



# In silico study of phosphatidylethanolamine-binding proteins in oil palm, coconut, and date palm: evolutionary insights into genetic relations, sequence conservation, and functional implications

Aqwin Polosoro · Wening Enggarini · Kusumawaty Kusumanegara · Dani Satyawan · Toto Hadiarto · Miftahudin · Ence Darma Jaya Supena

Received: 21 January 2024 / Accepted: 12 March 2024  
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

**Abstract** The phosphatidylethanolamine-binding protein (PEBP) gene family plays a vital role in various aspects of plant biology, including development, differentiation, metabolism, and signaling pathways. This study examined the genetic relations between PEBP genes in palms (oil palm, coconut, and date) using various techniques of sequence retrieval, alignment, phylogenetic tree construction, and protein structure modeling. A significant genetic resemblance was found between coconut and oil palm, surpassing the similarity observed between either of

these species and date palm. Three primary groups of PEBP proteins were identified: FLOWERING LOCUS T (FT), TERMINAL FLOWER1 (TFL1), and MOTHER OF FT AND TFL1 (MFT). The FT group was further categorized into two subclades. Analysis of the FT protein structures in these subclades revealed significant similarities. A notable difference was observed in the absence of important amino acids in specific segments of the second subclade, making it more similar to the TFL group. Furthermore, mRNA sequence analysis within and between palm species indicated significant genetic similarities in the evolutionary trajectories of these organisms, suggesting the possibility of gene duplication in their ancestral genomes. In silico expression analysis revealed that gene expression patterns differed from those of RNA sequences, possibly because data were not collected at a specific time. This study provides significant insights into the genetic evolution and diversification of the palm PEBP family, highlighting a possible link between the regulatory mechanisms that govern the flowering processes, which may be beneficial for accelerating breeding programs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10722-024-01960-1>.

A. Polosoro (✉) · Miftahudin · E. D. J. Supena (✉)  
Plant Biology Graduate Program, Department of Biology,  
Faculty of Mathematics and Natural Sciences, IPB  
University, Dramaga, Bogor, West Java 16680, Indonesia  
e-mail: aqwinpolosoro@apps.ipb.ac.id

E. D. J. Supena  
e-mail: encedarmo@apps.ipb.ac.id

A. Polosoro · W. Enggarini · D. Satyawan · T. Hadiarto  
Research Center for Genetic Engineering, National  
Research and Innovation Agency (BRIN), Jalan Raya  
Jakarta-Bogor Km. 46, Cibinong, Bogor, West Java 16911,  
Indonesia

K. Kusumanegara  
Research Center for Horticulture, National Research  
and Innovation Agency (BRIN), Jalan Raya Jakarta-Bogor  
Km. 46, Cibinong, Bogor, West Java 16911, Indonesia

**Keywords** Palm · Genetic evolution · PEBP · FT · MFT · TFL

## Introduction

The Arecaceae family, which includes palms, is derived from a single ancestral lineage and forms a closely related phylogenetic group (Comer et al. 2016). Cássia-Silva et al. (2020) identified notable evolutionary alterations in growth patterns and sizes within this family. Nadot et al. (2016) focused on the development of sexual systems in palms, taking advantage of recent advances in our phylogenetic understanding. Furthermore, advanced phylogenomic techniques have facilitated comprehensive investigations into the evolution of gene families in palms (Barrett et al. 2016). However, there are still gaps in our understanding of the interrelations among different palm genera and species. Barrett et al. (2019) proposed that historical polyploidy events may have affected the genomic evolution of palms, an area of research that remains incompletely understood.

Palms such as oil palm, coconut, and date palm show some differences in flower development. Palms usually produce separate male and female flowers organized in specific clusters (Castaño et al. 2014). They can exhibit either monoecious or dioecious characteristics, with male and female flowers present on the same or different plants. The oil palm is monoecious, with a spadix inflorescence containing both male and female flowers (Adam et al. 2007). Coconut commonly exhibit both male and female flowers on the same inflorescence, a condition known as monoeciousness (Nayar 2018). However, there are dioecious variations in which separate male and female plants exist, with the wind playing a role in pollination. Date palm exhibit dioecious characteristics, with distinct male and female plants (Al-Ameri et al. 2016). The observed variations in palm flower structures and reproductive systems among these economically substantial species underscore their diversity. Preserved gene coding regions play a crucial role in evolutionary research by offering valuable information about the structure and function of genes. The conserved sequences described by Moutinho and Eyre-Walker (2022) and Shan et al. (2022) contribute to our understanding of codon organization and coevolutionary dynamics. Yıldırım and Vogl (2023) argue that these tools also aid in the identification of regulatory elements and transcription factor-binding sites, which play a crucial role in gene control and the development of phenotypic traits. Comparative genomic

analyses of gene orders provide insights into evolutionary processes (Herrig et al. 2023; Xu et al. 2022), whereas similarities in gene structure aid in tracing their coevolution during speciation (Fuertes et al. 2019).

Polosoro et al. (2021) recently discovered phosphatidylethanolamine-binding proteins (PEBPs) in oil palm. These proteins are evolutionarily conserved and play important roles in diverse biological processes (Wang et al. 2019). These pathways play crucial roles in the transduction of cellular signaling networks (Hoogenboom et al. 2016; Leeggangers et al. 2018; Tribhuvan et al. 2020; Zhang et al. 2016). In addition to their signaling role, PEBPs affect various aspects of plant growth and development, including the regulation of flowering time, circadian rhythms, and seed development (Wang et al. 2016; Zhang et al. 2016).

Yang et al. (2019) classified plant PEBP into three main clades: FLOWERING LOCUS T (FT)-like, TERMINAL FLOWER 1 (TFL1)-like, and MOTHER OF FT AND TFL1 (MFT)-like. FT-like proteins, such as FT and TWIN SISTER OF FT (TSF), play a vital role in promoting flowering, as supported by research conducted by Ho and Weigel (2014) in *Arabidopsis* and Li et al. (2015) in cotton plants. On the other hand, TFL1-like proteins, including TFL1, BROTHER OF FT AND TFL1 (BFT), and CENTRORADIALIS (CEN), act as inhibitors of flower development, as demonstrated in apple and *Arabidopsis* (Haberman et al. 2016; Ahn et al. 2006; Yoo et al. 2010). Chen et al. (2018) found that MFT-like proteins in longan (*Dimocarpus longan*) primarily participate in the stress response and seed germination rather than flowering.

The categorization of the PEBP subfamilies aids in understanding their varied functions and regulatory roles in the processes of plant growth and flowering. Numerous investigations have been conducted on the contrasting roles of FT and TFL1 in flowering in *Arabidopsis thaliana*, despite their high DNA similarity (Bennett and Dixon 2021; Nakano et al. 2015; Z. Wang et al. 2017; Hanzawa et al. 2005; Ahn et al. 2006). Certain amino acids found in FT paralogs can hinder the flowering process in species, such as sugar beets (*Beta vulgaris*). In addition, genes resembling FT and TFL1 have been observed to control the growth and termination of meristems in perennial and cyclically growing species (Pin et al. 2010). This study seeks to enhance our understanding of the

evolutionary dynamics of PEBP genes in palm species. We aimed to understand the evolutionary path of palm PEBP-like genes in important palm species, such as oil palm, coconut, and date palm, by examining publicly available genomic DNA, mRNA, and protein sequences. Our study reveals the evolutionary connections and genetic associations among PEBP genes in these species, including the identification of FT-like genes originating from at least four distinct gene groups. The presence of consistent relational patterns in specific sequences observed in oil palm, coconut, and date palm implies a potential association with the regulatory mechanisms that control their flowering processes. This study signifies a potential association with the regulatory mechanisms that control their flowering processes, which are useful for accelerated breeding programs.

## Methods

### Sequence retrieval

This study involved a thorough collection of PEBP protein sequences from relevant references and databases. To obtain a comprehensive set of PEBP protein sequences in the palm family, we used the sequences of *Arabidopsis thaliana* PEBP, AtFT1 (BA77838), AtTFL1 (At5g03840), and AtMFT (NP\_173250.1) from the National Center for Biotechnology Information (NCBI) database to perform a targeted Blastp search, specifically for the Arecaceae family (taxid: 4710). Due to the scarcity of coconut mRNA data in the NCBI database, we employed an alternative methodology for analyzing several PEBP sequences from this species. They were obtained from the Coconut Genome Database (GCA\_008124465) through a Blastn search, utilizing PEBP sequences from oil palm and date palm (Supplementary 1). This ensured the incorporation of the coconut data into our analysis.

### Construction of phylogenetic trees and sequence alignment

The MUSCLE alignment tool was used to align protein and mRNA sequences. Phylogenetic analysis was performed using the maximum likelihood tree method, employing the Molecular Evolutionary

Genetics Analysis software package. We used Interactive Tree of Life post-analysis to visually interpret the phylogenetic trees. Furthermore, we conducted a pairwise distance analysis to compile mRNA sequences using the Molecular Evolutionary Genetics Analysis software to assess the genetic distances among the PEBP genes, thereby facilitating the quantification of their genetic divergence.

### Protein structure homology modeling with SWISS-MODEL

The SWISS-MODEL platform was used to analyze protein sequences and predict their functional characteristics. Our primary objective was to enhance the reliability of the model quality by utilizing Qualitative Model Energy Analysis. The SWISS-MODEL process involves identifying compatible templates, aligning target sequences with these templates, constructing a model based on the alignment, and thoroughly evaluating the quality of the model. The metrics of the SWISS-MODEL include sequence similarity data, Global Model Quality Estimate (GMQE), and QMEANDisCo. These metrics facilitated a thorough assessment of the models and provided insights into the structural and functional characteristics of the proteins under investigation.

### In silico expression analysis of coconut and oil palm PEBP genes

An in silico expression study was performed to investigate whether the expression of PEBP genes was correlated within the palm family. Coconut and oil palm were selected because of the availability of transcriptome data for several organs at different stages. Transcriptomic Sequence Read Archive (SRA) data for mature leaves (ERR3588913 and SRR25119995), roots (SRR22255955 and SRR7812013), male flowers (DRR129238 and DRR053157), and female flowers (DRR045028 and DRR053155) were obtained from the NCBI database. SRA data were extracted using Fastq and quantified using Kallisto to obtain transcripts per million levels for PEBP genes in coconut and oil palm. Transcripts per million values were examined using Heatmapper (University of Alberta, Canada) to explore differential expression.

## Results and discussion

### Grouping of PEBP genes in the palm family

We conducted thorough BLASTp and BLASTn searches to identify 27 unique PEBP-like proteins in the three palm tree species. The comprehensive dataset is presented systematically in Table 1. To ensure the accuracy of our findings, we comprehensively examined each predicted gene. Multiple MUSCLE alignments were performed to verify the presence of a fully intact and identifiable PEBP domain. After identification and verification of the sequences, we categorized the PEBP linked to palm

species into different clades and subclades. This classification is essential for understanding the evolutionary relations among proteins. Table 1 displays the unique gene names and protein accession codes assigned to each protein, which were obtained from annotations in the NCBI protein database. We used the NCBI gene ID information to identify the precise locations of these proteins, thereby offering a comprehensive framework for our research outcomes. Additional predicted proteins are annotated as gene names with asterisks. This approach improved data reliability and facilitated a deeper understanding of the genetic architecture and evolutionary dynamics of PEBP-like proteins in the palm family.

**Table 1** PEBP identification in coconut, oil palm, and dates using Blastp in the NCBI database

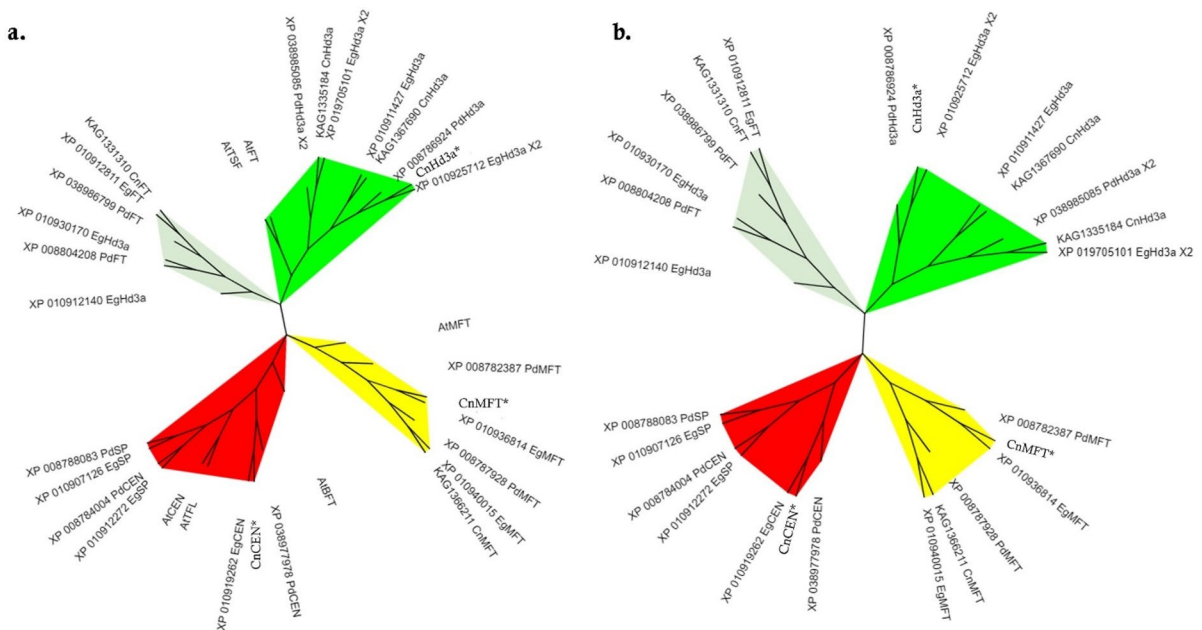
No	Protein group	Protein code	Genes name	Gene ID	Species	Length	Location
1	FT-like subclade I	XP_010925712	<i>EgHd3a X2</i>	LOC105048191	<i>Elaeis guineensis</i>	176 aa	Chromosome 7
2		XP_010911427	<i>EgHd3a</i>	LOC105037463	<i>Elaeis guineensis</i>	176 aa	Unknown
3		XP_019705101	<i>EgHd3a X2</i>	LOC105040685	<i>Elaeis guineensis</i>	184 aa	Chromosome 3
4		XP_008786924	<i>PdHd3a</i>	LOC103705069	<i>Phoenix dactylifera</i>	176 aa	Chromosome 9
5		XP_038985085	<i>PdHd3a X2</i>	LOC103702982	<i>Phoenix dactylifera</i>	176 aa	Chromosome 8
6		CnHd3a*	<i>CnHd3a</i>	COCNU_02G002890	<i>Cocos nucifera</i>	176 aa	Chromosome 2
7		KAG1367690	<i>CnHd3a</i>	COCNU_14G001580	<i>Cocos nucifera</i>	176 aa	Chromosome 14
8		KAG1335184	<i>CnHd3a</i>	COCNU_03G013030	<i>Cocos nucifera</i>	184 aa	Chromosome 3
9	FT-like subclade II	XP_010912140	<i>EgHd3a</i>	LOC105038131	<i>Elaeis guineensis</i>	173 aa	Chromosome 2
10		XP_010930170	<i>EgHd3a</i>	LOC105051422	<i>Elaeis guineensis</i>	179 aa	Chromosome 9
11		XP_010912811	<i>EgFT</i>	LOC105038642	<i>Elaeis guineensis</i>	179 aa	Chromosome 2
12		XP_008804208	<i>PdFT</i>	LOC103717552	<i>Phoenix dactylifera</i>	179 aa	Chromosome 14
13		XP_038986799	<i>PdFT</i>	LOC103708554	<i>Phoenix dactylifera</i>	179 aa	Chromosome 10
14		KAG1331310	<i>CnFT</i>	COCNU_02G012780	<i>Cocos nucifera</i>	179 aa	Chromosome 2
15	MFT	XP_010936814	<i>EgMFT</i>	LOC105056345	<i>Elaeis guineensis</i>	175 aa	Chromosome 13
16		XP_010940015	<i>EgMFT</i>	LOC105058700	<i>Elaeis guineensis</i>	175 aa	Chromosome 15
17		XP_008782387	<i>PdMFT</i>	LOC103701938	<i>Phoenix dactylifera</i>	175 aa	Chromosome 6
18		XP_008787928	<i>PdMFT</i>	LOC103705831	<i>Phoenix dactylifera</i>	175 aa	Chromosome 15
19		CnMFT*	<i>CnMFT</i>	COCNU_05G009470	<i>Cocos nucifera</i>	175 aa	Chromosome 5
20	KAG1366211	<i>CnMFT</i>	COCNU_13G000010	<i>Cocos nucifera</i>	175 aa	Chromosome 13	
21	TFL	XP_010919262	<i>EgCEN</i>	LOC105043421	<i>Elaeis guineensis</i>	173 aa	Chromosome 4
22		XP_010912272	<i>EgSP</i>	LOC105038236	<i>Elaeis guineensis</i>	173 aa	Chromosome 2
23		XP_010907126	<i>EgSP</i>	LOC105033865	<i>Elaeis guineensis</i>	173 aa	Unknown
24		XP_008784004	<i>PdCEN</i>	LOC103703083	<i>Phoenix dactylifera</i>	173 aa	Chromosome 10
25		XP_008788083	<i>PdSP</i>	LOC103705942	<i>Phoenix dactylifera</i>	173 aa	Chromosome 14
26		XP_038977978	<i>PdCEN</i>	LOC103714156	<i>Phoenix dactylifera</i>	173 aa	Chromosome 2
27		CnCEN*	<i>CnCEN</i>	COCNU_07G006960	<i>Cocos nucifera</i>	173 aa	Chromosome 7

\* Collected directly from the coconut database (GCA\_008124465) as the predicted protein

## Phylogenetic analyses

The maximum likelihood method was used to conduct phylogenetic reconstructions in our study. We aligned PEBP sequences with mRNA sequences derived from Arabidopsis and several palm species, including oil palm, date, and coconut. The phylogenetic trees in Fig. 1a, b exhibited consistent topologies, confirming the reliability of our alignment and reconstruction methods. The 27 PEBP proteins were classified into three main groups: the FT-clade, TFL-clade, and MFT-clade. In the FT clade, 14 FT-like proteins were categorized into two subclades: Subclade I, consisting of eight members, and Subclade II, consisting of six members (refer to Fig. 1a, b, and Table 1 for more details). None of the FT-like proteins in Subclade I aligned with the reference proteins AtFT and AtTSF, suggesting that these proteins have distinct sequence patterns. PEBP subclades in sugarcane, soybean, and sorghum show similar variations, with two, three, and three subclades, respectively (Książkiewicz et al. 2016; Lee et al. 2021; Venail et al. 2022). Lee et al. (2021) found that most FT proteins in various subclades play a role in promoting flowering, albeit with varying degrees of impact.

Proteins from various palm species tend to form distinct subclades within the TFL clade (Fig. 1). These subclades exhibited significant differences compared to their Arabidopsis counterparts, such as AtTFL, AtCEN, and AtBFT. Divergence was observed within the MFT clade, with significant differences between the palm TFL clade proteins and the Arabidopsis representative AtMFT. These findings underline the evolutionary diversity and adaptability of PEBP family members. Certain proteins display conserved characteristics among different species. Oil palm proteins exhibited greater similarity to coconut proteins compared to date palm proteins. Furthermore, our comparative analysis included mRNA sequences, as depicted in Fig. 1b, which contributed to a more comprehensive understanding of the evolutionary dynamics within the palm family. Protein and mRNA sequence analyses provided reliable results to enhance our understanding of the evolutionary history and functional differentiation of these proteins within the palm family.



**Fig. 1** Phylogenetic analysis of the **a** PEBP protein family with comparative references to FT, TFL, and MFT protein sequences and **b** PEBP mRNA in oil palm, coconut, and

dates. Both proteins and mRNA sequences clustered into three clades: FT (grey and green), TFL (red), and MFT (yellow). (Color figure online)

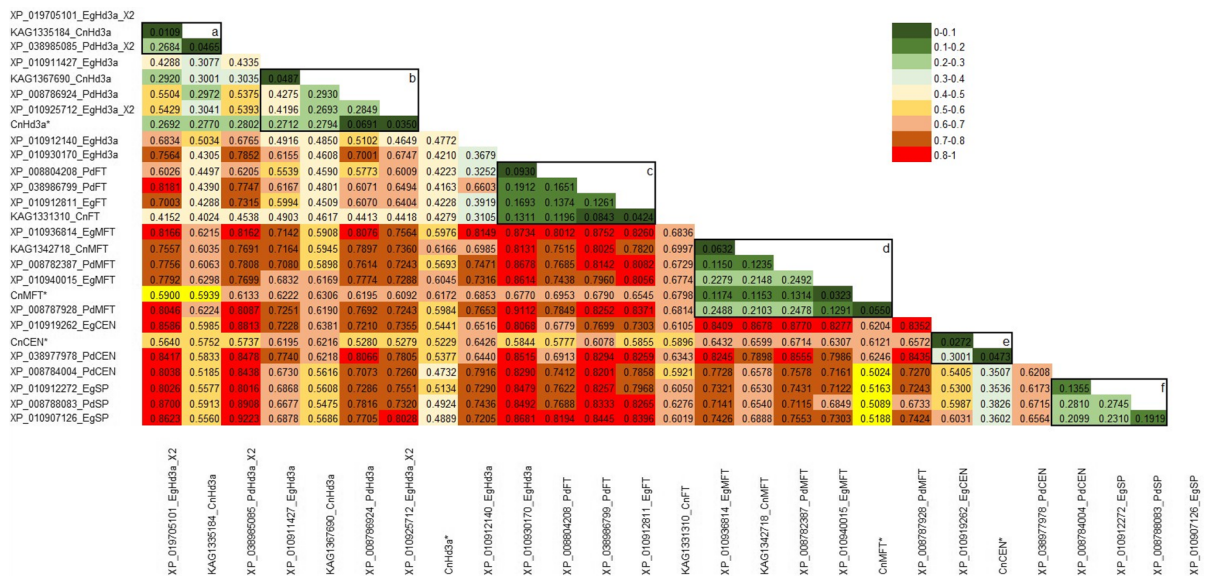
## Evolutionary divergence of PEBP mRNA

Figure 2 shows the pairwise distance results for the mRNA sequences of the PEBP family. Lower values indicate closer sequence proximity, and higher values indicate greater divergence. These results confirmed the phylogenetic patterns, suggesting significant sequence similarity within specific groups. In Fig. 2, box a, subclade I of FT-like exhibited a high degree of sequence homogeneity among its members. The coconut sequences exhibited an intermediate evolutionary position relative to the oil palm and date palm sequences, as indicated by their lower divergence values for both species. This suggests a closer evolutionary relation between coconut and these two palms. In contrast, the divergence between oil palm and date palm sequences was considerably greater, suggesting a substantial evolutionary gap. Based on these findings, the PEBP family was classified into six groups: FT-like subclade I, one group from FT-like subclade II, one group from MFT-like, and two groups from the TFL clade. This categorization is depicted in Fig. 2a–f. This pattern of sequence proximity extends beyond comparisons between different species to encompass duplications within the same species. Furthermore, we identified genes specific to

certain species and observed significant variations in their functions. One example is the oil palm gene XP\_010912140, which exhibits a DNA sequence alteration resulting in a modified protein function, despite the presence of similar variants in other species. The details of these findings are presented in Supplementary Table 2.

## Characterization of PEBP genes in the palm family

Oil palm, coconut, and date palm, which belong to the Arecaceae family, exhibit genetic similarities due to their shared botanical ancestry. Khan et al. (2018) found evidence supporting a closer genetic relation between oil palm and coconut compared to oil palm and date palm. Genetic proximity between PEBP-coding genes indicated that oil palm contains a higher copy number of all PEBP protein members. This study has several crucial implications. Our research supports these findings, indicating significant genetic similarities between oil palm and coconut. However, the PEBP genes in these species showed less resemblance to those found in date palm. This disparity underlines the variation in genetic evolution within the Arecaceae family. The Arecaceae family, which includes oil palm, coconut, and date palm, exhibited



**Fig. 2** The depiction of the evolutionary divergence of PEBP mRNA sequences from oil palm, coconut, and dates is presented as a distance matrix. Box a–f are groups of mRNA with

closely similar sequences, which suggest they originated from the same ancestral genes

gene duplication (Barrett et al. 2019). This tendency may be crucial for understanding the sources of comparable PEBP gene sequences in these plants. Gene duplication in the Arecaceae family indicates the intricate genetic mechanisms that contribute to the diversity of PEBP sequences in different palm species, highlighting the intricate nature of genetic evolution and adaptation in this family.

The similarity in the PEBP protein and mRNA sequences between coconut and oil palm, as opposed to date palm, indicates a close evolutionary relation. Hypothetically, the evolutionary relation among these three species may be related to the changes observed in the transition of flower structure throughout their evolution. Coconut plants generate hermaphrodite flowers, whereas oil palm produce separate male and female flowers, rendering them monoecious. On the other hand, date palms are dioecious. The concepts of monoecy and dioecy are closely related to evolutionary processes. According to Cronk (2022), the occurrence of monoecy where male and female flowers exist on the same plant, is considered to be the primary pathway leading to dioecy, which involves the division of sexual functions among distinct individuals. According to a recent study by Muyle et al. (2021), the transition from monoecy to dioecy can be achieved through mutations in a single gene. This evidence indicates that oil palm might be a transitional form between coconut and date palm.

#### Crucial amino acid residues

The MUSCLE alignment technique used to identify crucial amino acid residues in the PEBP family displayed notable variations within the PEBP domain alignment of the 27 proteins (Fig. 3). The importance of tyrosine residues in the structure of the FT/Hd3a protein was observed based on the findings of Hanzawa et al. (2005) (Fig. 3, box a). Hanzawa et al. conducted a study in which a tyrosine residue was substituted with a histidine in AtFT1, resulting in enhanced protein activity and decreased flowering, similar to the effects observed when AtTFL1 was overexpressed. A consistent pattern was observed in segment A (Fig. 3, box a), wherein all members of the FT subclade I exhibited the presence of tyrosine, which promotes flowering. In contrast, the majority of the proteins in the FT subclade II exhibited histidine residues resembling TFL proteins, except

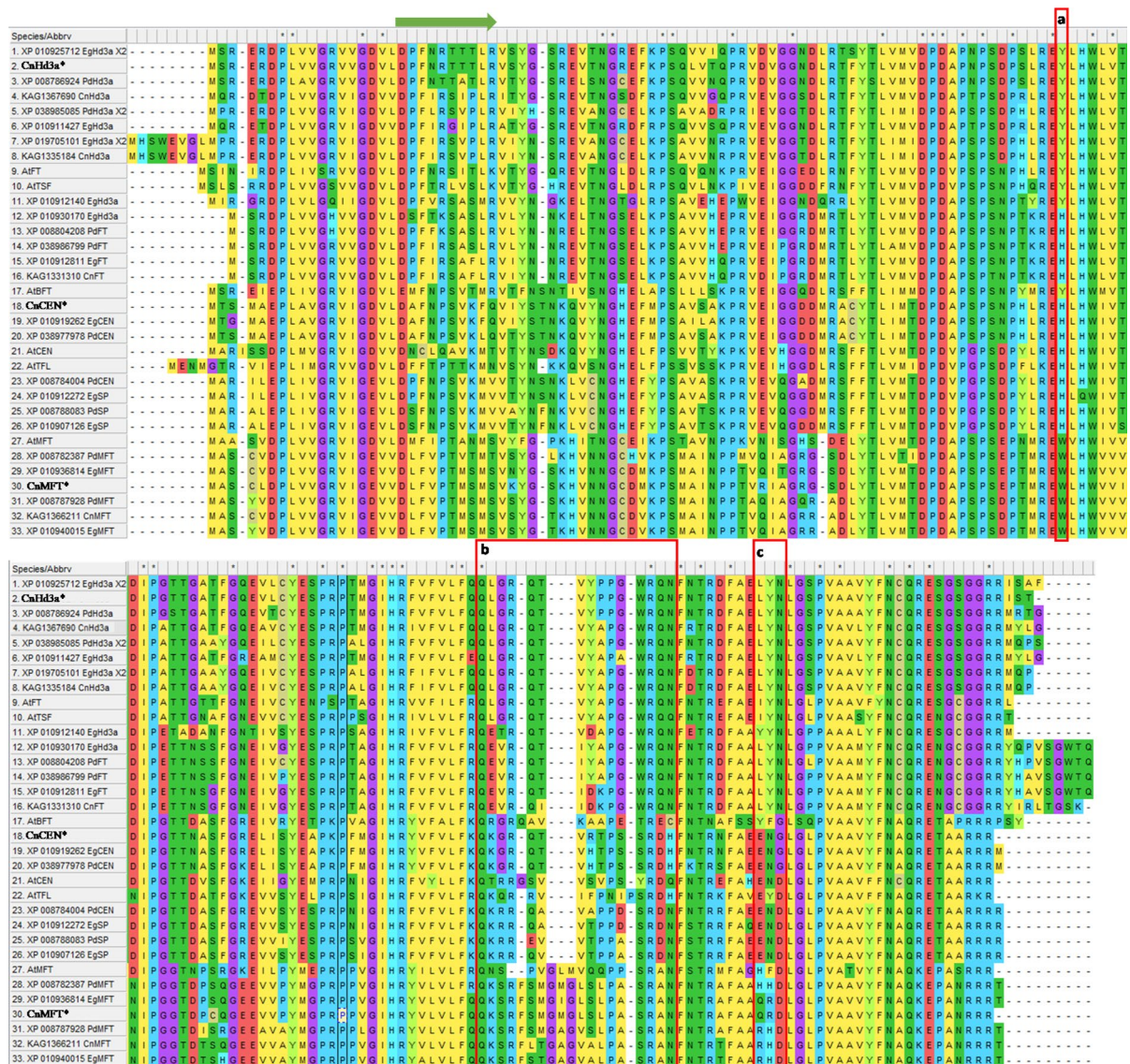
for XP\_010912140, which retained a tyrosine residue. All MFT proteins contained tryptophan at this specific site, indicating potential functional divergence due to single amino acid variations. Segment B (Fig. 3, box b) displayed consistent amino acid sequences in all FT proteins, which is consistent with the findings of Ahn et al. (2006). In the FT subclade I, a specific region contained 13 of 15 identical amino acids, indicating a possible conserved region.

In contrast, subclade II demonstrated less uniformity in this region. Segment C (Fig. 3, box c), which contained a distinct arrangement of LYN amino acids in FT proteins, but was not found in TFL or MFT proteins, was observed in both subclades of FT proteins, except for XP\_010912140. The conservation of segments B and C is important due to their role in forming a helix on the protein surface, which contributes to the formation of the central  $\beta$ -sheet. Segment C exhibited greater homology between TFL1 and FT homologs compared to segment B. However, a particular triad within FT remained predominantly consistent, distinguishing it from TFL1. This emphasizes the difficulty of modifying segment C and its crucial significance in protein structure and function (Ahn et al. 2006).

#### Modeling the protein structure

In this study, we employed the SWISS-MODEL platform to predict the structure of PEBP proteins, specifically those belonging to the FT subclades I, II, and TFL. To validate the accuracy of these models, we conducted a comparative analysis between the predicted structures and AtFT, a widely recognized protein structure (Table 2). The amino acid sequence similarity within FT-like subclade I varied from 70.83% to 73.4%, while the GMQE values ranged from 0.84 to 0.87. The GMQE scores for subclade II, similar to FT, ranged from 0.82 to 0.87, corresponding to sequence similarities ranging from 66.27% to 71.17%. The local quality estimates provided additional support for these results, as all amino acid residues scored above 0.6, indicating their suitability for the modeling process (Santhoshkumar and Yusuf 2020).

The high GMQE values and local quality estimates indicated structural similarity between the FT subclades I and II, which was also validated by protein modeling (Fig. 4a–c). The presence of



**Fig. 3** Complete Protein Sequence Alignments of reference protein sequences and all PEBP proteins in the palm. Horizontal arrows indicate the position of the PEBP domain. Block a indicates tyrosine residues, according to Hanzawa et al. (2005). Block b indicates the collection of fourteen amino acid



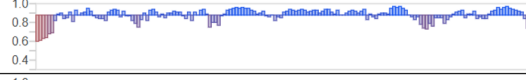




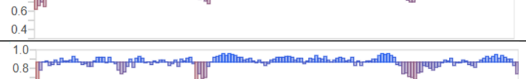

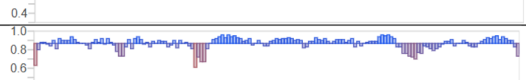





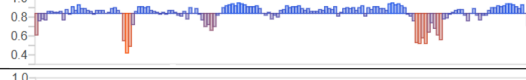


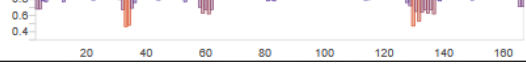

variations in amino acid residues within the three crucial conserved segments of these proteins suggests possible functional differences. Functional variations among homologous genes are frequently observed across a wide range of species. Mulki et al. (2018) discovered that the *Flowering Locus T3* in barley, which is associated with the *FT* gene, triggers spikelet growth but does not stimulate

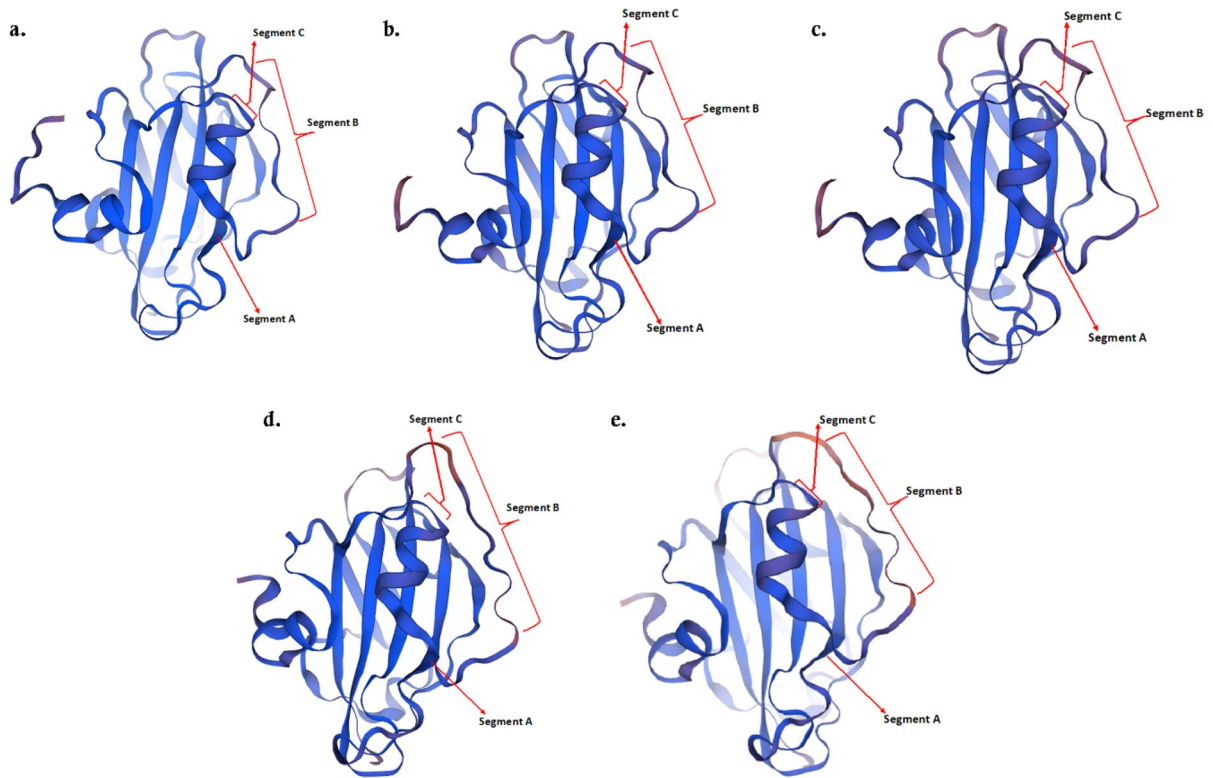
sequences reported by Ahn et al. (2006). Block c shows the regions corresponding to the LYN triad, as reported by Ahn et al. (2006). These characteristics are utilized to discern proteins exhibiting FT/Hd3a functionality

flowering. In contrast, Pieper et al. (2021) demonstrated that the *Flowering Locus T4* could postpone the flowering process in barley. These examples demonstrate the functional diversity that arises from variations in homologous gene families. Protein modulation mechanisms provide an explanation for changes in protein function that do not involve structural changes. One notable mechanism is



**Table 2** PEBP protein modeling of FT-like subclade I, FT-like subclade II, TFL-like, and MFT-like proteins compared to AtFT proteins

Group	Accessions	AtFT similarity	GMQE	QMEANDisCo Local
FT-like subclade I	XP_010925712 <i>EgHd3a</i> X2	72.89%	0.87	
	XP_010911427 <i>EgHd3a</i>	71.69%	0.86	
	XP_019705101 <i>EgHd3a</i> X2	70.83%	0.84	
	XP_038985085 <i>PdHd3a</i> X2	71.08%	0.86	
	XP_008786924 <i>PdHd3a</i>	71.08%	0.85	
	KAG1367690 <i>CnHd3a</i>	73.49%	0.87	
	CnHd3a*	72.89%	0.87	
FT-like subclade II	XP_010912140 <i>EgHd3a</i>	66.27%	0.87	
	XP_010930170 <i>EgHd3a</i>	69.33%	0.83	
	XP_010912811 <i>EgFT</i>	70.55%	0.82	
	XP_038986799 <i>PdFT</i>	71.17%	0.83	
	XP_008804208 <i>PdFT</i>	71.17%	0.83	
	KAG1331310 <i>CnFT</i>	68.71%	0.82	
TFL-like	XP_010907126 <i>EgSP</i>	73.10%	0.87	
	XP_010912272 <i>EgSP</i>	71.51%	0.87	
	XP_010919262 <i>EgCEN</i>	64.33%	0.85	
	XP_008788083 <i>PdSP</i>	73.10%	0.86	
	XP_008784004 <i>PdCEN</i>	72.67%	0.86	
	XP_038977978 <i>PdCEN</i>	66.08%	0.85	
	CnCEN*	65.12%	0.85	



**Fig. 4** Predicted protein structure models for **a** AtFT, **b** XP\_010925712 EgHd3a, **c** XP\_010912140 EgHd3a, **d** AtTFL1, and **e** XP\_010907126 EgSP were generated using SWISS-MODEL. A template search was performed against the

SWISS-MODEL template library through Blast and HHblits. Blue color denotes regions relatively similar to the reference proteins, while red indicates significant differences. (Color figure online)

allosteric modulation, in which the binding strength of a ligand at one site (allosteric site) influences the interaction of a different molecule at a distant site (active site) (McLeish et al. 2015). This process involves conformational changes in the protein structure, allowing it to respond dynamically to environmental cues or signaling molecules.

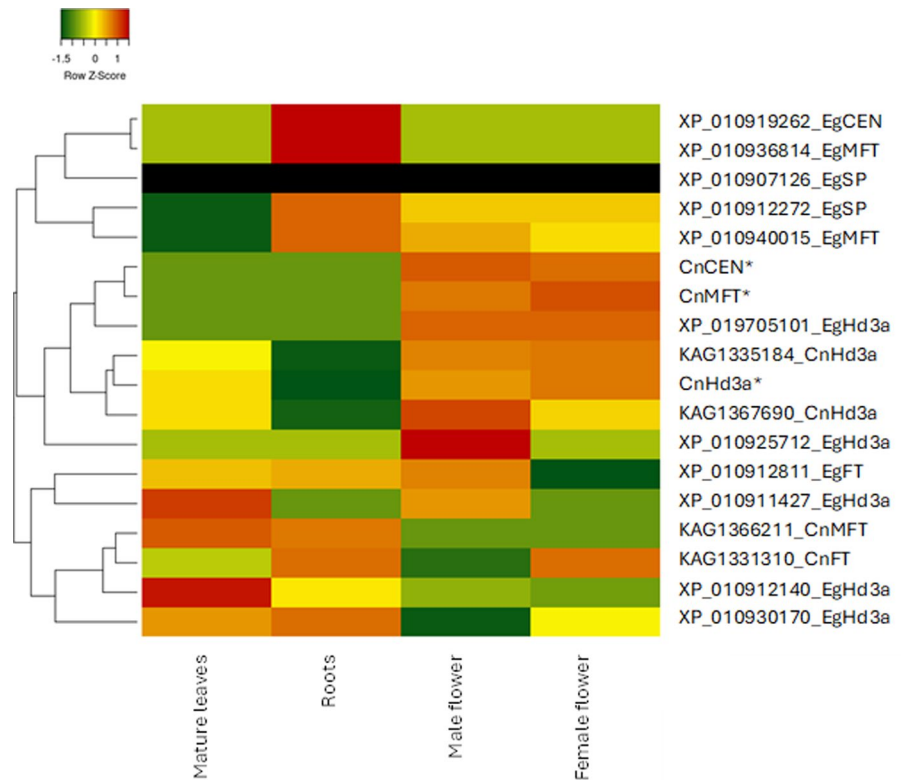
The GMQE values for the TFL members ranged from 0.85 to 0.87, with sequence similarities ranging from 64.33% to 73.10%. The analyzed protein structures exhibited a high degree of similarity to the reference structure AtTFL1, as indicated by the GMQE scores. Further analysis of the local quality estimate values revealed notable structural disparities in certain amino acid sequences, specifically at amino acid residues 33–36 and 134–138 (segment B). In these regions, the values dropped significantly below the threshold of 0.6, indicating significant structural deviations from the reference validated by protein modeling (Fig. 4d, e). These results confirm that either

segment (B or C) is adequate for TFL activity (Ahn et al. 2006).

#### In silico expression analysis

In silico expression analysis showed that the groups formed based on the expression patterns of genes (Fig. 5) were different from those formed based on the RNA sequences (Fig. 1a). For example, the gene encoding XP\_010911427 (*EgHd3a*) exhibited the highest expression levels in mature leaves, whereas its close relative in coconut (KAG1367690) showed high expression levels in male flowers. Data from four different tissues showed that three genes (XP\_010911427, KAG1366211, and XP\_010912140) were highly expressed in mature leaves, but their expression was very low in flowers. Six genes (XP\_010919262, XP\_010936814, XP\_010912272, XP\_010940015, KAG1331310, and XP\_010930170) were highly expressed in the roots, whereas the other

**Fig. 5** In silico heatmap expression profiling of PEBP genes in coconut and oil palm across seedling leaves, mature leaves, root, male flower, and female flower. Heat map color: red related to high expression, yellow to medium expression, and green to low expression. (Color figure online)



genes showed low expression. In addition, eight genes were highly expressed in the flowers, indicating their potential role in the reproductive phase to fertilization. XP\_010907126 showed very low or no expression. As the orthologs of rice Hd3a proteins, they may function in determining the initiation of flowering and panicle development (Zhao et al. 2015; Endo-Higashi and Izawa 2011). On the other hand, as the FT-antagonizing proteins, TFLs competed with Hd3a for complex formation to initiate or inhibit flowering. TFLs also redundantly functioned in inflorescence meristem development (Kaneko-Suzuki et al. 2018; Yoo et al. 2010). In addition, *MFT* induced earlier flowering in overexpressed *Arabidopsis* mutants (Yoo et al. 2003). Interestingly, *MFT* was highly expressed in seeds, especially in mid- to late-seed development in *Jatropha curcas* (Tao et al. 2014). The difference in expression patterns compared to nucleotide sequence patterns probably occurred because the data obtained from the NCBI database were not taken at a specific time, even though the timing of the initiation of flower development occurs precisely for these flowering genes. Most PEBP genes exhibit diurnal oscillation patterns (Yoo et al. 2010).

## Conclusion

This study significantly enhances our understanding of the genetic and evolutionary dynamics of the PEBP gene family in *Arecaceae*, particularly among oil palm, coconut, and date palm. It revealed a closer genetic affinity between oil palm and coconut than between oil palm and date palm, suggesting distinct evolutionary paths. This study categorizes PEBP genes into three main groups: *FT*, *TFL*, and *MFT*, and notes key structural similarities and functional differences within these groups. In silico expression analysis revealed that gene expression patterns differed from those of the RNA sequences. The difference in expression patterns may be due to the data not being obtained at specific times. Structural and sequence variations in these proteins may have important implications for their functional roles, highlighting the intricate relation between protein structure and function in plant biology. This comprehensive analysis of both protein and mRNA sequences offers valuable insights into the intricate nature of genetic evolution, highlighting the importance of gene duplication and diversification in these economically and ecologically

significant palm species and paving the way for future research in plant biology and genetics.

**Author contributions** EDJS supervised; AP, TH, M, S, and EDJS conceptualized; AP conducted the experiments; AP, DS, and TH analyzed the data; AP, TH, WE, and KK composed the manuscript; and EDJS oversaw. This article was reviewed and authorized by all authors.

**Funding** The authors declare that they received no funding or any other financial support during the preparation of this manuscript.

**Data availability** The authors affirm that the supplementary materials and/or articles contain data that support the findings of this study. The raw data used in this study can be obtained upon request to the corresponding author.

#### Declarations

**Conflict of interest** We declare no conflicts of financial or any other type of interest.

#### References

- Adam H, Jouannic S, Orieux Y, Morcillo F, Richaud F, Duval Y, Tregear JW (2007) Functional characterization of MADS box genes involved in the determination of oil palm flower structure. *J Exp Bot* 58(6):1245–1259. <https://doi.org/10.1093/jxb/erl263>
- Ahn JH, Miller D, Winter VJ, Banfield MJ, Lee JH, Yoo SY, Henz SR, Brady RL, Weigel D (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J* 25(3):605–614. <https://doi.org/10.1038/sj.emboj.7600950>
- Al-Ameri AA, Al-Quraini F, Gaafar ARZ, Khan S, Nadeem M (2016) Male specific gene expression in dioecious *Phoenix dactylifera* (Date palm) tree at flowering stage. *Pak J Bot* 48(1):131–135
- Barrett CF, Bacon CD, Antonelli A, Cano A, Hofmann T (2016) An introduction to plant phylogenomics with a focus on palms. *Bot J Linn Soc* 182(2):234–255. <https://doi.org/10.1111/boj.12399>
- Barrett CF, McKain MR, Sinn BT, Ge XJ, Zhang Y, Antonelli A, Bacon CD (2019) Ancient polyploidy and genome evolution in palms. *Genome Biol Evol* 11(5):1501–1511. <https://doi.org/10.1093/gbe/evz092>
- Bennett T, Dixon LE (2021) Asymmetric expansions of FT and TFL1 lineages characterize differential evolution of the EuPEBP family in the major angiosperm lineages. *BMC Biol* 19(1):181. <https://doi.org/10.1186/s12915-021-01128-8>
- Cássia-Silva C, Freitas CG, Lemes LP, Paterno GB, Dias PA, Bacon CD, Collevatti RG (2020) Higher evolutionary rates in life-history traits in insular than in mainland palms. *Sci Rep* 10(1):21125. <https://doi.org/10.1038/s41598-020-78267-5>
- Castañó F, Stauffer F, Marquinez X, Crèvecoeur M, Collin M, Pintaud JC, Tregear J (2014) Floral structure and development in the monoecious palm *Gaussia attenuata* (Arecaceae; Arecoideae). *Ann Bot* 114(7):1483–1495. <https://doi.org/10.1093/aob/mcu133>
- Chen Y, Xu X, Chen X, Chen Y, Zhang Z, Xuhan X, Lin Y, Lai Z (2018) Seed-specific gene *MOTHER of FT and TFL1 (MFT)* involved in embryogenesis, hormones and stress responses in *Dimocarpus longan* Lour. *Int J Mol Sci* 19(8):3–6. <https://doi.org/10.3390/ijms19082403>
- Comer JR, Zomlefer WB, Barrett CF, Stevenson DW, Heyduk K, Leebens-Mack JH (2016) Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). *Mol Phylogenet Evol* 97:32–42. <https://doi.org/10.1016/j.ympev.2015.12.015>
- Cronk QCB (2022) The distribution of sexual function in the flowering plant: from monoecy to dioecy. *Philos Trans Royal Soc London Ser B Biol Sci*. <https://doi.org/10.1098/rstb.2021.0486>
- Endo-Higashi N, Izawa T (2011) Flowering time genes *Heading date 1* and *Early heading date 1* together control panicle development in rice. *Plant Cell Physiol* 52(6):1083–1094. <https://doi.org/10.1093/pcp/pcr059>
- Fuertes MA, Rodrigo JR, Alonso C (2019) Conserved critical evolutionary gene structures in orthologs. *J Mol Evol* 87(2):93–105. <https://doi.org/10.1007/s00239-019-09889-1>
- Haberman A, Ackerman M, Crane O, Kelner JJ, Costes E, Samach A (2016) Different flowering response to various fruit loads in apple cultivars correlates with degree of transcript reaccumulation of a *TFL1*-encoding gene. *Plant J: Cell Mol Biol* 87(2):161–173. <https://doi.org/10.1111/tj.13190>
- Hanzawa Y, Money T, Bradley D (2005) A single amino acid converts a repressor to an activator of flowering. *PNAS* 102(21):7748–7753
- Herrig DK, Vertacnik KL, Ridenbaugh RD, Everson KM, Sim SB, Geib SM, Weisrock DW, Linnen CR (2023) Whole genomes reveal evolutionary relationships and mechanisms underlying gene-tree discordance in *Neodiprion sawflies*. *BioRxiv*. <https://doi.org/10.1101/2023.01.05.522922>
- Ho WWH, Weigel D (2014) Structural features determining flower-promoting activity of Arabidopsis *FLOWERING LOCUS T*. *Plant Cell* 26(2):552–564. <https://doi.org/10.1105/tpc.113.115220>
- Hoogenboom J, Fiers M, Immink R, Zuilhof H, Wennekes T (2016) Synthesis and evaluation of locostatin-based chemical probes towards PEBP-proteins. *Tetrahedron Lett* 57(22):2406–2409. <https://doi.org/10.1016/j.tetlet.2016.04.071>
- Kaneko-Suzuki M, Kurihara-Ishikawa R, Okushita-Terakawa C, Kojima C, Nagano-Fujiwara M, Ohki I, Tsuji H, Shimamoto K, Taoka KI (2018) TFL1-like proteins in rice antagonize rice FT-like protein in inflorescence development by competition for complex formation with 14–3-3 and FD. *Plant Cell Physiol* 59(3):458–468. <https://doi.org/10.1093/pcp/pcy021>
- Khan AL, Asaf S, Lee JJ, Al-Harrasi A, Al-Rawahi A (2018) First chloroplast genomics study of *Phoenix dactylifera* (var. Naghal and Khanezi): a comparative analysis. *PLoS*

- ONE 13(7):e0200104. <https://doi.org/10.1371/journal.pone.0200104>
- Książkiewicz M, Rychel S, Nelson MN, Wyrwa K, Naganowska B, Wolko B (2016) Expansion of the phosphatidylethanolamine binding protein family in legumes A case study of *Lupinus angustifolius* L. FLOWERING LOCUS T homologs, LanFTc1 and LanFTc2. BMC Genom 17(1):820. <https://doi.org/10.1186/s12864-016-3150-z>
- Lee SH, Choi CW, Park KM, Jung WH, Chun HJ, Baek D, Cho HM, Jin BJ, Park MS, Lim LH, Shim SI, Chung JI, Kim MC (2021) Diversification in functions and expressions of soybean *FLOWERING LOCUS T* genes fine-tunes seasonal flowering. Front Plant Sci 12:613675. <https://doi.org/10.3389/fpls.2021.613675>
- Leeggangers HACF, Rosilio-Brami T, Bigas-Nadal J, Rubin N, van Dijk ADJ, de Caceres N, Gonzalez FF, Saadon-Shirit S, Nijveen H, Hilhorst HWM, Immink RGH, Zaccari M (2018) *Tulipa gesneriana* and *Lilium longiflorum* PEBP genes and their putative roles in flowering time control. Plant Cell Physiol 59(1):90–106. <https://doi.org/10.1093/pcp/pcx164>
- Li C, Zhang Y, Zhang K, Guo D, Cui B, Wang X, Huang X (2015) Promoting flowering, lateral shoot outgrowth, leaf development, and flower abscission in tobacco plants overexpressing cotton *FLOWERING LOCUS T* (*FT*)-like gene *GhFT1*. Front Plant Sci 6:454. <https://doi.org/10.3389/fpls.2015.00454>
- McLeish TCB, Cann MJ, Rodgers TL (2015) Dynamic transmission of protein allostery without structural change: spatial pathways or global modes? Biophys J 109(6):1240–1250. <https://doi.org/10.1016/j.bpj.2015.08.009>
- Moutinho AF, Eyre-Walker A (2022) The silent impact: codon usage bias and protein evolution in bacteria. BioRxiv. <https://doi.org/10.1101/2022.07.07.499116>
- Mulki MA, Bi X, von Korff M (2018) Flowering locus T3 controls spikelet initiation but not floral development. Plant Physiol 178(3):1170–1186. <https://doi.org/10.1104/pp.18.00236>
- Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, Silva A, Bataillon T, Widmer A, Glémin S, Touzet P, Marais GAB (2021) Dioecy is associated with high genetic diversity and adaptation rates in the plant genus silene. Mol Biol Evol 38(3):805–818. <https://doi.org/10.1093/molbev/msaa229>
- Nadot S, Alapetite E, Baker WJ, Tregear JW, Barfod AS (2016) The palm family (Arecaceae): a microcosm of sexual system evolution. Bot J Linn Soc 182(2):376–388. <https://doi.org/10.1111/boj.12440>
- Nakano Y, Higuchi Y, Yoshida Y, Hisamatsu T (2015) Environmental responses of the *FT/TFL1* gene family and their involvement in flower induction in *Fragaria xananassa*. J Plant Physiol 177:60–66. <https://doi.org/10.1016/j.jplph.2015.01.007>
- Nayar NM (2018) Opportunistic flowering and pollinating system as a survival strategy of the coconut (*Cocos nucifera* L.), a monotypic species of the Arecaceae family. Genet Resour Crop Evolut 65(1):333–342. <https://doi.org/10.1007/s10722-017-0537-x>
- Pieper R, Tomé F, Pankin A, von Korff M (2021) FLOWERING LOCUS T4 delays flowering and decreases floret fertility in barley. J Exp Bot 72(1):107–121. <https://doi.org/10.1093/jxb/eraa466>
- Pin PA, Benlloch R, Bonnet D, Wremeth-Weich E, Kraft T, Gielen JJLL, Nilsson O (2010) An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. Science 330(6009):1397–1400. <https://doi.org/10.1126/science.1197004>
- Polosoro A, Enggarini W, Hadiarto T, Supena EDJ, Suharsono (2021) In silico screening of oil palm early and continuously flowering gene candidates for faster breeding program. IOP Conf Ser: Earth Environ Sci 762(1):012063. <https://doi.org/10.1088/1755-1315/762/1/012063>
- Santhoshkumar R, Yusuf A (2020) In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of *Curcuma longa*. J Genet Eng Biotechnol 18(24):1–9. <https://doi.org/10.1186/s43141-020-00041-x>
- Shan K, Wu C, Tang X, Lu R, Tan W, Lu J (2022) Molecular evolution of protein sequences and codon usage in monkeypox viruses. BioRxiv. <https://doi.org/10.1101/2022.12.23.521708>
- Tao YB, Luo L, He LL, Ni J, Xu ZF (2014) A promoter analysis of *MOTHER OF FT AND TFL1 1* (*JcMFT1*), a seed-preferential gene from the biofuel plant *Jatropha curcas*. J Plant Res 127(4):513–524. <https://doi.org/10.1007/s10265-014-0639-x>
- Tribhuvan KU, Das A, Srivastava H, Kumar K, Durgesh K, Mithra SVA, Sandhya, Jain PK, Gaikwad K (2020) Identification and characterization of PEBP family genes reveal CeFT8 a probable candidate for photoperiod insensitivity in *C. cajan*. 3 Biotech 10(5):1–12. <https://doi.org/10.1007/s13205-020-02180-x>
- Venail J, Da Silva Santos PH, Manechini JR, Alves LC, Scarpari M, Falcão T, Romanel E, Brito M, Vicentini R, Pinto L, Jackson SD (2022) Analysis of the PEBP gene family and identification of a novel FLOWERING LOCUS T orthologue in sugarcane. J Exp Bot 73(7):2035–2049. <https://doi.org/10.1093/jxb/erab539>
- Wang X, Liu Y, Yang K, Wei X, Wang X (2016) Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic *FLOWERING LOCUS T* gene homologs in gymnosperms. New Phytol 1:1–15. <https://doi.org/10.1111/nph.14066>
- 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>
- Wang M, Tan Y, Cai C, Zhang B (2019) Identification and expression analysis of phosphatidylethanolamine-binding protein (PEBP) gene family in cotton. Genomics 111(6):1373–1380. <https://doi.org/10.1016/j.ygeno.2018.09.009>
- Xu K, Lin C, Lee SY, Mao L, Meng K (2022) Comparative analysis of complete *Ilex* (Aquifoliaceae) chloroplast genomes: insights into evolutionary dynamics and phylogenetic relationships. BMC Genom 23(1):203. <https://doi.org/10.1186/s12864-022-08397-9>
- Yang Z, Chen L, Kohnen MV, Xiong B, Zhen X, Liao J, Oka Y, Zhu Q, Gu L, Lin C, Liu B (2019) Identification and characterization of the PEBP family genes in moso bamboo

- (*Phyllostachys heterocycla*). Sci Rep 9(1):14998. <https://doi.org/10.1038/s41598-019-51278-7>
- Yıldırım B, Vogl C (2023) Purifying selection against spurious splicing signals contributes to the base composition evolution of the polypyrimidine tract. J Evol Biol 36(9):1295–1312. <https://doi.org/10.1101/2022.01.29.478184>
- Yoo SY, Kardailsky I, Lee JS, Weigel D, Ahn JH (2003) Acceleration of flowering by overexpression of *MFT* (*MOTHER OF FT AND TFL1*). Mol Cells 17(1):95–101
- 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: Cell Mole Biol 63(2):241–253. <https://doi.org/10.1111/j.1365-313X.2010.04234.x>
- Zhang X, Wang C, Pang C, Wei H, Wang H, Song M, Fan S, Yu S (2016) Characterization and functional analysis of PEBP Family genes in upland cotton (*Gossypium hirsutum* L). PLoS ONE 11(8):1–20. <https://doi.org/10.1371/journal.pone.0161080>
- Zhao J, Chen H, Ren D, Tang H, Qiu R, Feng J, Long Y, Niu B, Chen D, Zhong T, Liu YG, Guo J (2015) Genetic interactions between diverged alleles of *early heading date 1* (*Ehd1*) and *Heading date 3a* (*Hd3a*)/*RICE FLOWERING LOCUS T1* (*RFT1*) control differential heading and contribute to regional adaptation in rice (*Oryza sativa*). New Phytol 208(3):936–948. <https://doi.org/10.1111/nph.13503>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.