Plant Mol Biol (2017) 94:361–379
DOI 10.1007/s11103-017-0612-x

Analysis of global gene expression profiles during the flowering initiation process of *Lilium*×*formolongi*

Yu-Fan Li[1] · Ming-Fang Zhang[1] · Meng Zhang[1] · Gui-Xia Jia[1]

Received: 24 May 2016 / Accepted: 11 April 2017 / Published online: 20 April 2017

Abstract The onset of flowering is critical for the reproductive development of plants. *Lilium*×*formolongi* is a lily hybrid that flowers within a year after sowing. We successfully identified four important stages during vegetative growth and flowering initiation of *L.*×*formolongi* under long day conditions. The plant tissues from the four stages were used in a genome-wide transcriptional analysis to investigate stage-specific changes of gene expression in *L.*×*formolongi*. In total, the sequence reads of the four RNA-sequencing libraries were assembled into 52,824 unigenes, of which 37,031 (70.10%) were differentially expressed. The global expression dynamics of the differentially expressed genes were predominant in flowering induction phase I and the floral differentiation stage, but down-regulated in flowering induction phase II. Various transcription factor families relevant to flowering were elucidated, and the members of the MADS-box, SBP and CO-like transcription factor families were the most represented. There were 85 differentially expressed genes relevant to flowering. *CONSTANS-LIKE, FLOWERING LOCUS T, TREHALOSE-6-PHOSPHATE SYNTHASE* and *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* homologs were discovered and may play significant roles in the flowering induction and transition process of *L.*×*formolongi*. A putative gene regulatory network, including photoperiod, age-dependent and trehalose-6-phosphate flowering pathways, was constructed. This is the first expression dataset obtained from a transcriptome analysis of photoperiod-mediated flowering pathway in lily, and it is valuable for the exploration of the molecular mechanisms of flowering initiation and the short vegetative stage of *L.*×*formolongi*.

Keywords *Lilium*×*formolongi* · Flowering induction · Photoperiod · Gene expression · RNA-seq

Introduction

Lilies are important ornamental plants that belong to the genus *Lilium*, which contains abundant species and cultivars. Commercial potted and cut lily cultivars are predominantly vegetatively-propagated (bulbs, bulb divisions and aerial axillary stem bulbils) rather than sexually (seed) (Beattie and White 1993; Griffiths 1934), and the seedlings of most lily species require a long vegetative growth stage of more than 1 year before flowering.

Seed-propagated white trumpet lilies, *Lilium*×*formolongi*, are important commercial cut flowers that are obtained from repeated interspecific crosses of *Lilium formosanum* Wallace×*Lilium longiflorum* (Okazaki 1994). This lily hybrid possesses the characteristic of flowering within a year after sowing (200–240 days) without vernalization, but is sensitive to photoperiod (Anderson et al. 2012; Sakamoto 2005). These characteristics are important to the development of seed-propagated,

Yu-Fan Li and Ming-Fang Zhang have contributed equally to this study and share first authorship.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-017-0612-x) contains supplementary material, which is available to authorized users.

✉ Gui-Xia Jia
gxjia@bjfu.edu.cn

[1] Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment and College of Landscape Architecture, Beijing Forestry University, Beijing 100083, China

non-vernalization-requiring colored lily hybrids (Optiz et al. 2009). Because of the desirable characteristics of *L.*×*formolongi*, most reported research focuses on the cultivation and cross-breeding of this plant (asl Hamid and Kim 2011; Horita et al. 2003). However, the molecular regulatory mechanism of flowering is largely unknown, as is the genetic mechanism of blossoming in 1 year of *L.*×*formolongi*.

Studies on the biogenetics of flowering time in the model plant *Arabidopsis thaliana* have shown that the flower development period can be divided into several stages, with the first being flowering induction, in which the shoot apical meristem (SAM) remains in a vegetative state (Meyerowitz et al. 1991). In this process, the plant responds to internal and environmental signals, including the age of the plant, its nutritional state, day length and temperature. These physiological and genetic activities determine the change of SAM from vegetative to floral (Meyerowitz et al. 1991). According to our previous research, the seedlings of *L.*×*formolongi* are sensitive to the long day (LD) flowering pathway. In addition, the characteristic of prompt blossoming in 1 year from seeds indicates that the age-dependent and carbohydrate nutritional flowering pathways may also play important roles in the flowering initiation of *L.*×*formolongi*.

Genetic analyses of *A. thaliana*, a facultative long-day plant, have revealed many photoperiodic genes, including *CRYPTOCHROME 1* (*CRY1*), *CONSTANS* (*CO*), *GIGANTEA* (*GI*), *FLOWERING LOCUS T* (*FT*) and *FLOWERING LOCUS D* (*FD*) (Turck et al. 2008). The circadian rhythm expression pattern of *CO*/*FT* is crucial in triggering flowering induction (Suárez-López et al. 2001; Meng et al. 2011). *CO* is a member of the *CONSTANS-LIKE* (*COL*) gene family, whose members encode putative zinc finger transcription factors (TFs) (Putterill et al. 1995). *CO* transcripts are regulated by the circadian clock and are diurnally regulated with a peak at night under both LDs and short days (SDs) (Andrés and Coupland 2012; Suárez-López et al. 2001). However, there is an additional peak of *CO* mRNA that occurs 12–16 h after dawn (in the light) under LD conditions, which is essential for the day-length-dependent promotion of flowering (Suárez-López et al. 2001). The CO protein in the dark is degraded by an ubiquitin ligase complex. Thus, only the peak of *CO* mRNA that occurs in the light at the end of a LD leads to CO protein accumulation (Andrés and Coupland 2012; Suárez-López et al. 2001). Subsequently, the accumulation of CO protein drives transcript accumulation of *FT* mRNA at the end of day under LD conditions, but not SD conditions, and triggers the flowering induction under LDs (Andrés and Coupland 2012; Kardailsky et al. 1999).

Recently, a new age-dependent flowering pathway, which is regulated by two microRNAs (miR156, miR172) and their target genes *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) and *APETALA2* (*AP2*)-like, was revealed in model plants (Bergonzi et al. 2013; Khan et al. 2014). Furthermore, sugar is an endogenous signal for the vegetative to reproductive stage transition in plants, and trehalose-6-phosphate (T6P) has been suggested to function as a proxy for carbohydrate status (O'Hara et al. 2013; Proveniers 2013). A recent study revealed that the *TREHALOSE-6-PHOSPHATE SYNTHASE 1* (*TPS1*) gene catalyzes the formation of T6P and the loss of *TPS1* causes *A. thaliana* to flower extremely late, even under inductive environmental conditions (Wahl et al. 2013).

The genome size of lily is very large, at ~36 Gb, which is ~300 times that of the *A. thaliana* genome (Du et al. 2015). Whole transcriptome sequencing using next-generation RNA sequencing (RNA-seq) technology is a convenient and rapid means to study gene expression at the whole-genome level and defines putative gene functions (Jain 2011; Ozsolak and Milos 2011). Several transcriptome studies have focused on the relevance of vernalization of lily bulbs to flowering (Huang et al. 2014; Liu et al. 2014; Villacorta-Martin et al. 2015). One report on the transcriptome profiles of *L. longiflorum* bulbs at several time points during cold exposure presented a detailed analysis of gene expression dynamics during vernalization and revealed some candidate vernalization genes in lily (Villacorta-Martin et al. 2015).

In addition to vernalization, photoperiod is also important for the flowering transition of *Lilium*. LD conditions could hasten the flowering of *L. longiflorum*. Additionally, small non-vernalized *L. longiflorum* bulbs could flower under LD conditions, but not under SD conditions (Lazare and Zaccai 2016). However, the molecular mechanism of the photoperiodic flowering pathway in lily has not been explored yet. And the internal transition stage from vegetative to reproductive growth, when the morphology of the SAM still remains in a vegetative status, has not been identified.

In this study, the different developmental stages of vegetative juvenile, flowering induction and floral differentiation of *L.*×*formolongi* seedlings were successfully identified under LD conditions. Additionally, plant tissues of the four stages were sampled for high-throughput Illumina sequencing to analyze the global gene expression profiles during the flowering initiation process. Based on the extensive data analyses, critical flowering homologs from several regulatory pathways were identified for the first time. The data set will serve as a foundation to understand the photoperiod-mediated, age-dependent and carbohydrate nutritional molecular mechanisms of flowering regulation, and will help to discern the molecular characteristics of the fast transition from vegetative to reproductive growth within a year from seeds of *L.*×*formolongi*.

Material and method

Plant materials and crucial stages confirmation during vegetative growth and flowering initiation

The seeds of *L.*×*formolongi* cv. Raizan 2 were sown in seedling-raising trays after stratification storage at 4 °C for a month. The seeds germinated and grew in a light growth chamber at 25/18 °C day/night with 70% humidity and were illuminated with white light (320 μmol m^{-2} s^{-1} photosynthetic photon flux density) under a 16/8-h photoperiod. At approximately 2 months after sowing, after the seedlings grew 2–3 rosette leaves, each seedling was transplanted into an 8-cm pot. At nearly 4 months after sowing, when the average number of rosette leaves was 5–6, the seedlings were transplanted into 15-cm pots. All of the seedlings were carefully monitored until flowering.

To identify the different stages of *L.*×*formolongi* seedlings during vegetative growth and flowering initiation under LD conditions, the phenotypic traits and anatomical structures of the SAM were investigated weekly from seeds germination to visible floral bud appearance. The phenotypic traits included the average number of rosette leaves before bolting and internodes after bolting, as well as the average plant height. The phenotypic traits were observed and recorded among 30 seedlings each week. Additionally, three stem apices were sampled from three seedlings with conforming phenotypic traits, and the samples were stored in Formalin–acetic acid–alcohol liquid (70% ethyl alcohol:glacial acetic acid:methanol; 18:1:1) for permanent paraffin sections to observe the course of flower differentiation at each week. Additionally, each 2 g sample of fresh leaves and stem apices were collected from the conforming seedlings and sampled for three biological replicates as the candidate materials for RNA-seq and qRT-PCR at each week. The leaves and stem apices were frozen in liquid nitrogen and stored at −80 °C separately.

Moreover, according to studies in *A. thaliana*, the circadian rhythm expression patterns of *CO* and *FT* are important biomarkers of flowering induction and transition (Andrés and Coupland 2012; Meng et al. 2011). The *COL* homolog *LfCOL9* (KJ744206) and *FT* homolog *LfFT1* (KJ744207) were cloned from *L.*×*formolongi* cv. Raizan 3 in our previous study. At the 2–3 and 9–10 rosette leaf stages and the 1–2 internode stage, the *LfCOL9* and *LfFT1* mRNA accumulations in the leaf blades were investigated by qRT-PCR every 4 h over 3 days. At each time point, leaves from three seedlings were collected as biological replicates and stored at −80 °C for RNA extraction.

Total RNA was extracted using the kits from Aidlab Biotechnology Co., LTD., China, according to the manufacturer's instructions. The quality of RNA was characterized on a 1% agarose gel electrophoresis and verified using an ND-1000 Spectrophotometer (NanoDrop, Wilmington, DE). The standards applied were $1.8 \leq OD260/280 \leq 2.2$ and $OD260/230 \geq 1.8$. cDNA was synthesized using a PrimeScript Double Strand cDNA synthesis kit (Takara, Dalian, China) according to the manufacturer's instructions. *Aquaporin TIP4-1* was selected as reference gene to standardize the results because it had shown the most stable expression in different developmental stages in lily according to our previous experiment. All the primers are listed in Table S1. qRT-PCR was performed in a Bio-Rad Connect Real-Time PCR Detection System using SYBR Premix EX Taq II Kit (TaKaRa) according to the manufacturer's protocol. All the samples were run in triplicates in 96-well optical reaction plates (Bio-rad, USA).Relative expression was determined by CT values and calculated by $2^{-\Delta\Delta CT}$ algorithm based on the expression of *Aquaporin TIP4-1*.

Illumina sequencing and gene functional annotation

The leaves and stem apexes from the four identified stages, including vegetative juvenile (VJ), flowering induction phase I (FI-1), flowering induction phase II (FI-2) and floral differentiation (FLD) stage were used for constructing four cDNA libraries for RNA-seq. Total RNA of the leaves and stem apexes from the same stage were extracted separately and then mixed in equal amounts, using a TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. RNA concentration and integrity number ($RIN > 7.5$) were evaluated on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Four cDNA libraries were constructed from the samples of four stages in which equal amounts of total RNA were pooled from three biological replicates per stage. The cDNAs were synthesized according the protocols of the Beijing Genome Institute (BGI; Shenzhen, China), then the four cDNA libraries were subjected to paired-end sequencing on an Illumina HiSeq™ 2000 platform. The clean reads sequenced from every cDNA library were assembled de novo to become unigenes using Trinity platform with parameters of 'K-mer = 25, group pairs distance = 250' (Grabherr et al. 2011). TGICL software was used to remove redundant sequences from unigenes and assembled them as long as possible (Pertea et al. 2003). Finally, all unigenes from the four gene expression libraries were assembled again to a single set of non-redundant library called All-Unigene, which was to be a reference library for differential gene expression analysis among the four libraries.

Furthermore, All-Unigene was aligned to a series of protein databases (E-value $<1e^{-5}$) including NCBI non-redundant protein database (Nr), Swiss-Prot database, Kyoto Encyclopedia of Genes and Genomes protein database (KEGG) and Clusters of orthologous Groups of proteins

database (COG) by BLASTx and an NCBI nucleotide database (Nt) (E-value $<1e^{-5}$) by BLASTn to obtain functional annotation for all unigenes. The best aligned results were used to decide the direction and protein coding-region (CDS) of the unigene sequences. Furthermore, all the unigenes were subjected to Blast2GO program to obtain gene ontology (GO) annotations and classifications (http://www.blast2go.org/).

Differential gene expression analysis, and gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses

Clean reads from each library of the four stages were mapped to All-Unigene, which is a reference database, using the short oligonucleotide analysis package SOAP aligner/soap2. The number of fragments that were uniquely aligned was normalized using the fragments per kilobase of transcript per million mapped reads (FPKM) method to calculate the unigene expression levels (Mortazavi et al. 2008). Then, the unigene expression levels between different libraries were compared using the edgeR package (Robinson et al. 2010) for differential expression analyses in the OmicShare tools, a free online platform for data analysis (http://www.omicshare.com/tools). The differentially expressed genes (DEGs) between the different gene expression libraries of four stages were screened restrictedly by *Q*-value ≤0.05 and the absolute value of log2 fold chang ≥1. As there was no replicate in this study, biological coefficient of variation which was the square-root of dispersions was set to 0.01 following the suggestion of edgeR official manual. The gene expression libraries of four stages were organized into six groups, VJ vs. FI-1, FI-1 vs. FI-2, FI-2 vs. FLD, VJ vs. FI-2, VJ vs. FLD and FI-1 vs. FLD, for pairwise comparisons to identify the DEGs during different stages.

Furthermore, the DEGs between any pair of stages were extracted and clustered by Short Time-series Expression Miner software (STEM) to analyze their expression profiles during the four stages (Ernst and Bar-Joseph 2006). The gene expression data of the DEGs in the four stages were normalized to 0, $\log_2$ (FI-1/VJ), $\log_2$ (FI-2/FI-1) and $\log_2$ (FLD/FI-2). The clustered profiles with *P*-value ≤0.05 were considered as significantly expressed. Then, the DEGs of significantly clustered profiles were subjected to GO and KEGG pathway enrichment analyses to identify significantly enriched GO terms and metabolic pathways compared with the whole genome background. The calculated *P*-value underwent a false discovery rate correction, taking the corrected *P*-value (*Q*-value) ≤0.1 as a threshold. GO terms and KEGG pathways fulfilling this condition were defined as significantly enriched. Furthermore, a bubble map was used to visualize the enriched KEGG results, which showed the significantly enriched KEGG pathways, the number of enriched unigenes, the enrichment factor and the significance.

Homologs related to flowering and real-time PCR validation

This RNA-seq analysis was performed to confirm the important flowering homologs in *L.*×*formolongi*. Several approaches were followed, including an initial screening of the KEGG pathways and GO terms related to flowering. Then, the coding sequences (CDSs) of the candidate unigenes involved in these KEGG pathways and GO terms were queried against the UniProtKB database using the BLAST algorithm (Villacorta-Martin et al. 2015). The unigenes with top hits were then identified as putative homologs based on the highest percentage of query coverage (≥70%) and the protein identity (≥60%), as well as the BLAST e-value (≤0.00001). Another approach, based on the functional annotations of all of the unigenes, selected those with annotation terms involved in flowering. Then, their CDSs were queried against the UniProtKB database using the BLAST algorithm, and the unigenes with top hits were selected and filtered as described previously.

Furthermore, the flowering homologs that showed differential expression were defined by a rigorous algorithm with an *Q*-value ≤0.05 and an absolute value of log2 fold change ≥1 as thresholds. Additionally, Cytoscape 3.2 software was used to generate and visualize co-expression clusters of the DEGs related to flowering using a correlation coefficient value of 0.8. Furthermore, protein sequences of important flowering homologs from *L.*×*formolongi, A. thaliana* and *Oryza sativa* were aligned using ClustalW in MEGA6 with default parameters. The phylogenic tree of MADS-box family genes was generated using the Bayesian method by MrBayes-3.2.6. Phylogenic analyses of other flowering homologous families used the Maximum Likelihood or Neighbor-Joining methods in MEGA6. The bootstrap analysis employed 1000 replicates, and the Poisson model was used for substitutions.

To validate the reliability of the RNA-seq analysis and the expression profiles of the important homologs related to flowering, 12 homologs were selected for qRT-PCR. The plant tissues of three individuals from the four stages were sampled at the same time with those used for RNA-seq.

Furthermore, in order to evaluate the biological variation of the gene expression levels in new material, three important homologs were chosen to evaluate their gene expression levels in three individuals as biological replicates different from the material used for RNA-seq. A new batch of seeds were sown, and the plant tissues of three biological replicates were sampled from eight development stages from seeds germination to the appearance of floral

buds under LD conditions. The total RNA of the leaves and stem apices from each individual were extracted separately and then mixed in equal amounts to synthesize the cDNAs for qRT-PCR tests. The RNA extraction and cDNA synthesis methods were described previously, as were the preparation and amplification program for qRT-PRC. *Aquaporin TIP4-1* was used as the internal control to standardize the results. All of the primers of the validated genes are listed in Table S1.

Results

Confirmation of the crucial stages during vegetative growth and flowering initiation

The phenotypic traits, anatomical SAM structures and the circadian rhythm expression patterns of *LfCOL9* and *LfFT1* in *L.*×*formolongi* cv. Raizan 2 seedlings were investigated from seed germination to visible floral bud appearance to define the crucial stages during vegetative growth and flowering initiation under LDs (Fig. 1a–c).

In seedlings with 2–3 to 9–10 rosette leaves, the morphology of the SAM exhibited a hemispheric vegetative shape (Fig. 1b). To ensure the accuracy of the VJ stage used for RNA-seq, the 2–3 rosette leaf stage was selected for a gene expression analysis. The mRNA accumulations of *LfCOL9* and *LfFT1* did not show diurnal rhythmic expression patterns for 3 days at that stage (Fig. 1c). Thus, the 2–3 rosette leaf stage was selected as the representative of the VJ stage for RNA-seq.

As the seedlings grew to have 9–10 rosette leaves at the 23rd week, the morphology of the SAM remained vegetative (Fig. 1b), while the expressions of *LfCOL9* and *LfFT1* started to display ambiguous circadian rhythms and low amplitudes (Fig. 1c). When the seedlings underwent bolting at the 24th week, they possessed 1–2 internodes and the morphology of the SAM was still hemispheric vegetative (Fig. 1b). However, the expressions of *LfCOL9* and *LfFT1* showed distinct and strong circadian rhythm patterns for 3 days with transcript levels peaking 4 h after dark. In addition, the diurnal expression of *LfCOL9* also had a transcript peak 12 h after illumination (in the light) (Fig. 1c). Meanwhile, the transcript accumulation of *LfFT1* was especially up-regulated compared with at the 9–10 rosette leaves stage (Fig. 1c). Therefore, the period from 9 to 10 rosette leaves to 1–2 internodes (just before and after bolting) was identified as flowering induction phase I (FI-1), and the tissues of the 1–2 internode stage (just bolting) was selected to represent FI-1 for RNA-seq.

When the seedlings had 5–6 internodes during the 25th week, paraffin sections showed that the morphology of the SAM appeared to be elongated and broad (Fig. 1b). This indicated that the SAM had undergone the morphological transition from vegetative to flowering, which was identified as flowering induction phase II (FI-2), and plant tissues of this stage were used for RNA-seq.

When the internode numbers reached 9–10 during the 26th week, the SAM enlarged and developed into inflorescence primordial (Fig. 1b). When the seedlings grew to 16–17 internodes, the SAM became concave and developed into small floral primordial (Fig. 1b). Thus, at the 9–10 internode stage, the seedlings entered into the FLD stage, the plant tissues of this stage were also used for RNA-seq.

Differential gene expression dynamics, GO and KEGG pathway enrichment analysis

To profile the global gene expression profile during the flowering initiation process of *L.*×*formolongi*, the four cDNA libraries of tissues from the VJ, FI-1, FI-2 and FLD stages were separately sequenced on the Illumina Hiseq 2000 platform. A total of 52,824 unigenes were generated from the four libraries, of which 37,032 were differentially expressed. The DEGs contained two expression patterns, up-regulated and down-regulated. Most of the DEGs were up-regulated in the FI-1 (14,713) and FLD (8310) stages, but down-regulated in the FI-2 stage (10,792) compared with the VJ stage (Fig. 2a).

To analyze the stage-specific expression profiles of the identified DEGs, the 37,032 DEGs were clustered by STEM according to their expression profiles. The result indicated that all of the DEGs could be clustered into 10 profiles, and just 4 profiles (colored blocks) contained 29,910 DEGs having significant expression patterns (P-value ≤ 0.05) (Fig. 2b). And profiles 1, 4 and 7 were the most significantly clustered.

The 29,910 DEGs of four significantly clustered profiles were subjected to GO and KEGG pathway enrichment analysis. The results of all the enriched GO terms are listed in Tables S2–S4. A bubble map indicates that 15 pathways were significantly enriched and the most enriched KEGG pathway was phenylalanine metabolism (Fig. 2c).

In addition, the DEGs of the significantly clustered profile 7 were subjected to a GO enrichment analysis. Some significantly enriched GO terms in the biological process category were related to reproductive transition and flower development (Table 1), of which reproduction (GO:0000003), DNA methylation (GO:0006306), floral organ morphogenesis (GO:0048444), flower development (GO:0009908) and reproductive shoot system development (GO:0090567) were the most significantly enriched GO terms. The results suggest that most of the DEGs related to flowering showed the expression pattern of profile 7, namely up-regulated in the FI-1 stage, then down-regulated in the FI-2 stage, and finally up-regulated again

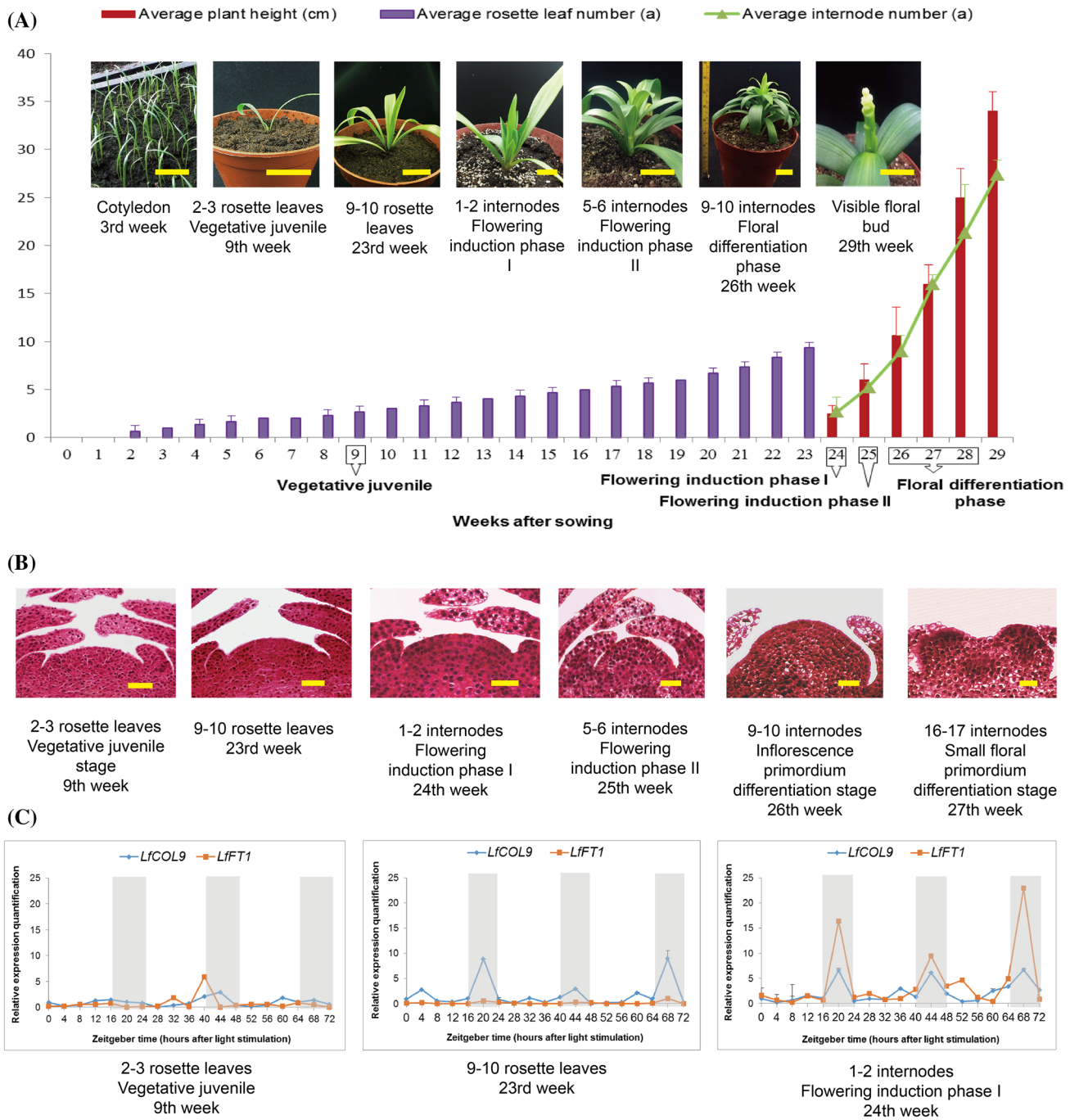


Fig. 1 Confirmation of different developmental stages during vegetative growth and flowering initiation process. **a** The morphological dynamic profile of *L.*×*formolongi* during different stages of vegetative growth and flower development. *Bar* 30 mm. **b** The anatomic structures of SAM using paraffin section during 2–3 rosette leaves (VJ), 9–10 rosette leaves, 1–2 internodes (FI-1), 5–6 internodes (FI-2) and 9–10 internodes (FLD). *Bar* 0.5 mm. **c** The diurnal circadian rhythm expression patterns of *LfCOL9* and *LfFT1* in leaves under LDs during the stages of 2–3 rosette leaves (VJ), 9–10 rosette leaves and 1–2 internodes (FI-1). Relative expression levels were determined by qRT-PCR (*y-axis*). The *x-axis* represents time point (hours). Data points represent an average of three biological replicates with three technical replicates. *Error bars* represent SE. The *shaded bars* over each chart represent dark periods

in the FLD stage. Thus, we hypothesized that FI-1 may be the key switching stage in which many important genes that respond during flowering were induced and up-regulated. The genes that have their highest transcript accumulations in the FLD stage may be related to floral meristem development.

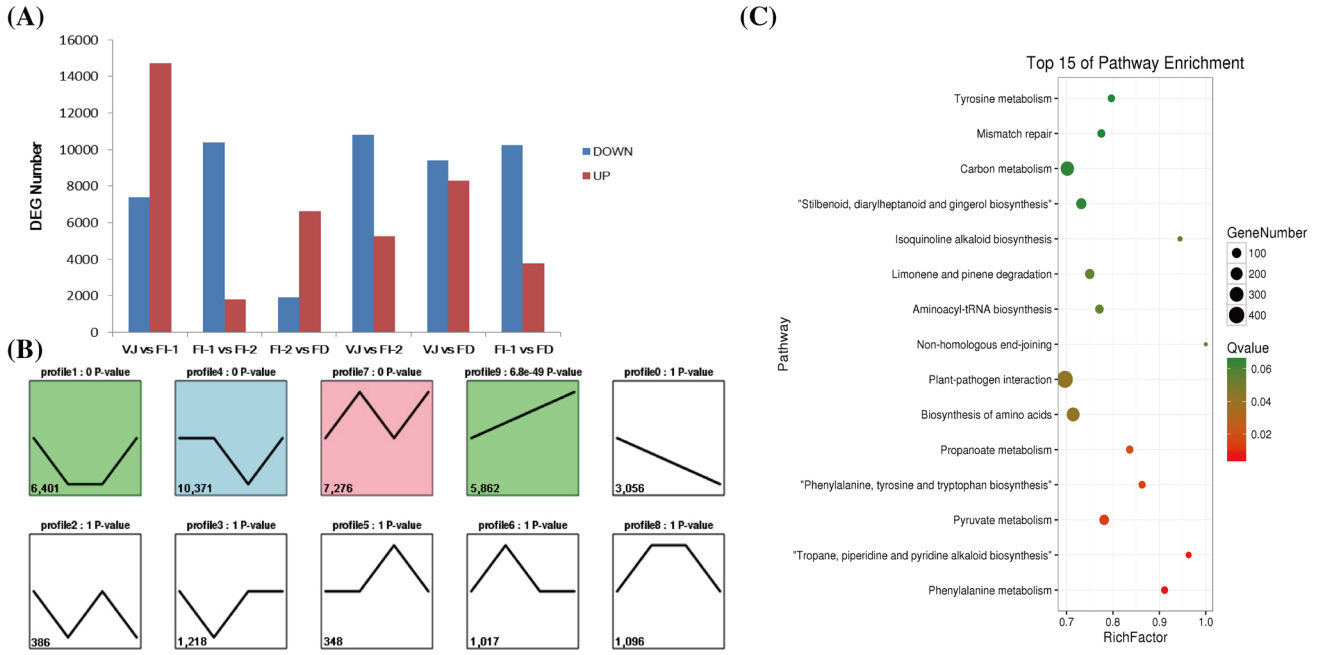


Fig. 2 Differential gene expression patterns and the KEGG pathway enrichment analysis results. **a** The number of the DEGs obtained from the pairwise comparison of four stages. The *red* columns represent the numbers of up-regulated DEGs in each pairwise comparison, and *blue* columns represent the down-regulated numbers. **b** Cluster analysis of the expression profiles of the DEGs. The profile ID and *P*-value are on the upper side of each profile block. The *number in the bottom left* indicates the number of genes assigned to the profile during clustering. Profiles blocks with colored background are significant in terms of the *P*-value. **c** A bubble map shows the significantly enriched results of KEGG pathways among the DEGs of significantly clustered profiles. The *y-axis* shows the top 15 enriched KEGG pathways. And the 15 pathways were sorted by the significance level from the bottom to the top on the *y-axis*. The *x-axis* shows the enrichment factor (the enriched gene number proportion of background gene number) of each enriched pathway. The *color of the round dots* represents the *Q*-value of each pathway and the size represents the number of enriched unigenes

Differential expression of putative homologs related to flowering in *L.*×*formolongi*

In total, 122 homologs implicated in flowering were identified from the database, and 85 homologs showed significantly differential expression during the four stages. These identified DEGs could be distributed into five genetic flowering pathways, photoperiod, age-dependent, trehalose-6-phosphate (T6P), gibberellin and ambient temperature (Table 2). The DEGs of the photoperiod pathway (41) were the most represented, followed by the age-dependent pathway (13) and the T6P pathway (9). Only one putative homolog each was identified in the gibberellin and the ambient pathways. Moreover, another 20 homologs were implicated in floral meristem identity. Table 2 shows the numbers and descriptions of important DEGs related to flowering in *L.*×*formolongi*, as well as the accession number and expression profile of the gene with the highest homology. The other putative flowering homologs are described in Table S5.

A co-expression network was used to depict the differential expression profiles of all of the flowering homologs during flowering initiation (Fig. 3). The homologs could be clustered into three expression types in which closely related genes displayed conserved expression profiles. The predominant type I cluster included genes that showed the highest expression in the FI-1 stage, like the circadian clock homologs *CRY1* and *PHYTOCHROME A* (*PHYA*), several *COL* and *FT* homologs from the photoperiod pathway, as well as several *TPS* homologs from the T6P pathway (Fig. 3). Cluster type II included the genes that exhibited the highest transcription levels in the VJ stage, such as the rest of the circadian clock and *COL* homologs from the photoperiod pathway, and the other *TPS* family members from the T6P pathway (Fig. 3). Cluster type III consisted of genes up-regulated in the FLD stage, like the floral meristem identity homolog *APETALA1, LEAFY*, and several *AGAMOUS-LIKE* (*AGL*) and *SPL* homologs (Fig. 3).

Differential expression of the TF genes related to flowering

TFs are the key proteins that mediate transcriptional regulation and also play crucial roles as molecular switches for gene expression in flowering transition. According to the previous study in *A. thaliana*, most of the important

Table 1 The significantly enriched GO terms related to reproductive transition and flower development in biological process of profile 7

GO ID	Description	Number of enriched DEGs classified in biological process/(% of 2621)	Number of DEGs classified in biological process/(% of 17,080)	*Q*-value
GO:0000003	Reproduction	153 (5.84%)	605 (3.54%)	7.25E−11
GO:0006306	DNA methylation	56 (2.14%)	167 (0.98%)	3.62E−09
GO:0016570	Histone modification	104 (3.97%)	399 (2.34%)	1.56E−08
GO:0048449	Floral organ formation	42 (1.60%)	131 (0.77%)	1.22E−06
GO:0048444	Floral organ morphogenesis	44 (1.68%)	145 (0.85%)	3.61E−06
GO:0009908	Flower development	128 (4.88%)	587 (3.44%)	1.52E−05
GO:0090567	Reproductive shoot system development	132 (5.04%)	612 (3.58%)	1.96E−05
GO:0044702	Single organism reproductive process	286 (10.91%)	1517 (8.88%)	6.13E−05
GO:2000241	Regulation of reproductive process	89 (3.40%)	393 (2.30%)	7.25E−05
GO:0048437	Floral organ development	93 (3.55%)	426 (2.49%)	0.000204
GO:0048569	Post-embryonic organ development	113 (4.31%)	538 (3.15)	0.000232
GO:0019953	Sexual reproduction	46 (1.76%)	187 (1.09%)	0.000614
GO:0009909	Regulation of flower development	76 (2.90%)	346 (2.03%)	0.000617
GO:0022414	Reproductive process	343 (13.09%)	1921 (11.25%)	0.000804
GO:0048608	Reproductive structure development	275 (10.49%)	1533 (8.98%)	0.002074
GO:0061458	Reproductive system development	275 (10.49%)	1533 (8.98%)	0.002074
GO:0003006	Developmental process involved in reproduction	321 (12.25%)	1833 (10.73%)	0.003981
GO:0010050	Vegetative phase change	21 (0.80%)	80 (0.47%)	0.008032
GO:0048442	Sepal development	10 (0.38%)	35 (0.20%)	0.033452
GO:0048464	Flower calyx development	10 (0.38%)	35 (0.20%)	0.033452

regulatory genes implicated in flowering transition encode TFs (Pařenicová et al. 2003; Singh et al. 2013). Therefore, we searched for TF genes related to flowering in *L.*×*formolongi* and studied their expression dynamics. Among the identified putative homologs, 52 genes encoded for TFs and showed differential expression, including the members of the bHLH, bZIP, RAV, MYB, Dof, CO-like, AP2/EREBP, SBP, GRAS, FLO_LFY and MADS-box TF families. The co-expression network also depicted the differential expression profiles of the TF genes during the four stages (Fig. 3).

In the network, the MADS-box TF family genes (orange circles) were the most represented among the various families, with 20 genes showing differential expression. They were mainly distributed in clusters I and III, representing up-regulated profiles in the FI-1 and FLD stages, respectively (Fig. 3). The SBP TF family genes (*SPLs*, yellow circles) (11) were the second most represented and were also mainly distributed in clusters I and III (Fig. 3). They were followed by CO-like TF family members (blue circles, eight), which were distributed in clusters I and II (Fig. 3), and exhibited their highest transcription levels in the FI-1 and VJ stages. The other TF family genes (pink circles) were mainly distributed in cluster I and were namely up-regulated in the FI-1 stage (Fig. 3).

These findings suggested that the FI-1 stage was the key stage of flowering transition and that most TF genes were up-regulated in this stage. Remarkably, the up-regulated MADS-box, SBP and CO-like TF genes in the FI-1 stage may be the critical trigger factors for flowering induction and transition.

In addition, we analyzed the differential expression profiles of the identified MADS-box homologs, including *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*), *AGL, AP1, SHORT VEGETATIVE STAGE* (*SVP*) (Fig. 4b). An *SVP* homolog and four *SOC1* homologs that showed the highest transcription levels in the FI-1 stage may be involved in the flowering transition. Two *AGL* homologs, *LfAGL44* and *LfAGL19*, and another three *SOC1* homologs, *LfSOC1.1, LfSOC1.2* and *LfSOC1.3*, were not only up-regulated in the FI-1 stage, but also showed high expression levels in the FLD stage, which means these genes may be related to the flowering transition process, but also participated in the floral development. Furthermore, six *AGL* homologs, *LfAGL1, LfAGL2, LfAGL3, LfAGL4, LfAGL24, LfAGL15* and three *AP1* homologs, *LfAP1.1, LfAP1.2* and *LfAP1.3*, were all up-regulated in the FLD stage and may function as floral meristem identity genes (Fig. 4b).

Table 2 The putative homologs related to flowering and their expression profiles in mixed sample of leaves and stem apexes during four stages in *L.*×*formolongi*. Gene expression during different stage is schematically represented based on the log-fold change from VJ to FLD stage

Flowering pathway and floral meristem identity genes	Putative flowering orthologs	Homologous species/Protein Identity (%)	Number of the gene members in the same family	Accession number	Flowering Enhancer/ Repressor	Expression profile of the flowering ortholog during four stages
Photoperiod	*LfCRY1*	*O. sativa* / 72	1	KX162660	Enhancer	
	LfPHYA1	*O. sativa* / 70	3	KX162661	Enhancer	
	LfGI	*O. sativa* / 72	1	KX162662	Enhancer	
	LfCOL9	*L.* × *formolongi. cv. Raizan 3* / 97	8	KJ744206	Enhancer	
	LfFT1	*L.* × *formolongi. cv. Raizan 3* / 99	4	KJ744207	Enhancer	
	LfFKF1	*A. thaliana* / 78	1	KX162665	Enhancer	
	LfCDF3.1	*Elaeis guineensis* / 60	3	KX162667	Repressor	
	LfCOP1	*O. sativa* / 74	1	KX162668	Repressor	
T6P	*LfTPS1*	*A. thaliana* / 76	9	KX162669	Enhancer	
Age-dependent	*LfAP2.1*	*A. thaliana* / 67	2	KX162670	Repressor	
	LfSPL3	*Elaeis guineensis* / 65	12	KX162671	Enhancer	
Gibberellin	*LfGAI*	*O. sativa* / 66	1	KX162672	Enhancer	
Ambient	*LfSVP*	*A. thaliana* / 63	1	KX162673	Repressor	
floral meristem identity genes	*LfLFY*	*L. longiflorum* / 99	1	KX162674	Enhancer	
	LfSOC1.1	*Elaeis guineensis* / 62	8	KX162675	Enhancer	
	LfAGL24	*L. longiflorum* / 100	4	KX162676	Enhancer	
	LfAP1.1	*L. longiflorum* / 99	3	KX162677	Enhancer	

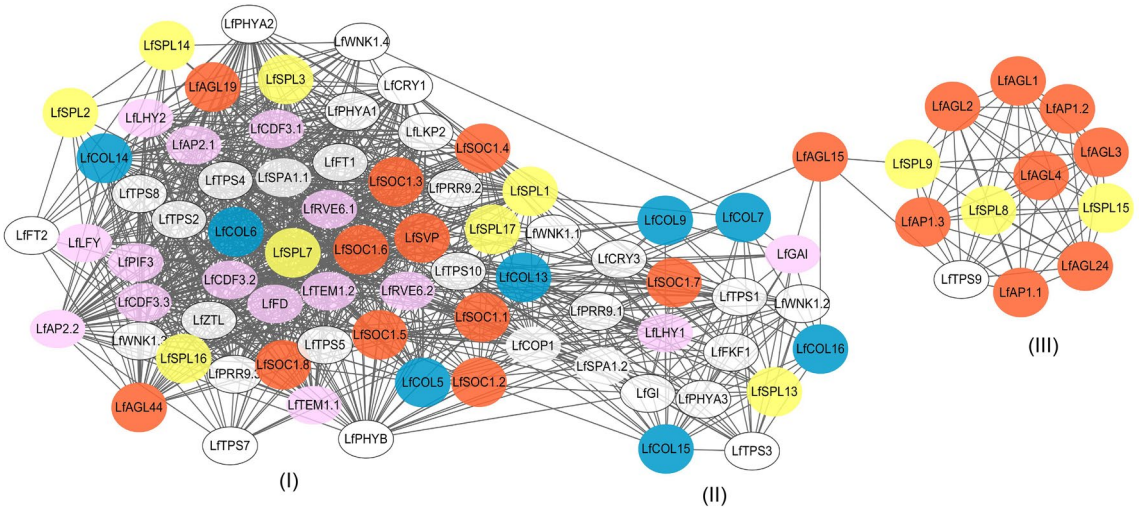


Fig. 3 The co-expression network map depicting the differential expression profiles of the flowering homologs and TF genes during flowering initiation. Cluster type I represents the up-regulated homologs in the FI-1 stage, cluster type II represents the homologs showed the highest expression in the VJ stage and cluster type III represents the up-regulated homologs in the FLD stage. The *orange circles* represent the members of MADS-box TF family, *yellow circles* represent the genes of SBP TF family, and the *blue circles* represent the genes of CO-like TF family. Remaining *pink* and *blank circles* represent the other TF family members and general homologs respectively

Differential expression of flowering homologs in the photoperiod pathway

A total of 41 DEGs related to circadian rhythm and photoperiod-mediate flowering were identified, including the circadian rhythm genes *CRY1, PHYA, TIMING OF CAB EXPRESSION 1* (*TOC1*), *FLAVIN-BINDING KELCH REPEAT F-BOX 1* (*FKF1*), the rhythm output genes *GI, CO* and *FT*, and the downstream bZIP transcription factor *FD* (Table 2). A heatmap showed the differential expression profiles of all 41 homologs (Fig. 5c). Almost all of the circadian clock homologs exhibited high transcript accumulations in the VJ or FI-1 stages, including *LfGI*, the homolog of *GI*, which had its highest expression in the VJ stage but also maintained a high expression level in the FI-1 stage (Fig. 5c).

In *A. thaliana, CO* is a member of *COL* gene family that contains 16 other members (Griffiths et al. 2003). In this study, eight *COL* homologs members were identified. Based on the phylogenetic relationships and conserved protein domains of the *COL* family members between *L.*×*formolongi, A. thaliana* and *O. sativa*, the eight *COL* homologs were divided into three types (Fig. 5a). The first type contained two B-box motifs and only included *LfCOL5*. The second type exhibited one B-box motif, and consisted of *LfCOL6, LfCOL7, LfCOL13, LfCOL14, LfCOL15* and *LfCOL16*. The last type contained a normal B-box motif and a second divergent B-box motif, and only included *LfCOL9* (Fig. 5a). Thus, the *LfCOLs* homologs with one B-box motif were predominant. Additionally, the *LfCOL*s were phylogenetically close related to the *O. sativa COLs*, most of which also contained only one B-box motif (Fig. 5a).

Moreover, among the eight *LfCOL*s, five homologs, including *LfCOL7* and *LfCOL9*, had their highest expression levels during the VJ stage (Fig. 5c). The other three homologs, *LfCOL5, LfCOL6* and *LfCOL14*, were up-regulated in the FI-1 stage (Fig. 5c). In addition, three *CYCLIC DOF FACTOR 3* (*CDF3*) homologs, a *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*) homolog and two *SUPPRESSOR OF PHYA-105 1* (*SPA1*) homologs, which as the repressors of *COL* or the COL protein, exhibited high expression levels in the FI-1 stage (Fig. 5c).

FT is a key flowering activator, and *protein TWIN SISTER of FT-like* (*TSF*) is the most homologous member of the same gene family as *FT* in *A. thaliana*. Moreover, *Heading date 3a* (*Hd3a*) is the rice (*O. sativa*) homolog of the *A. thaliana FT* (Kojima et al. 2002). Two *L.*×*formolongi* putative homologs of *FT-like* were identified in this study. The phylogenetic relationship of the *FT* and *TSF* homologs among *L.*×*formolongi, A. thaliana* and *O. sativa* were analyzed using a maximum likelihood tree (Fig. 5b). The results indicated that *LfFT1* was most closely related to *Hd3a* and that *LfFT2* was closely related to *OsTSF* (Fig. 5b). *LfFT1*was up-regulated in the FI-1 stage, and *LfFT2* was highly expressed in both the FI-1 and FLD stages (Fig. 5c).

In addition, two *TEMPRANILLO 1* (*TEM1*) homologs, *LfTEM1.1* and *LfTEM1.2*, which are repressors of *FT* and encode members of the RAV transcription factor

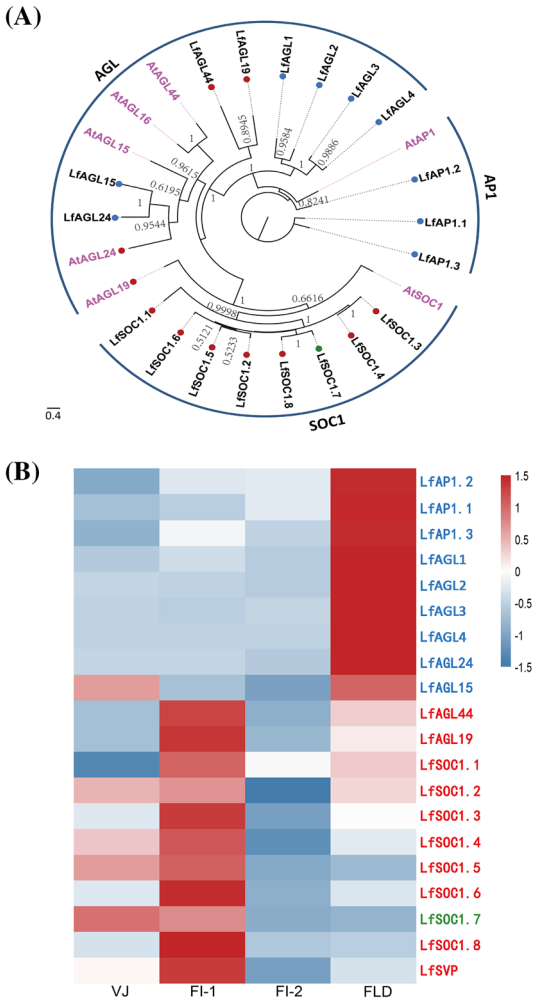


Fig. 4 The phylogenetic relationship and differential expression profiles of the putative MADS-box homologs. **a** Bayesian tree of the identified MADS-box TF genes of *L.*×*formolongi* and *A. thaliana*. The gene names colored in *pink* were *A. thaliana* MADS-box TF genes and the *black* were homologs from *L.*×*formolongi*. The prefixed *green dots* represent the highest expression in the VJ stage, *red dots* represent the up-regulated expression in the FI-1 stage, and *blue dots* represent up-regulated expression in the FLD stage. **b** A heatmap exhibiting the differential expression profiles of all the MADS-box homologs. Gene names are colored according to their expression patterns. *Green color* represents the highest expression in the VJ stage, *red* represents the up-regulated expression in the FI-1 stage, and *blue* represents the up-regulated expression in the FLD stage

family in *A. thaliana*, were up-regulated in the FI-1 stage (Fig. 5c). The *FD* homolog, *LfFD*, showed the highest expression in the FI-1 stage (Fig. 5c). *FD* is an important activator in the triggering of flowering at the stem apex (Wellmer et al. 2006).

Differential expression of the age-dependent and T6P pathway homologs

A set of 13 DEGs belonging to the age-dependent pathway were identified, including 11 SBP TF family member *SPL*s, whose homologs in *A. thaliana* act as the target genes of miR156, and 2 AP2/EREBP TF family member *AP2*s, whose homologs act as the target genes of miR172. In addition, nine *TPS* homologs belonging to the T6P signal pathway were identified.

Among the 11 *SPL* homologs, seven homologs were up-regulated in the FI-1 stage, such as *LfSPL3*, three homologs were up-regulated in the FLD stage, such as *LfSPL9*. Only *LfSPL13* not only up-regulated in the FI-2 stage, but also showed a high expression level in the FLD stage (Fig. 6c). Additionally, among the seven up-regulated *LfSPL*s in the FI-1 stage, five, such as *LfSPL16* and *LfSPL17*, also showed high expressions in the FLD stage (Fig. 6c). The two *AP2* homologs were all up-regulated in the FI-1 stage (Fig. 6c).

Of the nine *TPS*s homologs, *LfTPS1* was clustered with *AtTPS1* and *OsTPS1* in the phylogenetic tree and exhibited the highest transcription level in the VJ stage (Fig. 6a, c). The other *LfTPS*s either showed their highest expression levels in the VJ stage, or were up-regulated in the FI-2 stage (Fig. 6c). However, only *LfTPS9* showed its highest expression level in the FLD stage (Fig. 6c).

Verification of the gene expression profiles by qRT-PCR

To validate the accuracy and reproducibility of the expression profiles of genes in the RNA-seq analysis, 12 important DEGs related to flowering transition and floral meristem identity were selected for qRT-PCR. The candidate DEGs included *LfCOL*s, *LfFT1, LfSPL*s and *LfTPS*s, which acted as the key regulators in the photoperiod, age and T6P pathways. The linear regression analysis of the fold change of the gene expression between qRT-PCR and RNA-seq showed significantly positive correlation (correlation coefficient of 0.72, *p*-value was 1.16E−10) (Fig. 7a). And all of the homologs showed expression patterns in qRT-PCR analysis that were consistent with the RNA-seq data, confirming the reliability of our RNA-seq analysis (Fig. 7b).

Furthermore, three important flowering homologs of *LfCOL5, LfCOL9* and *LfFT1* were chosen to evaluate their expression levels in three individuals during eight developmental stages in of *L.*×*formolongi* under LD conditions. The materials were different from the RNA-seq, and sampled from a new batch of seedlings from eight development stages, including 2–3 rosette leaves (VJ), 5–6 rosette leaves, 9–10 rosette leaves, 1–2 internodes (FI-1), 5–6 internodes (FI-2), 7–8 internodes, 9–10 internodes (FLD) and 20–21 internodes (Fig. 7c). The three replicates of each homolog all showed consistent expression patterns during the eight

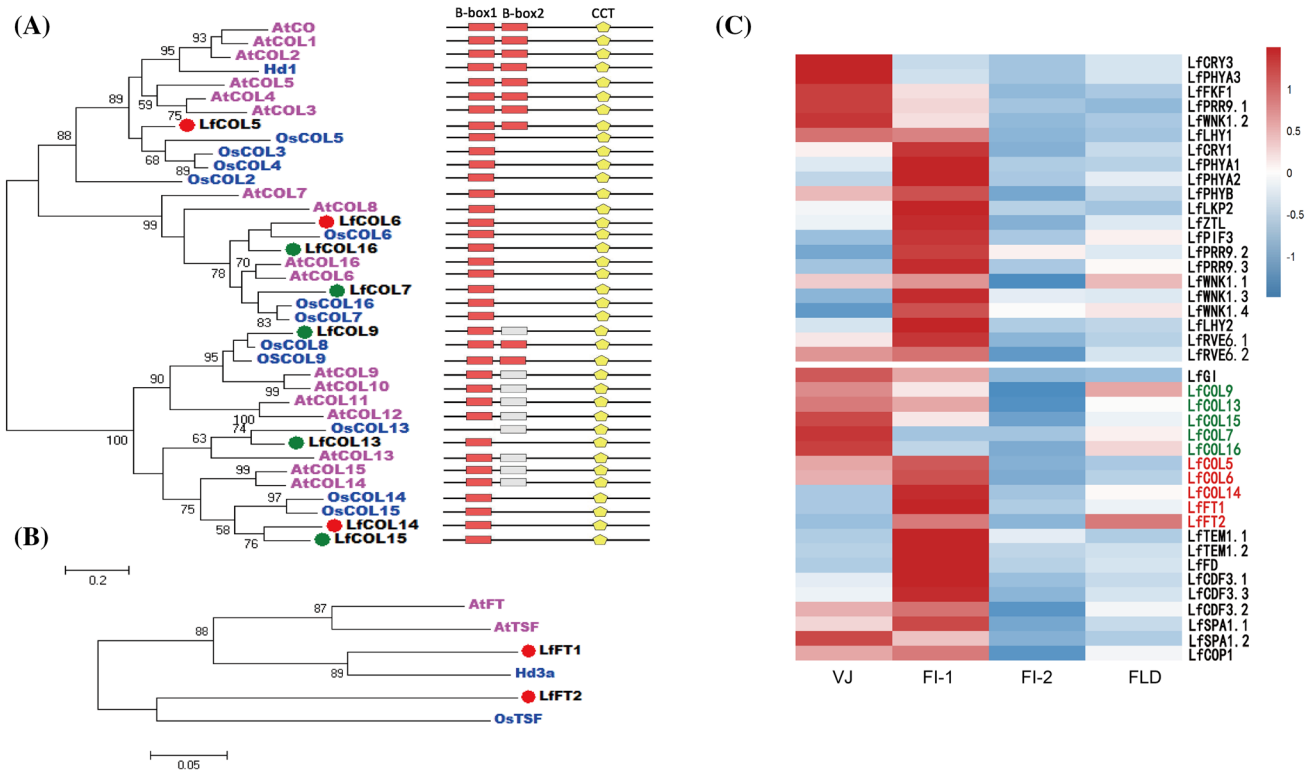


Fig. 5 The phylogenetic relationship and differential expression profiles of the putative photoperiod homologs. **a, b** Maximum likelihood trees of the identified *COL* and *FT-like* homologs of *L.*×*formolongi, A. thaliana* and *O. sativa*. Bootstrap values from 1000 replicates are used to assess the robustness of the tree. The *scale* indicates the average number of substitutions per site. The gene names colored in *pink* are *A. thaliana* genes, *blue* are *O. sativa* genes and the *black* are the homologs from *L.*×*formolongi*. The prefixed *green dots* represent the highest expression in the VJ stage, *red dots* represent the up-regulated expression in the FI-1 stage, and *blue dots* represent up-regulated expression in the FLD stage. The structure diagrams in the *right side* show the domain structures of B-box motif (*red rectangles*), second divergent B-box motif (*gray rectangles*), CCT motif (*yellow pentagons*) of the *COL* genes from *L.*×*formolongi, A. thaliana* and *O. sativa*. **c** A heatmap exhibiting the differential expression profiles of all the photoperiodic homologs. Gene names of *LfCOL*s and *LfFT*s are colored according to their expression patterns during the four stages refer to the description in Fig. 4

stages. The three genes expression patterns of the three individuals in the VJ, FI-1, FI-2 and FLD stages were consistent with the RNA-seq data of the same stages (Fig. 7c). Thus, the biological variation of the gene expression levels in the RNA-seq is reliable to some extent.

Discussion

Confirmation of different stages during flowering initiation

Lilium×*formolongi* is hallmarked of its ability to flower within a year after sowing. The confirmation of the flowering induction stage is very important for understanding the genetic mechanism of the prompt transition from vegetative growth to flowering in *L.*×*formolongi*. However, the flowering induction stage is hard to identify because the SAM is still morphologically vegetative, while the internal physiological mechanism and molecular dynamics have changed dramatically and switched to the reproductive stage (Meyerowitz et al. 1991). According to studies in the typical LD plant *A. thaliana*, the circadian rhythm expression patterns of *CO* and *FT* are crucial in triggering the flowering induction and transition (Andrés and Coupland 2012; Suárez-López et al. 2001). Our study suggests that the circadian rhythm expression patterns of *LfCOL9* and *LfFT1* in *L.*×*formolongi* could be applied as biomarkers of the flowering induction stage under LD conditions.

The circadian rhythm expression patterns of *LfCOL9* in the FI-1 stage almost conformed to the diurnal expression pattern of *CO* in *A. thaliana* (Andrés and Coupland 2012). Nevertheless, the expression pattern of *LfFT1*, with a transcript abundance peak at night, was not in agreement with the discovery in *A. thaliana that* showed the transcript abundance peak of *FT* occurred at dusk (Andrés and Coupland 2012). This may be due to the different origins, evolutionary histories and habitats of the different species. The analyses of *CO* and *FT* expressions in aspen trees had already detected variations in their relative peaks during

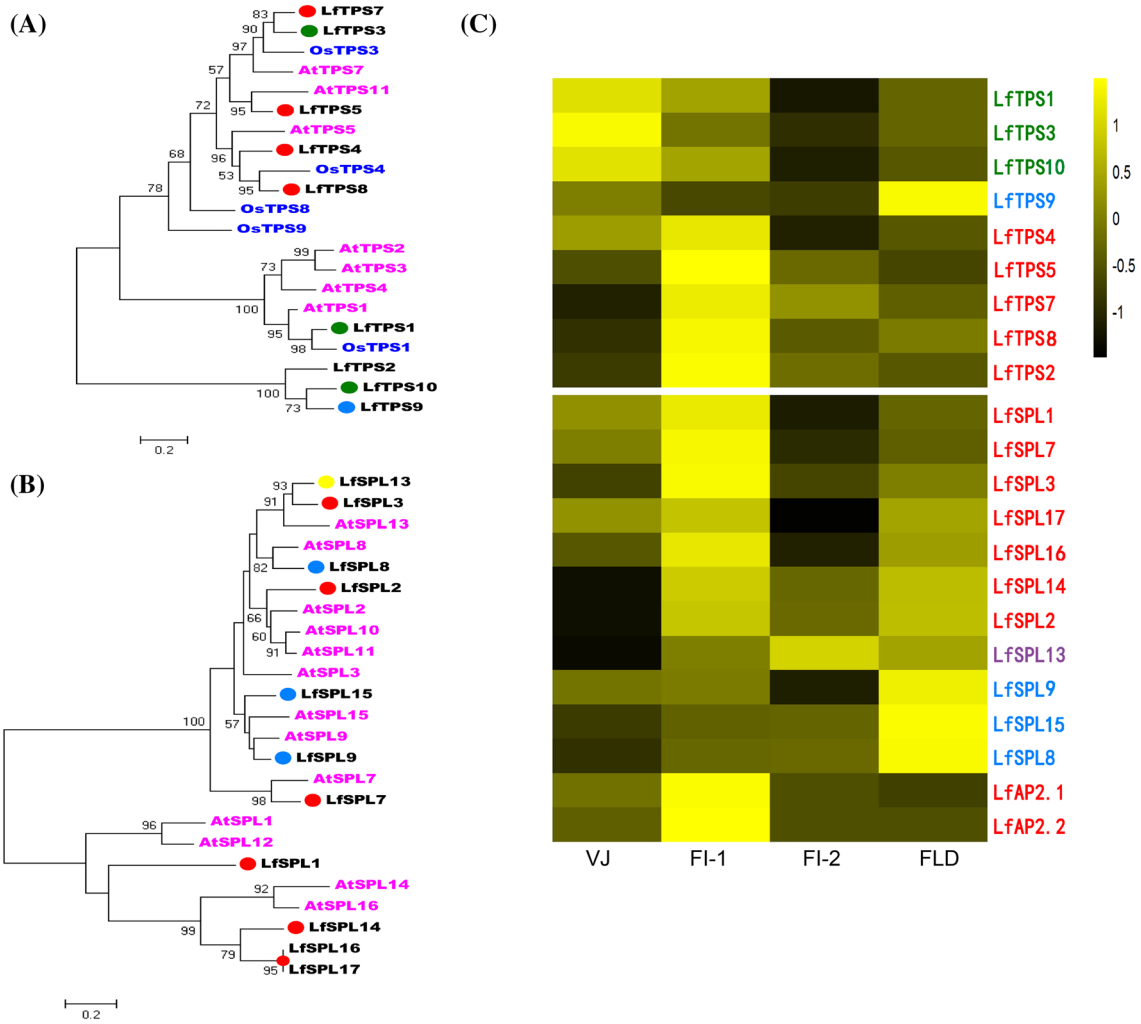
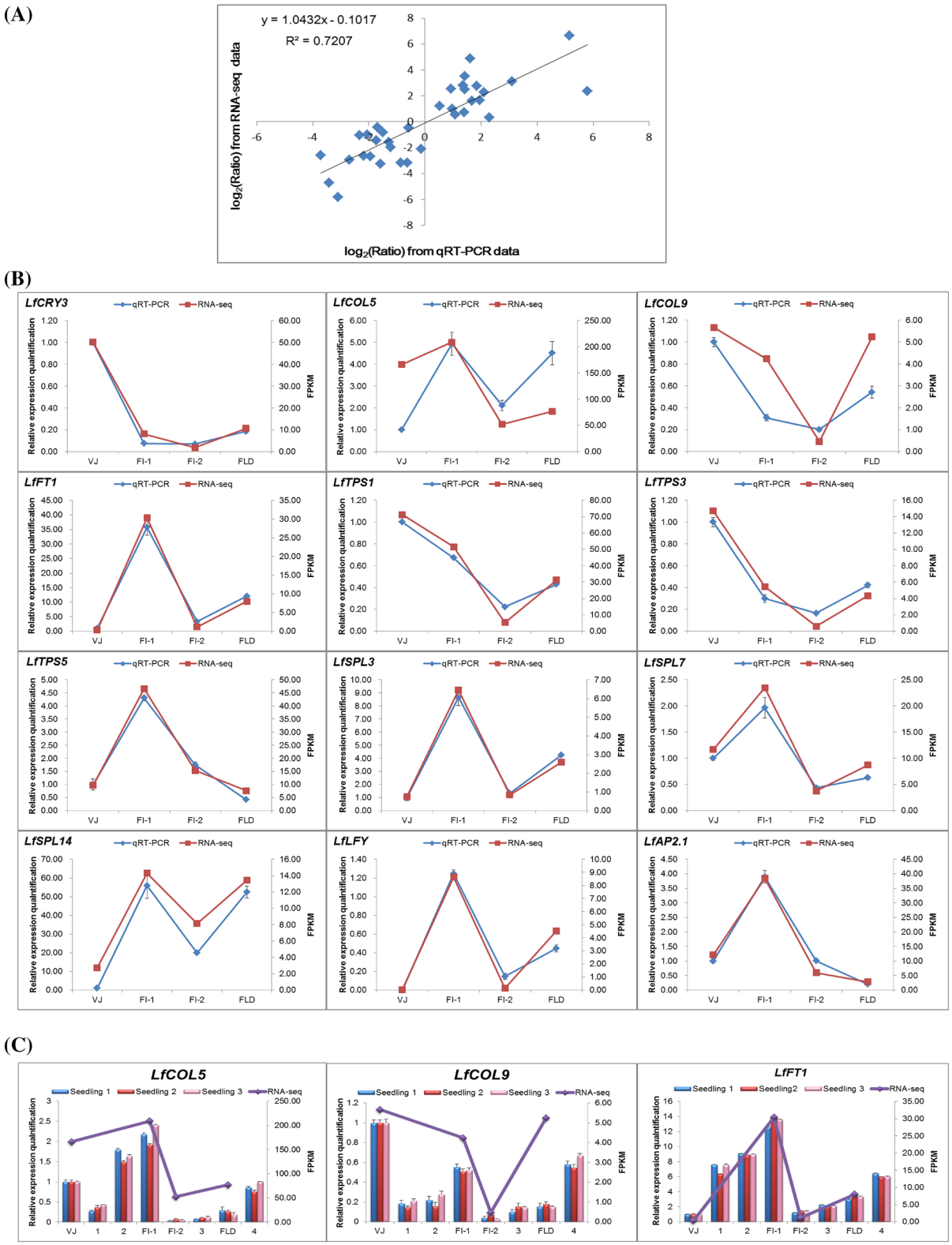


Fig. 6 The phylogenetic relationship and differential expression profiles of the putative *SPL, AP2* and *TPS* homologs. **a** Maximum likelihood tree of the identified *TPS* homologs between *L.*×*formolongi, A. thaliana* and *O. sativa*. The *color* of the gene names and the prefixed dots refer to the description in the previous Fig. 5. **b** Neighbor-joining tree of the identified *SPL* homologs between *L.*×*formolongi* and *A. thaliana*. The gene names colored in *pink* are *A. thaliana* genes, and the black are the homologs from *L.*×*formolongi*. The prefixed *red dots* represent the up-regulated expression in the FI-1 stage, *yellow dot* represents up-regulated expression in the FI-2 stage and *blue dots* represent up-regulated expression in the FLD stage. **c** A heatmap exhibiting the differential expression profiles of the identified *LfTPS, LfSPL* and *LfAP2* homologs. Gene names are colored according to their expression patterns. *Green color* represents the highest expression in the VJ stage, *red* represents the up-regulated expression in the FI-1 stage, *purple* represents the up-regulated expression in the FI-2 stage and *blue* represents the up-regulated expression in the FLD stage

the day–night cycle, such that southern trees showed an earlier peak than northern trees (Böhlenius et al. 2006).

In the flowering time control study of *A. thaliana*, the timing of flowering was carefully measured by counting the number of leaves formed before the initiation of flowering (Andrés and Coupland 2012). In this study, based on the investigation of the internal gene expression patterns and external anatomical SAM structures during different growth stages of the *L.*×*formolongi* seedlings, the corresponding phenotypic traits, including the average number of rosette leaves and internodes, could be used as the favorable morphological markers to judge the VJ, FI and FLD stages. The rosette stage, from 2–3 to 7–8 leaves, represented the vegetative growth stage of the seedlings. The stage just before and after seedling bolting was identified as FI-1. When the seedlings grew up to 5–6 internodes, the stage was identified as the morphological transition stage from vegetative to reproductive growth, namely FI-2. The FLD was identified using an average internode number of 9–10. Confirmation of the convenient morphological markers of the different growth and developmental stages would be helpful for the cultivation and flowering regulation analysis of *L.*×*formolongi*.



The differential gene expression profile in the FI-2 stage

In our RNA-seq study, 37,032 DEGs were generated from the four developmental stages of *L.*×*formolongi* seedlings under LD conditions. The global expression profiles of most DEGs were most predominant in the FI-1 and FLD stages, which demonstrated that FI-1 was the key stage involved in flowering induction and that

◂**Fig. 7** Coefficient analysis of fold change data between qRT-PCR and RNA-seq. **a** Correlation of gene expression results obtained from qRT-PCR and RNA-seq of 12 selected genes during four stages. Scatterplots were generated by the $\log_2$ expression ratios from RNA-seq (*y-axis*) and qRT-PCR (*x-axis*). **b** Expression profiles of 12 putative flowering homologs revealed by qRT-PCR (*left y-axis*) and RNA-seq (*right y-axis*) during the four stages. Data from qRT-PCR were means of three biological replicates with three technical replicates and bars represent SE. **c** Expression profiles of three putative flowering homologs in three individuals revealed by qRT-PCR (*left y-axis*) during eight stages. The notes of *x-axis* represent various stages *VJ* 2–3 rosette leaves; *1* 5–6 rosette leaves; *2* 9–10 rosette leaves; *FI-1* 1–2 internodes; *FI-2* 5–6 internodes; *3* 7–8 internodes; *FLD* 9–10 internodes; *4* 20–21 internodes. Data from qRT-PCR were means of three technical replicates and *bars* represent SE

the FLD stage involved the development of the floral meristem.

Nevertheless, the RNA-seq analysis revealed that most DEGs were down-regulated in the FI-2 stage in which the SAM morphology reached the early reproductive transition status. In *A. thaliana*, a global gene expression analysis during early flower development revealed that an important step for the onset of flower formation was characterized by a massive down-regulation of genes in incipient floral primordia, followed by a predominance of gene activation during the differentiation of floral organs (Wellmer et al. 2006). The results were in agreement with our transcriptome's DEG dynamics during the flowering initiation process. The DEGs were down-regulated in the FI-2 stage and exhibited higher transcription again in the FLD stage.

In addition, some important flowering activators identified in our study showed low expression levels in the FI-2 stage, and even in the FLD stage. Recent studies in *A. thaliana* showed that *AP1* acts predominantly as a transcriptional repressor during the initial stage of flower development. In particular, *AP1* down-regulates genes that participate in flowering transition, including some of its own activators, such as *SOC1, AGL24, SPL9, LFY* and *FD* (Wellmer and Riechmann 2010). This may also explain why *LfLFY* and several *LfSOC1*s genes showed lower transcript accumulations in the FI-2 and FLD stages. Thus, FI-2 is the important transition stage for floral development, even though most DEGs were down-regulated in this stage. Moreover, some flowering activators may be repressed by *LfAP1*s, therefore, showing low expression levels in the FI-2 and FLD stages.

Crucial flowering pathways and regulatory genes identified in *L.*×*formolongi*

In our RNA-seq study of the flowering initiation process of *L.*×*formolongi* seedlings, 85 DEGs related to flowering were identified. This was the first discovery of most of them in lily, and they were distributed among the photoperiod, age-dependent and T6P pathways. Remarkably, some homologs may be the key activators and TF genes reported by preceding studies in *A. thaliana*, such as members of the *LfCOL* and *LfFT* families from photoperiod, and the *LfSPL* and *LfTPS* families from the age-dependent and T6P pathways, respectively.

In *A. thaliana*, the members of the *COL* family contain one or two zinc finger B-boxes at the amino terminus and a CCT domain at the carboxy terminus, which could be divided into three types by the conservation of the B-box domains (Robson et al. 2001). The first type has two B-box motifs and includes *CO* and *COL1–COL5*. The second type contains only one B-box motif, as in *COL6–COL8* and *COL16*. The last type shows one B-box motif and a second divergent B-box motif, and consists of *COL9–COL15* (Griffiths et al. 2003). Recent studies showed that not all of the *COL* gene members could regulate the flowering time. *COL1* and *COL2* were predicted to encode zinc finger proteins with almost 67% amino acid identities to the protein encoded by *CO*. However, unlike *CO*, the altered expression of *COL1* or *COL2* in transgenic plants has little effect on flowering time (Ledger et al. 2001). Moreover, *COL3* is a positive regulator of red light signaling and root growth, and the over-expression of *COL5* can induce flowering in SD grown *A. thaliana* (Datta et al. 2006; Hassidim et al. 2009).

Based on the rice genomic sequence, 16 *OsCOL* genes have been identified (Griffiths et al. 2003). However, we only discovered 13 *OsCOL*s in the NCBI and plant transcription factor database (http://planttfdb.cbi.pku.edu.cn/family.php?sp=Osj&fam=CO-like) that showed integrated protein domains. The functions of these *OsCOL*s were different. *Hd1* is the homolog of *A. thaliana CO* and a major determinant of photoperiod sensitivity in *O. sativa*. A report indicated that *OsCOL3* negatively controlled flowering time under SDs by repressing the expression of *Hd3a*, independent of the SD-promotion pathway (Kim et al. 2008).

Lily, like *O. sativa*, is a monocotyledon. In our study, the eight identified *LfCOL* homologs could also be divided into three types by the conservation of the B-box domains, as in *A. thaliana*. However, most of the *LfCOL*s were classified into the second type that contained only one B-box motif, similar to *O. sativa*. Furthermore, the expression profiles of the eight *LfCOL*s were different. Several homologs showed their highest expression levels in the VJ stage, while others were up-regulated in the FI-1 stage or showed high expression levels in both the VJ and FI-1 stages. *LfCOL9*, which had been cloned in our previous study from *L.*×*formolongi* and showed a circadian rhythm expression pattern in the FI-1 stage, exhibited high expression levels in both the VJ and FI-1 stages. Thus, *LfCOL*s may great influence the induction of flowering, but the functions of various *LfCOL*

family members may be different and complex. Determining which putative *LfCOL* homolog is the most homologous to *A. thaliana CO* or *O. sativa Hd1*, and as the critical flowering activators that promote flowering under LDs, will require more intensive studies in the future.

FT is a pivotal integrator in *A. thaliana* because almost all of the flowering regulatory pathways converge on this gene, and *FT* transmits the floral inductive signal to downstream floral identity genes, including *LFY* and *AP1* (Mouradov et al. 2002). The *FT/TERMINAL FLOWER 1* (*TFL1*) family is a small gene family whose members, including *FT, TFL1, TSF, protein CENTRORADIALIS-like* (*ATC*), *MOTHER OF FT AND TFL1* (*MFT*) and *BROTHER OF FT AND TFL1* (*BFT*) (Yoo et al. 2010). Among these family members, *TSF* is most homologous to *FT*, and it shows a similar diurnal oscillatory expression pattern and produces an early-flowering phenotype when overexpressed (Michaels et al. 2005; Yamaguchi et al. 2005).

Lilium.×formolongi is a LD monocot plant, and *LfFT1* and *LfFT2* were both up-regulated in the FI-1 stage. *LfFT1* was phylogenetically closely related to *O. sativa Hd3a*, while *LfFT2* was closely related to *O. sativa TSF*. Thus, determining which *LfFT* homolog predominantly functions like the *O. sativa Hd3a* or *A. thaliana FT* to trigger flowering under LDs will require more studies in the future.

Putative gene networks controlling flowering induction and development in *L.×formolongi*

Studies on the genetic mechanism of flowering time in the model plant *A. thaliana* have shown that the responses to various external and internal conditions are integrated by a complex gene regulatory network, which includes several regulatory pathways that control flowering transition (Wellmer and Riechmann 2010). The genetic basis of flowering induction and floral development in the model plant *A. thaliana* appears to be similar to those in other plants, but species differences among the different regulatory pathways still exist (Khan et al. 2014). A hypothetical schematic diagram was created to depict the genetic and molecular interactions involved in photoperiod, age and T6P responses in *L.×formolongi* based on the annotation of the flowering homologs and their stage-specific expression profiles obtained from the RNA-seq data (Fig. 8a, b).

Arabidopsis thaliana is a facultative LD plant, in which inductive photoperiods are perceived in the leaves and the

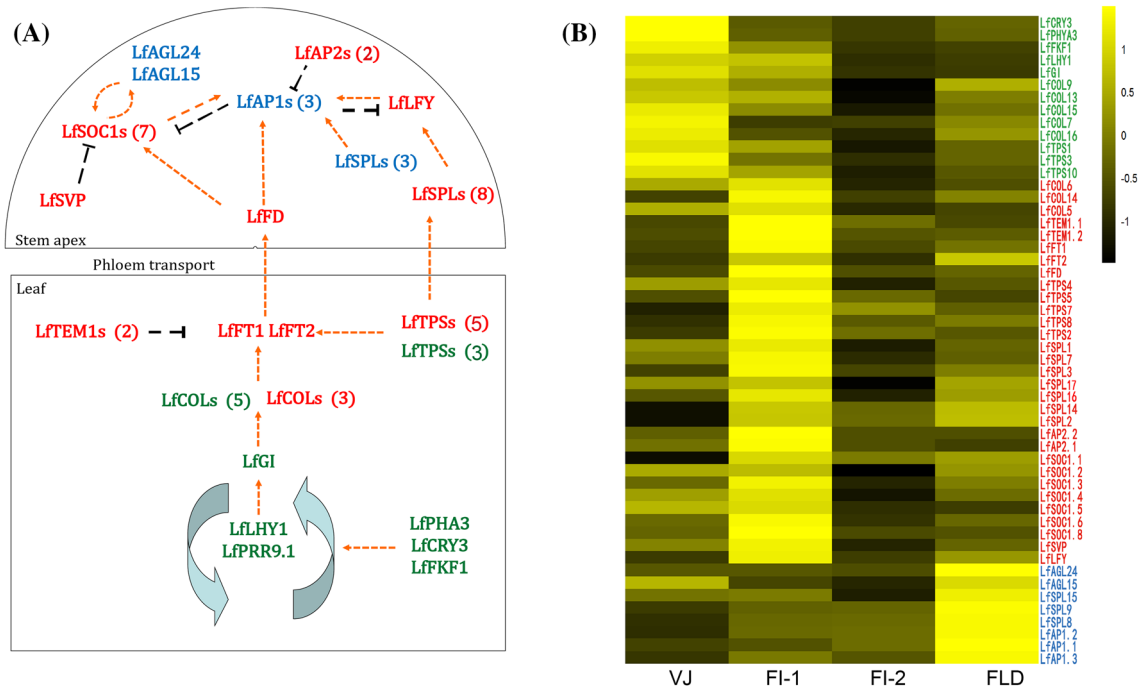


Fig. 8 Putative gene regulatory network of flowering induction and development, as well as the differential expression profiles of most flowering homologs of *L.×formolongi*. **a** The putative network of photoperiodic, age-dependent and T6P flowering pathways in *L.×formolongi*. Gene names are colored according to their expression patterns. *Green color* represents the highest expression in the VJ stage, *red* represents the up-regulated expression in the FI-1 stage, and *blue* represents the up-regulated expression in the FLD stage. And gene names are followed by the numbers of differentially expressed genes. *Arrows in red dotted lines* represent gene activation and *black blunted dotted lines* represent gene repression. **b** A heatmap exhibiting the differential expression profiles of most identified homologs in photoperiod, age-dependent and T6P pathways. Gene names are colored according to their expression patterns refer to the previous description in Fig. 6

circadian clock acts as a timekeeping to measure the day length (Samach and Coupland 2000). The simplest circadian clock models consist of photoreceptors, core oscillators and rhythm output pathways (Gardner et al. 2006). *GI, CO* and *FT* are distinct regulatory genes of the circadian clock output pathway, and *GI* promotes the expression of *CO* and *FT* under LD conditions (Mizoguchi et al. 2005). The expression of *FT* in the leaf vasculature is a main determinant of the timing of flowering. The FT protein is transported through the vasculature to the apex (Turck et al. 2008), where it interacts with the bZIP transcription factor *FD* and, coordinately, they up-regulate the MADS-box genes *SOC1, AP1* and *LFY* (Wellmer and Riechmann 2010). The miR156-*SPL*/age pathway acts both in the leaves and SAM. SPL proteins promote *FT* expression in leaves by regulating the expression of some MADS-box TFs, such as *SOC1* (Wahl et al. 2013), while at the shoot apex, *SPL*s and *FT*/*FD* converge on an overlapping set of targets, where *SPL*s directly activate the floral meristem identity genes *AP1* and *LFY* (Wang et al. 2009; Yamaguchi et al. 2009). The AP2 domain-containing protein genes normally act as floral repressors that down-regulate the expression of *SOC1* and *AP1* (Aukerman and Sakai 2003). Additionally, the T6P pathway regulates flowering at two sites in the plants. In the leaves, *TPS1* promotes flowering by activating the expression of *FT*. Furthermore, the T6P pathway controls the expression of several *SPL*s in the SAM (Wahl et al. 2013).

Thus, the photoperiod, age-dependent and T6P pathways are inter-connected both in the leaves and SAM, and converge through the key integrator *FT* gene, promoting the same targets, the floral meristem identity genes *SOC1, LFY* and *AP1*, to trigger flowering in the SAM.

In comparison with the genetic basis of flowering regulation in *A. thaliana*, the molecular regulatory network of flowering in *L.*×*formolongi* exhibited particular temporal and spatial expression patterns. The photoperiod pathway's circadian clock homologs may be the first gene set to perceive the light signals and circadian rhythms in the leaves and promote the expression of *LfGI* and *LfCOLs* in the VJ and FI-1 stages, which may be related to the prompt transition of flowering in *L.* × *formolongi*. Then, *LfCOLs* activate the high expression levels of *LfFT1* and *LfFT2*. The proteins of *LfFT1* and *LfFT2* were transported into the SAM and interacted with *LfFD* to activate *LfLFY* and some *LfSOC1*s in the FI-2 stage. Meanwhile, the *LfTPS* family members from the T6P pathway may also promote the expression of *FT* in the leaves, and further activate the expression of some *LfSPL* members in the SAM during the FI-1 stage. The up-regulated *LfSPL* members in the FI-1 stage may promote the expression of *LfLFY*, while the other *LfSPL*s that are up-regulated in the FLD stage may interact with the floral meristem identity *LfAP1* genes to promote flowering in the SAM. While the identified *LfAP2*s and *LfSVP* homologs may act as the repressors of the floral meristem identity *LfAP1* and *LfSOC1* genes (Fig. 8).

Conclusion

To our knowledge, this study is the first report showing genome-wide transcriptome dynamics during the photoperiod-mediated flowering initiation process in lily. In this study, we presented a detailed analysis of gene expression dynamics in *L.*×*formolongi*, covering the stages of vegetative and reproductive development, and the transition stages identified from molecular dynamics and morphological changes. This analysis revealed a large set of putative flowering genes, which were mostly discovered in lily for the first time and may play critical roles in the flowering transition of *L.*×*formolongi*. The detailed investigations of the candidate genes in the critical photoperiod, age-dependent and carbohydrate nutritional T6P pathways in this study will be very helpful in understanding the molecular mechanisms involved in flowering. Furthermore, the candidate homologs will provide references for the breeding of lilies with short vegetative periods and for reverse genetics approaches.

Acknowledgements This work has been supported by the National Natural Science Foundation of China (Grant No. 31470106) and the National Forestry Industry Research Special Funds for Public Welfare Projects (Grant No. 201204609).

Author contributions G-X Jia and M-F Zhang designed the experiments. Y-F Li, M-F Zhang and M Zhang performed the experiments. Y-F Li performed the bioinformatics and the statistical analyses of the transcriptiome and wrote the article. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Anderson NO, Berghauer E, Harris D, Johnson K, Lönnroos J, Morey M (2012) Discovery of novel traits in seed-propagated Lilium: non-vernalizationrequiring, day-neutral, reflowering, frost-tolerant, winter-hardy *L.*×*formolongi*. I. Characterization. Floric Ornam Biotechnol 6:63–72

Andrés F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 13:627–639

asl Hamid B, Kim JH (2011) Cross compatibility between *Lilium*×*fomolongi* group and *Lilium brownii*. Afr J Agric Res 6:968–977

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2-like* target genes. Plant Cell 15:2730–2741

Beattie D, White J (1993) Lilium-hybrids and species. In: De Hertogh A, Le Nard M (eds) The physiology of flower bulbs. Elsevier, Amsterdam, pp 423–454

Bergonzi S et al (2013) Mechanisms of age-dependent response to winter temperature in perennial flowering of *Arabis alpina*. Science 340:1094–1097

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) *CO*/*FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040–1043

Datta S, Hettiarachchi G, Deng X-W, Holm M (2006) *Arabidopsis CONSTANS-LIKE 3* is a positive regulator of red light signaling and root growth. Plant Cell 18:70–84

Du F et al (2015) De novo assembled transcriptome analysis and SSR marker development of a mixture of six tissues from *Lilium* Oriental hybrid 'Sorbonne'. Plant Mol Biol Rep 33:281–293

Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 7:1

Gardner MJ, Hubbard KE, Hotta CT, Dodd AN, Webb AA (2006) How plants tell the time. Biochem J 397:15–24

Grabherr MG et al (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29:644–652

Griffiths D (1934) Bulbs from seed. US Government Printing Office, Washington, D.C.

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS-like* gene families in barley, rice, and *Arabidopsis*. Plant Physiol 131:1855–1867

Hassidim M, Harir Y, Yakir E, Kron I, Green RM (2009) Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. Planta 230:481–491

Horita M, Morohashi H, Komai F (2003) Production of fertile somatic hybrid plants between Oriental hybrid lily and *Lilium*×*formolongi*. Planta 217:597–601

Huang J, Liu X, Wang J, Lü Y (2014) Transcriptomic analysis of Asiatic lily in the process of vernalization via RNA-seq. Mol Biol Rep 41:3839–3852

Jain M (2011) Next-generation sequencing technologies for gene expression profiling in plants. Brief Funct Genomics 11:63–70

Kardailsky I et al (1999) Activation tagging of the floral inducer *FT*. Science 286:1962–1965

Khan MRG, Ai XY, Zhang JZ (2014) Genetic regulation of flowering time in annual and perennial plants. Wiley Interdisc Rev RNA 5:347–359

Kim S-K, Yun C-H, Lee JH, Jang YH, Park H-Y, Kim J-K (2008) *OsCO3*, a *CONSTANS-LIKE* gene, controls flowering by negatively regulating the expression of *FT-like* genes under SD conditions in rice. Planta 228:355–365

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 43:1096–1105

Lazare S, Zaccai M (2016) Flowering pathway is regulated by bulb size in *Lilium longiflorum* (Easter lily). Plant Biol 1–8. doi:10.1111/plb.12440

Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. Plant J 26:15–22

Liu X, Wang Q, Gu J, Lü Y (2014) Vernalization of Oriental hybrid lily 'Sorbonne': changes in physiology metabolic activity and molecular mechanism. Mol Biol Rep 41:6619–6634

Meng X, Muszynski MG, Danilevskaya ON (2011) The *FT-like ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. Plant Cell 23:942–960

Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, Weigel D (1991) A genetic and molecular model for flower development in *Arabidopsis thaliana*. Development 113:157–167

Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM (2005) Integration of flowering signals in winter-annual. *Arabidopsis*. Plant Physiol 137:149–156

Mizoguchi T et al (2005) Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in *Arabidopsis*. Plant Cell 17:2255–2270

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-SEq. Nat Methods 5:621–628

Mouradov A, Cremer F, Coupland G (2002) Control of flowering time interacting pathways as a basis for diversity. Plant Cell 14:S111–S130

O'Hara LE, Paul MJ, Wingler A (2013) How do sugars regulate plant growth and development? New insight into the role of trehalose-6-phosphate. Mol Plant 6:261–274

Okazaki K (1994) *Lilium* species native to Japan, and breeding and production of *Lilium* in Japan. Acta Hortic 414:81–92

Optiz E, Anderson N, Younis A (2009) Development of colored, non-vernalization-requiring seed propagated lilies. In: 23rd International Eucarpia Symposium, Section Ornamentals: Colourful Breeding and Genetics 836:193–198

Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87–98

Pařenicová L et al (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis* new openings to the MADS world. Plant Cell 15:1538–1551

Pertea G et al (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652

Proveniers M (2013) Sugars speed up the circle of life. Elife 2:e00625

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847–857

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140

Robson F, Costa MMR, Hepworth SR, Vizir I, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. Plant J 28:619–631

Sakamoto H (2005) Acceleration of flowering by night break and heating treatment for harvesting in April and May in *Lilium*×*foromolongi* cv. Hayachine. Hortic Res 4:191–195

Samach A, Coupland G (2000) Time measurement and the control of flowering in plants. Bioessays 22:38–47

Singh VK, Garg R, Jain M (2013) A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. Plant Biotechnol J 11:691–701

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410:1116–1120

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. Annu Rev Plant Biol 59:573–594

Villacorta-Martin C, de Cáceres González FFN, de Haan J, Huijben K, Passarinho P, Hamo ML-B, Zaccai M (2015) Whole transcriptome profiling of the vernalization process in *Lilium longiflorum* (cultivar White Heaven) bulbs. BMC Genomics 16:1

Wahl V et al (2013) Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. Science 339:704–707

Wang J-W, Czech B, Weigel D (2009) miR156-regulated *SPL* transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. Cell 138:738–749

Wellmer F, Riechmann JL (2010) Gene networks controlling the initiation of flower development. Trends Genet 26:519–527

Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM (2006) Genome-wide analysis of gene expression during early *Arabidopsis* flower development. PLoS Genet 2:e117

Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T (2005) *TWIN SISTER OF FT* (*TSF*) acts as a floral pathway integrator redundantly with *FT*. Plant Cell Physiol 46:1175–1189

Yamaguchi A, Wu M-F, Yang L, Wu G, Poethig RS, Wagner D (2009) The microRNA-regulated SBP-Box transcription factor *SPL3* is a direct upstream activator of *LEAFY, FRUITFULL*, and *APETALA1*. Dev Cell 17:268–278

Yoo SJ, Chung KS, Jung SH, Yoo SY, Lee JS, Ahn JH (2010) *BROTHER OF FT AND TFL1* (*BFT*) has *TFL1-like* activity and functions redundantly with *TFL1* in inflorescence meristem development in *Arabidopsis*. Plant J 63:241–253