



OPEN

# Alternative splicing of flowering time gene *FT* is associated with halving of time to flowering in coconut

Wei Xia<sup>2,5</sup>, Rui Liu<sup>1,5</sup>, Jun Zhang<sup>1,5</sup>, Annaliese S. Mason<sup>3</sup>, Zhiying Li<sup>1</sup>, Shufang Gong<sup>1</sup>, Yazhu Zhong<sup>1</sup>, Yajing Dou<sup>1</sup>, Xiwei Sun<sup>1</sup>, Haikuo Fan<sup>1</sup> & Yong Xiao<sup>1,4</sup>

Coconut palm has two distinct types—“tall” and “dwarf”—which differ morphologically. Tall coconut varieties need 8–10 years to start flowering, while dwarf coconut varieties only require 3–5 years. We compared seedling and reproductive stage transcriptomes for both coconut types to determine potential molecular mechanisms underlying control of flowering time in coconut. Several key genes in the photoperiod pathway were differentially expressed between seedling and reproductive leaf samples in both tall and dwarf coconut. These genes included *suppressor of overexpression of constans* (*SOC1*), *flowering locus T* (*FT*), and *Apetala 1* (*AP1*). Alternative splicing analysis of genes in the photoperiod pathway further revealed that the *FT* gene produces different transcripts in tall compared to dwarf coconut. The shorter alternative splice variant of *FT* [which included a 6 bp deletion, alternative 3' splicing sites (A3SS)] was found to be exclusively present in dwarf coconut varieties but absent in most tall coconut varieties. Our results provide a valuable information resource as well as suggesting a probable mechanism for differentiation of flowering time onset in coconut, providing a target for future breeding work in accelerating time to flowering in this crop species.

Coconut palm (*Cocos nucifera* L., 2n = 32) belongs to the monocotyledonous family Arecaceae (Palmaceae) and is an important tropical oil and fruit crop. Presently, coconut palm is cultivated in 93 countries, including Central and South America, East and West Africa, Southeast Asia and the Pacific island, where it accounts for over 12 million hectares<sup>1</sup>. Coconut palm is divided into “tall” and “dwarf” types based on distinct morphological and breeding features between the two groups. Tall coconuts take 8–10 years to complete the vegetative to reproductive transition, can grow to a height of about 20–30 m, and have medium-to-large nuts. Tall coconuts are hardy and can adapt to a wide range of environmental conditions. Dwarf coconuts are early-flowering (4–6 years after planting), can grow to a height of about 10–15 m, and can produce a number of small nuts<sup>2</sup>. The large difference in flowering time between these two types is of particular interest for breeding and production: flowering time limits not only generation speed but also time to first harvest after planting, such that shorter flowering times are almost always desirable. However, to date little is known about regulation of flowering time in coconut.

In the last several decades, a great deal of work has been done in elucidating the molecular mechanisms underlying the transformation from vegetative growth to reproductive growth in plants. Four flowering regulation pathways have been described in model plant *Arabidopsis thaliana*: the photoperiod pathway<sup>3</sup>, the vernalization pathway<sup>4</sup>, the gibberellin pathway<sup>5</sup> and the autonomous pathway<sup>6</sup>. Time to flowering in *Arabidopsis thaliana* is predominantly determined by expression of the *Flowering Locus T* (*FT*) gene<sup>7</sup>. In *Arabidopsis thaliana* *FT* is negatively regulated by expression of *Flowering Locus C* (*FLC*), which is a transcriptional repressor, and positively regulated by expression of *Constans* (*CO*), which is a transcriptional activator<sup>7</sup>. Overexpression of *CO*

<sup>1</sup>National Engineering Research Center of Coconut/Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, People's Republic of China. <sup>2</sup>College of Tropical Crops, Hainan University, Haikou, People's Republic of China. <sup>3</sup>Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. <sup>4</sup>Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Haikou, People's Republic of China. <sup>5</sup>These authors contributed equally: Wei Xia, Rui Liu and Jun Zhang. ✉email: vanheco@163.com; xiaoyong1980@catas.cn

Gene ID	Forward primer	Reverse primer
AP1_CCG008583.1	AGGGAAGACGGAGGTGAGAC	GGGAGAGAAGACGATGAGGG
AP1_CCG008584.1	GGGGAAGACGGAGGTGAGACG	CCCTTGGGGGAGAAGACGATG
AP1_CCG022175.1	CCAGGTCTCCGTCGTCATCT	CTCTCGTGCTTGGCATCCC
AP1_CCG027514.1	AGGGGAAGACGGAGATGAA	GAGAAGACGATGACGGCGAC
CO_CCG005523.2	CGAAGGGAGGGTTATTGGGG	GTCGCAGGTGACGCAGAGC
FT_CCG009539.1	TTCATCAGGTCCGCATTCTT	AGGGTCCACCATCACGAGC
FT_CCG011565.1	GTGGTAGGGAGGGTCATAGGC	GGACAGTTGGAGGGTTGATAGC

**Table 1.** Primer sequence for real-time qPCR.

up-regulates *FT* and causes an early flowering phenotype<sup>8</sup>, as does direct overexpression of *FT*<sup>9,10</sup> *CO* is itself negatively regulated by *Suppressor of Overexpression of Constans 1 (SOC1)* and *PhyB*<sup>11</sup>, the latter of which is a photoreceptor<sup>12</sup>. *PhyB* loss of function mutants also therefore result in early flowering<sup>12</sup>. Flowering time phenotypes have also been observed as a result of mutation of other photoreceptor genes, such as *CRY2*<sup>13</sup>. In total at least 29 major genes are involved in regulation of flowering time in *Arabidopsis thaliana*, including *Flowering Locus D (FLD)* and *Leafy (LFY)*<sup>14–18</sup>.

The function of *FT* has also been well studied in many species other than model plant *Arabidopsis thaliana*. In rice, a total of thirteen *FT*-like homologs were identified, of which three *FT*-like genes can promote flowering (*OsFTL1-OsFTL3*)<sup>18–21</sup>. Of these genes, *OsFTL2* was found to be a mobile signal that moves from the leaf to the shoot apical meristem to induce flowering<sup>22,23</sup>. In barley, a total of five *FT*-like genes were found, of which *HvFT1* was highly expressed under long-day conditions in transition from vegetative to reproductive growth, *HvFT3* was identified as a candidate gene under a major flowering-time QTL, *HvFT2* and *HvFT4* were highly expressed at later developmental stages and *HvFT5* showed low expression level under short-day conditions<sup>24</sup>, highlighting the possible complexity of regulatory mechanisms involving multiple *FT* copies. In rapeseed, which has a history of polyploidy, 164 gene copies were identified from 29 core *Arabidopsis* flowering time gene homologs, an average of 4.7 copies per *Arabidopsis* gene<sup>25</sup>; 101 genome regions in this species were significantly associated with flowering time<sup>26</sup> and 12 haplotypes differentiated winter and spring-flowering rapeseed<sup>7,27</sup>. By contrast, although a total of four *FT*-like genes were isolated in poplar, only one (*PtFT2*) was shown to shorten juvenile phase and promote seasonal flowering<sup>28</sup>.

In coconut palm, we have as yet no information about the molecular basis of flowering time control, despite the importance of the flowering time trait in this long-lived agricultural tree species. In the present study, we used comparative transcriptomics to elucidate the molecular mechanisms underlying the differences in flowering time observed between tall and dwarf coconut types. Coconut orthologs of 174 *Arabidopsis* genes belonging to the flowering-time pathway were identified and characterized for expression in different tissues from seedling and reproductive stages. Among these genes, we detected and validated alternative splicing of *CnFT* transcripts as the putative causal mechanism for the observed flowering time difference between tall and dwarf coconut types.

## Materials and methods

**Plant materials and RNA isolation.** The leaf tissues of seedling and mature stage of two coconut plants were used for RNA-seq: 10-year-old Yellow Dwarf (YD) individuals; 1-year-old YD seedlings; 10-year-old Hainan Tall (HT) individuals and 1-year-old HT seedlings. Each sample included a mixture of two biological replicates. The eight coconut individuals were growing in the coconut germplasm resource of the Coconut Research Institute of the Chinese Academy of Tropical Agricultural Sciences. The average temperature and humidity in this location are approximately 23.9 °C and 89%. Total mRNA was extracted from spear leaves of these coconut individuals according to the MRIP method<sup>29</sup>.

**Illumina sequencing.** Purified mRNA was separately fragmented with divalent cations under increased temperature. These short fragments were taken as templates to synthesize the first-strand cDNA using hexamer primers and superscript III (Invitrogen, Carlsbad, CA, USA). Second-strand cDNA was then synthesized in a solution containing buffer, dNTP, RNaseH and DNA polymerase I and subsequently purified using a QiaQuick PCR extraction kit (Qiagen). EB buffer was used to resolve these short fragments for end reparation and poly (A) addition. The sequence adaptors were linked to two ends of short cDNA sequences and suitably sized cDNA fragments were selected out for PCR amplification based on the agarose gel electrophoresis results. Finally, the library established was sequenced using an Illumina HiSeq 200. The paired-end library was developed according to the paired-end sample Preparation kit protocol (Illumina, USA).

**Real-time qPCR assays.** Real-time qPCR was performed following a standard SYBR Premix Ex Taq Kit (TaKaRa) protocol in 96-well optical plates (Axygen) using a volume of 10 µl. The primer pairs were designed by using Snapgene Viewer 1.0 software (Table 1). The reactions were incubated in 0.2 ml tubes of a Mastercycler Ep Realplex4 (Eppendorf) machine as follows: 95 °C for 5 s, 55 °C for 15 s and 68 °C for 20 s. The procedure ended by a melt-curve ramping from 60 to 95 °C for 20 min to check the PCR specificity. All qPCR reactions were carried out in biological and technical triplicates. The final Ct values were the means of nine values. The comparative expression levels of the selected expressed sequences were normalized to that of *actin*.

**Calculating gene expression level.** RPKM (Reads Per kb per Million reads) was used to calculate gene expression level using the following formula<sup>30</sup>:

$$RPKM = \frac{10^6 C}{NL/10^3}$$

where C is the number of reads that exclusively aligned to one expressed sequence, N is the total number of reads that aligned to all expressed sequences, and L is the basic number in the CDS of the corresponding expressed sequence. The CDS and genome sequence (the first version of coconut genome) is available in the GigaDB database (GigaDB, RRID:SCR 004002, <https://gigadb.org/dataset/100347>).

The statistical significance of the differential expression was determined according to the method documented by Audic and Claveries<sup>31</sup>. The statistical significance  $p(x)$  can fit the Poisson distribution

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Arabic alphabet x in the formula represents number of reads mapped to each unigene.  $\lambda$  refers to the number of real transcripts of the unigene.

When thousands of hypothesis tests are performed, the p-value suitable for a single test is not sufficient to guarantee a low rate of false discovery. Thus, a false discovery rate (FDR) control method was applied using multiple hypothesis testing to correct the p-value results<sup>32</sup>. Subsequently, the RPKM ratio was used to compute the fold change of gene expression for each pair of samples simultaneously. The differentially expressed genes were selected using a threshold of  $FDR \leq 0.001$  and an absolute value of  $\log_2^{\text{ratio}} \geq 1$ <sup>33</sup>.

*The functional annotation of differential expressed genes.* The predicted genes were used for BLAST and annotation against NCBI non-redundant sequence database (Nr) at a cut-off value of  $10^{-5}$ . The differential expressed genes were also aligned by BLASTX to protein databases such as GO, KEGG, Swissprot, and InterPro database (transcription factor), and PRG database (PRG). The Blast2GO program was used to obtain GO annotations for differential expressed genes. The WEGO software was then used to perform GO functional classification of differential expression genes. Meanwhile, Blastall software was used to align differential expressed genes against KEGG (Kyoto Encyclopedia of Genes and Genomes) database. KEGG enrichment was performed by using phyper of R package. Enrichment of KEGG pathway were performed at a cut-off FDR of 0.01.

**Detection of alternative splicing.** We used rMATS 4.0.1 software to detect alternative splicing events based on RNA-seq data and coconut genome sequence (GigaDB, RRID:SCR 004002, <https://gigadb.org/dataset/100347>)<sup>34</sup>. The p-value and False Discovery Rate (FDR) were used to determine the isoform ratio of one gene at a cut-off PDR of 0.05. Detailed information is shown in Fig. 1. To exclude the influence of insertion/deletion variation in AS detection, we used STAR to map RNA-seq clean reads and GATK4 to call indel variations. The detected AS events that overlapped with transcript indel variations were filtered out for further analysis.

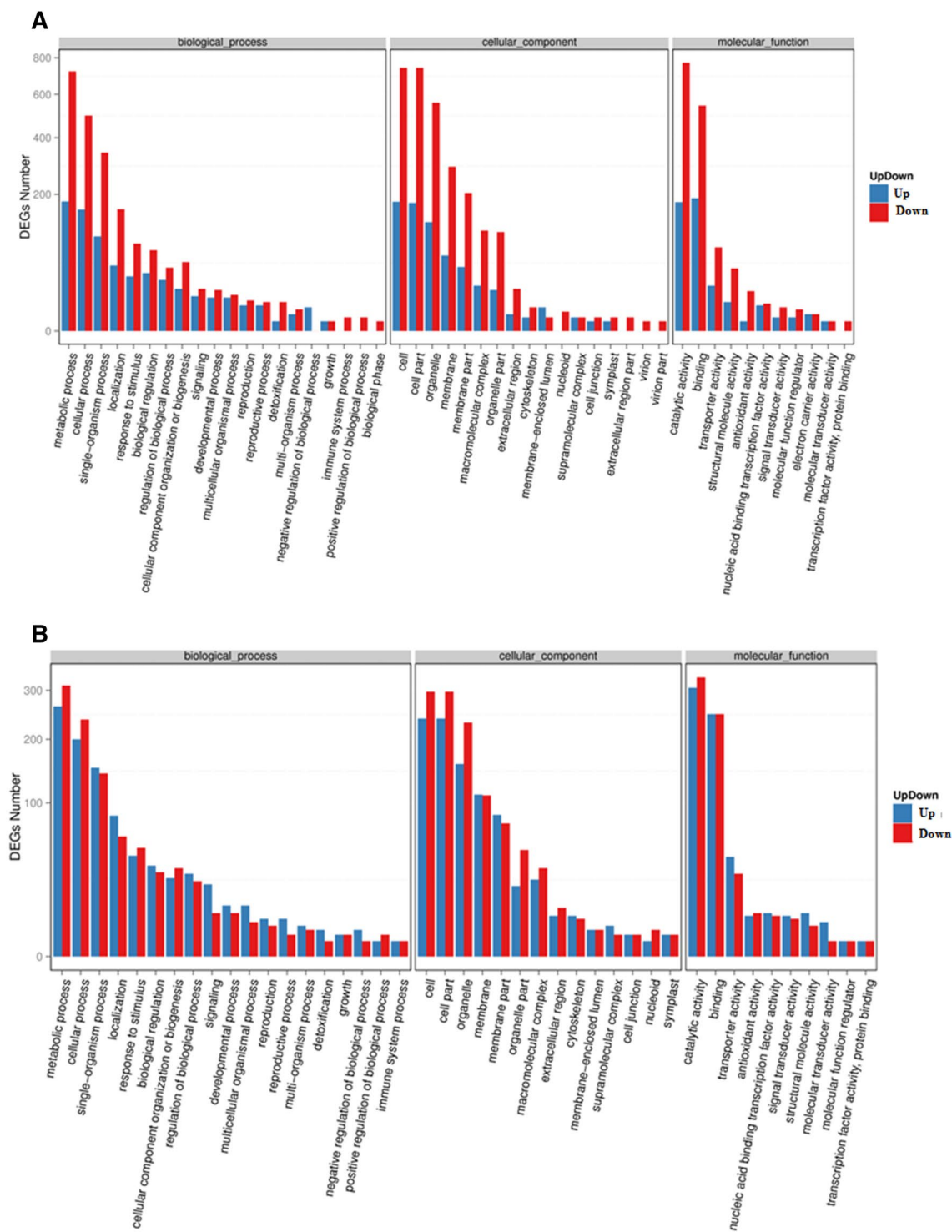
**PCR amplification and sequencing of 20 coconut individuals.** Primers were designed according to the exon sequences flanking alternative splicing junctions using Snapgene Viewer 1.0 software. The primer sequences were listed in Table 2. PCR amplifications were performed in 50  $\mu$ l reaction mixtures containing a 500 ng cDNA sample from the oil palm mesocarp, 1  $\times$  PCR buffer, 2 mM  $MgCl_2$ , 5 U of TaqDNA polymerase (TaKaRa, China), 0.5  $\mu$ M of each primer, and 0.2 mM dNTP mix. The PCR program included denaturation at 98  $^\circ$ C for 30 s; followed by 30 cycles at 98  $^\circ$ C for 10 s, 55  $^\circ$ C for 30 s, and 72  $^\circ$ C for 60 s; and a final extension at 72  $^\circ$ C for 5 min. The PCR products were electrophoretically visualized on a 1% agarose gel.

## Results

**Transcriptome sequencing for tall and dwarf coconut.** Approximately 6.6 Gb of unmapped clean reads were generated for each of four samples: tall coconut seedlings, tall coconut mature trees, dwarf coconut seedlings and dwarf coconut mature trees (Table 3). All clean reads have been deposited in the European Nucleotide Archive (Submission Number: PRJEB34852). After mapping all clean reads to the coconut reference genome using the software HISAT, a total of 36,558 transcripts were predicted. Among them, a total of 7,037 transcripts were predicted to be non-coding RNA. The remaining 29,521 were identified as protein-coding transcripts, 25,114 (85%) of which were consistent with previously known proteins predicted based on the coconut genome and the remaining 4,407 of which were novel protein-coding transcripts.

**Analysis and functional annotation of differential expression genes between tall and dwarf coconut.** Global expression changes were identified between seedlings and mature trees of tall and dwarf coconut using the software program PossionDis. Compared to seedling period of tall coconut, 1729 transcripts were up-regulated and 4,150 were down-regulated ( $abs(\log_2^{\text{Foldchange}}) < 1$  or  $FDR > 0.001$ ) in mature period of tall coconut. Meanwhile, compare to seedling period of dwarf coconut, 1964 transcripts were up-regulated and 2009 were down-regulated ( $abs(\log_2^{\text{Foldchange}}) < 1$  or  $FDR > 0.001$ ) in seedling period of dwarf coconut.

In order to annotate differentially expressed unigenes, unigene sequences were aligned against the NCBI non-redundant (Nr) protein database using a cut-off E-value of  $10^{-5}$  (Fig. 2). A large proportion of differentially expressed unigenes were assigned into the metabolic process (tall coconut: 902 unigenes and dwarf coconut: 576 unigenes) and the cellular process (tall coconut: 655 unigenes and dwarf coconut: 438 unigenes) categories



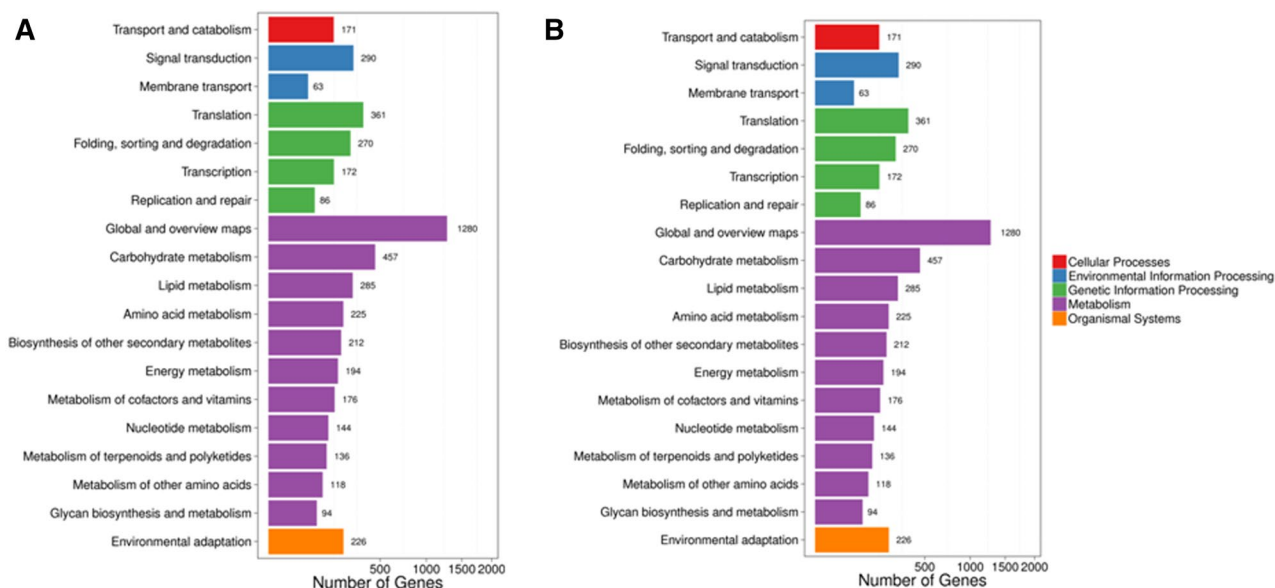
**Figure 1.** Gene ontology annotation of differentially expressed genes between seedling and mature trees of (A) tall coconut and (B) dwarf coconut.

Primer name	Primer sequence
FT_AS_475-F	GTGGTAGGGAGGGTCATAGGC
FT_AS_475-R	GGGCATTGAAGTAGACCGCAG

**Table 2.** Primer sequences of FT gene for alternative splice site.

	Clean reads (Gb)	Number of reads (million)	Q20 (%)
<b>Tall coconut</b>			
Seedling	6.59	65.94	97.58
Mature	6.57	65.74	97.52
<b>Dwarf coconut</b>			
Seedling	6.63	66.34	97.62
Mature	6.68	66.75	97.5

**Table 3.** Transcriptome raw data obtained from seedlings and mature tall and dwarf coconut trees.



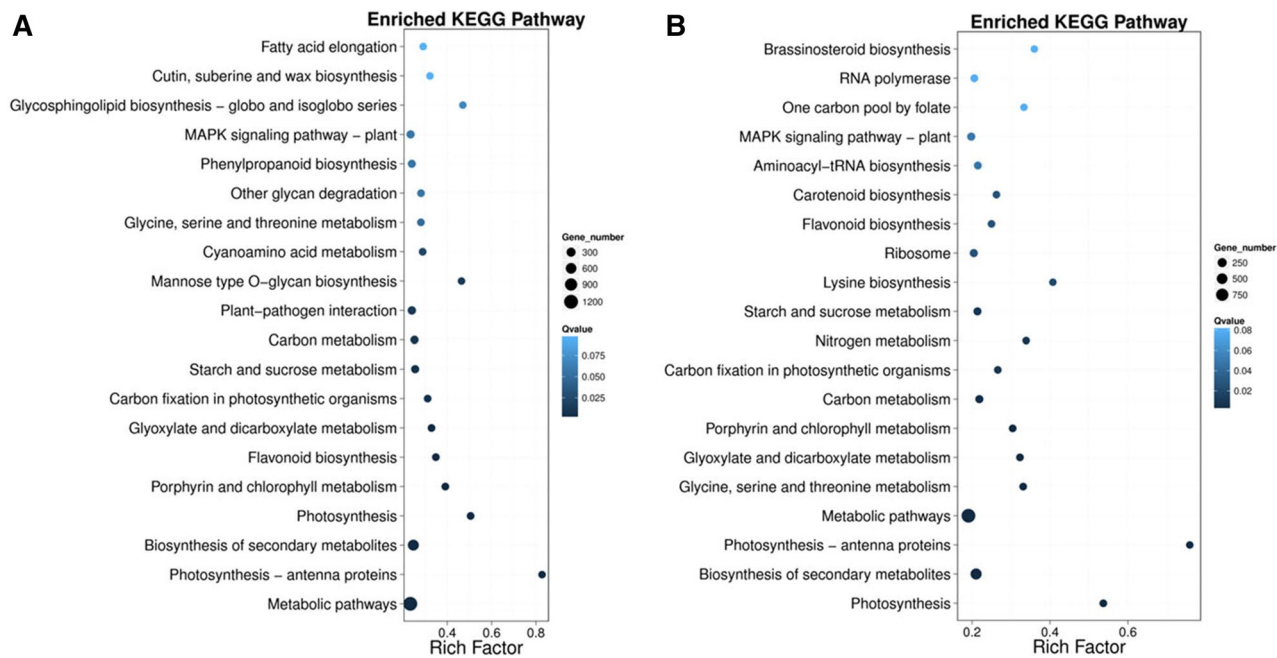
**Figure 2.** KEGG annotation of differentially expressed genes between seedling and mature trees of (A) tall coconut and (B) dwarf coconut.

for the “biological process” categories, cell (tall coconut: 921 unigenes and dwarf coconut: 537 unigenes) and cell part (tall coconut: 918 unigenes and dwarf coconut: 537 unigenes) in the “cellular component” categories, and catalytic activity (tall coconut: 949 unigenes and dwarf coconut: 636 unigenes) and binding (tall coconut: 733 unigenes and dwarf coconut: 498 unigenes) in the molecular function categories. Only down-regulated unigenes were detected in some functional categories in mature trees of dwarf coconut compared to dwarf coconut seedlings, including immune system processes, positive regulation of biological processes and biological phases under the biological processes categories, membrane-enclosed lumen, extracellular regions, and virion in the cellular component categories, and transcription factor activity and protein binding in the molecular function categories. By contrast, in each of these functional categories, both up-regulated and down-regulated unigenes were detected in mature trees of tall coconut compared to tall coconut seedlings.

In order to further identify the active biochemical pathways of differentially expressed unigenes, we mapped these differential expressed genes ( $abs(\log_2^{\text{Foldchange}}) < 1$  or  $FDR > 0.001$ ) to the reference canonical pathways in the Kyoto encyclopedia of genes and genomes (KEGG). A total of 2,580 differential expressed genes between seedling and mature stage of tall coconut were mapping the 115 KEGG pathways (Fig. 2). Moreover, in dwarf coconut, 1,704 differential expressed genes were mapped onto 113 KEGG pathways. Enrichment results for KEGG pathways showed that the highest enrichment factors were detected in the photosynthesis-antenna protein pathway in dwarf coconut, followed by the photosynthesis pathway (Fig. 3). Meanwhile, in tall coconut, the highest enrichment factors were detected in the photosynthesis-antenna protein pathway, followed by Mannose type O-glycan biosynthesis.

Meanwhile, Transcription factors were also predicted by aligning differential expressed genes with COG database. A total of 685 differential expressed genes in tall coconut were identified and mapped onto 46 gene families. While, in dwarf coconut, 509 differential expressed genes were identified to be transcription factors and mapped into 45 gene families. Meanwhile, differential expressed genes were also blasted against PRG database. A total of 340 differential expressed genes were identified to be resistant genes in tall coconut. Moreover, in dwarf coconut, 208 differential genes were identified to resistant genes.





**Figure 3.** Enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway of differentially expressed genes between seedlings and mature trees in (A) tall coconut and (B) dwarf coconut.

### Identification of candidate genes involved in flowering-time pathways and their expression profiles.

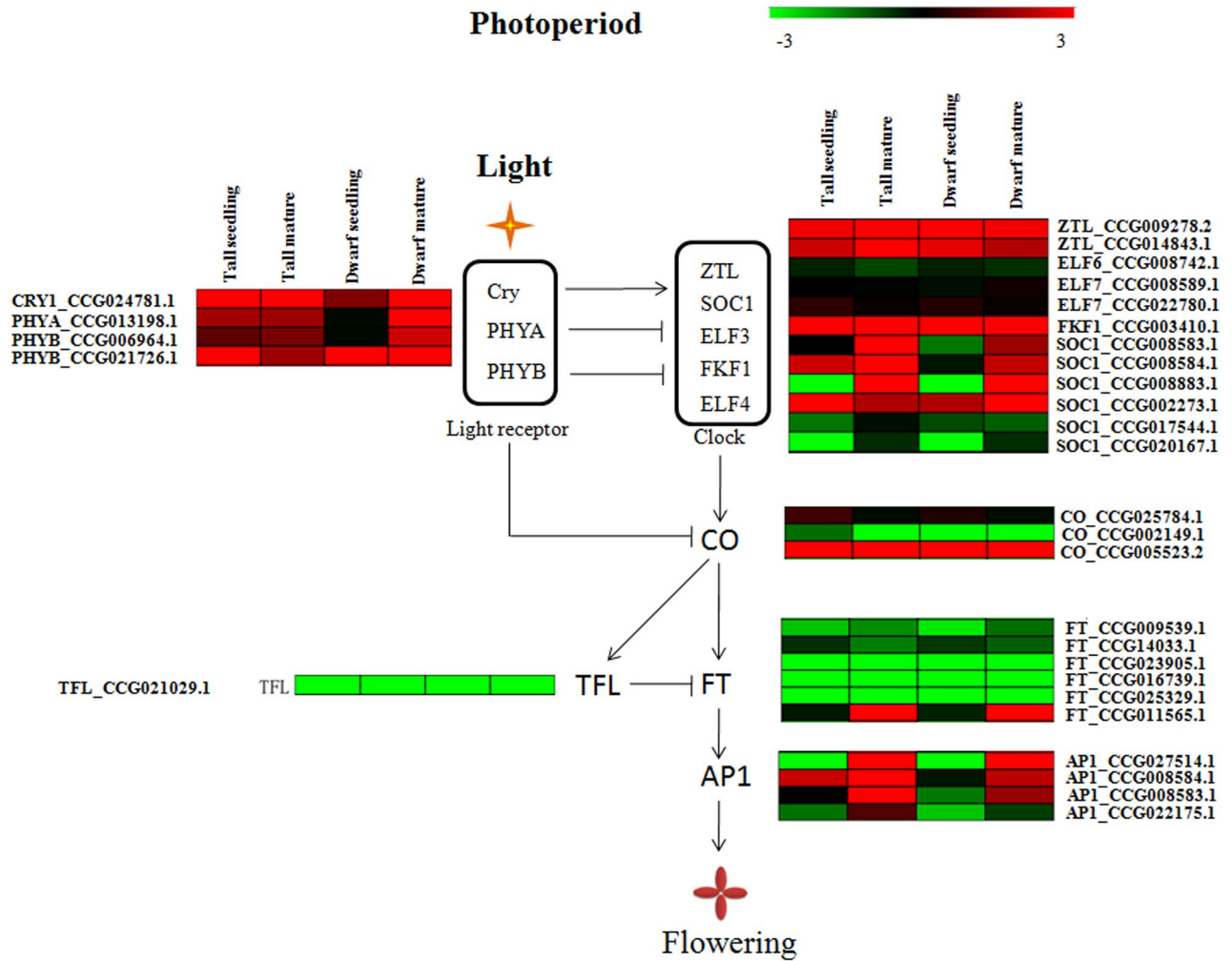
In order to identify candidate genes involving in flowering-time pathways, protein sequences of 174 *Arabidopsis* genes involving in flowering time were downloaded from the National Center for Biotechnology information (NCBI). Detailed information for these genes is listed in Supplementary Table 1. Subsequently, these *Arabidopsis* protein sequences were aligned using BLAST with the “tblastn” function against the CDS (coding sequences) database of coconut<sup>35</sup>. A total of 198 genes in coconut were identified as homologues of flowering-time genes in *Arabidopsis*. Detailed information for these coconut genes is listed in Supplementary Table 2. The 198 genes were classified into the four flowering-time pathways [photoperiod pathway, autonomous pathway, vernalization pathway, and gibberellin (GA) pathway] based on their *Arabidopsis* annotations.

From this data we identified four light receptor genes, including one copy of *CRY1* (*cryptochrome 1*), one copy of *PHYA* (*phytochrome A*), and two copies of *PHYB* (*phytochrome B*), all of which showed high expression levels both in seedlings and mature trees of tall and dwarf coconut (Fig. 4) and are known to regulate clock genes. We identified 12 putative clock genes in the coconut genome: two *ZTL* (*zeitlupe*), one *ELF6* (*early flowering 6*), two *ELF7* (*early flowering 7*), one *FKF1* (*flavin binding F-box 1*), and six *SOC1* genes. Among them, three *SOC1* genes (CCG008583, CCG008584, and CCG008883) showed higher expression levels in mature trees compared to seedlings in both tall and dwarf coconut. CCG008883 in particular had FPKM values of 0.22 and 0.34 in tall and dwarf coconut seedlings respectively, but FPKM values of 55.16 and 50.12 in mature trees of tall and dwarf coconut respectively. *SOC1* is known to induce *CO* gene expression in *Arabidopsis*: we also found three *CO* orthologs in coconut (CCG025784, CCG002149, and CCG005523), one of which (CCG005523) was highly expressed at both seedling and reproductive stages of tall and dwarf coconut, while the other *CO* genes had low expression levels at both developmental stages. *CO* genes are known up-regulators of the *Flowering Locus T (FT)* gene, which triggers the transition from vegetative to reproductive growth. We identified six *FT* homologs in coconut: CCG009539.1, CCG014033.1, CCG023905.1, CCG016739.1, CCG025329.1, and CCG011565.1. Among them, CCG011565.1 was more highly expressed at the reproductive stage than the seedling stage in both tall and dwarf coconut. Finally, *FT* can also enhance the expression of the *API* gene. We found four *API* homologs in coconut: CCG027514.1, CCG008584.1, CCG008583.1, and CCG022175.1. Among them, higher expression was detected for CCG027514.1 in mature trees compared to in seedlings of tall and dwarf coconut.

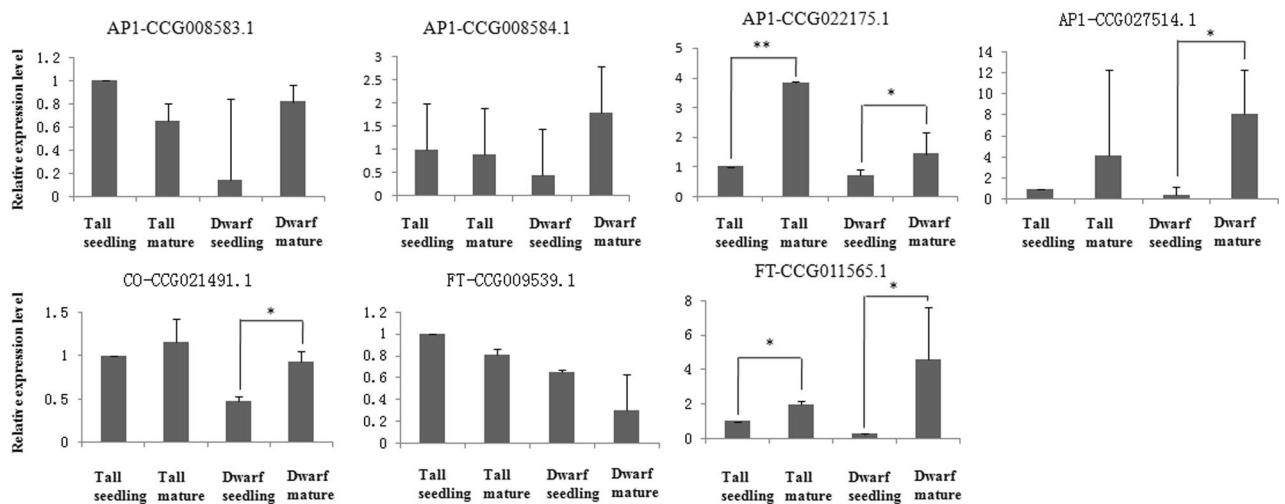
Meanwhile, quantitative real-time PCR (qPCR) was used to validate the gene expression profile of several genes involved in the photoperiod pathway: *CO* (CCG021491), *FT* (CCG009539 and CCG011565), and *API* (CCG008583, CCG008584, CCG022175, and CCG027514) gene copies (Fig. 5). The qPCR results were similar to the transcriptome data results. Higher expression levels in *CO* (CCG021491), *FT* (CCG011565), and *API* (CCG022175 and CCG027514) were detected in mature trees compared to in seedlings of tall and dwarf coconut.

### Alternative splicing of Flowering Locus T (FT) and association with flowering time.

A total of 11,376, 12,124, 11,676, and 12,186 novel transcripts were detected in mature tall coconut trees, tall coconut seedlings, mature trees of dwarf coconut, and dwarf coconut seedlings, respectively. These transcripts were analysed in conjunction with gene predictions to determine if alternative splicing events were occurring. We detected five types of alternative splicing: alternative 3′ splicing sites (A3SS), alternative 5′ splicing sites (A5SS), mutually exclusive exons (MXE), retained introns (RI), and skipped exons (SE). The largest proportion of novel alterna-



**Figure 4.** Gene expression changes involving the photosynthesis pathway which govern flowering time between seedlings and mature trees in tall and dwarf coconut.



**Figure 5.** Quantitative real-time PCR validation of candidate genes involved in the photosynthesis pathway, including *AP1*, *FT* and *CO*. \*Significant difference of gene expression between mature and seedling period of tall and dwarf coconut ( $P < 0.05$ ). \*\*High significant difference of gene expression between mature and seedling period of tall and dwarf coconut ( $P < 0.001$ ).





this pathway, *FT* is a crucial gene involved in signal transduction; overexpressing *FT* can result in an extremely early-flowering phenotype<sup>7</sup>. In our study, we detected alternative splicing of the *FT* gene between tall and dwarf coconut, as well as a number of key genes putatively involved in regulation of flowering time in coconut.

One of the most crucial phases in flowering plant development is the decision to flower, and flowering time has a major influence on plant fitness. The transition from the vegetative to the reproductive stage is controlled by several external factors, including photoperiod<sup>14</sup>, temperature, and abiotic stress, as well as by internal factors such as hormone production, C/N nutrient ratios and the age of the plant<sup>37</sup>. A complex set of interactions between the internal and external environment together decide when the plant transforms from the vegetative to the reproduction stage. In woody plants, it is difficult to investigate the molecular mechanisms controlling flowering time, due to their long life-cycles. Woody plants, due to their characteristic long life cycles and perenniality, generally have specific strategies to adapt to the ever-changing environment, such as bud dormancy. Despite this, several studies have shown a high level of conservation of genetic networks regulating flowering time between woody and herbaceous plants<sup>38–40</sup>. In our study, some coconut genes also showed high similarity with *Arabidopsis* genes known to be involved in the regulation of flowering time, including *CRY1*, *PHYA*, *PHYB*, *ZTL*, *SOC1*, *ELF3*, *FKF1*, *ELF4*, *CO*, and *FT* homologs. *CRY1* has been validated to up-regulate expression of clock genes in the *Arabidopsis* photoperiod pathway, including *ZTL* (*zeitlupe*), *SOC1* (*suppressor of overexpression of constans*), *ELF3* (*early flowering 3*), *FKF1* (*flavin binding F-box 1*), and *ELF4* (*early flowering 4*)<sup>11</sup>. By contrast, *PHYA* and *PHYB* can suppress the expression of clock genes in the *Arabidopsis* photoperiod pathway<sup>20</sup>. We also observed *SOC1* genes were up-regulated expressed in mature period compared to seedling period from tall and dwarf coconut.

Coconut is mainly cultivated in the tropical and subtropical regions, including Central and South America, East and West Africa, Southeast Asia and the Pacific Islands. Hence, it is not possible that flowering time could be regulated by the vernalization pathway in coconut. Based on our transcriptome data, we think that flowering time in coconut is most probably regulated by the photoperiod pathway. We observed high levels of expression of some crucial genes involved in the photoperiod pathway at the mature tree stage compared to the seedling stage.

Most genes in flowering plants contain introns which will be spliced from pre-mRNA in the nucleus before mature mRNA is formed. Splice site variation in the junction between introns and exons will result in multiple mRNA transcripts or isoforms from a single gene<sup>41</sup>. Alternative splicing can contribute to gene regulation and proteome diversity, and is also known to be involved in stress response<sup>42–45</sup>, development and reproductive growth<sup>46–48</sup>. In *Arabidopsis*, *FT* plays an important role in the transition from vegetative growth to reproduction growth<sup>49,50</sup>. Alternative splicing of *FT* in *Chrysanthemum morifolium* results in one *FT* gene encoding up to four different transcripts<sup>51</sup>. The expression level of these four *FT* isoforms depends on the developmental stage of flower, with alternative splicing associated with the degree of early flowering<sup>51</sup>. We also detected alternative splicing of a *CnFT* gene between tall and dwarf coconut. In dwarf coconut, which requires approximately 3–5 years to enter its reproductive stage, we detected only one transcript type, which lacked six nucleotides compared to the transcript type found to be most abundant in tall coconut (which requires 8–10 years to enter its reproductive stage). These results suggest that the *FT* transcript lacking the six nucleotides produce by A3SS may have a positive contribution to the early flowering phenotype in coconut. The other isoform of *CnFT* was detected in 8/11 tall coconut individuals: the remaining three may potentially result from open-pollination/heterozygosity resulting from cross-pollination with dwarf coconut varieties. Our results suggest that this alternative splice variant of *FT* could be a major player in flowering time differentiation between the dwarf and tall coconut types, but further investigation and validation is warranted to confirm this effect.

## Data availability

The raw data of transcriptome is from seedling and mature period of tall and dwarf coconut available in the European Nucleotide Archive (Submission Number: PRJEB34852). The CDS and genome sequence (the first version of coconut genome) is available in the GigaDB database (GigaDB, RRID:SCR 004002, <https://gigadb.org/dataset/100347>)<sup>35</sup>.

Received: 20 December 2019; Accepted: 24 June 2020

Published online: 15 July 2020

## References

- Batugal, P., Rao, V., Oliver, J., eds. Coconut Genetic Resources. IPGRI-Regional Office for Asia, the Pacific and Oceania, Serdang, Malaysia (2005).
- Patel, J. Coconut breeding. *Proc. Assoc. Econ. Biol.* **5**, 1–16 (1937).
- Koornneef, M., Hanhart, C. J. & van der Veen, J. H. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229**, 57–66 (1991).
- Clarke, J. H. & Dean, C. Mapping *FRI*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **242**, 81–89 (1994).
- Langridge, J. Effect of day-length and gibberellic acid on the flowering of *Arabidopsis*. *Nature* **180**, 36–37 (1957).
- Sheldon, C. C. *et al.* The FLE MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**, 445–458 (1999).
- Schiessl, S. V., Huettel, B., Kuehn, D., Reihardt, R. & Snowdon, R. J. Flowering time gene variation in Brassica species shows evolutionary principles. *Front. Plant Sci.* **8**, 1742 (2017).
- Samach, A. & Coupland, G. Time measurement and the control of flowering in plants. *BioEssays* **22**, 38–47 (2000).
- Onouchi, H. & Coupland, G. The regulation of flowering time of *Arabidopsis* in response to daylength. *J. Plant Res.* **111**, 271–275 (1998).
- Tanaka, C. *et al.* Direct interaction between VRN1 protein and the promoter region of the wheat FT gene. *Genes Genet. Syst.* **93**(1), 25–29 (2017).

11. Halliday, K. J., Koornneef, M. & Whitelam, G. C. Phytochrome B and at least one other phytochrome mediate the accelerated flowering response of *Arabidopsis thaliana* L. to low red/far-red ratio. *Plant Physiol.* **104**, 1311–1315 (1994).
12. Guo, Y., Wang, D., Jia, W. & Song, J. Effects of seed vernalisation and photoperiod on flowering induction in the haplophyte *Thellungiella halophila*. *Aust. J. Bot.* **60**, 743 (2012).
13. Lin, C. Photoreceptors and regulation of flowering time. *Plant Physiol.* **123**, 39–50 (2000).
14. Baurle, I. & Dean, C. The timing of developmental transitions in plants. *Cell* **125**, 655–664 (2006).
15. Boss, P. K., Bastow, P. M., Mylne, J. S. & Dean, C. Multiple pathways in the decision to flower: Enabling, promoting, and resetting. *Plant Cell.* **16**, 18–31 (2004).
16. Jack, T. Molecular and genetic mechanisms of floral control. *Plant Cell.* **16**, S1–17 (2004).
17. Moon, J., Lee, H., Kim, M. & Lee, I. Analysis of flowering pathway intergrators in *Arabidopsis*. *Plant Cell Physiol.* **46**, 292–299 (2005).
18. Putterill, J., Laurie, R. & Macknight, R. It's time to flower: The genetic control of flowering time. *BioEssays* **26**, 363–373 (2004).
19. Chardon, F. & Damerval, C. Phylogenomic analysis of the *PEBP* gene family in cereals. *J. Mol. Evol.* **61**, 1030–1033 (2005).
20. Izawa, T. *et al.* Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. *Genes Dev.* **6**, 2006–2020 (2002).
21. Paterson, A. H., Bowers, J. E., Peterson, D. G., Estill, J. C. & Chapman, B. A. Structure and evolution of cereal genomes. *Curr. Opin. Genet. Dev.* **13**, 644–650 (2003).
22. Salse, J., Piegu, B., Cooke, R. & Delseny, M. New in silico insight into the synteny between rice (*Oryza sativa* L.) highlights reshuffling and identifies new duplications in the rice genome. *Plant J.* **38**, 396–409 (2004).
23. Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S. & Shimamoto, K. Hd3a protein is a mobile flowering signal in rice. *Science* **316**, 1033–1040 (2007).
24. Faure, S., Higgins, J., Turner, A. & Laurie, D. A. The *Flowering Locus T*-like gene family in barley (*Hordeum vulgare*). *Genetics* **176**, 599–609 (2007).
25. Coelho, C. P., Minow, M. A. A., Chalfun-Junior, A. & Colasanti, J. Putative sugarcane *FT/TFL 1* genes delay flowering time and alter reproductive in *Arabidopsis*. *Front. Plant Sci.* **5**, 221 (2014).
26. Schiessl, S., Iniguez-Luy, F., Qian, W. & Snowdon, R. Diverse regulatory factors associated with flowering time and yield responses in winter-type *Brassica napus*. *BMC Genom.* **16**(1), 737 (2015).
27. Schiessl, S., Samans, B., Huttel, B., Reinhard, R. & Snowdon, R. J. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* **5**, 404 (2014).
28. Hsu, C. Y., Liu, Y., Luthe, D. S. & Yuceer, C. Poplar *FT2* shortens the juvenile phase and promotes seasonal flowering. *Plant Cell.* **18**, 1846–1861 (2006).
29. Xiao, Y. *et al.* Efficient isolation of high quality RNA from tropical palms for RNA-seq analysis. *Plant Omics.* **5**, 584–589 (2012).
30. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptome by RNA-seq. *Nat. Methods* **5**, 621–628 (2018).
31. Audic, S. & Claverie, J. M. The significance of digital gene expression profiles. *Genome Res.* **7**, 986–995 (1997).
32. Peiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368–375 (2003).
33. Wu, J., Zhang, Y., Zhang, H., Huang, H. & Folta, K. M. Whole genome wide expression profiles of *Vitis amurens* grape responding to downy mildew by using Solex sequencing technology. *BMC Plant Biol.* **10**, 234 (2010).
34. Shen, S. *et al.* Robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Acad. Sci. USA* **111**, E5593–E5601 (2014).
35. Xiao, Y. *et al.* The genome draft of coconut (*Cocos nucifera*). *Gigascience.* **6**, 1–11 (2017).
36. Menon, K. P. V. & Pandalai, K. M. *The Coconut. A Monograph.* Indian Central Coconut Committee 86–102 (Government Press, Madras, 1958).
37. Otiz-Marchena, M. I. Photoperiodic control of Carbon distribution during the floral transition in *Arabidopsis*. *Plant Cell.* **26**, 565–584 (2014).
38. Dong, X. *et al.* Genetic control of flowering time in woody plants: Roses as an emerging model. *Plant Divers.* **39**, 104–110 (2017).
39. Mohamed, R. *et al.* *Populus CEN/TFL 1* regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. *Plant J.* **62**, 674–688 (2010).
40. Chai, G. *et al.* *R2R3-MYB* gene pairs in *Populus*: Evolution and contribution to secondary wall formation and flowering time. *J. Exp. Bot.* **65**, 4255–4269 (2014).
41. Barbazuk, W. B., Fu, Y. & McGinnis, K. M. Genome-wide analyses of alternative splicing in plants: Opportunities and changes. *Genome Res.* **18**, 1381–1392 (2008).
42. Kwon, Y. J., Park, M. J., Kim, S. G., Baldwin, I. T. & Park, C. M. Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in *Arabidopsis*. *BMC Plant Biol.* **14**, 136 (2014).
43. Feng, J. *et al.* *SKIP* confers osmotic tolerance during salt stress by controlling alternative gene splicing in *Arabidopsis*. *Mol. Plant.* **8**, 1038–1052 (2015).
44. Shikata, H. *et al.* Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **111**, 18781–18786 (2014).
45. Wang, H. F. *et al.* Alternative splicing during *Arabidopsis* flower development results in constitutive and stage-regulated isoforms. *Front. Genet.* **5**, 25 (2014).
46. Zhang, B., Liu, Z. X., Ma, J., Song, Y. & Cheng, F. J. Alternative splicing of the *AGAMOUS* orthologous gene in double flower of *Magnolia stellata* (Magnoliaceae). *Plant Sci.* **241**, 277–285 (2015).
47. Yuan, Y. X. *et al.* A naturally occurring splicing site mutation in the *Brassica rapa FLC1* gene is associated with variation in flowering time. *J. Exp. Bot.* **60**, 1299–1308 (2009).
48. Lee, J. H. *et al.* Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* **342**, 628–632 (2013).
49. Corbesier, L. *et al.* *FT* protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007).
50. Notaguchi, M. *et al.* Long-distance, graft-transmissible action of *Arabidopsis* FLOWERING LOCUS T protein to promote flowering. *Plant Cell Physiol.* **49**, 1645–1658 (2008).
51. Mao, Y. *et al.* Functional analysis of alternative splicing of the FLOWERING LOCUS T orthologous gene in *Chrysanthemum morifolium*. *Hortic. Res.* **3**, 16058 (2016).

## Acknowledgements

This study has been supported by Hainan major science and technology project and the National Nonprofit Institute Research Grants of 1630152020005, 1630152019006 and 1630152017007, and the Species and variety Protection of the Ministry of Agriculture and Rural Affairs.

### Author contributions

Y.X., W.X., H.F.—designed the study. R.L., J.Z.—performed transcriptome sequencing and RNA extraction. Z.L., S.G., Y.Z. and J.Z.—performed alternative splicing analysis. Y.X., A.S.M. and W.X.—drafted and revised the manuscript. Y.D. and X.S.—assisted in the data analysis.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-68431-2>.

**Correspondence** and requests for materials should be addressed to H.F. or Y.X.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**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) 2020