

SCIENTIFIC REPORTS



OPEN

Stages identifying and transcriptome profiling of the floral transition in *Juglans regia*

Shaowen Quan^{1,2}, Jianxin Niu^{1,2}, Li Zhou^{1,2}, Hang Xu^{1,2}, Li Ma^{1,2} & Yang Qin^{1,2}

Using paraffin sections, the stages of walnut female flower bud differentiation were divided into the predifferentiation period (F_1), initial differentiation period (F_2) and flower primordium differentiation period (F_3). Leaf buds collected at the same stage as F_2 were designated JRL. Transcriptomic profiling was performed, and a total of 132,154 unigenes were obtained with lengths ranging from 201 bp to 16,831 bp. The analysis of differentially expressed genes (DEGs) showed that there were 597, 784 and 532 DEGs in the three combinations F_1vsF_2, F_1vsF_3, and F_2vsF_3, respectively. The comparison F_2vsJRL showed that 374 DEGs were differentially expressed between female buds and leaf buds. Thirty-one DEGs related to flowering time were further used to construct coexpression networks, and CRY2 and NF-YA were identified as core DEGs in flowering time regulation. Eighteen DEGs related to flowering time were subjected to real-time quantitative analysis. Our work provides a foundation for further research on the walnut floral transition and provides new resources for future research on walnut biology and biotechnology.

Juglans regia is widely cultivated for its commercially valuable timber and nuts in China and other temperate parts of the world¹. In 2016, global production of walnuts (in shell) was 3.7 million metric tons, and China was the leading producer worldwide with 41% of total production (FAO, 2017)². In contrast to annual plants, walnut trees require several years before flowering and to bear fruit after seed germination, and little has been reported about the mechanism of the floral transition in walnut.

Many studies have been performed on the floral transition in *Arabidopsis thaliana* and other plants^{3–12}. Six genetic pathways (including the photoperiod, vernalization, temperature, gibberellin, autonomy and age pathways) and several flowering integrator genes (including *FLOWERING LOCUS T (FT)*, *LEAFY (LFY)*, and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*) have been shown to regulate flowering time in *Arabidopsis thaliana*^{3,6–8,12–18}. However, few flowering time genes have been identified in walnut.

In this study, sections of paraffin-embedded tissues were used to observe the morphogenesis of floral buds in *Juglans regia*, and a high-throughput sequencing platform was used to sequence cDNA libraries at each stage of the floral transition. We investigated the differentially expressed genes (DEGs) involved in the developmental process of the walnut floral transition, and our results demonstrated that DEGs between the stages before and during the floral transition played a key role in regulating the floral transition. These results provide essential information on the genes and pathways involved in floral transition development in the walnut and may help in further researching the molecular mechanism of walnut flowering.

Results

Morphological characteristics of walnut floral transition. Morphological differentiation of female flower buds in walnut was continuously observed by paraffin section. We divided walnut female flower bud differentiation into stages as follows: F_1 (predifferentiation period), F_2 (initial differentiation period) and F_3 (flower primordium differentiation period). In the predifferentiation period (F_1), the growth points of the flower bud appeared flat, and the bud scales appeared light green. In the initial differentiation period (F_2), the flower stalk primordium was raised, and the bud scales were yellow-green. In the flower primordium differentiation

¹Department of Horticulture, College of Agriculture, Shihezi University, Shihezi, 832003, Xinjiang, China. ²Xinjiang Production and Construction Corps Key Laboratory of Special Fruits and Vegetables Cultivation Physiology and Germplasm Resources Utilization, Shihezi, 832003, Xinjiang, China. Correspondence and requests for materials should be addressed to J.N. (email: njx105@163.com)

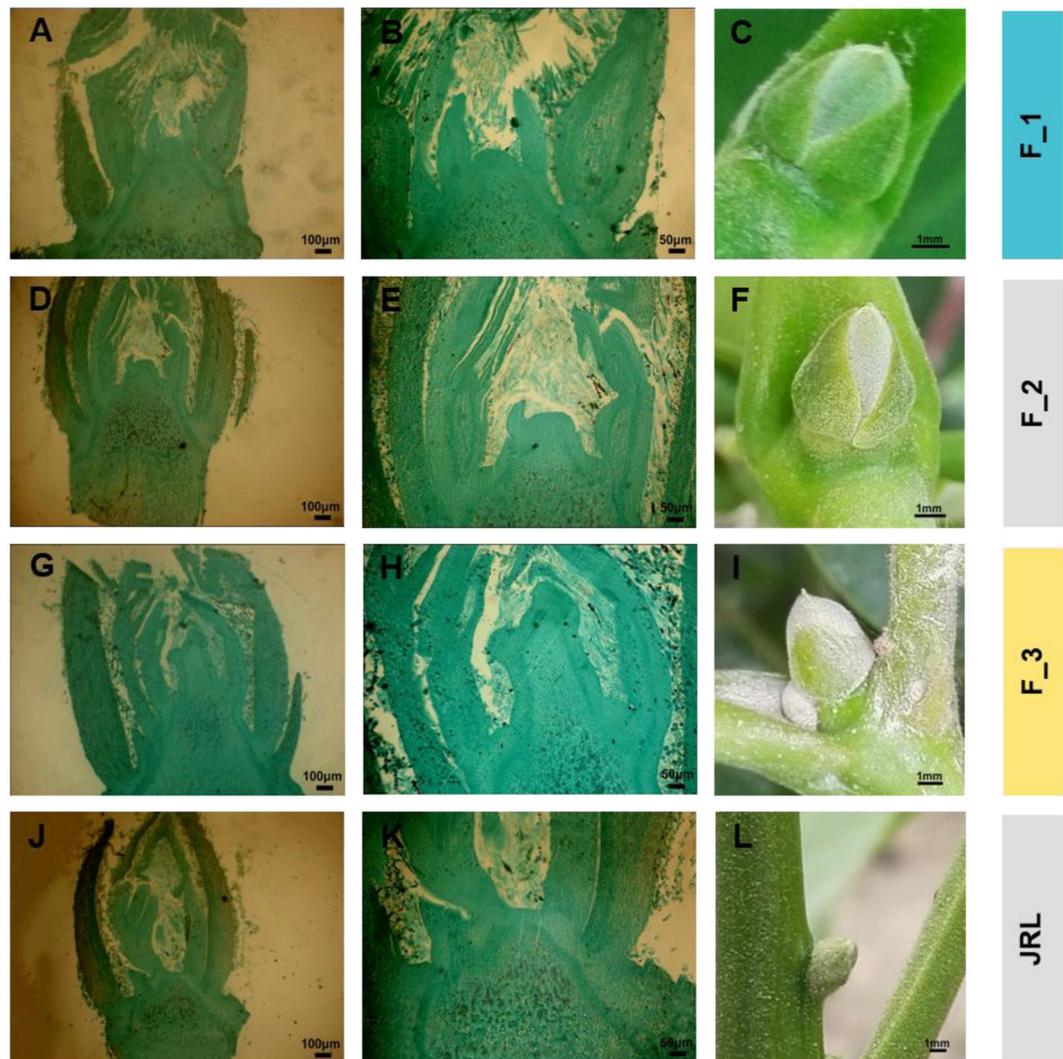


Figure 1. Morphological characteristics of walnut female flower buds. (A–C) show the female flower buds in the predifferentiation period identified in the undifferentiated stage (F_1); (D–F) show the female flower buds in the initial differentiation period (F_2); (G–I) show the female flower buds identified at the flower primordium differentiation period (F_3); (J–L) show the leaf buds (JRL) collected at the same stage of F_2. (A,D,G,J) are images of paraffin sections with a 100 μm scale bar; (B,E,H,K) are images of paraffin sections with a 50 μm scale bar; (C,F,I,L) show the external morphology of the buds at different stages with a 1 mm scale bar.

| Sample | Raw reads | Clean reads | Clean bases | Error (%) | Q20 (%) | Q30 (%) | GC (%) |
|--------|-----------|-------------|-------------|-----------|---------|---------|--------|
| F_1 | 57270794 | 56064740 | 8.41G | 0.02 | 96.86 | 92.20 | 44.83 |
| F_2 | 42190694 | 40686164 | 6.1G | 0.03 | 94.41 | 86.66 | 44.92 |
| F_3 | 48518386 | 46937334 | 7.04G | 0.03 | 94.79 | 87.59 | 45.15 |
| JRL | 46946334 | 45353312 | 6.8G | 0.03 | 94.63 | 87.12 | 44.94 |

Table 1. Quality of data output.

period (F_3), the primordium of the flower stalk continued to extend, the pistil primordium began to appear, and the bud scales appeared yellow-green (Fig. 1). The leaf buds (JRL) were collected at the same stage as F_2.

General analysis of walnut transcriptome data. *Sequencing, splicing, and annotation of transcriptome data.* A total of 194,926,208 raw reads were obtained through the sequencing of the transcript libraries (Table 1). By eliminating chimeric and low-quality reads, 189,041,550 clean reads were finally screened, and 132,154 unigenes were obtained with a mean length of 956 bp; the lengths of the unigenes ranged from 201 bp to 16831 bp (Fig. 2A).

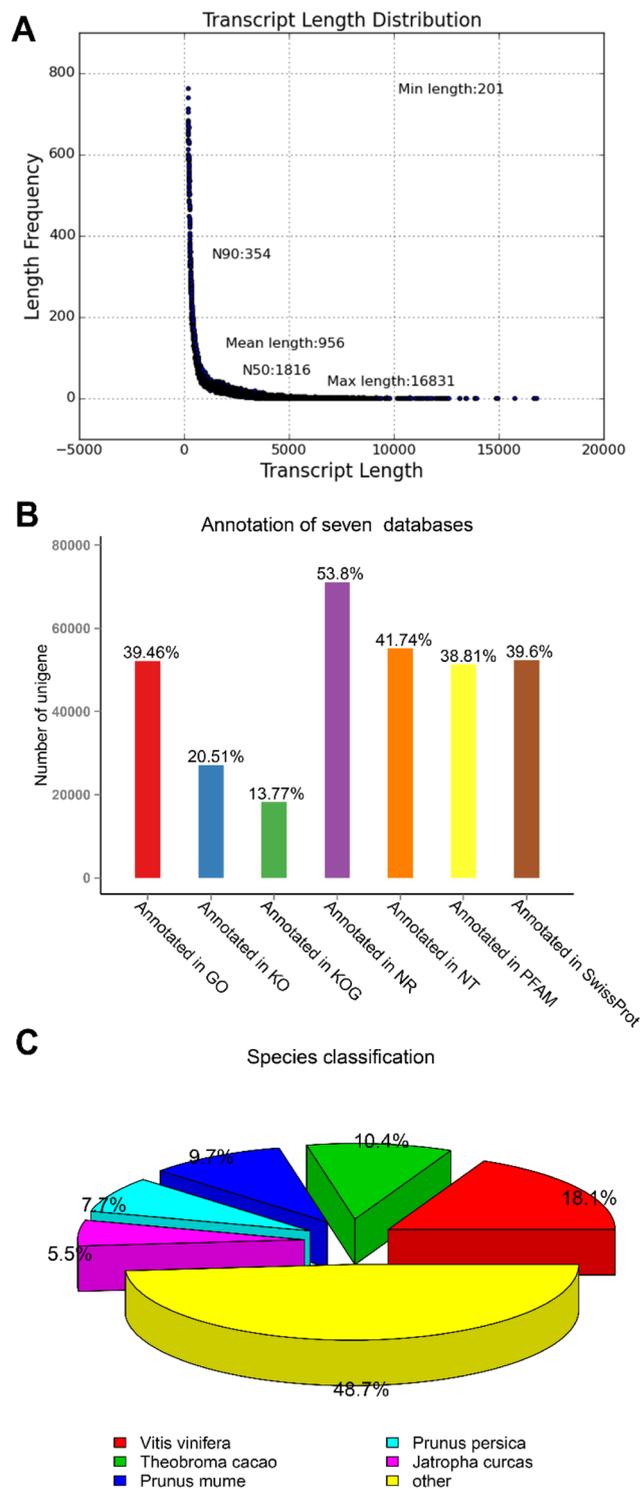


Figure 2. Analysis of the transcriptome data. (A) length distribution of transcripts; (B) transcriptome data annotated in seven databases; (C) the species classification of the walnut transcriptome data annotated by the NR database.

All unigenes were compared with 7 databases (NR, NT, KEGG, Swiss-Prot, PFAM, GO, and KOG/COG) with an E value cutoff of e^{-5} for homology. Then, the unigenes were annotated according to the results: 39.46% of unigenes were annotated by the GO database, 20.51% of unigenes were annotated by the KO database, 13.77% of unigenes were annotated by the KOG database, 53.8% of unigenes were annotated by the NR database, 41.74% of unigenes were annotated by the NT database, 38.81% of unigenes were annotated by the PFAM database, and 39.6% of unigenes were annotated by the Swiss-Prot database. In total, 78,718 unigenes of the 132,154 total

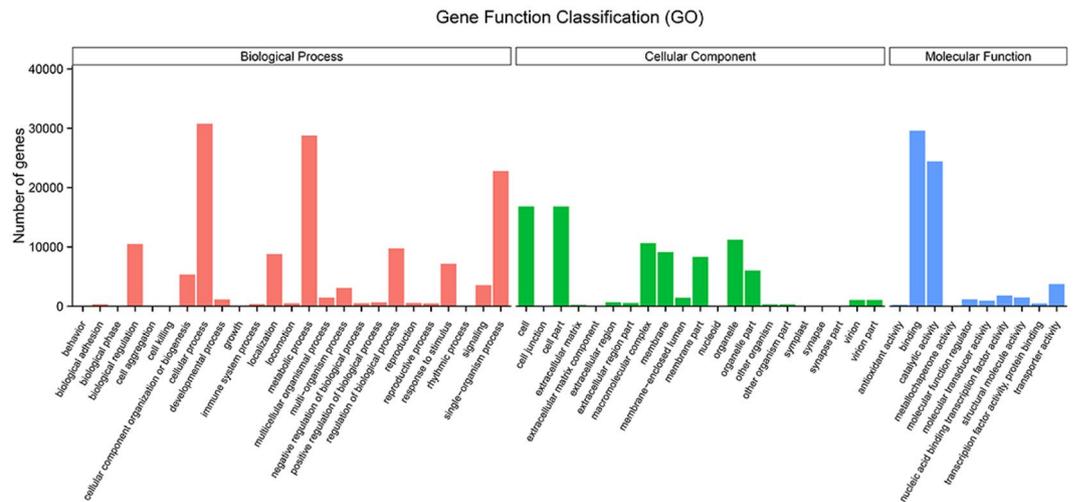


Figure 3. Gene numbers of the enriched GO terms.

unigenes were annotated to at least one database, accounting for 59.56% of all unigenes, and 11,165 unigenes were annotated by all seven major databases, accounting for 8.44% of all unigenes (Fig. 2B).

The annotation results of the NR database show that 11372 unigenes (18.1% of the total annotated unigenes) were compared to *Vitis vinifera*, 6522 unigenes (10.4% of the total annotated unigenes) were compared to *Prunus mume*, 6141 unigenes (9.7% of the total annotated unigenes) were compared to *Prunus persica*, 4834 unigenes (7.7% of the total annotated unigenes) were compared to *Jatropha curcas*, 3457 unigenes (5.5% of the total annotated unigenes) were compared to *Citrus sinensis*, and the remaining 48.7% unigenes were compared to other species (Fig. 2C).

GO enrichment. The gene functional classifications were identified in a GO enrichment analysis based on the 52,157 unigenes annotated in the GO database. In the biological process category, cellular process, metabolic process, and single organism process were highly represented. In the cellular component category, significantly enriched genes were associated with cell, cell part, and organelle. In the molecular function category, GO terms related to binding, catalytic activity and transporter activity were significantly enriched (Fig. 3). **KEGG pathway enrichment.** KEGG pathway enrichment analysis showed that transport and catabolism, signal transduction, transition, carbohydrate metabolism and environmental adaptation had the most unigenes in classifications of Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes and Organismal Systems, respectively (Fig. 4). In particular, Environmental Information Processing and Organismal Systems, which included many subterms associated with flowering, such as the pathways of plant hormone signal transduction (ko04075) and Circadian rhythm-plant (ko04712), were enriched (Table S1).

Transcription factor (TF) enrichment analysis. TFs function alone or with other proteins in complexes by promoting or blocking the recruitment of RNA polymerase to specific genes^{19–21}. Some TFs have been shown to play key roles in plant flowering^{22–24}. Among the 132,154 unigenes, 4,835 of the unigenes (approximately 3.7%) were annotated as TFs, and they fell into diverse categories covering nearly all TF families, with MYB, NAC, and AP2-EREBP being the most highly represented (Fig. 5).

DEGs analysis of walnut transcriptome data. Identification of DEGs. The four databases from F_1, F_2, F_3, and JRL were used for paired comparisons, and four combinations (F_1vsF_2, F_1vsF_3, F_2vsF_3, F_2vsJRL) were chosen and analyzed.

The three combinations of female flower bud developmental stages (F_1vsF_2, F_1vsF_3, F_2vsF_3) had 597 DEGs, 784 DEGs and 532 DEGs, respectively. In addition, 16 DEGs were shared by all three combinations, which means they were expressed differently in the three periods of female flower buds (Fig. 6). The expression levels of DEGs between F_1, F_2 and F_3 were plotted in a heatmap (Fig. 7A).

In addition, the female flower buds at the beginning of differentiation (F_2) and the leaf buds of the same stage (JRL) were used for comparison (F_2vsJRL). The results showed that 374 DEGs were expressed differently in female buds (F_2) and leaf buds (JRL), of which 90 unigenes in the female flower buds (F_2) showed a higher expression level than that in the leaf buds (JRL), and the other 284 unigenes showed the opposite trend (Fig. 6E). The expression levels of 374 DEGs between female buds (F_2) and leaf buds (JRL) were plotted in a heatmap (Fig. 7B).

GO functional analysis of DEGs. We conducted GO functional analysis with the upregulated and downregulated DEGs of F_1vsF_2, F_1vsF_3, and F_2vsF_3, and the results are shown in Fig. 8. The results indicated that the DEGs of the three stages of female flower bud differentiation had functions enriched in the regulation of RNA metabolic process (GO:0051252, BP); regulation of nucleobase-containing compound metabolic process (GO:0019219, BP); apoplast (GO:0048046, CC); nucleic acid binding transcription factor activity (GO:0001071,

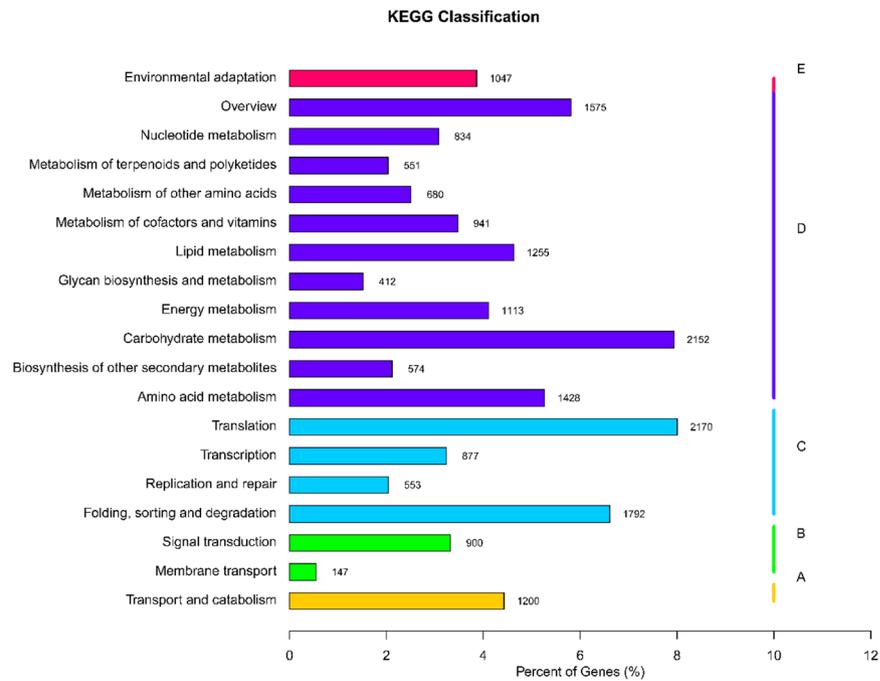


Figure 4. KEGG pathway analysis of the transcriptome data. The abscissa is the ratio of the single pathway gene number to the total annotated gene number, and the ordinate is the name of the KEGG metabolic pathway. We divided the involved genes into 5 branches: (A) Cellular Processes; (B) Environmental Information Processing; (C) Genetic Information Processing; (D) Metabolism; (E) Organismal Systems.

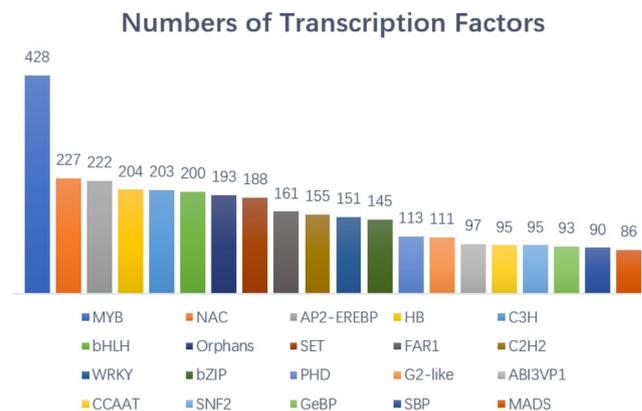


Figure 5. Transcription factor enrichment analysis of the walnut transcriptome data.

MF); transcription factor activity, sequence-specific DNA binding (GO:0003700, MF); and sequence-specific DNA binding (GO:0043565, MF). In addition, the GO functions of the DEGs between female flower buds (F_2) and leaf buds (JRL) were enriched in photosynthesis (GO:0015979, BP), photosystem I reaction center (GO:0009538, CC), photosystem I (GO:0009522, CC), photosystem (GO:0009521, CC), and thylakoid part (GO:0044436, CC).

KEGG pathway analysis of DEGs: KEGG pathway analysis of the DEGs of F_1VVSF_2, F_1VVSF_3, F_2VVSF_3 showed that most pathways were associated with flowering, and three pathways (Photosynthesis-antenna proteins, Plant hormone signal transduction, Porphyrin and chlorophyll metabolism) were shared by the three combinations. In addition, the pathway enrichment analysis of female flower buds (F_2) and leaf buds (JRL) also included many flowering-related pathways (Fig. 9).

Analysis of DEGs related to flowering time: The flowering time genes that induce plant flowering are usually attributed to the photoperiod, vernalization, autonomous, GA, and age-regulated pathways, as well as the integrated factors of the intersection of these pathways^{3,6,13,25}.

Based on previous studies, we summarized the important flowering time genes in the model plant *Arabidopsis thaliana*, and these genes were used as query genes (Table 2). Finally, we screened 31 DEGs associated with flowering time in walnut, and eighteen of the flowering time DEGs were chosen for qRT-PCR.

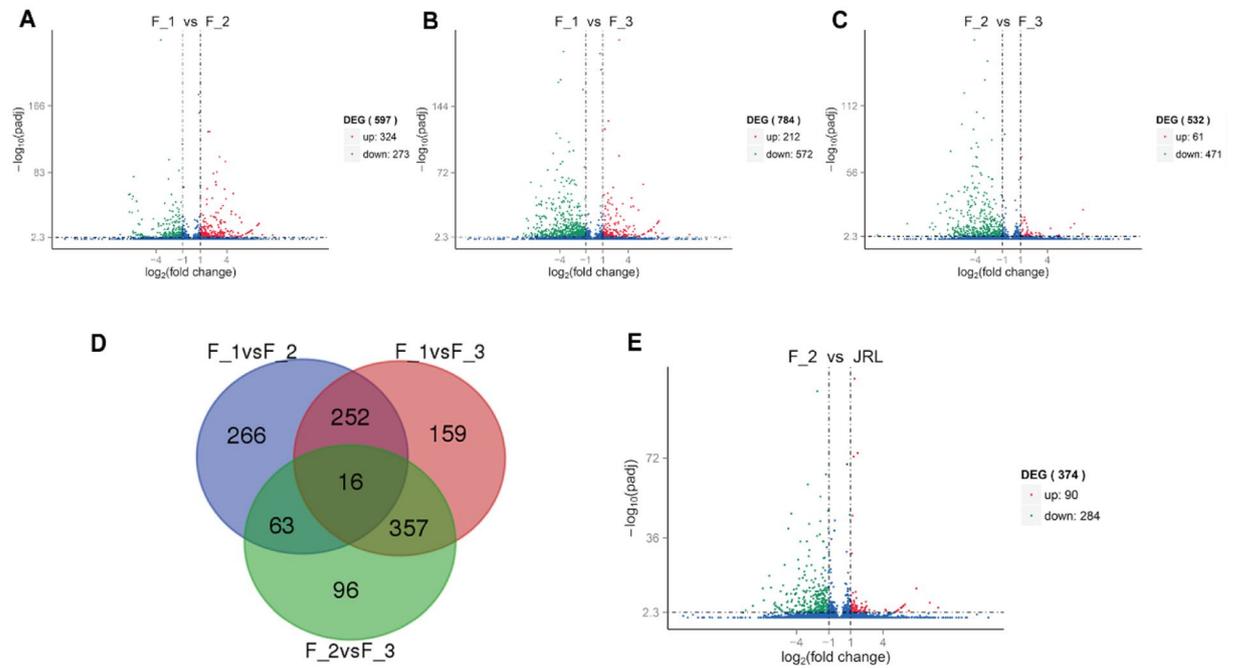


Figure 6. Numbers of DEGs in F_1, F_2, F_3 and JRL. (A–C,E) volcano plot of the DEGs in F_1vsF_2, F_1vsF_3, F_2vsF_3 and F_2vsJRL. The abscissa represents the multiple expression of genes in different samples. The ordinate represents the statistical significance of the change in gene expression. The scattered points in the map represent the genes, and genes without significant differences are presented as blue dots. Upregulated genes and downregulated genes with significant differences are presented as red and green dots, respectively; (D) Venn diagram of F_1vsF_2, F_1vsF_3 and F_2vsF_3.

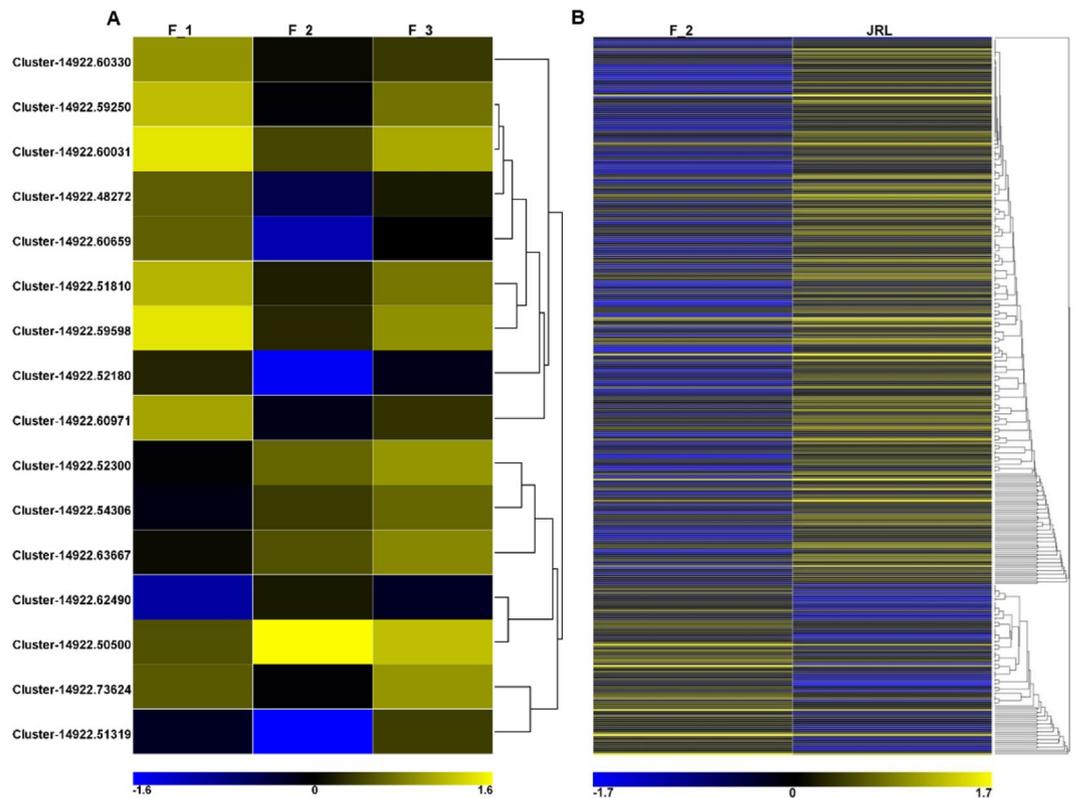


Figure 7. (A) Heatmap of 16 DEGs expressed differently in three stages of female flower buds (F_1, F_2, and F_3); (B) Heatmap of DEGs between female flower buds (F_2) and leaf buds (JRL).

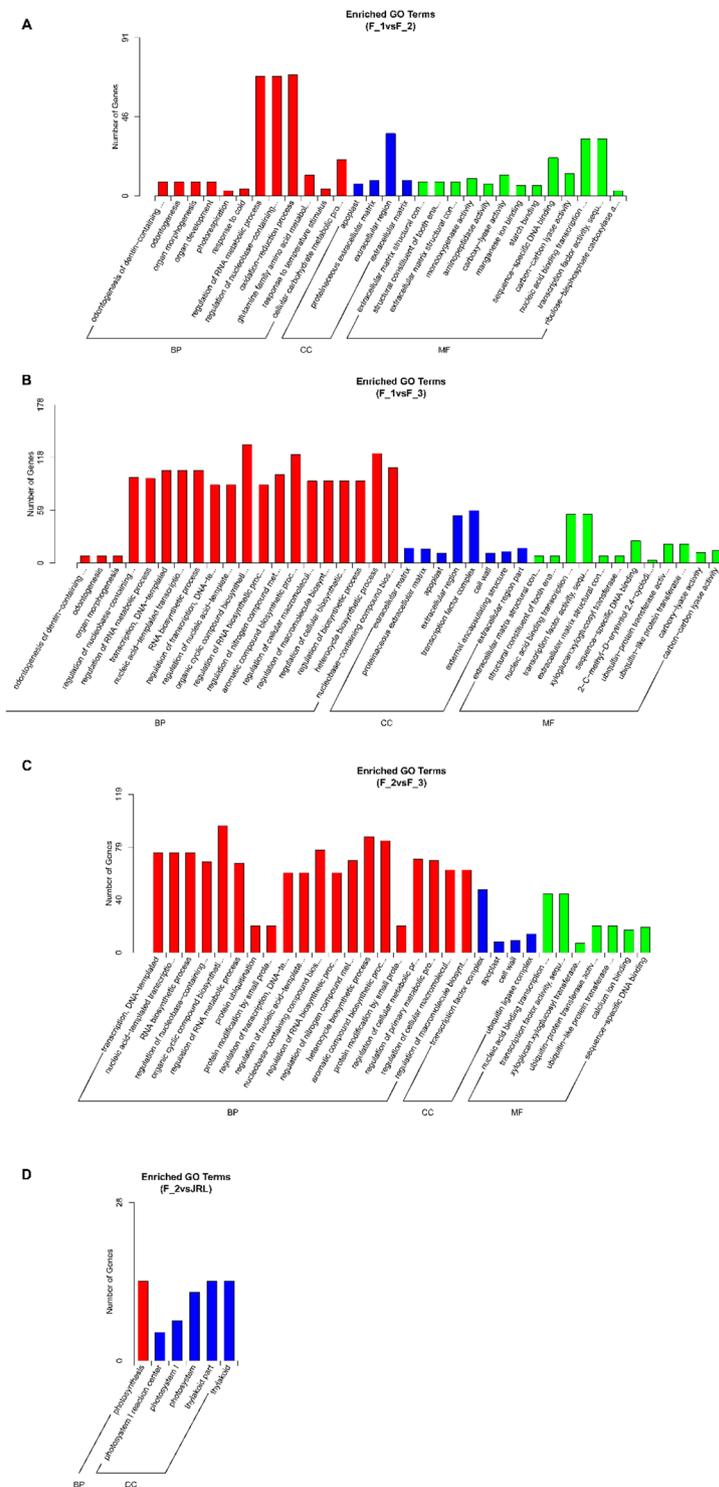


Figure 8. (A–D) represent the enriched GO terms of the DEGs in F_1vsF_2, F_1vsF_3, F_2vsF_3 and F_2vsJRL, respectively. The upregulated GO terms are shown with red bars, and the downregulated GO terms are shown with blue bars.

DEGs related to the walnut floral transition. A total of 31 unigenes were involved in the photoperiod (circadian rhythm), vernalization pathway, GA, and age-related pathways. To investigate the complex molecular mechanisms underlying floral transition in walnut, we set the 31 flowering time-related DEGs in the flowering time regulation network (Fig. 10).

Most of the DEGs were annotated in the photoperiod pathway: In total, twenty-four DEGs were annotated in this pathway, fifteen DEGs were highly expressed in the predifferentiation period (F_1), including *CRY2* (1), *FKF1* (1), *GI* (2), *NF-YA* (2), *PIF3* (1), *PPR5* (2), *PPR7* (3), *PPR9* (1), *PPR73* (1), and *PPR95* (1). *CDF2* (1), *CDF3*

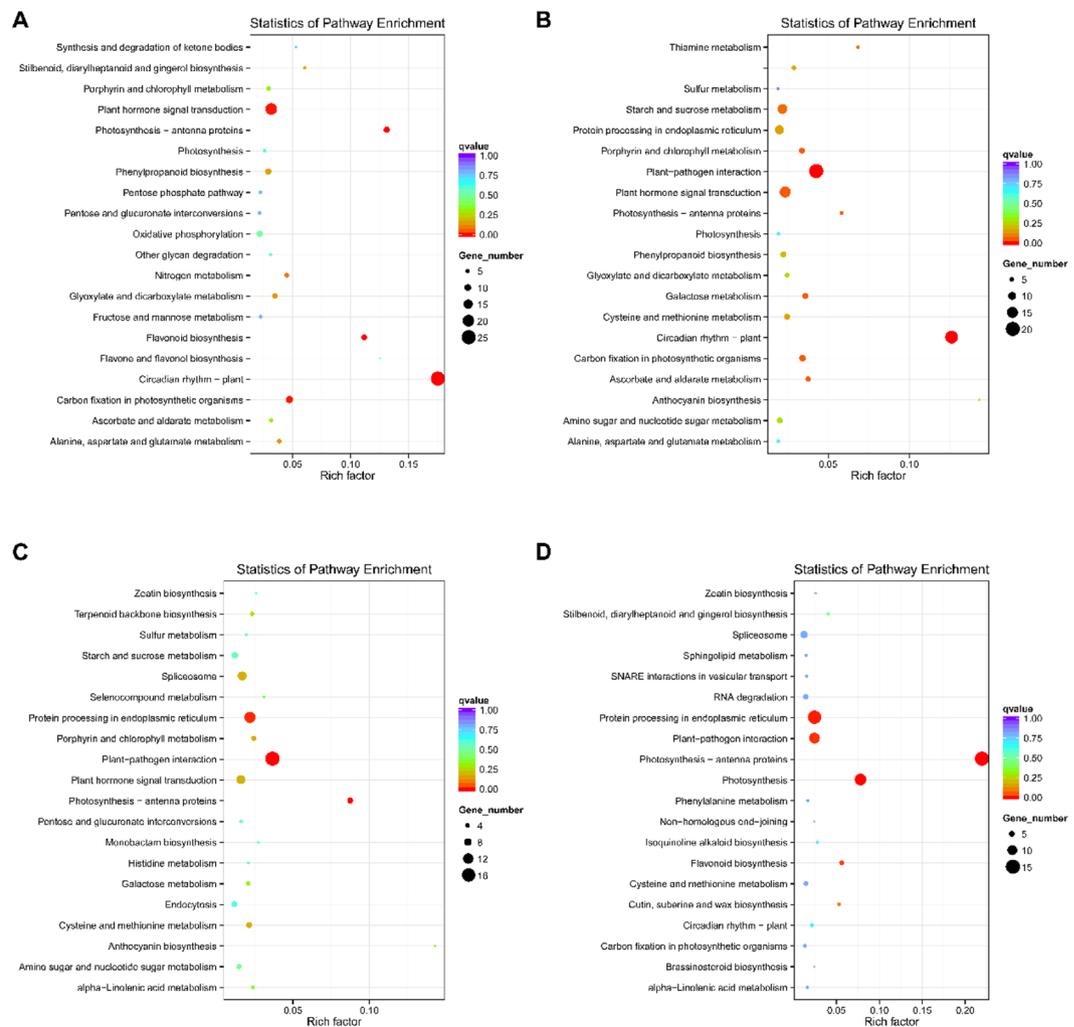


Figure 9. (A–D) represent scatterplots of the top 20 enriched KEGG pathways of the DEGs in F_1vsF_2, F_1vsF_3, F_2vsF_3 and F_2vsJRL, respectively. The vertical axis represents the pathway names, the horizontal axis represents the rich factors corresponding to the pathways, the size of the q value is represented by the color of the dots, and the number of differentially expressed genes in each pathway is expressed by the size of the dots.

| Flowering Pathways | Involved Genes (Number) |
|----------------------|--|
| Photoperiod | <i>CCA1</i> (0), <i>CDFs</i> (2), <i>CO</i> (0), <i>COP1</i> (0), <i>CRY2</i> (1), <i>DNF</i> (0), <i>ELF3</i> (0), <i>ELF4</i> (0), <i>FD</i> (0), <i>FKF1</i> (1), <i>FLD</i> (0), <i>FLK</i> (0), <i>GI</i> (2), <i>LHY</i> (2), <i>NF-Ys</i> (3), <i>PHYA</i> (0), <i>PHYB</i> (0), <i>PHR2</i> (2), <i>PIF3</i> (1), <i>PRRs</i> (8), <i>SMZ</i> (0), <i>SNZ</i> (0), <i>SPA</i> (0), <i>TEM1</i> (2), <i>TEM2</i> (0), <i>TOC1</i> (0) |
| Vernalization | <i>FLC</i> (0), <i>FRI</i> (0), <i>LHP1</i> (0), <i>VIL2</i> (1), <i>VIN3</i> (0), <i>VRN2</i> (0), <i>VRN1</i> (0), <i>VRN2</i> (0) |
| Autonomous | <i>FCA</i> (0), <i>FLC</i> (0), <i>FLD</i> (0), <i>FLK</i> (0), <i>FLM</i> (0), <i>FPA</i> (0), <i>FVE</i> (0), <i>FY</i> (0), <i>LD</i> (0), <i>SVP</i> (0), <i>REF6</i> (0) |
| Ambient temperatures | <i>ARP6</i> (0), <i>HOS1</i> (0), <i>SVP</i> (0), <i>HVP</i> (0), <i>miR156</i> (0), <i>SPL3</i> (0) |
| GA | <i>GA1</i> (0), <i>GAI</i> (0), <i>GID</i> (1), <i>GNC</i> (0), <i>GNL</i> (0), <i>RGA</i> (0), <i>RGL1</i> (0), <i>RGL2</i> (0), <i>RGL3</i> (0), <i>PY</i> (0), <i>FUL</i> (0), <i>RGA</i> (0) |
| Age regulated | <i>SPLs</i> (0), <i>MIR156</i> (0), <i>MIR172</i> (0), <i>JMJ18</i> (0), <i>TPS1</i> (3) |
| Integrators | <i>FT</i> (0), <i>LFY</i> (0), <i>SOC1</i> (0) |
| Supplement | <i>API</i> (2) |

Table 2. Important genes involved in flowering time pathways. The number of walnut DEGs is shown after each gene name.

(1), and *LHY* (2) were highly expressed in the initial differentiation period (F_2). *NF-YA* (1) and *TEM1* (2) were highly expressed in the flower primordium differentiation period (F_3). Two *PHR2* genes had higher expression levels in leaf buds (JRL) than in flower buds (F_2), while *PRR9* exhibited the opposite trend.

In *Arabidopsis thaliana*, *CYCLING DOF FACTOR 2* (*CDF2*) and *CYCLING DOF FACTOR 3* (*CDF3*) act redundantly to reduce *CONSTANS* (*CO*) expression and are degraded by the complex of *GIGANTEA* (*GI*) and *FLAVIN-BINDING KELCH REPEAT F-BOX 1* (*FKF1*), thus releasing repression of *CO* and *FT* transcription^{26–28}.

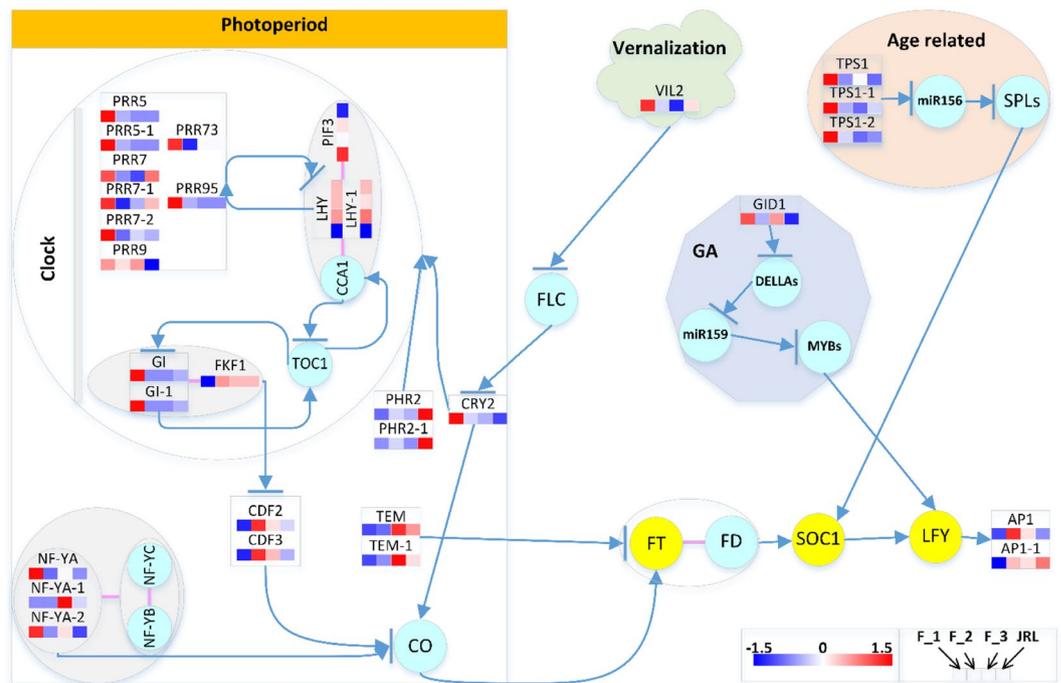


Figure 10. Heatmap of flowering-time-related DEGs. The DEG expression levels are indicated by color code according to the scale. The squares from left to right indicate the samples in F_1, F_2, F_3, and JRL. Genes within the circle were not differentially expressed in our study; circles filled with azure blue indicate normal genes, and circles filled with yellow indicate integral genes. Arrows indicate promotion, and arrows with a bar indicate inhibition. Red lines between genes indicate that these genes can combine to form a complex. The heatmap was drawn with HEMI according to the DEG expression, and the flow chart was drawn with Visio.

Blue-light photoreceptors such as *PHOTOLYASE-RELATED 2* (*PHR2*) are essential light detectors for the early development of plants and mediate phototropism and the expression of specific genes. Loss of a blue-light photoreceptor in the *hy4* mutants of *Arabidopsis thaliana* substantially delayed flowering^{29,30}. *CRYPTOCHROME 2* (*CRY2*) can promote the expression of *FT* and is negatively regulated by *FLOWERING LOCUS C* (*FLC*)³¹. The NUCLEAR FACTOR Y (NF-Y) transcription factors (heterotrimeric complexes composed of NF-YA and the dimer of NF-YB/NF-YC) can initiate photoperiod-dependent flowering by cooperatively interacting with *CO* to drive the expression of *FT*³².

Phytochrome Interacting Factor 3 (*PIF3*) can bind the promoters of *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) to form an *in vitro* ternary complex. Additionally, the complex of *CCA1* and *LHY* can bind with *TIMING OF CAB 1* (*TOC1*) to form a feedback loop that is necessary for the circadian clock in *Arabidopsis thaliana*³³. In addition, the *CCA1* and *LHY* complexes were repressed by the *Pseudo Response Regulators* (*PRRs*) encoded by *TOC1*^{6,34}. The activator *CO* and the repressor *TEMPRANILLO* (*TEM*) have a quantitative balance that determines the *FT* transcription level, as it shifts the *CO/TEM* balance in favor of *CO* activity, allowing *FT* transcription to reach the threshold level required to trigger flowering³⁵.

In the age-regulated pathway, three DEGs were annotated as *TREHALOSE-6-PHOSPHATE SYNTHASE 1* (*TPS1*), which is essential for normal vegetative growth and transition to flowering^{36,37}. They were differentially expressed between the different flower bud formation stages, and the flower buds showed no difference from the leaf buds in their expression level.

In the vernalization pathway, vernalization can cause epigenetic changes in *FLC* and indirectly promote flower formation by histone modification, and *MADS AFFECTING FLOWERING 5* (*MAF5*) is an *FLC*-related gene^{38–43}. *VERNALIZATION INSENSITIVE 3-LIKE 2* (*VIL2*), which can repress *MAF5* and permit more rapid flowering during noninductive photoperiod conditions in *Arabidopsis thaliana*⁴⁴. In this study, *VIL2* was highly expressed in the predifferentiation period (F_1), decreased in the initial differentiation period (F_2), and decreased to the lowest point at the flower primordium differentiation period (F_3). In the leaf buds (JRL), its expression level was higher than that in the flower buds during the same period (F_2).

In the GA pathway, GA regulates *LEAFY* (*LFY*) via *DELLAs*, *miR159* and *MYBs*^{45–47}. The combination of gibberellin and *GIBBERELLIN-INSENSITIVE DWARF1* (*GID1*) can combine with the *DELLA* protein to form a GA-*GID1*-*DELLA* trimer, and then the *SKP1-CUL1-F-box* (*SCF*) polymer can tag the trimer to induce the ubiquitin 26S proteasome to degrade the *DELLA* protein, relieving the inhibitory effect of the *DELLA* protein on plant growth and producing gibberellin effects^{48,49}. In this study, *GID1* was highly expressed in the predifferentiation period (F_1), and its expression was downregulated in the initial differentiation period (F_2). In addition, the expression of *GID1* was higher in flower buds (F_2) than in leaf buds (JRL).

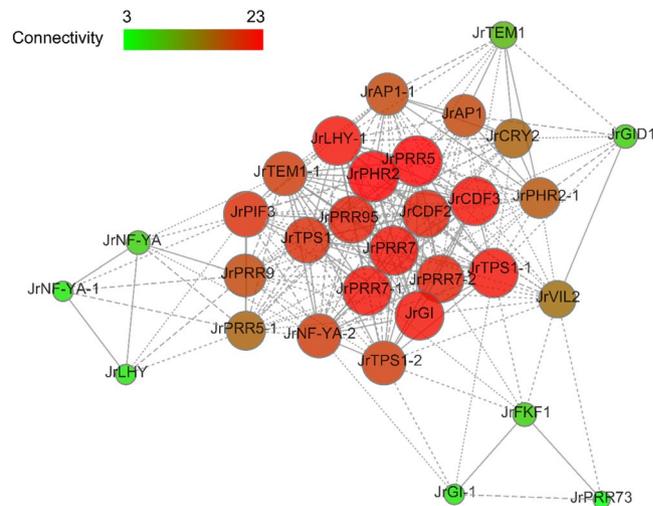


Figure 11. Coexpression networks of 31 DEGs related to flowering time. In the drawn weight network graph, the weight between genes is divided into four parts, which are represented by point lines, short dotted lines, long dotted lines and solid lines from light to heavy weights. Larger nodes and redder colors indicate greater connectivity of the genes in the network graph. The coexpression networks were drawn on the OmicShare website (<http://www.omicshare.com/tools/Home/>) based on the correlation coefficients between the genes.

In addition to the DEGs mentioned above, there are two DEGs annotated as *APETALA1* (*AP1*). In *Arabidopsis thaliana*, *AP1* is required for the floral transition⁵⁰. In this study, the two *AP1* genes were nearly undetectable in the predifferentiation period (F_1), and their expression showed a significant upregulation in the initial differentiation period (F_2).

Coexpression networks. Weighted gene coexpression network analysis (WGCNA) is a biology method for interaction analysis and resolving correlation networks⁵¹. To search for the genes involved in flowering time regulation in walnut, thirty-one flowering time-related DEGs were used to construct a coexpression network using the WGCNA method, and the results are presented in Fig. 11. In the coexpression network, many of the hub genes that participate in flowering time regulation were identified, such as *JrCDF-2*, *JrCDF-3*, *JrPRR7*, *JrPRR7-1*, *JrPRR7-2*, *JrTPS-1-1*, and the hub genes with the highest edge numbers were *JrCRY2* and *JrNF-YA-2*.

Verification of DEG expression by qRT-PCR. To further verify the gene expression levels shown by RNA-Seq, we chose eighteen flowering time-related DEGs to perform qRT-PCR. Except for *JrTPS1-2*, the expression trends of the DEGs showed high similarity between the qPCR data and the RNA-Seq data (Fig. 12).

Discussion

Identification of the critical periods of female flower bud differentiation in walnut. Little is known about the development of walnut female flower buds, and initial differentiation is usually identified by a duration of one month after the female flowers bloom or four to six weeks after the medium and short twigs finish elongating. It is essential to ensure that experimental materials can be collected at precise times; however, due to regional differences in phenology, previous research conclusions cannot be used directly. In this experiment, we have preliminarily identified the initiation period of female flower bud differentiation in walnut through two years of observation of tissue sections, external morphological characteristics and local phenology. Generally, the stage of floral differentiation was divided into the physiological differentiation stage and morphological differentiation stage. Although we identified the initiation period of the morphological differentiation stage, the initiation period of the physiological differentiation stage and the DEGs involved in floral bud determination are still unknown in walnut and will be investigated in a future study. In addition, we suggest that some simple experimental methods, such as freehand sectioning or counting the number of bud scales can be used as preliminary judgment methods in field conditions (Fig. S1).

Functional classification of genes. To functionally categorize the unigenes, the GO and KEGG databases were used to annotate the functions of the unigenes. GO enrichment showed that single organism process (GO:0044699, level 2, BP) was enriched, which included the GO subterm GO:0048449 (level 4) with the function of floral organ formation. The GO analysis of the DEGs indicated that the regulation of RNA metabolism, nucleic acid binding transcription factor activity and transcription factor activity were important during female flower bud differentiation. Interestingly, the GO terms of the DEGs between the flower buds (F_2) and leaf buds (JRL) were mainly enriched in functions related to photosynthesis. In total, 14 DEGs were associated with GO terms related to photosynthesis between F_2 and JRL, and the expression of these 14 DEGs were all upregulated in the leaf buds (Table S3), which suggested that the related genes upregulated in the leaf buds may participate in the process of carbohydrate assimilation. KEGG pathway analysis indicated that most pathways were associated with flowering and were shared by the four combinations of F_1vsF_2, F_1vsF_3, F_2vsF_3 and F_2vsJRL.

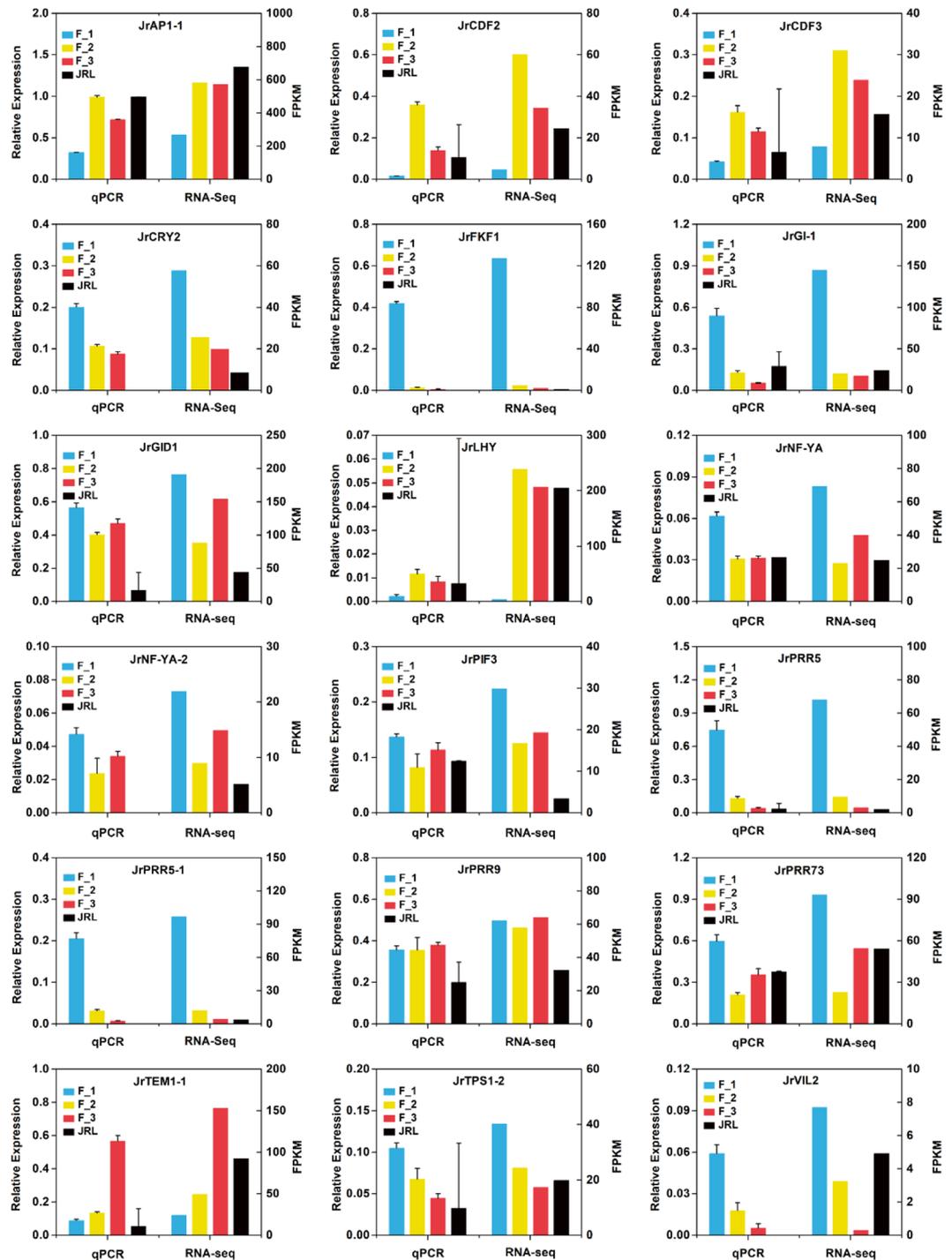


Figure 12. Verification of 18 flowering time-related DEGs by qRT-PCR.

Flowering pathways in walnut during the floral transition. Environmental cues (photoperiod, vernalization) and internal cues (autonomous, GA and age) result in the floral transition³. Most of the flowering time-related DEGs were enriched in the photoperiod pathway, and none of the DEGs were involved in the autonomous pathway or ambient temperature pathway. In the photoperiod pathway, the circadian rhythm genes, including *CCA1*, *FKF1*, *GI*, and *LHY*, can activate the photoperiodic hub gene *CO*. *CO* can activate *SOCI*, which can induce the expression of meristem identity genes, such as *LFY* and *API*, to activate the floral transition.

Interestingly, in the dimer of GI-FKF1 and trimer of CCA1-LHY-PIF3, the expression levels of the components in each complex were not consistent, such as *GI*, which showed a contrasting expression pattern to *FKF1*, as did *LHY* and *PIF3*. In addition, *PHR2* genes are annotated as blue-light photoreceptors, and *CRY2* genes also function as blue-light receptors. They are flavoproteins similar in sequence and repressors of the CLOCK/BMAL1 heterodimer^{29,52,53}. However, they showed contrasting expression patterns for reasons that remain unclear.

In the development of the floral transition, *PRRs*, *GI*, and *FKF1* were all downregulated from F_1 to F_2, while *LHY* and *CDFs* were upregulated from F_1 to F_2, suggesting that photoperiod pathway (circadian rhythm) genes may participate in the regulation of the floral transition.

In this study, the three stages of the female floral transition were identified by the morphological characteristics of tissue sections, and transcriptome-wide investigation of the gene expression profiles in walnut flower buds and leaf buds was conducted during the floral transition. Thirty-one DEGs related to flowering time were identified, and among them, *CRY2* and *NF-YA* were screened as core DEGs in flowering time regulation. Eighty-eight of the thirty-one DEGs (including *CRY2* and *NF-YA*) were confirmed by real-time quantitative analysis, and most of the RNA-Seq data were validated by qPCR data. Our work provides a foundation for further research on the walnut floral transition.

Materials and Methods

Plant materials. Walnut (*Juglans regia* L.) trees were grown under natural conditions in the southern part of the Xinjiang Uyghur Autonomous Region, China. Leaf buds were collected during the floral transition period (JRL), and female flower buds were collected before, during, after the floral transition period (F_1, F_2 and F_3). Each sample was pooled from 3 buds, and 3 biological repeats were performed, for a total of 9 buds for each stage of the floral transition. Three mixed samples from 9 buds were collected for sequencing, and the total RNA of each sample was extracted individually.

Microscope observations. We peeled off the outer scales of the buds and fixed the buds in FAA fixative solution. Then, the fixed buds were dehydrated with a continuous gradient of ethanol and embedded in paraffin. Samples were cut into 8–12 μm slices (Leica Microtome, Germany), deparaffinized with xylene, and hydrated in a decreasing ethanol series. The sections were stained with Safranin and Fast Green and mounted with neutral gum. Finally, we observed the slices under a Motic microscope (Motic AE31, China).

Transcriptome sequencing and library construction. Sequencing was carried out by the Novogene company, Beijing, China. Total RNA was extracted using RNAout 1.0 (Tianenze, Beijing, China). A total of 1.5 μg RNA per sample was used as input material for RNA sample preparation. Sequencing libraries were generated, and the clustering of the index-coded samples was performed. After cluster generation, the library preparations were sequenced on an Illumina HiSeq 2000 platform, and paired-end reads were generated.

Quantitative real-time PCR. Total RNA was extracted using RNAout 1.0 (Tianenze, Beijing, China) by Novogene, Beijing, China. We synthesized the cDNA using the PrimeScript RT Reagent Kit (TaKaRa, Dalian, China). We performed real-time quantification using CFX Manager (Bio-Rad, USA) with SYBR Green Realtime PCR Master Mix (Toyobo, Osaka, Japan). The protocol of the real-time PCR was as follows: an initiation step at 95 °C for 5 min; followed by 40 cycles of 30 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C; and the melting curve step was processed from 65 to 95 °C. Each reaction was repeated for three times. The walnut Actin gene (forward primer: 5'-CCATCCAGGCTGTTCTCTC-3', and reverse primer: 5'-GCAAGGTCAGACGAAGG-3') and walnut GAPDH gene (forward primer: 5'-ATTTGGAATCGTTGAGGGTCTTATG-3' and reverse primer: 5'-AATGATGTTGAAGGAAGCAGCAC-3') were used as the internal controls. The results were analyzed by the $2^{-\Delta\Delta C_t}$ method⁵⁴.

Differential expression analysis. For differential gene expression analysis, the read counts of each sequenced library were adjusted with EdgeR software. Differential expression analysis of two samples was performed using the DEGseq (2010) R package. The P values were adjusted using the q values⁵⁵, and thresholds for significantly differential expression were set as q value < 0.005 and $|\log_2(\text{foldchange})| > 1$.

References

- Pollegioni, P. *et al.* Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions. *PLoS One* **12**, e0172541, <https://doi.org/10.1371/journal.pone.0172541> (2017).
- FAO, 2017, <http://www.fao.org/faostat/en/#data/QC> (2018).
- Khan, M. R., Ai, X. Y. & Zhang, J. Z. Genetic regulation of flowering time in annual and perennial plants. *Wiley Interdiscip Rev RNA* **5**, 347–359, <https://doi.org/10.1002/wrna.1215> (2014).
- Bernier, G., Havelange, A., Houssa, C., Petitjean, A. & Lejeune, P. Physiological Signals That Induce Flowering. *Plant Cell* **5**, 1147–1155, <https://doi.org/10.1105/tpc.5.10.1147> (1993).
- Simpson, G. G. & Dean, C. Environmental-dependent acceleration of a developmental switch: the floral transition. *Science's STKE: signal transduction knowledge environment* **2000**, pe1, <https://doi.org/10.1126/stke.2000.18.pe1> (2000).
- Fornara, F., de Montaigu, A. & Coupland, G. SnapShot: Control of flowering in Arabidopsis. *Cell* **141**(550), 550 e551–552, <https://doi.org/10.1016/j.cell.2010.04.024> (2010).
- Wigge, P. A. *et al.* Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* **309**, 1056–1059, <https://doi.org/10.1126/science.1114358> (2005).
- Hames, C. *et al.* Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *Embo j* **27**, 2628–2637, <https://doi.org/10.1038/emboj.2008.184> (2008).
- Irish, V. F. The flowering of Arabidopsis flower development. *Plant J* **61**, 1014–1028, <https://doi.org/10.1111/j.1365-313X.2009.04065.x> (2010).
- Chandler, J. W. Floral meristem initiation and emergence in plants. *Cell Mol Life Sci* **69**, 3807–3818, <https://doi.org/10.1007/s00018-012-0999-0> (2012).
- Pose, D., Yant, L. & Schmid, M. The end of innocence: flowering networks explode in complexity. *Curr Opin Plant Biol* **15**, 45–50, <https://doi.org/10.1016/j.pbi.2011.09.002> (2012).
- Torti, S. & Fornara, F. AGL24 acts in concert with SOC1 and FUL during Arabidopsis floral transition. *Plant Signal Behav* **7**, 1251–1254, <https://doi.org/10.4161/psb.21552> (2012).
- Srikanth, A. & Schmid, M. Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci* **68**, 2013–2037, <https://doi.org/10.1007/s00018-011-0673-y> (2011).

14. Kardailsky, I. *et al.* Activation tagging of the floral inducer FT. *Science* **286**, 1962–1965 (1999).
15. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
16. Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056, <https://doi.org/10.1126/science.1115983> (2005).
17. Corbesier, L. *et al.* FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* **316**, 1030–1033, <https://doi.org/10.1126/science.1141752> (2007).
18. Moyroud, E., Kusters, E., Monniaux, M., Koes, R. & Parcy, F. LEAFY blossoms. *Trends in plant science* **15**, 346–352, <https://doi.org/10.1016/j.tplants.2010.03.007> (2010).
19. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327–335 (1996).
20. Nikolov, D. B. & Burley, S. K. RNA polymerase II transcription initiation: A structural view. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 15–22 (1997).
21. Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annual review of genetics* **34**, 77–137, <https://doi.org/10.1146/annurev.genet.34.1.77> (2000).
22. Brambilla, V. *et al.* Antagonistic Transcription Factor Complexes Modulate the Floral Transition in Rice. *Plant Cell* **29**, 2801–2816, <https://doi.org/10.1105/tpc.17.00645> (2017).
23. Goslin, K. *et al.* Transcription Factor Interplay between LEAFY and APETALA1/CAULIFLOWER during Floral Initiation. *Plant Physiol* **174**, 1097–1109, <https://doi.org/10.1104/pp.17.00098> (2017).
24. Smaczniak, C. *et al.* Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci USA* **109**, 1560–1565, <https://doi.org/10.1073/pnas.1112871109> (2012).
25. Simpson, G. G. & Dean, C. Arabidopsis, the Rosetta stone of flowering time? *Science* **296**, 285–289, <https://doi.org/10.1126/science.296.5566.285> (2002).
26. Fornara, F. *et al.* Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell* **17**, 75–86, <https://doi.org/10.1016/j.devcel.2009.06.015> (2009).
27. Imaizumi, T., Schultz, T. F., Harmon, F. G., Ho, L. A. & Kay, S. A. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science* **309**, 293–297, <https://doi.org/10.1126/science.1110586> (2005).
28. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science* **318**, 261–265, <https://doi.org/10.1126/science.1146994> (2007).
29. Kanai, S., Kikuno, R., Toh, H., Ryo, H. & Todo, T. Molecular evolution of the photolyase-blue-light photoreceptor family. *Journal of molecular evolution* **45**, 535–548 (1997).
30. Bagnall, D. J., King, R. W. & Hangarter, R. P. Blue-light promotion of flowering is absent in hy4 mutants of Arabidopsis. *Planta* **200**, 278–280 (1996).
31. El-Din El-Assal, S. *et al.* The role of cryptochrome 2 in flowering in Arabidopsis. *Plant Physiol* **133**, 1504–1516, <https://doi.org/10.1104/pp.103.029819> (2003).
32. Siriwardana, C. L. *et al.* NUCLEAR FACTOR Y, Subunit A (NF-YA) Proteins Positively Regulate Flowering and Act Through FLOWERING LOCUS T. *PLoS Genet* **12**, e1006496, <https://doi.org/10.1371/journal.pgen.1006496> (2016).
33. Viczian, A. *et al.* Functional characterization of phytochrome interacting factor 3 for the Arabidopsis thaliana circadian clockwork. *Plant & cell physiology* **46**, 1591–1602, <https://doi.org/10.1093/pcp/pci175> (2005).
34. Nakamichi, N. *et al.* PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. *Plant Cell* **22**, 594–605, <https://doi.org/10.1105/tpc.109.072892> (2010).
35. Castillejo, C. & Pelaz, S. The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. *Curr Biol* **18**, 1338–1343, <https://doi.org/10.1016/j.cub.2008.07.075> (2008).
36. van Dijken, A. J., Schluepmann, H. & Smeekens, S. C. Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. *Plant Physiol* **135**, 969–977, <https://doi.org/10.1104/pp.104.039743> (2004).
37. Wahl, V. *et al.* Regulation of flowering by trehalose-6-phosphate signaling in Arabidopsis thaliana. *Science* **339**, 704–707, <https://doi.org/10.1126/science.1230406> (2013).
38. Chandler, J., Wilson, A. & Dean, C. Arabidopsis mutants showing an altered response to vernalization. *Plant J* **10**, 637–644 (1996).
39. Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell* **107**, 525–535 (2001).
40. Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* **297**, 243–246, <https://doi.org/10.1126/science.1072147> (2002).
41. Sung, S. & Amasino, R. M. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. *Nature* **427**, 159–164, <https://doi.org/10.1038/nature02195> (2004).
42. Greb, T. *et al.* The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Curr Biol* **17**, 73–78, <https://doi.org/10.1016/j.cub.2006.11.052> (2007).
43. Ratcliffe, O. J., Nadzan, G. C., Reuber, T. L. & Riechmann, J. L. Regulation of flowering in Arabidopsis by an FLC homologue. *Plant Physiol* **126**, 122–132 (2001).
44. Kim, D. H. & Sung, S. The Plant Homeo Domain finger protein, VIN3-LIKE 2, is necessary for photoperiod-mediated epigenetic regulation of the floral repressor, MAF5. *Proc Natl Acad Sci USA* **107**, 17029–17034, <https://doi.org/10.1073/pnas.1010834107> (2010).
45. Blazquez, M. A. & Weigel, D. Integration of floral inductive signals in Arabidopsis. *Nature* **404**, 889–892, <https://doi.org/10.1038/35009125> (2000).
46. Eriksson, S., Bohlenius, H., Moritz, T. & Nilsson, O. GA4 is the active gibberellin in the regulation of LEAFY transcription and Arabidopsis floral initiation. *Plant Cell* **18**, 2172–2181, <https://doi.org/10.1105/tpc.106.042317> (2006).
47. Alonso-Peral, M. M. *et al.* The microRNA159-regulated GAMBY-like genes inhibit growth and promote programmed cell death in Arabidopsis. *Plant Physiol* **154**, 757–771, <https://doi.org/10.1104/pp.110.160630> (2010).
48. Griffiths, J. *et al.* Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. *Plant Cell* **18**, 3399–3414, <https://doi.org/10.1105/tpc.106.047415> (2006).
49. Harberd, N. P., Belfield, E. & Yasumura, Y. The angiosperm gibberellin-GID1-DELLA growth regulatory mechanism: how an “inhibitor of an inhibitor” enables flexible response to fluctuating environments. *Plant Cell* **21**, 1328–1339, <https://doi.org/10.1105/tpc.109.066969> (2009).
50. Mandel, M. A., Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature* **360**, 273–277, <https://doi.org/10.1038/360273a0> (1992).
51. Presson, A. P. *et al.* Integrated Weighted Gene Co-expression Network Analysis with an Application to Chronic Fatigue Syndrome. *BMC Systems Biology* **2**, 95–95, <https://doi.org/10.1186/1752-0509-2-95> (2008).
52. Biernat, M. A. *et al.* A baculovirus photolyase with DNA repair activity and circadian clock regulatory function. *J Biol Rhythms* **27**, 3–11, <https://doi.org/10.1177/0748730411429665> (2012).
53. Cashmore, A. R., Jarillo, J. A., Wu, Y. J. & Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284**, 760–765 (1999).
54. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods (San Diego, Calif.)* **25**, 402–408, <https://doi.org/10.1006/meth.2001.1262> (2001).
55. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440–9445, <https://doi.org/10.1073/pnas.1530509100> (2003).

Acknowledgements

We thank Dr. Gale (Shihezi University) for careful editing of this manuscript. This work was supported by the important National Science and Technology Specific projects of Xinjiang (No. 201130102-1-4) and the National Natural Science Foundation of China (No. 30560090).

Author Contributions

J.X.N. led and coordinated the project. J.X.N. and S.W.Q. designed the study. S.W.Q., L.Z., H.X., L.M. and Y.Q. collected the plant materials and isolated the RNA. S.W.Q. and L.Z. conducted real-time quantitative P.C.R. S.W.Q. conducted the bioinformatics analysis and wrote the paper. All authors have read and agree with the final manuscript. J.X.N. is the corresponding author and is responsible for all contact and correspondence.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-43582-z>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019