1-2, suggesting that the Rosaceae TFL-like genes complement TFL1 function in Arabidopsis [65–67]. MdCENA, the closest homologs of MdTFLs, was shown expressed in proliferating tissues, while no transcription of MdCENb was detected in most organs in apple [65]. The multifaceted role of FT/TFL-like genes was also observed in other tree species [18]. In poplar, PtFT1 functions as a floral promoter activated by chilling temperature, while vegetative growth and dormancy break is promoted by PtFT2 [38]. In gymnosperms, FT-like genes exhibited a contrasting role in regulating growth cycling and bud set [68]. For example, FT/TFL-like genes in Norway spruce (PaFTL2) and Scots pine (PsFTL2) increases during bud set in autumn and decreased during bud burst in the next spring, [43, 69, 70]. Thus, FT/TFL like genes may undertake some novel functions concerning floral transition, vegetative growth cycling, and branching in Rosaceae tree species.

Co-expression network of FT in regulating flowering in P. mume

The flowering time regulation pathway has been studied in many species, but is best characterized in Arabidopsis [57, 71]. It is well understood that FT is a hub gene integrating four major pathways, including photoperiodic pathway, temperature pathway, autonomous pathway, and gibberellin pathway [56]. The expression of FT is facilitated by transcription factor CONSTANS (CO), which peaks at the end of the day under long photoperiod [72, 73]. Other genes involved in temperature pathway, SVP (SHORT VEGETATIVE PHASE), FLC (FLOWERING LOCUS C), and PIF4 (PHYTOCHROME INTERACTING FACTOR 4) can also regulate FT transcription through directly binding FT promoter or intronic regions [72, 74–76]. Upon induction, FT travels from the leaves to the shoot apical
meristem and binds to FD to regulate meristem identify genes such as AP1 (APETALA 1) to produce reproductive organs [77, 78]. Though flowering regulatory module concerning FT is well understood in annuals or biennial plants, it is still unclear in temperature tree species. Unlike annual or biennials, floral initiation in temperate trees starts in the preceding summer, floral buds enter dormancy during winter, and then bloom in the spring [79]. Thus, flowering in temperate woody perennials involves chilling-induced dormancy release, resumption of floral organ growth, and production of reproductive gametes, which is more complex than flowering in annuals or biennials [79, 80].

To gain insights into the regulatory network of FT during floral induction in perennials, we used WGCNA and identified a number of candidate genes whose expression pattern strongly correlated with FT in P. mume. Among these candidate, PmDAM1, PmDAM4, PmDAM5, and PmDAM6 were found down-regulated as the progression of dormancy break and blooming. DAM genes, resembling Arabidopsis FLC, act as major chilling-dependent regulators of bud dormancy and flowering time in apple [81, 82], pear [83, 84] and peach [85]. Another MADS-box gene PmSVP displaying similar expression pattern as PmDAMs was reported to maintain bud dormancy in apple [81]. Thus, PmDAMs and PmSVP may function as FT repressors in the same manner as in Arabidopsis by binding to the CArG box in the promoter region of PmFT [86]. A number of genes previously identified upstream FT, including PmCOL, PmGI, PmCIB1, were found induced by chilling in endodormant bud before the activation of PmFT. These genes may act directly or indirectly to activate FT expression during dormancy release in P. mume. We also observed that some known FT regulated genes, namely AP1, SOC1 (SUPPRESSION OF OVEREXPRESSION OF CONSTANS 1), , and LFY (LEAFY) peaked before the induction of FT, which can be explained by the fact that flower bud differentiation and development in Prunus perennials initiated in summer, continued into winter bud dormancy, and completed in the next spring [79, 87].
Additionally, we identified a number of co-expressed genes falling into functional categories that were not shown relatedness to dormancy or flowering in previously studies. Future functional studies is required to further dissect their roles in floral induction and characterize their interaction with FT in vivo.

Conclusion

In this study, we systemically characterized PEBP gene family in seven Rosaceae species and examined their gene structure, molecular evolution, protein features, and expressional profiling. The total 45 PEBP genes identified can be divided into three major clades, including FT-like, TFL-like, and MFT-like genes. We observed highly conserved protein motif and gene structure among PEBP genes. The selection scans revealed positive selection acting on FT/TFL clades while strong purifying selection restraining diversification within lineages. The tissue-specific expression pattern of PEBP genes is generally conservative across species. We identified a number of putative candidate genes co-expressed with FT, revealing a FT-related flowering regulatory model in Prunus species different from that in annual or biennial plants. In summary, the comprehensive analysis of PEBP family in our study presented evidence of structural and functional conservation of PEBP genes among Rosaceae woody perennials and provided an insight into the adaptive evolution of the PEBP gene family over the evolutionary history of flowering plants.

Materials And Methods

Identification of PEBP gene family

We obtained the most recent version of genomes for P. persica [88], P. mume [89], P. yedoensis [90], P. avium [91], P. dulcis, M. domestica [92] and Pyrus communis [93] from GDR (Genome Database For Rosaceae) [94]. To identify PEBP genes for each species, we
retrieved the HMM model PF01161 for PBP domain from the Pfam database
(https://pfam.xfam.org) and searched the genome protein databases with e-value cutoff
1.0×e⁻⁵ using HMMER 3.1 software [95]. In addition, we used protein sequences of AtFT
(At1g65480.1), AtTSF (At4g20370.1), AtTFL1 (At5g03840.1), AtBFT (At5g62040.1), AtMFT
(At1g18100.1) downloaded from TAIR (The Arabidopsis Information Resource)
(www.arabidopsis.org) as query to blast against local protein databases of seven species
and we only retained putative PEBP proteins with identity > 40% and e-value ≤ 1.0×e⁻⁵.
The genes identified with both methods were considered as candidate PEBP family genes
and were then verified with SMART [96], Pfam [97], CDD database [98] to ensure the
completeness of PBP domain.

Phylogenetic analysis

Multiple sequence alignment was performed using protein sequences of characterized
PEBP genes with software MUSCLE v3.8 [99] and visualized with GeneDoc v2.6 [100].
Phylogenetic trees were constructed using neighbor-joining (NJ) method with MEGA7 [101],
maximum likelihood (ML) analysis with RAxML v8.1 [102], and Bayesian inference (BI) with
MrBayes 3.1 [103]. BI method was performed with 100,000 generations of MCMC
processes. The Bayesian inference was performed with 100,000 generations of Markov
chain Monte Carlo (MCMC) simulations discarding the first 2500 trees as ‘burn-in’. With
consistent tree topologies inferred with these three approaches, neighbor-joining tree was
chosen to display the phylogenetic relationship of Rosaceae PEBP proteins.

Gene structure and protein motif detection of PEBP gene family

The exon and intron coordinate of PEBP genes were analyzed by comparing the coding
sequences with their genome sequences. MEME (Multiple Expectation Maximization for
Motif) online tool was used to predict protein motifs [104]. The protein motifs were further
annotated with the Pfam [97], SMART [96] and CDD [98] online tools. Chromosome
distribution of PEBP genes were obtained based on genome GFF3 files. Finally, gene
structure, protein motifs, and their chromosome location were visualized with software
TBtools [105].

Molecular evolution of PEBP genes
To investigate signatures of positive selection on Rosaceae PEBP genes, we first extracted
the coding sequence for PEBP genes and aligned them with MUSCLE v3.8 [99]. The
sequence alignment was then trimmed with Gblocks [106] in ‘condon’ mode and the
resulting alignments were used to infer phylogenetic relationships with RAxML [102]. The
ratio ($\omega$) of synonymous substitution sites ($dS$) and nonsynonymous substitution sites ($dN$)
were computed for each lineage and site of PEBP family genes using branch model, site
model and branch-site model with codeml package in PAML 4.0 [107]. To test the
hypothesis of adaptive evolution in specific PEBP lineages and across sites, we performed
likelihood ratio tests to evaluate the fit of branch models (FT, TFL, CEN, BFT set as fore-
ground branch), site models (M0, M1a, M2a, M7 and M8) and branch site models. The
positive selected sites were detected with Bayes Empirical Bayes analysis in PAML 4.0
[107].

Cis-elements analysis of FT promoter region
To investigate the conservation of cis-regulatory model of FT genes across species, we
extracted the 2kb upstream region of the start codon (ATG) and submitted the sequences
to PlantCARE [108] and PlantPan 2.0 database [109]. The cis-acting elements predicted
with these two methods were integrated and considered as putative cis-acting element.

Tissue-specific expression profile of PEBP genes
The RNA-seq data for different tissues of P. mume, P. yedoensis, and P. persica was
retrieved from three independent studies: GSE40162 from GEO database [89], SRP136962
and SRA053230 from NCBI SRA database [90, 110]. The raw SRA files were first dumped to FASTQ format using SRA-toolkit and preprocessed with Trimmomatic v0.38 [111] to trim off poor quality reads. Clean paired reads were aligned with reference genome of P. mume, P. yedoensis, and P. persica respectively with software HISAT2 [112]. The genic count was computed with HTSeq [113] and normalized to FPKM with R package ‘edgeR’ [114]. To investigate tissue specific expression of PEBP genes, we extracted the FPKM value for each PEBP gene across different tissues of P. mume, P. persica and P. yedoensi. The expression profile of PEBP genes was visualized using the ‘pheatmap’ package in R.

Co-expression network of FT during floral induction in P. mume

To investigate the functional role of FT during floral induction, we obtained the transcriptome data of four successive stages during floral bud flushing in P. mume from a previous study reported by Zhang et al. (2018). The procedure of sample collection, RNA extraction, sequencing library construction, quality control, and gene expression quantification was described in details [55]. We normalized the gene expression into FPKM (Fragments Per Kilobase Exon model per Million mapped fragments) and performed weighted gene co-expression network analysis with WGCNA v1.67 package in R [115]. The Dynamic Tree Cut algorithm was applied to identify modules (power $\beta$ of 4; height cutoff of 0.3; minimal module size: 30). To identify the key modules co-expressed with FT, we calculated the module-trait association and ranked genes by their correlation with the FPKM value of PmFT. Finally, top 50 candidate genes ($R^2 > 0.6$) co-expressed with PmFT and 11 putative FT interacting factors in Arabidopsis flowering pathway [56, 57] were selected to construct the co-expression network of FT. The FT regulatory network was visualized with Cytoscape 3.1 [116].

Declarations
Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and materials
The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that no competing interests exist.

Funding
This work was financially supported by the Fundamental Research Funds for the Central Universities (NO. BLX201828) and Special Fund for Beijing Common Construction Project.

Authors’ contributions
Zhang M designed the study; Li P and Yan XL performed the blast analysis and sequence curation; Wang J and Cheng T provided help with transcriptome analysis; Zhang Q supervised the project and revised the manuscript.

Acknowledgements
We thank Dr. Jason Holliday (Virginia Tech) for his help in revising the manuscript.

Reference
1. Chautard H, Jacquet M, Schoentgen F, Bureaud N, Bénédetti H: Tfs1p, a Member of the PEBP Family, Inhibits the Ira2p but Not the Ira1p Ras GTPase-Activating Protein in
Saccharomyces cerevisiae. Eukaryotic Cell 2004, 3(2):459-470.

2. Yeung K, Seitz T, Li S, Janosch P, McFerran B, Kaiser C, Fee F, Katsanakis KD, Rose DW, Mischak H et al: Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. Nature 1999, 401(6749):173-177.

3. Zheng X-M, Wu F-Q, Zhang X, Lin Q-B, Wang J, Guo X-P, Lei C-L, Cheng Z-J, Zou C, Wan J-M: Evolution of the PEBP gene family and selective signature on FT-like clade. Journal of Systematics and Evolution 2016, 54(5):502-510.

4. Serre L, Vallée B, Bureaud N, Schoentgen F, Zelwer C: Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. Structure 1998, 6(10):1255-1265.

5. Banfield MJ, Barker JJ, Perry ACF, Brady RL: Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. Structure 1998, 6(10):1245-1254.

6. Tsuji H, Taoka K-i: Chapter Five - Florigen Signaling. In: The Enzymes. Edited by Machida Y, Lin C, Tamanoi F, vol. 35: Academic Press; 2014: 113-144.

7. Vallee´e BS, Coadou G, Labbe´ H, Sy D, Vovelle F, Schoentgen F: Peptides corresponding to the N- and C-terminal parts of PEBP are well-structured in solution: new insights into their possible interaction with partners in vivo. The Journal of Peptide Research 2008, 61(2):47-57.

8. Hengst U, Albrecht H, Hess D, Monard D: The Phosphatidylethanolamine-binding Protein Is the Prototype of a Novel Family of Serine Protease Inhibitors. Journal of Biological Chemistry 2001, 276(1):535-540.

9. Keller ET, Fu Z, Brennan M: The role of Raf kinase inhibitor protein (RKIP) in health and disease. Biochemical Pharmacology 2004, 68(6):1049-1053.

10. Lorenz K, Lohse MJ, Quitterer U: Protein kinase C switches the Raf kinase inhibitor from
Raf-1 to GRK-2. *Nature* 2003, 426(6966):574–579.

11. Odabaei G, Chatterjee D, Jazirehi AR, Goodglick L, Yeung K, Bonavida B: Raf-1 Kinase Inhibitor Protein: Structure, Function, Regulation of Cell Signaling, and Pivotal Role in Apoptosis. 2004, 91:169–200.

12. Bradley D, Carpenter R, Copsey L, Vincent C, Rothstein S, Coen E: Control of inflorescence architecture in Antirrhinum. *Nature* 1996, 379(6568):791–797.

13. Karlgren A, Gyllenstrand N, Kallman T, Sundstrom JF, Moore D, Lascoux M, Lagercrantz U: Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiology* 2011, 156(4):1967–1977.

14. Vaistij FE, Barros-Galvão T, Cole AF, Gilday AD, He Z, Li Y, Harvey D, Larson TR, Graham IA: MOTHER-OF-FT-AND-TFL1 represses seed germination under far-red light by modulating phytohormone responses in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 2018, 115(33):8442–8447.

15. Mackenzie KK, Coelho LL, Lütken H, Müller R: Phylogenomic Analysis of the PEBP Gene Family from *Kalanchoë*. *Agronomy* 2019, 9(4):171.

16. Hedman H, Källman T, Lagercrantz U: Early evolution of the MFT-like gene family in plants. *Plant Molecular Biology* 2009, 70(4):359–369.

17. Liu Y-Y, Yang K-Z, Wei X-X, Wang X-Q: Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic FLOWERING LOCUS T gene homologs in gymnosperms and sheds new light on functional evolution. *New Phytologist* 2016, 212(3):730–744.

18. Wickland Daniel P, Hanzawa Y: The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant* 2015, 8(7):983–997.

19. Wang Z, Zhou Z, Liu Y, Liu T, Li Q, Ji Y, Li C, Fang C, Wang M, Wu M et al: Functional
Evolution of Phosphatidylethanolamine Binding Proteins in Soybean and Arabidopsis. The Plant Cell Online 2015, 27(2):323–336.

20. Ahn JH, Miller D, Winter VJ, Banfield MJ, Lee JH, Yoo SY, Henz SR, Brady RL, Weigel D: A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. The EMBO Journal 2006, 25(3):605–614.

21. Bai S, Tuan PA, Saito T, Ito A, Ubi BE, Ban Y, Moriguchi T: Repression of TERMINAL FLOWER1 primarily mediates floral induction in pear (Pyrus pyrifolia Nakai) concomitant with change in gene expression of plant hormone-related genes and transcription factors. Journal of Experimental Botany 2017, 68(17):4899–4914.

22. Hanzawa Y, Money T, Bradley D: A single amino acid converts a repressor to an activator of flowering. Proceedings of the National Academy of Sciences 2005, 102(21):7748–7753.

23. Koskela EA, Sønsteby A, Flachowsky H, Heide OM, Hanke M-V, Elomaa P, Hytönen T: TERMINAL FLOWER1 is a breeding target for a novel everbearing trait and tailored flowering responses in cultivated strawberry (Fragaria × ananassa Duch.). Plant Biotechnology Journal 2016, 14(9):1852–1861.

24. Baumann K, Venail J, Berbel A, Domenech MJ, Money T, Conti L, Hanzawa Y, Madueno F, Bradley D: Changing the spatial pattern of TFL1 expression reveals its key role in the shoot meristem in controlling Arabidopsis flowering architecture. Journal of Experimental Botany 2015, 66(15):4769–4780.

25. Chen Q, Payyavula RS, Chen L, Zhang J, Zhang C, Turgeon R: FLOWERING LOCUS T mRNA is synthesized in specialized companion cells in Arabidopsis and Maryland Mammoth tobacco leaf veins. Proceedings of the National Academy of Sciences 2018, 115(11):2830–2835.

26. Luccioni L, Krzymuski M, Sánchez-Lamas M, Karayekov E, Cerdán PD, Casal JJ:
CONSTANS delays Arabidopsis flowering under short days. *The Plant Journal* 2019, 97(5):923-932.

27.Ho WWH, Weigel D: *Structural Features Determining Flower-Promoting Activity of Arabidopsis FLOWERING LOCUS T*. *The Plant Cell* 2014, 26(2):552-564.

28.Footitt S, Ölçer-Footitt H, Hambidge AJ, Finch-Savage WE: A laboratory simulation of Arabidopsis seed dormancy cycling provides new insight into its regulation by clock genes and the dormancy-related genes DOG1, MFT, CIPK23 and PHYA. *Plant, Cell & Environment* 2017, 40(8):1474-1486.

29.Jin S, Jung HS, Chung KS, Lee JH, Ahn JH: *FLOWERING LOCUS T has higher protein mobility than TWIN SISTER OF FT*. *Journal of Experimental Botany* 2015, 66(20):6109-6117.

30.Huang N-C, Jane W-N, Chen J, Yu T-S: *Arabidopsis thaliana CENTRORADIALIS homologue (ATC) acts systemically to inhibit floral initiation in Arabidopsis*. *The Plant Journal* 2012, 72(2):175-184.

31.Liu X, Zhang J, Abuahmad A, Franks RG, Xie D-Y, Xiang Q-Y: *Analysis of two TFL1 homologs of dogwood species (Cornus L.) indicates functional conservation in control of transition to flowering*. *Planta* 2016, 243(5):1129-1141.

32.Ryu JY, Lee H-J, Seo PJ, Jung J-H, Ahn JH, Park C-M: *The Arabidopsis Floral Repressor BFT Delays Flowering by Competing with FT for FD Binding under High Salinity*. *Molecular Plant* 2014, 7(2):377-387.

33.Faure S, Higgins J, Turner A, Laurie DA: *The FLOWERING LOCUS T-Like Gene Family in Barley (Hordeum vulgare)*. *Genetics* 2007, 176(1):599-609.

34.Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M: *Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition to Flowering Downstream of Hd1 under Short-Day Conditions*. *Plant and Cell Physiology* 2002, 43(10):1096-1105.

35.Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K: *Hd3a and RFT1 are essential for...*
flowering in rice. *Development* 2008, 135(4):767–774.

36. Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J: The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences* 2006, 103(51):19581–19586.

37. Chen Y, Jiang P, Thammannagowda S, Liang H, Wilde HD: Characterization of Peach TFL1 and Comparison with FT/TFL1 Gene Families of the Rosaceae. *Journal of the American Society for Horticultural Science* 2013, 138(1):12–17.

38. Hsu CY, Adams JP, Kim H, No K, Ma C, Strauss SH, Drnevich J, Vandervelde L, Ellis JD, Rice BM et al: FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. *Proceedings of the National Academy of Sciences* 2011, 108(26):10756–10761.

39. Kotoda N, Hayashi H, Suzuki M, Igarashi M, Hatsuyama Y, Kidou S-i, Igasaki T, Nishiguchi M, Yano K, Shimizu T et al: Molecular Characterization of FLOWERING LOCUS T-Like Genes of Apple (*Malus × domestica* Borkh.). *Plant and Cell Physiology* 2010, 51(4):561–575.

40. Wu L, Liu D, Wu J, Zhang R, Qin Z, Liu D, Li A, Fu D, Zhai W, Mao L: Regulation of FLOWERING LOCUS T by a MicroRNA in Brachypodium distachyon. *The Plant Cell* 2013, 25(11):4363–4377.

41. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K: Hd3a Protein Is a Mobile Flowering Signal in Rice. *Science* 2007, 316(5827):1033–1036.

42. Bohlenius H: CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science* 2006, 312(5776):1040–1043.

43. Karlgren A, Gyllenstrand N, Clapham D, Lagercrantz U: FLOWERING LOCUS T/Terminal FLOWER1-Like Genes Affect Growth Rhythm and Bud Set in Norway Spruce. *Plant Physiology* 2013, 163(2):792–803.
44. Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jiménez-Gómez JM, Lippman ZB: Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nature Genetics* 2016, 49(1):162-168.

45. Springer NM, Jiang K, Liberatore KL, Park SJ, Alvarez JP, Lippman ZB: Tomato Yield Heterosis Is Triggered by a Dosage Sensitivity of the Florigen Pathway That Fine-Tunes Shoot Architecture. *PLoS Genetics* 2013, 9(12):e1004043.

46. Iwata H, Gaston A, Remay A, Thouroude T, Jeaffre J, Kawamura K, Oyant LH-S, Araki T, Denoyes B, Foucher F: The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. *The Plant Journal* 2012, 69(1):116-125.

47. Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arus P, Dandekar AM et al: Multiple Models for Rosaceae Genomics. *Plant Physiology* 2008, 147(3):985-1003.

48. Soundararajan P, Won SY, Kim JS: Insight on Rosaceae Family with Genome Sequencing and Functional Genomics Perspective. *BioMed Research International* 2019, 2019:1-12.

49. Yamamoto T, Terakami S: Genomics of pear and other Rosaceae fruit trees. *Breeding Science* 2016, 66(1):148-159.

50. Xiang Y, Huang C-H, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H: Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Molecular Biology and Evolution* 2016:msw242.

51. Gottschalk C, van Nocker S: Diversity in Seasonal Bloom Time and Floral Development among Apple Species and Hybrids. *Journal of the American Society for Horticultural Science* 2013, 138(5):367-374.

52. Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H et al: The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research* 2012, 23(2):396-408.
53. Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H et al: A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nature Communications 2019, 10(1).

54. Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O: An Antagonistic Pair of FT Homologs Mediates the Control of Flowering Time in Sugar Beet. Science 2010, 330(6009):1397-1400.

55. Zhang Z, Zhuo X, Zhao K, Zheng T, Han Y, Yuan C, Zhang Q: Transcriptome Profiles Reveal the Crucial Roles of Hormone and Sugar in the Bud Dormancy of Prunus mume. Scientific Reports 2018, 8(1).

56. Amasino R: Seasonal and developmental timing of flowering. The Plant Journal 2010, 61(6):1001-1013.

57. Song YH, Ito S, Imaizumi T: Flowering time regulation: photoperiod- and temperature-sensing in leaves. Trends in Plant Science 2013, 18(10):575-583.

58. Komeda Y: Genetic regulation of time to Flower in Arabidopsis thaliana. Annual Review of Plant Biology 2004, 55(1):521-535.

59. Teotia S, Tang G: To Bloom or Not to Bloom: Role of MicroRNAs in Plant Flowering. Molecular Plant 2015, 8(3):359-377.

60. Książkiewicz M, Rychel S, Nelson MN, Wyrwa K, Naganowska B, Wolko B: Expansion of the phosphatidylethanolamine binding protein family in legumes: a case study of Lupinus angustifolius L. FLOWERING LOCUS T homologs, LanFTc1 and LanFTc2. BMC Genomics 2016, 17(1).

61. Xu G, Ma H, Nei M, Kong H: Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences 2009, 106(3):835-840.

62. Banfield MJ, Brady RL: The structure of Antirrhinum centroradialis protein (CEN)
suggests a role as a kinase regulator. *Journal of Molecular Biology* 2000, 297(5):1159-1170.

63. Si Z, Liu H, Zhu J, Chen J, Wang Q, Fang L, Gao F, Tian Y, Chen Y, Chang L et al: *Mutation of SELF-PRUNING homologs in cotton promotes short-branching plant architecture. Journal of Experimental Botany* 2018, 69(10):2543-2553.

64. Park S, Srinivasan C, Dardick C, Callahan A, Scorza R: *Plum (Prunus domestica) Trees Transformed with Poplar FT1 Result in Altered Architecture, Dormancy Requirement, and Continuous Flowering. PLoS ONE* 2012, 7(7):e40715.

65. Mimida N, Kotoda N, Ueda T, Igarashi M, Hatsuyama Y, Iwanami H, Moriya S, Abe K: *Four TFL1/CEN-Like Genes on Distinct Linkage Groups Show Different Expression Patterns to Regulate Vegetative and Reproductive Development in Apple (Malus × domestica Borkh.). Plant and Cell Physiology* 2009, 50(2):394-412.

66. Esumi T, Kitamura Y, Hagihara C, Yamane H, Tao R: *Identification of a TFL1 ortholog in Japanese apricot (Prunus mume Sieb. et Zucc.). Scientia Horticulturae* 2010, 125(4):608-616.

67. Hanano S, Goto K: *Arabidopsis TERMINAL FLOWER1 Is involved in the regulation of flowering time and inflorescence development through transcriptional repression. The Plant Cell* 2011, 23(9):3172–3184.

68. Klintenäs M, Pin PA, Benloch R, Ingvarsson PK, Nilsson O: *Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. New Phytologist* 2012, 196(4):1260-1273.

69. Avia K, Kärkkäinen K, Lagercrantz U, Savolainen O: *Association of FLOWERING LOCUS T/TERMINAL FLOWER 1-like gene FTL2 expression with growth rhythm in Scots pine (Pinus sylvestris). New Phytologist* 2014, 204(1):159-170.
70. Gyllenstrand N, Clapham D, Kallman T, Lagercrantz U: A Norway Spruce FLOWERING LOCUS T Homolog Is Implicated in Control of Growth Rhythm in Conifers. Plant Physiology 2007, 144(1):248-257.

71. Song YH, Shim JS, Kinmonth-Schultz HA, Imaizumi T: Photoperiodic Flowering: Time Measurement Mechanisms in Leaves. Annual Review of Plant Biology 2015, 66(1):441-464.

72. Samach A, Gover A: Photoperiodism: The consistent use of CONSTANS. Current Biology 2001, 11(16):R651-R654.

73. Yang S, Weers BD, Morishige DT, Mullet JE: CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biology 2014, 14(1):148.

74. Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES: FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. Proceedings of the National Academy of Sciences 2011, 108(16):6680-6685.

75. Hartmann U, Hohmann S, Nettesheim K, Wisman E, Saedler H, Huijser P: Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. The Plant Journal 2000, 21(4):351-360.

76. Kumar SV, Lucynshyn D, Jaeger KE, Alós E, Alvey E, Harberd NP, Wigge PA: Transcription factor PIF4 controls the thermosensory activation of flowering. Nature 2012, 484(7393):242-245.

77. Lee J, Lee I: Regulation and function of SOC1, a flowering pathway integrator. Journal of Experimental Botany 2010, 61(9):2247-2254.

78. Teper-Bamnolker P: The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL Accumulation in Arabidopsis Leaves. The Plant Cell Online 2005, 17(10):2661-2675.

79. Wilkie JD, Sedgley M, Olesen T: Regulation of floral initiation in horticultural trees. Journal of Experimental Botany 2008, 59(12):3215-3228.

80. Lloret A, Badenes ML, Ríos G: Modulation of Dormancy and Growth Responses in
Reproductive Buds of Temperate Trees. *Frontiers in Plant Science* 2018, 9.

81. Wu R, Tomes S, Karunairetnam S, Tustin SD, Hellens RP, Allan AC, Macknight RC, Varkonyi-Gasic E: *SVP-like MADS Box Genes Control Dormancy and Budbreak in Apple*. *Frontiers in Plant Science* 2017, 08.

82. Mimida N, Saito T, Moriguchi T, Suzuki A, Komori S, Wada M: *Expression of DORMANCY-ASSOCIATED MADS-BOX (DAM)-like genes in apple*. *Biologia Plantarum* 2015, 59(2):237–244.

83. Niu Q, Li J, Cai D, Qian M, Jia H, Bai S, Hussain S, Liu G, Teng Y, Zheng X: *Dormancy-associated MADS-box genes and microRNAs jointly control dormancy transition in pear (Pyrus pyrifolia white pear group) flower bud*. *Journal of Experimental Botany* 2016, 67(1):239–257.

84. Tuan PA, Bai S, Saito T, Ito A, Moriguchi T: *Dormancy-Associated MADS-Box (DAM) and the Abscisic Acid Pathway Regulate Pear Endodormancy Through a Feedback Mechanism*. *Plant and Cell Physiology* 2017, 58(8):1378–1390.

85. Leida C, Conesa A, Llácer G, Badenes ML, Ríos G: *Histone modifications and expression of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner*. *New Phytologist* 2012, 193(1):67–80.

86. Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, Fornara F, Schneeberger K, Krajewski P, Coupland G: *Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis*. *Genome Biology* 2015, 16(1).

87. Kurokura T, Mimida N, Battey NH, Hytönen T: *The regulation of seasonal flowering in the Rosaceae*. *Journal of Experimental Botany* 2013, 64(14):4131–4141.

88. Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, Paris R, Aramini V, Gazza L, Rossini L et al: *The Peach v2.0 release: high-resolution linkage mapping and deep
resequencing improve chromosome-scale assembly and contiguity. BMC Genomics 2017, 18(1).

89. Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G et al: The genome of Prunus mume. Nature Communications 2012, 3(1).

90. Baek S, Choi K, Kim G-B, Yu H-J, Cho A, Jang H, Kim C, Kim H-J, Chang KS, Kim J-H et al: Draft genome sequence of wild Prunus yedoensis reveals massive inter-specific hybridization between sympatric flowering cherries. Genome Biology 2018, 19(1).

91. Shirasawa K, Isuzugawa K, Ikenaga M, Saito Y, Yamamoto T, Hirakawa H, Isobe S: The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding. DNA Research 2017, 24(5):499–508.

92. Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R et al: High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics 2017, 49(7):1099–1106.

93. Tinker NA, Chagné D, Crowhurst RN, Pindo M, Thrimawithana A, Deng C, Ireland H, Fiers M, Dzierzon H, Cestaro A et al: The Draft Genome Sequence of European Pear (Pyrus communis L. ‘Bartlett’). PLoS ONE 2014, 9(4):e92644.

94. Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbott A, Tomkins J, Main D: GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BMC Bioinformatics 2004, 5(1):130.

95. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD: HMMER web server: 2018 update. Nucleic Acids Research 2018, 46(W1):W200-W204.

96. Letunic I, Doerks T, Bork P: SMART: recent updates, new developments and status in 2015. Nucleic Acids Research 2015, 43(D1):D257-D260.

97. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A,
Hetherington K, Holm L, Mistry J et al: Pfam: the protein families database. Nucleic Acids Research 2014, 42(D1):D222-D230.

98. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Ger RC, Gonzales NR et al: CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research 2017, 45(D1):D200-D203.

99. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 2004, 32(5):1792-1797.

100. Nicholas KB, Nicholas HB: GeneDoc: Analysis and visualization of genetic variation. In: Embnetnews: 1997.

101. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 2016, 33(7):1870-1874.

102. Stamatakis A: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014, 30(9):1312-1313.

103. Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003, 19(12):1572-1574.

104. Bailey TL, Johnson J, Grant CE, Noble WS: The MEME Suite. Nucleic Acids Research 2015, 43(W1):W39-W49.

105. Chen C, Chen H, He Y, Xia R: TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. bioRxiv 2018:289660.

106. Castresana J: Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution 2000, 17(4):540-552.

107. Yang Z: PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 2007, 24(8):1586-1591.

108. Lescot M: PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 2002,
30(1):325-327.

109. Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, Chiang-Hsieh Y-F, Hou P-F, Yang T-Y, Chang W-C: *PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants*. Nucleic Acids Research 2016, 44(D1):D1154-D1160.

110. Lazzari B, Caprera A, Vecchietti A, Stella A, Milanesi L, Pozzi C: *ESTree db: a Tool for Peach Functional Genomics*. BMC Bioinformatics 2005, 6(Suppl 4):S16.

111. Bolger AM, Lohse M, Usadel B: *Trimmomatic: a flexible trimmer for Illumina sequence data*. Bioinformatics 2014, 30(15):2114-2120.

112. Kim D, Langmead B, Salzberg SL: *HISAT: a fast spliced aligner with low memory requirements*. Nature Methods 2015, 12(4):357-360.

113. Anders S, Pyl PT, Huber W: *HTSeq—a Python framework to work with high-throughput sequencing data*. Bioinformatics 2014, 31(2):166-169.

114. Robinson MD, McCarthy DJ, Smyth GK: *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics 2009, 26(1):139-140.

115. Langfelder P, Horvath S: *WGCNA: an R package for weighted correlation network analysis*. BMC Bioinformatics 2008, 9(1).

116. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang P-L, Lotia S, Pico AR, Bader GD, Ideker T: *A travel guide to Cytoscape plugins*. Nature Methods 2012, 9(11):1069-1076.

Tables

Table 1. Detailed information of PEBP family genes from *A. thaliana, M. domestica, Pyrus communis, P. persica, P. mume, P. yedoensis, P. avium* and *P. dulcis*. Gene notation was assigned to each PEBP based on their *Arabidopsis* ortholog.

| Gene | Species          | Gene          | NCBI accession       |
|------|------------------|---------------|----------------------|
| FT   | *A. thaliana*    | AT1G65480.1   | AT4G20370.1          |
|      | *M. domestica*   | MD12G1262000  | BAD08340.1/NP_00128079 |
|      | *Pyrus communis* | PCP004421.1   | PCP023373.1          |
|      | *P. persica*     | Prupe.6G364900.1 | XP_007206002.1      |
| Plant Species | Accession Numbers |
|---------------|------------------|
| P. mume       | Pm003733         |
| P. yedoensis  | Pyn_C1040.7     |
| P. avium      | Pav_co4051015.1_g010.1.mk |
| P. mume       | Pm001309         |
| P. yedoensis  | Pyn_C2138.1     |
| P. avium      | Pav_sc0000363.1_g840.1.mk |
| M. domestica  | MD14G1021100 MD12G1023900 |
| M. domestica  | MD11G1163500 MD03G1143000 |
| Pryus communins | PCP003730.1 PCP025869.1 |
| P. persica    | Prupe.7G112600.1 |
| P. mume       | Pm026188         |
| P. yedoensis  | Pyn_C2138.1     |
| P. avium      | Pav_sc0000363.1_g840.1.mk |
| P. dulcis     | Prudul26A021958 |
| CEN           |                 |
| A. thaliana   | AT2G27550.1      |
| M. domestica  | MD11G1163500 MD03G1143000 |
| Pryus communins | PCP019918.1 PCP022206.1 |
| P. persica    | Prupe.6G128400.1 |
| P. mume       | Pm001309         |
| P. yedoensis  | Pyn_C0319.14    |
| P. avium      | Pav_sc0000977.1_g050.1.mk |
| P. dulcis     | Prudul26A027558 |
| MFT           |                 |
| A. thaliana   | AT1G18100.1      |
| M. domestica  | MD06G1229900     |
| Pryus communins | PCP033759.1     |
| P. persica    | Prupe.5G230900.1 |
| P. mume       | Pm025099         |
| P. yedoensis  | Pyn_C1114.5     |
| P. avium      | Pav_sc0000103.1_g490.1.mk |
| P. dulcis     | Prudul26A015523 |
| BFT           |                 |
| A. thaliana   | AT5G62040.1      |
| M. domestica  | MD01G1198400 MD07G1265900 |
| Pryus communins | PCP007682.1 PCP03682.1 |
| P. persica    | Prupe.2G291900.1 |
| P. mume       | Pm019359         |

33
Table 2. Parameter estimates and likelihood values for branch-site models among sites and lineages of PEBP. Significant chi-square comparisons were indicated with * (pLRT<0.05), ** (pLRT<0.01), *** (pLRT<0.001). Positive selected sites in foreground lineages with its probability were detected with Bayes Empirical Bayes analysis.

| Branch-site model | Test | Fore-ground branch | Estimate of parameters |
|-------------------|------|--------------------|------------------------|
| Model A           | 1    | FT                 | proportion p=0.544, 0.445, 0.006, 0.005 |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   | 2    | TFL                | proportion p=0.770, 0.213, 0.013, 0.004 |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   | 3    | CEN                | proportion p=0.550, 0.449, 0.0006, 0.0005 |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   | 4    | BFT                | proportion p=0.772, 0.227, 0, 0 |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   | 5    | (FT, TFL, CEN, BFT)| proportion p=0.437, 0.359, 0.112, 0.092 |
|                   |      |                    |                         |
|                   |      |                    |                         |
|                   |      |                    |                         |

Table 3. Summary statistics of putative cis-element present in 2kb upstream promoter
region of *FT* across eight species. The organism column indicates in which organism the motif was discovered.

| Cis-element | Organism     | Motif           | Description                                      | A. thaliana | Populus tricocarpa | do |
|-------------|--------------|-----------------|-------------------------------------------------|-------------|-------------------|----|
| CCACA box   | A. thaliana  | CCACA           | Binding site for CO                              | 5           | 4                 | 0  |
| CArG box    | A. thaliana  | CC[A/T]_6GG     | Binding site for MADS-domain transcription factor | 10          | 1                 | 3  |
| MYB         | A. thaliana  | CAACAG          | Binding site for MYB transcription factor        | 5           | 3                 | 1  |
| bHLH        | A. thaliana  | ATGTG/AGGTG     | Binding site for MYC                             | 7           | 3                 | 3  |
| TCT-motif   | A. thaliana  | TCTTAC          | Part of a light responsive element               | 2           | 2                 | 0  |
| GATA-motif  | A. thaliana  | AAGATAAGATT     | Part of a light responsive element               | 0           | 1                 | 0  |
| AE-box      | A. thaliana  | AGAAAACAA       | Part of a module for light response              | 2           | 0                 | 1  |
| G-box       | A. thaliana  | TACGTG          | Cis-acting regulatory element involved in light responsiveness | 2           | 1                 | 1  |
| GT1-motif   | A. thaliana  | GGTTAA          | Light responsive element                         | 2           | 1                 | 2  |
| MSA-like    | (T/C)C(T/C)AACG G(T/C)(T/C)A | Cis-regulatory element involved in cell cycle regulation | 0           | 0                 | 0  |
| GARE-motif  | *Brassica oleracea* | TCTGTTG         | Gibberellin-responsive element                   | 0           | 0                 | 0  |
| P-box       | *Oryza sativa* | CCTTTTG         | Gibberellin-responsive element                   | 1           | 0                 | 1  |
| AP2;ERF     | A. thaliana  | CCGAC           | Ethylene-responsive transcription factor         | 7           | 3                 | 5  |
| ABRE        | A. thaliana  | ACGTG           | Cis-acting element involved in the abscisic acid responsiveness | 1           | 3                 | 2  |
| LTR         | *Hordeum vulgare* | CCGAAA          | Cis-acting element involved in low-temperature responsiveness | 1           | 0                 | 0  |
| W box       | A. thaliana  | CCGAAA          | Binding site for WRKY transcription factor       | 0           | 0                 | 2  |

**Figures**
Protein sequence alignment of PEBP family protein from Rosaceae species and A. thaliana. The sequences were aligned using Muscle. The conserved protein motif 14-3-3 interaction interface and anion-binding site were underlined in pink and purple respectively (Mackenzie KK et al., 2019). A, B, C, and D representing four segments in exon 4 (Ahn et al., 2006) were underlined in orange, blue, green and brown.
Phylogenetic tree of PEBPs from Rosacea and A. thaliana constructed by neighbor-joining method. All PEBP proteins can be clustered into three clades and five subfamilies.
MEME identified 5 major motifs among PEBP proteins. Conserved motifs were predicted with MEME and visualized with TBtools.
Figure 4

Tissue-specific expression of PEBP genes in three Prunus species. (a) P. persica, (b) P. mume, (c) P. yedoensis.
Co-expression network of FT during floral induction in P. mume. (a) Cytoscape visualization of candidate genes co-expressed with PmFT during dormancy release. Candidate genes from modules ‘darkorange’, ‘turquoise’, ‘paleturquoise’, and ‘darkgrey’ were colored in orange, blue, lightblue and grey respectively. The circle size represents the significance of gene expression correlation with PmFT. (b) Expression pattern of PmFT and its putative co-expressed genes during dormancy release process.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Figure S2.pdf
Figure S1.pdf
Figure S3.pdf
Table S2.pdf
Figure S5.pdf
Table S1.pdf
Figure S4.pdf
Table S3.pdf
