Sarmast, J Proteomics Bioinform 2017, 10:11
DOI: 10.4172/jpb.1000451

Journal of
Proteomics & Bioinformatics

Research Article Open Access

In silico Functional Analysis of FLC and FT-Genes Responsible for Postponing and Accelerating the Onset of Flowering

Mostafa Khoshhal Sarmast

Department of Horticultural Science, Faculty of Plant Production, Gorgan University of Agricultural Sciences and Natural Resources, Iran

Abstract

In silico functional analyses of FLOWERING LOCUS C (FLC) and FLOWERING LOCUS T (FT) genes were investigated herein. The recognition of cis-regulatory modules and their organization are a prerequisite to information on the regulation of gene expression. Accordingly, this study carried out a promoter analysis on FT and FLC genes in Arabidopsis and its orthologs in *Prunus persica* by PlantPan and PLANCARE servers in which the integrating transcription factor binding site (TFBs) was involved in hormonal and stress signals. Light responsive elements were also studied. Among the responsive elements identified in FT and FLC, Arabidopsis had the highest number of responsive elements. This is partly because *P. persica* is day-neutral, whereas Arabidopsis is a facultative long day plant. The results of motif recognition in FT and FLC proteins with the MEME server showed that FT is highly conserved, but the number of motifs suggests that the FLC protein had far more numerous motifs, which is likely due to its higher amino acid length. Some specified TFs like AGAMOUS (AG), AGAMOUS LIKE (AGL), LEAFY (LFY) and APETALA1 (AP1) were found by the PlantPan web server in FT and FLC promoters' analysis. They were also present in the protein-protein interaction network analysis which raises the question whether or not FT or FLC can be positive or negative regulators of their own expression in a feedback loop.

Keywords: Bioinformatics; Flowering; Gene network; Promoter analysis; FT

Introduction

The importance of plants to humans and the earth is unquestionable. Literature and arts are mostly inspired by the beauty of flowers and plants. Environmental conditions and the plant's internal factors are two key components that affect plants' flowering. The number of factors causing flowering may be different but they depend on the plant species, since flower induction can simultaneously be dependent on several factors. Tracking the seasonal changes during the years, type and perception of environmental signals and transducing environmental cues can precede developmental changes of flowering. However, these are not completely understood. The transitions to flowering, many genes have been identified with their pivotal role in the induction and production of floral organs. Postponed flowering in tree species can occur due to their extended juvenility, and this can be a substantial obstacle in breeding programs. The delay in flowering could result not only in a higher amount of biomass but also could occur coincidentally through cross-pollination between genetically modified and native cross-breeding plants. Therefore, it is less likely to have concerns about the movement of transgenes via pollen flow [1]. Two of the most important types of genes that are involved in floral development can be discussed herein. Floral meristem identity genes such as APETALA1 (AP1) and LEAFY (LFY) act as a trigger to push shoot apical meristems (SAM) toward flowering, and floral organ identity genes such as APETALA1 (AP1), APETALA2 (AP2), APETALA3 (AP3), PISTILLATA (PI) and AGAMOUS (AG) have products that are transcriptional factors. The latter is most likely to control the expression of other genes which has products that are involved in the formation and/or function of sepals, petals, stamens and carpels according to the ABC model at the anlagen of SAM [2]. The question concerning the mechanism through which the stimulation of flowering occurs as a result of the long day deserves further investigation. The circadian clock-mediated inductions of CONSTANS (CO) encode zinc finger proteins that regulate the transcription of other genes. This occurs as a result of the long day in Arabidopsis. The sharp increase in CO protein level, as a result of the long day, acts as a transcriptional regulator to stimulate expression in a downstream target gene called FLOWERING LOCUS T (FT) which is a RAF-kinase inhibitor-like protein for accelerating the onset of flowering [3,4]. Genetic screens have yielded the identification of this gene. Both of the aforementioned genes have been identified to be specifically expressed when being in companion with the cells of the leaf. CO–mediated induction of FT can produce a small globular protein (23 kDa) which moves in the phloem from leaves to the SAM [5,6]. FT's interaction with FD (FLOWERING LOCUS D) up-regulates floral meristem identity genes to induce reproductive development [7-10]. Recent research results indicated that FT can be a candidate in promoting flowering in trees [10]. Meanwhile, the FLC (FLOWERING LOCUS C) is a member of the small family of closely related MADS-domain key protein in plants [11,12] which that acts as a floral repressor by down regulation of the FT gene and mediating autonomous and vernalization pathways. Nonetheless, the FLC has a key role in the initiation of flowering and is involved in the reproductive structure, but its expression during all developmental stages in most parts of the plant propose some new regulatory functions [13]. There are a large number of DNA-binding proteins in plant genomes known as transcription factors (TFs) that bind to the short conserved motifs of 5 to 20 nucleotides called cis-acting regulatory elements (CAREs). They orchestrate the initiation of transcription by RNA polymerase II and also which function as gene expression regulators. CAREs are usually found in the vicinity of the

***Corresponding author:** Mostafa Khoshhal Sarmast, Department of Horticultural Science, Faculty of Plant Production, Gorgan University of Agricultural Sciences and Natural Resources (GUASNR), Basij SQ, Gorgan 49138-43464, Golestan, Iran, Tel: +98-17-32437618; E-mail: mkhsarmast@gau.ac.ir; mkhsarmast@ucdavis.edu

Received October 07, 2017; **Accepted** November 15, 2017; **Published** November 21, 2017

Citation: Sarmast MK (2017) *In silico* Functional Analysis of FLC and FT-Genes Responsible for Postponing and Accelerating the Onset of Flowering. J Proteomics Bioinform 10: 267-276. doi: 10.4172/jpb.1000451

5' end of the genes - usually upstream of the gene transcription start site (TSS), known as a promoter. The identification of plant promoters may provide fundamental information in understanding the regulation of gene expression [14]. Most promoter elements regulating TSS selection are localized in the proximal promoter. Many plant promoter databases have been developed based on cis-regulatory elements including PlantCARE [15], http://www.bioinformatics.psb.ugent.be/webtools /plantcare/html/], PLACE [16], http://www.dna.affrc.go.jp/PLACE/] or TRANSFAC [17], http://www.gene-regulation.com/pub/databases.html] and ppdb [18], http://www.ppdb.gene.nagoya-u.ac.jp] and PlantPan [19], http://plantpan2.itps.ncku.edu.tw.

The identification of direct physical binding of proteins or indirect protein-protein interactions could give some insight into the discovery of new molecular players. It further promotes or strengthens the interpretation of screens relating to genome-wide association, thereby giving clues about the biological function of some protein and other properties observed thereof.

Previously, researchers have used the STRING database [20] to find direct and indirect interactions of proteins with FT and FLC. The objectives of the present work were to recognize the cis-elements modules and their organization in the regulatory promoter rejoin of FT and FLC genes in Arabidopsis and peach. This was done in order to generate a comprehensive understanding of the regulation of gene expression. Furthermore, this research endeavored to identify the common motifs between FT and FLC proteins. Their functions through different data banks were studied and, finally, the proteins that interacted with FT and FLC were recognized. The proteins were hypothesized to be involved in the acceleration or suppression of flowering.

Materials and Methods

Promoter analysis of FT and FLC genes

The genomic DNA of FT and FLC gene in Arabidopsis (NM_105222, NM_001085094, respectively) was accessed through the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) web server. It was applied as a platform in order to recognize the promoter region of the FT and FLC genes by using the BLAST search through the Phytozome database (http://www.phytozome.net/). These were used for the purposes of investigation in this study. After identifying the genes on the chromosome using the BLAST-N algorithm, the region around the 1500 bp upstream of the start codon (ATG) in FT and FLC genes of *P. persica* and Arabidopsis were taken as a promoter. The upstream region of the FT and FLC in Arabidopsis and their corresponding orthologs in *P. persica* were analyzed by using the PLANT CARE and PlantPan database. For this purpose, upstream region sequences of FT and FLC in Arabidopsis and their corresponding orthologs were applied to predict their key cis-acting regulatory elements and the precise location of these elements.

Motifs identification and their functional analysis

Referring to the FLC and FT proteins obtained from the NCBI, these sequences were aligned by the MEME web server (http://meme.sdsc.edu/ meme4_6_1/cgi-bin/meme.cgi) so as to identify their common motifs [21]. In the case of FT, 91 sequences with an average length of 176.6 amino acids were analyzed. In FLC, however, 58 sequences with an average length of 184.6 amino acids were analyzed. In both proteins, the minimum and maximum motif size was 6 aa and 50 aa, respectively. The default value of 15 and 10 has been determined empirically for the maximum motif number concerning the FLC and FT, respectively. The predictions of probable functions of conserved domains of FT and FLC within the protein sequences was performed by the ELM program (http://elm.eu.org) and SMART (http://smart.embl-heidelberg.de). The UniProtKB (http://www.uniprot.org/) was used for the identification of some gene ontology characteristics of FT and FLC. Primary sequence analysis was done by ProtParam (http://expasy.org/cgi-bin/protparam). Moreover, similarity was assessed by using different programs at NCBI, such as the BLAST-P and PSI-BLAST (http://BLAST.ncbi.nlm.nih.gov/BLAST.cgi). Multiple sequence alignments were performed by using the Vector NTI Suit 9.

Secondary and tertiary structure prediction

The Swiss model web server (http://swissmodel.expasy.org/) was employed for the prediction of secondary structure of FT and FLC in *A. thaliana*.

Protein-protein interaction networks

A well-defined protein–protein interaction network in Arabidopsis gives good reason for the use of FLC (NP_001078563.1) and FT protein (NP_176726.1) as a query. STRING 9.0 (http://string-db.org) was used to predict all the proteins that interact with the FT and FLC proteins.

Results and Discussion

Analysis of FT and FLC promoters

The results of FT and FLC promoter analysis on the 1.5 kb sequence upstream of the ATG (start codon) in Arabidopsis and their orthologs in *P. persica* by PLANT CARE revealed that different transcription factors (TF) are attached to specific DNA binding sites. The TF mediate FT and FLC gene expression (Figure 1) in which there are the light-responsive elements (LRE), hormone-responsive element (HRE), stress-responsive element (SRE) and some miscellaneous-responsive elements (MRE). Each of these elements has its own specific cis-acting regulatory element.

For example, the subdivisions of HREs include the abscisic acid-responsive element, gibberellins-responsive element, auxin-responsive element, ethylene-responsive elements and the jasmonic acid and salicylic acid-responsive elements (Table 1). This indicates that the responses to photoperiodic and circadian rhythms closely correspond with TFs that affect the stress- and hormone-responsive elements. It may suggest that the integration of TFs involved in hormone and stress can control FT and FLC expression which lead to modulating light responses [22]. This hypothesis has to be experimentally supported by a HRE defected mutant. Results originating from bioinformatics results about cis-acting regulatory elements on FT and FLC promoters that are involved in endosperm and meristem expression are consistent with previous reports [13,23]. The specific elements presented in the FT and FLC promoters of Arabidopsis and peach were counted and are depicted in Tables 1 and 2. It has been stated that the PLANT CARE promoter database focuses on cis-regulatory elements rather than the core promoter structure. This approach will lead to a better understanding of the gene expression profile [18]. The core promoter is comprised of the least sequence required for gene expression which usually occupies around 80 nucleotides surrounding the transcription start site [2]. The particular time and place for expressive activity are designated by the transcription factor binding site (TFBs). It is known that the TFBs are located in the non-coding sequence upstream of the transcription start site called the regulatory promoter region. Accordingly, recognizing the mentioned cis-regulatory modules and their organization is a manifest need to shed light on the regulation of gene expression [24].

Figure 1: FT promoter sequence of *Arabidopsis thaliana*. 1500 bp upstream of start codon were used to represent proximal transcriptional DNA binding site (TDBS). Grey section represent CAAT and TATA box core promoter region. Red cis-regulatory elements involved in light responsive element. The colorless box indicated cis-regulatory element involved in other responses. The functions of these specific arenas are addressed in Table 1.

Site name	Element core sequence	Mean N. element (*A. thaliana*)	Mean N. element (*P. persica*)	Function
TATA-box	TATA	71	61	Core promoter element around -30 of transcription start
CAAT-box	CAAAT-TGCCAAC	33	23	Common cis-acting element in promoter and enhancer regions
TA-rich region	TATATATATATATATATATATA	0	5	Enhancer
Light-responsive element				
G-box	CACGTA	3	4	cis-acting regulatory element involved in light responsiveness
GT1-motif	GGTTAAT	0	3	Light responsive element
TCT-motif	TCTTAC	2	2	Part of a light responsive element
GAG-motif	GGAGATG	0	1	Part of a light responsive element
Sp1	CC(G/A)CCC	0	1	Light responsive element
I-box	ATGATATGA	3	1	Part of a light responsive element
AE-box	AGAAACAA	2	1	Part of a module for light response
3-AF1 binding site	TAAGAGAGGAA	0	1	Light responsive element
GA-motif	AAAGATGA	1	1	Part of a light responsive element
Box 4	ATTAAT	2	0	Part of a conserved DNA module involved in light responsiveness
chs-Unit 1 m1	ACCTACCACAC	1	0	Part of a light responsive element
chs-CMA2a	GCAATTCC	1	0	Part of a light responsive element
Pc-CMA2a	CAACCAATGAAAA	1	0	Part of a light responsive element
ACE	AAAACGTTTA	1	0	cis-acting element involved in light responsiveness
LAMP-element	CTTTATCA	1	0	Part of a light responsive element
MRE	AACCTAA	1	0	MYB binding site involved in light responsiveness
circadian	CAAAGATATC	3	2	cis-acting regulatory element involved in circadian control
Mean		**1.29[a] (1.02)**	**1[a] (1.1)**	

PGRs-responsive element				
ABRE	TACGTG	0	1	cis-acting element involved in the abscisic acid responsiveness
ERE	ATTTCAAA	1	0	Ethylene-responsive element
GARE-motif	AAACAGA	2	0	Gibberellin-responsive element
TGA-element	AACGAC	1	0	Auxin-responsive element
TCA-element	CCATCTTTTT	1	0	cis-acting element involved in salicylic acid responsiveness
P-box	CCTTTTG	0	1	Gibberellin-responsive element
CGTCA-motif	CGTCA	2	2	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	TGACG	2	2	cis-acting regulatory element involved in the MeJA-responsiveness
Mean		**1.125[a] (0.83)**	**0.75[a] (0.88)**	
Stress-responsive element				
MBS	CAACTG	1	2	MYB binding site involved in drought-inducibility
ARE	TGGTTT	2	2	cis-acting regulatory element essential for the anaerobic induction
Box-W1	TTGACC	0	1	Fungal elicitor responsive element
HSE	AGAAAATTCG	4	1	cis-acting element involved in heat stress responsiveness
TC-rich repeats	ATTTTCTTCA	0	1	cis-acting element involved in defense and stress responsiveness
LTR	CCGAAA	1	0	cis-acting element involved in low-temperature responsiveness
TC-rich repeats	ATTTTCTTCA	2	0	cis-acting element involved in defense and stress responsiveness
Mean		**1.43[a] (1.3)**	**1[a] (0.81)**	
Meristem- responsive element				
CAT-box	GCCACT	1	2	cis-acting regulatory element related to meristem expression
Skn-1_motif	GTCAT	1	0	cis-acting regulatory element required for endosperm expression
O2-site	GATGATGTGG	0	1	cis-acting regulatory element involved in zein metabolism regulation
Box III	CATTTACACT	0	1	Protein binding site
HD-Zip 3	GTAAT(G/C)ATTAC	1	0	Protein binding site
5UTR Py-rich stretch	TTTCTTCTCT	2	0	cis-acting element conferring high transcription levels
AT-rich element	ATAGAAATCAA	1	0	Binding site of AT-rich DNA binding protein (ATBP-1)
ATGCAAAT motif	ATACAAAT	1	0	cis-acting regulatory element associated to the TGAGTCA motif
Mean		**0.88[a] (0.64)**	**0.5[a] (0.75)**	
Total		152	123	

[†]In each column, means with the same letters are not significantly different at $P \leq 0.001$ level of probability using *T-test*. Means are followed by Standard Deviation (SD) in parenthesis

Table 1: Comparison between FT cis-regulatory elements of Arabidopsis and peach resulted from 1500 bp upstream of the ATG.

Site name	Element core sequence	Mean N. element (*A. thaliana*)	Mean N. element (*P. persica*)	Function
TATA-box	TAATA	83	63	Core promoter element around -30 of transcription start
CAAT-box	CAAAT	32	31	Common cis-acting element in promoter and enhancer regions
Light-responsive element				
Box 4	ATTAAT	6	0	Part of a conserved DNA module involved in light responsiveness
G-box	CACGAC	5	4	cis-acting regulatory element involved in light responsiveness
as-2-box	GATAatGATG	1	0	Involved in shoot-specific expression and light responsiveness
MRE	AACCTAA	1	1	MYB binding site involved in light responsiveness
circadian	CAANNNNATC	1	1	cis-acting regulatory element involved in circadian control
GATA-motif	AAAAAATTTC	1	1	Part of a light responsive element
Sp1	CC(G/A)CCC	1	2	Light responsive element
box II	AAAACGTTTA	1	1	Part of a light responsive element
ACE	AAAACGTTTA	1	0	cis-acting element involved in light responsiveness

MNF1	GTGCCC(A/T)(A/T)	1	1	Light responsive element
Box 4	ATTAAT	0	2	Part of a conserved DNA module involved in light responsiveness
GA-motif	ATAGATAA	0	1	Part of a light responsive element
TCT-motif	TCTTAC	0	1	Part of a light responsive element
Box I	TTTCAAA	0	1	Light responsive element
ATCT-motif	AATCTAATCT	0	1	Part of a conserved DNA module involved in light responsiveness
L-box	AAATTAACCAAC	0	1	Part of a light responsive element
3-AF1 binding site	AAGAGATATTT	0	1	Light responsive element
CATT-motif	GCATTC	0	1	Part of a light responsive element
Mean		**1.06[a] (1.6)**	**1.1[a] (0.90)**	
PGRs-responsive element				
ABRE	GGACACGTGGC	5	1	cis-acting element involved in the abscisic acid responsiveness
ERE	ATTTCAAA	0	1	Ethylene-responsive element
TCA-element	CAGAAAAGGA	1	1	cis-acting element involved in salicylic acid responsiveness
TGACG-motif	TGACG	2	2	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	CGTCA	2	2	cis-acting regulatory element involved in the MeJA-responsiveness
P-box	GCCTTTTGAGT	0	2	Gibberellin-responsive element
AuxRR-core	GGTCCAT	0	1	cis-acting regulatory element involved in auxin responsiveness
Mean		**1.43[a] (1.8)**	**1.43[a] (0.53)**	
Stress-responsive element				
ARE	TGGTTT	2	4	cis-acting regulatory element essential for the anaerobic induction
MBS	TAACTG	1	1	MYB binding site involved in drought-inducibility
HSE	AAAAAATTTC	1	1	cis-acting element involved in heat stress responsiveness
TC-rich repeats	ATTTTCTTCA	0	3	cis-acting element involved in defense and stress responsiveness
Box-W1	TTGACC	0	1	fungal elicitor responsive element
GC-motif	CCCCCG	0	1	Enhancer-like element involved in anoxic specific inducibility
LTR	CCGAAA	0	1	cis-acting element involved in low-temperature responsiveness
Mean		**0.57[a] (0.78)**	**1.71[a] (1.2)**	
Meristem- responsive element				
GCN4_motif	CAAGCCA	0	1	cis-regulatory element involved in endosperm expression
Skn-1_motif	GTCAT	2	1	cis-acting regulatory element required for endosperm expression
O2-site	GATGATGTGG	1	0	cis-acting regulatory element involved in zein metabolism regulation
ATGCAAAT motif	ATACAAAT	1	0	cis-acting regulatory element associated to the TGAGTCA motif
5UTR Py-rich stretch	TTTCTTCTCT	1	0	cis-acting element conferring high transcription levels
MSA-like	TCAAACGGT	1	0	cis-acting element involved in cell cycle regulation
Mean		**1[a] (0.63)**	**0.33[a] (0.50)**	
Total		156	138	

†In each column, means with the same letters are not significantly different at $P \leq 0.001$ level of probability using *T-test*. Means are followed by Standard Deviation (SD) in parenthesis.

Table 2: Comparison between *FLC* cis-regulatory elements of Arabidopsis and peach resulted from 1500 bp upstream of the ATG.

Results indicated that the maximum number of elements in both plants is not allocated to LRE according to PLAN CARE. This magnifies the role of other cis-elements during the flowering process. Among the LRE elements in both genes in the two species of Arabidopsis and *P. persica*, the G-box had the highest cis-acting regulatory element involved in light responsiveness. G-box is not only a target site of phytochrome interacting factors (PIFs), which is required for phytochrome-regulated transcription in photoperiod response [25], but it is also involved in stress and defense responses [13,26]. Of the total number of responsive elements identified in FT and FLC, Arabidopsis had the highest number. This may not be due to the amino acid length of FT and FLC in Arabidopsis (Table 3). Perhaps it is due to the day neutral characteristic of *P. persica* in contrast to that of Arabidopsis which is a long-day species.

Proteins	Species	Amino acid number		Amino acid number	P-value
FT	*Arabidopsis thaliana*	175a	†Other species	175.22a	0.76
FLC	*Arabidopsis thaliana*	196a	Other species	199.6a	0.25

†In each row, means with the same letters are not significantly different at P ≤ 0.001 level of probability using T-test. †Species from Table 6 were used to generate this average.

Table 3: Evaluation of the FT and FLC amino acids length between *Arabidopsis thaliana* and other species.

The research results by the PlantPan database indicate that there are several common TFBS in FT and FLC promoters which were not recognizable by PLAN CARE. Some of the specified TF in regulatory promoter regions of the FT and FLC gene were AG [27], LFY [28] and

AP1 [29]. FLC and its target (FT) contained a CArG box [30] in their binding regions. Too many MYB and MYC TF were found also at FT and FLC promoters by the PlantPan [31,32]. AG and AGLs are TFs that work with SOC1 in flowering. In the presence of external or internal stimuli, they may motivate floral meristem identity genes like LFY and AP1. This often results in flower development at the anlagen of SAM [2].

Motifs identification

Flowering locus T is a probable component of the mobile flower-promoting signal (floral stimulus or florigen). It promotes the transition from vegetative growth to flowering. Its subcellular location is the cytoplasm and the nucleus. In contrast to FLC that is mostly localized in shoot apexes (add references or reason); the FT is mostly localized in the vasculature of leaves. The FT protein product moves from leaf to shoot apex and acts as a long-distance signal that induces Arabidopsis flowering [9,33]. The MEME algorithms that have been widely used for the discovery of DNA and protein sequence motifs [34] were used to identify conserved motifs of the FT proteins deposited in the NCBI data base. It is clear from Figure 2 that three domains are devoted largely to the total protein. The evaluation of the aforementioned domain function by using the SMART web server revealed that the first motif is obviously a PEBP (Phosphatidyl ethanolamine-binding protein) family – a highly conserved group of proteins that have been identified in numerous tissues in a wide variety of organisms, including bacteria, yeast, nematodes, plants, drosophila and mammals [3]. Various functions have been described for the members of this family, including the control of the morphological switch between the shoot growth and flower structures as well as the regulation of several signaling pathways such as the MAP kinase pathway [35]. The domains that were mentioned above can be evaluated by using the eukaryotic linear motif resource for functional sites in proteins (ELM) that have been revealed following the functional site class: Motif 1: APCC-binding Destruction motifs act by the anaphase-promoting ubiquitin ligase complex APC/C and selectively target numerous cell cycle-regulatory proteins for ubiquitin-mediated proteasome-dependent degradation, BRCT phosphopeptide ligands (or the BRCT domains in Eukaryotes which are present in proteins that are associated with the DNA damage-response. They recognize and bind specific phosphorylated serine (pS) sequences. This phospho-protein mediated interaction of the BRCT domain has a central role in the check points of the cell-cycle. Motif 2: the PKA Phosphorylation site (of motifs phosphorylated by a subset of AGC group kinases including PKA that all have similar sequence specificity). The WXXXYF motif is repeated in the Pex5p protein and is bound non-conventionally by an SH3 domain in the Pex13p peroxisomal membrane protein. It is involved in the import of peroxisomal matrix proteins. Motif 3: AP2 alpha ligands (concerning motifs responsible for the binding of accessory endocytic proteins to the alpha-subunit of adaptor protein AP-2 and their recruitment to the site of clathrin coated vesicle formation). This protein contains two copies of an approximately 70 amino acid domain termed the AP2 repeat because of its initial description in the floral homeotic protein APETALA2 (AP2) [36]. Evidence shows the connection of the AP2-domain in both ethylene and JA signaling. This suggests that ethylene and JA may cross-talk via these transcription factors [37].

The PIKK phosphorylation site often known as the phosphoinositide-3-OH-kinase related kinases (PIKKs) are atypical protein kinases exclusive to eukaryotes. The PIKK members are large proteins with Ser/Thr kinase activity serving important roles in DNA repair and DNA damage checkpoints, and also in the PKA Phosphorylation site (concerning motifs phosphorylated by a subset of AGC group kinases including PKA, all of which have similar sequence specificity).

Flowering locus C (FLC) is a MADS-box eukaryotic family of transcriptional regulators that share a stereotypical MIKC structure [38]. They have a central role in the regulation of flowering time in late-flowering phenotypes. This can enable the blocking of the transition from the vegetative to the reproductive development by repressing the 'SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1' and FT gene [39]. Results of the electronic Fluorescent Pictograph (eFP) Browser showed a high FLC expression in the vegetative shoot apex and also in the root tissues. A lower level of expression was observed in

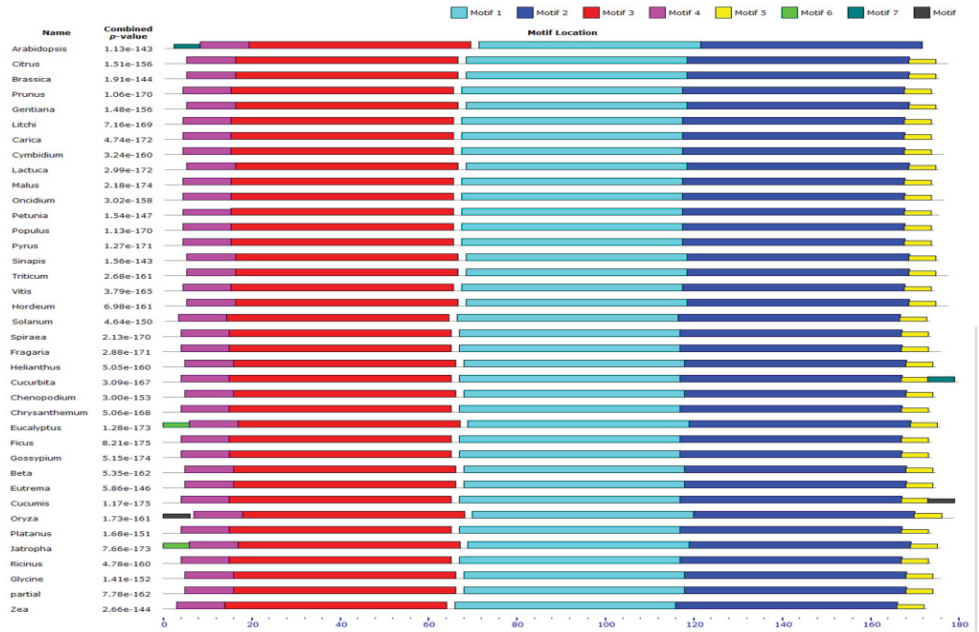

Figure 2: Distribution of FT protein motifs using MEME web server.

leaves and stems (rephrase this). Its subcellular localization is probably the nucleolus based on the eFP Browser.

It appears that in plants, FLC has more frequent domains compared to FT (Figure 3 and Table 4). According to the ASMART web server, FLC is a MADS-box domain that seems to be more closely related to SRF (Human serum response factor) domains which is a ubiquitous nuclear protein important for cell proliferation and differentiation as well as MEF2 domains in fungi. Alvarez-Buylla et al. [38] have suggested that animal and fungal MEF2-like sequences are more closely related to the t plant MADS-domain sequences than that of animal SRF-like sequences. Proteins belonging to the MADS family function as dimmers. MADS genes in plants encode key developmental regulators of flower, fruit, leaf and root development. Eukaryotic regulatory proteins with the highly conserved DNA-binding MADS domain include the MCM1 which is the regulator of cell type-specific genes in fission yeast, the DSRF which is a Drosophila trachea development factor, the MEF2 family of myocyte-specific enhancer factors and the Agamous and Deficiens families of plant homeotic proteins. According to the ELM server, the functional site classes were obtained with the FLC motif, including the Y-based sorting signal which is responsible for the interaction with the mu subunit of AP (Adaptor Protein) complex, the PKB Phosphorylation site which hosts a AGC group of kinases that act as a phosphorylation factor, the Cyclin recognition site, the PP1 docking motif, the SID, the PLK phosphorylation site, the Clathrin box, the MAPK docking motif, NLS classical Nuclear Localization Signals, di Lysine ER retrieving signal, FHA phosphopeptide ligands and the PIKK phosphorylation site. Consensus sequences in each motif of the FT and FLC, in addition to the amino acid frequency in FT and FLC, are presented in Tables 5 and 6.

Analysis of protein-protein interaction network

GIGANTEA (GI) is presented as a FB in Figure 4. It has a protein-protein interaction with CO and FT, but acts earlier than CO and FT in a circadian clock-controlled flowering pathway. This is because the GI is mediated in the regulation of phytochrome B signaling along with the CO to promote the FT gene when exposed to the photoperiod of long-day [40]. In addition to GI, CO can be influenced by two other transcription factors including HRB1 and SPINDLY is a zinc finger domain that could be involved in red and blue light signal transduction. Meanwhile, (SPY) acts as a repressor of GA responses. Furthermore, the positive regulation of cytokinin signaling can have an indirect and direct interaction with CO and FT, respectively [41,42].

Regarding the FLC protein, as is clear from Figure 4, the FLA represents FRI (FRIGIDA) and is a dominant allele that directly interacts with FLC to keep the plant in its vegetative state. According to the STRING web server, EMBRYONIC FLOWER 2 (EMF2) is shown as CYR1, CURLY LEAF (CLF) and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE). It is similar to the polycomb group protein that may be involved in flowering processes by repressing FLC promoters. The FLOWERING LOCUS D (FLD) is similar to the aforementioned proteins as it suppresses the FLC function. FLC can also interact directly with the SHORT VEGETATIVE PHASE (SVP) by binding to its promoter region to delay flowering (references). However, it has been stated that when the SVP loses its function, it does not fully suppress the delay of flowering because, according to Figure 4, there are some other proteins with which the FLC may interact [43]. Another key TF that interacted with FLC is VRN2 (Figure 4). It has been reported that the activation of VRN genes, after a long period of cold, can encode a DNA-binding protein and, especially, a homologue of one of the polycomb group proteins. The encoding is skewed towards the downregulation of FLC by dimethylation of lysines 9 and 27 on histone [44]. According to string results, the FRIGIDA-LIKE1 (FRL1) can directly or indirectly up-regulate FLC proteins [45].

The research results gained by protein-protein interaction indicated that the domains involved in the developmental and reproductive

Species	Number of amino acid (FLC)	Species	Number of amino acid (FT)
Arabidopsis thaliana	196	*Arabidopsis thaliana*	175
Brassiaca napus	197	*Medicago sativa*	176
B. rapa	196	*Lactuca sativa*	175
B. nigra	197	*Hordium vulgar*	177
Coffea arabica	206	*Populus nigra*	174
Eutrema japonica	197	*Eutrema japonica*	175
Raphanus sativus	197	*Litchi chinensis*	174
Vitis vinifera	210	*Camellia sativa*	175
Pyrus pyrifolia	199	*Prunus persica*	174
Mean	**†199.2[a] (4.8)**	**Mean**	**175.2[b] (1.13)**

†In each column, means with the same letters are not significantly different at P ≤ 0.001
level of probability using T-test. Means are followed by Standard Deviation (SD) in parenthesis

Table 4: FT and FLC amino acid length comparison, originated from NCBI.

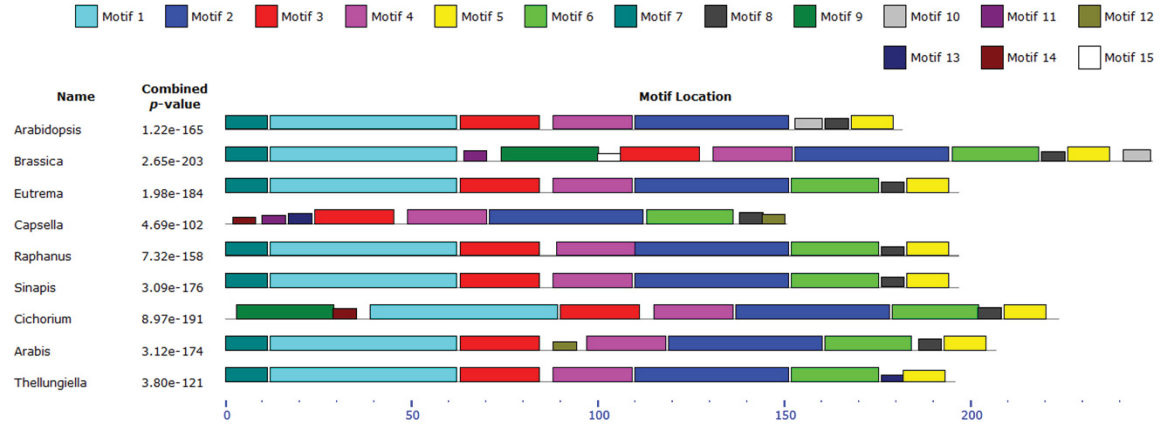

Figure 3: Distribution of FLC protein motifs using web server of MEME.

pathways throughout the plant life are closely connected with the other proteins that act as a switch from the vegetative stage to the reproductive development like the bZIP transcription factor FD, AP2, SVP and so forth. The research results of secondary structure predictions are purely indicative of a high rate of helix in FLC protein structures than the FT (Figure 5). Frequency of versatile regulatory elements on FT and FLC promoters suggest that these genes recruit several TFs to regulate specific gene expression. This issue is evident in their protein-protein interaction with STRING web server's results. The AP2-motif that was revealed in motif analysis of the FT protein is shown to be involved in stress and hormone-responsive gene expression [37] which, to some extent, explains the presence of hormonal and stress responsive elements on the regulatory promoter region of FT gene. The AP2 domain can be shown to correspond with ethylene and the MeJA responsive element, as also reported previously [37].

Furthermore, the analysis of genes involved in the flowering process confirmed that the response to photoperiodic and circadian rhythms may closely correspond with TFs that affect stresses and hormone-responsive elements. Accordingly, it can be reasonably concluded that

Motif number of FT	Consensus sequence
Motif 1	MVDPDAPSPS[DN]P[NH]LREYLHWLVTDIPATTGA[ST] FGQE[IV]VCYE[SN]PRP[TS][VM]GIH
Motif2	RFV[FL]VLFRQLGRQTVYAPGWRQNFNTRDFAELYNLG[LS]PVAAVYFNCQRES
Motif3	DVLDPFTRS[IV][SN]LRVTY[GN]N[RK]EV[NS]NGCEL[KR] PSQVVNQPRV[ED][IV]GG[DN]DLRTFYT
Motif4	RDPLVVGRV[IV]G
Motif5	GSGGRR
Motif6	MPRD[QR][DF]
Motif7	[HR][AM][GS][DI][EN][CI]
Motif8	[MR][AV][GQ][DS][DG][RY]
Motif number of FLC	Consensus sequence
Motif1	NKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVS[AS] SGKLYSFSSGDN
Motif2	NVS[VI][DG][SA]LVQLE[ED]HLETALS[VL]TRA[RK] KTELMLKLV[ED][NS]LKEKEK
Motif3	VKILDRYGKQH[AD]DDLKALD[LHR]Q
Motif4	N[YC]GSH[HY]ELLELV[ED]SKL[VE][EG][SP]NV
Motif5	[IQ][SI][DS][ID]NLPVTLP
Motif6	L[KE]EEN[QH]VLASQMEKN[HNT][LH]V[GRV]AEA[DE]
Motif7	MGRKKLEIKRI
Motif8	ME[IMV]SP[AG]
Motif 9	F[KY]V[KL]LC[GS][AF][EV]L[ST][RT][HI][DN][AI][GV][AQ] [EF][QV][LM][EG][MR][FR][IV][HY][VY]
Motif10	[GH][HL]VG[AV]E[AF]
Motif11	A[LS]G[KT][LP][NY]
Motif12	GQ[DI][LS][DQ][NS]
Motif13	[AF][LS]S[GP][AD][NS]
Motif14	S[QV][AN][DL]LV
Motif15	[MS][EG][DR]R[KS][LV]

Table 5: Motif distribution.

Amino acid	FT	FLC
Ala	4.6427	6.0593
Cys	1.5943	0.7696
Asp	5.844	4.932
Glu	4.4967	9.9019
Phe	4.7493	1.2303
Gly	8.5499	4.6718
His	1.5943	1.9782
Ile	3.1494	4.3087
Lys	1.1284	10.005
Leu	8.0671	14.801
Met	1.8638	2.634
Asn	5.1423	4.8561
Pro	7.6629	1.42
Gln	3.868	3.848
Arg	9.7962	5.3005
Ser	5.8833	10.742
Thr	6.1247	3.1597
Val	10.728	7.5768
Trp	1.1284	0.0759
Tyr	3.9859	1.7289

Table 6: Frequency of FT and FLC amino acids in plant exist in NCBI data bank.

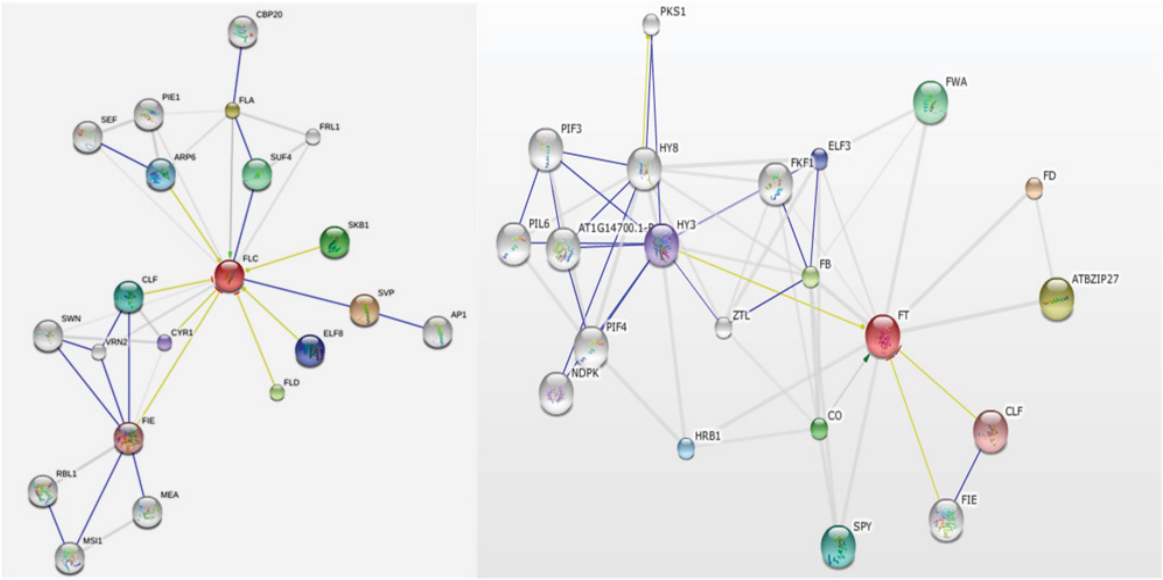

Figure 4: Protein-protein interaction network analysis of FLC (left) and FT (right) using STRING 9.0 in *Arabidopsis thaliana*.

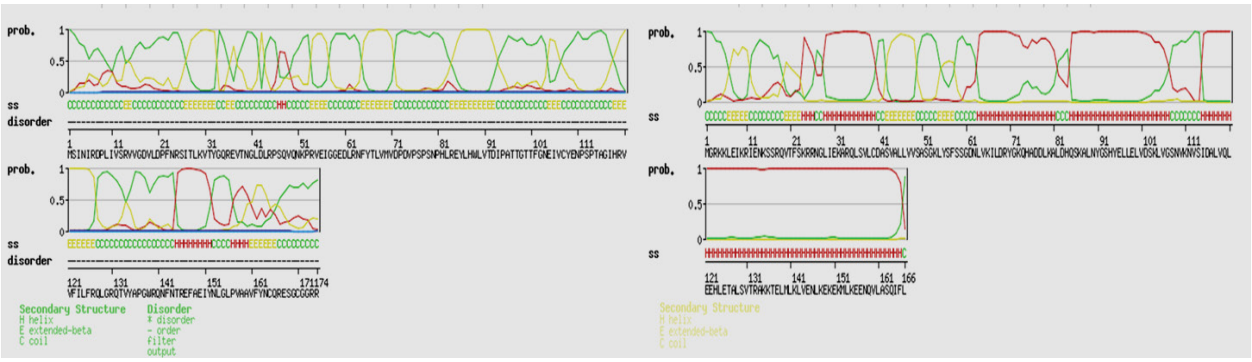

Figure 5: Secondary structure prediction of FT (left) and (FLC) in *Arabidopsis thaliana* using http://swissmodel.expasy.org/.

manipulating the plant flowering process precisely, without affecting the other normal physiological processes, can be a difficult ambition to achieve.

Conflict of Interest

Author declares no competing financial interest.

References

1. Salehi H, Seddighi Z, Krevchenko AN, Sticklen MB (2005) Experision of the cry1AC in 'Arizona Common' Common Bermudagrass *via* Agrobacterium-mediated transformation and control of Black Cutworm. J American Society for Horticultural Science 130: 619-623.
2. Taiz L, Zeiger E, Møller IM, Murphy A (2015) Plant physiology and development. 6th edn. Sinauer Associates, Inc, USA.
3. Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T, et al. (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960-1962.
4. Lee J, Lee I (2010) Regulation and function of SOC1, a flowering pathway integrator. J Exp Bot 61: 2247-2254.
5. An H, Roussot C, Suarez-Lopez P, Corbesier L, Vincent C, et al. (2004) CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development 131: 3615-3626.
6. Ayre BG, Turgeon R (2004) Graft transmission of a floral stimulant derived from CONSTANS. Plant Physiol 135: 2271-2278.
7. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, et al. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309: 1052-1056.
8. Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, et al. (2005) Integration of spatial and temporal information during floral induction in Arabidopsis. Science 309: 1056-1059.
9. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316: 1030-1033.
10. Zhang H, Harry DE, Ma C, Yuceer C, Hsu CY, et al. (2010) Precocious flowering in trees: The FLOWERING LOCUS T gene as a research and breeding tool in Populus. J Exp Bot 61: 2549-2560.
11. Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell 11: 949-956.
12. Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, et al. (1999) The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11: 445-458.
13. Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, et al. (2011) FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. Proc Natl Acad Sci USA 108: 6680-6685.
14. Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV, et al. (2003) PlantProm: A database of plant promoter sequences. Nucleic Acids Res 31: 114-117.
15. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, et al. (2002) PLAN CARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30: 325-327.
16. Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27: 297-300.
17. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34: 108-110.
18. Yamamoto YY, Obokata J (2008) ppdb: A plant promoter database. Nucleic Acids Res 36: 977-981.
19. Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant Promoter Analysis Navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene group. BMC Genomics 9: 561.
20. Szklarczyk D, Morris JHH, Cook H, Kuhn M, Wyder S, et al. (2017) The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45: D362-D368.
21. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28-36.
22. Lei H, Su S, Wen L, Wang X (2017) Molecular cloning and functional characterization of CoFT1, a homolog of FLOWERING LOCUS T (FT) from *Camellia oleifera*. Gene 626: 215-226.
23. Chiang GC, Barua D, Kramer EM, Amasino RM, Donohue K, et al. (2009) Major flowering time gene, flowering locus C, regulates seed germination in *Arabidopsis thaliana*. Proc Natl Acad Sci 106: 11661-11666.
24. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory network. Current Opinion in Structural Biology 14: 283-291.
25. Menkens AE, Schindler U, Cashmore AR (1995) The G-box: A ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. Trends Biochem Sci 20: 506-510.
26. Arias JA, Dixon RA, Lamb CJ (1993) Dissection of the functional architecture of a plant defense gene promoter using a homologous in vitro transcription initiation system. Plant Cell 5: 485-496.
27. Huang H, Tudor M, Weiss CA, Hu Y, Ma H (1995) The Arabidopsis MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol Biol 28: 549-567.
28. Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, et al. (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. Cell 105: 793-803.
29. Tilly JJ, Allen DW, Jack T (1998) The CArG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects. Development 125: 1647-1657.
30. Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: An *in vitro* and *in vivo* study. J Biol Chem 278: 28154-28159.
31. Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, et al. (1997) Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9: 1859-1868.

32. Hosoda K, Imamura A, Katoh E, Hatta T, Tachiki M, et al. (2002) Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. Plant Cell 14: 29.

33. Huang T, Bohlenius H, Eriksson S, Parcy F, Nilsson O (2005) The mRNA of the Arabidopsis Gene FT Moves from Leaf to Shoot Apex and Induces Flowering. Science 309: 1694-1696.

34. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research 34: 369-373.

35. Trakul N, Posner MR (2005) Modulation of the MAP kinase signaling cascade by Raf kinase inhibitory protein. Cell Research 15:19-23.

36. Jofuku KD, den Boer BG, Montagu MV, Okamuro JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. The Plant Cell 9: 1211-1225.

37. Menke FLH, Champion A, Kijne JW, Memelink J (1999). A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitorinducible AP2-domain transcription factor, ORCA2. EMBO J 18: 4455-4463.

38. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci 97: 5328-5333.

39. He Y, Michaels SD, Amasino RM (2003) Regulation of Flowering Time by Histone Acetylation in Arabidopsis. Science 302: 1751-1754.

40. Huq H, Tepperman JM, Quail PH (2000) GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. Proc Natl Acad Sci 97: 9789-9794.

41. Tseng TS, Swain, Olszewski SM (2001) Ectopic expression of the tetratricopeptide repeat domain of SPINDLY causes defects in gibberellin response. Plant Physiol 126: 1250-1258.

42. Kim J, Yi H, Choi G, Shin B, Song PS, et al. (2003) Functional Characterization of Phytochrome Interacting Factor 3 in Phytochrome-Mediated Light Signal Transduction. The Plant Cell 15: 2399-2407.

43. Li D, Liu C, Shen L, Wu Y, Chen H, et al. (2008) A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental Cell 15: 110-120.

44. Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, et al. (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427: 164-167.

45. Michaels SD, Bezerra IC, Amasino RM (2004) FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. Proc Natl Acad Sci 101: 3281-32853.