JARQ 57 (2), 139-144 (2023) https://www.jircas.go.jp

Characterization of Leaf Transcriptome in a Tropical Tree Species, *Shorea curtisii*, over a Flowering Season

Ahmad Husaini SUHAIMI[1], Masaki J. KOBAYASHI[2], Akiko SATAKE[3], Soon Leong LEE[4], Norwati MUHAMMAD[4], Tatsuya OTANI[5], Toshiaki KONDO[6], Naoki TANI[2,7]*and Suat Hui YEOH[1]*

[1] Institute of Biological Sciences, Faculty of Science, Universiti Malaya, Lembah Pantai, Kuala Lumpur, Malaysia
[2] Forestry Division, Japan International Research Center for Agricultural Sciences, Tsukuba, Japan
[3] Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Japan
[4] Forestry Biotechnology Division, Forest Research Institute Malaysia, Kepong, Selangor, Malaysia
[5] Shikoku Research Center, Forestry and Forest Products Research Institute, Forestry Research and Management Organization, Kochi, Japan
[6] Bio-Resources and Utilization Division, Japan International Research Center for Agricultural Sciences, Tsukuba, Japan
[7] Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Abstract

General flowering (GF) is a synchronous flowering phenomenon in the Southeast Asian tropical rainforests that occurs at irregular intervals of multiple years. While in this study, the leaf transcriptome of a GF species, *Shorea curtisii* obtained from three-time points – before and after floral initiation, and post-flowering stage – was sequenced, the unpredictable intervals of GF raise conservation concerns for these under-researched forests with rich economically and ecologically important species. We assembled 243,759,478 sequencing reads into 39,943 non-redundant unigenes including 677 putative homologs of *Arabidopsis thaliana* flowering-related genes. Differential expression analysis conducted on pairwise comparisons of the time points identified 930 differentially expressed unigenes, which includes 17 flowering-related homologs. The differential expression of unigenes with significant enrichments of functions related to drought corroborated the involvement of drought as an environmental cue for GF, with the outcomes of this study offering an insight into the conservation of floral regulatory genes and pathways in *Shorea* and possibly being used as a model to better understand the floral initiation cues and regulation of GF trees.

Discipline: Forestry
Additional key words: Dipterocarpaceae, flowering time, mass flowering, next-generation sequencing, transcriptomics

Introduction

General flowering (GF) is a synchronous flowering phenomenon that occurs at irregular intervals of one to ten years in the Southeast Asian tropical rainforests (Appanah 1993, Chen et al. 2018) and involves multiple plant families including Dipterocarpaceae (Ashton et al. 1988). As understanding the triggers of GF and predicting its timing is important to solve these problems, the unpredictable intervals of GF cause difficulty in sufficient seed collection and consequently their conservation (Kettle 2010).

Despite the lack of clear seasonality in the tropics (Kume et al. 2011), many studies have proposed that GF is triggered by environmental factors such as drought and low temperature, which occurs supra-annually in this region (Chen et al. 2018, Ushio et al. 2020, Yeoh et al. 2017). Understanding the molecular basis of floral transition is the first step in predicting the timing of GF. However, our understanding of this topic is still limited (Kobayashi et al. 2013, Yeoh et al. 2017).

In the model plant, *Arabidopsis thaliana*, the process of determining flowering time is regulated by complex genetic networks that respond to multiple

*Co-corresponding authors: suathui_yeoh@um.edu.my, ntani@affrc.go.jp
Received 12 April 2022; accepted 19 July 2022.

endogenous and external environmental factors (Ausin et al. 2005). Six major floral regulatory pathways have been identified in *A. thaliana* namely photoperiod, ambient temperature, vernalization, aging, autonomous, and gibberellin pathways (reviewed in Parcy 2005). The signals perceived by these pathways are integrated by floral pathway integrators (Simpson & Dean 2002), which then activate the transcription of floral meristem identity genes in the shoot apical meristem, hence initiating flowering (Wagner 2017). Where it is not clear whether such conservation of flowering genes is observed in other GF species and which genes are differentially expressed during GF in other tissues, we saw in a previous study in buds of *Shorea beccariana* that a number of genes in the floral regulatory pathways were conserved at the sequence and functional levels (Kobayashi et al. 2013). Besides buds, leaves are important in floral initiation as floral signals are perceived and transmitted from leaves (reviewed in Johansson & Staiger 2015). Hence, differentially expressed genes in leaf tissues over a flowering season can help to discern the environmental cues and molecular mechanism of floral induction in GF species.

To elucidate the extent of flowering gene and pathway conservation among GF species as well as the environmental cues that trigger flowering, we investigated transcriptome from leaves in another GF species, *S. curtisii*. Using mathematical modelling on meteorological data and expression level of a floral integrator gene as an indicator of floral initiation, Yeoh et al. (2017) previously found that the synergistic effects of drought and low temperature best explained the occurrence of GF in this species. Based on the gene expression reported in the previous study as a guide to the progression of flowering and the transcriptome analysis in the current study, we discussed transcriptional changes in leaves during flowering and the conservation of floral regulatory pathways in *S. curtisii*.

Materials and methods

1. Study species and sampling

Shorea curtisii (Dipterocarpaceae), from *Shorea* section *Mutica*, have been recorded as reliable indicators of GF (Ashton et al. 1988). In this study, leaf samples at the top layer of canopies from two *S. curtisii* individuals, C1 and C2, with diameter at breast height of 72 cm and 89 cm, respectively, were collected from Semangkok Forest Reserve (2°58′N, 102°18′E). The samples were from the collection described in a previous study (Yeoh et al. 2017). We selected three-time points (TPs) corresponding to before and after floral initiation, and abortion/fruiting stage based on the expression pattern of an *FT* homolog reported in the same study (Yeoh et al. 2017) and their developmental stages i.e., vegetative (TP-A, 18 December 2013), inflorescence (TP-B; 02 April 2014), and fruiting or abortion (TP-C; 17 June 2014) observed during collection (Supplementary Table S1). Samples were collected at each TP around midday, soaked in RNAlater reagent (Ambion, USA), and stored at –80°C.

2. Transcriptome sequencing and annotation

Total RNA was extracted from leaf samples using the method described by Kobayashi et al. (2013). RNA-seq was performed using Illumina HiSeq 4000 Sequencer (Illumina, USA). After removal of low quality bases by Trimmomatic v0.36 (Bolger et al. 2014), the sequence reads from each sample were combined and *de novo* assembled using Trinity v2.8.5 (Grabherr et al. 2011), where it was seen that only non-redundant transcripts with complete open reading frame were retained by using TransDecoder v5.5.0 (http://transdecoder.github.io) and CD-HIT v4.8.1 (Fu et al. 2012). The transcripts in the final set of assembly are hereafter referred to as unigenes. The quality of the assembly was assessed with BUSCO v5.2.1 (Simao et al. 2015). The unigenes were queried against the proteome of *A. thaliana* (Cheng et al. 2017) using BLASTx v2.10 (Camacho et al. 2009) with *E*-value cut-off: 1E–10. The annotated unigenes were searched against *A. thaliana* flowering genes database (Bouché et al. 2016) to identify their homologs in *S. curtisii*. Translated unigenes were also queried against other public protein databases using BLASTp v2.10 (Camacho et al. 2009) and InterProScan v5.36 (Jones et al. 2014). We also searched the unigenes against Gene Ontology (GO; Ashburner et al. 2000) and KEGG pathways (Kanehisa & Goto 2000) to characterize the unigenes. GO classification is divided into cellular component (CC), molecular function (MF), and biological process (BP) categories. Details of the methods are shown in Supplementary Figure S1.

3. Differential gene expression and enrichment analyses

The unigenes were quantified by Salmon v1.5.1 (Patro et al. 2016) and the read counts were utilized for differential expression analysis using DESeq2 v1.32.0 (Love et al. 2014). The samples were compared in a pairwise manner and differentially expressed unigenes (DEUs) with absolute $\log_2$ fold change ≥ 1 and FDR < 0.05 were identified. Significantly enriched GO terms ($P < 0.05$) and KEGG pathways (FDR < 0.05) in the DEUs were also identified using topGO package

v2.40.0 (Alexa & Rahnenfuhrer 2020) and KOBAS v3.0 (Bu et al. 2021), respectively.

4. Supplementary

All supplementary tables and figures are in a public data depository, Dryad (https://datadryad.org/). The digital object identifier (DOI) is https://doi.org/10.5061/dryad.69p8cz94f.

Results and discussion

1. Assembly and annotation of *Shorea curtisii* leaf transcriptome

The RNA-seq generated 243,759,478 sequences with an average length of 150 bp. After the removal of redundant transcripts, the resulting *S. curtisii* leaf transcriptome comprises 39,943 unigenes with an average length of 2,073 bp (Supplementary Fig. S2) and N50 of 2,393 bp. The transcriptome has a GC content of 41.92%, which is comparable to other trees in the order Malvales (Gao et al. 2018, Teh et al. 2017). The quality assessment of the assembly using BUSCO indicates an almost complete transcriptome assembly (Supplementary Table S2). Of the total unigenes, 37,852 unigenes (94.77%) were annotated with at least one of the protein databases showing that only a small percentage of the unigenes (5.23%) have yet to be identified (Table 1, Supplementary Table S3). All read files are available in the NCBI SRA database under BioProject accession number PRJNA768952. The assembly data is available in the NCBI TSA database under TSA Accession No. GJMJ00000000.

Among the 36,253 homologs of *A. thaliana* in *S. curtisii* (Table 1), we identified 677 homologs that correspond to 204 (67% out of the total 306) non-redundant *A. thaliana* flowering-related genes distributed across all of the floral regulatory pathways (Table 2, Supplementary Table S4). Among the homologs of flowering-related genes, and also implying that the sequence of these genes is conserved in *Shorea* spp, 77 were also identified in *Shorea beccariana* (Kobayashi et al. 2013).

2. Differential expression of flowering-related homologs across vegetative and flowering stages

Differential expression analysis of pairwise TPs to identify candidate unigenes that are involved in floral regulation of *S. curtisii* found 930 non-redundant DEUs (Table 3, Supplementary Table S5). The identified DEUs included 17 flowering-related homologs in *S. curtisii* (Table 4, Supplementary Table S4). Among the differentially expressed flowering-related homologs, two

Table 1. Summary of *Shorea curtisii* unigenes annotations against public protein databases

Database	Number of unigenes
Araport11	36,253 (90.76%)
UniProt (v 2019-09)	32,003 (80.12%)
Pfam v32.0	30,916 (77.40%)
PANTHER v14.1	36,491 (91.36%)
GO (v 2019-12)	36,156 (90.52%)
KEGG (v 2019-12)	16,899 (42.30%)
Annotated in at least one database	37,852 (94.77%)
Annotated in all the databases	15,603 (39.06%)

Number in parentheses indicates the percentage of annotated unigenes to the total number of unigenes.

Table 2. Summary of *Arabidopsis thaliana* flowering-related homologs found in the leaf transcriptome of *Shorea curtisii* based on the floral regulatory pathways

Pathway	Number of *A. thaliana* genes	Number of *S. curtisii* homologs
Aging	22	20 (5)
Ambient temperature	7	16 (4)
Autonomous	7	28 (7)
Floral meristem identity	9	13 (4)
General	117	304 (94)
Hormones	28	28 (15)
Integrator	8	23 (4)
Photoperiod and circadian clock	103	225 (65)
Sugar	9	26 (7)
Vernalization	28	69 (19)

Number in parentheses indicates the number of nonredundant *A. thaliana* genes that correspond to *S. curtisii* homologs.

Table 3. Number of differentially expressed unigenes (DEUs) for each pairwise time-point comparisons of leaf transcriptome in *Shorea curtisii*

Time points	Number of DEUs			
	High expression at TP-A	High expression at TP-B	High expression at TP-C	Total
TP-A and TP-B	147	497	-	644
TP-B and TP-C	-	66	62	128
TP-A and TP-C	150	-	201	351

unigenes, expression of both homologs being consistent with the previous study (Yeoh et al. 2017), of encoding floral promoter, *ScFT1* and *ScFT2,* were highly expressed at floral initiation (TP-B, Table 4), which indicates the reliability of the RNA-seq data.

Surprisingly, the differential expression analysis showed more than half of the differentially expressed flowering-related homologs (9 out of 17) were associated with the circadian clock and photoperiod pathway (Table 4, Supplementary Table S4) although all sampling activities were conducted around noon. As such, the fluctuations in the expression levels of these unigenes could not simply be explained by a diurnal effect. Other factors such as temperature, nutritional status (reviewed in Inoue et al. 2017) and drought (Ng et al. 2021, Wang et al. 2021) have been reported to cause transcriptional changes to circadian clock genes previously.

3. Functional analysis of differentially expressed unigenes in *Shorea curtisii*

To further understand the functions of the DEUs, we conducted GO and KEGG enrichment tests. The analysis showed that several GO terms associated with drought response such as "response to reactive oxygen species" (BP) and "response to water deprivation" (BP) (Carvalho 2008, Huang et al. 2008) were enriched in the DEUs that were highly expressed during the floral transition (Supplementary Table S6B). This suggests that the individuals were responding to drought prior to flowering, which is congruent with meteorological data at the study site (Yeoh et al. 2017) and previous studies in dipterocarps that hypothesized drought as one of the GF triggers (Chen et al. 2018, Kobayashi et al. 2013, Satake et al. 2019).

Moreover, we found genes involved in "circadian rhythm" enriched during the floral transition (Supplementary Table S7B) which is consistent with reports in other plants (Liu et al. 2016, Quan et al. 2019). Previous studies have reported that, where the observed changes in the expression levels of circadian clock-related homologs in this study may be caused by drought during the floral transition, the expressions of circadian clock genes are affected by drought in Dipterocarpaceae (Supplementary Table 22 in Ng et al. 2021) and other plants (Legnaioli et al. 2009, Wang et al. 2021). However, further studies are needed to understand transcriptional regulations of the circadian clock and photoperiodic genes under drought and their involvements in the regulation of GF.

Conclusion

In our transcriptome analysis using leaves of a GF species, *S. curtisii*, we found multiple homologs of flowering-related genes from all of the *A. thaliana* flowering pathways. Partial overlap of these homologs with those in *S. beccariana* suggests therefore that conservation of flowering-related genes at the sequence level are shared among the genus *Shorea*. Furthermore, it is seen that the results of differential expression analysis as well as GO and KEGG enrichment tests supported the notion of drought as a trigger of GF. While we hope

Table 4. Expression of differentially expressed flowering-related homologs in *Shorea curtisii* and the corresponding flowering pathways

Unigene	Flowering pathway	TP-A	TP-B	TP-C
ScFT1	Integrator	0.66	75.51	149.49
ScFT2		0.00	109.63	84.13
ScFKF1	Photoperiod/ Circadian Clock	62.97	19.48	39.64
ScLHY1		335.10	88.74	231.23
ScLHY2		1,013.80	294.30	772.70
ScLHY3		515.73	126.18	400.45
ScLWD1		49.74	4.83	6.24
ScPRR9-1		245.09	78.04	240.54
ScPRR9-2		241.88	39.60	236.55
ScRAV2		12.94	128.4	36.24
ScSPL4		165.52	64.82	41.29
ScVRN2	Vernalization	0.00	10.3	1.42
ScAPS1	Sugar	6.56	29.6	10.54
ScPGI1		3.54	17.67	27.49
ScSUS4		51.37	78.27	14.60
ScELF6	General	0.21	2.23	5.61
ScFCA		7.84	0.00	4.67

The expression values were normalized using Trimmed Mean of M-values (TMM, Robinson & Oshlack 2010).

that the insights obtained from the transcriptome data of *S. curtisii* in this study will become the basis for future molecular studies of tropical plant phenology, further studies using more TPs will help us understand the relationship between drought and GF.

Acknowledgements

We acknowledge the Selangor State Forest Department for granting permission to conduct this research in Semangkok Forest Reserve. The authors appreciate B Yasri and P Ramli for their assistance in the fieldwork, and KKS Ng for his assistance in RNA extraction. The study was supported by grants from MoE-HIR (UM.C/625/1/HIR/MOE/SCI/18) and BKP (BKS074-2017) awarded to YSH as well as Environment Research and Technology Development Fund (Grant No. RFd-1101) from The Ministry of the Environment, Japan and KAKENHI (26251042) grants awarded to AS. AHS was funded by Universiti Malaya's SLAB fellowship.

References

Alexa, A. & Rahnenfuhrer, J. (2020) topGO: Enrichment analysis for Gene Ontology.

Appanah, S. (1993) Mass flowering of dipterocarp forests in the aseasonal tropics. *J. Biosci.*, **18**, 457-474.

Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25-29.

Ashton, P. S. et al. (1988) Staggered flowering in the Dipterocarpaceae - new insights into floral induction and the evolution of mast fruiting in the aseasonal tropics. *Am. Nat.*, **132**, 44-66.

Ausin, I. et al. (2005) Environmental regulation of flowering. *Int. J. Dev. Biol.*, **49**, 689-705.

Bolger, A. M. et al. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

Bouché, F. et al. (2016) FLOR-ID: An interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.*, **44**, D1167-D1171.

Bu, D. et al. (2021) KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.*, **49**, W317-W325.

Camacho, C. et al. (2009) BLAST plus: Architecture and applications. *BMC Bioinform.*, **10**.

Chen, Y. Y. et al. (2018) Species-specific flowering cues among general flowering *Shorea* species at the Pasoh Research Forest, Malaysia. *J. Ecol.*, **106**, 586-598.

Cheng, C. Y. et al. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.*, **89**, 789-804.

Consortium, T. U. (2020) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480-D489.

Cruz de Carvalho, M. H. (2008) Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signal. Behav.*, **3**, 156-165.

El-Gebali, S. et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427-D432.

Fu, L. M. et al. (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150-3152.

Gao, Y. et al. (2018) *De novo* genome assembly of the red silk cotton tree (*Bombax ceiba*). *GigaScience*, **7**, giy051.

Grabherr, M. G. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644-U130.

Huang, D. et al. (2008) The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *J. Exp. Bot.*, **59**, 2991-3007.

Inoue, K. et al. (2017) Integration of input signals into the gene network in the plant circadian clock. *Plant Cell Physiol.*, **58**, 977-982.

Johansson, M. & Staiger, D. (2015) Time to flower: Interplay between photoperiod and the circadian clock. *J. Exp. Bot.*, **66**, 719-730.

Jones, P. et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, **30**, 1236-1240.

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27-30.

Kanehisa, M. et al. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.*, **28**, 726-731.

Kettle, C. J. (2010) Ecological considerations for using dipterocarps for restoration of lowland rainforest in Southeast Asia. *Biodivers. Conserv.*, **19**, 1137-1151.

Kobayashi, M. J. & Shimizu, K. K. (2013) Challenges in studies on flowering time: interfaces between phenological research and the molecular network of flowering genes. *Ecol. Res.*, **28**, 161-172.

Kobayashi, M. J. et al. (2013) Mass flowering of the tropical tree *Shorea beccariana* was preceded by expression changes in flowering and drought-responsive genes. *Mol. Ecol.*, **22**, 4767-4782.

Kume, T. et al. (2011) Ten-year evapotranspiration estimates in a Bornean tropical rainforest. *Agric For Meteorol*, **151**, 1183-1192.

Legnaioli, T. et al. (2009) TOC1 functions as a molecular switch connecting the circadian clock with plant responses to drought. *EMBO J.*, **28**, 3745-3757.

Liu, K. et al. (2016) Transcriptome analysis and identification of genes associated with floral transition and flower development in sugar apple (*Annona squamosa* L.). *Front. Plant Sci.*, **7**, 1695.

Love, M. I. et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 1-21.

Ng, K. K. S. et al. (2021) The genome of *Shorea leprosula* (Dipterocarpaceae) highlights the ecological relevance of drought in aseasonal tropical rainforests. *Commun. Biol.*, **4**, 1166.

Parcy, F. (2005) Flowering: A time for integration. *Int. J. Dev. Biol.*, **49**, 585-593.

Patro, R. et al. (2016) Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *BioRxiv*, 021592.

Quan, S. et al. (2019) Stages identifying and transcriptome profiling of the floral transition in *Juglans regia*. *Sci. Rep.*, **9**, 7092.

Robinson, M. D. & Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.

Satake, A. et al. (2019) Drought and cool temperature cue general flowering synergistically in the aseasonal tropical forests of Southeast Asia. *Ecol. Res.*, **34**, 40-49.

Simao, F. A. et al. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212.

Simpson, G. G. & Dean, C. (2002) Arabidopsis, the Rosetta stone of flowering time? *Science*, **296**, 285-289.

Teh, B. T. et al. (2017) The draft genome of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.*, **49**, 1633-1641.

Thomas, P. D. et al. (2003) PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.*, **13**, 2129-2141.

Ushio, M. et al. (2020) Dynamic and synergistic influences of air temperature and rainfall on general flowering in a Bornean lowland tropical forest. *Ecol. Res.*, **35**, 17-29.

Wagner, D. (2017) Key developmental transitions during flower morphogenesis and their regulation. *Curr. Opin. Genet. Dev.*, **45**, 44-50.

Wang, K. et al. (2021) Two homologous *LHY* pairs negatively control soybean drought tolerance by repressing the abscisic acid response. *New Phytol.*, **229**, 2660-2675.

Ye, J. et al. (2006) WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.*, **34**, W293-W297.

Yeoh, S. H. et al. (2017) Unravelling proximate cues of mass flowering in the tropical forests of South-East Asia from gene expression analyses. *Mol. Ecol.*, **26**, 5074-5085.