Journal of Experimental Botany, Vol. 75, No. 13 pp. 3946–3958, 2024
https://doi.org/10.1093/jxb/erae176 Advance Access Publication 20 April 2024

RESEARCH PAPER

Evolution of *FLOWERING LOCUS T-like* genes in angiosperms: a core *Lamiales*-specific diversification

Jiu-Xia Zhao[1,4,†], Shu Wang[2,4,5,†], Jing Wen[1,4,†], Shi-Zhao Zhou[1,4], Xiao-Dong Jiang[1], Mi-Cai Zhong[1], Jie Liu[1,4], Xue Dong[1,3], Yunfei Deng[2,5,*], Jin-Yong Hu[1,*], and De-Zhu Li[1,3,*]

[1] Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[2] State Key Laboratory of Plant Diversity and Specialty Crops, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China
[3] Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[4] University of the Chinese Academy of Sciences, Beijing 100049, China
[5] Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, Guangzhou 510650, China

[†] These authors contributed equally to this work.
* Correspondence: dzl@mail.kib.ac.cn, hujinyong@mail.kib.ac.cn, or yfdeng@scbg.ac.cn

Received 14 October 2023; Editorial decision 11 April 2024; Accepted 18 April 2024

Editor: Usha Vijayraghavan, Indian Institute of Science, India

Abstract

Plant life history is determined by two transitions, germination and flowering time, in which the phosphatidylethanolamine-binding proteins (PEBPs) FLOWERING LOCUS T (FT) and TERMINAL FLOWER1 (TFL1) play key regulatory roles. Compared with the highly conserved *TFL1-like* genes, *FT-like* genes vary significantly in copy numbers in gymnosperms, and monocots within the angiosperms, while sporadic duplications can be observed in eudicots. Here, via a systematic analysis of the PEBPs in angiosperms with a special focus on 12 representative species featuring high-quality genomes in the order *Lamiales*, we identified a successive lineage-specific but systematic expansion of *FT-like* genes in the families of core *Lamiales*. The first expansion event generated *FT1-like* genes mainly via a core *Lamiales*-specific whole-genome duplication (cL-WGD), while a likely random duplication produced the *FT2-like* genes in the lineages containing *Scrophulariaceae* and the rest of the core *Lamiales*. Both *FT1-* and *FT2-like* genes were further amplified tandemly in some families. These expanded *FT-like* genes featured highly diverged expression patterns and structural variation, indicating functional diversification. Intriguingly, some core *Lamiales* contained the relict *MOTHER OF FT AND TFL1 like* 2 (*MFT2*) that probably expanded in the common ancestor of angiosperms. Our data showcase the highly dynamic lineage-specific expansion of the *FT-like* genes, and thus provide important and fresh evolutionary insights into the gene regulatory network underpinning flowering time diversity in *Lamiales* and, more generally, in angiosperms.

Keywords: Diversity, evolution, *FLOWERING LOCUST T* (*FT*), core *Lamiales*, phosphatidylethanolamine-binding protein (PEBP) gene family, whole-genome duplication (WGD).

Introduction

Flowering time, the transition from vegetative to reproductive growth, is pivotal for reproductive success, and is tightly regulated by a complex interaction between endogenous developmental signals and exogenous environmental factors (Blazquez and Weigel, 2000; Michaels *et al.*, 2005). Flowering time in the model *Arabidopsis thaliana* is regulated by a complicated intrinsic gene-regulatory network (GRN) containing >300 floral regulators, among which both the florigen-encoding *FLOWERING LOCUS T* (*FT*) and its antagonistic anti-florigen-encoding *TERMINAL FLOWER 1* (*TFL1*) belong to the same phosphatidylethanolamine-binding protein (PEBP) gene family (Kardailsky *et al.*, 1999; Turck *et al.*, 2008; Fornara *et al.*, 2010; Wickland and Hanzawa, 2015; Bouche *et al.*, 2016). PEBPs are highly conserved in bacteria, animals, and plants, in which the *MOTHER OF FT AND TFL1* (*MFT*) *like* genes probably serve as the evolutionary ancestor of both *FT* and *TFL1* genes (Liu *et al.*, 2016; Bennett and Dixon, 2021; Jin *et al.*, 2021; Tsoy and Mushegian, 2022). In addition to flowering time regulation, *MFT-like* genes feature seed-specific expression and modulate seed oil/protein contents and germination via the abscisic acid (ABA) and gibberellic acid (GA) signaling pathways (Xi *et al.*, 2010; Nakamura *et al.*, 2011; Chen *et al.*, 2018; Cai *et al.*, 2023). In gymnosperms and some angiosperms, *MFT-like* genes seem to be duplicated independently into *MFT1* and *MFT2* subfamilies (Hedman *et al.*, 2009; Bennett and Dixon, 2021). However, only some angiosperm species maintain the *MFT2-like* genes.

The *TFL1*- and *FT-like* genes are exclusive to seed plants including gymnosperms and angiosperms (Karlgren *et al.*, 2011; Klintenas *et al.*, 2012; Liu *et al.*, 2016). In the *TFL1-like* lineage, *TFL1*, *BROTHER OF FT AND TFL1* (*BFT*), and the *CENTRORADIALIS* (*ATC*) homologs are floral repressors in Arabidopsis (Kobayashi *et al.*, 1999; Mimida *et al.*, 2001; Yoo *et al.*, 2010). The gymnosperm spruce contains one *TFL1-like* gene and two *FT-like* genes, which however do not promote flowering in Arabidopsis (Klintenas *et al.*, 2012; Liu *et al.*, 2016). Only in the angiosperm Arabidopsis and many other species do *FT* and the *FT* paralog, *TWIN SISTER OF FT* (*TSF*), promote flowering (Michaels *et al.*, 2005; Yamaguchi *et al.*, 2005), hence remaining a hot target gene in evolutionary analysis.

Current phylogenetic analyses identify a particular *FT-like* gene expansion into five core lineages with further duplication into 12 conserved clades in grasses or monocots (Bennett and Dixon, 2021). In dicots, duplication of *FT-like* genes has not been identified at family or order level, though sporadic reports on its amplifications and functional diversification can be found (Kikuchi *et al.*, 2009; Kong *et al.*, 2010; Huang *et al.*, 2012; Wickland and Hanzawa, 2015; Jiang *et al.*, 2022). Besides their essential roles in flowering time regulation, *FT-like* genes play essential roles in stomata movement, and tuber and bulb formation (Kinoshita *et al.*, 2011; Gonzalez-Schain *et al.*, 2012; Lee *et al.*, 2013; Jing *et al.*, 2023). Copy number variation in *FT-like* genes seems to correlate tightly with domestication in crops such as rice, maize, and soybean (Danilevskaya *et al.*, 2008; Itoh *et al.*, 2010; Wu *et al.*, 2017; Cai *et al.*, 2020).

The order *Lamiales* comprises ~26 families, ~24 000 species, and ~12% of eudicot plants, and has high value in terms of economy, horticulture, and medicine, as well as morphological innovations (Tallent-Halsell and Watt, 2009; Li *et al.*, 2021; F. Zhao *et al.*, 2023). The snapdragon (*Antirrhinum majus*, *Plantaginaceae*) serves as a model species for understanding molecular mechanisms underlying corolla zygomorphy and other traits (Zhong and Kellogg, 2015), while *Strobilanthes* plants (*Acanthaceae*) are known for their natural indigo and traditional medicinal properties, together with the remarkable diversity in flowering time behavior (Janzen, 1976; Splitstoser *et al.*, 2016; J.X. Zhao *et al.*, 2023). However, very limited information is available about the flowering genes in *Lamiales*. Here, based on both sequences and genome syntenies, we identified a core *Lamiales* (including *Plantaginaceae* and successive families)-specific expansion of *FT* genes into *FT1*- and *FT2-like* genes with 12 high quality genomes. The expansion of *FT1* genes is tightly associated with a core *Lamiales*-specific whole-genome duplication (cL-WGD), while the amplification of *FT2* genes seems to be random at first followed by independent tandem duplications. Both expansion events are followed by strong expression and gene structure diversification. These data are crucial for the understanding of the molecular and genetic mechanisms underlying the flowering time diversity in *Lamiales*, and more generally, in dicots of angiosperms.

Materials and methods

Comparative genomics and whole-genome duplication analyses

The most recent versions of genomes were used to construct orthologous gene families by OrthoFinder (v.2.0) (Emms and Kelly, 2019) (Supplementary Table S1). Mafft (v7.490) was utilized to construct multiple sequence alignments of 316 single-copy orthologs among 15 species (Katoh and Standley, 2013). RAxML software (v 8.2.12) was used to construct the maximum-likelihood tree with the PROTGAMMAAUTO model by employing sequence alignments with *A. thaliana*, *Rosa wichuraiana*, and *Solanum tuberosum* as outgroups (Stamatakis, 2014). The MCMCTree program of PAML (v 4.9h) was applied to estimate divergence time using protein alignments (Yang, 2007). Two calibration values were selected from the TimeTree website (http://www.timetree.org). For WGD analysis, the syntenic regions were found by MCscanX based on all-to-all BLASTP results (Wang *et al.*, 2012). WGDI and MultiAxon Paleopolyploidy Search (MAPS) were also used to infer the occurrence of WGD (Li *et al.*, 2018; Li and Barker, 2020; Sun *et al.*, 2022).

Identification and analyses of the *PEBP* genes

To identify the *PEBP* genes of each *Lamiales* species, we retrieved the HMM model (PF01161) of the PEBP domain from the Pfam database (https://pfam.xfam.org) and searched the genome protein databases with an e-value cut-off of $1.0\times e^{-5}$ using HMMER 3.1 software (Potter *et al.*, 2018). In addition, we used protein sequences AtFT (At1g65480.1),

AtTSF (At4g20370.1), AtTFL1 (At5g03840.1), AtBFT (At5g62040.1), and AtMFT (At1g18100.1) downloaded from TAIR (The Arabidopsis Information Resource, www.arabidopsis.org) as query sequences to blast against the local protein databases of 12 species (identities >30% and e-values ≤$1.0\times e^{-10}$) to identify putative PEBP sequences. The genes identified by both methods were considered as candidate PEBP family genes and were then verified with Pfam and the CDD database to ensure the completeness of the PEBP domain. Redundant sequences or sequences with an incomplete PEBP domain were excluded from the following analyses. For the other 98 angiosperms species, *PEBP-like* genes were identified by *BLAST* searches (e-values ≤1×10^{-3}) against complete genomes from Phytozome (https://phytozome.jgi.doe.gov). The Pfam database was used to confirm that the target gene had a PEBP domain (e-value <$1e^{-10}$).

Phylogenetic clustering analyses

PEBP-like sequences from 111 angiosperm species including 12 *Lamiales* species were aligned using the protein sequences with the software MAFFT (Katoh and Standley, 2013) with default parameters, and were visualized and edited with Jalview (https://www.jalview.org/). Alignend amino acid sequences were toggled to nucleotide sequences for phylogenetic analyses. Both IQ-TREE (v1.6.10) and RAxML were used to construct the maximum-likelihood tree with the best-fit model (Kalyaanamoorthy *et al.*, 2017). The best-fit model for the *PEBP-like* gene in *Lamiales* is GTR+F+R8, while the GTRGAMMAI model was used for the *PEBP-like* sequence from seed plants.

Gene structure, conserved protein domains, and motif analyses

The exon and intron locations of *PEBP* genes were determined by comparing the coding sequences with their genome sequences. To predict protein motifs, the MEME (Multiple Expectation Maximization for Motif) online tool (http://meme-suite.org/tools/meme) with optimum motifs set from 6 to 15 and a maximum number of 10 motifs was used. Conserved protein domains were analyzed using the NCBI CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) with the PSSM model (maximum number of hits <500) and the *Pfam* database with an e-value <$1e^{-10}$. The chromosome distributions of *PEBP* genes were obtained based on genome gene model annotation files. Finally, the gene structures, protein motifs, and chromosome locations were visualized using the software TBtools (Chen *et al.*, 2020).

Duplication and synteny analyses

To identify the synteny of *PEBP* family genes among species, synteny analysis was conducted by performing all-to-all BLASTP comparisons between the genomes of *Lamiales* species and the selected reference plant. Additionally, self-blast was performed by comparing protein-coding genes against their respective genomes using BLASTP. BLASTP hits with e-values <$1e^{-10}$ were utilized as input for MCScanX (Multiple Collinearity Scan toolkit) to identify potential collinear blocks within and between genomes of different species (Wang *et al.*, 2012). Based on the self-blast results, the duplicate_gene_classifier function was used to predict the *PEBP* duplication type and collinearity according to protocols described in the pipeline manuals. The phylogenetically clustered gene sets were defined as syntenic only when a minimum of five genes were collinear.

cis-Motif analyses

To investigate the conservation of the *cis*-regulatory model of *FT* genes across different clades, the 2 kb upstream region of the start codon (ATG) was extracted and sequences were submitted the to PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for prediction of *cis*-regulator elements.

Expression analyses

Reference genomes, gene model annotation files, and RNA-seq data for root, and two stages of stem and leaf were taken from a previous report (Xu *et al.*, 2020). The RNA reads were aligned to *S. cusia* genomes with *HISAT2* (Kim *et al.*, 2019). Then the TPM (transcripts per kilobase million) of each gene was calculated based on the length of the gene and read counts mapped to this gene.

Sample collection, RNA extraction, and RT–qPCR analysis

Sample of *S. cusia* were collected in June 2023 in Kunming Botanical Garden (KBG), Kunming Institute of Botany, Chinese Academy of Sciences (KIB, CAS). Young leaves, mature leaves, root, shoot apical meristems (SAMs), and stems with three biological replicates were collected and immediately frozen in liquid nitrogen for RNA extraction and reverse transcription followed by quantitative PCR (RT–qPCR) assays. The OminiPlant RNA Kit (DNase I) (CW2598S; Cwbio, Beijing, China) was used for extraction of total RNA. NovoScript® Plus All-in-one 1st Strand cDNA Synthesis SuperMix (gDNA Purge) (E047-01; Novoprotein, Shanghai, China) was used for reverse transcription. The quantitative PCRs were performed with NovoStart®SYBR qPCR SuperMix Plus (E096-01; Novoprotein, Shanghai, China) on a QuantStudio™ 7 Flex Real-Time PCR System (ThermoFisher). *ScuPP2A* (*EVM0021025*) was used to normalize *FT-like* gene expression in *S. cusia*, while *ScuACT2* (*EVM0028367*) was used to validate the reproducibility of *FT* expression. The relative expression levels were calculated following the procedures previously described (Hu *et al.*, 2014;Yu *et al.*, 2023). Primers of the five *FT-like* sequences were designed using TBtools (Chen *et al.*, 2020) and optimized with a trial test.

Results

PEBP gene evolution in angiosperms

In this study, we obtained 777 PEBP sequences from 117 species (covering six gymnosperms and all major eudicot groups including 18 monocot species) with genomes sequenced (Supplementary Table S2). A phylogenetic clustering revealed that, besides the known duplications in monocots, *FT* expansions were readily identified in *Fabales*, *Malvales*, *Lamiales*, and *Sapindales*, agreeing with previous reports (Fig. 1; Supplementary Fig. S1; Supplementary Table S2) (Nishikawa *et al.*, 2007; Klintenas *et al.*, 2012; Liu *et al.*, 2016; Wu *et al.*, 2017; Bennett and Dixon, 2021; Jiang *et al.*, 2022). For the *MFT* genes, consistent with previous analyses (Hedman *et al.*, 2009; Bennett and Dixon, 2021), two distinct clades (*MFT1-* and *MFT2-like*) were identified in the basal and some of the core eudicots.

Next, we focused on 12 species in the order of *Lamiales* (Fig. 2A; Supplementary Table S1). With Arabidopsis (*Brassicales*), rose (*Rosales*), and potato (*Solanales*) as outgroups, a phylogenetic clustering grouped the 104 *PEBP-like* genes into three major lineages, the *MFT-*, the *TCB-*, and the *FT-like* genes (Fig. 2A; Supplementary Fig. S2; Supplementary Table S3), agreeing well with the high conservation of these lineages in

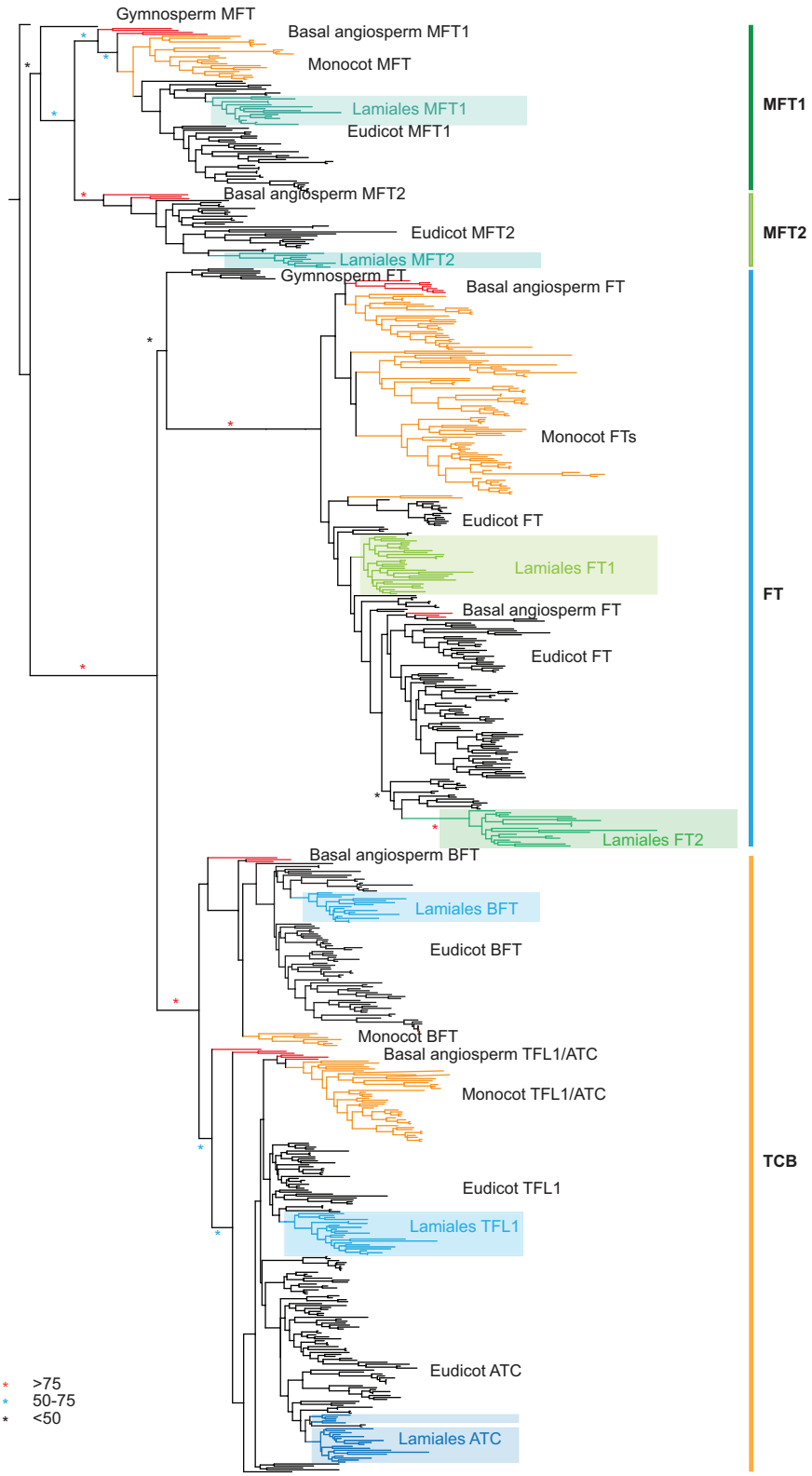

Fig. 1. Evolutionary pattern of PEBP proteins in angiosperms. A maximum-likelihood (ML) phylogeny based on 720 *PEBP* genes in 111 angiosperm plants with *MFT* genes of *Ginkgo bloba* as the outgroup. Bootstrap values are shown with red (≥75%), blue (between 50% and 75%), and black (<50%) asterisks for major lineages. Colored shading marks the *Lamiales* MFT1s (dark moderate cyan), MFT2s (dark cyan), FT1s (strong green), FT2s (dark lime green), BFT (bright blue), TFL1 (pure blue), and ATC (strong blue), and the lineages with colored lines show the genes from gymnosperms (pure red), basal angiosperms (dark moderate orange), and monocots (pure orange), respectively.

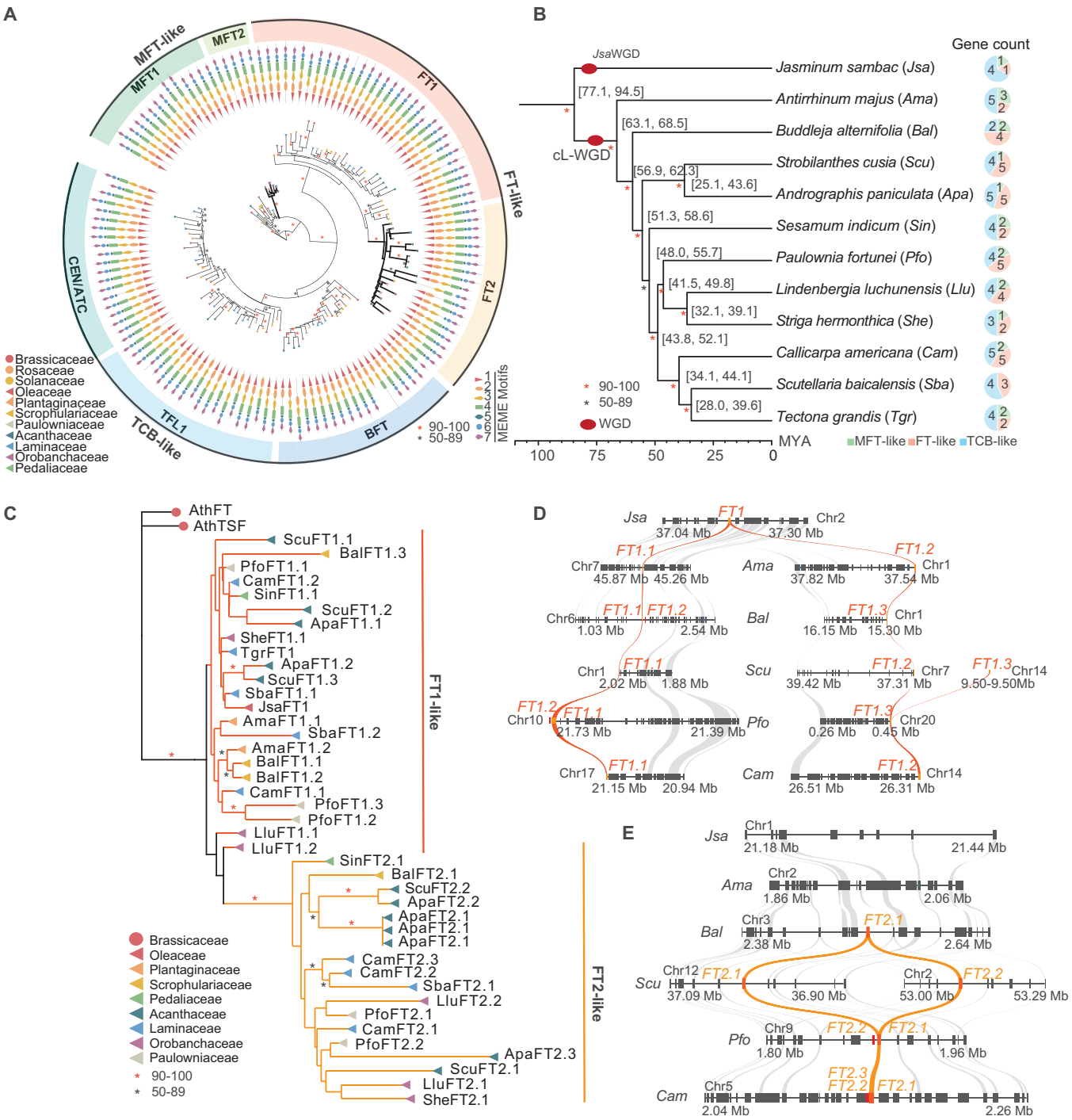

Fig. 2. Lineage-specific expansion of *FT-like* genes in core *Lamiales*. (A) An ML tree showing the clustering of PEBPs in *Lamiales* plants. Within major clades (MFT-, FT-, and TCB-like), the *Lamiales* feature two specific branches (MFT2 and FT2) but share five (MFT1, FT1, BFT, TFL1, and CEN/ATC) with Arabidopsis. A schematic representation drawn to scale of conserved protein motifs is shown next to each gene. (B) Evolutionary trajectory of 12 *Lamiales* species shown with an ML tree based on 316 single-copy orthologous genes using three outgroup species (not shown here). Red circles mark the *Jasminum*-specific (JsaWGD) and core *Lamiales*- (including species from *Antirrhinum majus* in *Plantaginaceae* and species of other families) specific cL-WGD events, respectively. Estimated divergence times at 95% probability are shown along each node. *PEBP* gene counts for each clade (*MFT* genes in light green, *FT* genes in light red, and *TCB* genes in light blue) are given in the right-hand panel. (C) Phylogenetic clustering of the FTs in *Lamiales*. (D) Collinearity relationships of the *FT1-like* genes in *Jsa*, *Ama*, *Bal*, *Scu*, *Pfo*, and *Cam*. (E) Microsynteny pattern for *FT2* genes in six *Lamiales* species. Note that, despite the good synteny of surrounding regions, no sign of *FT2-like* genes is detected in *Ama* and *Jsa*. In (A–C), asterisks in red (≥90%) and black (between 50% and 89%) mark the bootstrap support values. In (D) and (E), some species may randomly or tandemly duplicate their *FT* genes.

angiosperms (Fig. 1). We detected six *PEBP* genes in *Jasminum sambac* (*Jsa*) and the obligate parasite *Striga hermonthica* (*She*) (Qiu *et al.*, 2022), while we identified 12 genes in *Andrographis paniculate* (*Apa*), *Paulownia fortunei* (*Pfo*), and *Callicarpa americana* (*Cam*) (Fig. 2B). A further analysis for conserved MEME motifs identified seven motifs (1–7) ranging from 15 to 50 amino acids in all lineages, with the numbers of exon ranging from two to seven (CXN00020111 and CXN00025255 in *Apa*) and most of the genes featuring four (Fig. 2A; Supplementary Table S4). Synteny analyses showed good collinearity for each sub-branch gene among six species, while no synteny was detected between the expanded branch (*FT2*- and *MFT2-like*) and the conserved branch (*FT1*- and *MFT1-like*) (Figs 2C–E, 3). In general, *Lamiales* plants featured a highly variable number of *PEBP* genes with special expansions in *FT* (*FT1* and *FT2*) and *MFT* (*MFT1* and *MFT2*) genes.

FT1 gene expansion correlates with a core *Lamiales*-specific whole-genome duplication

Fourty-one *FT-like* genes present in the 12 *Lamiales* genomes could be easily clustered into two groups: *FT1*- and *FT2-like* genes (Fig. 2C). All 12 species featured at least one *FT1-like* in *J. sambac* (*JsaFT1*; *Oleaceae*) and a maximum of three in *B. alternifolia* (*BalFT1.1*, *BalFT1.2*, and *BalFT1.3*; *Plantaginaceae*) and *S. cucia* (*ScuFT1.1*, *ScuFT1.2*, and *ScuFT1.3*; *Acanthaceae*) (Fig. 2B, C; Supplementary Table S5). A gene collinearity analysis revealed a high level of synteny at the *FT1-like* loci between *Jsa* and the other five species (*Ama*, *Bal*, *Scu*, *Pfo*, and *Cam*), while no synteny could be identified between any *FT1*- and *FT2-like* genes (Supplementary Figs S3, S4).

WGDs are known to affect genome sizes, chromosomes numbers, and gene copy numbers in *Lamiales*, though whether *Lamiales* share a common WGD remains unresolved (Lyko and Wicke, 2021; Xu *et al.*, 2022). Intriguingly, our detailed paleopolypoid analyses using a stable phylogeny framework revealed that besides the known gamma (γ) whole-genome triplication (WGT) in the common ancestor of core eudicots, a specific WGD in the common ancestor resulting in core *Lamiales* (cL-WGD) but after speciation from *J. sambac* was observed (Fig. 2B; Supplementary Fig. S5). Furthermore, all synteny blocks surrounding *FT1-like* genes co-localized within the cL-WGD blocks; thus the cL-WGD might have underlain the expansion of the *FT1-like* genes (Supplementary Fig. S6; Supplementary Table S5). Interestingly, *FT1-like* genes were additionally and tandemly duplicated in *Bal* and *Pfo* (Fig. 2D; Supplementary Table S5).

FT2 gene expansion differs from that of *FT1* genes

The origin and duplication pattern for *FT2-like* genes differed from those of the *FT1-like* genes. Except for *Jsa* and *Ama*, which diverged early from the ancestors leading to *Lamiales* (Fig. 2), the other 10 species had at least one *FT2-like* gene (Fig. 2C). Besides the fact that *Apa* and *Scu* had five and two copies of *FT2-like* genes located on different chromosomes, respectively, all *FT2-like* genes featured high levels of synteny among species (Fig. 2E; Supplementary Fig. S3). However, no synteny could be observed between any *FT1*- and *FT2-like* genes (Supplementary Fig. S4). Furthermore, several *FT2-like* expansions via additional independent tandem duplications were observed in *Apa*, *Pfo*, *Llu*, and *Cam* (Supplementary Fig. S7; Supplementary Table S5). It seems that the *FT2-like* gene on chromosome 5 of *Cam* is tandemly duplicated three times (Supplementary Fig. S7). These data suggest that *FT2-like* genes might have been duplicated independently first in the common ancestor before the split of *Scrophulariaceae* and other core *Lamiales* plants, and followed by additional independent tandem duplications in some lineages. This differs significantly from both the *MFT2-like* genes, which were very probably relict due to their high sequence similarity to the basal angiosperms (Fig. 3) (Hedman *et al.*, 2009; Bennett and Dixon, 2021), and the *TCB* genes, which were probably duplicated via WGD-mediated events (Supplementary Figs S8, S9). Finally, a detailed synteny analysis did not detect any gene collinearity between the *FT2-like* genes in core *Lamiales* and the duplicated *FT* genes in soybean and monocots (Supplementary Fig. S10), again suggesting an independent expansion in these species.

Relict nature of *MFT2-like* genes in *Lamiales*

Consistent with previous reports (Hedman *et al.*, 2009; Bennett and Dixon, 2021), we identified two groups of *MFT-like* genes in *Lamiales* (*MFT1*- and *MFT2-like*; Fig. 3A). A sequence alignment revealed that MFT2 featured identical amino acid changes in the conserved PEBP domain, suggesting that the these *MFT2* genes might have undergone the same replication or experienced highly similar selection pressure (Fig. 3B). Interestingly, *Lamiales MFT2* genes were phylogenetically clustered with the *MFT2* genes from two basal angiosperms, *Nymphaea colorata* (*Nco*) and *Amborella trichopoda* (*Atr*) and those from eudicot plants. A synteny analysis detected no collinearity between the *MFT1*- and *MFT2-like* genes, while a weak synteny for both *MFT1-like* and *MFT2-like* genes could be identified between species within the *Lamiales* (Fig. 3C, D). Intriguingly, there was a strong gene collinearity between *Ama* and the two basal angiosperms. A further synteny analysis of the *AmaMFT2*-harboring segment in the *Sin* genome identified another DNA segment with high sequence collinearity on chromosome 12 but without *SinMFT2* (Supplementary Fig. S11), an indication that, via an unknown mechanism, *SinMFT2* might have translocated from chromosome 8 to 12. Additionally, six, among the 12 *Lamiales* species, independently had *MFT2*, suggesting that the maintenance of *MFT2* genes was very probably random in *Lamiales*. Hence, in contrast to the fact that every species has maintained the *MFT1-like* genes, some *Lamiales* have specifically lost the angiosperm-specific

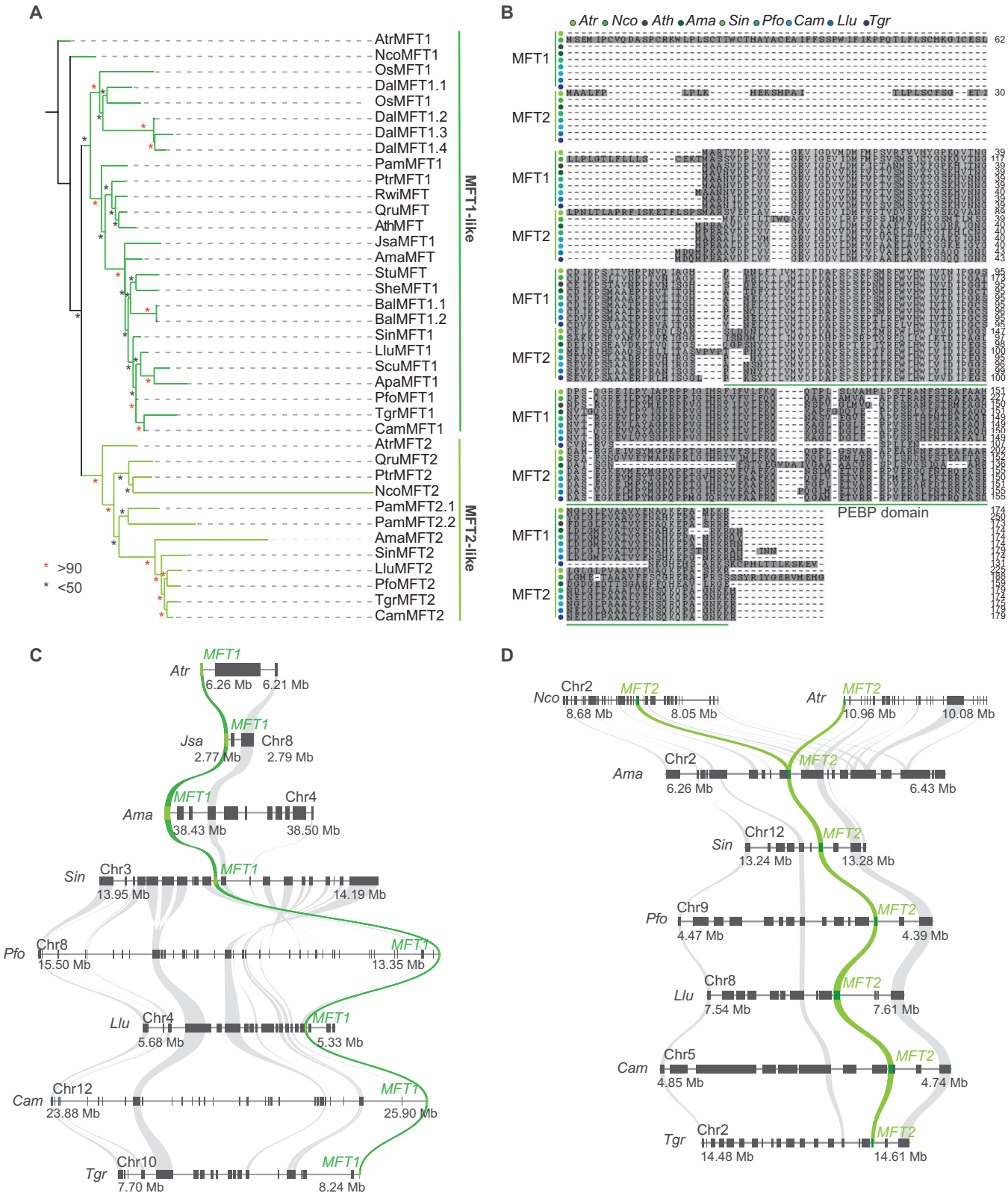

Fig. 3. Evolutionary pattern of *MFT-like* genes in core *Lamiales*. (A) A phylogeny indicated the *MFT2* duplication in basal angiosperms and some *Lamiales* plants. (B) Protein sequence alignments of MFT1- and MFT2-likes. (C) Microsynteny relationships for *MFT1* blocks between *Amborella trichopoda* (*Atr*) and the seven *Lamiales* species. (D) Collinearity relationships of the *MFT2-like* genes in two basal angiosperms and six *Lamiales* species.

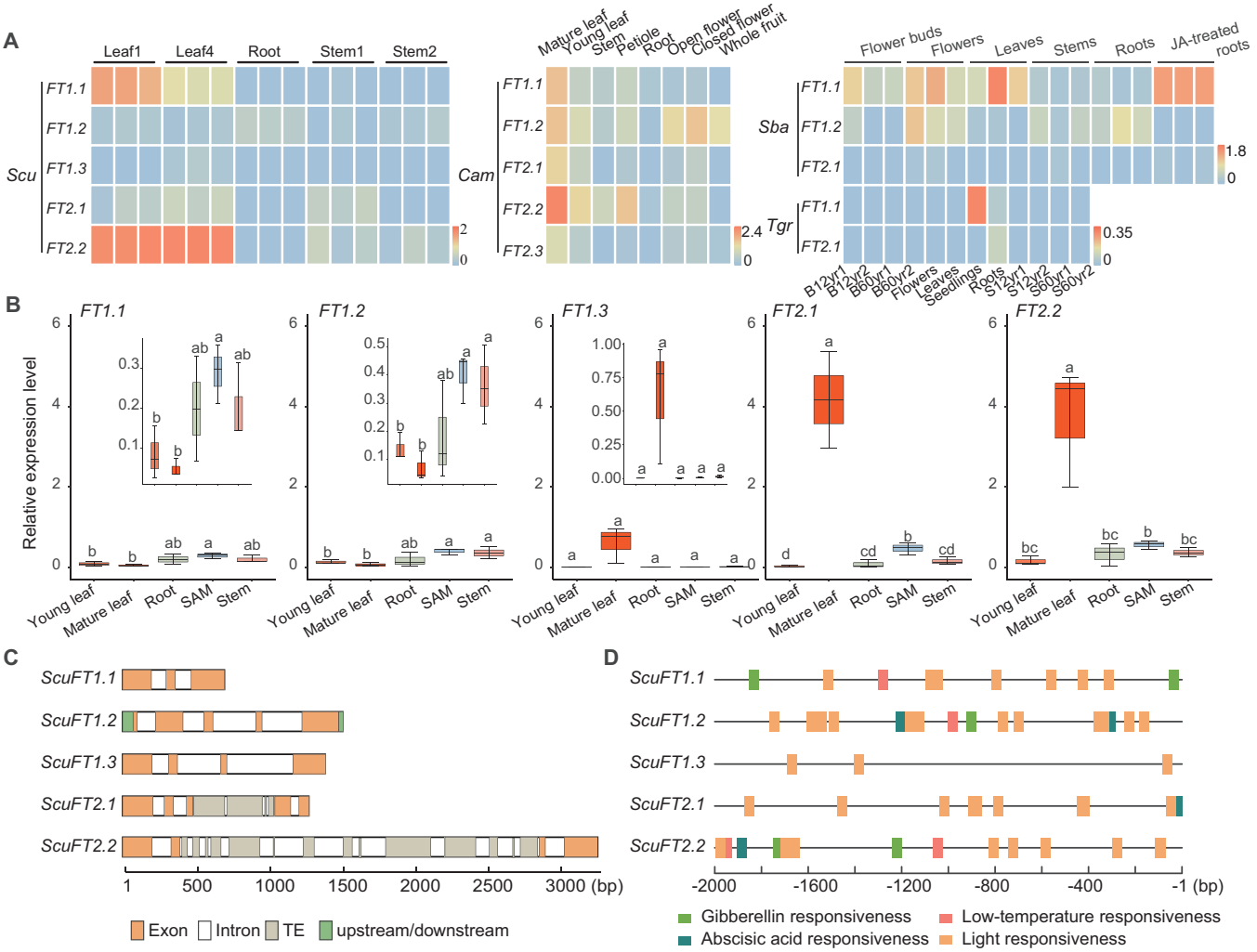

Fig. 4. High diversification in gene expression and structures for *FT* genes. (A) *FT1-* and *FT2-like* genes exhibit a highly diversified expression pattern. The normalized expression data for *Cam*, *Sba*, and *Tgr* are from previous reports (see the Materials and methods). (B) RT–qPCR assay confirms the highly diversified expression of the *ScFT* genes in five types of tissues. Data normalized to *ScuPP2A* are shown here, while the results referenced to *ScuACT2* are included in Supplementary Fig. S11. Error bars mark the SD of three biological replicates collected from at least three plants grown under natural conditions. Letters show the significant differences tested with ANOVA, $P<0.05$. (C) Gene structure and TE insertion patterns for *FT* genes in *S. cusia*. Boxes in orange, white, gray, and green show the exons, introns, TEs, and up-/downstream untranslated regions, respectively. (D) Distribution of conserved *cis*-motifs related to responses to gibberellins (green), abscisic acid (dark green), light (orange), and low temperature (pink) within the 2 kb promoter regions upstream of the translational start site in the five Sc*FT* genes.

MFT2-like genes, a pattern reported in a previous study (Bennett and Dixon, 2021).

High variation in expression and gene structures followed *FT* gene expansion

Gene duplication followed by variations in gene expression and even gene structure changes in promoters and exons–introns is not rare in either model or non-model species (Blackman, 2013; Mao *et al.*, 2016). To examine whether the *FT-like* gene expansion was followed by variation in gene expression, we explored *S. cusia* (Nees) Kuntze (*Scu*), an important herb and one of the main natural indigo sources (Ballard, 2007; Gu *et al.*, 2014). This species features perennial polycarpic flowering behavior, while its sister species *S. biocullata* Y.F. Deng & J.R.I. Wood shows perennial monocarpic mass flowering (Deng *et al.*, 2010; F. Zhao *et al.*, 2023). *S. cusia* had three *FT1-like* genes (*FT1.1*, *FT1.2*, and *FT1.3*) and two *FT2-like* genes (*FT2.1* and *FT2.2*) (Fig. 4A–D). As expected, these five *FT* genes indeed featured highly diversified expression in the five tissues examined (Xu *et al.*, 2020). *ScuFT2.2* showed the highest expression in both young and old leaves, while *ScuFT1.1* was expressed at a relatively high level in young leaves followed by in old leaves (Fig. 4A). The other *FT* genes were barely detected or had very low expression in all five tissues, indicating a strong expression variation following the expansion. To verify this expression

diversity, we collected five types of tissues and performed gene expression analyses with RT–qPCR with both *ScuPP2A* and *ScuACT2* as the references (Supplementary Table S6). Indeed, these *ScFT* genes featured highly variable and tissue-specific expression (Fig. 4B; Supplementary Fig. S12). An additional exploration of gene expression in other species also confirmed this pattern (Fig. 4A).

ScuFT genes featured significant sequence variation in both coding and non-coding regions (Fig. 4C, D). Substantial amino acid changes were present between the two groups of FTs especially within the conserved PEBP domain, but both featured the Arabidopsis Y85, a position distinguishing the FT/TFL1 functions (Supplementary Figs S13, S14) (Hanzawa *et al.*, 2005; Ahn *et al.*, 2006). However, *FT1* genes differed significantly in exon 4 from *FT2* genes in segment B that encodes an external P-loop essential for functional activity (Ahn *et al.*, 2006). Additionally, FT2s showed a moderate variation in the conserved domain, where the interaction with 14-3-3 bridge proteins occurred, and in the 'LYN triad' motif that distinguishes FT from the TFL1 in Arabidopsis (Taoka *et al.*, 2011; Li *et al.*, 2015).

Intriguingly, *ScuFT1.2* and *ScuFT2.1* had five exons, while the other three copies featured four exons, like the Arabidopsis *FT* (Fig. 4C). Furthermore, various types of transposable elements (TEs) were present in introns of both *ScuFT2* genes (Fig. 4C; Supplementary Table S7). However, the presence of enriched TEs seemed not to be specific for *FT2-like* genes, as several *FT1-like* genes (*SbaFT1*, *SinFT1.1*, and *PfoFT1*) also featured different types of TEs.

The expression variation was accompanied by changes in sequences and *cis*-motifs within the 2 kb promoter region of these *ScFT* genes (Fig. 4D). The *ScuFT1.3* promoter contained three light-responsive elements, while that of *ScuFT2.2* featured the most enriched elements responding to all four types of phytohormones together with two low-temperature-responsive and nine light-related elements. Like *ScuFT2.2*, *ScuFT1.2* had two elements related to ABA, indicating that both genes might respond to fluctuation in water availability, an environmental factor probably limiting the natural distribution of *Strobilanthes* as most species in this genera live neighbouring streams (Hu *et al.*, 2011). However, whether these elements play essential roles in expression variation awaits further functional testing.

Discussion

Due to their essential roles in regulation of flowering time and other developmental processes, *PEBP* gene expansions are not really rare. Gymnosperms have two *FT-like* and one *TFL1-like* group, but both *FT*- and *TFL1-like* genes in spruce act as flowering repressors in *A. thaliana* (Karlgren *et al.*, 2011; Klintenas *et al.*, 2012; Liu *et al.*, 2016). In agreement with a recent report (Bennett and Dixon, 2021), our detailed phylogenetic clustering revealed that *FT-like* genes are significantly expanded in major lineages of monocots, especially in grasses (Fig. 1). In *Lamiales* of the eudicots, two successive expansion events generating the *FT1*- and *FT2-like* clades seem to involve independent mechanisms (Figs 1, 2, 5).

The first expansion into *FT1-like* genes occurred in the common ancestor leading to core *Lamiales* and was very probably due to the cL-WGD as all these *FT1-like* genes shared rather good collinearity and the duplicated *FT1* genes co-localized within the genomic segments retained following the cL-WGD (Fig. 2). This type of expansion happens frequently, as seen in both *Fabales* and some species of *Brassicales* in angiosperms and in gymnosperms (Fig. 1) (Wang *et al.*, 2015; Liu *et al.*, 2016; Panchy *et al.*, 2016; Li *et al.*, 2023). Accompanying the *FT1-like* gene expansions, both *ATC*- and *BFT-like* genes were co-expanded in core *Lamiales* (Fig. 5).

The second *FT2-like* gene expansion differed from that of the *FT1-like* genes. This event was observed in lineages including *Scrophulariaceae* and the rest of the core *Lamiales*, but not in the *Plantaginaceae* (Figs 2, 5). *FT2-like* genes shared no synteny to either *FT1* genes of *Lamiales* or other angiosperm *FT* genes, an indication of random duplication, though the exact mechanisms needs further analysis. However, all *FT2-like* genes shared high levels of genome synteny, indicating that the generation of *FT2-like* genes might be via one event present in the common ancestor after the split of *Plantaginaceae* from core *Lamiales* and inherited since then. Of course, random duplication of *FT* genes is not really rare, as a similar *Rosoidae*-specific expansion has been observed in rose and its relatives (Jiang *et al.*, 2022).

Following the copy number expansion, both *FT1* and *FT2* genes featured significant sequence, gene structure, *cis*-motif, and gene expression variation, a pattern which has been frequently reported for both *PEBP* and other genes (Fig. 4) (Liu *et al.*, 2016; Jiang *et al.*, 2022; Niu *et al.*, 2022). Our analyses demonstrated a strong diversity in the five *FT* genes during different developmental stages and in different tissues of *S. cusia* and several other species examined, and this was accompanied occasionally by variation in both gene structure and *cis*-motifs, and hence very probably represents functional differentiation (Mao *et al.*, 2016). However, these raise the questions of which *FT* encodes the real florigen and what the other *FT* genes do. In contrast to the *ScuFT1* genes, the relatively high expression of both *ScuFT2* genes in mature leaves seems to be consistent with the pattern for Arabidopsis florigen-encoding *FT*, and thus might act as potential florigen(s) in *S. cusia*. It is noteworthy that, except position Y85 of AthFT, ScuFT2s differ from ScuFT1s in the key motifs of all important domains (Supplementary Figs S13, S14), hence whether they act as the real florigen needs further experimental assays.

In general, both cL-WGD and random events are likely to underlie the *PEBP* gene expansions in *Lamailes*, while tandem duplications additionally enrich this diversity. Though yet to be verified further, some evidence has shown that *FT* can be

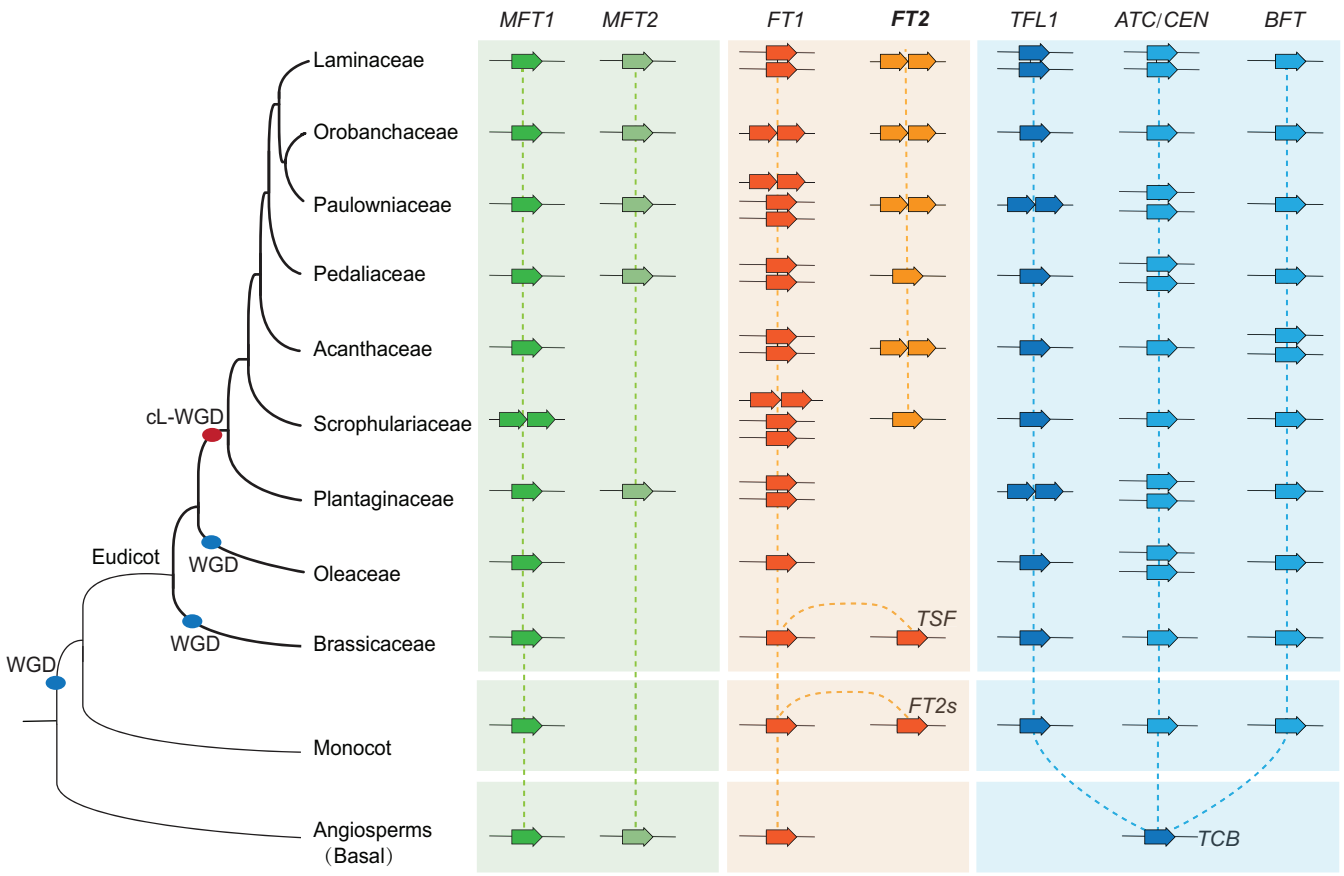

Fig. 5. A hypothetic model for the evolution of *PEBP* genes in angiosperms with a focus on *Lamiales*. Genes exhibiting conserved synteny are represented by colored arrows and connected by vertical dotted lines. Two horizontal arrows indicate genes expanded via tandem duplication, while two vertical arrows mark genes expanded due to WGDs and a single arrow shows no expansion. The red filled circle labels the cL-WGD event, while blue filled circles indicate known WGD events in *Brassicaceae* and all angiosperms.

induced by low temperatures, consequently promoting out-of-season flowering in *Olea europaea* (Haberman *et al.*, 2017). Therefore these variations including the *cis*-motifs in promoters of different *FT* genes clearly constitute a molecular basis for the high diversification of *Lamiales* plants and their adaptation to the ever-changing environment.

Supplementary data

The following supplementary data are available at *JXB* online.

Table S1. The genome data used in this study.

Table S2. *PEBP-like* sequences identified in this study.

Table S3. Identified conserved protein domains using the NCBI CD-Search Tool.

Table S4. Detailed information of *PEBP-like* genes in *Lamiales* genomes.

Table S5. *PEBP-like* genes and their duplication types predicted with MCScanX in 12 *Lamiales* genomes.

Table S6. Primers used in this study.

Table S7. Statistics of TE distribution within *PEBP* genes of six *Lamiales* species.

Fig. S1. A maximum-likelihood (ML) phylogeny based on 777 *PEBP* genes in 117 species with *MFT* genes from *Ginkgo bloba* as outgroups.

Fig. S2. Distribution of the syntenic *PEBP-like* genes in 12 *Lamiales* species.

Fig. S3. Macro- and microsynten between *Jsa* and other *Lamiales* plants.

Fig. S4. Synteny of *PEBP-like* genes in *Paulownia fortunei* (*Pfo*) and *Callicarpa americana* (*Cam*).

Fig. S5. Whole-genome duplication (WGD) analysis for *Lamiales*.

Fig. S6. WGD-derived and tandem duplications play key roles in *PEBP* gene expansion.

Fig. S7. Tandem duplication drives expansion of *FT2-like* genes in core *Lamiales*.

Fig. S8. Evolutionary pattern for the *TCB-like* clade.

Fig. S9. *BFT* genes expand in *Acanthaceae*.

Fig. S10. Synteny for *FT2* blocks between the *Lamiales* species and species from *Caryophyllales*, *Fabales*, and monocots.

Fig. S11. Synteny for *MFT2* blocks between *Antirrhinum majus* (*Ama*) and *Sesamum indicum* (*Sin*).

Fig. S12. RT–qPCR assay with *ScuAct2* as a reference confirms the highly diversified expression of the *ScFT* genes in five different tissue types.

Fig. S13. Protein sequence alignments with structural features annotated for the conserved position Arabidopsis Y85 in exon 2, 14-3-3 interaction surfaces (in two parts, a and b), the P-loop motif, and the 'LYN triad'.

Fig. S14. Alignments of the FT1s and FT2s identify conserved motifs.

Acknowledgements

We sincerely thank Professors Chunlei Xiang and Yongpeng Ma for interesting discussions, and Dr Fu Gao and Mr Hongyuan Yu for nursing some of the plants used in this study.

Author contributions

J-YH, D-ZL, and YD: conceptualization and coordination of the project; J-XZ and SW: data analysis and visualization; JW: performing the RT–qPCR analysis with the help of J-XZ; J-YH and J-XZ: writing the paper with help from all other authors. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no conflicts of interest.

Funding

Our research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS; XDB31000000) to J-YH and D-ZL, the National Nature Science Foundation of China (32170213 to YD and 32370578 to XD), and the Guangdong Flagship Project of Basic and Applied Basic Research (2023B0303050001) to YD. This work was partially facilitated by the Germplasm Bank of Wild Species and the Kunming Botanical Garden, Kunming Institute of Botany, CAS.

Data availability

Relevant data can be found within the paper and are available at BlueOmics: http://blueomics.iflora.cn/.

References

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

Ballard MW. 2007. Natural dyes: sources, tradition, technology and science. Studies in Conservation **52**, 318–319.

Bennett T, Dixon LE. 2021. Asymmetric expansions of FT and TFL1 lineages characterize differential evolution of the EuPEBP family in the major angiosperm lineages. BMC Biology **19**, 181.

Blackman BK. 2013. Interacting duplications, fluctuating selection, and convergence: the complex dynamics of flowering time evolution during sunflower domestication. Journal of Experimental Botany **64**, 421–431.

Blazquez MA, Weigel D. 2000. Integration of floral inductive signals in Arabidopsis. Nature **404**, 889–892.

Bouche F, Lobet G, Tocquin P, Perilleux C. 2016. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. Nucleic Acids Research **44**, D1167–D1171.

Cai Y, Wang L, Chen L, *et al*. 2020. Mutagenesis of GmFT2a and GmFT5a mediated by CRISPR/Cas9 contributes for expanding the regional adaptability of soybean. Plant Biotechnology Journal **18**, 298–309.

Cai Z, Xian P, Cheng Y, Zhong Y, Yang Y, Zhou Q, Lian T, Ma Q, Nian H, Ge L. 2023. MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. New Phytologist **239**, 905–919.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant **13**, 1194–1202.

Chen Y, Xu X, Chen X, Chen Y, Zhang Z, Xuhan X, Lin Y, Lai Z. 2018. Seed-specific gene MOTHER of FT and TFL1 (MFT) involved in embryogenesis, hormones and stress responses in *Dimocarpus longan* Lour. International Journal of Molecular Sciences **19**, 2403.

Danilevskaya ON, Meng X, Hou Z, Ananiev EV, Simmons CR. 2008. A genomic and expression compendium of the expanded PEBP gene family from maize. Plant Physiology **146**, 250–264.

Deng YF, Wood JRI, Ying F. 2010. *Strobilanthes biocullata* (Acanthaceae), a new species from Hunan, China. Novon **20**, 406–411.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology **20**, 238.

Fornara F, de Montaigu A, Coupland G. 2010. SnapShot: control of flowering in Arabidopsis. Cell **141**, 550–550.e2.

Gonzalez-Schain ND, Diaz-Mendoza M, Zurczak M, Suarez-Lopez P. 2012. Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner. The Plant Journal **70**, 678–690.

Gu W, Zhang Y, Hao XJ, Yang FM, Sun QY, Morris-Natschke SL, Lee KH, Wang YH, Long CL. 2014. Indole alkaloid glycosides from the aerial parts of *Strobilanthes cusia*. Journal of Natural Products **77**, 2590–2594.

Haberman A, Bakhshian O, Cerezo-Medina S, Paltiel J, Adler C, Ben-Ari G, Mercado JA, Pliego-Alfaro F, Lavee S, Samach A. 2017. A possible role for flowering locus T-encoding genes in interpreting environmental and internal cues affecting olive (*Olea europaea* L.) flower induction. Plant, Cell & Environment **40**, 1263–1280.

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

Hedman H, Kallman T, Lagercrantz U. 2009. Early evolution of the MFT-like gene family in plants. Plant Molecular Biology **70**, 359–369.

Hu J-q, Deng Y, Wood JRI. 2011. Strobilanthes Blume. In: Wu ZY, Raven PH, Hong DY, eds. Flora of China. **Vol. 19**. Beijing and St. Luis: Science Press and Missouri Botanical Garden Press, 381–429.

Hu JY, Zhou Y, He F, Dong X, Liu LY, Coupland G, Turck F, de Meaux J. 2014. miR824-regulated AGAMOUS-LIKE16 contributes to flowering time repression in Arabidopsis. The Plant Cell **26**, 2024–2037.

Huang W, Fang Z, Zeng S, Zhang J, Wu K, Chen Z, Teixeira da Silva JA, Duan J. 2012. Molecular cloning and functional analysis of three FLOWERING LOCUS T (FT) homologous genes from Chinese Cymbidium. International Journal of Molecular Sciences **13**, 11385–11398.

Itoh H, Nonoue Y, Yano M, Izawa T. 2010. A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nature Genetics **42**, 635–638.

Janzen DH. 1976. Why Bamboos wait so long to flower. Annual Review of Ecology and Systematics **7**, 347–391.

Jiang XD, Zhong MC, Dong X, Li SB, Hu JY. 2022. Rosoideae-specific duplication and functional diversification of FT-like genes in Rosaceae. Horticulture Research **9**, uhac059.

Jin S, Nasim Z, Susila H, Ahn JH. 2021. Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. Seminars in Cell & Developmental Biology **109**, 20–30.

Jing S, Jiang P, Sun X, Yu L, Wang E, Qin J, Zhang F, Prat S, Song B. 2023. Long-distance control of potato storage organ formation by SELF PRUNING 3D and FLOWERING LOCUS T-like 1. Plant Communications **4**, 100547.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods **14**, 587–589.

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D. 1999. Activation tagging of the floral inducer FT. Science **286**, 1962–1965.

Karlgren A, Gyllenstrand N, Kallman T, Sundstrom JF, Moore D, Lascoux M, Lagercrantz U. 2011. Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiology **156**, 1967–1977.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution **30**, 772–780.

Kikuchi R, Kawahigashi H, Ando T, Tonooka T, Handa H. 2009. Molecular and functional characterization of PEBP genes in barley reveal the diversification of their roles in flowering. Plant Physiology **149**, 1341–1353.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology **37**, 907–915.

Kinoshita T, Ono N, Hayashi Y, *et al*. 2011. FLOWERING LOCUS T regulates stomatal opening. Current Biology **21**, 1232–1238.

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

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. Science **286**, 1960–1962.

Kong FJ, Liu BH, Xia ZJ, *et al*. 2010. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant Physiology **154**, 1220–1231.

Lee R, Baldwin S, Kenel F, McCallum J, Macknight R. 2013. FLOWERING LOCUS T genes control onion bulb formation and flowering. Nature Communications **4**, 2884.

Li C, Lin H, Dubcovsky J. 2015. Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. The Plant Journal **84**, 70–82.

Li J, Wang Y, Dong Y, Zhang W, Wang D, Bai H, Li K, Li H, Shi L. 2021. The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. Horticulture Research **8**, 53.

Li Y, Xiao L, Zhao Z, Zhao H, Du D. 2023. Identification, evolution and expression analyses of the whole genome-wide PEBP gene family in *Brassica napus* L. BMC Genomic Data **24**, 27.

Li Z, Barker MS. 2020. Inferring putative ancient whole-genome duplications in the 1000 Plants (1KP) initiative: access to gene family phylogenies and age distributions. GigaScience **9**, giaa004.

Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, Rundell RJ, Barker MS. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. Proceedings of the National Academy of Sciences, USA **115**, 4713–4718.

Liu YY, Yang KZ, Wei XX, Wang XQ. 2016. 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 **212**, 730–744.

Lyko P, Wicke S. 2021. Genomic reconfiguration in parasitic plants involves considerable gene losses alongside global genome size inflation and gene births. Plant Physiology **186**, 1412–1423.

Mao Y, Sun J, Cao P, Zhang R, Fu Q, Chen S, Chen F, Jiang J. 2016. Functional analysis of alternative splicing of the FLOWERING LOCUS T orthologous gene in *Chrysanthemum morifolium*. Horticulture Research **3**, 16058.

Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM. 2005. Integration of flowering signals in winter-annual Arabidopsis. Plant Physiology **137**, 149–156.

Mimida N, Goto K, Kobayashi Y, Araki T, Ahn JH, Weigel D, Murata M, Motoyoshi F, Sakamoto W. 2001. Functional divergence of the TFL1-like gene family in Arabidopsis revealed by characterization of a novel homologue. Genes To Cells **6**, 327–336.

Nakamura S, Abe F, Kawahigashi H, *et al*. 2011. A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. The Plant Cell **23**, 3215–3229.

Nishikawa F, Endo T, Shimada T, Fujii H, Shimizu T, Omura M, Ikoma Y. 2007. Increased CiFT abundance in the stem correlates with floral induction by low temperature in Satsuma mandarin (*Citrus unshiu* Marc.). Journal of Experimental Botany **58**, 3915–3927.

Niu T, Wang X, Abbas M, Shen J, Liu R, Wang Z, Liu A. 2022. Expansion of CONSTANS-like genes in sunflower confers putative neofunctionalization in the adaptation to abiotic stresses. Industrial Crops and Products **176**, 114400.

Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. Plant Physiology **171**, 2294–2316.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Research **46**, W200–W204.

Qiu S, Bradley JM, Zhang P, Chaudhuri R, Blaxter M, Butlin RK, Scholes JD. 2022. Genome-enabled discovery of candidate virulence loci in *Striga hermonthica*, a devastating parasite of African cereal crops. New Phytologist **236**, 622–638.

Splitstoser JC, Dillehay TD, Wouters J, Claro A. 2016. Early pre-Hispanic use of indigo blue in Peru. Science Advances **2**, e1501623.

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

Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. 2022. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Molecular Plant **15**, 1841–1851.

Tallent-Halsell NG, Watt MS. 2009. The invasive *Buddleja davidii* (butterfly bush). The Botanical Review **75**, 292–325.

Taoka K, Ohki I, Tsuji H, *et al*. 2011. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature **476**, 332–335.

Tsoy O, Mushegian A. 2022. Florigen and its homologs of FT/CETS/PEBP/RKIP/YbhB family may be the enzymes of small molecule metabolism: review of the evidence. BMC Plant Biology **22**, 56.

Turck F, Fornara F, Coupland G. 2008. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology **59**, 573–594.

Wang Y, Tang H, Debarry JD, *et al*. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research **40**, e49.

Wang Z, Zhou Z, Liu Y, *et al*. 2015. Functional evolution of phosphatidylethanolamine binding proteins in soybean and Arabidopsis. The Plant Cell **27**, 323–336.

Wickland DP, Hanzawa Y. 2015. The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. Molecular Plant **8**, 983–997.

Wu F, Sedivy EJ, Price WB, Haider W, Hanzawa Y. 2017. Evolutionary trajectories of duplicated FT homologues and their roles in soybean domestication. The Plant Journal **90**, 941–953.

Xi W, Liu C, Hou X, Yu H. 2010. MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis. The Plant Cell **22**, 1733–1748.

Xu W, Zhang L, Cunningham AB, Li S, Zhuang H, Wang Y, Liu A. 2020. Blue genome: chromosome-scale genome reveals the evolutionary and molecular basis of indigo biosynthesis in *Strobilanthes cusia*. The Plant Journal **104**, 864–879.

Xu Y, Zhang J, Ma C, Lei Y, Shen G, Jin J, Eaton DAR, Wu J. 2022. Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism. Molecular Plant **15**, 1384–1399.

Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T. 2005. TWIN SISTER OF FT (TSF) acts as a floral pathway integrator redundantly with FT. Plant and Cell Physiology **46**, 1175–1189.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution **24**, 1586–1591.

Yoo SJ, Chung KS, Jung SH, Yoo SY, Lee JS, Ahn JH. 2010. BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence meristem development in Arabidopsis. The Plant Journal **63**, 241–253.

Yu D, Dong X, Zou K, Jiang XD, Sun YB, Min Z, Zhang LP, Cui H, Hu JY. 2023. A hidden mutation in the seventh WD40-repeat of COP1 determines the early flowering trait in a set of Arabidopsis *myc* mutants. The Plant Cell **35**, 345–350.

Zhao F, Liu B, Liu S, *et al*. 2023. Disentangling a 40-year-old taxonomic puzzle: the phylogenetic position of *Mimulicalyx* (Lamiales). Botanical Journal of the Linnean Society **201**, 135–153.

Zhao JX, Wang S, Liu J, *et al*. 2023. A comparative full-length transcriptomic resource provides insight into the perennial monocarpic mass flowering. The Plant Journal **116**, 1842–1855.

Zhong J, Kellogg EA. 2015. Duplication and expression of CYC2-like genes in the origin and maintenance of corolla zygomorphy in Lamiales. New Phytologist **205**, 852–868.