



OPEN ACCESS

EDITED BY

Rajeev K. Varshney,
International Crops Research Institute
for the Semi-Arid Tropics (ICRISAT),
India

REVIEWED BY

Heng Sun,
Wuhan Botanical Garden, Chinese
Academy of Sciences (CAS), China
Ze Peng,
South China Agricultural University,
China
Hui Song,
Qingdao Agricultural University, China

*CORRESPONDENCE

Kauê Santana,
kaue.costa@ufopa.edu.br
Thiago André,
thiagojandre@gmail.com

[†]These authors have contributed equally
to this work.

SPECIALTY SECTION

This article was submitted to Plant
Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 26 May 2022

ACCEPTED 29 August 2022

PUBLISHED 29 September 2022

CITATION

Almeida de Jesus D, Batista DM,
Monteiro EF, Salzman S, Carvalho LM,
Santana K and André T (2022), Structural
changes and adaptative evolutionary
constraints in FLOWERING LOCUS T
and TERMINAL FLOWER1-like genes of
flowering plants.
Front. Genet. 13:954015.
doi: 10.3389/fgene.2022.954015

COPYRIGHT

© 2022 Almeida de Jesus, Batista,
Monteiro, Salzman, Carvalho, Santana
and André. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Structural changes and adaptative evolutionary constraints in FLOWERING LOCUS T and TERMINAL FLOWER1-like genes of flowering plants

Deivid Almeida de Jesus ^{1†}, Darlisson Mesquista Batista ^{2†},
Elton Figueira Monteiro ², Shayla Salzman ³,
Lucas Miguel Carvalho ⁴, Kauê Santana ^{5*} and
Thiago André ^{6*}

¹Institute of Biology Genetics Graduate Program, Federal University of Rio de Janeiro Rio de Janeiro, Rio de Janeiro, Brazil, ²Programa de Pós-Graduação em Biodiversidade, Universidade Federal do Oeste do Pará Santarém, Pará, Santarém, Brazil, ³School of Integrative Plant Sciences. Section of Plant Biology. Cornell University Ithaca, New York, NY, United States, ⁴Center for Computing in Engineering and Sciences, State University of Campinas. Campinas, São Paulo, Brazil, ⁵Institute of Biodiversity, Federal University of Western Pará Santarém Pará, Santarém, Brazil, ⁶Botany Department, University of Brasília, Brasília, Brazil

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 through *in silico* modeling to identify determinant sites that evolved in both proteins and allowed the adaptative diversification in the flowering phenology and developmental processes. In addition, we demonstrate that the occurrence of destabilizing mutations in residues located at the phosphatidylcholine binding sites of FT structure are under 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.

KEYWORDS

phenology, structural Biology, phylogenetics, protein structure prediction, adaptative evolutionary constraints, natural selection

Article highlights

- FT and TFL1 have mostly been evolving under purifying selection as evidenced by substitutions in the third position of the codons that encode key residues involved with FT/TFL1 function do not alter the encoded amino acids.
- Residues from the P-loop domain of the analyzed FT structures show predominantly high destabilizing mutations which is consistent with constant selective pressure found for this region.
- Protein conformation more than sequence appears to be under strong selective pressure in that amino acid substitutions that would have resulted in structural changes in over evolutionary time show matching and stabilizing mutations.
- The appearance of destabilizing mutations in residues of phosphatidylcholine binding sites are under negative selection, while some residues of the 4th exon are under positive selection, these structural changes are 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 indicate their structural role in maintaining the overall stability of FT structure.

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 stimulus (Ahn et al., 2006a; Pin and Nilsson, 2012) 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 are 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* (*API*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) genes, leading to flower development (Wigge, 2005; Collani et al., 2019; Zhu et al., 2020). Contrasting FT activity, TFL1 proteins are known to repress flowering by inhibiting the 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, 2005a; 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., 2005a), *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, 2005b). 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 (Chardon and Damerval, 2005; Carmona et al., 2007; Zheng et al., 2016). 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., 2011a). 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 adaptative 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 the occurrence of duplication events in the evolutionary history of angiosperms (Hedman et al., 2009; Karlgrén et al., 2011a; Wickland and Hanzawa, 2015).

The FT and TFL1 of *Arabidopsis thaliana* exhibit conserved structures (Ahn et al., 2006a; Pin and Nilsson, 2012) 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 named the P-loop (*A. thaliana* position 139–152), confers an antagonistic activity to the floral regulators (Ahn et al., 2006a; Pin and Nilsson, 2012). 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*. Our results shed light

on the role of natural selection in the adaptive evolution of flowering proteins in angiosperms and can help in the improvement of crops with economic interest and for the flower and fruit industries.

Materials and methods

Phylogenetic analysis and ancestral sequence inference

First, we performed a sequence alignment search to find *FT/TFL1*-like homologs (Supplementary Information S1) using as a start point the coding DNA sequences (CDSs) of *FT* and *TFL1* from *Arabidopsis thaliana* (GeneBank accession codes: *FT*: NM_105222.3; *TFL1*: NM_120465.3). The similarity sequence search was performed through the BLASTn tool using the GenBank database (Benson et al., 2017). Sequences with the highest identities, percent similar identities above 82%, were included in further analysis. In addition, sequences from different species were included in our analysis based on previous findings (Ahn et al., 2006b; Ho and Weigel, 2014).

To infer gene phylogenetic trees, we used a total of 103 coding sequences of *FT* (100 sequences from angiosperms, belonging to monocots, eudicots, asterids, and brassicales groups; and 3 sequences from gymnosperms); and 82 sequences of *TFL1* (78 sequences from angiosperms belonging to monocots and eudicots; and 4 sequences from gymnosperms). In both phylogenetic inferences, gymnosperm sequences of *FT* and *TFL1* were used as outgroups (Supplementary Information S1). Multiple sequence alignments were performed in MUSCLE (Edgar, 2004) as implemented in MEGA 7 (Kumar et al., 2016). These alignments were further manually checked and edited, mainly to maintain reading frames. The jModel Test 2 program was used to select the nucleotide model with the best Bayesian information criterion (BIC) score (Darrriba et al., 2012). We used MEGA7 (Kumar et al., 2016) for ancestral sequence reconstruction (ASR) using the Maximum Likelihood method to estimate the ancestral state of each node in a phylogeny where the state is chosen to be the one that maximizes the probability of the sequence data based on the selected evolutionary model of nucleotide or amino acid substitutions. 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. Seventeen fossil-based times to the most recent common ancestor (mrca) calibrations were used based on published data. Calibration dates and associated citations can be found in Supplementary Information S2. 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).

In silico prediction of FT and TFL1 structures

Putative structures of FT/TFL1 proteins were generated through 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 (20 × 20) matrix. The models of both proteins were optimized at the atomic level using the random parameter in the range [0,400] and resistance [0,20] e 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), both using default parameters. 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) using the parameters of the Amber forcefield.

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 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 (Guerois 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 Eq. 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 the 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

Our results shed a light on the evolutionary history and structural importance of *FT/TFL1*-like genes. Using fossil records, we calibrated our phylogenetic trees, providing more accurate indications of the divergence of both genes during the evolution of angiosperms. 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, such as the electrostatic potential map of the phosphatidylcholine site, as well as, the structural implications of mutations at key residues involved with the molecular function of both FT and TFL1 proteins.

Reconstruction of evolutionary history of *FT* and *TFL1* in angiosperms

In the present study, when species presented two or more copies for the same gene (interspecies paralogs) within the *FT* group or in *TFL1*, we used both copies in our analyses as previous studies highlighted the impact of ignoring paralogs in generating bias in the prediction of protein function (Stambouliau et al., 2020). 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 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. 181 million years ago and *TFL1*-like genes ca. 163 million years ago. Within the angiosperm clade, the *FT*-like genes diversified approximately 134 million years ago (Figure 1) and the *TFL1*-like genes around

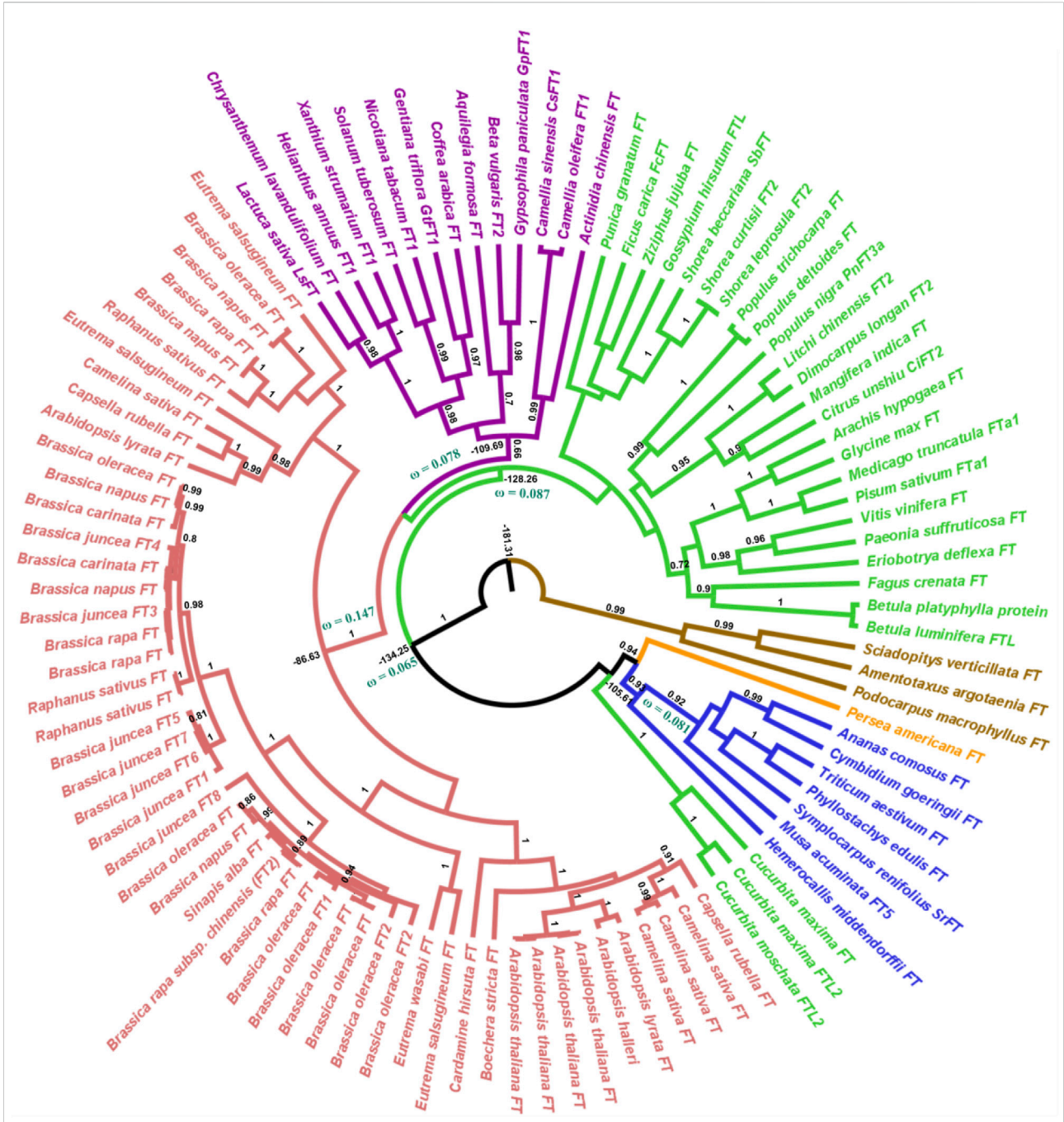


FIGURE 1
 Bayesian phylogeny of angiosperm FT sequences. The major branches are indicated by posterior probability. Molecular dating and the dN/dS ratios represented by the ω value are observed in the major clades. Colors indicate the clades, red = brassicales, purple = asterids, green = eudicots, blue = monocots, orange = Magnoliidae, brown = external group. The brassicales, asterids, and eudicots clades (red, purple, and green) form the ancestor of eudicots.

131 million years ago (Figure 2). (Klintenäs et al., 2012a) suggested that *FT* is found exclusively in flowering plants. Here, we used the *FT* and *TFL1* genes from gymnosperms solely to root the phylogenetic tree. The choice of different species was due to the absence of sequences of both genes

from the same species available in public databases. However, it has been evidenced in previous studies that gymnosperms carry one or two copies of the *FT/TFL1* genes, that represent a clade that is sibling or ancestor of the *FT/TFL1* genes found in angiosperms (Karlgrén et al., 2011a; Klintenäs et al., 2012a).

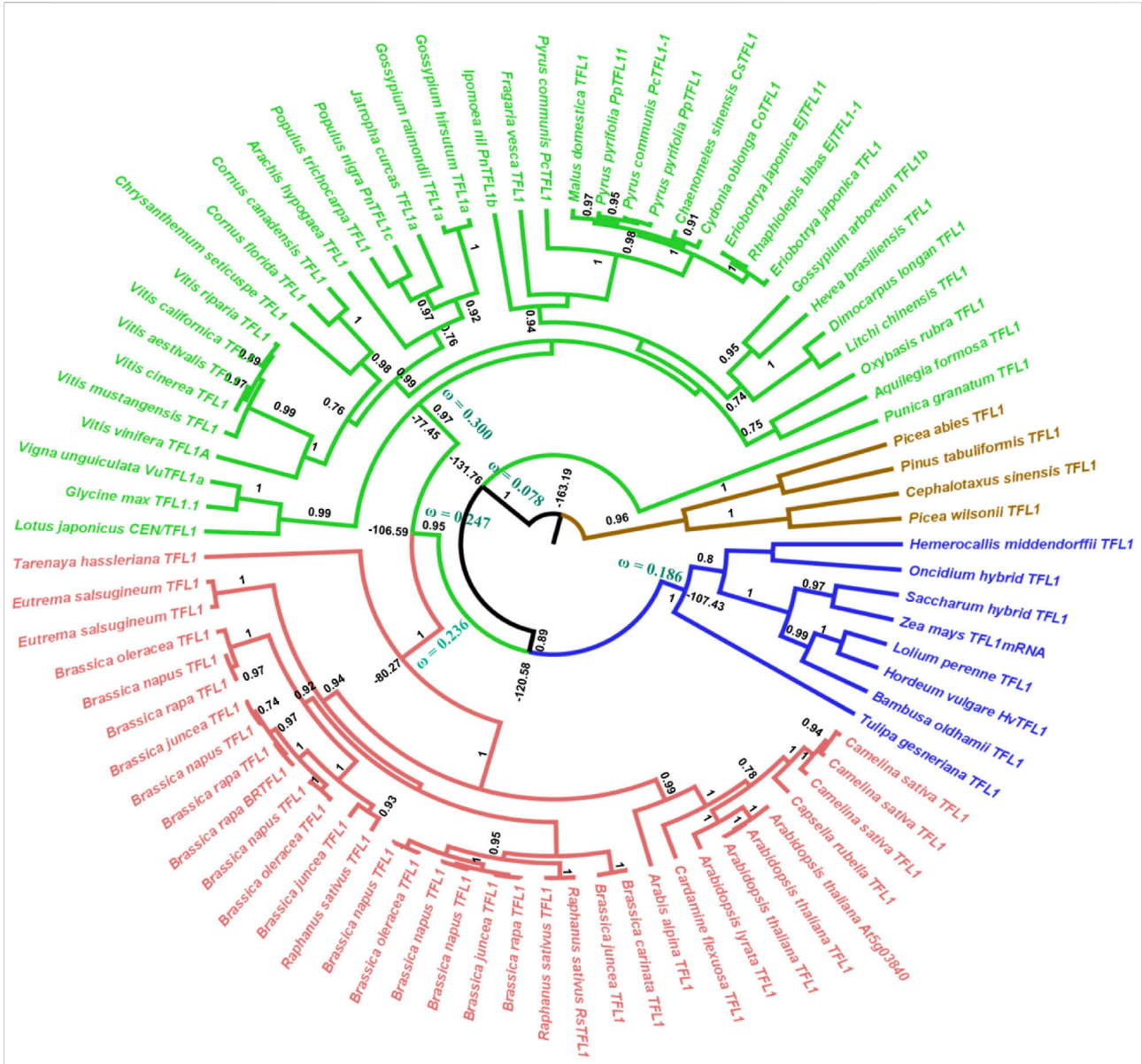


FIGURE 2
 Bayesian phylogeny of angiosperm TFL1 sequences. The major branches are indicated by posterior probability. Molecular dating and the dN/dS ratios represented by the ω value are observed in the major clades. Colors indicate the clades, red = brassicales, green = non-brassicales, blue = monocots, brown = outgroup. The brassicales and non-brassicales clades (red and green) form the ancestor of eudicots.

In the angiosperm clade, we recovered four monophyletic groups of *FT*-like sequences: monocots (posterior probability = 0.93), eudicots (0.90), asterids (0.66), and brassicales (1.0). Regarding the *TFL1* sequences, we also recovered distinct clades for the monocots (1.0), eudicots (0.95), brassicales (1.0), and an unresolved group which we named non-brassicales (0.97) (Figure 1). The estimated time of evolutionary divergence of *FT*-like sequences is approximately 105 million years ago for monocots, 128 for eudicots, 109 for asterids, and 86 for brassicales. Divergence time for *TFL1*

sequences was relatively similar, if not slightly earlier. Monocot *TFL1* diverged approximately 107, eudicots 106, brassicales 80, and the non-brassicales group 77 million years ago (Figure 2). It is important to note that a previous study used genetic distances to analyze the divergences of both genes and the authors reported the presence of *FT*-like genes in gymnosperms (Liu et al., 2016). In contrast, our study used fossil records to estimate the divergences between the main groups of *FT* and *TFL1* genes, which may contribute to more accurate findings regarding the divergence of these genes in plant evolution.

Herein, we showed that both FT and TFL1 genes had similar divergence times.

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., 2005b; Lifschitz et al., 2006; Xi et al., 2010; Pin and Nilsson, 2012). Pleiotropy has a functional diversity regulating different phenotypic characteristics, thus influencing the natural selection pressures leading to the appearance or elimination of new characteristics (Auge et al., 2019).

We also noted that the sequence of *Aquilegia* sp. Remained in the Asterids clade and this result could be related due to the coalescence evolution of this gene in relation to the taxonomic group (Gatesy et al., 2019). Furthermore, it is important to highlight that the phylogenetic relationships of the *FT* and *TFL1* genes from angiosperms and the basal group of angiosperms have been well reported in previous studies (Karlgrén et al., 2011b; Klintonäs et al., 2012a). These studies have demonstrated that the *FT* and *TFL1* genes from these groups do not imply the formation of independent clades which is similar to our reported phylogenetic results. In addition, 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., 2019a).

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*). The ancestral sequence reconstruction used here, refers to the reconstruction of the ancestral state of each node in a phylogenetic tree. The state is chosen to maximize the probability of the sequence data given under the nucleotide evolution likelihood probability model. The result obtained is a consensus nucleotide sequence used as input for molecular modeling and in the natural selection analysis software. Moreover, we analyzed the current representatives of protein structures (*FT* and *TFL1*) in *A. thaliana* (reference structure).

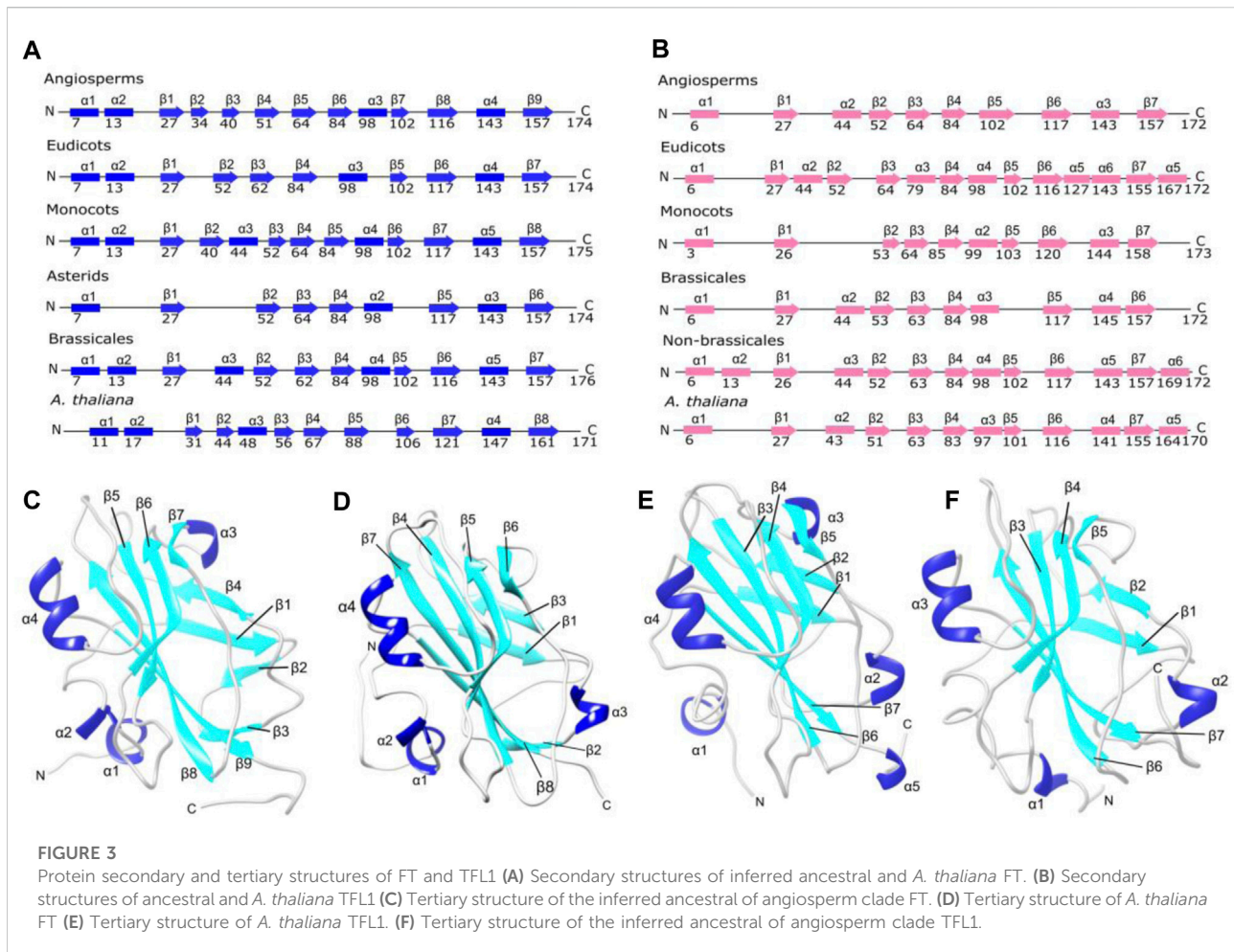
Structurally, *FT* shows seven β -strands and four α -helices, except the ancestral structure of the asterids clade that exhibits three α -helices and six β -strands, *FT* structure from ancestral

angiosperms shows four α -helices and nine β -strands (Figure 3, panel A and C). In contrast, the *TFL1* structures contain predominantly three α -helices and seven β -strands (Figures 3B–F).

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 and 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 Gln140 (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 4B). 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. A possible explanation for this change occurring only in monocots is based on the findings by (Ospina-Zapata et al., 2020) in which monocots show greater recruitment of homologous copies of *FT* that play a role in repressing flowering, as it is attributed to *TFL1*. In contrast, the *TFL1*-like sequences are kept in a few copies, and in some species, they can be lost or maintained with neutral mutations (see section mutational analysis) to allow mutations like the ones presented here. Analysis of the transcriptome of different species of monocots combined with mutagenesis approaches may provide better evidence.

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 (Supplementary Information S3). 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 of the TFL1 domain, even though variations in the frequency of some codons were observed (Figure 4B).

Modeled structures of ancestral FT and TFL1 and *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, Supplementary Information S4, S5), and a satisfactory energetic profile as exhibited by the local quality estimation of Qmean (Supplementary Information S6, S7), which indicate reliable structures. We performed a short MD

simulation to reach the final conformation of the modeled proteins structures. This computational procedure was necessary to correct some stereochemical inconsistencies obtained from the comparative modeling of the FT and TFL1 structures. In addition, the RMSD plots obtained over the MD simulation showed that modeled structures reached a stable conformation after 18ns (Supplementary Information S8).

Previous studies have demonstrated that the segment C located at the C-terminal region of FT/TFL1 structures is crucial for the molecular function of both proteins (Ahn et al., 2006a; Hedman et al., 2009). Based on this assumption, we performed a comparison using the structural alignment of the C-terminal region of the ancestors of FT and TFL1 following the divergence time of the phylogenies (Figures 1, 2). We observed that the ancestral protein of FT in eudicots, which was the first ancestral state to diverge from the main groups of flowering plants, showed high conservation regarding the segment C when compared with its ancestor in angiosperms (Table 1). The ancestor of monocots showed a slightly more divergent RMSD-C α value between the analyzed FT structures. The FT protein of asterids is an ancestor that evolved within the lineage

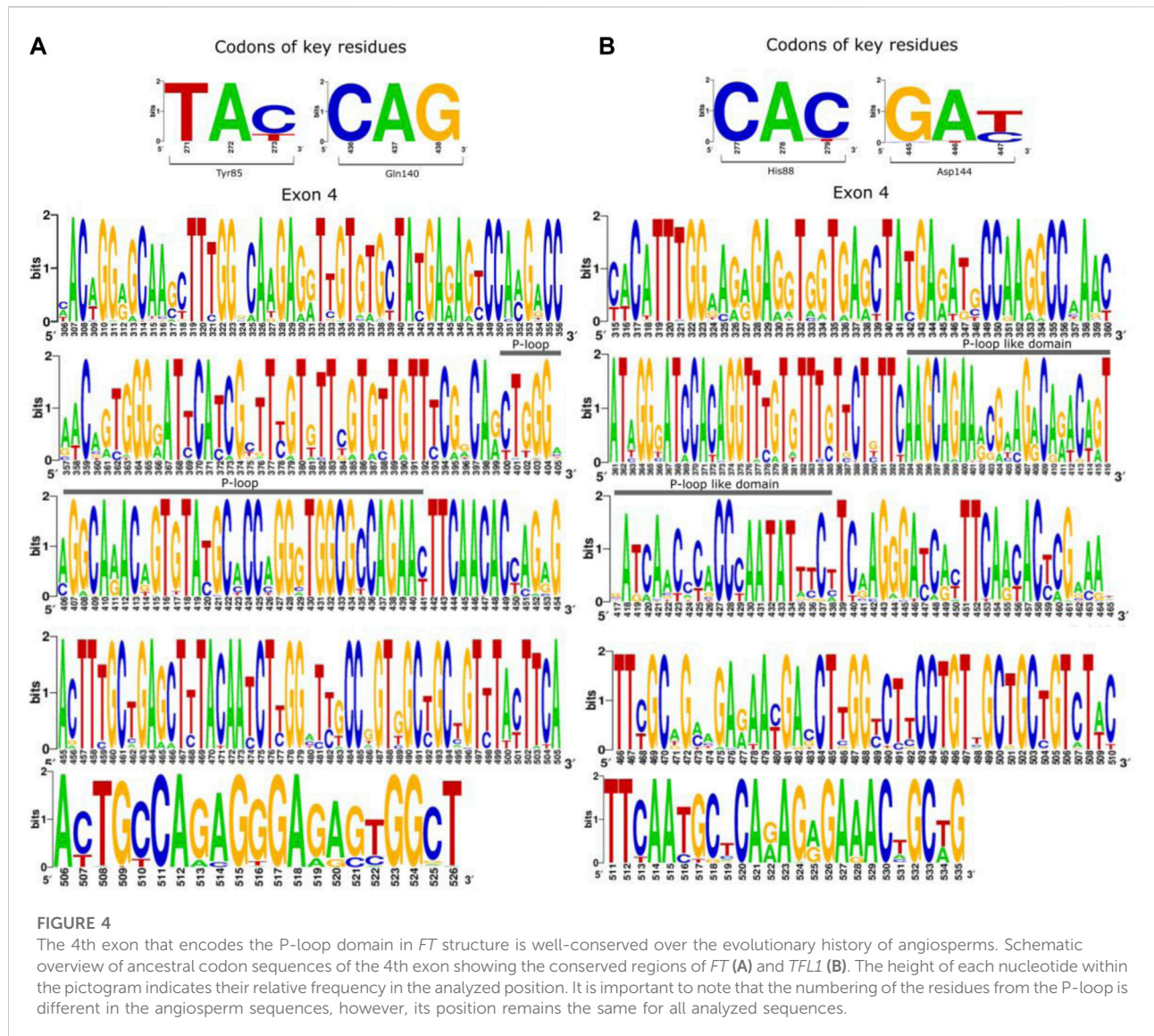


TABLE 1 Structural comparison between segments C of modeled ancestral structures of FT and TFL1. RMSD-C_α values exhibited in angstroms.

FT-like structures		TFL1-like structures	
Taxonomic groups	RMSD-C _α (Å)	Taxonomic groups	RMSD-C _α (Å)
Angiosperms with Eudicots	0.845	Angiosperms with Monocots	1.618
Angiosperms with Monocots	1.153	Angiosperms with Eudicots	1.927
Eudicots with Asterids	0.870	Eudicots with Brassicales	1.513
Eudicots with Brassicales	1.051	Eudicots with non-Brassicales	2.047
Brassicales with <i>A. thaliana</i>	0.869	Eudicots with <i>A. thaliana</i>	1.932

of eudicots and it also showed high conservation for the same segment. In contrast to eudicots ancestral structure, the brassicales structure, showed a higher RMSDs-C_α value when

aligned a eudicots. Finally, the analysis of the FT structure of *A. thaliana*, revealed that the segment C evolved experiencing few mutations which could explain its conserved function over time.

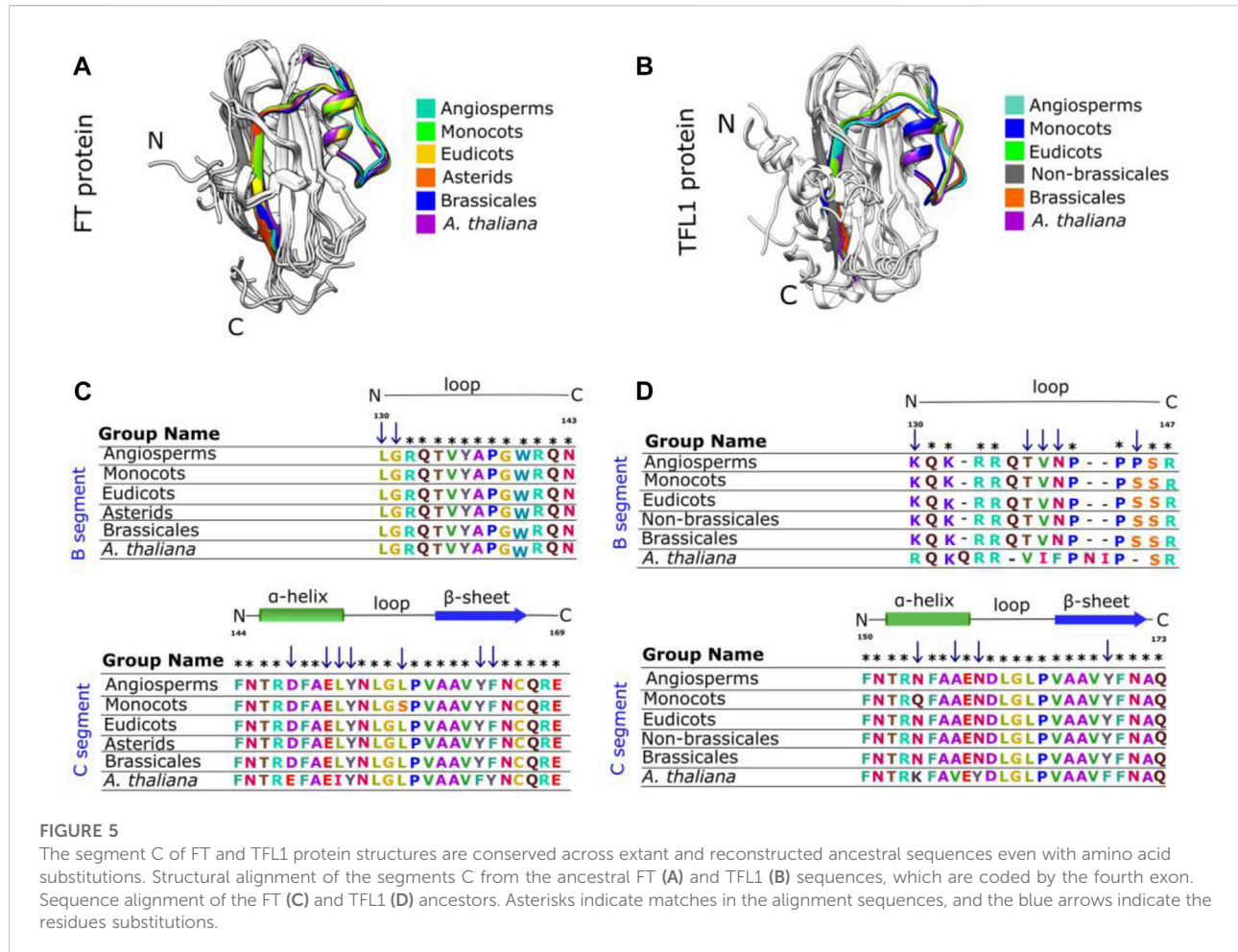


FIGURE 5

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

The structural alignment of FT showed RMSDs-C α values \leq 1,153 Å. In contrast, the TFL1 showed RMSDs-C α values \geq 1,513 Å (Table 1). The segment C of the FT ancestral protein of monocots, when compared to the structure of the ancestral angiosperms, showed to be more conserved than its phylogenetic most evolutionary related group, the eudicots. Comparing the ancestors derived from the lineage of eudicots, brassicales had the lowest RMSDs-C α value and therefore, the most conserved segment C among all the analyzed ancestors, while non-brassicales showed the highest RMSDs-C α value, and thus the higher structural variation, followed by eudicots and *A. thaliana* FT structures, respectively. We hypothesize that both proteins evolved independently with natural selection maintaining the most important regions, such as the segment C. We also observed that the ancestral structures of FT proteins showed to be conserved regarding its folding (Figure 5A), exhibiting an α -helix, and a β -strand in segment C (Ahn et al., 2006a).

The sequence of segment C of FT from *A. thaliana* showed mutations at three residue positions (Asp146Glu, Leu150Ileu, and Phe162Tyr) when compared with the other 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 slight variations in the RMSD-C α values between the ancestral states analyzed over the evolutionary history of angiosperms, which demonstrated that those structures have been conserved during evolution (Table 1). However, it is interesting to note some mutations in four residues in TFL1 sequences of *A. thaliana* at the positions Asn150Lys, Ala153Val, Asn155Tyr, and Tyr165Phe when compared with the ancestral sequences.

Comparing the modeled structures of the ancestral lineages of angiosperms, monocots, eudicots, asterids, and brassicales clades, we noted that these proteins remained well-conserved, showing only one substitution at the residue Ser155 in the ancestral monocot structure when compared with the other ancestral sequences. In contrast, compared with *A. thaliana*, this position contains a leucine substitution (Figure 5C). Comparing the *in silico* modeled TFL1 ancestral structures, we identified a mutation at the position Gln146Asn in the ancestral sequence of monocots (Figure 5D), we further discuss possible explanations of these changes below in the mutation analysis

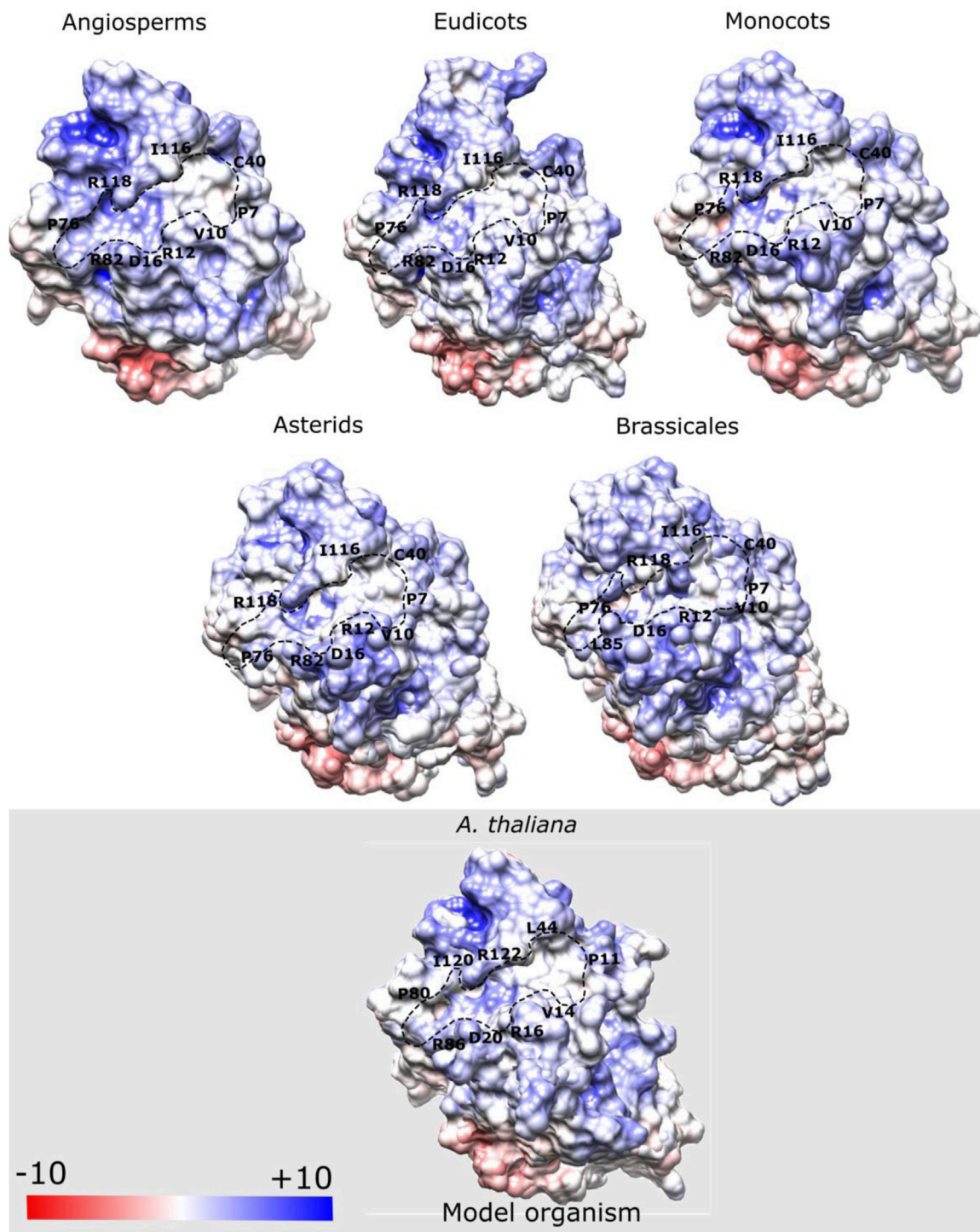
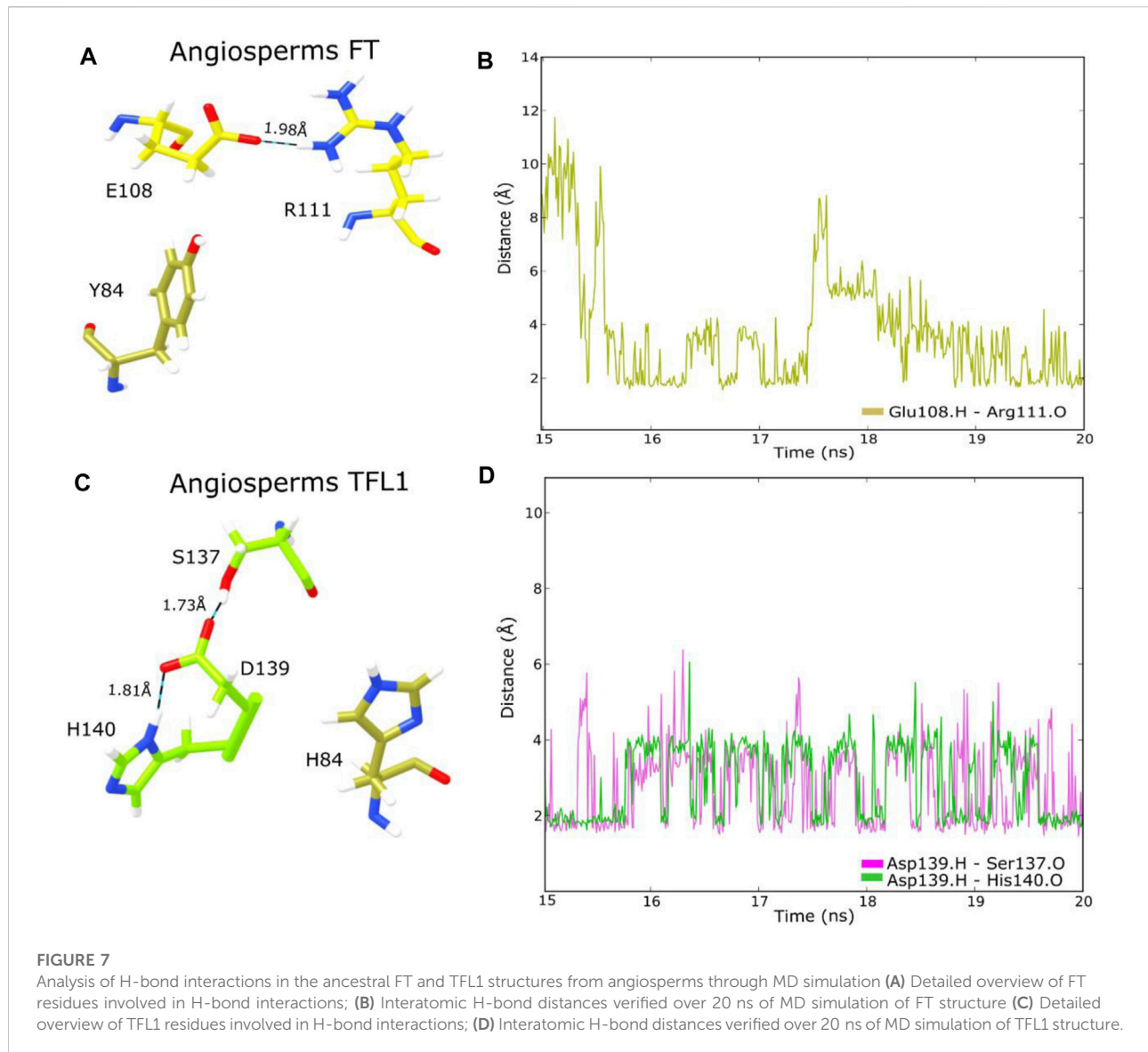


FIGURE 6
 Electrostatic potential map (kBT/e) of phosphatidylcholine (PC) binding site of modeled ancestral FT structures. Some residues located at the PC binding site, such as Pro8, Val11, Arg13, Asp17, Leu41, Pro77, Arg83, Ile117, and Arg119 exhibit a predominantly positive potential. Blue regions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

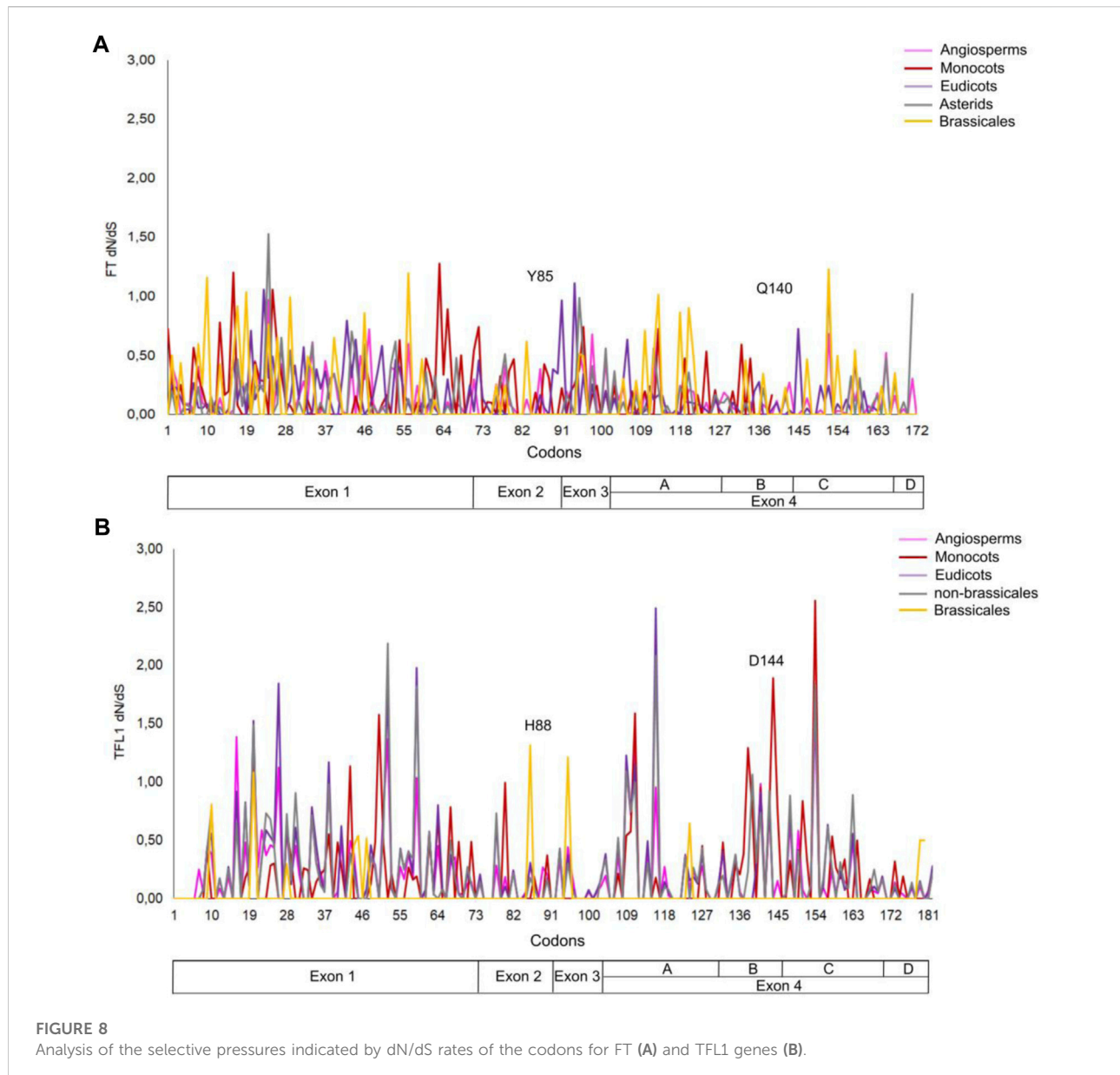


section. The residues Tyr85 and Gln140 (*A. thaliana* numbering) involved in the repression of FT activity and His88 and Asp144 (*A. thaliana* numbering) involved in the activation of TFL1, remained conserved in all inferred ancestral sequences of angiosperms, which is consistent with the molecular function previously described for these residues (Ahn et al., 2006a; Ho and Weigel, 2014).

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 (in kBT/e) 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). Some residues considered important for P-loop domain function, such as Glu109 and Gln140 maintained predominantly negative charges which corroborate with the results reported by Ho and Weigel (2014). However, in the present study, we performed a complete analysis of the potential charge of the P-loop domain and found that this domain shows a potential charge conserved in the ancestral and current FT structures.

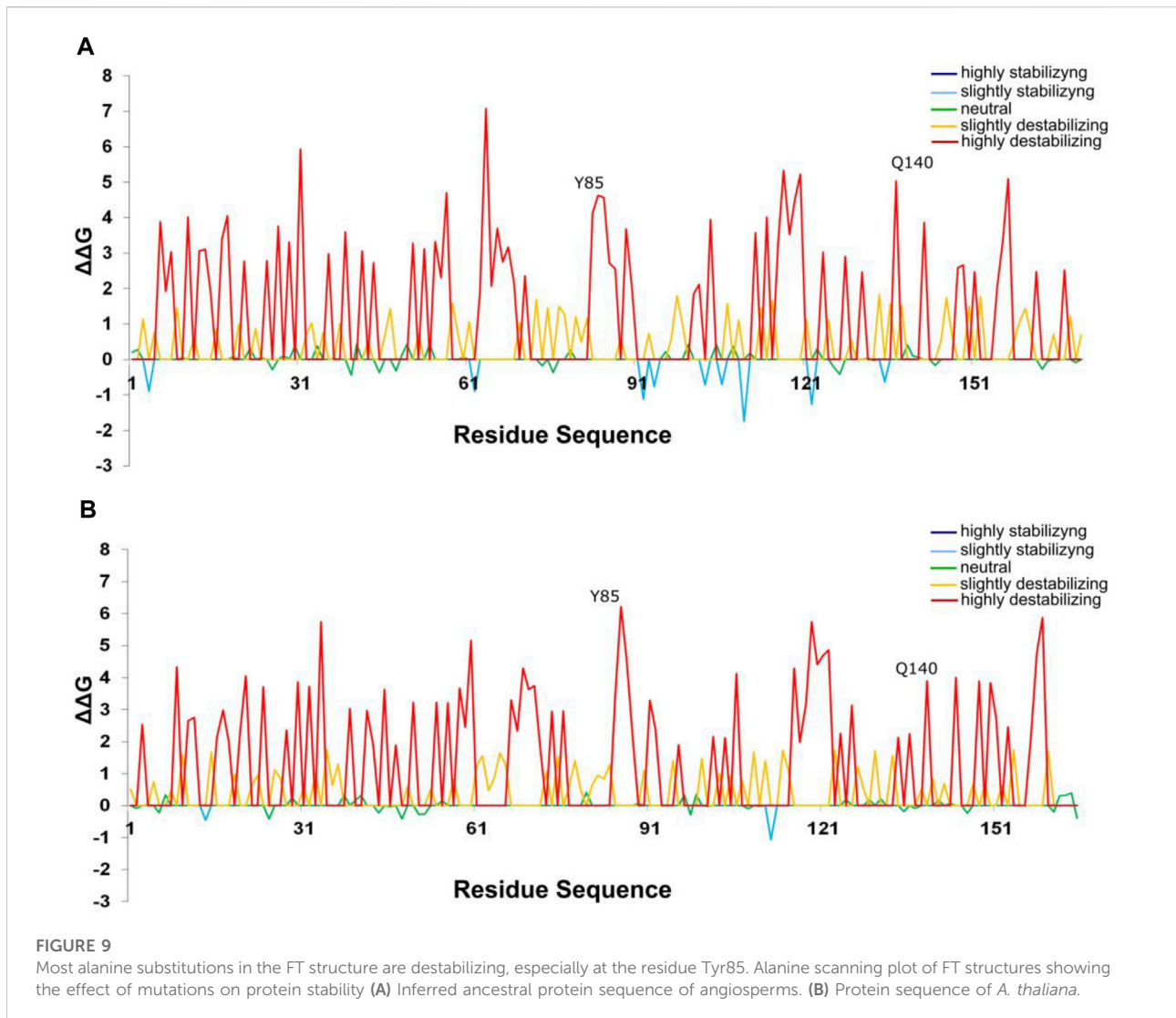
Two important structural features of the FT protein regarding the regulation of flowering include the surface-exposed loop region, named segment B (residues 128–141) encoded by the fourth exon which is involved in PC binding,



and 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 (Figures 7A–D), thus indicating the conservation of these residue interactions over evolutionary time.

Our structural analysis showed 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 F/TFL1 evolution and their effects on protein structural stability

The number of mutations that confer advantageous changes during the evolutionary process is limited due to the crossing of 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 adaptive evolutionary constraints in proteins structures (Studer et al., 2014a; 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 8B) 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., 2006a; 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 (Figure 9), 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 9). Moreover, we found that the codons that encode these residues are under constant selective pressure, thus corroborating with previous findings for the FT gene (Klintonas et al., 2012b; Mackenzie et al., 2019b). In contrast, TFL1 demonstrated a predominant neutral natural selection (Figure 8).

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 ($\Delta\Delta G > 0$), suggesting purifying selection over evolutionary time. These findings also corroborate the 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 (Figures 1, 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 (Supplementary Information S9).

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) (Supplementary Information S10, S11). 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.

During the evolutionary process, new mutations can be retained in the genomic pool by the relative strengths of 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 (Sen et al., 2011; Studer et al., 2014b; Piot et al., 2017) have demonstrated that the appearance of destabilizing mutations in residues directly involved in the active sites could be selected by positive selection, and that such mutations may be functionally necessary as they may contribute to increasing the conformational flexibility of some regions and allowing adaptation under different conditions, which may be the case of amino acid changes in the 4th exon region of the TFL1 protein in *A. thaliana*. In addition, this destabilization can be compensated for by the occurrence of stabilizing mutations in other sites such as the formation of a new H-bond between amino acids Asp139 and Ser137 (Figure 7B) that contributes to maintaining the overall protein stability (Buller and Townsend, 2013; Studer et al., 2014a; 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.

Conclusion

Here, we have reconstructed the evolutionary history of FT/TFL1-like proteins in the main flowering plant groups and by combining ancestral sequence inference with structural and mutational analyses, we have identified the main residue 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 selective pressures in residues located at the phosphatidylcholine binding site involved with H-bond formation indicate their structural role to maintain FT overall stability throughout evolution. In addition, residues of the 4th exon are found to be under positive selection and they may be involved in the conformational alterations of proteins. However, the formation of H-bonds that confer stability to the structure may be indicative that structural changes are compensated by the occurrence of stabilizing mutations to maintain the overall stability of the protein. Finally, our study opens up new insights to understand the roles of natural selection in the adaptive evolution of flowering proteins in angiosperms and could further help crop improvement with an economic interest in flower and fruit industries.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

DAJ, DMB, and EFM developed the research in databases and conducted the evolutionary and structural tests under the

References

Ahn, J. H., Miller, D., Winter, V. J., Banfield, M. J., Jeong, H. L., So, Y. Y., et al. (2006a). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J.* 25, 605–614. doi:10.1038/sj.emboj.7600950

supervision of TA, KS, LMC, and SS; DAJ, DMB, SS, KS, and TA wrote the text of the manuscript; all authors reviewed and contributed to the final version of the manuscript.

Funding

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) and the Fundação Amazônia de Amparo a Estudos e Pesquisas do Pará (FAPESPA), through the grant AUXPE 88881.159139/2017-01 to TA. KS is also grateful for the scholarship from the Brazilian funding agency CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). DAJ is also grateful for the scholarship from the Brazilian funding agency CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPERJ (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro). SS was supported by the National Science Foundation Postdoctoral Research Fellowship in Biology (1906333).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.954015/full#supplementary-material>

Ahn, J. H., Miller, D., Winter, V. J., Banfield, M. J., Jeong, H. L., So, Y. Y., et al. (2006b). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J.* 25, 605–614. doi:10.1038/sj.emboj.7600950

- Anisimova, M., and Kosiol, C. (2007). Investigating protein-coding sequence evolution with Probabilistic codon substitution models. doi:10.1093/molbev/msn232
- Auge, G. A., Penfield, S., and Donohue, K. (2019). Pleiotropy in developmental regulation by flowering-pathway genes: Is it an evolutionary constraint? *New Phytol.* 224, 55–70. doi:10.1111/nph.15901
- Banfield, M. J., and Brady, R. L. (2000). The structure of Antirrhinum centroradialis protein (CEN) suggests a role as a kinase regulator. *J. Mol. Biol.* 297, 1159–1170. doi:10.1006/jmbi.2000.3619
- Benkert, P., Künzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37, W510–W514. doi:10.1093/nar/gkp322
- Benner, S. a., Caraco, M. D., Thomson, J. M., and Gaucher, E. A. (2002). Planetary biology--paleontological, geological, and molecular histories of life. *Science* 296 (1979), 864–868. doi:10.1126/science.1069863
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2017). GenBank. *GenBank. Nucleic Acids Res.* 45, D37–D42. doi:10.1093/nar/gkw1070
- Bernier, G., and Périlleux, C. (2005a). A physiological overview of the genetics of flowering time control. *Plant Biotechnol. J.* 3, 3–16. doi:10.1111/j.1467-7652.2004.00114.x
- Bernier, G., and Périlleux, C. (2005b). A physiological overview of the genetics of flowering time control. *Plant Biotechnol. J.* 3, 3–16. doi:10.1111/j.1467-7652.2004.00114.x
- Bhattacharya, S., Banerjee, A., Sah, P. P., Mal, C., and 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. doi:10.1016/j.compbiolchem.2018.09.013
- Brandt, W., Backenköhler, A., Schulze, E., Plock, A., Herberg, T., Roese, E., et al. (2014). Molecular models and mutational analyses of plant specifier proteins suggest active site residues and reaction mechanism. *Plant Mol. Biol.* 84, 173–188. doi:10.1007/s11103-013-0126-0
- Buller, A. R., and Townsend, C. A. (2013). Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc. Natl. Acad. Sci. U. S. A.* 110, E653–E661. doi:10.1073/pnas.1221050110
- Carmona, M. J., Calonje, M., and Martínez-Zapater, J. M. (2007). The FT/TFLL1 gene family in grapevine. *Plant Mol. Biol.* 63, 637–650. doi:10.1007/s11103-006-9113-z
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi:10.1002/jcc.20290
- Chardon, F., and Damerval, C. (2005). Phylogenomic analysis of the PEBP gene family in cereals. *J. Mol. Evol.* 61, 579–590. doi:10.1007/s00239-004-0179-4
- Collani, S., Neumann, M., Yant, L., and Schmid, M. (2019). FT Modulates genome-wide DNA-binding of the bZIP transcription factor FD. *Plant Physiol.* 180, 367–380. doi:10.1104/pp.18.01505
- da Costa, K. S., Galúcio, J. M. P., Leonardo, E. S., Cardoso, G., Leal, É., Conde, G., et al. (2017). Structural and evolutionary analysis of Leishmania Alba proteins. *Mol. Biochem. Parasitol.* 217, 23–31. doi:10.1016/j.molbiopara.2017.08.006
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi:10.1038/nmeth.2109
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667. doi:10.1093/nar/gkh381
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). *Bayesian Phylogenetics with BEAUti and the BEAST 1.7* Research article, 29, 1969–1973. doi:10.1093/molbev/mss075
- Edgar, R. C. (2004). Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340
- Faber, M. S., Wrenbeck, E. E., Azouz, L. R., Steiner, P. J., and Whitehead, T. A. (2019). Impact of *in Vivo* protein folding probability on local fitness landscapes. *Mol. Biol. Evol.* 36, 2764–2777. doi:10.1093/molbev/msz184
- Fan, Z.-Y., He, X.-H., Fan, Y., Yu, H.-X., Wang, Y.-H., Xie, X.-J., et al. (2020). Isolation and functional characterization of three MiFTs genes from mango. *Plant Physiol. biochem.* 155, 169–176. doi:10.1016/j.plaphy.2020.07.009
- Fiser, A., and Šali, A. (2003). Modeller: Generation and Refinement of Homology-based protein structure models. *Methods Enzymol.* 374, 461–491. doi:10.1016/S0076-6879(03)74020-8
- Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi:10.1111/j.1469-8137.2009.02923.x
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9, 3891–3898. doi:10.1002/ece3.5015
- Gatesy, J., Sloan, D. B., Warren, J. M., Baker, R. H., Simmons, M. P., and Springer, M. S. (2019). Partitioned coalescence support reveals biases in species-tree methods and detects gene trees that determine phylogenomic conflicts. *Mol. Phylogenet. Evol.* 139, 106539. doi:10.1016/j.ympev.2019.106539
- Guerois, R., Nielsen, J. E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* 320, 369–387. doi:10.1016/S0022-2836(02)00442-4
- Hanzawa, Y., Money, T., and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7748–7753. doi:10.1073/pnas.0500932102
- Hedman, H., Källman, T., and Lagercrantz, U. (2009). Early evolution of the MFT-like gene family in plants. *Plant Mol. Biol.* 70, 359–369. doi:10.1007/s11103-009-9478-x
- Hestenes, M. R., and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* (1977). (1934) 49, 409. doi:10.6028/jres.049.044
- Hiraoka, K., Yamaguchi, A., Abe, M., and Araki, T. (2013). The florigen genes FT and TSF Modulate lateral shoot Outgrowth in Arabidopsis thaliana. *Plant Cell Physiol.* 54, 352–368. doi:10.1093/pcp/pcs168
- Ho, W. W. H., and Weigel, D. (2014). Structural features determining flower-Promoting activity of Arabidopsis FLOWERING LOCUS T. *Plant Cell* 26, 552–564. doi:10.1105/tpc.113.115220
- Jin, S., Nasim, Z., Susila, H., and Ahn, J. H. (2020). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin. Cell Dev. Biol.* 109, 20–30. doi:10.1016/j.semcdb.2020.05.007
- Jørgensen, W. L., Maxwell, D. S., and 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. doi:10.1021/ja9621760
- Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., et al. (1999). Activation tagging of the floral inducer FT. *Science* 286 (1979), 1962–1965. doi:10.1126/science.286.5446.1962
- Karlgen, A., Gyllenstrand, N., Källman, T., Sundström, J. F., Moore, D., Lascoux, M., et al. (2011a). Evolution of the PEBP gene family in plants: Functional diversification in seed plant evolution. *Plant Physiol.* 156, 1967–1977. doi:10.1104/pp.111.176206
- Karlgen, A., Gyllenstrand, N., Källman, T., Sundström, J. F., Moore, D., Lascoux, M., et al. (2011b). Evolution of the PEBP gene family in plants: Functional diversification in seed plant evolution. *Plant Physiol.* 156, 1967–1977. doi:10.1104/pp.111.176206
- Kim, S. Y., Yu, X., and Michaels, S. D. (2008). Regulation of CONSTANS and FLOWERING LOCUS T expression in response to changing light quality. *Plant Physiol.* 148, 269–279. doi:10.1104/pp.108.122606
- Kinoshita, T., Ono, N., Hayashi, Y., Morimoto, S., Nakamura, S., Soda, M., et al. (2011). FLOWERING LOCUS T regulates stomatal opening. *Curr. Biol.* 21, 1232–1238. doi:10.1016/j.cub.2011.06.025
- Klintonäs, M., Pin, P. A., Benlloch, R., Ingvarsson, P. K., and Nilsson, O. (2012a). 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. doi:10.1111/j.1469-8137.2012.04332.x
- Klintonäs, M., Pin, P. A., Benlloch, R., Ingvarsson, P. K., and Nilsson, O. (2012b). 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. doi:10.1111/j.1469-8137.2012.04332.x
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. *Science* 286, 1960–1962. doi:10.1126/science.286.5446.1960
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for Bigger datasets Brief communication. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054
- Laskowski, R. a., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi:10.1107/S002188982009944
- Lee, R., Baldwin, S., Kenel, F., McCallum, J., and Macknight, R. (2013). FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nat. Commun.* 4, 2884. doi:10.1038/ncomms3884

- Lifschitz, E., Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., Amsellem, Z., et al. (2006). The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6398–6403. doi:10.1073/pnas.0601620103
- Liu, Y.-Y., Yang, K.-Z., Wei, X.-X., and Wang, X.-Q. (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. doi:10.1111/nph.14066
- Mackenzie, K. K., Coelho, L., Lütken, H., and Müller, R. (2019a). Phylogenomic analysis of the PEBP gene family from *Kalanchoë*. 1–16.
- Mackenzie, K. K., Coelho, L., Lütken, H., and Müller, R. (2019b). Phylogenomic analysis of the PEBP gene family from *Kalanchoë*. 1–16.
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “2010 Gateway computing Environments Workshop,” in GCE 2010. 2010 Gateway Computing Environments Workshop (New Orleans: GCE 2010).
- Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y. C., Dörmann, P., and Coupland, G. (2014). Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nat. Commun.* 5, 3553–3558. doi:10.1038/ncomms4553
- Nakamura, Y., Lin, Y. C., Watanabe, S., Liu, Y. C., Katsuyama, K., Kanehara, K., et al. (2019). High-resolution crystal structure of Arabidopsis FLOWERING LOCUS T illuminates its Phospholipid-binding site in flowering. *iScience* 21, 577–586. doi:10.1016/j.isci.2019.10.045
- Navarro, C., Abelenda, J. A., Cruz-Oró, E., Cuéllar, C. A., Tamaki, S., Silva, J., et al. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* 478, 119–122. doi:10.1038/nature10431
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426. doi:10.1093/oxfordjournals.molbev.a040410
- Neves Cruz, J., da Costa, K. S., de Carvalho, T. A. A., and de Alencar, N. A. N. (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.* 38, 1425–1434. doi:10.1080/07391102.2019.1607560
- Ogiso-Tanaka, E., Matsubara, K., Yamamoto, S., Nonoue, Y., Wu, J., Fujisawa, H., et al. (2013). Natural variation of the RICE FLOWERING LOCUS T 1 contributes to flowering time divergence in Rice. *PLoS One* 8, e75959. doi:10.1371/journal.pone.0075959
- Ospina-Zapata, D. A., Madrigal, Y., Alzate, J. F., and Pabón-Mora, N. (2020). Evolution and expression of reproductive Transition regulatory genes FT/TFL1 with Emphasis in selected Neotropical Orchids. *Front. Plant Sci.* 11, 469. doi:10.3389/fpls.2020.00469
- Pin, P. A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J. L., et al. (2010). An antagonistic pair of FT homologs Mediates the control of flowering time in sugar beet. *Science* 330 (1979), 1397–1400. doi:10.1126/science.1197004
- Pin, P. A., and Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant Cell Environ.* 35, 1742–1755. doi:10.1111/j.1365-3040.2012.02558.x
- Piot, A., Hackel, J., Christin, P. A., and Besnard, G. (2017). One-third of the plastid genes evolved under positive selection in PACMAD grasses. *PLanta* 247, 255–266. doi:10.1007/S00425-017-2781-X
- Rallapalli, P. M., Orenco, C. A., Studer, R. A., and Perkins, S. J. (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. doi:10.1093/molbev/msu248
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior Summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi:10.1093/sysbio/syy032
- Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3, 198–210. doi:10.1002/wcms.1121
- Sen, L., Fares, M. A., Liang, B., Gao, L., Wang, B., Wang, T., et al. (2011). Molecular evolution of rbcL in three gymnosperm families: Identifying adaptive and coevolutionary patterns. *Biol. Direct* 6, 29–19. doi:10.1186/1745-6150-6-29
- Sharir-Ivry, A., and 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. doi:10.1093/molbev/msy177
- Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi:10.1016/j.gde.2015.11.003
- Stamboulian, M., Guerrero, R. F., Hahn, M. W., and Radivojac, P. (2020). The ortholog conjecture revisited: The value of orthologs and paralogs in function prediction. *Bioinformatics* 36, i219–i226. doi:10.1093/BIOINFORMATICS/BTAA468
- Studer, R. A., Christin, P. A., Williams, M. A., and Orenco, C. A. (2014a). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2223–2228. doi:10.1073/pnas.1310811111
- Studer, R. A., Christin, P. A., Williams, M. A., and Orenco, C. A. (2014b). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2223–2228. doi:10.1073/pnas.1310811111
- Tokuriki, N., and Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19, 596–604. doi:10.1016/j.sbi.2009.08.003
- Wang, H., Ni, X., and Harris-Shultz, K. (2019). Molecular evolution of the plant ECERIFERUM1 and ECERIFERUM3 genes involved in aliphatic hydrocarbon production. *Comput. Biol. Chem.* 80, 1–9. doi:10.1016/j.compbiolchem.2019.02.009
- Wang, Z., Yang, R., Devisetty, U. K., Maloof, J. N., Zuo, Y., Li, J., et al. (2017). The divergence site-specific functional divergence among clades of TFL1 is independent to their interaction and binding Activities. *Front. Plant Sci.* 8, 697–716. doi:10.3389/fpls.2017.00697
- Wendick, C. J., and Chang, B. S. W. (2012). An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.* 29, 1297–1300. doi:10.1093/molbev/msr311
- Wiberg, K. B. (1965). A Scheme for Strain energy minimization. Application to the Cycloalkanes 1. *J. Am. Chem. Soc.* 87, 1070–1078. doi:10.1021/ja01083a024
- Wickland, D. P., and Hanzawa, Y. (2015). The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: Functional evolution and molecular Mechanisms. *Mol. Plant* 8, 983–997. doi:10.1016/j.molp.2015.01.007
- Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and Temporal information during floral Induction in Arabidopsis. *Science* 309 (1979), 1056–1059. doi:10.1126/science.1114358
- Xi, W., Liu, C., Hou, X., and 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. doi:10.1105/tpc.109.073072
- Xu, D., and 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. doi:10.1016/j.bpj.2011.10.024
- Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., and Araki, T. (2005a). TWIN SISTER of FT (TSF) acts as a floral pathway integrator redundantly with FT. *Plant Cell Physiol.* 46, 1175–1189. doi:10.1093/pcp/pci151
- Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., and Araki, T. (2005b). TWIN SISTER of FT (TSF) acts as a floral pathway integrator redundantly with FT. *Plant Cell Physiol.* 46, 1175–1189. doi:10.1093/pcp/pci151
- Yoo, S. J., Chung, K. S., Jung, S. H., Yoo, S. Y., Lee, J. S., and Ahn, J. H. (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. doi:10.1111/j.1365-313X.2010.04234.x
- Yu, X., Liu, H., Sang, N., Li, Y., Zhang, T., Sun, J., et al. (2019). Identification of cotton MOTHER OF FT AND TFL1 homologs, GhMFT1 and GhMFT2, involved in seed germination. *PLoS One* 14, e0215771. doi:10.1371/journal.pone.0215771
- Zhang, X., An, L., Nguyen, T. H., Liang, H., Wang, R., Liu, X., et al. (2015). The Cloning and functional characterization of Peach CONSTANS and FLOWERING LOCUS T homologous genes PpCO and PpFT. *PLoS One* 10, e0124108. doi:10.1371/journal.pone.0124108
- Zheng, X.-M., Wu, F.-Q., Zhang, X., Lin, Q.-B., Wang, J., Guo, X.-P., et al. (2016). Evolution of the PEBP gene family and selective signature on FT-like clade. *J. Syst. Evol.* 54, 502–510. doi:10.1111/jse.12199
- Zhu, Y., Klasfeld, S., Jeong, C. W., Jin, R., Goto, K., Yamaguchi, N., et al. (2020). TERMINAL FLOWER 1-FD complex target genes and competition with FLOWERING LOCUS T. *Nat. Commun.* 11, 5118. doi:10.1038/s41467-020-18782-1