Songklanakarin J. Sci. Technol.
41 (2), 332-340, Mar. – Apr. 2019

http://www.sjst.psu.ac.th

Original Article

An *EgHd3a-like* and its alternatively spliced transcripts in the oil palm (*Elaeis guineensis*)

Sukhuman Whankaew[1], Kedsirin Ruttajorn[1], Claus Krogh Madsen[2], Torben Asp[2], Li Xu[1], Alisa Nakkaew[1, 3], and Amornrat Phongdara[1, 3*]

[1] *Department of Molecular Biotechnology and Bioinformatics, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110 Thailand*

[2] *Department of Molecular Biology and Genetics, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, Slagelse, 4200 Denmark*

[3] *Center for Genomics and Bioinformatics Research, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110 Thailand*

Received: 6 December 2016; Revised; 24 October 2017; Accepted: 17 November 2017

Abstract

Heading date 3a (*Hd3a*) is an important flowering regulator of short-day plants, and plays a role in several developmental processes. The oil palm is commercially grown for the edible oil derived from its fruit. Flowering and fruit development are key factors for successful production, contributing to productivity of the oil palm fruit. Because little is known about the gene organization of *Hd3a* in the oil palm, the present study aimed at isolating the *Hd3a* and its regulatory region. Furthermore, alternatively spliced transcripts were cloned, and their expression in different tissues was also investigated. The structure of the oil palm *Hd3a* (*EgHd3a-like*) gene and its promoter were established based on the isolation of *EgHd3a-like* cDNA and genomic DNA. The promoter analysis revealed that it contains two key regulatory elements, CCAAT boxes and an ARR1 motif, which are the binding sites of *Hd3a* inducers. Other *cis*-elements corresponding to flowering and organ development were also represented. Expression of the transcript variants was investigated by RT-PCR in the anther, pistil, mesocarp and leaf. *EgHd3a-like* was expressed in all tissues tested, which supports its multifaceted roles in several developmental processes. The variants were found in all tested tissue types but at different levels, showing some level of tissue-specificity by variant. Taken together, these results indicate that *EgHd3a-like* is regulated under various conditions and that transcript variants might play an important role in gene function and regulation.

Keywords: alternative splicing, floral transition, flowering locus T, heading date, promoter

1. Introduction

The oil palm (*Elaeis guineensis*) is commercially cultivated as an important source of edible oil. Palm oil is widely used in food, biofuel and cosmetic industries (Murphy, 2014). Among oil crops, the oil palm is an efficient crop in terms of oil yield and land usage without requiring replanting. Driven by economic development, the demand and pricing of palm oil are increasing (Murphy, 2014). Accordingly, finding ways to increase the yield is an important challenge and is necessary for increasing palm oil production.

The oil palm is considered a short-day (SD) plant, which accelerates floral transition or flowering under SD conditions (Legros *et al.*, 2009). Flowering is a potent factor contributing to fruit productivity. It is controlled by floral

*Corresponding author
Email address: pamornra@yahoo.com

regulatory genes. Under SD conditions, the long duration of night time induces *Heading date 3a* (*Hd3a*) expression by *Heading date 1* (*Hd1*) gene. Independently, *Early heading date* (*Ehd1*) is integrated into the photoperiodic control of *Hd3a* (Doi *et al.*, 2004; Kojima *et al.*, 2002). In previously studied plants, the Hd3a protein, a florigen, is synthesized in the vascular tissue of leaves, then moves to the shoot apical meristem and interacts with FD via a 14-3-3 protein. The resulting complex binds to the promoter of floral meristem identity genes to promote flowering (Taoka *et al.*, 2013).

Among floral regulatory genes under SD regulation, *Hd3a* has been detected as a heading-date-related quantitative trait locus (Kojima *et al.*, 2002). It encodes an ortholog protein of *Arabidopsis* FLOWERING LOCUS T (FT), which regulates floral transition in response to long-day conditions (Kojima *et al.*, 2002). FT is a mobile signal recognized as a florigen or flowering hormone. The role of FT in floral promotion is conserved. In addition, FT-like proteins also play a role in the regulation of several developmental processes, such as fruit set, vegetative growth and stomatal control (Pin & Nilsson, 2012). These observations support the possibility that *Hd3a* might be a key to promoting the floral transition of the oil palm and may play many important roles in several processes related to fruit yield.

Little is known about the genetics and genomic organization of *Hd3a* in the oil palm (*EgHd3a*). A better understanding of *EgHd3a-like* and its regulation might facilitate improvement of early flowering, of the flowering under long-day conditions, and of fruit development. To provide the genetic information on *EgHd3a-like*, the present study aimed to isolate and characterize *EgHd3a-like*. The promoter was also cloned and analyzed to obtain a better understanding of how *EgHd3a-like* is regulated. The present study also observed tissue-specific expression and identified novel alternatively spliced variants of *EgHd3a-like* transcripts. An understanding of the genetic regulation and genomic organization of *EgHd3a-like* and of its tissue specificity is important not only from the viewpoint of basic biological knowledge but also from the perspective of developing agronomical applications that increase oil palm productivity.

2. Materials and Methods

2.1 Sample preparation

Tenera palms were grown in the field of the Faculty of Natural Resources, Prince of Songkla University (Songkhla, Thailand). DNA was extracted from leaf samples of oil palm using the DNeasy plant mini kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. For RNA extraction, mature anthers, mature pistils, mesocarp (from 10 to 12 weeks after anthesis fruit), and leaves from frond-17 were collected from 5 to 7-year-old plants in triplicate. The leaf samples were collected at 8 a.m. in January. Leaf sampling was repeated in October in order to confirm the expression results. Total RNA was extracted using a CTAB (Cetyltrimethylammonium bromide) protocol modified from that of Jensen *et al.* (2015). The total DNA and RNA were quantified using a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., DE, USA). The cDNA was synthesized by Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, USA).

2.2 Isolation of *EgHd3a-like* and its promoter

FT homologous gene sequences from *Arabidopsis thaliana* (AB027504.1), *Oryza sativa* (AB433515.1), *Oncidium* Gower Ramsey (EU583502.1), *Chrysanthemum morifolium* (GQ925916.1), *Phyllostachys meyeri* (AB498761.1), *Helianthus annuus* (GQ884981.1), *Cymbidium faberi* (HQ164434.1), and *Eucalyptus globulus* (HQ453992.1) matched by alignment with a *Phoenix dactylifera* sequence (ACYX02010487.1) and the following PCR primers were designed from this alignment to be used for the amplification of *EgHd3a-like* gene in oil palm; *EgHd3a*F_1: 5'-GCAGCA CCCAGTGACAAATCTATA-3', *EgHd3a*F_2: 5'-GGATCC TTTGGTTGTGGGAAG-3', *EgHd3a*F_3: 5'-TGACCTCAGG ACCTTCTACAC-3', *EgHd3a*R_1: 5'-GTCATCCGAGGTTG TAGAGC-3' and *EgHd3a*R_2: 5'-GCCATAGCTTACGGTT GCATC-3'. The PCR mixture was composed of 1X Herculase II Reaction Buffer containing 2 mM $MgCl_2$, 0.4 mM dNTPs, 0.4 µM of each primer, 0-6% Dimethylsulfoxide (DMSO), and 0.5 µl Herculase II Fusion DNA Polymerases (Agilent Technologies, CA, USA). The PCR conditions were: an initial denaturation at 95°C for 2 min; then 40 cycles of 94°C for 1 min, 58°C for 30 s, and 72°C for 3 min; followed by a final extension at 72°C for 3 min. The fragments were then purified and cloned into the pCR™4Blunt-TOPO® vector (Invitrogen, USA), followed by sequencing. The isolated sequences were compared to the public NCBI database by Blastx (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The amino acid sequence homology was also confirmed by SWISS-MODEL workspace (Biasini *et al.*, 2014). The sequence alignment was conducted using BioEdit v7.2.5 and ClustalX. Sequences of cDNA and gDNA from oil palm were compared for structure of the gene. To obtain the promoter of *EgHd3a-like*, inverse PCR (iPCR) was performed. The transcription start site was predicted using Neural Network Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html) with a score cutoff of 0.85. PlantPAN software (Chow *et al.*, 2015) was used to define regulatory elements with a similar score cutoff, 0.95.

2.3 Tissue specific gene expression and alternatively spliced variants

To explore tissue-specific expression and observe alternatively spliced variants of the *EgHd3a-like*, RT-PCR was used. cDNAs from anther, pistil, mesocarp and leaf mRNAs were used as templates for PCR amplification with forward primer on exon 1 (*EgHd3a*1F: 5'-ATGTGAGCTT AAACCCTCGGC-3') and reverse primer on exon 4 (*EgHd3a*1R: 5'-AGGTTGCATCCTTCTCCCGC-3'). The PCR products were separated on a 2% agarose gel. The observed alternatively spliced transcripts were cloned into the pGEM-T Easy Vector (Promega, Madison, WI, USA), followed by sequencing and analysis.

3. Results and Discussion

3.1 Isolation of the *EgHd3a-like* gene

Hd3a, an ortholog of the *ArabidopsisFT*, is known to promote floral transition under SD condition (Kojima *et al.*, 2002). FT-like proteins also play broad functions in several developmental processes such as floral organ development,

fruit set, and fruit yield (Pin & Nilsson, 2012). In the present study, *EgHd3a-like* was isolated. The genomic sequence length from start codon to stop codon was 1,751 nucleotides containing 4 exons and 3 introns. The coding sequence (CDS) of *EgHd3a-like* was 555 nucleotides in length and encodes a putative protein of 184 amino acids. At the same time, the oil palm genome sequence was released (Singh *et al.*, 2013). Blastx analysis revealed homology to predicted protein HEADING DATE 3A-like isoform X2 (XP_019705101.1) of *E. guineensis*, HEADING DATE 3A-like isoform (XP_01 9705100.1) of *E. guineensis*, FLOWERING LOCUS T of *Symplocarpus renifolius* (BAV32136.1), and predicted HEADING DATE 3A-like isoform X1 of *P. dactylifera* (XP_008787029.1) with 100%, 96%, 88% and 86% identities, respectively. In depth analysis of oil palm genome showed that the predicted protein HEADING DATE 3A-like isoform X2 and X1 are located on chromosome number 3 at NC_025995.1 (10789981..10792272). In addition, there are the other 5 Hd3a homologs located on chromosomes number 2, 3, 7, 9 and "unknown".

The protein structure homology modeling showed that it is structurally homologous to the rice Hd3a protein with high sequence identity (77.58%). Amino acid sequence alignments with Hd3a from *Phoenix dactylifera* and *Oryza sativa* are shown in Figure 1. These analyses confirm that the protein corresponds to the Hd3a protein.

3.2 Characterization of *EgHd3a-like* promoter

Promoters are important regulatory regions controlling the development and physiology of organisms by regulating gene expression (Wittkopp & Kalay, 2012). Therefore, *EgHd3a-like* promoter was investigated. The putative transcription start site for *EgHd3a-like* is an adenine (A) at 837 bp and is marked as +1. In the 836 bp upstream of the transcription start site, several *cis*-acting elements have been detected (Appendix Table). In the floral regulatory network of rice, there are two major binding proteins that act as inducers of *Hd3a*, named *Hd1* (Hayama *et al.*, 2003), and *Ehd1* (Doi *et al.*, 2004). Therefore, it was investigated whether there are any *Hd1* and *Ehd1* binding sites on the *Hd3a* promoter. The CCAAT box is accepted as a binding site for *Hd1* (the rice ortholog of *Arabidopsis* CONSTANS) (Wenkel *et al.*, 2006). Based on a similarity score of 0.95 from PlantPAN (Chow *et al.*, 2015), three putative CCAAT boxes were identified at -36, -403 and -663 bp. Regarding the binding site of the Ehd1 protein, designated the ARR1 binding motif (Doi *et al.*, 2004), one such ARR1 site was found at -420 bp. The TATA box, a regulatory element involved in the formation of the transcription initiation complex, is generally located close to the transcription start site (Bernard *et al.*, 2010). Here, the TATA box was found at -43 bp. In addition, *cis*-acting elements corresponding to several transcription factors (TF) were observed. For example, the binding sites for the TF involved in flowering, plant hormone responsiveness, embryo development, cotyledon identity, and specification of polarity of the ovule outer integument were detected, as well as the binding site for the transcriptional repressor of flowering time under long-day conditions and the negative regulator in the phytochrome-mediated light responses. Moreover, several binding sites of TF responding to plant hormones and ions such as auxin, cytokinin, ethylene and calcium were detected. The discovery of these structural features indicates that it is controlled under various conditions and it may be possible to apply some of these features in treatments to induce the expression of *EgHd3a-like*.

3.3 Tissue distribution of *EgHd3a-like* and its alternatively spliced transcripts

The expression of *EgHd3a-like*, a multi-functional gene, was investigated in the anther, pistil, mesocarp and leaf. The results showed that *EgHd3a-like* was expressed in all tested tissues (Figure 2). It was mainly expressed in anther, pistil, and mesocarp, whereas only slight expression levels were detected in leaf. The results suggest that this *EgHd3a-like* might play an important role in floral and fruit development, and may thereby relate to oil palm fruit yield. To act as a florigen, it is essentially transcribed in leaf tissue (Pin and Nilsson, 2012). The low-level expression of *EgHd3a-like* in leaves might be because acting as a florigen in oil palm does not require a high amount of *EgHd3a-like* expression, or the environmental conditions during sample collection were perhaps not suitable for expression. The question of *EgHd3a-like* expression in different leaf fronds and different phases of flowering induction could be addressed in a further investigation.

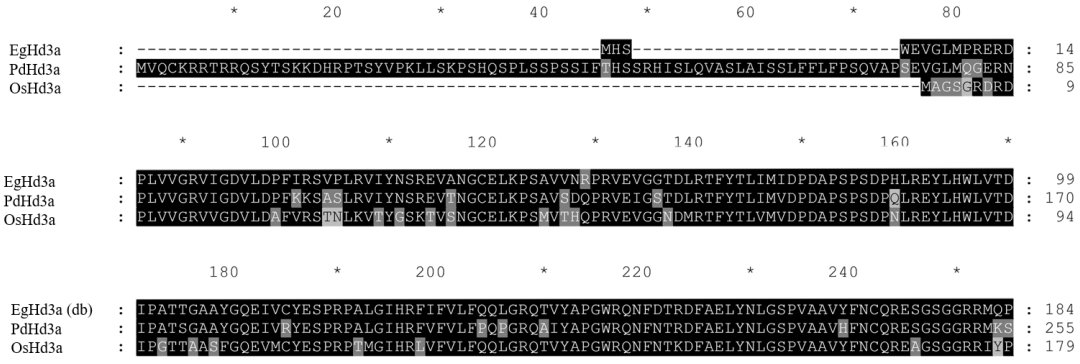
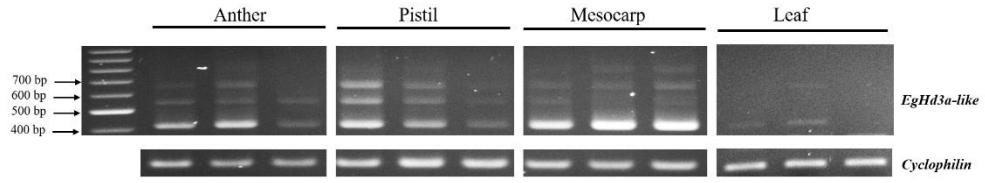
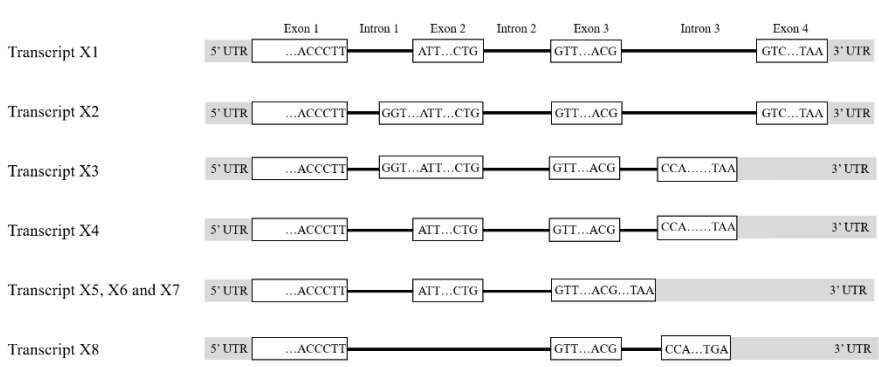


Figure 1. Sequence alignment of EgHd3a (Hd3a from *E guineensis*), PdHd3a (Hd3a from *Phoenix dactylifera*) and OsHd3a (Hd3a from *Oryza sativa*). White letters on a dark background represent amino acids with complete homology, while those positions sharing similar but non-identical amino acids are shadowed in gray.



Transcript variants	Expected size
Transcript X1	412 bp
Transcript X2	436 bp
Transcript X3	520 bp
Transcript X4	496 bp
Transcript X5	543 bp
Transcript X6	459 bp
Transcript X7	640 bp
Transcript X8	434 bp

Figure 2. Expression of alternatively spliced transcripts of *EgHd3a-like* in anther, pistil, mesocarp and leaf tissues in biological triplicate samples. *Cyclophilin*, which was used as a control, is shown in the bottom panel.



Peptide length	Expected PCR product size
184 aa	412 bp
192 aa	436 bp
118 aa	520 bp
110 aa	496 bp
120 aa	543 bp, 459 bp, 640 bp
99 aa	434 bp

Figure 3. Sequence variation of *EgHd3a-like* alternatively spliced transcripts. Sequences in boxes indicate the coding sequence region of each exon, whereas lines indicate each intron. Eight different transcripts encoding 6 isoforms are designated as transcripts X1–8. The numbers of corresponding amino acids and expected sizes of PCR products are indicated at the end of each transcript.

During the analysis of the cDNA clones, evidence for alternatively spliced transcripts was found. To demonstrate *in vivo* expression of all transcript forms, RT-PCR was employed. Five transcript variants that had different product sizes were observed by agarose gel electrophoresis. To confirm that the transcript variants were produced by the same gene, the RT-PCR products were cloned and sequenced. The sequencing of sampled transcript clones revealed that at least eight *EgHd3a-like* transcripts with the PCR product sizes of 412 bp, 434 bp, 436 bp, 459 bp, 496 bp, 520 bp, 543 bp and 640 bp, encoding putatively proteins of sizes 184 aa, 99 aa, 192 aa, 120 aa, 110 aa, 118 aa, 120 aa, and 120 aa, respectively, identified as in Figure 3. The transcript type X1 (184 aa) acted as the major type based on the intensity of RT-PCR products (in comparison with an internal marker) and the number of sequencing clones. From the results, only transcript type X2 (192 aa) was able to produce a longer polypeptide than transcript type X1(184 aa), while the others produced truncated proteins. An alignment of the coding variants and genomic sequences revealed that the transcript variants resulted from alternative splicing. All splicing junctions of *EgHd3a-like* contain a conserved splicing signal in agreement with the GT–AG type (Sheth *et al.*, 2006). The novel transcript variants have been reported to GenBank (Accession no. KX242327-KX242334). Alternative splicing can generate multiple mRNA variants from a single gene, and this phenomenon is common in some gene families (Barbazuk *et al.*, 2008). In plants, this phenomenon is known to play an important role in modulating gene expression and regulating plant processes (Reddy, 2007; Syed *et al.*, 2012). It is particularly prevalent in responses to abiotic and biotic stress and in circadian regulation (Simpson *et al.*, 2016). For flowering genes, alternative splicing of *FCA*, an ABA receptor with an RNA binding domain, was reported to control the developmental switch from the vegetative to the reproductive phase and is important in the transition to flowering (Reddy, 2007). For *FT* gene, Zhang *et al.* (2011) reported that 21 spliced forms were produced by alternative splicing of *PaFT1* and *PaFT2* in London plane (*Platanus acerifolia* Willd). In adult trees of London plane, levels of the alternative spliced variants were related to flower development stage and tree dormancy. Taken in its entirety, the variation of the transcripts might contribute to stress responses and to circadian regulation. In some cases, alternative splicing has resulted in premature termination and led to dominant-negative or deleterious gain-of-function activities of the truncated proteins (Chang *et al.*, 2007). This overall suggests that the variants are functionally important.

4. Conclusions

In conclusion, the promoter analysis revealed that several transcription factors likely affect the expression of *EgHd3a-like*. A better understanding of the regulation of *EgHd3a-like* expression could contribute to controlling floral and fruit developmental processes. This is the first report of alternatively spliced isoforms of *EgHd3a-like*. The variation of isoforms raises intriguing questions, notably whether the expression of transcript variants correlates with floral and fruit developmental stages. Further research could target a better

understanding of *EgHd3a-like* regulation: how to control and how to promote the expression of *EgHd3a-like*. These investigations will be beneficial for understanding the regulation of floral and fruit development, and for breeding applications.

Acknowledgements

This work was supported by the Strategic Scholarships Fellowships Frontier Research Networks (Specific for the Southern region) and the Office of the Higher Education Commission of Thailand. The Postdoctoral Fellowship from Prince of Songkla University, the Scholarship for Oversea Thesis Research Study (for Ph.D. candidates) and the scholarship to Mr. Li Xu, supported by the Graduated School, Prince of Songkla University were also acknowledged. Plant materials were provided by the Faculty of Natural Resources, Prince of Songkla University. The authors are thankful to Mrs. Anna Chatthong for improving the English of this manuscript.

References

Barbazuk, W. B., Fu, Y., & McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Research, 18*, 1381-1392. doi:10.1101/gr.053678.106

Bernard, V., Brunaud, V., & Lecharny, A. (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics, 11*, 1-15. doi:10.1186/1471-2164-11-166

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research, 42*, W252-W258. doi:10.1093/nar/gku340

Chang, Y. F., Imam, J. S., & Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry, 76*, 51-74. doi:10.1146/annurev.biochem.76.050106.093909

Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H. D., Lee, T. Y., . . . Chang, W. C. (2015). PlantPAN 20: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Research, 44*(Database issue), D1154–D1160. doi:10.1093/nar/gkv1035

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., . . . Yoshimura, A. (2004). Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes and Development, 18*, 926–936. doi:10.1101/gad.1189604

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., & Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature, 422*. doi:101038/nature01549

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., & Yano, M. (2002). *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant and cell physiology, 43*, 1096-1105. doi:10.1093/pcp/pcf156

Legros, S., Mialet-Serra, I., Caliman, J. P., Siregar, F. A., Clément-Vidal, A., & Dingkuhn, M. (2009). Phenology and growth adjustments of oil palm (*Elaeis guineensis*) to photoperiod and climate variability. *Annals of botany, 104*, 1171-1182. doi:10.1093/aob/mcp214

Murphy, D. J. (2014). The future of oil palm as a major global crop: Opportunities and challenges. *Journal of Oil Palm Research, 26 (1)*, 1-24.

Pin, P. A., & Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant, Cell and Environment, 35*, 1742-1755. doi:10.1111/j.1365-3040.2012.02558.x

Reddy, A. S. N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology,58*, 267-294. doi:10.1146/annurev.arplant.58.032806.103754

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research, 34*, 3955-3967. doi:10.1093/nar/gkl556

Simpson, C. G., Fuller, J., Calixto, C. P. G., McNicol, J., Booth, C., Brown, J. W. S., & Staiger, D. (2016). Monitoring alternative splicing changes in Arabidopsis circadian clock genes. In P. Duque (Ed.), *Environmental Responses in Plants: Methods and Protocols* (pp119-132). New York, NY: Springer.

Singh, R., Ong-Abdullah, M., Low, E. T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., & Sambanthamurthi, R. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature, 500(7462)*, 335-339. doi:10.1038/nature12309

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., & Brown, J. W. S. (2012). Alternative splicing in plants – coming of age. *Trends in Plant Sciences, 17*, 616-623. doi:10.1016/j.tplants.2012.06.001

Taoka, K. I., Ohki, I., Tsuji, H., Kojima, C., & Shimamoto, K. (2013). Structure and function of florigen and the receptor complex. *Trends in Plant Sciences, 18*, 287-294. doi:10.1016/j.tplants.2013.02.002

Tsuji, H., Tamaki, S., Komiya, R., & Shimamoto, K. (2008). Florigen and the photoperiodic control of flowering in rice. *Rice, 1*, 25-35. doi:10.1007/s12284-008-9005-8

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., & Coupland, G. (2006). CONS TANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell, 18*, 2971-2984. doi:10.1105/tpc.106.043299

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics, 13,* 59-69. Retrieved from http://www.nature.com/nrg/journal/v13/n1/suppinfo/nrg3095_S1.html

Zhang, J., Liu, G., Guo, C., He, Y., Li, Z., Ning, G., Shi, X., & Bao, M. (2011). The *FLOWERING LOCUS T* orthologous gene of *Platanus acerifolia* is expressed as alternatively spliced forms with distinct spatial and temporal patterns. *Plant Biology, 13,* 809-820. doi:10.1111/j.1438-8677.2010.00432.x

Appendix

Transcription factor family	Gene name	Hit Sequence	Position according to transcription start site	Strand	Associated conditions
Alpha-amylase	X16509	TATCCat	-5	+	Important for breakdown of endosperm starch during germination.
AP2;ERF	RAP2.2	ATCTA	-8	+	Involved in carotenoid biosynthesis regulation. Involved in ethylene response and resistance to necrotrophic pathogens. Acts as a downstream regulator in the ethylene signaling pathway.
AP2;RAV;B3	TEM1; EDF1	TGTTG	-772	-	Transcritptional repressor of flowering time on long day plants.
		CAACA	-171	+	
B3	Os04g0676600	GCATG	-602	-	Target barley iron deficiency-inducible gene IDS2
		CATGC	-546	+	
		CATGC	-58	+	
	LEC2	gcCATGCacaa	-60	+	Plays a central role in embryo development. Required for the maintenance of suspensor morphology, specification of cotyledon identity, progression through the maturation phase and suppression of premature germination. Ectopic expression is sufficient to promote somatic embryogenesis.
B3;ARF;	ARF11; ARF19;	gAGACA	-15	-	Modulate early auxin response genes expression. Involved in ethylene responses. Regulates lateral root formation through direct regulation of LBD16 and/or LBD29.
	IAA22	ctCGACAac	-393	+	
bHLH	MYC2; ATMYC2; RD22BP1; JAI1;JIN1; ZBF1	CACATg	-77	+	Common transcription factor of light, abscisic acid (ABA), and jasmonic acid (JA) signaling pathways.
	BIM1	CATTTg	-532	+	Positive brassinosteroid-signaling protein
		cATTTG	-532	-	
		CAACTg	-291	+	
		cAACTG	-291	-	
		CACATg	-77	+	
		cACATG	-77	-	

Transcription factor family	Gene name	Hit Sequence	Position according to transcription start site	Strand	Associated conditions
bZIP	AT1G77920; TGA8	TGACG	-367	+	Binding to the as-1-like cis elements mediate auxin- and salicylic acid-inducible transcription. May be involved in the induction of the systemic acquired resistance (SAR) via its interaction with NPR2
			-264	+	
C2H2	STZ; ZAT10	gACACTg	-781	+	Involved in abiotic stress responses. Probably involved in jasmonate (JA) early signaling response.
		gACACTa	-610	+	
CG-1;CAMTA	CAMTA3; SR4	GCGCG	-576	+	Regulates transcriptional activity in response to calcium signals.
		gCGCGT	-576	-	
		gCGCGTccg	-576	+	
		atcgCGCGT	-579	+	
Dof	AT1G29160	AAAGA	-821	+	Acts as a negative regulator in the phytochrome-mediated light responses. Controls phyB-mediated end-of-day response and the phyA-mediated anthocyanin accumulation. Not involved in direct flowering time regulation.
		TCCTT	-808	-	
		CCTTT	-807	-	
		ACTTT	-763	-	
		ACTTT	-758	-	
		TCCTT	-744	-	
		CCTTT	-743	-	
		ACTTT	-702	-	
		AAAGT	-674	+	
		AAGGG	-562	+	
		TCTTT	-487	-	
		TCTTT	-479	-	
		TCTTT	-475	-	
		TCTTT	-471	-	
		TCTTT	-464	-	
		TCTTT	-460	-	
		CCCTT	-446	-	
		CCTTT	-445	-	
		GCCTT	-435	-	
		AAAGC	-424	+	
		AAAGA	-380	+	
		AAAGA	-342	+	
		AAAGT	-334	+	
		TCCTT	-306	-	
		CCTTT	-305	-	
		AAAGG	-271	+	
		AAGGA	-270	+	
		AAAGG	-239	+	
		AAGGA	-238	+	
Dof	AT1G29160	AAGGG	-201	+	Acts as a negative regulator in the phytochrome-mediated light responses. Controls phyB-mediated end-of-day response and the phyA-mediated anthocyanin accumulation. Not involved in direct flowering time regulation.
		AAGGC	-157	+	
		GCTTT	-147	-	
		AAAGA	-119	+	
		ACTTT	-104	-	
		AAAGA	-51	+	
		AAAGA	-28	+	

Transcription factor family	Gene name	Hit Sequence	Position according to transcription start site	Strand	Associated conditions
Homeodomain;TALE	KNAT6;	TGTCA	-739	-	Plays a role in meristem function.
	KNAT6L;	AGTCA	-697	-	Contributes to the shoot apical
	KNAT6S	TGTCA	-409	-	meristem (SAM) maintenance and
		TGACA	-373	+	organ separation by controlling
		TGACG	-367	+	boundary establishment in embryo in
		TGTCA	-281	-	a CUC1, CUC2 and STMdependent
		TGACG	-264	+	manner. Involved in maintaining cells
		AGTCA	-219	-	in an undifferentiated, meristematic
		TGACA	-174	+	state.
Myb/SANT;		TGACC	-98	+	
ARR-B	ARR1; RR1	ttcTGTCAgc	-412	-	
	ARR2	cttGAATCtt	-420	+	
		ttgaATCTTc	-419	+	
		aAAGATactc	-119	-	
		tccaTATCTaaaa	-3	+	Functions as a response regulator
	ARR18;	AGATT	-819	+	involved in His-to-Asp phosphorelay
	RR18	AATCC	-810	-	signal transduction system.
	ARR14;	AATCC	-661	-	
	RR14	AATCT	-416	+	
		AATCT	-9	-	
Myb/SANT; G2-like;MYB	KAN2	cttgAATCTt	-420	+	Regulates lateral organ polarity. Promotes abaxial cell fate during lateral organd formation. Functions with KAN1 in the specification of
	KAN; KAN1	aatATATTcc	-44	-	polarity of the ovule outer integument.
		aataTATTCc	-44	+	Transcriptional repressor that regulates lateral organ polarity. Promotes lateral organ abaxial identity by repressing the adaxial regulator ASYMMETRIC LEAVES2 (AS2) in abaxial cells. Required for abaxial identity in both leaves and carpels. Functions with KAN2 in the specification of polarity of the ovule outer integument. Regulates cambium activity by repressing the auxin efflux carrier PIN1. Plays a role in lateral root formation and development.
Myb/SANT; MYB-related	EPR1; RVE7	acaATATCa	-720	+	Involved in phytochrome A-mediated cotyledon opening. Controlled by the central oscillator mediated by LHY and CCA1. Part of a regulatory circadian feedback loop.
MYB-related	CCA1	tccaTATCTa	-3	-	Involved in the circadian clock and in the phytochrome regulation.
MYB-related	RVE1	ccATATCtaa	-2	+	Morning-phased transcription factor integrating the circadian clock and auxin pathways. Does not act within the central clock, but regulates free auxin levels in a time-of-day specific manner. Positively regulates the expression of YUC8 during the day, but has no effect during the night. Negative regulator of freezing tolerance.

Transcription factor family	Gene name	Hit Sequence	Position according to transcription start site	Strand	Associated conditions
NAC;NAM	EMB2749; VND4; ANAC007; NAC007	actcACGCAa	-113	-	Transcription factor probably involved in xylem formation.
NAC;NAM	CUC3;	tCACGCaact	-111	-	Involved in molecular mechanisms
	NAC368; ANAC031	actCACGCaa	-113	-	regulating shoot apical meristem (SAM) formation during embryogenesis and organ separation. Required for axillary meristem initiation and separation of the meristem from the main stem. May act as an inhibitor of cell
NF-YB;	NF-YB4	ATTGG	-817	-	Component of the NF-Y/HAP
NF-YA;		CCAAT	-663	+	transcription factor complex. The NF-
NF-YC		CCAAT	-403	+	Y complex stimulates the transcription
		ATTGG	-353	-	of various genes by recognizing and
		CCAAT	-36	+	binding to a CCAAT motif in promoter
SBP	AtSPL3;	TGTAC	-692	-	Promotes both vegetative phase
	SPL3	GTACA	-691	+	change and flowering. Regulates
		TGTAC	-653	-	phase-specific patterns of leaf
		GTACG	-652	+	epidermal differentiation and
		GGTAC	-526	-	flowering time, but does not seem to
		GTACA	-525	+	affect leaf shape.
		AGTAC	-521	-	
		GTACT	-520	+	
TBP		ATATAtt	-43	-	
WRC;GRF	Transcription factor TFIID	tCTGAAac	-800	+	Transcription factor TFIID (or TATA-binding protein, TBP)
ZF-HD	AtGRF6; GRF6	ATAAT	-207	-	Transcription activator that plays a role in the regulation of cell expansion in leaf and cotyledons tissues
	AtHB33; HB33; ZHD5				Regulates floral architecture and leaf development. Regulators in the abscisic acid (ABA) signal pathway that confers sensitivity to ABA in an ARF2-dependent manner.