Current Plant Biology 18 (2019) 100097

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Transcriptomic analysis related to the flowering of the citrus hybrid Microcitrangemonia

Eliana Maria Rocha Sousa[a], Caio Suzart[a], Shirley Nascimento Costa[a], Marcio Gilberto Cardoso Costa[a], Alex-Alan Furtado de Almeida[a], Mauricio Antonio Coelho Filho[c], Felipe Rodrigues da Silva[b], Walter dos Santos Soares Filho[c], Fabienne Micheli[a,d], Abelmon da Silva Gesteira[a,c,*]

[a] Departamento de Ciências Biológicas, Centro de Genética e Biologia Molecular, Universidade Estadual de Santa Cruz, Ilhéus, Bahia, Brazil
[b] Embrapa Informática Agropecuária, Campinas, SP, Brazil
[c] Embrapa Mandioca e Fruticultura, Cruz das Almas, Bahia, Brazil
[d] CIRAD –UMR AGAP, Montpellier, France

ARTICLE INFO

Keywords:
Citrus spp
precocious
molecular
reproduction

ABSTRACT

Flowering in citrus plants grown from seeds occurs between four to ten years; therefore, the reduction of the juvenile period is a challenge for the genetic improvement of citrus. However, the Embrapa Mandioca e Fruticultura created a hybrid called Microcitrangemonia [Lemon Tree 'Rangpur' (LCR) X Citrus 'Yuma' (CTYM) – 005) X *Microcitrus papuana* (MCP) – 011] which presents a short juvenile period (one year) and flowering in all seasons of the year. It has a rare genotype for citrus cultivation. Other varieties with similar characteristics are only transgenic. This work describes the genes expressed in flowering-related pathways in the citrus hybrid Microcitrangemonia, which shows early and constant flowering and compares the gene function of each pathway with the existing *Arabidopsis thaliana* homologues. We performed RNA extraction, RNA-sequence library construction and sequencing from leaves, reproductive buds and stems of the H011 plant. We identified flowering-related transcripts and cladograms of similarity for the SOC1, AP1 and FLC proteins. Sequence annotation showed that most of the hits are similar to *Citrus sinensis*, 57%, and *Citrus clementina*, 38.9%. Different sequences similar to 29 proteins belonging to different flowering-related pathways have been identified, such as gibberellin, photoperiod, autonomous and vernalization. The analysis of similarity allowed inferring that H011 has greater evolutionary kinship with *Citrus sinensis*. The outcomes contribute to the understanding of the early and constant flowering of H011 as well as provide an important resource for breeding programs in citrus aimed at juvenile period reduction.

1. Introduction

In the plant life cycle, flowering is one of the most important events representing the transition from the vegetative to the reproductive period. If the plant has market value due to fruit commercialization, flowering represents the beginning of production and profits [1].

In citrus, which are perennial, woody plants, with several species with commercial value, flowering is an event of late occurrence. Orchards produced from seeds may take more than five years to breakdown from the long juvenile period, which is a problem for production and a challenge in citrus genetic improvement programs [2].

Researchers and producers have adopted different strategies in an attempt to reduce juvenile period in citrus such as the use of rootstocks, application of phyto-regulators and plant submission to the abiotic stresses [3,4]. However, there is still a need to better understand the genetic and molecular mechanisms that regulate flowering in citrus so that the measures may become more effective or even to create new actions to reduce the vegetative period and consequently promote flowering.

The citrus hybrid Microcitrangemonia (H011), developed at EMBRAPA Cassava and Fruits, exhibits characteristics related to the juvenile period and flowering, which are unique compared to other citrus species of zygotic origin. H011 has a very short juvenile period; there are individuals that begin to bloom in the first months of life.

* Corresponding author at: Departamento de Ciências Biológicas, Centro de Genética e Biologia Molecular, Universidade Estadual de Santa Cruz, Ilhéus, Bahia, Brazil.
E-mail address: abelmon.gesteira@embrapa.br (A. da Silva Gesteira).

https://doi.org/10.1016/j.cpb.2018.12.003
Received 2 October 2018; Received in revised form 11 December 2018; Accepted 15 December 2018

Table 1
Proteins involved with flowering in Arabidopsis and citrus.

Initials	Name	Organism	Access	Database
AP1	Apetala 1	*Citrus sinensis*	NM001288899	NCBI
BAM 1	Barely and Meristem 1	*Arabidopsis thaliana*	O49545	UNIPROT
BAM 2	Barely and Meristem 2	*Arabidopsis thaliana*	Q9M2Z1	UNIPROT
CCA1	Circadian Clock Associated 1	*Arabidopsis thaliana*	AAB40525.1	NCBI
CIFT2	Flowering Locus T2	*Citrus unshiu*	A9ZND2	UNIPROT
CIFT3	Flowering Locus T3	*Citrus unshiu*	A9ZND3	UNIPROT
CO	Constans	*Arabidopsis thaliana*	CAA71587.1	NCBI
CRY1	Cryptochrome-1	*Arabidopsis thaliana*	AEE82696.1	NCBI
CRY2	Cryptochrome-2	*Arabidopsis thaliana*	AEE27693.1	NCBI
CRY3	Cryptochrome-3	*Arabidopsis thaliana*	AED93369.2	NCBI
DELLA	Della	*Citrus sinensis*	006482132.1	UNIPROT
EMF1	Embryonic Flower1	*Arabidopsis thaliana*	A0A178UAP8	UNIPROT
EMF2	Embryonic Flower2	*Arabidopsis thaliana*	BAB58956.1	NCBI
FCA	Flowering time control	*Poncirus trifoliata*	A0A1D8GWG0	UNIPROT
FES1	Protein Frigida Essential1	*Arabidopsis thaliana*	OAP09490.1	NCBI
FLC	Flowering Locus C	*Poncirus trifoliata*	ACB72867.1	NCBI
FLD	Flowering Locus D	*Arabidopsis thaliana*	OAP07104.1	NCBI
FRI	Frigida	*Citrus sinensis*	A0A067DU43	UNIPROT
GI	Gigantea	*Arabidopsis thaliana*	AKX69393.1	NCBI
LFY	Leafy	*Citrus sinensis*	Q6EEV8	UNIPROT
LHY	Late Elongated Hypocotyl	*Arabidopsis thaliana*	Q6R0H1	UNIPROT
LFR	Leaf and Flower Related	*Arabidopsis thaliana*	OAP02817.1	NCBI
PHYA	Phytochrome A	*Arabidopsis thaliana*	OAP14284.1	NCBI
PHYB	Phytochrome B	*Arabidopsis thaliana*	OAP11187.1	NCBI
SOC1	Suppressor of constans overexpression 1	*Citrus sinensis*	NM001288843.1	NCBI
SUF3	Suppressor of FRI 3	*Arabidopsis thaliana*	A0A178U6Y1	UNIPROT
SUF4	Suppressor of FRI 4	*Arabidopsis thaliana*	OAP11907.1	NCBI
SVP	Short Vegetative Phase	*Poncirus trifoliata*	B6VC86	UNIPROT
TFL	Terminal Flower 1	*Citrus sinensis*	Q69F37	UNIPROT
TOC1	Timing of CAB Expression 1	*Arabidopsis thaliana*	OAO90757.1	NCBI
VIN3	Vernalization Insensitive 3	*Arabidopsis thaliana*	A0A178UF19	UNIPROT
VIP3	Vernalization Independence3	*Arabidopsis thaliana*	A0A178V016	UNIPROT
VIP4	Vernalization Independence4	*Arabidopsis thaliana*	OAO91028.1	NCBI
VIP5	Vernalization Independence5	*Arabidopsis thaliana*	A0A178WDP6	UNIPROT
ZTL	Zeitlupe	*Arabidopsis thaliana*	A0A178UDC5	UNIPROT

Thus, in addition to being interesting to producers, this genotype is a rich molecular and genetic source for information on mechanisms controlling flowering in citrus.

Most studies related to floral development take place on plant models such as *Arabidopsis*. Signaling pathways such as gibberellin (GA), photoperiod, vernalization and autonomous were discovered, as well as the identification of different genes that are part of these pathways and which genes induce or repress flowering [5,6]. Therefore, the information found in *Arabidopsis* can be used to carry out similar investigations in citrus. Previous studies in citrus which show that genes such as *LEAFY (LFY)*, *APETALAI (API)*, *TERMINAL FLOWER (TFL), APETALA3* (*AP3)*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and *WUSCHEL (WUS)* are linked to flowering and are homologous to those found in *Arabidopsis* [7,8,9]. These studies are among the pioneers on genes related to flowering in citrus, which is a complex process that currently requires further investigation, specially on the genes expressed initially in each pathway. Most of the studies regarding citrus integrate meristem-identity genes. However, identifying representatives of other pathways may help in better understanding the interaction between the gene products and induction/repression processes that occur on the transition from the vegetative to the reproductive stage. This study investigated the genes expressed in the pathways related to flowering in H011. We conducted the functional annotation of leaf, stem and bud transcripts focusing at genes previously related to the photoperiod, vernalization, autonomous and gibberellin pathways and compared the function of the genes of each pathway with the existing homologues in *Arabidopsis thaliana.*

2. Methodology

2.1. Collecting samples of the hybrid Microcitrangemonia, RNA extraction, library construction and sequencing

Leaves, reproductive buds and stems were collected from the hybrid citrus H011 at the EMBRAPA Cassava and Fruits, Germoplasma Active Bank (BAG), Cruz das Almas, BA. We obtained five samples of each tissue, conditioned in suitably identified aluminum foil bags and immediately frozen in liquid nitrogen for RNA preservation. The samples were transported to the Laboratory of Molecular Biology (LBM) and stored at −80 °C for two weeks.

We extracted total RNA from the samples from approximately 75 mg of plant material from each tissue replicate, macerated with liquid nitrogen, using the *RNAqueous® Kit (Applied Biosystems, Ambion)* following the manufacturer's recommendations. RNA quality was verified in electrophoresis with 1% agarose gel and, for the band visualization, we used fluorescent dye Gel Red®. The images were captured using the Kodak Gel Logic® photodocumentator.

Qubit Fluorometer (Life Technologies) measured RNA concentration using a Qubit RNA Assay kit (Life Technologies) for RNA labeling. RNA samples from each tissue (stem, leaf and bud) were combined in three equimolar pools (LF (leaf field)), BF (bud field), SF (stem field).

The RNA-seq library was prepared with a Illumina TruSeq kit and protocol, then quantified by qPCR (real time PCR) using a Kapa® Kit and the Applied Biosystems 7500 Fast apparatus. Library sequencing occurred on the Illumina MiSeq platform.

2.2. Sequence processing and annotation

After the sequencing, FastQC tool [10] was used to verify the *reads*

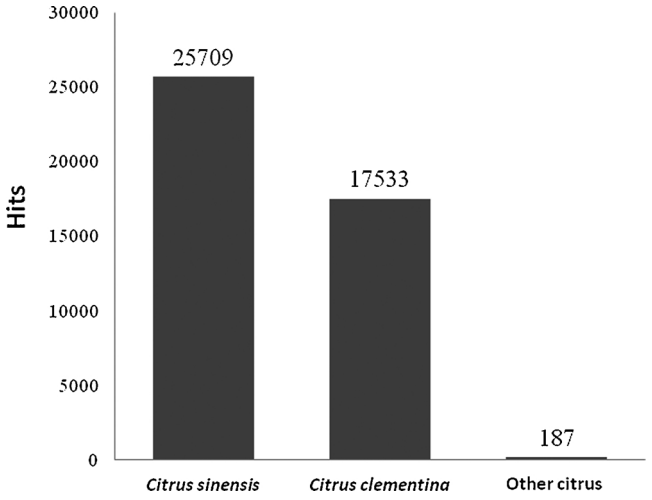

Fig. 1. Distribution of best hits of the most prevalent species of citrus in relation to the hybrid Microcitrangemonia H011.

quality. Then, adapters of the reads, sequencing artifacts and fragments with low quality were removed with the Trimmomatic version 0.32 program [11]. The filtered *reads* of the library were aligned to assemble the transcripts with the Trinity tool [12].

Diamond version 3 software was used to analyze the sequences [13]. The alignment was performed with the Blastx algorithm that allowed finding similarity for the functional annotation of the sequences translated into six reading frames when compared with vegetable-derived proteins available at Swiss-Prot and TrEMBL do Uniprot (http://www.uniprot.org/) and Refseq do NBCI (http://www.ncbi.nlm.nih.gov).

Afterwards, the Blast2GO version 5 tool [14] received the results (output) of the alignment of the transcripts to perform the functional annotation of the sequences by mapping in the Gene Ontology (GO) database. The Biological Process, Molecular Function and Cellular Component characteristics of the proteins provided by the functional annotation were analyzed in general (all Microcitrangemonia sequences) and in a specific way (by searching for the keyword "flower" in the annotation terms).

2.3. Identifying transcripts related to flowering

At the NCBI and Uniprot databases, we acquired sequences of 32 different citrus or Arabidopsis proteins associated with the pathways involved in the flowering process (Table 1). These pathways are gibberellin, photoperiod, autonomous, vernalization, floral integrators and meristematic identity. Subsequently, we used the BLAST tool to identify homologous H011 transcripts. The cut-off used for consideration of sequence similarity was the alignment identity and coverage of equal to or greater than 50%.

2.4. Cladogram of similarity

Sequences of different plant species were obtained from the NCBI and Uniprot databases for the SOC1, AP1 and FLC proteins belonging respectively to the floral integrator pathways, meristematic and autonomous identity. To develop the cladogram to the protein FLC, we used 25 sequences belonging to 12 species, to the protein SOC1, 33 sequences reporting to 31 species; and to protein AP1, 25 sequences belonging to 21 species. In the alignment for each protein, we included the homologous sequence of the H011 protein, using the program Muscle [15]. The cladogram was made using the MEGA 7.0 [16] tool.

3. Results and discussion

There were different GO terms for the H011 sequences distributed in the three different GO classes (Biological Process, Molecular Function and Cellular Component). A restricted search for the processes presenting the word flower in the GO terms allowed us to obtain specific results related to the transition from the vegetative to the reproductive period. The proteins involved in the pathways related to flowering were identified in the H011sequences and different hypotheses were created on the performance of these proteins. The similarity study between plant species and the H011 occurred for three proteins that were of great importance for flowering and the evolutionary proximity of the H011 was demonstrated.

3.1. Annotations of the sequences

The 5,204,591 paired-end reads obtained by sequencing and filtered by TRIMMOMATIC were processed by Trinity tool which assembly generated 45,080 transcripts. The alignment to plant sequences from Refseq Database with BLASTN identified 43,430 transcripts similar to sequences belonging to citrus and 1650 similar to sequences from other plant species.

Citrus sinensis was the species with a higher prevalence in relation to the similarity with the H011 sequences (57%; 25,700 hits). *Citrus clementina* had 38.9% (17,533) of hits and other species had 0.4% (187) of hits (Fig. 1). The high prevalence of similarity with *C. sinensis* and *C. clementina* is a consequence of genome sequencing of the two species, since most of the information available in citrus-related databases belong to both species [17,18].

The functional annotations of the transcripts through Blast2GO made it possible to obtain 26,073 GOs terms distributed among the three existing categories. Molecular function (MF), biological process (BP) and cellular component (CC) had 21,268, 19,280 and 18,908 annotated transcripts, respectively (Fig. 2). The same transcript may have more than one GO term simultaneously. Fig. 3 shows the most found terms between levels 2 to 6 for the three main categories - BP, CC and MF.

For BP, most of the transcripts relate to the metabolic processes and to the cellular processes, presenting 54.7% (14,265) and 54% (14,081), respectively (Fig. 3). The abundance of transcripts associated with the metabolic process occurs because of the importance and dimension of the process as it corresponds to the catabolism and anabolism of cellular molecules such as lipids, proteins, carbohydrates and nucleotides. The regulation of the metabolic processes generates the cellular homeostasis necessary for the correct organism functioning [19,20]. The prevalence of transcripts linked to cellular processes shows the cells need to perform activities such as intercellular communication, differentiation, apoptosis, exocytosis, phagocytosis, growth and multiplication [21,22]. Therefore, metabolic and cellular processes are among the main ones

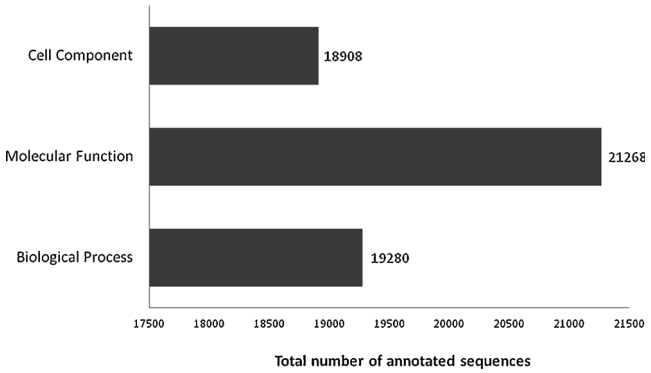

Fig. 2. Annotation of GOs levels found through the analyses performed in Blast2GO.

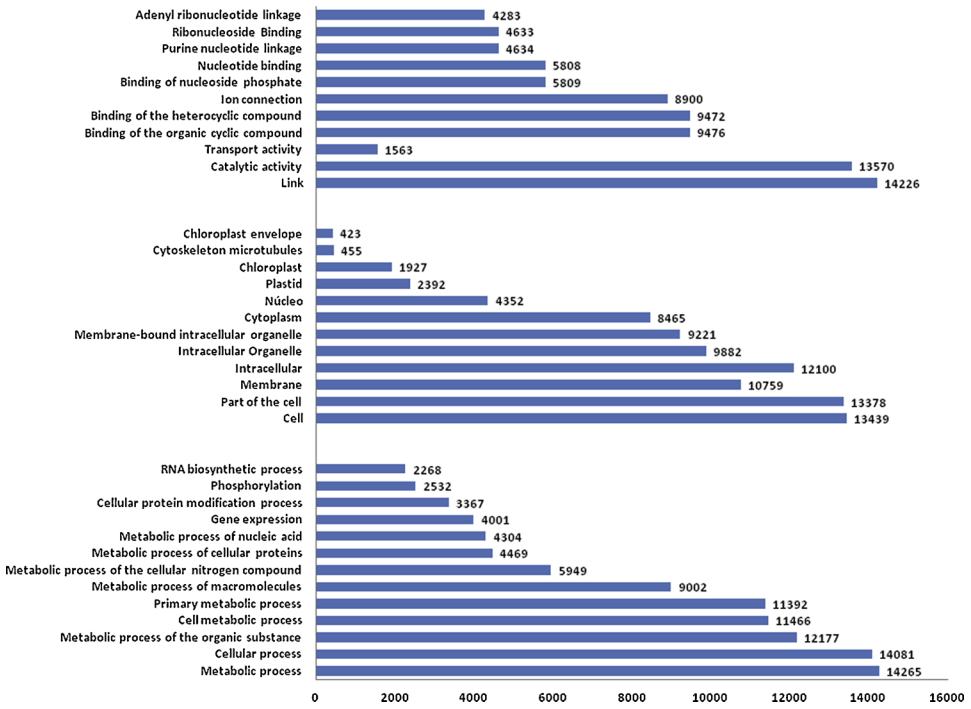

Fig. 3. Graphic with the distribution of the sequences between the GO terms for each category.

for the organism survival, which justifies the classification as the most common and processes in functional notes made for different living beings.

In the MF category, the highest prevalence of transcripts was for binding, 54.5% (14,226), and catalytic activity, 51.5% (13,439) (Fig. 3). The binding allows cells to create functional pairs such as receptor of signals and ligands, enzyme and substrate, protein and deoxyribonucleic acid (DNA). Such bindings are essential for the functioning of the different pathways of cellular metabolism [23]. Similar to the action of catalysts, specific enzymes whose function is to change the speed of a chemical or biochemical reaction that are not consumed during the process. A large diversity of enzymes is constantly produced in the cells, so that they act by accelerating different steps of the metabolic pathways. The absence of catalytic activity would allow extremely slow chemical and biochemical reactions in organisms [24].

Regarding CC, 51.5% (13,439) of transcripts were associated with the cell and 51.3% (13,378), to the cell part (Fig. 3). Transcripts arranged in the cellular cytoplasm assist in the metabolic processes and are not necessarily translated specifically for performance in organelles are the aggregates to the term cell [25]. The cell part corresponds to the spatial delimitation in eukaryotes, cytoplasm, nucleus, plasma membrane and organelles. Thus, the large number of sequences bound to such compartments associate with structure and function [26].

After the general functional annotation, a specific one occurred by searching for the word flower between the GO terms to obtain the most related category to flowering in relation to the H011. We recorded 156 transcripts belonging only to BP (Fig. 4). The term floral development obtained the greatest number of different transcripts, 36% (56), due to the development stage in which the H011 presented at collection time, reproductive period, with constant flower production. In addition, the flowering process in citrus is complex and requires expression of different proteins for pathway activation and repression [4]. Then, negative regulation of floral development, floral development and positive regulation of floral development had 12% (19), 12% (18) and 8% (12) different transcripts, respectively. Such sequences are possibly associated with floral induction genes such as GI, CO, FT, AP1, SOC1 and flowering repression genes such as FLC, SVP, FLC. According to Pajon et al., [27], Yamagishi et al., [28] and Deng et al., [29] plants of different genetically modified species, in relation to the flowering phenotype, can demonstrate positive regulation of genes that induce flowering, and/or negative regulation results for genes that repress flowering demonstrate phenotypic pattern similar to the H011.

However, an affinity of flowering in the H011 was evidenced with photoperiodism found in seven different terms (Fig. 4), though with lower percentages of transcripts. There is a direct relationship between photoperiod and flowering, so that plants are classified according to short, long or neutral days. Thus, for plants to germinate, grow and bloom it depends on the duration of the night and day period, detected by the circadian clock of the vegetable [30]. Flower morphogenesis was the last term associated with the H011, 1% (2) transcripts, which corroborates the data obtained in relation to the H011 sequences similar to proteins associated with flowering pathways. Only the AP1 and BAM proteins were identified in the meristematic identity pathway.

3.2. *Microcitrangemonia* sequences similar to flowering-related proteins

By means of tblasn, we identified 82.8% (29) different sequences in the H011 (Table 2) similar to the investigated proteins (Table 1) that belong to flowering-related pathways.

In studies developed mainly with *Arabidopsis*, flowering induction is a complex process involving the expression and repression of several genes. Gene regulation occurs from external and internal stimuli in the plant. Thus, pathways acting in the control of flowering can be classified as response to external factors: photoperiod and vernalization, and response to internal factors such as gibberellin (GA) and endogenous [31,32].

We found DELLA protein in the gibberellin pathway for the H011 (Table 2). This protein represses GA action by acting as a negative regulator for the transcription factors involved in GA signaling [33]. GA in citrus is associated with repression of flowering and its greater quantity and prevalence over time promotes increased vegetative development, consequently long juvenile period. Thus, the beginning of the reproductive period occurs late, from three to seven years depending on the species [4]. DELLA protein culminates with the main H011 characteristic, which is the short juvenile period, possibly the least GA amount. Fig. 5 shows H011 seedlings with two and three

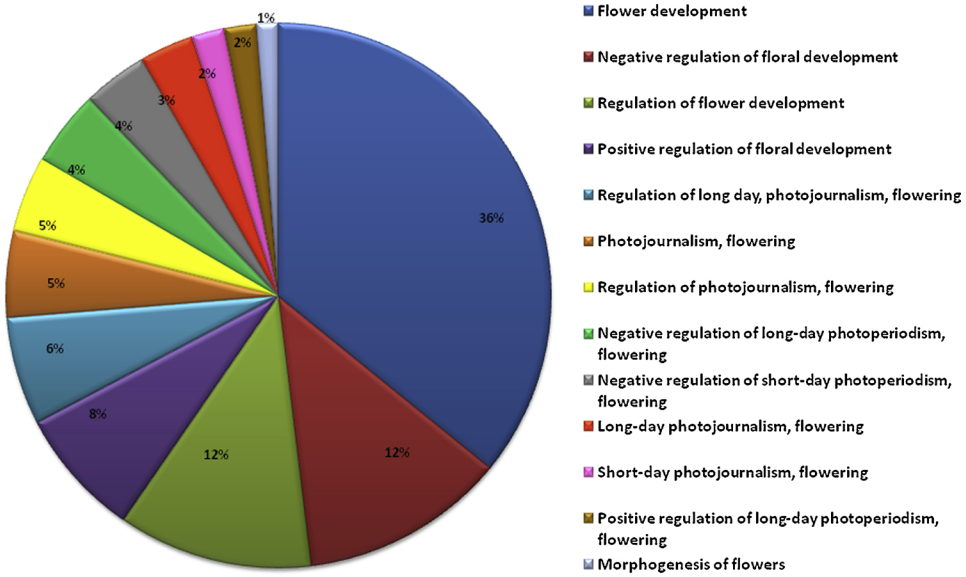

Fig. 4. Graphic illustrating the categorization of GO terms in the biological process class (BP) when looking for the word flower.

months of life with floral bud, which was not previously registered for citrus plants from classic genetic improvement.

In the photoperiod pathway, sequences homologous to the proteins were found in the H011: PHYA, PHYB, CRY1, CRY2, CRY3, TOC1, LHY, ZTL, CCA, GI, CO (Table 2). These proteins associate with the induction of flowering by promoting cascade molecular responses resulting in the expression of genes belonging to the floral integrator pathway (Fig. 6). Thus, the photoperiod is one of the most relevant factors for the occurrence of flowering in the plant [34].

Phytochromes A and B (PHYA and PHYB) and cryptochromes 1, 2 and 3 (CRY1, CRY2 and CRY3) act as red and blue light photoreceptors, respectively. Light as stimulator, photoreceptors and LHY, TOC1 and CCA work on the synchronization and functioning of the plant circadian clock. At dawn, the available light assists in the gene expression of the proteins CCA1 and LHY, which activate other genes needed during the daytime period and make impossible the expression of the genes expressed in the nocturnal period as the TOC1. In contrast, at dusk, the amount of LHY and CCA1 proteins decreases, consequently increasing the level of TOC1 [35]. The control of TOC1 degradation is performed by ZTL protein thus assisting in the support of the normal circadian period [36]. Periodic information from circadian clock promotes the expression of the GI gene; consequently, the GI protein activates the CO gene encoding the CO protein responsible for promoting the expression of the FT gene that makes up the pathway of the floral integrators (Fig. 6). These proteins were found in other species of citrus and other woody plants [37], suggesting that the genes involved with the photoperiod remain in these plants.

The H011 sequences also showed similarities with FCA, FLC, SVP and TFL proteins belonging to the autonomous flowering pathway (Table 2). These proteins act as negative regulators of flowering preventing the transcription of genes encoding the pathway proteins of floral integrators [38]. There is an inverse relationship between the presence and amount of FLC protein in relation to FT protein, so that when FLC protein exists, the FT protein is not expressed [29]. However, in the H011 both proteins appeared (Fig. 6, Table 2). The hypothesis is that the proteins related to the repression of flowering do not act effectively for the H011. This hypothesis is further reinforced by the phenotypic behavior of the genotype that exists in the Embrapa Cassava and Fruits (BA) for approximately eight years. H011 began to bloom after a year and to the present day it shows the production of flower and fruit concomitantly and daily; it is possible to obtain flower and fruit from the H011 in all seasons.

In addition, there were homologous sequences for the EMF2, SUF3, SUF4, VIP3, VIP4, VIP5 and VIP6 proteins for the vernalization pathway in the H011. The SUF3, SUF4, VIP3, VIP4 and VIP5 proteins act as positive regulators for the FLC expression during the low temperature period [39]. This indicates a possible high prevalence of FLC protein in the H011, strengthening the hypothesis that the repressor function of FLC flowering is not effective in the genotype under study. In a mutant *Poncirus trifoliata* with early blooming, *FLC* had higher expression during winter and no flower production in that period, followed by a decrease in expression in spring and summer, which are flowering periods [40].

In relation to EMF2, the action is mainly in the suppression of the early flowering by means of flowering genes repression. According to Yoshida et al., [41] mutants of *Arabidopsis* were silenced from *EMF2* and *EMF1* gene, so the decrease in the activity of the EMF proteins allowed the flowering. EMF protein is part of the *Polycomb group (PCG)* which is relatively conserved in plants. Thus, EMF2 protein possibly does not develop the function in the H011, although it is identified in the genotype, since the flowering occurs very early.

We also identified in the H011 proteins from the floral integrator pathway CIFT2, CIFT3, SOC1 and the meristematic identity pathway AP1, BAM1 and BAM2 (Table 2). These proteins are the most studied in relation to flowering in citrus. SOC1, CIFT2 and CIFT3 are fundamental to promote the modification of the vegetative meristem in the reproductive meristem. Proteins CIFT2 and CIFT3 are produced mainly in leaf tissue; transport to the apical region of the stem through the phloem [8,42]. When reaching the meristematic regions they activate the expression of the genes involved in the meristematic identity pathway and initiate the formation of flower structures such as petals [2]. According to Almeida [43], the H011 shows CIFT expression in the summer and in the winter, when the expression values were superior compared to the summer. These data, associated to the found proteins, corroborate with the characteristic of constant flowering in all the seasons of the year in the H011 induced mainly through the ICTP expression.

The identification of the large variety of flowering-related proteins in the H011 allows new information mainly on GA pathways, photoperiod, autonomous and citrus vernalization. Most of the available knowledge relates to *Arabidopsis*. From this study, the confirmation of the existence of such proteins in another citrus genotype subsidizes the development of a greater amount of work related to these pathways in citrus. In addition, this is the first work investigating the expression of

Table 2
Microcitrangemonia hybrid transcripts that share homology with proteins related to the gibberellin pathway, photoperiod, autonomous, vernalization, floral integrators and meristematic identity.

pathway	Protein	Transcripts	E-value	Coverage	Identity
Gibberellin	DELLA	TRINITY_DN17875_c0_g1_i1	0	100	99,653
		TRINITY_DN17280_c0_g1_i1	0	95	62,478
		TRINITY_DN15517_c0_g1_i1	1,4E-14	79	55,556
Photoperiod	LHY	TRINITY_DN18818_c0_g1_i2	9,87E-51	93	68,421
		TRINITY_DN18818_c0_g1_i8	2,32E-19	93	82
		TRINITY_DN18818_c0_g1_i1	1,09E-49	93	69,006
	ZTL	TRINITY_DN14483_c0_g1_i1	0	95	86,201
	GI	TRINITY_DN18149_c0_g1_i5	0	99	75,639
		TRINITY_DN18149_c0_g1_i2	0	99	75,639
		TRINITY_DN18149_c0_g1_i1	0	99	75,639
		TRINITY_DN18149_c0_g1_i6	0	98	75,92
		TRINITY_DN18149_c0_g1_i4	0	98	75,92
	CCA	TRINITY_DN18818_c0_g1_i2	2,36E-47	90	54,751
		TRINITY_DN18818_c0_g1_i1	3,91E-47	91	54,587
	TOC1	TRINITY_DN18114_c1_g2_i1	2,67E-157	95	50,164
	PHYA	TRINITY_DN15125_c0_g1_i1	0	68	55,802
	PHYB	TRINITY_DN17176_c0_g1_i2	0	94	78,558
	CRY1	TRINITY_DN17773_c0_g1_i1	0	98	98,595
	CRY2	TRINITY_DN16320_c0_g1_i1	0	92	66,9
	CRY3	TRINITY_DN16320_c0_g1_i2	0	92	66,9
	CONSTANS	TRINITY_DN16882_c0_g1_i1	5,39E-109	100	56,806
		TRINITY_DN16882_c0_g1_i2	3,43E-108	100	56,728
Autonomy	FCA	TRINITY_DN13003_c0_g1_i1	1,22E-78	63	98,37
	FLC	TRINITY_DN16126_c0_g1_i1	2,14E-89	94	87,958
		TRINITY_DN16126_c0_g1_i2	2,26E-91	97	93,464
	SVP	TRINITY_DN12354_c0_g1_i1	1,27E-113	82	99,441
		TRINITY_DN12354_c0_g1_i2	3,2E-111	82	98,883
		TRINITY_DN17088_c0_g1_i3	2,47E-71	99	59,211
		TRINITY_DN17088_c0_g1_i4	2,89E-71	99	59,211
		TRINITY_DN17088_c0_g1_i2	2,69E-71	99	59,649
		TRINITY_DN17088_c0_g1_i6	4,96E-88	99	59,664
		TRINITY_DN17088_c0_g1_i7	7,07E-82	99	53,184
	TFL	TRINITY_DN9698_c0_g2_i1	1,68E-60	99	54,857
		TRINITY_DN9698_c0_g1_i1	1,96E-38	61	58,333
		TRINITY_DN13672_c0_g1_i1	1,97E-87	94	70,414
Vernalization	EMF2	TRINITY_DN16261_c0_g2_i1	0	99	65,991
		TRINITY_DN16261_c0_g3_i2	2,02E-12	62	59,375
		TRINITY_DN16261_c0_g3_i1	2,03E-12	62	59,375
	SUF3	TRINITY_DN18474_c3_g1_i1	4,27E-152	70	70,684
	SUF4	TRINITY_DN16182_c0_g1_i2	1,16E-131	89	63,26
		TRINITY_DN16182_c0_g1_i1	2,16E-128	99	59,553
	VIP3	TRINITY_DN11875_c0_g1_i1	6,48E-73	100	75
	VIP4	TRINITY_DN16441_c0_g1_i2	9,17E-161	99	62,208
		TRINITY_DN16441_c0_g1_i1	1,03E-160	99	62,208
	VIP5	TRINITY_DN14172_c0_g1_i1	0	99	65,177
	VIP6	TRINITY_DN16513_c1_g2_i1	0	60	78,605
Floral integrators	CIFT2	TRINITY_DN13672_c0_g1_i1	2,67E-56	95	56,14
		TRINITY_DN9698_c0_g2_i1	1,12E-45	93	52,663
		TRINITY_DN9698_c0_g1_i1	7,88E-26	60	58,824
	CIFT3	TRINITY_DN13672_c0_g1_i1	3,11E-57	94	57,396
		TRINITY_DN9698_c0_g2_i1	1,12E-44	93	51,479
		TRINITY_DN9698_c0_g1_i1	1,58E-25	60	57,843
	SOC1	TRINITY_DN16069_c0_g1_i1	0	97	97,212
Meristematic identity	AP1	TRINITY_DN5887_c0_g2_i1	0	100	98,825
		TRINITY_DN5887_c0_g1_i1	0	100	98,825
	BAM1	TRINITY_DN11169_c0_g1_i1	0	74	99,338
	BAM2	TRINITY_DN15623_c0_g1_i1	0	71	98,539

flowering genes for the H011. Thus, from the knowledge and the hypotheses, new investigations are possible on the H011, demonstrating the expression of important genes related to flowering and that can be better explored for the creation of new varieties in programs of genetic improvement of citrus through of the manipulation of these genes and consequently of the manifestation of the characteristic.

3.3. Similarity between different plant species for SOC1, AP1 and FLC

There have been genomic, transcriptomic and proteomic studies related to flowering for over a decade; however, the availability of complete gene and protein sequences linked to floral induction for most plant genera in databases is still relatively low. Proteins belonging to repressor pathways such as FLC, integrators such as SOC1 and meristematic identity such as AP1 are among the proteins with the greatest amount of sequences for plants in the databases, being thus, selected for this study.

The analysis of similarity for each protein allowed inferring about the evolutionary proximity of the H011 in relation to other species of citrus and other plant species such as *Arabidopsis thaliana*. Different studies have reported the conservation of genes related to floral regulation among plant families [27,44]. Data from Pillitteri et al., [7] show that for the Washington Navel orange the amino acid sequences from the genes CsLFY and CsAP1 showed evolutionary proximity and

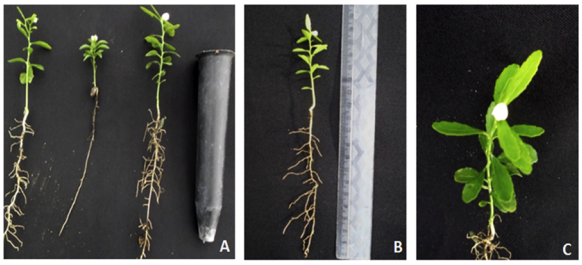

Fig. 5. Seedling of the hybrid Microcitrangemonia in flowering; A- Three replications with three months of life and with floral buds; B- Seedling with 19 cm long with floral bud; C- Seedling with two months of life and floral bud at the apical end.

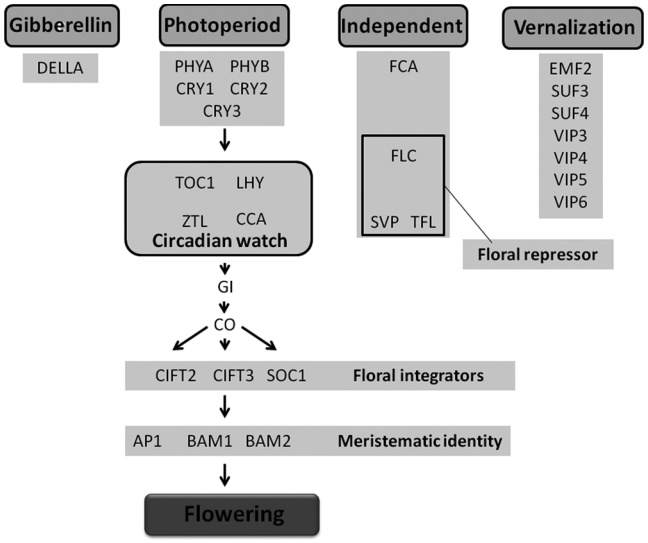

Fig. 6. Schematic view of the interaction between flowering pathways (Zhang et al., 2011), proteins identified in the hybrid Microcitrangemonia.

similarity greater than 65% for *A. thaliana*. H011 showed different results with greater evolutionary similarity with other species.

In relation to AP1, SOC1 and FLC, the H011 had higher identity with *Citrus sinensis* (Fig. 7), sharing an ancestor with less evolutionary time when compared to the other species. After *C. sinensis* for AP1, the species with the highest molecular similarity were *Paeonia lactiflora* (peony), *Malus domestica* (apple) and *Eriobotrya japonica* (yellow or loquat plum). In relation to SOC1, the species with the highest molecular similarity were *Manihot esculenta* (cassava) and *Jatropha curcas* and for FLC, the species were *Poncirus trifoliata* (citrus) and *Dimocarpus longan*. *A. thaliana* was further apart in the cladogram in relation to the three investigated proteins, showing greater evolutionary divergence.

The evolutionary proximity of the H011 with *C. sinensis* to the studied proteins relates to the H011 origin, which originates from the crosses of [Lemon Tree Rangpur (LCR) X Citrus 'Yuma' (CTYM) – 005) X *Microcitrus papuana* (MCP) – 011]. The female parental (LCR X CTYM-005) shows a gene composition of *Citrus limonia* (LCR), *C. sinensis* and *P. trifoliata,* since CTYM-005 is a hybrid (*C. sinensis* x *P. trifoliata)*. The male genitor provided the genetic material of *M. papuana* [45,46], so all the species involved in the creation of the H011, as well as the H011 itself belong to the Rutaceae family.

The other plant species that show evolutionary proximity to the H011 belong to different families, except *P. trifoliate*. *P. lactiflora* is a perennial herb from the family Paeoniaceae [47], *M. domestica* and *E. japonica* are part of the family Rosaceae [48], M. *esculenta* and *J. curcas* are from the family Euphorbiaceae [49] and *D. longan* from the family Sapindaceae [50]. However, all families, including the Rutaceae, are within the same clade, the Rosids. A heterogeneous clade supported by molecular analysis of different genes, which justifies the similarity in the composition of the AP1, SOC1 and FLC sequences between the H011 and the previously mentioned species belonging to different families.

The knowledge obtained on the evolutionary similarity of the H011 transcripts promotes a foundation for future studies on gene expression or transformation that may involve the H011 to species close in the evolutionary context and proteins involved with the flowering process. Therefore, this study is an important source of information for future research.

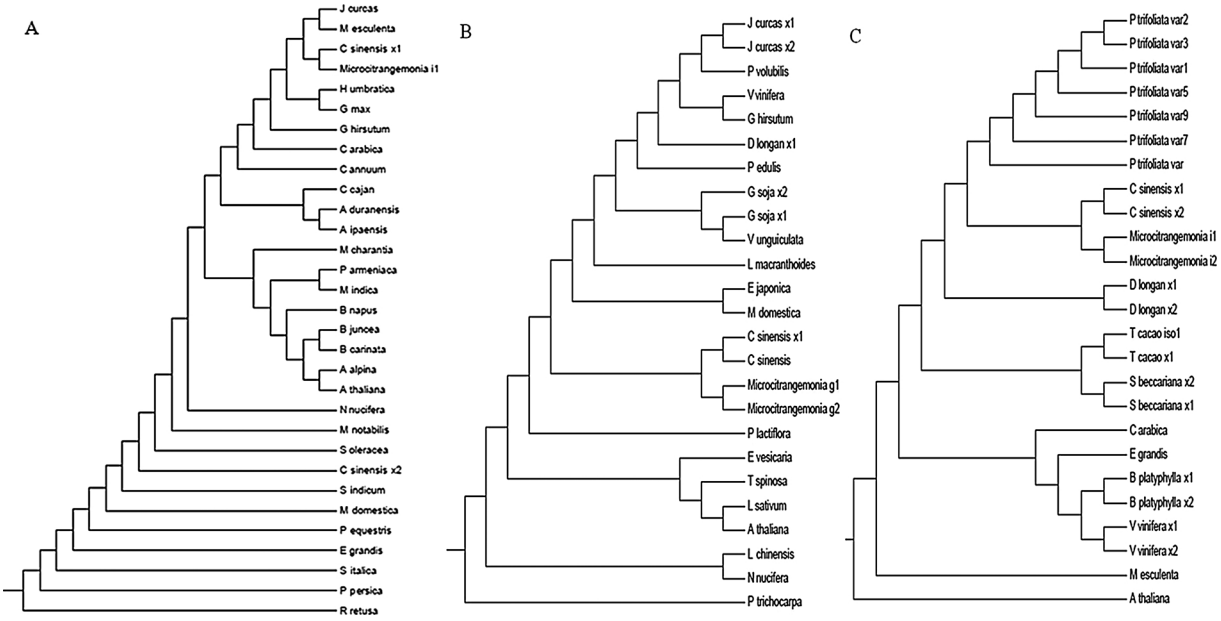

Fig. 7. Cladogram of different species of citrus and other plants for three flowering-related proteins. (A) SOC- floral integrator (B) AP1- meristematic identity (C) FLC - flowering repressor (autonomous pathway). Grouping occurred using the UPGMA method.

4. Conclusion

In the PB category are the largest number of H011 transcripts involved in the flowering process, classified as floral repression and induction, but phenotypically repression was not observed. Genes involved in the flowering repression process, such as FLC, demonstrated a non-effective action for FT repression, unlike the FLC action in Arabidopsis and other citrus species. In the gibberellin, photoperiod, autonomic and vernalization pathways it was proven the expression of different genes that make up each pathway, the transcripts of such genes were identified in H011, most previously found more frequently only in Arabidopsis. Most of the transcripts showed evolutionary similarity with Citrus sinensis mainly for proteins SOC1, AP1 and FLC e the information obtained contributes to the understanding of the physiological behavior of H011 in relation to the transition from the vegetative to the reproductive period and it becomes an important source of knowledge mainly for citrus breeding programs that seek to study the reduction of the juvenile period and gene transformation.

Conflict of Interest

The authors have declared that no competing interests exist.

Acknowledgements

National Council for Scientific and Technological Development - CNPq (grant: 301356/2012-2 and 472733/2013-3). Scholarship of E.M.R. Sousa financed by CAPES

References

[1] F. Nishikawa, Regulation of floral induction in citrus, Journal of the Japanese Society for Horticultural Science. 82 (2013) 283–292, https://doi.org/10.2503/jjshs1.82.283.
[2] K. Velázquez, J. Agüero, M.C. Vives, P. Aleza, J.A. Pina, P. Moreno, L. Navarro, J. Guerri, Precocious flowering of juvenile citrus induced by a viral vector based on Citrus leaf blotch virus: a new tool for genetics and breeding, Plant biotechnology journal 14 (2016) 1976–1985, https://doi.org/10.1111/pbi.12555.
[3] D.L.G. Correa, J.E. Vélez-Sánchez, J.O. Orduz Rodríguez, Influence of water deficit on growth and development of fruits valencia orange (*Citrus sinensis* Osbeck) in the piedmont of Meta department, Colombia, Acta Agronómica. 62 (2013) 136–147.
[4] R. Goldberg-Moeller, L. Shalom, L. Shlizerman, S. Samuels, N. Zur, R. Ophir, E. Blumwald, A. Sadka, Effects of gibberellin treatment during flowering induction period on global gene expression and the transcription of flowering-control genes in Citrus buds, Plant science 198 (2013) 46–57, https://doi.org/10.1016/j.plantsci.2012.09.012.
[5] Y. Komeda, Genetic regulation of time to flower in Arabidopsis thaliana, Annu. Rev. Plant Biol. 55 (2004) 521–535, https://doi.org/10.1146/annurev.arplant.55.031903.141644.
[6] D.S. Ó'Maoiléidigh, E. Graciet, F. Wellmer, Gene networks controlling *Arabidopsis thaliana* flower development, New Phytologist. 201 (2014) 16–30, https://doi.org/10.1111/nph.12444.
[7] L.J. Pillitteri, C.J. Lovatt, L.L. Walling, Isolation and Characterization of LEAFY and APETALA1 Homologues from *Citrus sinensis* L. OsbeckWashington', Journal of the American Society for Horticultural Science 129 (2004) 846–856.
[8] L.J. Pillitteri, C.J. Lovatt, L.L. Walling, Isolation and characterization of a TERMINAL FLOWER homolog and its correlation with juvenility in citrus, Plant Physiology 135 (2004) 1540–1551, https://doi.org/10.1104/pp.103.036178.
[9] F.C. Tan, S.M. Swain, Functional characterization of AP3, SOC1 and WUS homologues from citrus (*Citrus sinensis*), Physiologia plantarum 131 (2007) 481–495, https://doi.org/10.1111/j.1399-3054.2007.00971.x.
[10] R.K. Patel, M. Jain, NGS QC Toolkit: a toolkit for quality control of next generation sequencing data, PloS one 7 (2012) e30619, , https://doi.org/10.1371/journal.pone.0030619.
[11] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.
[12] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, Nature biotechnology 29 (2011) 644, https://doi.org/10.1038/nbt.1883.
[13] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nature methods 12 (2014) 59, https://doi.org/10.1038/nmeth.3176.
[14] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676, https://doi.org/10.1093/bioinformatics/bti610.
[15] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic acids research 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[16] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, MEGA 7: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol Biol Evol 28 (2011) 2731–2739, https://doi.org/10.1093/molbev/msr121.
[17] G.A. Wu, S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol, M.A. Takita, Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication, Nature biotechnology 32 (2014) 656, https://doi.org/10.1038/nbt.2906.
[18] Q. Xu, L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B.H. Hao, M.P. Lyon, J. Chen, The draft genome of sweet orange (*Citrus sinensis*), Nature genetics 45 (2013) 59, https://doi.org/10.1038/ng.2472.
[19] T. Gabaldón, A.A. Pittis, Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes, Biochimie 119 (2015) 262–268, https://doi.org/10.1016/j.biochi.2015.03.021.
[20] M. Müller, M. Mentel, J.J. van Hellemond, K. Henze, C. Woehle, S.B. Gould, R.Y. Yu, M. van der Giezen, A.G. Tielens, W.F. Martin, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes, Microbiology and Molecular Biology Reviews 76 (2012) 444–495, https://doi.org/10.1128/MMBR.05024-11.
[21] F.D. Mast, L.D. Barlow, R.A. Rachubinski, J.B. Dacks, Evolutionary mechanisms for establishing eukaryotic cellular complexity, Trends in cell biology 24 (2014) 435–442, https://doi.org/10.1016/j.tcb.2014.02.003.
[22] Y. Li, D. Shin, S.H. Kwon, Histone deacetylase 6 plays a role as a distinct regulator of diverse cellular processes, The FEBS journal 280 (2013) (2013) 775–793, https://doi.org/10.1111/febs.12079.
[23] J.E. Pérez-Ortín, P. Alepuz, S. Chávez, M. Choder, Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression, Journal of Molecular Biology. 425 (2013) 3750–3775, https://doi.org/10.1016/j.jmb.2013.02.029.
[24] M.K. Potdar, G.F. Kelso, L. Schwarz, C. Zhang, M.T. Hearn, Recent developments in chemical synthesis with biocatalysts in ionic liquids, Molecules. 20 (2015) 16788–16816, https://doi.org/10.3390/molecules200916788.
[25] F. Rashid, A. Shah, G. Shan, Long non-coding RNAs in the cytoplasm, Genomics, proteomics & bioinformatics 14 (2016) 73–80, https://doi.org/10.1016/j.gpb.2016.03.005.
[26] A.K. Krabberød, R.J. Orr, J. Bråte, T. Kristensen, K.R. Bjørklund, K. Shalchian-Tabrizi, Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in Rhizaria, Molecular biology and evolution 34 (2017) 1557–1573, https://doi.org/10.1093/molbev/msx075.
[27] Pajon M, V.J. Febres, G.A. Moore, Expression patterns of flowering genes in leaves of 'Pineapple'sweet orange [*Citrus sinensis* (L.) Osbeck] and pummelo (*Citrus grandis* Osb eck), BMC plant biology 17 (2017) 146, https://doi.org/10.1186/s12870-017-1094-3.
[28] N. Yamagishi, N. Yoshikawa, Expression of FLOWERING LOCUS T from Arabidopsis thaliana induces precocious flowering in soybean irrespective of maturity group and stem growth habit, Planta 233 (2011) 561–568, https://doi.org/10.1007/s00425-010-1318-3.
[29] W. Deng, H. Ying, C.A. Helliwell, J.M. Taylor, W.J. Peacock, E.S. Dennis, FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis, Proceedings of the National Academy of Sciences 108 (2011) (2011) 6680–6685, https://doi.org/10.1073/pnas.1103175108.
[30] Y.H. Song, J.S. Shim, H.A. Kinmonth-Schultz, T. Imaizumi, Photoperiodic flowering: time measurement mechanisms in leaves, Annual review of plant biology 29 (2015) 441–464, https://doi.org/10.1146/annurev-arplant-043014-115555.
[31] F. Fornara, K.C. Panigrahi, L. Gissot, N. Sauerbrunn, M. Rühl, J.A. Jarillo, G. Coupland, Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response, Developmental cell 17 (2009) 75–86, https://doi.org/10.1016/j.devcel.2009.06.015.
[32] F. Fornara, A. Montaigu, G. Coupland, SnapShot: control of flowering in Arabidopsis, Cell 141 (2010) 550, https://doi.org/10.1016/j.cell.2010.04.024.
[33] J. Lu, T. Wang, Z. Xu, L. Sun, Q. Zhang, Genome-wide analysis of the GRAS gene family in Prunus imume, Mol. Genet. Genomics 290 (2015) 303–317, https://doi.org/10.1007/s00438-014-0918-1.
[34] Z.H. Bian, Q.C. Yang, W.K. Liu, Effects of light quality on the accumulation of phytochemicals in vegetables produced in controlled environments: a review, Journal of the Science of Food and Agriculture. 30 (2015) 869–877, https://doi.org/10.1002/jsfa.6789.
[35] D. Alabadí, T. Oyama, M.J. Yanovsky, F.G. Harmon, P. Más, S.A. Kay, Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock, Science 293 (2001) 880–883, https://doi.org/10.1126/science.1061320.
[36] W.Y. Kim, S. Fujiwara, S.S. Suh, J. Kim, Y. Kim, L. Han, K. David, J. Putterill, H.G. Nam, D.E. Somers, ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light, Nature 449 (2007) 356, https://doi.org/10.1038/nature06132.
[37] M.C. Dornelas, R.L. Camargo, L.H. Figueiredo, M.A. Takita, A genetic framework for flowering-time pathways in Citrus spp, Genetics and Molecular Biology. 30 (2007) 769–779, https://doi.org/10.1590/S1415-47572007000500006.
[38] M.J. Kang, H.S. Jin, Y.S. Noh, B. Noh, Repression of flowering under a noninductive photoperiod by the HDA9-AGL19-FT module in Arabidopsis, New Phytologist. 206 (2015) 281–294, https://doi.org/10.1111/nph.13161.
[39] K. Choi, J. Kim, H.J. Hwang, S. Kim, C. Park, S.Y. Kim, I. Lee, The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by

recruiting chromatin modification factors, The Plant Cell. 23 (2011) (2011) 289–303, https://doi.org/10.1105/tpc.110.075911.
[40] J.Z. Zhang, Z.M. Li, L. Mei, J.L. Yao, C.G. Hu, PtFLC homolog from trifoliate orange (Poncirus trifoliata) is regulated by alternative splicing and experiences seasonal fluctuation in expression level, Planta 229 (2009) 847–859, https://doi.org/10.1007/s00425-008-0885-z.
[41] N. Yoshida, Y. Yanai, L. Chen, Y. Kato, J. Hiratsuka, T. Miwa, Z.R. Sung, S. Takahashi, EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis, The Plant Cell 13 (2001) (2001) 2471–2481, https://doi.org/10.1105/tpc.010227.
[42] F. Nishikawa, T. Endo, T. Shimada, H. Fujii, T. Shimizu, M. Omura, Y. Ikoma, Increased CiFT abundance in the stem correlates with floral induction by low temperature in Satsuma mandarin (*Citrus unshiu* Marc.), Journal of experimental botany 58 (2007) 3915–3927, https://doi.org/10.1093/jxb/erm246.
[43] L.S. Almeida, Estudo morfológico e análise da expressão do gene Flowering Locus T em um híbrido de citros com florescimento precoce, Dissertation (Masters in Genetics and Molecular Biology) Universidade Estadual de Santa Cruz, 2016.
[44] P. Winterhagen, P. Tiyayon, A. Samach, M. Hegele, J.N. Wünsche, Isolation and characterization of FLOWERING LOCUS T subforms and APETALA1 of the subtropical fruit tree *Dimocarpus longan*, Plant physiology and biochemistry 71 (2013) 184–190, https://doi.org/10.1016/j.plaphy.2013.07.013.
[45] W.S. Soares Filho, C.A.S. Ledo, M.P. Quintela, L.A. Mattos, O.S. Passos, A.S. Souza, Cruzamentos em citros: frequência e vigor de híbridos, Revista Brasileira de Fruticultura 29 (2007) 393–398.
[46] T. Penjor, M. Yamamoto, M. Uehara, M. Ide, N. Matsumoto, R. Matsumoto, Y. Nagano, Phylogenetic relationships of Citrus and its relatives based on matK gene sequences, PloS one 8 (2013) e62574, , https://doi.org/10.1371/journal.pone.0062574.
[47] M. Assadi, A taxonomic revision of the genus Paeonia (Paeoniaceae) in Iran, Iranian journal of botany 22 (2016) 75–78.
[48] D. Potter, T. Eriksson, R.C. Evans, S. Oh, J.E.E. Smedmark, D.R. Morgan, M. Kerr, K.R. Robertson, M. Arsenault, T.A. Dickinson, C.S. Campbell, Plant systematics and evolution 266 (2007) (2007) 5–43, https://doi.org/10.1007/s00606-007-0539-9.
[49] G.L. Webster, Classification of the Euphorbiaceae, Annals of the Missouri Botanical Garden. (1994) 3–32, https://doi.org/10.2307/2399908.
[50] S. Buerki, F. Forest, P. Acevedo-Rodríguez, M.W. Callmander, J.A. Nylander, M. Harrington, I. Sanmartín, P. Küpfer, N. Alvarez, Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae), Molecular Phylogenetics and Evolution. 51 (2009) 238–258, https://doi.org/10.1016/j.ympev.2009.01.012.