

Structural Changes and Adaptative Evolutionary Constraints In *FLOWERING LOCUS T* and *TERMINAL FLOWER1*-Like Genes of Flowering Plants

Deivid Almeida de Jesus

Universidade Federal do Oeste do Para - Campus Tapajos: Universidade Federal do Oeste do Para

Darlisson Mesquista Batista

Universidade Federal do Oeste do Para

Shayla Salzman

Cornell University

Lucas Miguel Carvalho

Universidade Estadual de Campinas

Kaue Santana (✉ kaue.costa@ufopa.edu.br)

Universidade Federal do Oeste do Para <https://orcid.org/0000-0002-2735-8016>

Thiago André

Universidade Federal do Oeste do Para

Research Article

Keywords: phenology, structural biology, phylogenetics, protein structure prediction, adaptative evolutionary constraints, natural selection.

Posted Date: March 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-270771/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Regulation of flowering is a crucial event in the evolutionary history of angiosperms. The production of flowers is regulated through the integration of different environmental and endogenous stimuli, many of which involve the activation of different genes in a hierarchical and complex signaling network. The *FLOWERING LOCUS T/TERMINAL FLOWER 1 (FT/TFL1)* gene family is known to regulate important aspects of flowering in plants. To better understand the pivotal events that changed FT and TFL1 functions during the evolution of angiosperms, we reconstructed the ancestral sequences of *FT/TFL1-like* genes and predicted protein structures to identify determinant sites that evolved in both proteins and allowed the adaptive diversification in the flowering phenology and developmental processes. Residues from the P-loop domain of the analyzed FT structures showed predominantly high destabilizing mutations which is consistent with constant selective pressure found for this region. In addition, we demonstrate that the occurrence of destabilizing mutations in residues located at the phosphatidylcholine binding sites of FT structure experience positive selection, and some residues of 4th exon are under negative selection, which is compensated by the occurrence of stabilizing mutations in key regions and the P-loop to maintain the overall protein stability. Our results shed light on the evolutionary history of key genes involved in the diversification of angiosperms.

Article Highlights

- FT-like genes were present in the common ancestor of gymnosperms and angiosperms approximately 190 million years ago.
- Substitution in the third position of the codons that encode the key residues involved with FT/TFL1 function did not alter the encoded amino acids and it is in accordance with the existence of purifying selection.
- Residues from the P-loop domain of the analyzed FT structures showed predominantly high destabilizing mutations which is consistent with constant selective pressure found for this region.
- The appearance of destabilizing mutations in residues of phosphatidylcholine binding sites were selected by negative selection, while some residues of the 4th exon were found to be under positive selection.
- Some of the structural changes in FT were compensated by the occurrence of stabilizing mutations in key regions to maintain the overall stability of the protein.
- The presence of destabilizing mutations and negative selective pressures in residues located at the phosphatidylcholine binding site involved with H-bond formation indicates their structural role to maintain the overall stability of FT structure.

Key Message

FT and TFL1 proteins have experienced mainly negative selection however, some of the structural changes in FT were compensated by the occurrence of stabilizing mutations in key regions to maintain the overall stability of the protein.

Introduction

Flowering is a major event in the angiosperms life cycle because it allows sexual reproduction (Jin et al. 2020). The production of reproductive meristems and flower organs is regulated through the integration of environmental and endogenous stimuli (Ahn et al. 2006) involving the activation of different genes and a complex and hierarchical signaling network. The *FLOWERING LOCUS T (FT)* and *TERMINAL FLOWER 1 (TFL1)* gene family is known to regulate important aspects of growth and flowering in plants (Pin and Nilsson 2012; Jin et al. 2020). FT proteins are key regulators activated by the transcription factor CONSTANS (CO) and are involved with the control of multiple flowering pathways in angiosperms (Kobayashi et al. 1999; Kim et al. 2008; Ogiso-Tanaka et al. 2013; Fan et al. 2020). When activated, the FT interacts with the FD, a basic leucine zipper domain transcription factor (bZIP), and induces the expression of *APETALA1 (AP1)* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)* genes, leading to the flower development (Wigge 2005; Collani et al. 2019; Zhu et al. 2020). Contrasting to FT activity, TFL1 proteins are known to repress flowering by inhibiting expression of key flowering pathway genes (Kardailsky et al. 1999; Lifschitz et al. 2006). Besides flowering regulation, FT/TFL1-like proteins have also been identified as regulatory factors in a wide range of developmental processes in plants that includes seed germination (Xi et al. 2010), stomatal opening (Kinoshita et al. 2011), response to extended cold winter temperatures (Pin et al. 2010), control of the lateral shoot development (Hiraoka et al. 2013), and formation of storage organs (Navarro et al. 2011).

FT and TFL1 proteins are homologs to phosphatidylethanolamine-binding proteins (PEBPs) which are all involved in the signaling pathways that control differentiation of stem apical meristem (Bernier and Périlleux 2005; Liu et al. 2016). Genes reported belonging to PEBPs superfamily include *CENTRORADIALIS (CEN)* (Banfield and Brady 2000), *TWIN SISTER OF FT (TSF)* (Yamaguchi et al. 2005), *BROTHER OF FT AND TFL1 (BFT)* (Yoo et al. 2010), and *MOTHER OF FT AND TFL1 (MFT)* (Yu et al. 2019), and others (Bernier and Périlleux 2005). Regarding the evolution of these genetic regulators, phylogenetic analyses revealed that these PEBPs-like genes are grouped in three main clades: FT-like, TFL1-like, and MFT-like genes. Similarly, gymnosperms possess two groups: MFT-like and a group that occupies an intermediate position between the FT- and TFL1-like (FT/TFL1-like) genes (Karlgrén et al. 2011). Recently studies have demonstrated that FT/TFL1-like sequences were present in gymnosperms lineages in duplicates, which could have occurred even prior to the emergence of seed plants (Liu et al. 2016). Genomic analyses revealed that gene duplications played an important role in the diversification of gene function in angiosperms, which were essential for adaptive evolution (Soltis et al. 2015). Different studies have reported that MFT genes are ancestral to FT and TFL1, and the origin of these orthologue genes is related to occurrence of duplication events in the evolutionary history of angiosperms (Hedman et al. 2009; Karlgrén et al. 2011; Wickland and Hanzawa 2015).

The FT and TFL1 of *Arabidopsis thaliana* exhibit conserved structures (Ahn et al. 2006) with small tractable changes differentiating them. Studies have demonstrated that mutations in four key residues, Glu109, Trp138, Gln140, and Asn152 could transform the activator function of FT into the suppressor

activity of TFL1 (Hanzawa et al. 2005; Ho and Weigel 2014). In addition, an external loop region of 14 residues in FT, located at *A. thaliana* position 139 to 144, named the P-loop, has been shown to confer an antagonistic activity to the floral regulators (Ahn et al. 2006). To better understand the pivotal events that changed FT and TFL1 functions during the evolution of flowering plants, and the structural role of residue sites in both proteins throughout the diversification of angiosperms, we reconstructed ancestral sequences of the FT/TFL1 genes, predicted the corresponding protein structures and performed structural mutational analyzes utilizing the genetic model of *A. thaliana* in addition to *Chamaecostus fusiformis* (Maas) C.D. Specht & D.W. Stev. (Costaceae, Zingiberales), an Amazonian understory ground herb that shows intermittent growth and tight regulation of flowering in response to climate seasonality (André et al. 2015).

Material And Methods

RNA extraction and RNA-seq analyses

The variability of the transcribed genome of *C. fusiformis* was obtained from a native population from Novo Progresso (Southern Amazonia, Pará, Brazil) using RNA-seq analyzes (Moumeni et al. 2011). Leaves and root tubers of four individuals were collected in April 2016, in the rainy season, and fixed in homemade RNA-later. During the dry season, in November 2016, only root tubers were present. Genomic libraries and RNA-seq in the Hi-Seq platform were conducted by Novogene® (Sacramento, CA, USA). Sequencing quality was accessed through FastQC (Wingett and Andrews 2018). We used Trimmomatic (Bolger et al. 2014) for the removal of low-quality and adapter sequences, with the following parameters: ILLUMINACLIP:TrueSeq3-PE.fa:2:30, LEADING:20, TRAILING:20, SLIDINGWINDOW:4:20, MINLEN:90. We used Trinity (Grabherr et al. 2011) for *de novo* transcript assembly with default parameters and contig minimum size of 200 bp. We annotated via BLAST (Altschul et al. 1990) using the UniRef90 database (Suzek et al. 2015).

Phylogenetic Analysis and Ancestral Sequence Inference

Coding DNA sequences (CDSs) of FT- and TFL1-like genes were retrieved from GenBank (Benson et al. 2017) through the BLASTn tool using as reference the FT and TFL1 sequences of *Arabidopsis thaliana* (GeneBank accession codes: FT: NM_105222.3; TFL1: NM_120465.3). To infer gene phylogenetic trees, we used a total of 57 coding sequences of FT (53 sequences from angiosperms, belonging to monocots, eudicots, asterids, and brassicales groups; and 3 sequences from gymnosperms), including *C. fusiformis*; and 50 sequences of TFL1 (46 sequences from angiosperms belonging to monocots and eudicots; and 4 sequences from gymnosperms), but no sequence for *C. fusiformis* was retrieved from the transcriptome. In both phylogenetic inferences, gymnosperm sequences of FT and TFL1 were used as outgroups (Supporting information 01). Multiple sequence alignments were performed in MUSCLE (Edgar 2004) as implemented in MEGA 7 (Kumar et al. 2016). These alignments were further manually edited, mainly to maintain reading frames. The jModel Test 2 program was used to select the best nucleotide model with the Bayesian information criterion (BIC) scores (Darriba et al. 2012). We used MEGA7 (Kumar et al. 2016) for ancestral sequence reconstruction. Bayesian phylogenetic inference was executed in BEAST version 1.8.4 (Drummond et al. 2012). A relaxed clock with an uncorrelated lognormal model of rate variation was used and the Yule speciation process for branching rates was selected. The GTR+I+G evolutionary model was selected. Seventeen fossil-based time to the most recent common ancestor (tmrca) calibrations were used based on published data. Calibration dates and associated citations can be found in Supporting information 02. A CTMC rate prior was selected and no monophyletic prior assignment was made. Markov chain Monte Carlo simulations were run for 5×10^7 generations and sampled every 1×10^3 . These analyses were performed in CIPRES Science Gateway server v3.3 (Miller et al. 2010). To analyze the continuous parameter values sampled from the Bayesian chains, we assessed the convergence of the models across independent runs by analyzing plots of the marginal later distributions in Tracer (version 1.7.1) (Rambaut et al. 2018). To ensure high effective sample size (ESS) values, we considered a value above or equal to 200 ($ESS \geq 200$). Tracer was also used to assess burn-in levels and a maximum clade-credibility tree was obtained from the later sample of trees using TreeAnnotator v. 1.7.1 (Drummond et al. 2012).

Prediction of FT and TFL1 Structures

Structures of FT/TFL1 proteins were modeled by comparative modeling in Modeller program version 9.19 (Fiser and Šali 2003), which uses the satisfaction of the spatial restraints method. The crystallographic structures of FT (PDB code: 1WKP, chain A; resolution: 2.6Å) and TFL1 (PDB code: 1WKO, chain A; resolution: 2.6Å) from *A. thaliana* were used as templates. To predict structures, we performed a pairwise sequence alignment using the BLOSUM62 (20x20) matrix. The models of both proteins were optimized at the atomic-level in the ModRefiner program (Xu and Zhang 2011) and then, minimized by 1,000 cycles of conjugated gradient and 1,000 cycles of steepest-descent algorithms in the Amber16 package (Salomon-Ferrer et al. 2013). The modeled protein structures were validated by the stereochemical quality using the Ramachandran plot obtained in Procheck program version 3.5.4 (Laskowski et al. 1993) and the energetic profile obtained by the Qmean plot (Benkert et al. 2009). Moreover, structural alignment and RMSD-C α values were used to evaluate the conservation of the modeled FT and TFL1 structures with the selected templates. Finally, to analyze the surface potentials of the FT and TFL1 regions involved in their molecular activity, we obtained the Poisson-Boltzmann electrostatic potential map using the PDB2PQR server (Dolinsky et al. 2004).

Molecular Dynamics Simulation

To analyze structural changes in FT and TFL1, molecular dynamics (MD) simulations were performed in the Amber16 package (Case et al. 2005). The all-atom forcefield Amberff14SB was used to parameterize the protein structures. The proteins were solvated in a truncated octahedral water-box with the explicit solvation model TIP3P (Jorgensen et al. 1996). We used a distance of 10 Å between the cell wall and the solvated atoms of the system, and a distance of 0.8 Å between water molecules and the solute. Counter-ions Cl⁻ were also added to neutralize the analyzed systems. Initially, all hydrogen atoms of the system were minimized for 3,000 cycles of the steepest-descent (Wiberg 1965) and 3,000 cycles of the conjugate gradient algorithm (Hestenes and Stiefel 1952).

Water and ions were minimized for 2,000 cycles of the steepest-descent and 3,000 cycles of the conjugate gradient. Then, the whole system was minimized using 2,000 cycles of the steepest descent and more 3,000 cycles of the conjugate gradient; and we performed seven repetitions, with progressive relaxing of restraints. After completion of minimization, the system was gradually heated to increase the temperature to 300 K during 4.25 ns of equilibration. Then, the MD was performed with the isobaric-isothermal ensemble with a total time of 20 ns. The temperature was maintained using the Langevin thermostat and the SHAKE algorithm was applied to all hydrogens of the system, which allowed us to use integration cycles of 2.0 fs and the constant isotropic pressure was maintained at 1 bar by using the Berendsen barostat. A cutoff of 10 Å was used for the minimum image convention. The values of the RMSD and RMSF, based on the heavy atoms of the protein backbone were used to determine the conformational changes over the MD simulations.

Mutational Analysis of Protein Structures

To analyze the effects of mutations in the coding sequences on the structural analysis of the proteins, alanine scanning was performed using the FoldX program (Gueriois et al. 2002). FoldX uses a linear combination of different empirical terms to calculate free energy (ΔG). Empirical terms include Coulomb terms for electrostatic interactions, van der Waals terms, hydrophobic and solvation forces, hydrogen bonds, and so on. The results of mutational analyses ($\Delta\Delta G_{\text{fold}}$) are expressed by the difference between the free energy of the wild-type (ΔG_{wt}) and the mutant (ΔG_{mut}) structures (kcal.mol^{-1}) according to Equation 1:

$$\Delta\Delta G_{\text{fold}} = \Delta G_{\text{fold,wt}} - \Delta G_{\text{fold,mut}} \quad (1)$$

Where $\Delta G_{\text{fold,wt}}$ is the free energy variation of wild-type structure and $\Delta G_{\text{fold,mut}}$ is the variation of the mutant structure. If $\Delta\Delta G < 0$, the mutation was considered stabilizing, and if $\Delta\Delta G > 0$, the mutation was considered structurally destabilizing (Morrison & Weiss, 2001). The mutations were classified in five different categories, depending on the implications to the stability of protein structure: highly stabilizing ($\Delta\Delta G < -1.84 \text{ kcal.mol}^{-1}$); slightly stabilizing ($-1.84 \text{ kcal.mol}^{-1} \leq \Delta\Delta G < -0.46 \text{ kcal.mol}^{-1}$); neutral ($-0.46 \text{ kcal.mol}^{-1} < \Delta\Delta G \leq +0.46 \text{ kcal.mol}^{-1}$); slightly destabilizing ($+0.46 \text{ kcal.mol}^{-1} < \Delta\Delta G \leq +1.84 \text{ kcal.mol}^{-1}$); highly destabilizing ($\Delta\Delta G > +1.84 \text{ kcal.mol}^{-1}$).

Calculations of dN/dS rates

To identify residues under natural selection, i.e., neutral, positive or negative selections, we calculated the non-synonymous mutation (dN) and the synonymous mutation (dS) rates using the clade and codon models available in EasyCodeML program (Gao et al. 2019), which implements the clade and codon-based models of CodeML. The clade model was based in the model C (CmC) to estimate a separate ω ratio for each clade, and it was compared against a null model 2a_rel (M2a_rel), in which ω is fixed among the analyzed clades (Anisimova and Kosiol 2007; Weadick and Chang 2012). FT's ω values were obtained for the most recent common ancestor of five major clades: angiosperms, monocots, eudicots, asterids, and brassicales. TFL1's ω values were obtained for angiosperms, monocots, eudicots, brassicales, and non-brassicales. The ω values were validated by likelihood scores.

Results And Discussion

We inferred the ancestral sequences and predicted ancestral structures of FT and TFL1 for the major clades of flowering plants. Our results allow us to identify the structural changes of proteins crucial to the flowering process throughout the evolutionary history of angiosperms, as well as, the structural implications of mutations at key residues for the molecular function of both FT and TFL1 proteins.

Reconstruction of Evolutionary History of FT and TFL1 in Angiosperms

In the present study, the TIM2ef+I+G was selected as the most appropriate evolutionary model for the FT and TFL1 sequences. Using a Bayesian relaxed-clock approach, we estimated the phylogeny and the divergence time for FT and TFL1 genes. Our phylogenetic analyses revealed that the natural selection acted differently for both genes, thus they did not show the same evolutionary divergence found in angiosperms (Flagel and Wendel 2009). These results corroborate previous phylogenetic trees obtained for both genes (Wang et al. 2017).

Our phylogenetic hypothesis for gene evolution demonstrates that FT-like genes diverged ca. 190 million years ago and TFL1-like genes ca. 168 million years ago. Within the angiosperm clade, the FT-like genes diversified approximately 135 million years ago (Figure 1) and the TFL1-like genes around 132 million years ago (Figure 2). Klintonäs et al. (2012) suggested that FT is found exclusively in flowering plants. However, FT-like genes are present in gymnosperms.

In the angiosperm clade, we recovered four monophyletic groups of FT-like sequences: monocots (posterior probability = 0.98), eudicots (0.98), asterids (0.78), and brassicales (0.70). Regarding the TFL1 sequences, we also recovered distinct clades for the monocots (1.0), eudicots (0.55), brassicales (1.0), and an unresolved group which we named non-brassicales (0.98) (Figure 1). The estimated time of evolutionary divergence of FT-like sequences is approximately 112 million years ago for monocots, 125 for eudicots, 100 for asterids, and 104 for brassicales. Divergence time for TFL1 sequences was relatively similar, if not slightly earlier. Monocot TFL1 diverged approximately 107, eudicots 110, brassicales 85, and the non-brassicales group 80 million years ago (Figure 2).

We conjecture that differences between the phylogenetic trees of FT/TFL1-like genes and the well-known angiosperm phylogeny could be caused by divergent evolution related to the pleiotropic effects exercised by the FT/TFL1-like genes in angiosperms, which influence a wide range of developmental stages of

plants, such as flowering, seed storage, and stomatal opening (Yamaguchi et al. 2005; Lifschitz et al. 2006; Xi et al. 2010; Pin and Nilsson 2012). Pleiotropy has a functional diversity regulating different phenotypic characteristics, thus influences the natural selection pressures leading to the appearance or elimination of new characteristics (Auge et al. 2019). Our phylogenetic analysis shows the formation of well-supported clades for the main monophyletic groups of angiosperms, thus indicating that the natural selection acts to preserve the FT/TFL1 functions in each clade during the evolutionary divergence of these genes. Indeed, a prior study of dN/dS rates in duplicated FT/TFL1-like genes showed that they suffer negative pressures (Mackenzie et al. 2019).

Inference of the Ancestral Sequences of FT and TFL1

Based on the phylogenetic trees, we selected five ancestral FT- and TFL1-like sequences, which are representative of five clades with satisfactory posterior support: monocots, eudicots, asterids (FT), brassicales, and non-brassicales (TFL1). Moreover, we analyzed the current representatives of protein structures (FT and TFL1) in *A. thaliana* (reference structure) and *C. fusiformis* (FT). The later species provides a unique view of the role of this protein in the life cycle timing due to the presence of phenologic and structural adaptations to climatic seasonality (André et al. 2015), particularly the senescence of shoot organs during the dry season.

Structurally, FT shows seven β -strands and four α -helices, except the ancestral structure of the asterids clade that exhibits three α -helices and six β -strands, FT structure from ancestral angiosperms shows four α -helices and nine β -strands (Figure 3, panel A and C) and the *C. fusiformis* shows three α -helices and five β -strands (Figure 3, panel A and E). This can be understood by the absence of the first exon, which is evident when we align the ancestral sequences and the sequences of the other monocots representatives and which will be further discussed below. The FT protein structure of *C. fusiformis* revealed structural changes when compared with the ancestral structure of the angiosperms and monocot clades, as well as that of *A. thaliana* (see Figure 3, panel D and E). In contrast, the TFL1 structures contain predominantly three α -helices and seven β -strands (Figure 3, panel B and F). It is interesting to note the absence of the TFL1 gene in the *C. fusiformis* transcriptome, which included both root tubers and leaves collected at both dormancy and growth phases of the life cycle. We propose that FT paralogs have assumed the floral repression function, as evidenced in a previous study of *Beta vulgaris* ssp. *vulgaris* in which the paralogs BvFT1 and BvFT2 have antagonistic functions (Pin et al. 2010).

Studies have demonstrated that a short loop segment located between the residues 128 to 145 (P-loop domain) is well conserved across plant families and is the major determinant of FT activity (Pin et al., 2010; Pin & Nilsson, 2012). Our reconstruction of ancestral angiosperm FT found both the highly conserved P-loop domain (ancestral sequences: Leu127 to Asn142; *A. thaliana*: Leu131 to Asn146) as well as the key residues Tyr85 and Gln40 (Figure 4, panel A). In contrast, the homolog region of P-loop from TFL1 ancestral sequences is less conserved showing more substitutions through the evolutionary history of angiosperms, which suggests the non-functionality of these regions in flowering repression (Figure 4, panel B). However, the key-residue His88 was found to be conserved in all analyzed ancestral sequences of angiosperms.

A previous study demonstrated that the flowering time in sugar beet (*Beta vulgaris* ssp. *vulgaris*; Eudicots clade) is controlled by the combined activity of two paralogs, which have an antagonistic function: BvFT1 and BvFT2 (Pin et al. 2010). The P-loop domain of BvFT1 and BvFT2 show substitutions in the residues Tyr134Asn and Trp138Gln (*B. vulgaris* spp. *vulgaris* numbering), which are involved in flowering repression (Pin and Nilsson 2012). In the present study, we noted a new residue position for the analyzed ancestral sequences of the P-loop domain when compared with the *A. thaliana* sequence. Within TFL1, the residues Pro134 and Arg138 found in *A. thaliana* remained conserved in the analyzed ancestral sequences and a substitution in the homolog region of the P-loop domain was found at Pro134Asn in the ancestral sequence of monocots.

Ho and Weigel (2014) identified that the position of the residue 134 and 138 in the FT structure of *A. thaliana* are located at the external P-loop domain and easily accessible to the surface of the phosphatidylcholine (PC) binding site, thus indicating interaction with other molecules (Ho and Weigel 2014). It is therefore expected that purifying selection acts to conserve this structure over evolutionary time. Analyzing the nucleotide frequency that encodes the key residues for FT/TFL1 function, as well as the C-terminal region of both proteins, we observed that substitution in the third position of codons that encode key residues of FT (Tyr85 and Gln140) and TFL1 (His88 and Asp144) involved with molecular function does not alter the encoded amino acids (Supporting information 03). This is in accordance with the dN/dS rates found for this region and indicates the existence of purifying selection on these codon sites for both genes. We observed negative selection for the homologous P-loop region the TFL1 domain, even though variations in the frequency of some codons were observed (Figure 4, panel B).

Structural Comparison of Ancestral FT and TFL1 with Their Counterparts of *A. thaliana*

The modeled structures of the FT / TFL1 proteins exhibited > 90% of residues in favorable regions of Ramachandran plot (sum of residues in most favorable and permitted regions, Supporting information 04 and 05), and a satisfactory energetic profile as exhibited by the local quality estimation of Qmean (Supporting information 06 and 07), which indicate reliable structures. In addition, the RMSD plots obtained over the MD simulation showed that modeled structures reached a stable conformation after 18ns (Supporting information 08).

Previous studies have demonstrated that the C segment located at the C-terminal region of FT/TFL1 structures is crucial for the molecular function of both proteins (Ahn et al. 2006; Hedman et al. 2009). Based on this assumption, we performed a comparison using structural and sequence alignment of the C-terminal region of the inferred ancestral FT and TFL1 to compare them with their counterparts in *A. thaliana*, used as a reference for the structural analyses. We observed that the ancestral structures of FT proteins are conserved (Figure 5A), exhibiting an α -helix, and a β -strand in segment C (Ahn et al. 2006). The structural alignment of the C segment of the ancestral structures of FT protein with the *A. thaliana* structure (reference) showed an RMSDs-C α value ≤ 1.153 Å, which demonstrates low structural changes for these analyzed groups in angiosperms (Table 1).

The residues sequences of the C segment of FT from *A. thaliana* showed mutations at three residues positions (Asp146Glu, Leu150Ileu, and Phe162Tyr), *C. fusiformis* showed mutations in four residues (Glu149Ala, Leu155Asp, Leu160Val, Tyr161Phe), when compared with the others analyzed ancestral sequences (Figure 5C). Structural conservation was also found in the C-terminal region of TFL1 which is composed of an α -helix and a β -strand (Figure 5B). The structural alignment of segment C showed RMSD-C α values ≤ 2.06 Å, which demonstrated that those structures were conserved during the evolutionary time (Table 1). We noted mutations in four residues in TFL1 sequences of *A. thaliana* at the positions Asn150Lys, Ala153Val, Asn155Tyr, and Tyr165Phe when compared with the ancestral sequences. Regarding the FT structure of *C. fusiformis*, we noticed the conservation of the C-terminal region which are located the key residues and segment C encoded by the 4th exon of FT gene (Figure 5 A and C).

Comparing only the ancestral structures of angiosperms, monocots, eudicots, asterids, and brassicales clades, we noted that these proteins remained well-conserved with a substitution, however, at the residue Ser155 in the ancestral monocot structure when compared with the other angiosperm ancestral sequences. In contrast, comparing with *A. thaliana*, this position contains a leucine substitution (Figure 5C). Comparing the modeled TFL1 ancestral structures, we recover a mutation at the position Gln146Asn in the ancestral sequence of monocots (Figure 5D). The residues Tyr85 and Gln140 (*A. thaliana* numbering) involved in the repression of FT activity and His88 and Asp144 (*A. thaliana* numbering) involved in the activation of TFL1, remained conserved in all inferred ancestral sequences of angiosperms (Ahn et al. 2006; Ho and Weigel 2014).

Table 1 Structural comparison between C segments of modeled ancestral structures of FT and TFL1 with their counterpart structure from *A. thaliana* (RMSD-C α values exhibited in angstroms)

FT-like structures		TFL1-like structures	
Taxonomic groups	RMSD-C α (Å)	Taxonomic groups	RMSD-C α (Å)
Angiosperms	0.86	Angiosperms	1.11
Monocots	1.06	Monocots	1.20
Eudicots	0.84	Eudicots	2.06
Brassicales	0.87	Brassicales	1.11
Asterids	0.90	Non-brassicales	1.35
<i>C. fusiformis</i>	1.62		

Flowering time in angiosperms is regulated in part by phosphatidylcholine (PC) interaction with FT protein (Nakamura et al. 2014). Therefore, we analyzed the distribution of electrostatic charges on the surface of the PC binding site of FT to understand how this region changes across the ancestral structures of the main clades of angiosperms (Figure 6). We observed that the electrostatic surface of the residues located at the PC binding site, such as Pro8, Val11, Arg13, Asp17, Leu41, Pro77, Arg83, Ile117, and Arg119 (*A. thaliana* numbering) exhibited a predominantly positive potential in the ancestral sequences, which is in accordance with previous findings for the FT structure of *A. thaliana* (Nakamura et al. 2019). However, we observed some structural changes in the residues of the PC binding site of *C. fusiformis* which only contains the residues Ser8, Pro44, Arg50, Ile84, and Arg86 (Figure 6). This structural difference is due to the absence of the 1st exon in the nucleotide sequence that encodes this protein region. Previous studies have demonstrated that molecular mechanisms of PC alter during the daytime to promote flowering (Nakamura et al. 2014; Nakamura et al. 2019), thus we conjecture that the seasonal changes and restrictions on the flowering time that *C. fusiformis* experiences could influence the activation of the PC binding site in the FT structure. Moreover, the electrostatic potential map not showed alterations in the protein surface, when compared with other FT structures, suggesting that this variant structure could preserve its molecular function even in the absence of some residues in the PC binding site.

Two regions of FT structure have been pointed to regulate the flowering; the first one corresponds to a surface-exposed loop region, named segment B (residues 128 to 141) encoded by the fourth exon which is involved in PC binding, and the second one corresponds to Tyr85, a key functional residue that differentiates FT activity from the floral repressor TFL1 (Lee et al. 2013; Zhang et al. 2015). Structural analyses of the PC binding site of the FT structure revealed that the residue Tyr85 formed an H-bond with the oxygen of residues Glu109 and Gln112, which confers stability to the PC binding site. Similarly, residue Gln140, located in segment B is also involved with the H-bond network, whereas His87 and Arg139 stabilize the spatial coordination of Tyr85 by van der Waals interaction (Nakamura et al. 2019). Moreover, studies have also demonstrated the formation of H-bonds between the residue His88 with Asp144 in the TFL1 structure of *A. thaliana* (Ho and Weigel 2014). Based on these assumptions, we performed a structural analysis of H-bond interactions in the adjacent residues to Tyr85 in FT and His88 in TFL1 structures through the MD simulation in the ancestral sequence representative to angiosperms clade of both proteins. Our analyses demonstrate that these interactions maintain stability over the MD simulation (Figure 7A-D), thus indicating the conservation of these residue interactions over evolutionary time.

The structural analysis shows that the FT Tyr85 residue (Tyr84 in the ancestral sequence from angiosperms) does not form an H-bond, however, Glu109 (Glu108 in angiosperms) does interact with Arg111 (Figure 7A). In the structure of TFL1, Asp144 (Asp139 in the ancestral sequence of angiosperms) interacts with the residues Ser137 and His140. Differently, His88 (His84 in the ancestral sequence of angiosperms) does not interact with the other adjacent residues (Figure 7C).

Analysis of Mutations Occurring During FT/TFL1 Evolution and Their Effect on Protein Structural Stability

The number of mutations that confer advantageous changes during the evolutionary process is limited due to the crossing an energetic barrier of the fitness landscape of protein structures that could lead to alterations in their stability and function (Tokuriki and Tawfik 2009; Faber et al. 2019). Mutational analysis has been widely applied to correlate the structure with protein function (Brandt et al. 2014; da Costa et al. 2017; Bhattacharya et al. 2018), organism phenotype (Neves Cruz et al. 2019), and to analyze protein evolution from ancestral sequences and adaptative evolutionary constraints in proteins structures (Studer et al. 2014; Sharir-Ivry and Xia 2018). In the present study, we performed a mutational analysis for FT and TFL1 structures using alanine scanning to investigate the influences of mutations in the structural stability in both proteins and correlate these results with the selective pressures indicated by dN/dS rates of the FT (Figure 8, panel A) and TFL1 (Figure 8, panel B) codons.

Our results reveal that most alanine substitutions in FT and TFL1 sequences are destabilizing, especially at residue Tyr85 that plays an important role in FT activity, and at residue His88 in TFL1 structure that is involved with the flowering repression (Ahn et al. 2006; Ho and Weigel 2014). The residues from the P-loop domain (Pin and Nilsson 2012) of FT structure from *A. thaliana* showed predominantly high destabilizing mutations, especially the residues Phe128 (+ 4.86 kcal.mol⁻¹), Gln130 (+ 2.25 kcal.mol⁻¹), Gly132 (+ 3.12 kcal.mol⁻¹), Gly140 (+ 2.12 kcal.mol⁻¹), Arg142 (+ 2.24 kcal.mol⁻¹), and Phe145 (+ 3.89 kcal.mol⁻¹) (Figure 8). Moreover, we find that the codons that encode these residues are under constant selective pressure, corroborating with previous finding for FT/TFL1 genes (Klinteräs et al. 2012; Mackenzie et al. 2019).

Analyzing dN/dS rates of the residues involved with the H-bond interactions in the FT structure present in *A. thaliana* and ancestral sequence of the angiosperm clade, we observed the presence of negative pressure (dN/dS<1), and the presence of destabilizing mutations ($\Delta\Delta G>0$), thus corroborating purifying selection over evolutionary time. These findings also corroborate with previous observation that the presence of H-bond interactions in the PC binding site maintains its structural stability (Nakamura et al. 2019). Similarly, mutations in the residues that formed H-bond interactions in the PC binding site were destabilizing for overall protein structure and showed purifying selection in ancestral angiosperm and *A. thaliana* TFL1.

We also performed analyses of the evolutionary rates of the ancestral proteins of each clade based on the inferred phylogenies (see Figures 1 and 2). Considering the FT sequences, we observe that the angiosperm clade showed a low dN/dS rate when compared with the other ancestral sequences ($\omega_{\text{angio}} = 0.065$). The class of substitution sites selected were the following: neutral ($\omega = 1$), sites under positive selection ($\omega > 1$), and sites under negative selection ($\omega < 1$) (Nei and Gojobori 1986). The ancestral sequences of FT probably evolved under negative selection with low rates or absence of non-synonymous substitutions, which conserved the protein structure during the evolutionary history that led to the appearance of flowering plants. The ancestral sequences of monocots exhibited a dN/dS rate $\omega = 0.081$, which suggests that the occurrence of synonymous mutations was influenced by natural selection (negative pressure). There is evidence that natural selection is purifying, which may explain little change in ancestral sequences. Similar results were also found for ancestral sequences of eudicots ($\omega = 0.087$) and asterids ($\omega = 0.078$). In contrast, the ancestral sequence of brassicales exhibited a high dN/dS rate ($\omega = 0.147$) when compared with the other analyzed sequences which could indicate the predominance of non-synonymous mutations at the residues sites. The negative pressure of the FT gene (rates dS > dN) could be related to its different molecular functions in the flowering process (Pin and Nilsson 2012). Analyzing the evolutive trajectory of FT in the monocots, we noticed that majority of the investigated species showed dN/dS rates compatible with purifying selection, similarly when compared with the ω value for the clade ($\omega = 0.081$), thus showing few residues under positive selection. In addition, our mutational analyses demonstrated that the FT structure of *C. fusiformis* showed high destabilizing mutations (Supporting information 09), probably due to the absence of the first exon. We hypothesize that during the evolutionary adaptation of this species, positive pressure in regions not directly involved with the encoding of key residues as well as the PC binding site acted to remove the disadvantageous mutations.

Regarding the ancestral TFL1 sequences of the angiosperm's clade, we obtained a $\omega = 0.115$, and for monocots $\omega = 0.186$, eudicots $\omega = 0.247$, non-brassicales $\omega = 0.300$ and brassicales $\omega = 0.236$ (Figure 2). As previously discussed, the key residues His88 and Asp144 involved with TFL1 function remained under negative selection, however, we document regions of exon 4 under positive selection (angiosperms with residues Val109, Ser111, Lys154; monocots: Gly111, Gln138, Ala144, Gln154; eudicots: Val109, Ser111, Arg116, Lis154; non-brassicales: Val109, Arg116, Tre139, Lis154;) (see Supporting information 10 and 11). This is reflected in the increase in dN/dS over time when compared to the values found for FT, which may indicate an adaptive modification of TFL1 protein structure (Benner et al. 2002), possibly due to its influences on different phenotypic traces of the flowering plants (Rallapalli et al. 2014). Additionally, TFL1 proteins probably evolved by a stability regime with a balance between stabilizing and destabilizing mutations that could lead to the divergence of large groups in the angiosperms (Dasmeh et al. 2013).

During the evolutionary process, new mutations can be retained in the genomic pool by the relative strengths of the natural selection and genetic drift. Regarding the natural selection forces, the rate of fixation of these mutations is accelerated by positive selection, under which favorable mutations to protein function or stability tend to be retained and in contrast, it is decelerated by the negative selection that tends to remove from the genomic pool disadvantageous mutations (Wang et al. 2019). Moreover, different studies have demonstrated that the appearance of destabilizing mutations in residues directly involved in the active sites could be selected by positive selection, and the destabilization can be compensated by the occurrence of stabilizing mutations in other sites to maintain the overall protein stability (Buller and Townsend 2013; Studer et al. 2014; Sharir-Ivry and Xia 2018). Indeed, our analysis of the evolutionary rates for codons regions of the 2nd and 4th exons of FT gene that encode the P-loop domain and PC binding site found dN/dS < 1, and for TFL1 some codons that encode residues from the 4th exon region found values equal to dN/dS = 1 and dN/dS > 1.

Conclusions

We reconstructed the evolutionary history of FT/TFL1-like proteins in the main flowering plant groups and combining ancestral sequence inference with structural and mutational analyses, we identify the main residues sites that evolved by evolutionary constraints, altered the protein stability, and inverted their function from activation or repression of flowering time. Our results show that the main sites are conserved under negative selection, which includes the P-loop domain and PC binding site of FT structures. The residue Tyr85 located at the PC binding site of FT structure forms an H-bond with the oxygen of the residues Glu109 and Gln112 and confers stability to protein structure. Similarly, the residue Gln140, located in segment B also forms H-bond interactions, whereas His87 and Arg139 stabilize the spatial coordination of Tyr85. The presence of destabilizing mutations and negative pressures in residues located at the phosphatidylcholine binding site involved with H-bond formation indicate their structural role to maintains the FT overall stability. In addition, residues from the P-loop domain of the analyzed FT structures show predominantly high destabilizing mutations which is consistent with constant negative pressure found for this region. We demonstrate that important regions of TFL1 and FT show destabilizing mutations under negative selection. However, some residues of the 4th exon are found to be under positive selection, indicating that structural changes are compensated by the occurrence of stabilizing mutations to maintain the overall stability of the protein. Our study opens up new insights to understand the roles of natural selection in the adaptative evolution of flowering proteins in angiosperms and could further help crop improvement with an economic interest in flower and fruit industries.

Declarations

Funding

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)

Fundação Amazônia de Amparo a Estudos e Pesquisas do Pará (FAPESPA, grant number: AUXPE 88881.159139/2017-01)

Conflicts of interest/Competing interests (include appropriate disclosures)

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Availability of data and material

All data of the present study can be request by e-mail of the corresponding authors.

Sequence under submission to Genbank. *Chamaecostus fusiformis* – Zingiberales FT:

AGTAAGGAAGTGACCAACGGATCCGAGCTCAAGCCGTCGATGGTGGTCAACGAGCCTAGGCTCGATATTCGAGGCCCCGACATGAGGTCTCTTTACACACTAGTTATGG

Code availability

Not applied.

Authors' contributions

Conceptualization: K.S. and TA.; investigation: DAJ, DMB, L.M.C; data curation: K.S., S.S.; writing—original draft preparation, SS, K.S., T.A.; writing—review and editing: T.A, K.S., L.M.C., supervision: T.A and K.S. All authors have read and agreed to the published version of the manuscript.

Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals

Not applied.

Ethics approval

Not applied.

Consent to participate

Not applied.

Consent for publication

Not applied.

References

Ahn JH, Miller D, Winter VJ, Banfield MJ, Jeong HL, So YY, Henz SR, Brady RL, Weigel D (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J* 25:605–614 . <https://doi.org/10.1038/sj.emboj.7600950>

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410 . [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

- André T, Salzman S, Wendt T, Specht C (2016) Speciation dynamics and biogeography of Neotropical spiral gingers (Costaceae). *Mol Phylogenet Evol* 103:55–63 . <https://doi.org/10.1016/j.ympev.2016.07.008>
- Anisimova M, Kosiol C (2007) Investigating Protein-Coding Sequence Evolution with Probabilistic Codon Substitution Models. <https://doi.org/10.1093/molbev/msn232>
- Auge GA, Penfield S, Donohue K (2019) Pleiotropy in developmental regulation by flowering-pathway genes: is it an evolutionary constraint? *New Phytol* 224:55–70 . <https://doi.org/10.1111/nph.15901>
- Banfield MJ, Brady RL (2000) The structure of *Antirrhinum centroradialis* protein (CEN) suggests a role as a kinase regulator. *J Mol Biol* 297:1159–1170 . <https://doi.org/10.1006/jmbi.2000.3619>
- Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res* 37:W510–W514 . <https://doi.org/10.1093/nar/gkp322>
- Benner S a, Caraco MD, Thomson JM, Gaucher E a (2002) and Molecular Histories of Life. *Science* (80-) 296:864–868
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2017) GenBank. *Nucleic Acids Res* 45:D36-42 . <https://doi.org/10.1093/nar/gkw1070>
- Bernier G, Périlleux C (2005) A physiological overview of the genetics of flowering time control. *Plant Biotechnol J* 3:3–16 . <https://doi.org/10.1111/j.1467-7652.2004.00114.x>
- Bhattacharya S, Banerjee A, Sah PP, Mal C, Ray S (2018) Mutations and functional analysis of 14-3-3 stress response protein from *Triticum aestivum*: An evolutionary analysis through in silico structural biochemistry approach. *Comput Biol Chem* 77:343–353 . <https://doi.org/10.1016/j.compbiolchem.2018.09.013>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120 . <https://doi.org/10.1093/bioinformatics/btu170>
- Brandt W, Backenköhler A, Schulze E, Plock A, Herberg T, Roese E, Wittstock U (2014) Molecular models and mutational analyses of plant specifier proteins suggest active site residues and reaction mechanism. *Plant Mol Biol* 84:173–188 . <https://doi.org/10.1007/s11103-013-0126-0>
- Buller AR, Townsend CA (2013) Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc Natl Acad Sci* 110:E653–E661 . <https://doi.org/10.1073/pnas.1221050110>
- Carmona MJ, Calonje M, Martínez-Zapater JM (2007) The FT/TFL1 gene family in grapevine. *Plant Mol Biol* 63:637–50 . <https://doi.org/10.1007/s11103-006-9113-z>
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–88 . <https://doi.org/10.1002/jcc.20290>
- Chardon F, Damerval C (2005) Phylogenomic analysis of the PEBP gene family in cereals. *J Mol Evol* 61:579–90 . <https://doi.org/10.1007/s00239-004-0179-4>
- Collani S, Neumann M, Yant L, Schmid M (2019) FT Modulates Genome-Wide DNA-Binding of the bZIP Transcription Factor FD. *Plant Physiol* 180:367–380 . <https://doi.org/10.1104/pp.18.01505>
- da Costa KS, Galúcio JMP, Leonardo ES, Cardoso G, Leal É, Conde G, Lameira J (2017) Structural and evolutionary analyses of *Leishmania Alba* proteins. *Mol Biochem Parasitol* 217:23–31 . <https://doi.org/10.1016/j.molbiopara.2017.08.006>
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772 . <https://doi.org/10.1038/nmeth.2109>
- Dasmeh P, Serohijos AWR, Kepp KP, Shakhnovich EI (2013) Positively Selected Sites in Cetacean Myoglobins Contribute to Protein Stability. *PLoS Comput Biol* 9:1–12 . <https://doi.org/10.1371/journal.pcbi.1002929>
- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665-7 . <https://doi.org/10.1093/nar/gkh381>
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian P hylogenetics with BEAUti and the BEAST 1 . 7 Research article. 29:1969–1973 . <https://doi.org/10.1093/molbev/mss075>
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797 . <https://doi.org/10.1093/nar/gkh340>
- Faber MS, Wrenbeck EE, Azouz LR, Steiner PJ, Whitehead TA (2019) Impact of In Vivo Protein Folding Probability on Local Fitness Landscapes. *Mol Biol Evol* 36:2764–2777 . <https://doi.org/10.1093/molbev/msz184>

- Fan Z-Y, He X-H, Fan Y, Yu H-X, Wang Y-H, Xie X-J, Liu Y, Mo X, Wang J-Y, Luo C (2020) Isolation and functional characterization of three MiFTs genes from mango. *Plant Physiol Biochem* 155:169–176 . <https://doi.org/10.1016/j.plaphy.2020.07.009>
- Fiser A, Šali A (2003) Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In: *Methods in enzymology*. pp 461–491
- Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. *New Phytol.* 183:557–564
- Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW (2019) EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol* 9:3891–3898 . <https://doi.org/10.1002/ece3.5015>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652 . <https://doi.org/10.1038/nbt.1883>
- Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320:369–87 . [https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4)
- Hanzawa Y, Money T, Bradley D (2005) A single amino acid converts a repressor to an activator of flowering. *Proc Natl Acad Sci* 102:7748–7753 . <https://doi.org/10.1073/pnas.0500932102>
- Hedman H, Källman T, Lagercrantz U (2009) Early evolution of the MFT-like gene family in plants. *Plant Mol Biol* 70:359–369 . <https://doi.org/10.1007/s11103-009-9478-x>
- Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. *J Res Natl Bur Stand (1934)* 49:409 . <https://doi.org/10.6028/jres.049.044>
- Hiraoka K, Yamaguchi A, Abe M, Araki T (2013) The Florigen Genes FT and TSF Modulate Lateral Shoot Outgrowth in *Arabidopsis thaliana*. *Plant Cell Physiol* 54:352–368 . <https://doi.org/10.1093/pcp/pcs168>
- Ho WWH, Weigel D (2014) Structural Features Determining Flower-Promoting Activity of *Arabidopsis* FLOWERING LOCUS T. *Plant Cell* 26:552–564 . <https://doi.org/10.1105/tpc.113.115220>
- Jin S, Nasim Z, Susila H, Ahn JH (2020) Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin. Cell Dev. Biol.*
- Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* 118:11225–11236 . <https://doi.org/10.1021/ja9621760>
- 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 (80-)* 286:1962–1965 . <https://doi.org/10.1126/science.286.5446.1962>
- Karlgrén A, Gyllenstrand N, Källman T, Sundström JF, Moore D, Lascoux M, Lagercrantz U (2011) Evolution of the PEBP gene family in plants: Functional diversification in seed plant evolution. *Plant Physiol* 156:1967–1977 . <https://doi.org/10.1104/pp.111.176206>
- Kim SY, Yu X, Michaels SD (2008) Regulation of CONSTANS and FLOWERING LOCUS T Expression in Response to Changing Light Quality. *Plant Physiol* 148:269–279 . <https://doi.org/10.1104/pp.108.122606>
- Kinoshita T, Ono N, Hayashi Y, Morimoto S, Nakamura S, Soda M, Kato Y, Ohnishi M, Nakano T, Inoue SI, Shimazaki KI (2011) FLOWERING LOCUS T regulates stomatal opening. *Curr Biol* 21:1232–1238 . <https://doi.org/10.1016/j.cub.2011.06.025>
- Klinterås 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 Phytol* 196:1260–1273 . <https://doi.org/10.1111/j.1469-8137.2012.04332.x>
- Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. *Science* 286:1960–2 . <https://doi.org/10583960>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7. 0 for Bigger Datasets Brief communication. 33:1870–1874 . <https://doi.org/10.1093/molbev/msw054>
- Laskowski R a., MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291 . <https://doi.org/10.1107/S0021889892009944>
- Lee R, Baldwin S, Kenel F, McCallum J, Macknight R (2013) FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nat Commun* 4:2884 . <https://doi.org/10.1038/ncomms3884>
- Lifschitz E, Eviatar T, Rozman A, Shalit A, Goldshmidt A, Amsellem Z, Alvarez JP, Eshed Y (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc Natl Acad Sci* 103:6398–6403 .

<https://doi.org/10.1073/pnas.0601620103>

Liu Y-YY, Yang K-ZZ, Wei X-XX, Wang X-QQ (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 Phytol* 212:730–744 . <https://doi.org/10.1111/nph.14066>

Mackenzie KK, Coelho L, Lütken H, Müller R (2019) Phylogenomic Analysis of the PEBP Gene Family from *Kalanchoë*. 1–16

Miller MA, Pfeiffer W, Schwartz T (2010) 2010 Gateway Computing Environments Workshop, GCE 2010. 2010 Gatew Comput Environ Work GCE 2010

Moumeni A, Satoh K, Kondoh H, Asano T, Hosaka A, Venuprasad R, Serraj R, Kumar A, Leung H, Kikuchi S (2011) Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC Plant Biol* 11:174 .

<https://doi.org/10.1186/1471-2229-11-174>

Nakamura Y, Andrés F, Kanehara K, Liu YC, Dörmann P, Coupland G (2014) Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nat Commun* 5:4–8 . <https://doi.org/10.1038/ncomms4553>

Nakamura Y, Lin YC, Watanabe S, Liu Y chi, Katsuyama K, Kanehara K, Inaba K (2019) High-Resolution Crystal Structure of Arabidopsis FLOWERING LOCUS T Illuminates Its Phospholipid-Binding Site in Flowering. *iScience* 21:577–586 . <https://doi.org/10.1016/j.isci.2019.10.045>

Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* 478:119–122 . <https://doi.org/10.1038/nature10431>

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–26

Neves Cruz J, da Costa KS, de Carvalho TAA, de Alencar NAN (2019) Measuring the structural impact of mutations on cytochrome P450 21A2, the major steroid 21-hydroxylase related to congenital adrenal hyperplasia. *J Biomol Struct Dyn* 1–10 . <https://doi.org/10.1080/07391102.2019.1607560>

Ogiso-Tanaka E, Matsubara K, Yamamoto S, Nonoue Y, Wu J, Fujisawa H, Ishikubo H, Tanaka T, Ando T, Matsumoto T, Yano M (2013) Natural Variation of the RICE FLOWERING LOCUS T 1 Contributes to Flowering Time Divergence in Rice. *PLoS One* 8:e75959 . <https://doi.org/10.1371/journal.pone.0075959>

Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JLL, Nilsson O (2010) An Antagonistic Pair of FT Homologs Mediates the Control of Flowering Time in Sugar Beet. *Science* (80-) 330:1397–1400 . <https://doi.org/10.1126/science.1197004>

Pin PA, Nilsson O (2012) The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant, Cell Environ* 35:1742–1755 . <https://doi.org/10.1111/j.1365-3040.2012.02558.x>

Rallapalli PM, Orengo CA, Studer RA, Perkins SJ (2014) Positive selection during the evolution of the blood coagulation factors in the context of their disease-causing mutations. *Mol Biol Evol* 31:3040–3056 . <https://doi.org/10.1093/molbev/msu248>

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 67:901–904 . <https://doi.org/10.1093/sysbio/syy032>

Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 3:198–210 . <https://doi.org/10.1002/wcms.1121>

Sharir-Ivry A, Xia Y (2018) Nature of long-range evolutionary constraint in enzymes: Insights from comparison to pseudoenzymes with similar structures. *Mol Biol Evol* 35:2597–2606 . <https://doi.org/10.1093/molbev/msy177>

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. *Curr Opin Genet Dev* 35:119–125 . <https://doi.org/10.1016/j.gde.2015.11.003>

Specht CD (2006) Systematics and Evolution of the Tropical Monocot Family Costaceae (Zingiberales): A Multiple Dataset Approach. *Syst Bot* 31:89–106 . <https://doi.org/10.1600/036364406775971840>

Studer RA, Christin PA, Williams MA, Orengo CA (2014) Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A* 111:2223–2228 . <https://doi.org/10.1073/pnas.1310811111>

Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31:926–932 . <https://doi.org/10.1093/bioinformatics/btu739>

Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19:596–604 . <https://doi.org/10.1016/j.sbi.2009.08.003>

Wang H, Ni X, Harris-Shultz K (2019) Molecular evolution of the plant ECERIFERUM1 and ECERIFERUM3 genes involved in aliphatic hydrocarbon production. *Comput Biol Chem* 80:1–9 . <https://doi.org/10.1016/j.compbiolchem.2019.02.009>

Wang Z, Yang R, Devisetty UK, Maloof JN, Zuo Y, Li J, Shen Y, Zhao J, Bao M, Ning G (2017) The Divergence of Flowering Time Modulated by FT/TFL1 Is Independent to Their Interaction and Binding Activities. *Front Plant Sci* 8:1–16 . <https://doi.org/10.3389/fpls.2017.00697>

- Weadick CJ, Chang BSW (2012) An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol Biol Evol* 29:1297–1300 . <https://doi.org/10.1093/molbev/msr311>
- Wiberg KB (1965) A Scheme for Strain Energy Minimization. Application to the Cycloalkanes 1. *J Am Chem Soc* 87:1070–1078 . <https://doi.org/10.1021/ja01083a024>
- Wickland DP, Hanzawa Y (2015) The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. *Mol Plant* 8:983–997 . <https://doi.org/10.1016/j.molp.2015.01.007>
- Wigge PA (2005) Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. *Science* (80-) 309:1056–1059 . <https://doi.org/10.1126/science.1114358>
- 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*. *Plant Cell* 22:1733–1748 . <https://doi.org/10.1105/tpc.109.073072>
- Xu D, Zhang Y (2011) Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophys J* 101:2525–2534 . <https://doi.org/10.1016/j.bpj.2011.10.024>
- 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 Cell Physiol* 46:1175–1189 . <https://doi.org/10.1093/pcp/pci151>
- 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. *Plant J* 63:241–253 . <https://doi.org/10.1111/j.1365-3113X.2010.04234.x>
- Yu X, Liu H, Sang N, Li Y, Zhang T, Sun J, Huang X (2019) Identification of cotton *MOTHER OF FT AND TFL1* homologs, GhMFT1 and GhMFT2, involved in seed germination. *PLoS One* 14:e0215771 . <https://doi.org/10.1371/journal.pone.0215771>
- Zhang X, An L, Nguyen TH, Liang H, Wang R, Liu X, Li T, Qi Y, Yu F (2015) The Cloning and Functional Characterization of Peach *CONSTANS* and *FLOWERING LOCUS T* Homologous Genes PpCO and PpFT. *PLoS One* 10:e0124108 . <https://doi.org/10.1371/journal.pone.0124108>
- 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 (2016) Evolution of the PEBP gene family and selective signature on FT -like clade. *J Syst Evol* 54:502–510 . <https://doi.org/10.1111/jse.12199>
- Zhu Y, Klasfeld S, Jeong CW, Jin R, Goto K, Yamaguchi N, Wagner D (2020) *TERMINAL FLOWER 1*-FD complex target genes and competition with *FLOWERING LOCUS T*. *Nat Commun* 11:5118 . <https://doi.org/10.1038/s41467-020-18782-1>

Figures

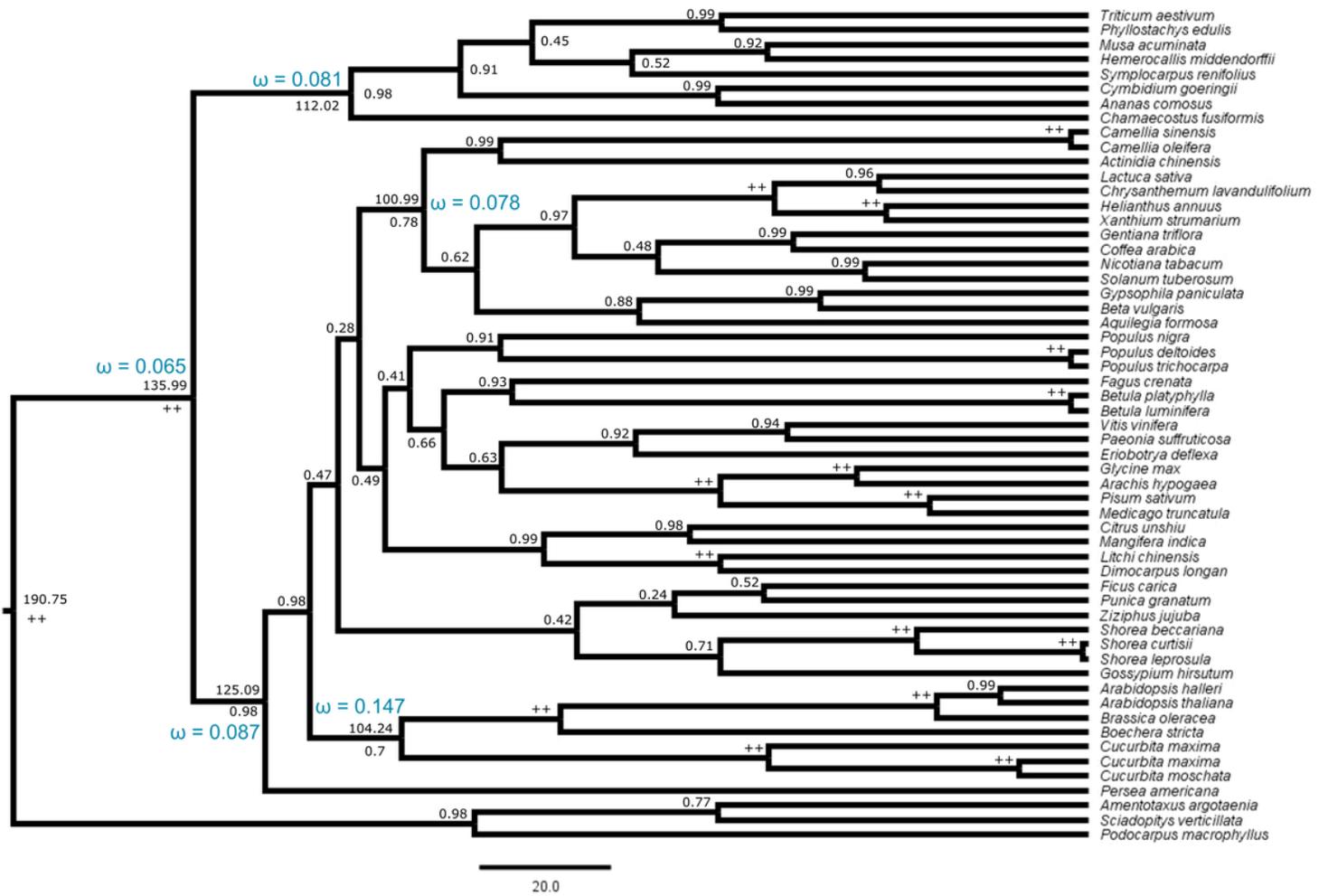


Figure 1

Bayesian phylogeny of FT sequences from angiosperms. Each branch is indicated by the posterior probability with ++ = 1.0 posterior. Molecular dating and dN/dS rates represented by ω value are noted at major clades

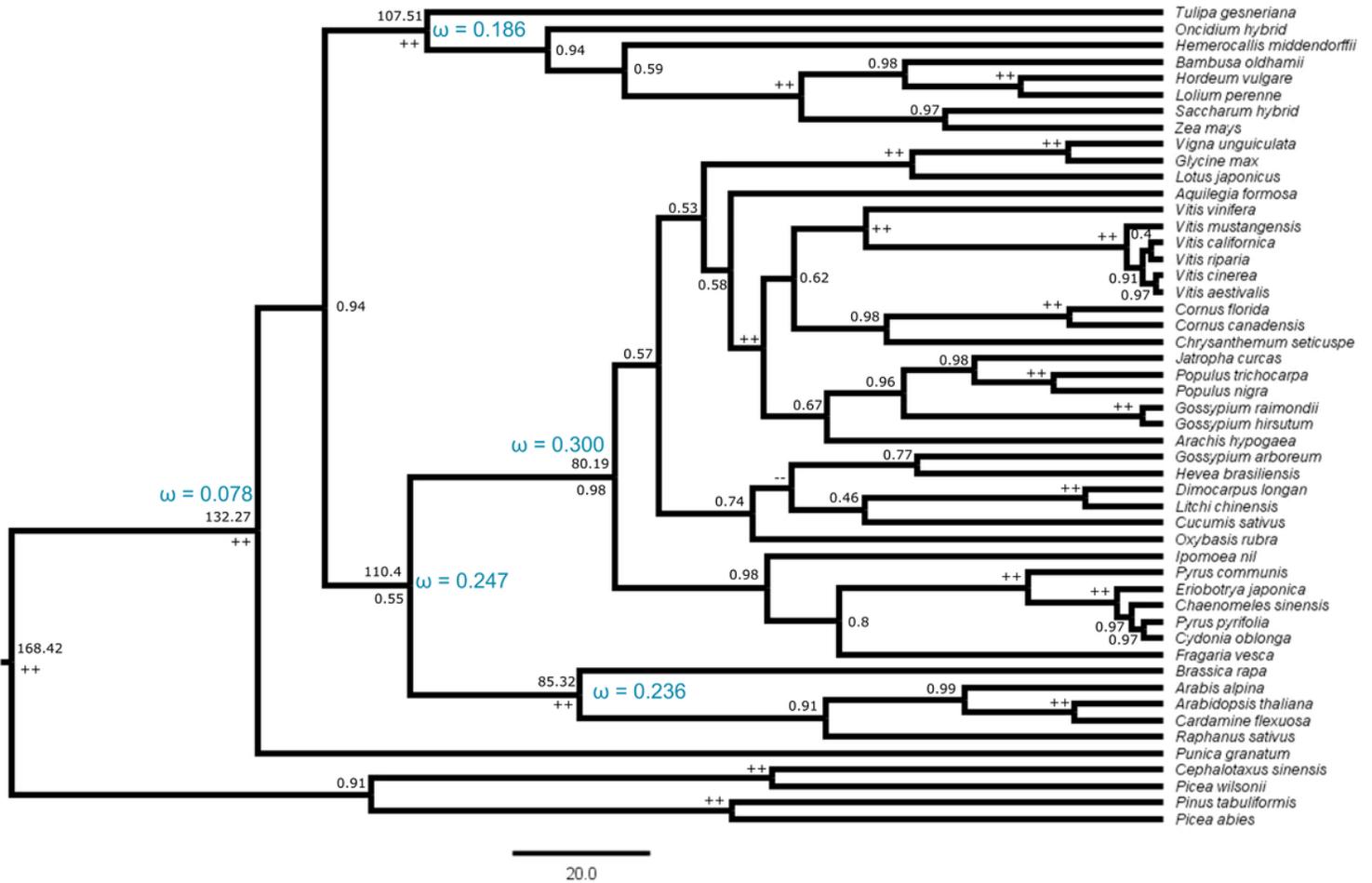


Figure 2
 Bayesian phylogeny of TFL1 sequences from angiosperms. Each branch is indicated by the posterior probability with ++ = 1.0 and - < 0.4 posterior. Molecular dating and dN/dS rates represented by ω value are noted at major clades

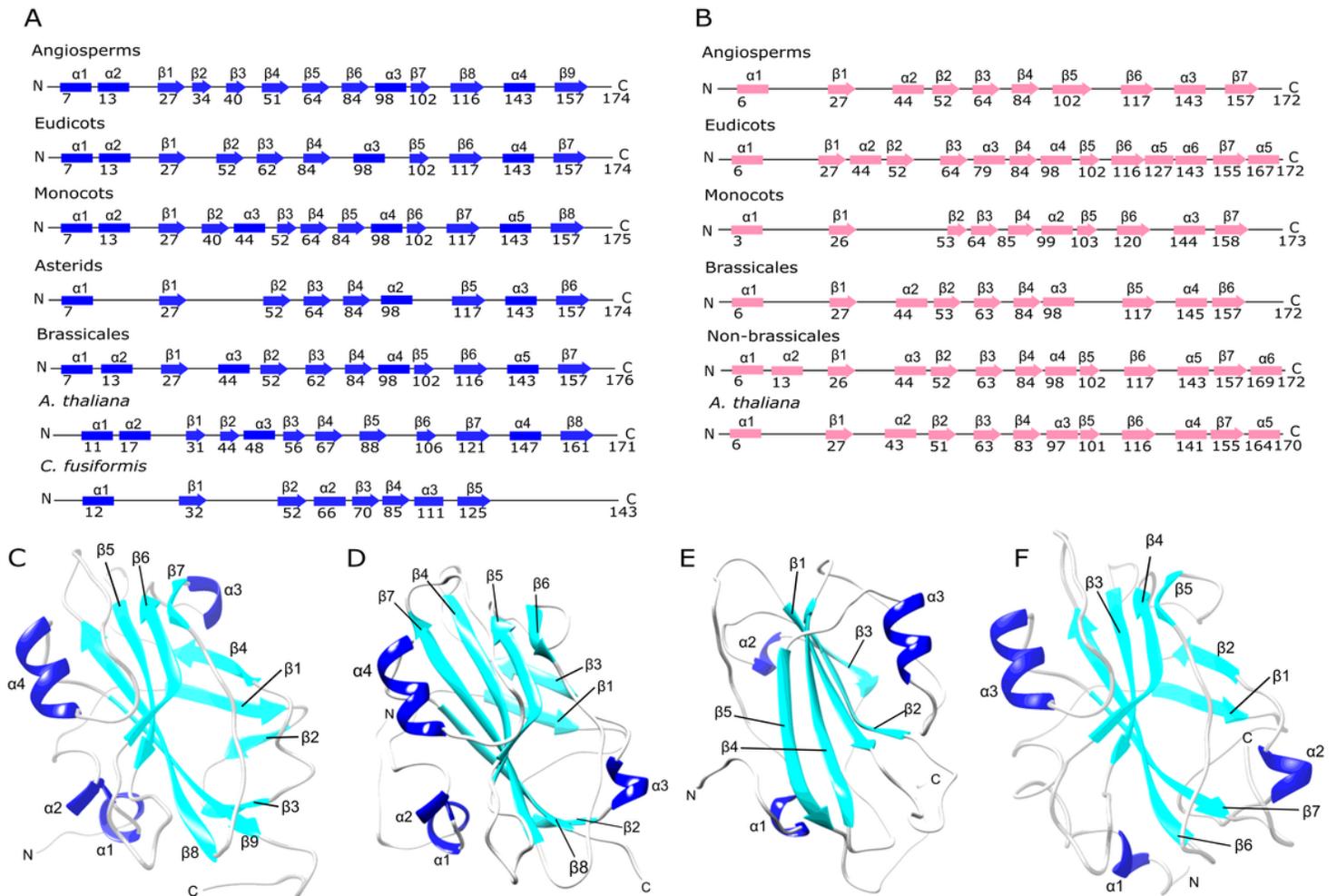


Figure 3
 Protein secondary and tertiary structures of FT and TFL1. (A) Secondary structures of inferred ancestral, *A. thaliana*, and *C. fusiformis* FT. (B) Secondary structures of ancestral and *A. thaliana* TFL1. (C) Tertiary structure of the inferred ancestral of angiosperm clade FT. (D) Tertiary structure of *A. thaliana* FT. (E) Tertiary structure of *C. fusiformis* FT. (F) Tertiary structure of the inferred ancestral of angiosperm clade TFL1

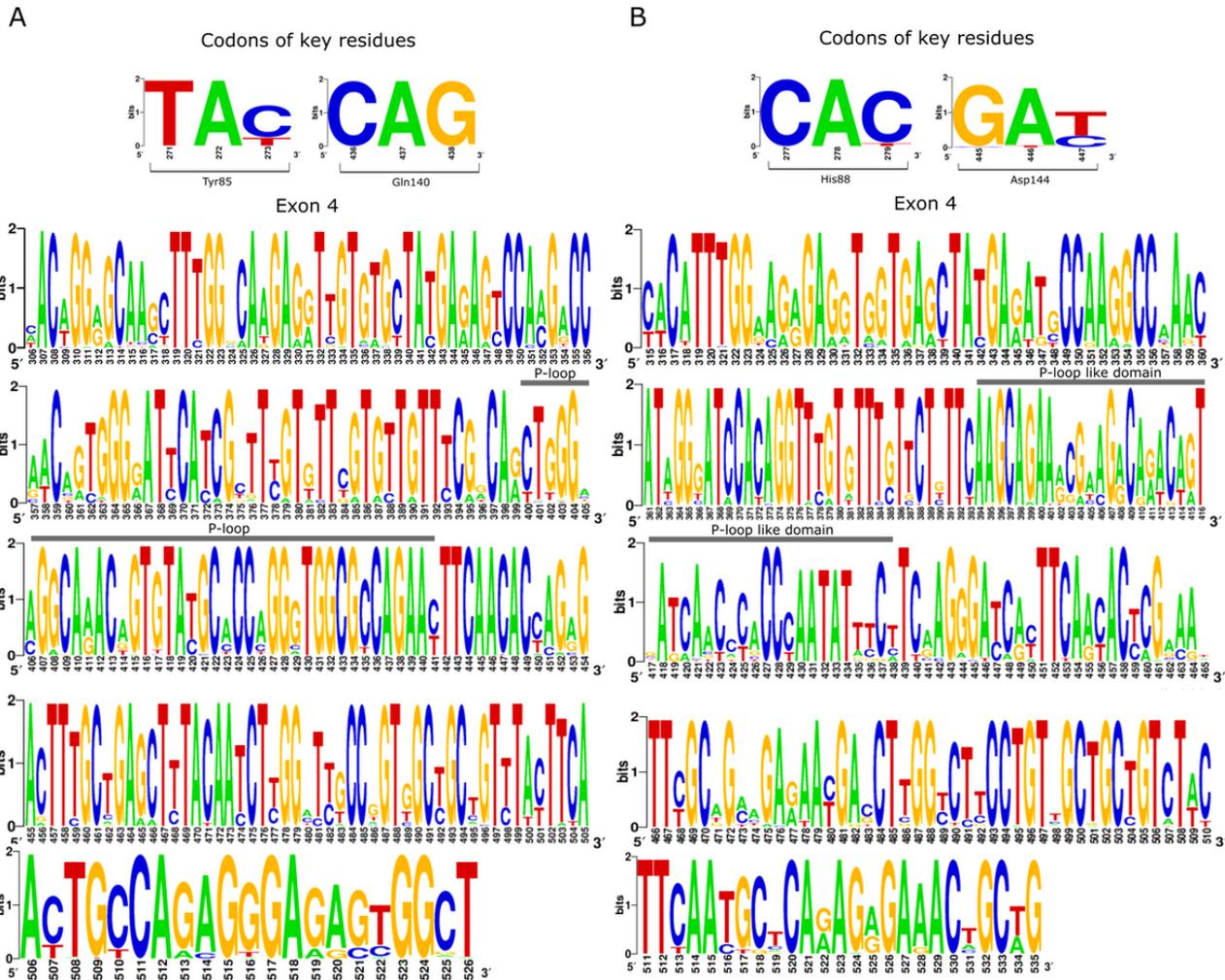


Figure 4

The 4th exon that encodes the P-loop domain in FT structure is well-conserved over the evolutionary history of angiosperms. Schematic overview of ancestral codon sequences of the 4th exon showing the conserved regions of FT (panel A) and TFL1 (panel B). The height of each nucleotide within the pictogram indicates their relative frequency in the analyzed position

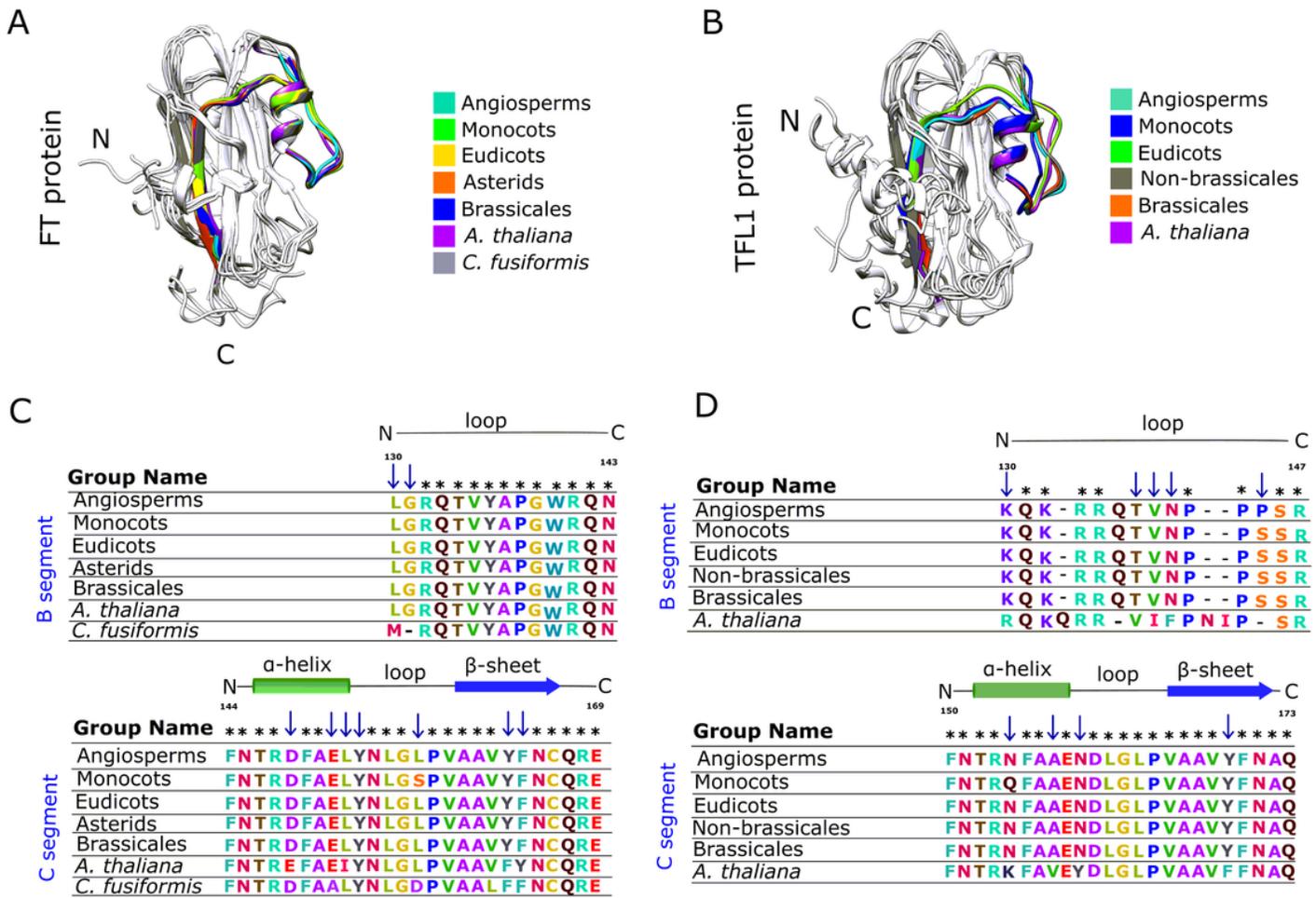


Figure 5

The C segment of FT and TFL1 protein structures are conserved across extant and reconstructed ancestral sequences even with amino acid substitutions. Structural alignment of the C segments from the ancestral FT (panel A) and TFL1 (B) sequences, which are coded by the fourth exon. Sequence alignment of the ancestral FT (C) and TFL1 (D) with their counterpart *A. thaliana* and *C. fusiformis*. Asterisks indicate matches in the alignment sequences, and the blue arrows indicate the residues substitutions

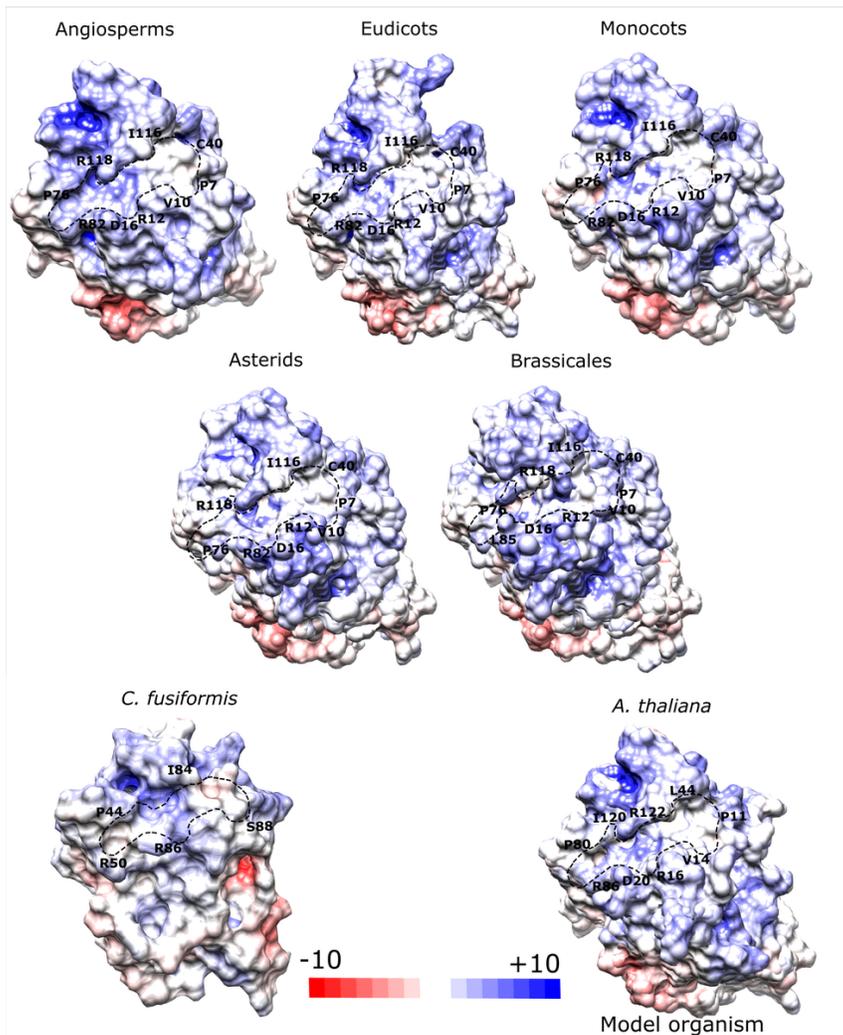


Figure 6
 Electrostatic potential map of phosphatidylcholine (PC) binding site of modeled ancestral FT structures. Some residues located at the PC binding site, such as Pro8, Val11, Arg13, Asp17, Leu41, Pro77, Arg83, Ile117, and Arg119 exhibit a predominantly positive potential. Blue regions indicate the positive potential, white indicates the neutral potential (no charge), and the red regions negative potential. Highlighted residues belong to the binding pocket

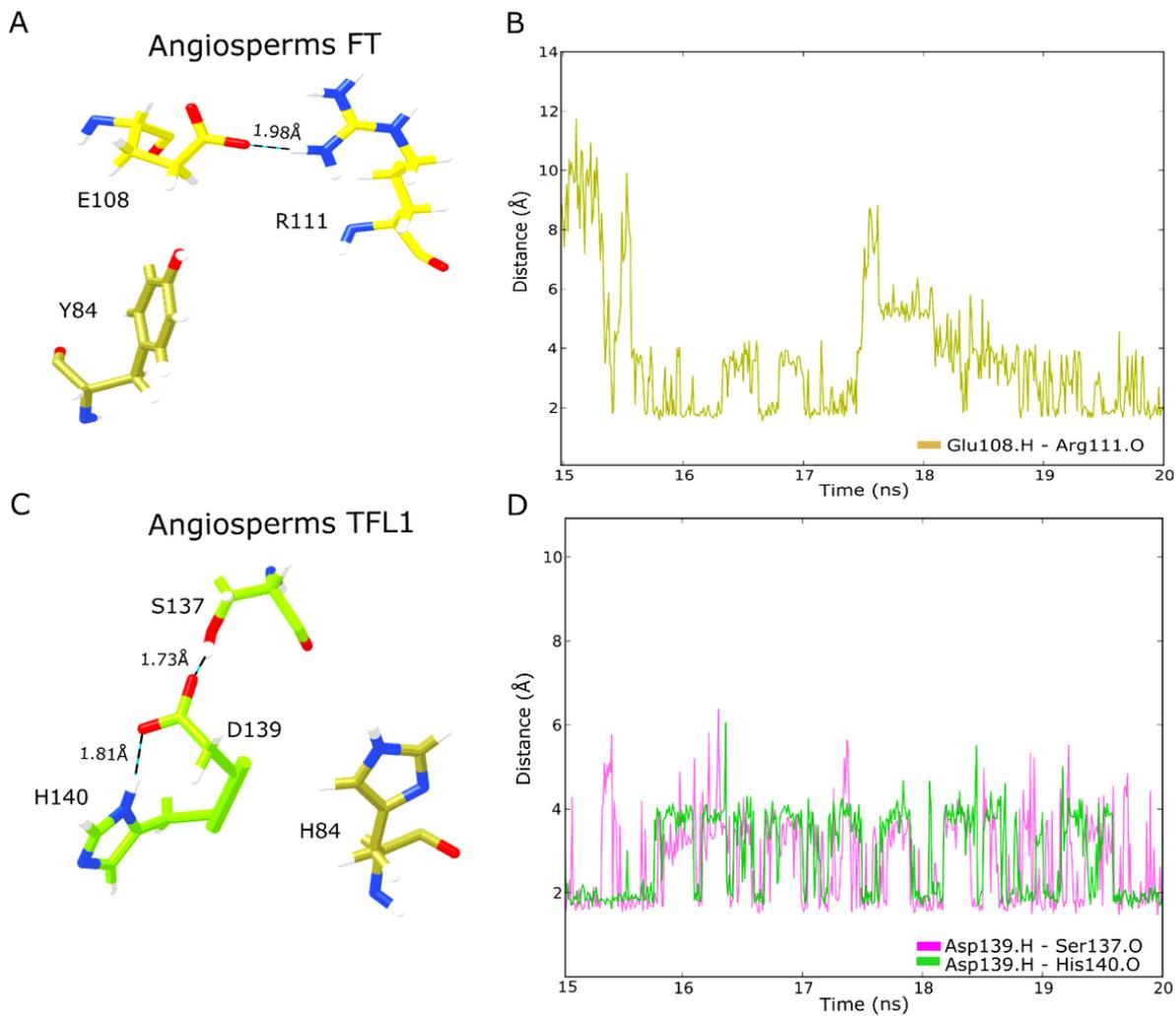


Figure 7
 Analysis of H-bond interactions in the ancestral FT and TFL1 structures from angiosperms through MD simulation. (A) Detailed overview of FT residues involved in H-bond interactions; (B) Interatomic H-bond distances verified over 20ns of MD simulation of FT structure. (C) Detailed overview of TFL1 residues involved in H-bond interactions; (D) Interatomic H-bond distances verified over 20ns of MD simulation of TFL1 structure

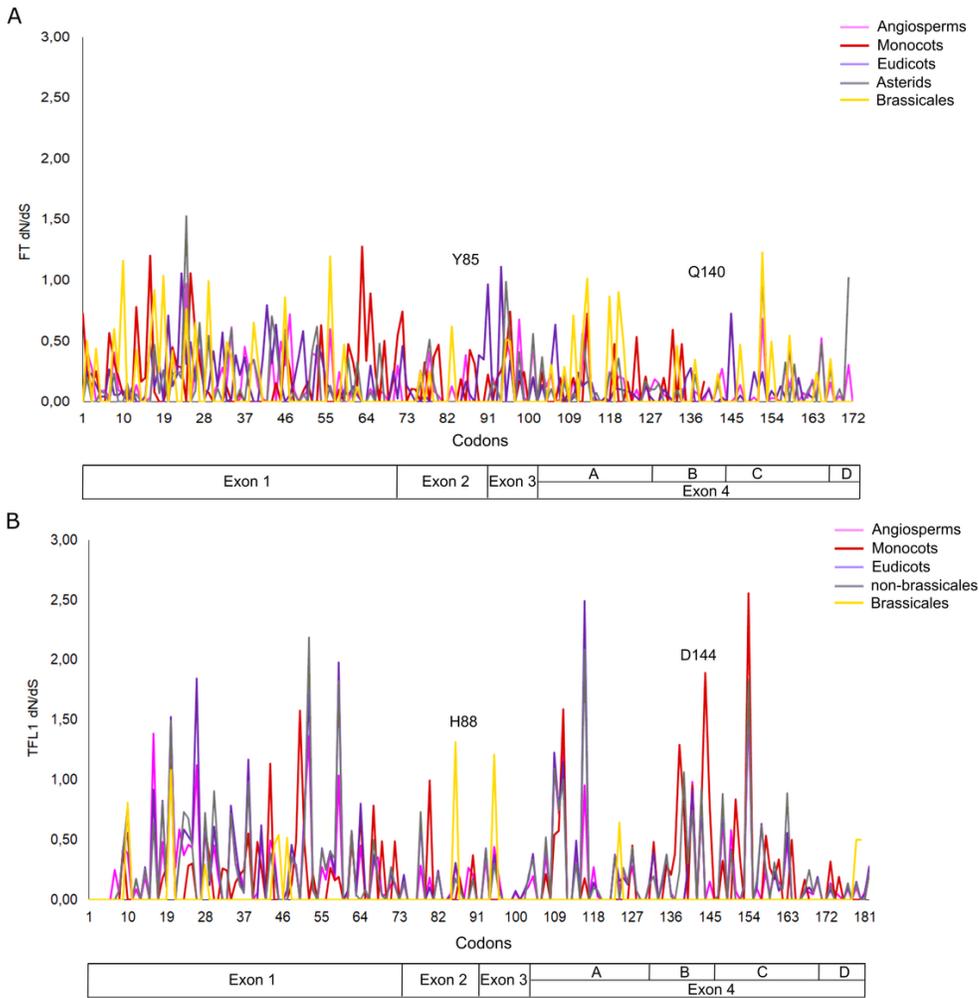


Figure 8

Analysis of the selective pressures indicated by dN/dS rates of the codons for FT (panel A) and TFL1 genes (panel B)

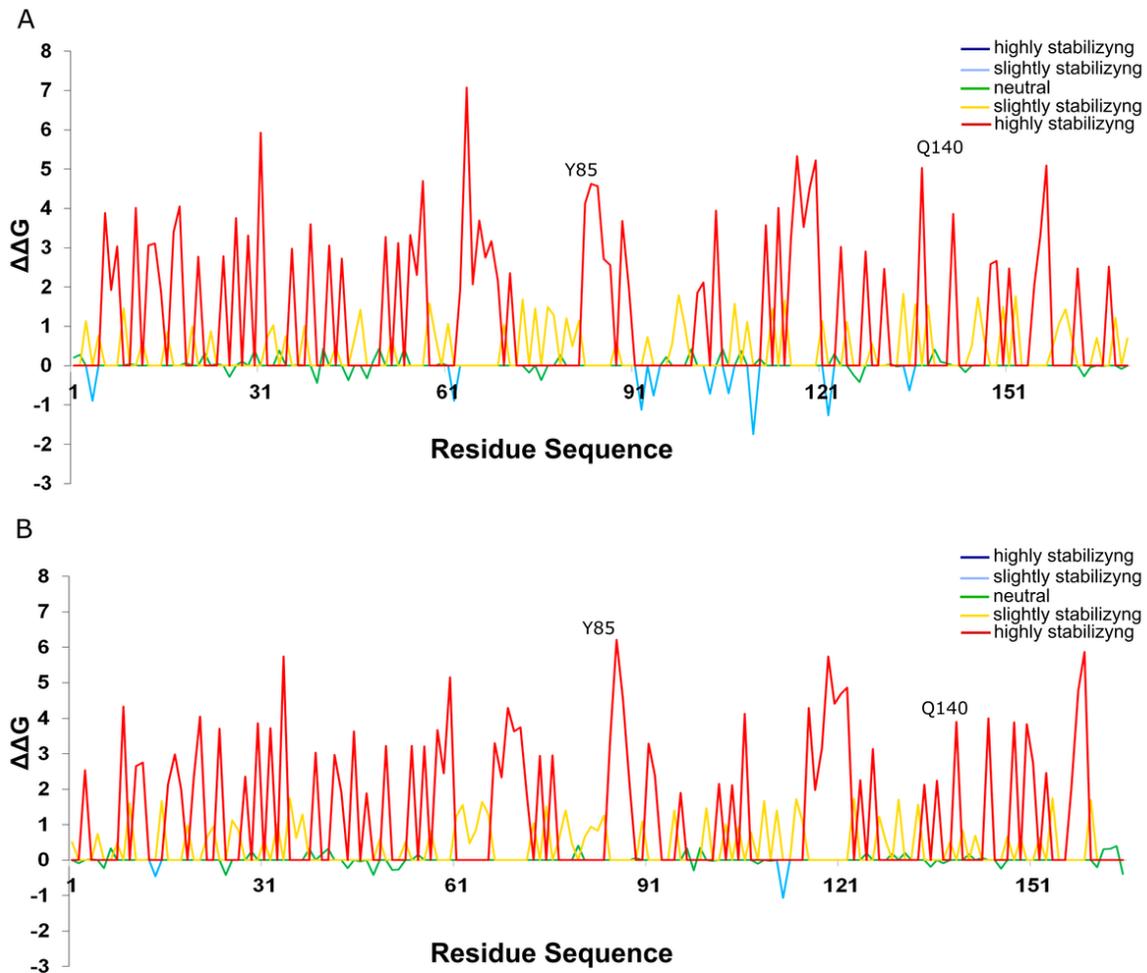


Figure 9

Most alanine substitutions in FT structure are destabilizing, especially at the residue Tyr85. Alanine scanning plot of FT structures showing the effect of mutations on protein stability. (A) Inferred ancestral protein sequence of angiosperms. (B) Protein sequence of *A. thaliana*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicalabstract.png](#)
- [Supportinginformation.docx](#)