Phylogenomic Analysis of the PEBP Gene Family from Kalanchoë

Mackenzie, Kathryn; Lopes Coelho, Lívia; Lütken, Henrik Vlk; Müller, Renate

Published in:
Agronomy

DOI:
10.3390/agronomy9040171

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Phylogenomic Analysis of the *PEBP* Gene Family from Kalanchoë

Kathryn Kuligowska Mackenzie 1,2,*, Lívia Lopes Coelho 1, Henrik Lütken 1 and Renate Müller 1

1 Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Højbakkegård Alle 9-13, 2630 Taastrup, Denmark; livia@plen.ku.dk (L.L.C.); hlm@plen.ku.dk (H.L.); ren@plen.ku.dk (R.M.)
2 Department of Agricultural Sciences, Faculty of Agriculture and Forestry, University of Helsinki, Latokartanonkaari 7, P.O. Box 27, 00014 Helsinki, Finland
* Correspondence: kathryn.mackenzie@helsinki.fi

Received: 15 February 2019; Accepted: 25 March 2019; Published: 30 March 2019

**Abstract:** The *PEBP* family comprises proteins that function as key regulators of flowering time throughout the plant kingdom and they also regulate growth and plant architecture. Within the *PEBP* protein family, three subfamilies can be distinguished in angiosperms: MOTHER OF FT AND TFL1-like (MFT), FLOWERING LOCUS T-like (FT-like), and TERMINAL FLOWER1-like (TFL1-like). Taking advantage of the genome sequences available from *K. fedtschenkoi* and *K. laxiflora*, we performed computational analysis to identify the members of the *PEBP* gene family in these species. The analyses revealed the existence of 11 *PEBP* genes in *K. fedtschenkoi* and 18 in *K. laxiflora*, which are clustered in two clades: FT-like and TFL1-like. The *PEBP* genes had conserved gene structure and the proteins had highly conserved amino acid sequences in the positions crucial for the protein functions. The analysis of Ka/Ks ratio revealed that most recently duplicated genes are under positive selection. Despite being an economically important genus, the genetics underlying the regulation of flowering in Kalanchoë is poorly understood. The results of this study may provide a new insight into the molecular control of flowering that will allow further studies on flowering control in Kalanchoë.

**Keywords:** BROTHER OF FT AND TFL1; gene duplication; Kalanchoë fedtschenkoi; Kalanchoë laxiflora; FLOWERING LOCUS T; molecular phylogeny; TERMINAL FLOWER 1

1. Introduction

The family of phosphatidylethanolamine-binding proteins (*PEBPs*) is a group of proteins present in all eukaryote kingdoms. Despite extensive sequence conservation, *PEBP* proteins act as regulators of various signaling pathways to control growth and differentiation [1,2]. Phylogenetic studies suggest that the *PEBP* gene family can be divided into three subfamilies: MOTHER OF FT AND TFL1-like (MFT-like), FLOWERING LOCUS T-like (FT-like), and TERMINAL FLOWER 1-like (TFL1-like) [1,3]. It is thought that the MFT-like clade is the evolutionary ancestor to the FT-like and TFL1-like clades. A duplication of an ancestral MFT-like gene might have given rise to the MFT-like clade and the FT/TFL1-like clade. Further diversification of function and a second duplication resulted in the emergence of separate FT-like and TFL1-like clades [3,4]. A recent study revealed that the MFT-like subfamily exists in both basal land plants (bryophytes and pteridophytes) and seed plants (gymnosperms and angiosperms), while FT-like and TFL1-like genes are only found in gymnosperm and angiosperms. This suggests that the first duplication event took place after the divergence of basal land plants from the common ancestor of seed plants, while the second duplication event occurred before the divergence of seed plants. Additionally, within the three
subfamilies, more recent gene and/or genome duplications were observed in both angiosperms and
gymnosperms further expanding the PEBP gene family [3].

The most studied functions of the PEBP gene family members concern the involvement of FT
and TFL1 homologs in controlling flowering time. Despite the high amino acid similarity FT and TFL1
proteins have opposite activities: the FT protein can act as florigen [5] moving in the phloem from
leaves to the shoot apex, while TFL1 functions as a repressor in the shoot apex [6]. This antagonistic
activity requires interaction with a common partner the bZIP transcription factor FD [7,8]. The
complex of FD with FT/TFL1 likely contains other proteins including 14-3-3 proteins that mediate
these interactions [9].

Although in angiosperms the FT- and TFL1-like genes were previously thought to be primarily
involved in the control of the transition to reproductive development, recent studies on perennial
species suggest a more general role in controlling the growth and termination of meristems. The TFL1
ancestor underwent two separate duplication events in the common ancestor of angiosperms, which
created three lineages corresponding to TFL1, BROTHER OF FT AND TFL1 (BFT), and the
CENTRORADIALIS homolog (ATC or CEN) [10]. The TFL1/BFT/CEN-like genes control inflorescence
meristem identity and delay transition to the reproductive phase [11] and are involved in growth and
dormancy cycles [12], and seasonality of flowering in perennials [13,14]. Furthermore, the proteins
from the TFL1/CEN/BFT-like subfamily can act as anti-florigen, a transmissible flowering repressor
[15]. In many plant species, numerous FT-like genes arose from gene duplication events that might
have led to subfunctionalization or neofunctionalization [16]. The FT-like genes were demonstrated
to function in flower repression [17,18], vegetative growth [19], and storage organ formation [20,21].
Among angiosperms, MFT-like genes are thought to have a conserved function in regulation of seed
germination via abscisic acid and gibberellic acid signaling pathways [22–25].

The Kalanchoë genus comprises ~140 species distributed on Madagascar, Southern and Eastern
Africa, and to some extent, tropical Africa, the Arabian Peninsula, and Southern Asia. The species are
mainly perennial succulent shrublets or shrubs, rarely small trees. However, they can also be
perennial to biennial or rarely annual herbs [26]. Economically, the genus ranks as the second most
important group of potted plants in Europe mainly due to high popularity of Kalanchoë blossfeldiana
and its interspecific hybrids. The genus presents a wide range of attractive traits that can be of
commercial value [26,27]. However, numerous species are difficult both, to induce flowering and
control the time of flowering [28–30]. Generally, the flowering induction in Kalanchoë is determined
by photoperiod. Within the genus two photoperiodic groups have been identified in respect to
requirement for induction of flowering. They include short day (SD) plants, i.e., plants that are flower
induced when exposed to a period of short days, and long-short day (LSD) plants, i.e., plants that
require a dual sequence of photoperiods, which include species belonging to the former Bryophyllum
genus sensu stricto [26]. There are, however, other factors, such as temperature and light intensity,
that can greatly influence flower induction in Kalanchoë [31,32]. Moreover, some species have a long
juvenile phase [33,34]. Apart from the significance of Kalanchoë as ornamental plants, the species of
this genus have long been viewed as important models for the study of ecologically relevant
modification of photosynthesis; the Crassulacean Acid Metabolism (CAM). Kalanchoë fedtschenkoi is
now viewed as an emerging model system for functional genomics of CAM. Currently, genome
sequences are available for K. fedtschenkoi and K. laxiflora, members of Bryophyllum section (Kalanchoe
fedtschenkoi v1.1 and Kalanchoe laxiflora v1.1, DOE-JGI) [35] being the first sequenced species in the
eudicot lineage Saxifragales. Interestingly, both induction of flowering and switch from C3
photosynthesis to CAM can be induced by photoperiod in Kalanchoë species; however, it is unknown
if the output pathways of these processes are interconnected [36]. The members of the Kalanchoë
genus include also several species with a broad range of ethnomedicinal uses. Therapeutic action of
Kalanchoë plants is attributed to bufadienolides, a group of poly-hydroxy steroid hormones,
displaying pharmacological activities such as anticancer, anti-inflammatory and cardioactive effects
[37].

In this study, we performed computational analysis to identify members of the PEBP gene family
from two Kalanchoë species taking advantage of available genome sequence data. We have
systematically analyzed the gene structure, gene family evolution and protein attributes to identify relevant targets for future functional genomic studies.

2. Materials and Methods

2.1. Identification of PEBP Sequences

The sequences were obtained by annotation and using full-length *A. thaliana* sequences as query sequences in BLASTp and TBLASTX searches against genomes and proteomes of *K. fedtschenkoi* Raym.-Hamel & H. Perrier and *K. laxiflora* Baker (Phytozome: *Kalanchoe fedtschenkoi* v1.1, *Kalanchoe laxiflora* v1.1, DOE-JGI). Additionally, recovered Kalanchoë protein sequences were used as query in BLASTp searches against Kalanchoë proteomes. The *A. thaliana* sequences were obtained from GenBank: *AtFT* (NM_001334207.1), *AtFTL1* (NM_120465.3), *AtTSF* (NM_118156.2), *AtATC* (NM_128315.4), *AtBFT* (NM_125597.2), *AtMFT* (NM_101672.4) and proteins: *AtFT* (BAA77838.1), *AtFTL1* (NP_196004.1), *AtTSF* (NP_193770.1), *AtATC* (NP_180324.1), *AtBFT* (NP_201010.1), and *AtMFT* (NP_173250.1). The Kalanchoë genes were named as *KfFT* to *KfFT7*, *KfFTL1* to *KfFTL1.3*, and *KfBFT1* for *K. fedtschenkoi*, and *KfTFL1* to *KfTFL1.1* to *KfTFL1.4*, and *KfBFT1* to *KfBFT2* for *K. laxiflora*, based on their accession IDs and database gene annotations (Table 1). Obtained protein sequences were analyzed for the presence of PEBP domains using the CDD database [38].

The sequences were divided into four groups according to the sequence identity for each domain (Table 1). The sequences were further divided into four groups according to the sequence identity for each domain (Table 1).

### Table 1. Characterization of PEBP genes and proteins from Kalanchoë.

<table>
<thead>
<tr>
<th>No.</th>
<th>Accession IDs</th>
<th>Gene Name</th>
<th>No. of Transcripts</th>
<th>Protein Length (aa)</th>
<th>Molecular Weight (kDa)</th>
<th>PI</th>
<th>Instability Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kaladp0007s0057</td>
<td><em>KfFT1</em></td>
<td>1</td>
<td>177</td>
<td>20.21</td>
<td>8.88</td>
<td>41.81; U</td>
</tr>
<tr>
<td>2</td>
<td>Kaladp0031s0098</td>
<td><em>KfFT2</em></td>
<td>1</td>
<td>181</td>
<td>20.31</td>
<td>8.75</td>
<td>44.11; U</td>
</tr>
<tr>
<td>3</td>
<td>Kaladp0055s0469</td>
<td><em>KfFT3</em></td>
<td>1</td>
<td>178</td>
<td>20.21</td>
<td>9.09</td>
<td>38.93; S</td>
</tr>
<tr>
<td>4</td>
<td>Kaladp0055s0470</td>
<td><em>KfFT4</em></td>
<td>1</td>
<td>176</td>
<td>19.88</td>
<td>7.72</td>
<td>38.51; S</td>
</tr>
<tr>
<td>5</td>
<td>Kaladp0060s0252</td>
<td><em>KfFT5</em></td>
<td>1</td>
<td>180</td>
<td>20.28</td>
<td>8.39</td>
<td>38.23; S</td>
</tr>
<tr>
<td>6</td>
<td>Kaladp0099s0141</td>
<td><em>KfFT6</em></td>
<td>1</td>
<td>181</td>
<td>20.41</td>
<td>7.73</td>
<td>51.05; U</td>
</tr>
<tr>
<td>7</td>
<td>Kaladp1016s0006</td>
<td><em>KfFT7</em></td>
<td>1</td>
<td>177</td>
<td>19.97</td>
<td>8.76</td>
<td>44.12; U</td>
</tr>
<tr>
<td>8</td>
<td>Kalax.0016s0040</td>
<td><em>KfFT1</em></td>
<td>1</td>
<td>180</td>
<td>20.27</td>
<td>8.39</td>
<td>38.23; S</td>
</tr>
<tr>
<td>9</td>
<td>Kalax.0052s0013</td>
<td><em>KfFT2</em></td>
<td>1</td>
<td>203</td>
<td>22.90</td>
<td>7.47</td>
<td>44.14; U</td>
</tr>
<tr>
<td>10</td>
<td>Kalax.0084s0028</td>
<td><em>KfFT3</em></td>
<td>1</td>
<td>177</td>
<td>20.21</td>
<td>8.88</td>
<td>41.81; U</td>
</tr>
<tr>
<td>11</td>
<td>Kalax.0100s0001</td>
<td><em>KfFT4</em></td>
<td>1</td>
<td>177</td>
<td>20.02</td>
<td>7.72</td>
<td>37.86; S</td>
</tr>
<tr>
<td>12</td>
<td>Kalax.0100s0002</td>
<td><em>KfFT5</em></td>
<td>1</td>
<td>177</td>
<td>19.97</td>
<td>9.27</td>
<td>41.29; U</td>
</tr>
<tr>
<td>13</td>
<td>Kalax.0114s0057</td>
<td><em>KfFT6</em></td>
<td>1</td>
<td>177</td>
<td>19.96</td>
<td>8.76</td>
<td>43.51; U</td>
</tr>
<tr>
<td>14</td>
<td>Kalax.0189s0069</td>
<td><em>KfFT7</em></td>
<td>1</td>
<td>177</td>
<td>20.25</td>
<td>8.88</td>
<td>42.24; U</td>
</tr>
<tr>
<td>15</td>
<td>Kalax.0399s0008</td>
<td><em>KfFT8</em></td>
<td>1</td>
<td>181</td>
<td>20.39</td>
<td>7.73</td>
<td>50.63; U</td>
</tr>
<tr>
<td>16</td>
<td>Kalax.0435s0026</td>
<td><em>KfFT9</em></td>
<td>1</td>
<td>181</td>
<td>20.41</td>
<td>7.73</td>
<td>51.05; U</td>
</tr>
<tr>
<td>17</td>
<td>Kalax.0644s0009</td>
<td><em>KfFT10</em></td>
<td>1</td>
<td>178</td>
<td>20.21</td>
<td>9.09</td>
<td>38.93; S</td>
</tr>
<tr>
<td>18</td>
<td>Kalax.0798s0001</td>
<td><em>KfTFL1</em></td>
<td>1</td>
<td>176</td>
<td>19.88</td>
<td>7.72</td>
<td>38.51; S</td>
</tr>
<tr>
<td>19</td>
<td>Kalax.0508s0010</td>
<td><em>KfTFL1</em></td>
<td>1</td>
<td>112</td>
<td>12.28</td>
<td>5.24</td>
<td>32.41; S</td>
</tr>
<tr>
<td>20</td>
<td>Kaladp0011s0878</td>
<td><em>KfTFL1.1</em></td>
<td>1</td>
<td>179</td>
<td>20.39</td>
<td>8.35</td>
<td>33.13; S</td>
</tr>
<tr>
<td>21</td>
<td>Kaladp0099s0128</td>
<td><em>KfTFL1.2</em></td>
<td>1</td>
<td>175</td>
<td>19.89</td>
<td>8.73</td>
<td>54.72; U</td>
</tr>
<tr>
<td>22</td>
<td>Kaladp0099s0167</td>
<td><em>KfTFL1.3</em></td>
<td>1</td>
<td>173</td>
<td>19.39</td>
<td>7.96</td>
<td>46.78; U</td>
</tr>
<tr>
<td>23</td>
<td>Kalax.0010s0108</td>
<td><em>KfTFL1.1</em></td>
<td>1</td>
<td>179</td>
<td>20.36</td>
<td>8.35</td>
<td>27.75; U</td>
</tr>
<tr>
<td>24</td>
<td>Kalax.0018s0096</td>
<td><em>KfTFL1.2</em></td>
<td>1</td>
<td>173</td>
<td>19.39</td>
<td>7.96</td>
<td>46.78; U</td>
</tr>
<tr>
<td>25</td>
<td>Kalax.0073s0028</td>
<td><em>KfTFL1.3</em></td>
<td>1</td>
<td>175</td>
<td>19.88</td>
<td>8.73</td>
<td>54.72; U</td>
</tr>
<tr>
<td>26</td>
<td>Kalax.0154s0031</td>
<td><em>KfTFL1.4</em></td>
<td>1</td>
<td>175</td>
<td>19.88</td>
<td>8.73</td>
<td>54.72; U</td>
</tr>
<tr>
<td>27</td>
<td>Kaladp0096s0031</td>
<td><em>KfBFT1</em></td>
<td>1</td>
<td>177</td>
<td>20.06</td>
<td>9.16</td>
<td>50.66; U</td>
</tr>
<tr>
<td>28</td>
<td>Kalax.0003s0185</td>
<td><em>KfBFT1</em></td>
<td>1</td>
<td>176</td>
<td>19.88</td>
<td>7.93</td>
<td>49.42; U</td>
</tr>
<tr>
<td>29</td>
<td>Kaladp0635s0001</td>
<td><em>KfBFT2</em></td>
<td>1</td>
<td>176</td>
<td>19.85</td>
<td>9.16</td>
<td>48.94; U</td>
</tr>
</tbody>
</table>

*U*: unstable, *S*: stable

2.2. Phylogenetic Analysis
Alignment of the PEBP domains from Kalanchoë PEBP genes was conducted using ClustalW. Multiple sequence alignment of PEBP proteins of Kalanchoë species, Arabidopsis thaliana (BAA77838.1, NP_196004.1, NP_201010.1, NP_173250.1, NP_193770.1, NP_180324.1), Oryza sativa (BAO02979.1, BAO03159.1, Q656A5, Q9ASJ1), Chrysanthemum seticuspe (BAL14659.1, BAN89465.1), Fragaria vesca (NP_001266951.1, AEP23097.1), Malus domestica (ADP69290.1, ACL98164.1, BAG31959.1, BAD06418.1, BAG31957.1, BAG31958.1), and Populus species (AFU08239.1, AFU08240.1, Q6TXM3, B9HPZ6, Q2PPJ2, B9ID58) was performed using Clustal Omega. MEGA 7.0 software was used to build the neighbor-joining phylogenetic tree from the protein and gene sequence alignment using the following parameters; p-distance model, pairwise gap deletion, and 1000 bootstraps. DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) analysis [39] was performed to determine the homologous relationships among PEBP proteins from Kalanchoë and other species.

2.3. Protein Characterization

The protein analysis was performed using ProtParam [40] for prediction of protein length, molecular weight, pi and instability index.

To identify conserved motifs in PEBP family proteins, the online MEME v.4.12.0 (Multiple Expectation Maximization for Motif Elicitation) [41] program was used with minimum motif length of 6 and maximum of 100, the maximum number of motifs was set to 20. Protein structure prediction was performed using SWISSMODEL and Swiss PdbViewer [42].

2.4. Gene Structure Analysis

The information about the gene length and distribution of exons and introns in PEBP genes was obtained from Phytozome. The intron/exon organization for PEBP genes was determined by aligning the CDS sequences to their corresponding genomic DNA sequences and using the result as the input for graphical display at the Gene Structure Display Server v2 [43].

Gene CDS sequences were aligned and percent identity matrix was generated using Clustal Omega with default parameters [44].

2.5. Gene Duplication Analysis

MEGA 7.0 software [45] was used to calculate rate of nonsynonymous substitutions (Ka) and synonymous substitutions (Ks). Codon-based testing of purifying selection for analysis between sequences was conducted using the modified Nei-Gojobori (assumed transition/transition bias = 2) method. All ambiguous positions were removed for each sequence pair. The dates of the duplication events (T—Duplication time; Mya—Million years ago) between the duplicated Kalanchoë genes associated with terminal branches in the tree species clades were calculated by the equation [46]

$$T = \frac{Ks}{2\lambda} \times 10^{-6} \text{Mya}, \text{the } \lambda = 1.5 \times 10^{-8}$$

Diversity analysis (nonsynonymous substitutions per nonsynonymous site; Ka/synonymous substitutions per synonymous site—Ks) was performed using the DnaSP v5.10 program [47] with a sliding window mode (window size 50, step 10).

3. Results

3.1. Identification of PEBP Sequences

BLAST searches using PEBP-like nucleotide and protein sequences from A. thaliana against plant protein and nucleotide databases of two Kalanchoë species resulted in 29 PEBP-like genes being retrieved; predicted to encode 30 proteins. There were 11 genes in K. fedtschenkoi that is a diploid species and 18 in K. laxiflora that is a tetraploid species. Based on annotations provided by the Phytozome database the genes and corresponding proteins were assigned to three groups: FT-like
(seven genes from *K. fedtschenkoi* and 12 from *K. laxiflora*), TFL1-like (three genes from *K. fedtschenkoi* and four from *K. laxiflora*), and BFT-like (one gene from *K. fedtschenkoi* and two from *K. laxiflora*). The detailed information about Phytozome gene IDs assigned gene names and number of transcripts are included in Table 1.

3.2. Comparative Phylogenetic Analysis

The unrooted neighbor-joining phylogenetic tree was constructed using predicted amino acid protein sequences from Kalanchoë and six other species in which functional assessment of FT and TFL1 functions was confirmed by transgenic approach [4,15,48] (Figure 1).

The comparative phylogenetic analysis confirmed that *PEBP*-like proteins from Kalanchoë are homologous to FT-like, TFL1-like and CEN-like, and BFT-like proteins from other species, and that MFT-like sequences could not be identified in Kalanchoë.
The gene homology analysis using DELTA-BLAST further confirmed that Kalanchoë *FT*-like, *TFL1*-like, and *BFT*-like genes are homologous to the *A. thaliana FT* gene (69–78% sequence similarity), the CEN/TFL genes (62–77% sequence similarity), and the *BFT* gene (66–67% sequence similarity), respectively (Table S1). Furthermore, the homology analysis comparing the Kalanchoë *PEBP* genes and sequences available in the Genbank for various species confirmed that Kalanchoë sequences have high homology with perennial species (Table S2).

3.3. Kalanchoë Protein Characterization

The length of the proteins ranged from 112 to 203 aa (average: 176 aa; median: 177 aa) in *FT*-like proteins and 145 to 179 aa (average: 172 aa; median: 175 aa) in *TFL1*-like proteins. The *BFT*-like proteins were either 176 or 177 aa. All identified proteins were characterized by a conserved *PEBP* domain. Domain analysis of primary transcript results from the CDD database confirmed the presence of the *PEBP* domain in the N-terminal regions of these proteins, except for KlFT1 that contained only a partial *PEBP* domain. The multiple alignment of domain sequences followed by phylogenetic analysis confirmed the presence of three protein groups corresponding to *FT*-like, *TFL1*-like, and *BFT*-like proteins (Figure 2).

![Figure 2. Amino acid alignment of deduced *PEBP* family protein sequences from Kalanchoë and *Arabidopsis thaliana FT* proteins. The colors indicate amino acids of different biochemical properties as obtained through MEGA 7.0. The sequences were aligned using ClustalW. Alignment of amino acid sequences of *PEBP* proteins at the 14-3-3 interaction interface is underlined in pink. The conserved DPDXP (Asp-Pro-Asp-X-Pro) and GIHR (Gly-Ile-His-Arg) motifs that compose anion-binding sites are underlined in orange. The conserved segment B (positions 128–141, exon 4) is underlined in blue and segment C (149–151, exon 4—XYN triad) is underlined in green. Amino acid positions indicated on the top of the sequence alignment are based on the *Arabidopsis thaliana FT* protein sequence BAA77838.1. Bold amino acid positions distinguish residues that are considered crucial for *PEBP* flowering inductive and repressive functions and 14-3-3 protein binding according to references [4,7,8,9,49]. In the present alignment arginine at position 62 (R62), proline at position 94 (P94), phenylalanine at position 101 (F101), and arginine at position 140 (R140) are predicted to participate in binding between *PEBP* proteins and 14-3-3 proteins. Tyrosine at position 85 (Y85) is present in all FT-like proteins, while histidine at position 85 (H85) is present in *TFL1*-like proteins. FT-like proteins contain tyrosine at position 134 (Y134), while *TFL1*-like proteins contain non-tyrosine amino acids. All FT-like proteins contain tryptophan at position 138 (W138), while *TFL1*-like contain non-tryptophan amino acids. At position 140, FT-like proteins contain glutamine (Q140), while *TFL1*-like proteins contain aspartic acid (D140) or glutamic acid (E140).

The Kalanchoë *PEBP* proteins shared 40% to 100% identity at the amino acid sequence level (Table S3). The FT-like protein identity ranged between 81% to 96% in *K. fedtschenkoi* and 71% to 99%
in K. laxiflora. There was between 72% to 100% sequence identity among FT-like proteins between both species. The TFL1-like protein identity ranged between 65% to 69% in K. fedtschenkoi and from 50% to 100% in K. laxiflora. There was between 51% to 100% sequence identity among TFL1-like proteins between both species. The BFT-like proteins in K. laxiflora showed 98% identity, while between species the identity was 98%.

The PEBP proteins had highly conserved amino acid sequences in the positions crucial for the protein functions (description in Figure 2). The differences in the conserved residues were observed in KfTFL1.1 and KfTFL1.1a at position 62 where lysine (K) was identified instead of arginine (R) that has, however, similar properties, and a deletion of the second aspartic acid (D) in the DPDXP motif that forms the anion-binding site was observed in the KfTFL1.4 protein. In addition, in BFT-like proteins the arginine residues were observed at positions 129 and 131, but not at position 130.

The molecular weights of the predicted molecules were ~19–23 kDa for all proteins except from KfTFL1 (12.28 kDa) and KfTFL1.1b (15.89 kDa). The pIs were 5.24 and 9.27 and instability indexes were between 27.75 and 54.72. Of 30 analyzed proteins, 10 were predicted as stable and 20 as unstable. The detailed information about the proteins’ molecular weight, isoelectric points and instability indexes are included in Table 1.

The MEME protein motif search tool identified a total of ten conserved motifs in Kalanchoë PEBP proteins ranging from 6 to 44 amino acids. The motif pattern appeared to be highly conserved among the proteins (Figure S1).

The protein models of KfFT3 and KfTFL1.3 were constructed based on the similarity with crystal structure of A. thaliana FT, 1wkpA, and TFL1, 1wkoA (Figure 3). KfFT3 shared 76% identity with 1wkpA, and TFL1 shared 69% identity with 1wkoA.

![Figure 3. Predicted structure of KfFT3 and KfTFL1.3 in comparison to AtFT (A. thaliana 1wkpA) and AtTFL1 (A. thaliana 1wkoA) proteins [8]. The external loop segment B is represented in blue, segment C (XYN triad) in green, putative 14-3-3-binding site in pink, and anion binding site in orange color. The conserved Y85/H85 residue crucial for FT/TFL1 antagonistic activities are shown in red. The conserved R130 residue of segment B important for 14-3-3 binding activity is represented in violet.](image)

### 3.4. Gene Structure

The length of the coding regions of PEBP-like genes in Kalanchoë ranged from 339 to 612 bp, with an average of 526 bp (FT-like: 531 bp; TFL1-like: 518 bp; BFT-like: 532 bp). All the genes of Kalanchoë were predicted to encode one transcript, except KfTFL1.1, that had two transcripts (Figure 4). Analysis of intron-exon distribution revealed that both FT-like and TFL1-like genes conserved the characteristic genomic organization for the gene family, with four exons and three introns, except KfTFL1 and KfTFL1.1b (secondary transcript of the gene). The length of the first exon in FT-like was between 198 and 285 bp (average: 211 bp, median: 207). The second and third exons were highly
conserved with the length of 41 and 62 bp, respectively, except from \textit{KIFTL1}, which had third exon of 64 bp. The fourth exon in \textit{FT}-like group was between 224 and 236 bp (average: 228 bp and median: 230). The \textit{TFL1}-like genes had a first exon between 198 and 216 bp (average: 207 bp, median 204 bp). Similarly, to \textit{FT}-like genes, the second and third exons were highly conserved with 62 and 41 bp, respectively, except the \textit{KIFTL1.1} gene (\textit{KIFTL1.1b}), which had a third exon of 160 bp. The fourth exon in \textit{TFL1}-like group was 221 bp in all the genes. All three \textit{BFT}-like genes were characterized by structures comprising three exons and two introns. The first exon was 207 or 210 bp, while second and third exons were 103 and 221 bp, respectively, in all the genes.

The introns had variable length in all the gene groups from 74 to 1396 bp. The 4-exon \textit{FT}-like genes were characterized by a structure with two short introns (<200 bp) and one long intron (>500 bp). The long intron was found either in the second (12 out 18 genes) or the third position (six out 18 genes). The 4-exon \textit{TFL1}-like genes had either all three short introns (<200 bp; five out of seven genes), or had a first medium length intron (~240 bp) and a second long intron (~600 bp) (two out of seven genes). The \textit{BFT}-like genes were characterized by a first short intron (<100 bp) and one medium length intron (~320 bp).

The identity of coding sequences for all \textit{PEBP}-like genes in Kalanchoë ranged between 50.5% and 100% (Table S3). The \textit{FT}-like gene identity ranged between 72% to 99% in \textit{K. fedtschenkoi} and 69% to 99% in \textit{K. laxiflora}. There was between 69% to 100% sequence identity among \textit{FT}-like genes between both species. The \textit{TFL1}-like gene identity ranged between 65% to 68% in \textit{K. fedtschenkoi} and 56% to 99% in \textit{K. laxiflora}. There was between 56% to 100% sequence identity among \textit{TFL1}-like genes between both species. The \textit{BFT}-like genes in \textit{K. laxiflora} showed 98% identity, while between species the identity was 97% and 99%.

3.5. Gene Duplication Analysis

The number of synonymous and nonsynonymous substitutions per site of the duplicated \textit{PEBP} genes in Kalanchoë associated with terminal branches in the tree species were determined using MEGA 7.0. We determined that average GC content in the third codon position that was 59% (Table S5). Thus, Ks values were used for calculation of the time of duplication events. The gene duplication events in \textit{K. fedtschenkoi} in \textit{FT}-like clade included one recent event dated approximately 0.9 MYA,
while other occurred between 5.5 and 16.0 MYA. The duplication events in TFL1/BFT-like clade might have occurred earlier between 20.5 and 25.4 MYA. In K. laxiflora the gene duplication events might have occurred more recently. The recent duplication events in FT-like and TFL1/BFT-like clades occurred approximately 0.9–3.2 MYA, with only one older duplication event dated approximately 20.7 MYA. The recent duplications in K. laxiflora can be associated with whole genome duplication and might have resulted in divergence and formation of a new species. The duplication events are presented in the species gene trees (Table 2 and Figure 5).

Figure 5. Species PEBP gene family trees in Kalanchoë fedtschenkoi (left) and Kalanchoë laxiflora (right) with proposed gene duplication events; red marks indicate probable respective gene duplication events (D1 to D3-4) that occurred before divergence of K. fedtschenkoi and K. laxiflora; blue marks indicate gene duplications observed only in K. laxiflora associated with probable WGD that lead to speciation; and the green mark indicates gene duplication in K. fedtschenkoi that occurred after split between K. fedtschenkoi and K. laxiflora.

Codon-based test of purifying selection revealed that the majority of duplicated gene pairs were under purifying selection, except one FT-like gene pair in K. fedtschenkoi (KfFT2–KfFT7) and three FT-like gene pairs in K. laxiflora (KlFT1–KfFTL1, KlFT2–KfFT6 and KlFT5–KfFT10) (Table 2).

Table 2. The Ks values and estimated absolute dates for the duplication events between the duplicated Kalanchoë genes associated with terminal branches in the tree species clades.

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>CDS Identity</th>
<th>Ks</th>
<th>Mya</th>
<th>Purifying Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>KfFT1</td>
<td>KfFT2</td>
<td>83.2</td>
<td>0.461</td>
<td>15.4</td>
<td>yes</td>
</tr>
<tr>
<td>KfFT1</td>
<td>KfFT7</td>
<td>82.5</td>
<td>0.480</td>
<td>16.0</td>
<td>yes</td>
</tr>
<tr>
<td>KfFT2</td>
<td>KfFT7</td>
<td>99.0</td>
<td>0.027</td>
<td>0.9</td>
<td>no</td>
</tr>
<tr>
<td>KfFT3</td>
<td>KfFT4</td>
<td>94.2</td>
<td>0.166</td>
<td>5.5</td>
<td>yes</td>
</tr>
<tr>
<td>KfFT5</td>
<td>KfFT6</td>
<td>88.2</td>
<td>0.222</td>
<td>7.4</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.1</td>
<td>KfTFL1.2</td>
<td>65.3</td>
<td>0.727</td>
<td>24.2</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.1</td>
<td>KfTFL1.3</td>
<td>67.6</td>
<td>0.650</td>
<td>21.7</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.2</td>
<td>KfTFL1.3</td>
<td>66.7</td>
<td>0.763</td>
<td>25.4</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.1</td>
<td>KfBFT1</td>
<td>64.8</td>
<td>0.614</td>
<td>20.5</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.2</td>
<td>KfBFT1</td>
<td>62.1</td>
<td>0.811</td>
<td>27.0</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1.3</td>
<td>KfBFT1</td>
<td>62.6</td>
<td>0.644</td>
<td>21.5</td>
<td>yes</td>
</tr>
<tr>
<td>KfTFL1</td>
<td>KfTFL1</td>
<td>95.0</td>
<td>0.063</td>
<td>2.1</td>
<td>no</td>
</tr>
<tr>
<td>KfFT2</td>
<td>KfFT6</td>
<td>98.3</td>
<td>0.053</td>
<td>1.8</td>
<td>no</td>
</tr>
<tr>
<td>KfFT3</td>
<td>KfFT7</td>
<td>97.8</td>
<td>0.095</td>
<td>3.2</td>
<td>yes</td>
</tr>
<tr>
<td>KfFT4</td>
<td>KfFT11</td>
<td>97.6</td>
<td>0.053</td>
<td>1.8</td>
<td>yes</td>
</tr>
<tr>
<td>KfFT5</td>
<td>KfFT10</td>
<td>98.1</td>
<td>0.026</td>
<td>0.9</td>
<td>no</td>
</tr>
<tr>
<td>KfFT8</td>
<td>KfFT9</td>
<td>99.3</td>
<td>0.026</td>
<td>0.9</td>
<td>yes</td>
</tr>
<tr>
<td>KfBFT1</td>
<td>KfBFT2</td>
<td>97.9</td>
<td>0.096</td>
<td>3.2</td>
<td>yes</td>
</tr>
<tr>
<td>KlTFL1.1a</td>
<td>KlTFL1.2</td>
<td>68.2</td>
<td>0.621</td>
<td>20.7</td>
<td>yes</td>
</tr>
<tr>
<td>KlTFL1.3</td>
<td>KlTFL1.4</td>
<td>99.4</td>
<td>0.041</td>
<td>1.4</td>
<td>yes</td>
</tr>
</tbody>
</table>

1 Ks—Rate of synonymous substitutions; 2 Mya—Estimated dates of duplication events (million years ago); 3 results of the codon-based test of purifying.
To further evaluate the sequence diversity in FT and TFL1 homologs, we performed a sliding window analysis of rates of pairwise nonsynonymous (Ka) and synonymous (Ks) substitutions (Figure 6). While exon 4 of the FT proteins had almost no amino acid substitutions indicating strong purifying selection, exon 4 of TFL1 had higher Ka/Ks ratio, suggesting more relaxed selection.

Figure 6. Average pairwise ratio of nonsynonymous (Ka) to synonymous substitutions (Ks) (sliding window, window size 50, step 10) of Kalanchoë PEBP genes. TFL1/BFT genes show an excess of nonsynonymous substitutions in exon 4.

4. Discussion.

The PEBP family is one of the most ancient gene families, with a highly conserved gene structure and high protein sequence similarities across species [1,2]. Its members include genes with very important functions in flowering induction and plant architecture [3,4]. The PEBP genes, particularly FT-like and TFL1-like genes, have been identified and their detailed functions were studied in many plant species including model plants, crop and vegetable species, and ornamental plants [7,9,15,21,50,51]. In this study, we focused on PEBP genes from Kalanchoë that is an important genus of flowering ornamental plants. So far, there is no information available about PEBP genes in Kalanchoë or other species of the Crassulaceae family. Therefore, using available genome sequence data we identified PEBP gene members in two species—K. fedtschenkoi and K. laxiflora—and characterized their gene structures, gene family evolution, and protein features.

The FT gene is expressed in the phloem companion cells of the leaves under flower inductive conditions and corresponds to an FT protein of approximately 20 kDa with a theoretical isoelectric point of ~8–9 [52–55]. The FT protein is a key component of the florigen complex, which is translocated from leaves to SAM where it promotes flowering [4,6,56]. The TFL1 gene is expressed in the SAM [7] and the TFL1 protein is a signal that is translocated from the inner SAM cells to the outer cells and coordinates the cell identity [57]. With regard to flowering, the FT protein promotes flowering in plants, while [6] TFL1 represses flowering [49]. However, other TFL1-like proteins can be expressed in the vascular tissues and translocate to SAM to repress flowering [15,49,58,59].

Previous functional studies have shown that the PEBP family can be divided into three major functional clades [2,3,24]. In the present study, phylogenetic analysis of the deduced protein sequences demonstrated that Kalanchoë proteins can be classified into an FT-like clade and a TFL1/BFT-like clade. The FT-like proteins were closely related to proteins of perennial species, which were reported to regulate induction of flowering [19,48,60–62]. Thus, these proteins are likely the FT homologs that can regulate flower transition and initiation in Kalanchoë. Similar to FT-like proteins, the TFL1-like proteins showed close relation to TFL1-like proteins that controls inflorescence
meristem identity and delays the transition to the reproductive phase at the SAM [63–65]. Interestingly, the BFT-like proteins showed high homology to a mobile floral inhibitor from Chrysanthemum seticuspe [15]. Old physiological studies suggested an existence of a flowering inhibitor produced in leaves of Kalanchoë plants [66,67]. Thus, BFT-like proteins might fulfill the function of a mobile flower repressor. In angiosperms, all three PEBP gene families were identified in all species (including Phoenix dactylifera which was previously reported to lack MFT-like genes) [3]. However, in our study we were unable to identify MFT-like sequences in neither K. fedtschenkoi nor K. laxiflora. The inability to identify MFT-like genes might be due to genome misassembly resulting in mosaic gene sequences or gene loss in the assembly due to collapse of the repetitive surroundings [68,69]. However, even though very unlikely, the loss of MFT-like genes cannot be ruled out. K. fedtschenkoi and K. laxiflora represent the only species with available genome data from the Saxifragales order. Thus, the comparison with other closely related species is currently not possible.

It has been demonstrated in A. thaliana that FT and TFL1 may have an interchangeable roles by replacing a single amino acid [7] or protein segment [8,16] (Figure 2). Differences in FT/TFL1-like protein activities might be due to their binding affinity towards FD and/or 14-3-3 proteins. In the Kalanchoë PEBP protein alignment, we identified four amino acids, i.e., R62, P94, F101, and R130, predicted to participate in binding of 14-3-3 proteins [9,21]. Amino acid alignment revealed that Kalanchoë FT-like proteins had conserved tyrosine at position 85 (Y85) characteristic for floral promoters, while TFL1-like proteins had conserved histidine at position 85 (H85) characteristic for floral repressors [7]. Generally, FT-like proteins contain a region called segment B, which forms a loop in the protein structure and is essential for FT-like proteins to function as floral promoters. FT-like proteins usually contain tyrosine at position 134 (Y134) and tryptophan at position 138 (W138) in segment B, whereas the flowering repressor proteins contain non-tyrosine and non-tryptophan amino acids in these positions, respectively [4,8,49]. Additionally, FT-like proteins contain a triad region—Segment C—that is required for full functionality of FT-like proteins but not TFL1-like proteins [4,8]. Therefore, the presence of the mentioned residues indicates that Kalanchoë FT-like can act as flower activators and TFL1/BFT-like as flower repressors.

The number of PEBP-like genes found in different plant species varies greatly with up to 19 genes found in Glycine max and 24 copies identified in Musa acuminate [3]. The average number of PEBP genes in monocots (~17) was shown to be roughly twice the number in dicots (~8) [3]. In our study, we identified 11 genes in K. fedtschenkoi and 18 in K. laxiflora. These numbers are similar to eudicot species, such as Brassica rapa (12 PEBP genes), Solanum lycopersicum (12 PEBP genes), and G. max (19 PEBP genes), which were demonstrated to experience additional whole genome duplication events throughout their evolutionary history [3]. Recent analysis of the K. fedtschenkoi genome provided strong evidence for two ancestral WGD events in this species [35]. The comparison between PEBP trees from K. fedtschenkoi and K. laxiflora suggests that the latter species experienced a recent WGD event that based on Ks values associated with lateral branches took place between 0.9 and 3.2 MYA (Table 2 and Figure 5). Thus, the number of PEBP genes is consistent with diploid/tetraploid nature of the analyzed Kalanchoë species.

The investigated PEBP genes had highly similar sequences and exon-intron structures (Table S3 and Figure 4). Particularly, the exonic structures were highly conserved with characteristic for PEBP gene family exon 2 (62 nt) and exon 3 (41 nt) invariable in size in FT-like and TFL1-like genes. The only exception from FT-like genes includes KIFTL1 that appears to be a pseudogene lacking the entire 4th exon. However, incomplete gene sequence can be also a result of a mistake during genome assembly. Furthermore, the BFT-like genes from both species demonstrated novel gene structure with three exons and two introns resulting from a fusion between exon 2 and exon 3 (103 nt). Even though PEBP genes have highly conserved gene structures, some species exhibit novel features such as additional intron/exon in Musa acuminate [3] and Chenopodium rubrum [70], and exon fusion in Zea mays of FT-like genes [24]. KITFL1.1 was predicted to have alternative transcript as a result of downstream alternative usage of transcription start site (TSS). Alternative usage of TSSs and alternative splicing are key mechanisms to generate gene variation in eukaryotes. Both mechanisms are known to play important roles in tissue-specific gene expression and functional variation, which
have significant impact on biological processes [71]. Alternative splicing in TFL1/CEN paralogs in saffron was reported to influence terminal flowering and flowering time [72]. Thus alternative transcript of the TFL1.1 gene in K. laxiflora may be relevant in spatiotemporal expression of the gene. However, it is also possible that downstream TSS might produce a truncated protein whose function is deteriorated or lost.

The evaluation of sequence diversity in FT and TFL1 homologs revealed that the majority of the most recently duplicated genes are under positive selection (Table 2). Furthermore, a more detailed sliding window analysis of Ka and Ks revealed strong differences in the substitution rates in exon 4 between FT- and TFL1-like genes. This is consistent with previous studies showing that segment B situated in exon 4 evolved very rapidly in TFL1 orthologs, but is almost invariant in FT orthologs. Thus, the residues encoded by the fourth exon of FT determine the function of the protein [1,8].

In Kalanchoë, flowering time is an important economic trait. Despite efforts to understand the mechanisms underlying the impact of photoperiod and temperature on the induction of flowering, little is known about the genetic basis of flower transition. The FT protein is a key integrator among different flowering pathways in angiosperms that promotes flowering [73]. In many plant species FT homologs regulate aspects of plant development in response to photoperiod and temperature [48,51,62,74,75]. Thus, the extended family of FT genes identified in Kalanchoë might be of significant relevance to flower induction in the response to different environmental cues. In breeding programs and commercial cultivation of many perennial plants a prolonged juvenility is one of the major problems. Modifying the FT/TFL1 ratio can change the flowering time [73]. Furthermore, TFL1 expression may be relevant for prevention of precocious flowering [76]. The overexpression of an FT homolog in poplar induced early flowering [62] and downregulation of a TFL1 homolog accelerated the flowering age [64]. The continuous flowering trait in roses and strawberries is associated with loss of function of the TFL1 homolog. The continuous flowering plants are characterized by short juvenile phase and rapid flowering after seed germination [77]. Thus, modification of FT/TFL1 expression may provide possibility to shorten the juvenile phase in Kalanchoë species and obtain plants that flower for a long period of time with no need of environmental control.

**Supplementary Materials:** The following are available online at www.mdpi.com/ [link], Figure S1: Alignments of the amino acid sequences of the PEBP proteins, Figure S2: Results of MEME motif analysis with PEBP proteins from Kalanchoë, Table S1: Results of homology analysis of Kalanchoë genes with Arabidopsis thaliana, Table S2: Results of homology analysis of Kalanchoë genes with various species, Table S3: CDS and protein identity in Kalanchoë, Table S4: Predicted subcellular localization of Kalanchoë PEBP proteins, Table S5: GC content (%) at first (P1), second (P2) and third (P3) codon position in PEBP family genes in Kalanchoë.


**Funding:** Part of this research was granted by CAPES (Coordination for the Improvement of Higher Level Personnel, Brazil, process number 9110-13-5). The authors did not receive funds for covering the publication costs.

**Acknowledgments:** The authors acknowledge the scholarship granted by CAPES (Coordination for the Improvement of Higher Level Personnel, Brazil, process number 9110-13-5) to LLC.

**Conflicts of Interest:** The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

**References**


