Isolation and characterization of LHY homolog gene expressed in flowering tissues of Tectona grandis (teak)

B. Norlia¹*, M. Norwati¹, A. Norwati¹, H. Mohd Rosli¹ and M. S. Norihan²

¹Genetic Unit, Forest Research Institute Malaysia (FRIM), 52109 Kepong, Selangor, Malaysia, Forest Biotechnology Division, Forest Research Institute Malaysia (FRIM), 52109 Kepong, Selangor, Malaysia.
²Cell and Molecular Biology Department, Faculty of Biotechnology and Biomolecular Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia.

Accepted 3 March, 2008

Floral initiation of teak through molecular biology approach is being studied for better understanding of teak flower development. Through PCR subtractive hybridization method, LHY homolog gene has been isolated from teak flowering tissues. The full-length cDNA of the gene was 2948 base pair (bp) and potentially encoded for 768 amino acids. It was named Tectona grandis LHY (Tg-LHY), as the gene was similar to the LHY gene of some species. Amino acid sequence alignment revealed that Tg-LHY was similar to LHY of Castanea sativa, LHY of Phaseolus vulgaris and LHY of Arabidopsis thaliana. The highly conserved region found in Tg-LHY was the MYB protein, which is the DNA-binding protein responsible in negative feedback loop reaction of central oscillator in plant circadian clock system. The level of gene expression was found to be high four hours after dawn in flowering shoots and flower. This paper reported the isolation and characterization of the gene.

Key words: LHY gene, circadian clock, floral gene, Tectona grandis.

INTRODUCTION

Flowering is a fundamental part of plants reproductive system. In teak, flowering is an important aspect in its growth and development. In natural habitat, teak flowers at the age of 5 - 6 years after planting (Boonkird 1966). However, teak planted in Malaysia under plantation condition has been observed to flower as early as 3 years after being planted (Krishnapillay, 2000). The early terminal flowering of teak causes forking and reduces vegetative growth due to energy utilization for the flowering process. This indirectly reduced the wood volume of the timber due to shorter and smaller clear bole harvested (Krishnapillay, 2000). Hence, in attempting to overcome this problem, study on floral initiation through molecular biology approach has been initiated in Forest Research Institute Malaysia (FRIM).

Flowering in teak is influenced by the environmental conditions, especially light (Nanda, 1962) and rain-fall (Tangmitcharoen, 1997; Palupi, 1996). Studies on environmental inputs that promote the floral transition showed that there were photoperiod, vernalization, light quality and availability of water and nutrients (Adams et al., 1997; Adams et al., 1998; Bernier et al., 1993). The understanding of floral transition is most advances in the environment cue photoperiod. Molecular genetic approaches have been applied to understand the control of flowering time, mainly in Arabidopsis, which recognised day length and promotes flowering under long day. A plant detects light duration via photoreceptor gene and couples the light accepted with an internal time or oscillator, which is known as circadian clock.

Circadian clock systems are complex signalling networks that allow organism to adjust cellular and physiological activities in anticipation of periodic changes.
in the environment. The systems governs many plant
processes including movement of organs like leaves and
petals, hypocotyls elongation, stomatal opening, expres-
sion of several genes, phosphorylation of some protein,
chloroplast movement, hormone production and flowering
time (Jarillo et al., 2004). Late elongated hypocotyls
(LHY) gene has been identified as one of the important
genes in the systems. It is one of the components in
central oscillator of the system that control the plant
clock genes, and is expressed in flowering tissues of
flowering shoots of Tectona grandis, which is a neutral
day plant is homolog gene expressed in flowering tissues of
Tectona grandis. The central oscillator of the system that
control the plant genes in the systems. It is one of the components in
Southern blotting and hybridization were conducted in order to
determine the copy number of
Tectona grandis Late Elongated Hypo-
cotyl (Tg-LHY) gene isolated. In this method, twenty micrograms of
genomic DNA extracted from young leaves were digested with Bam
HI, Eco RI, Hind III and Xho I enzymes. Each digestion mix
contained 20 µl of the appropriate 10 X enzyme reaction buffer and 50
units of restriction enzyme in a final volume of 200 µl. The
reactions were incubated at 37°C; after 2 h incubation an addition of
50 units restriction enzyme was made, mixed and further incubated
for overnight.

The standard Southern blotting method was carried out accord-
ing to Sambrook et al. (1989). The cDNA of Tg-LHY, which was
amplified using lhy1f and lhy2r primers, was used as probes during
hybridization. The probe was labelled with Digoxigenin-II-dUTP
using DIG-High Prime kit (Roche) according to procedure outlined
by the supplier. Hybridization and detection were conducted follow-
ing the procedure supplied in the DIG Starter Kit II (Roche).

Northern hybridization

Northern hybridization analysis of Tg-LHY was carried out using
radioactive labelled probe. The expression was examined on differ-
ent tissues of teak, which were flower, stem, leaf, vegetative shoot,
two-weeks-old seedling and root. Tissues from different stages of
flowering shoots were also examined, i.e. TFS1, TFS2, TFS3 and
TFS4. Since LHY gene has been reported as one of the circadian
clock genes, it could be interesting to investigate the rhythm of its
expression in teak; therefore TFS3 samples were collected every 2
h starting from 0600 to 1800 h. The signals were detected by
exposing the membrane to a FUJI Phosphor Imager plate over-
night. The image was scanned using FUJI FILM 3000
Phospholmager.

Probe was labeled by radioactive using Megaprime® DNA
labelling systems (Amersham Biosciences). In preparation, 25 ng of
Tg-LHY cDNA was used and made up to a final volume of 11 µl.
Each template cDNA was denatured by heating in boiling water for
10 min followed by rapid chilling in ice. The contents of the tube
were briefly spun to the bottom and 4 µl of High Prime solution (200
µl random primer mixture, 1 U/µl Klenow polymerase, 0.125 mM
dATP, 0.125 mM dGTP, 0.125 mM dTTP, 0.125 mM dCTP, 5X stabilized reaction
buffer in 50% (v/v) glycerol) and 5 µl of 50 Ci/mmol aqueous solution) were added. The contents were mixed
gently and briefly spun down. The mixture was then incubated at
37°C for at least 2 h and the labeling reaction was terminated by
addition of 2 µl of 0.2 M EDTA (pH 8.0). Before the probe was used,
the unincorporated nucleotides from the labeling reaction were
removed using a spin-column (Chroma Spin + TE10 Column,
Clontech).

Hybridization was conducted by pre-hybridizing the membrane in
hybridization buffer (0.5 M sodium phosphate, [pH 7.2], 7% [w/v]
SDS, 1 mM EDTA , 1% [w/v] BSA) added with 500 µl of 10 mg/ml
denatured herring sperm DNA (Invitrogen), for 4 h at 60°C. The
denatured probes were then added to the buffer and hybridized at
60°C overnight in hybridization oven. On the following day, the
membranes were washed twice in 40 mM sodium phosphate buffer
[pH 7.2], 1% (w/v) SDS for 10 min at room temperature. They were
then washed in 40 mM sodium phosphate (pH 7.2), 5% (w/v) SDS
at the hybridization temperature for 15 min. A series of stringency
washes in 2X SSC, 0.1% (w/v) SDS followed by 1X SSC, 0.1% (w/v) SDS were depending on the binding strength of the probes
to the templates. After the membranes were exposed overnight to a
FUJI Phosphor Imager plate, the images were scanned using FUJI
FILM 3000 Phospholmager.

RESULTS

PCR subtractive hybridization used to isolate genes ex-
pressed in flowering tissues stage 3 (TFS3) against vegetative shoots of teak has sequenced 94 clones. Out of 94 clones analysed, one clone was chosen for further analysis. The clone, TFS3-B7, which was about 891 bp in length, was showed to be 54% identical to LHY protein of *Castanea sativa*, 52% identical to Circadian Clock Associated 1 (CCA1) protein of *Mesembryanthemum crystal* and 56% identical to LHY protein of *Phaseolus vulgaris*. In order to obtain a full-length cDNA of TFS3-B7 fragment, specific primers were designed based on the sequence analysed. The primers, lhy1r (5'-ggcgaccttggatactga-3') and lhy1f (5'-caaaatccctgcagcctat-3'), were used to amplify the 5' and 3'-RACE of the TFS3-B7 fragment respectively. PCR product sizes of approximately 1.8 kb for 5'-RACE and 1.6 kb for 3'-RACE were isolated and cloned. BLASTX search on the sequence of both clones confirmed their identities as 5'- and 3'- end of LHY gene from a few species.

Full-length cDNA amplification was carried out using primers designed based on the non-coding regions from both ends of the cDNA fragment. The primers, lhy2f (5'-atctgattaagatcggaactt-3') and lhy2r (5'-ggtatcaacgcagagtacttt-3'), and the nested, lhy2nf (5'-ggctctctacaataacac-3') and lhy2nr (5'-tagcacacataactcacgaaa-3'), were used to amplify and confirm the full-length cDNA of TFS3-B7 fragment. A PCR product of approximately 2.3 kb was obtained (Figure 1) and subsequently cloned. To exclude sequencing errors, clone was sequenced in triplicate. When discrepancies occurred, the majority were determined to be the correct sequences. This full-length cDNA of TFS3-B7 was named *Tg-LHY* (*Tectona grandis* LHY).

The *Tg-LHY* cDNA clone obtained was 2303 bp long, while the complete cDNA (composite of the clone and the 5'- and 3'- untranslated regions obtained from 5'- and 3'-RACE) was 2948 bp. The 5'- non-coding region was 429 bp long and 3'- non-coding region was 219 bp upstream of the poly (A)+ tail. The *Tg-LHY* cDNA potentially encoded for 768 amino acids, which was 57% identical to LHY of *Castanea sativa*, 51% identical to CCA1 of *Mesembryanthemum crystal*, 52% identical to LHY of *Phaseolus vulgaris* and 43% identical to LHY of *Arabidopsis thaliana*.

The predicted amino acid sequences of *Tg-LHY* were compared with its homologs from other dicots to reveal conserved regions as well as variable ones. *Tg-LHY* shares a highly conserved region with all the LHY proteins at the N-terminus within lysine (K)-13 to threonine (T)-108. Within this region is a sequence similar to the repeat sequence highly conserved in Myb-related protein. The similarity showed the structure feature expected of Myb repeats in dicots, two of the three regularly spaced tryptophans present in most Myb repeats conserved in LHY gene, while the expected position of the third tryptophan was occupied by alanine in all the dicots analysed (Figure 2).

To ascertain the *Tg-LHY* copy number in teak genome, genomic DNA was digested and probed at high stringency with the DIG labeled 800 bp of 5'-end *Tg-LHY* cDNA. A single hybridizing restriction fragment was detected in each lane of genomic DNA digested with restriction enzymes *Bam* H1, *Eco* R1, *Hind* III and *Xho* I (Figure 3). All the restriction enzymes used had no cleavage site in the probe fragment, indicating that *Tg-LHY* cDNA was a product of a single copy gene.

Northern hybridization analysis showed that the expression of *Tg-LHY* was detected higher in the flower but lower in the leaf, 2-weeks-old seedling and root. No expression was evident in stem and vegetative tissues (Figure 4 [A]). A faint signal was detected in TFS1, TFS2 and TFS4 tissues collected at 0900 hr. In TFS3 tissues, transcription of *Tg-LHY* was detected higher at 0800 hr but lower at 0600 and 1000 hr. No transcription was evident on other tested tissues, which were collected at the times indicated (Figure 4[B]).

**DISCUSSION**

*Tg-LHY* was predicted to encode a DNA-binding protein related to MYB transcription factors. The 43 amino acids region of *Tg-LHY* showed strong homology to DNA-binding domains of previously reported MYB proteins and several of the feature characteristics of a MYB repeat.
<table>
<thead>
<tr>
<th>Plant</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. grandis</em></td>
<td>KTRKPYTITKGERWTEEHNRFLEALKLYGRAWQRIEEHIGTKTAHV</td>
</tr>
<tr>
<td><em>M. crystallinum</em></td>
<td>KTRKPYTITKGERWTEEHNRFLEALKLYGRAWQRIEEHIGTKTAHV</td>
</tr>
<tr>
<td><em>C. sativa</em></td>
<td>KAKRKPYTITKGERWTEDEHNRFLEALKLYGRAWQRIEEHIGTKTAHV</td>
</tr>
<tr>
<td><em>P. vulgaris</em></td>
<td>KTRKPYTITKGERWTEEHNRFLEALKELYGRAWQRIEEHIGTKTAHV</td>
</tr>
<tr>
<td><em>A. thaliana</em></td>
<td>KAKRKPYTITKGERWTEDEHNRFLEALKLYGRAWQRIEEHIGTKTAHV</td>
</tr>
</tbody>
</table>

**Figure 2.** CLUSTALW analysis of Tg-LHY protein at the highly conserved region (K-13 to T-108) with LHY protein from other dicots. The other dicots obtained from genebank are *M. crystallinum* (accession No. AY371287.1), *C. sativa* (accession No. AY611029.1), *P. vulgaris* (accession No. AJ420902.2) and *Arabidopsis thaliana* (accession No. NM179237.1). Arrows indicate the beginning and end of conserved MYB protein domains and asterisks indicate positions of respective tryptophan or alanine. Black boxes indicate single fully conserved residues and grey boxes indicate conservation of strong groups. Those without colour indicate no consensus.

**Figure 3.** Southern hybridization analysis of Tg-LHY in teak genomic DNA. 20 µg of teak genomic DNA was digested with (1) *Bam*HI, (2) *Eco* RI, (3) *Hind* III and (4) *Xho* I. The digests were run alongside a DNA marker of GeneRuler Ladder Mix (Fermentas).

MYB proteins were first isolated from animals and reported to play important roles in controlling regulatory decisions during proliferation and differentiation (Weston, 1998). The conserved MYB DNA-binding domain generally comprises up to three imperfect repeats, which are referred to as 1R, 2R and 3R. In each MYB repeat, three regularly spaced tryptophan residues form a tryptophan cluster in the three-dimensional helix-turn-helix structure of about 53 amino acids (Ogata et al., 1992).

MYB proteins isolated from plants were present in three, two or single repeat. The proteins are classified into three subfamilies depending on the number of adjacent repeats in the MYB domain. The subfamilies are termed as MYB1R, MYB2R3R and MYB3R factors for MYB with one repeat; two and three repeats respectively (Stracke et al., 2001). 2R3R-type MYB factor is the largest subfamily characterized in plants, and they are estimated to be over 100 members in *Arabidopsis* (Jin and Martin, 1999). The MYB1R factor, however, is a class of expanding importance in plants and has been shown to act as transcriptional activator (Baranowskij et al., 1994) and some are associated closely with the activity of circadian clock (Carre and Kim, 2002).

The Tg-LHY isolated from teak consisted of a single MYB repeat and therefore was predicted to be in a MYB1R subfamily. The predicted amino acid sequence of the Tg-LHY protein within the MYB domain was 100% identical to CCA1 of *Mysembranthemum crystallinum*, 97% identical to LHY of *Castanea sativa* and *Phaseolus vulgaris*; and 95% identical to LHY of *Arabidopsis thaliana*. The sequence of the MYB domain conserved in all these genes are from Lys(K)-13 to K-107. The domain has a structure expected of MYB repeats: two of the three regular tryptophans present in most MYB repeats are conserved in all the sequences while alanine is replacing the third tryptophan residue (Ogata et al., 1994). The high degree of homology shown by these genes suggested that they were particularly closely related proteins.

The structural characteristic common to all known MYB proteins is the DNA-binding domain, which binds to DNA in a sequence-specific manner. The DNA binding domain of MYB proteins in Tg-LHY showed a strong homology to the DNA binding domain of LHY and CCA1 gene of...
Figure 4. Northern analysis of Tg-LHY on different teak tissues with radioactive labelled probe. Each lane contains 20 µg of total RNA. (A) Hybridization on different organ tissues, (1) flower, (2) stem, (3) leaf, (4) vegetative shoot, (5) root and (6) 2-weeks-old seedling. (B) Hybridization on different flowering tissues and time harvested of TFS3, (1) TFS-1, (2) TFS-2, (3) TFS-4, (4 to 10) are TFS-3 tissues harvested at different time, 0600, 0800, 1000, 1200, 1400, 1600 and 1800 hr, respectively. The respective RNA stained with Etbr was used as a control for equal loading.

Arabidopsis. The high similarity within the DNA binding domain of these genes suggests that they share a functional similarity. Even though these genes encoded a single MYB repeat instead of two or three as in most MYB proteins, they exhibited specific DNA binding activities as in other MYB proteins (Baranowskij et al., 1994; Wang et al., 1997). The MYB binding domain of CCA1 has been reported bound to an element (consensus AAAaa/cAATCT) found in the promoter of Arabidopsis light-harvesting chlorophyll a/b protein (CAB) gene (Wang et al., 1997). The related consensus, AAAATATCT, is also found in the putative promoters of 31 evening-specific genes identified by microarray analysis (Harmer et al., 2000). Mutation of this element eliminates evening specific transcription, whereby as little as a single nucleotide change in a consensus element might result in transcription nearly out of phase (Harmer et al., 2000).

Southern hybridization analysis, using the whole fragment of Tg-LHY cDNA as a probe to hybridize genomic DNA digested with Bam HI, Eco RI, Hind III and Xho I detected only a single band. This indicates that the cDNA sequence of Tg-LHY in teak is a product of a single copy gene. The single copy LHY gene has also been reported in Arabidopsis (Schaffer et al., 1998). The expression of Arabidopsis LHY gene controlled by the circadian clock, whereby the rhythm expression persisted for at least three days under constant light or darkness (Schaffer et al., 1998). Plants lacking LHY and CCA1 functional genes have their rhythmic expressions abolished (Alabadi et al., 2002). Since the constitutive expression is correlated with arrhythmic phenotypes, the rhythmic expression of LHY together with CCA1 was suggested to be required for the expression of circadian rhythmicity in Arabidopsis (Carre, 2002).

Tg-LHY was predicted to be associated with the central oscillator of the circadian clock, as what has been reported in LHY of Arabidopsis by Alabadi et al. (2002). In Arabidopsis, LHY together with CCA1 gene was reported to fulfill some of the required criteria for oscillator components such as robust circadian oscillation of transcript and protein levels; control their own levels by feedback inhibition of its synthesis and mutant of these two genes stops all overt rhythmicity measured (Alabadi et al., 2002). LHY and CCA1 proteins were proposed to act redundantly in the early day, binding to an evening element (AAAATATCT) in the TOC1 and repressing its expression (Alabadi et al., 2001). When LHY/CCA1 levels fell late in the day, the TOC1 protein is proposed to activate the transcription of LHY/CCA1, thus forming the outline of a transcription feedback loop (Alabadi et al., 2001). Although further studies need to be done to ascertain the exact function of Tg-LHY in teak, it is reasonable to conclude that Tg-LHY is active in teak flo-
wering shoots stage 3 (TFS3) and might be involved in the early flower development of teak.

Transition to flowering is promoted by a variety of environmental and endogenous signals through the activation of the floral pathway integrators. The most advanced understanding of the environmental cues to floral promotion involves photoperiod. The change in a photoperiod in plant is detected and regulated by a circadian system. Tg-LHY, which was isolated from flowering tissues of teak, is probably involved in the circadian clock towards flowering development in teak. In Arabidopsis, mutation of LHY gene was shown to affect flower development, in LHY over expressed in constant light, the flower forming later than the wild type (Schaffer et al., 1998). The expression of Tg-LHY in floral tissues of teak indicated that flowering in teak was regulated through the circadian clock gene system. However, further investigation need to be carried out to ascertain the assumption.

Photoperiod is most likely not an environmental cue detected by the circadian clock system of teak, which grows in normal day environment. Hence, five years data of light intensity and temperature collected at Chupin Perlis, which is the nearest meteorological station to the sample collection site were analysed (data not shown). The analysis showed that temperature or light intensity might be a potential environmental factor that is detected by the circadian clock of teak. Recently in Arabidopsis, few studies have related flowering time to ambient temperature. A study of wild plants and different flowering time mutants of Arabidopsis grown under constant 23 or 16°C has suggested the role of autonomous pathway in sensing changes in ambient growth temperature that can affect the expression of FT gene (Blazquez et al., 2003). Another study by Halliday et al. (2003) has revealed that the early-flowering phenotype of the phyB mutant of Arabidopsis is temperature dependent, occurring at 22°C but not at 16°C. This temperature-sensitive phytochrome control of flowering also seems to operate through regulation of FT. The preliminary analysis carried out in teak suggested that higher temperature, which is due to higher light intensity, might be the cue perceived by the input pathway of the teak circadian clock and could further on triggered the integrator gene towards flower development. However, there is a need for further analysis to ascertain the effect of temperature and light intensity on the flower development of teak.

Circadian clock system in plants has also been reported to regulate many other processes in Arabidopsis, such as leaf movement (Engelmann et al., 1992), hypocotyl elongation (Dowson-Day and Millar, 1999), stomata opening (Somers et al., 1998), stem elongation (Jouve et al., 1998) and a few gene expressions (Harmer et al., 2000). Tg-LHY isolated might also function in other processes, as Northern analysis showed that the expression of Tg-LHY was also detected in leaf and floral tissues. The circadian clock in different tissues has been observed to act independently, but share common components of clock mechanism (Thain et al., 2000; Hall et al., 2002). Genes with different rhythm periods were also reported to control by different circadian clocks (Thain et al., 2002).

The molecular biology of flowering has been intensively studied in herbaceous model system. Many of proposed flowering pathways are being constructed and refined with the discoveries of new genes and their interactions. However, studies on tree species, especially tropical timber species, are very scarce. Thus, in the effort to understand the flowering process in an economically important tropical timber species, some preliminary has been established for teak in this study.

REFERENCES


