Sains Malaysiana 47(8)(2018): 1709–1723
http://dx.doi.org/10.17576/jsm-2018-4708-10

Cloning and Analysis of the *Eg4CL1* Gene and Its Promoter from Oil Palm (*Elaeis guineensis* Jacq.)

(Pengklonan dan Analisis Gen *Eg4CL1* dan Promoternya daripada Kelapa Sawit (*Elaeis guineensis* Jacq.))

YUSUF CHONG YU LOK, IDRIS ABU SEMAN, NOR AINI AB SHUKOR,
MOHD NORFAIZULL MOHD NOR & MOHD PUAD ABDULLAH*

ABSTRACT

The empty fruit bunches of oil palm have been used as the raw material to produce biofuel. However, the lignin present in oil palm tissues hampers the enzymatic saccharification of lignocellulosic biomass and lower the yield of biofuel produced. Hence, various efforts were taken to identify the lignin biosynthetic genes in oil palm and to investigate their regulation at the molecular level. In this study, a lignin biosynthetic gene, Eg4CL1 *and its promoter were isolated from the oil palm. Eg4CL1 contains the acyl-activating enzyme consensus motif and boxes I & II which are present in other 4CL homologs. Eg4CL1 was clustered together with known type I 4CL proteins involved in lignin biosynthesis in other plants. Gene expression analysis showed that* Eg4CL1 *was expressed abundantly in different organs of oil palm throughout the course of development, reflecting its involvement in lignin biosynthesis in different organs at all stages of growth. The presence of the lignification toolbox - AC elements in the 1.5 kb promoter of* Eg4CL1 *further suggests the potential role of the gene in lignin biosynthesis in oil palm. Together, these results suggested that* Eg4CL1 *is a potential candidate gene involved in lignin biosynthesis in oil palm.*

Keywords: Biofuel; lignin; oil palm; promoter; 4CL

ABSTRAK

Tandan kosong buah kelapa sawit telah digunakan sebagai bahan asas untuk menghasilkan biofuel. Walau bagaimanapun, lignin yang terdapat dalam tisu kelapa sawit menghalang proses sakarifikasi enzimatik biojisim lignoselulosa dan mengurangkan hasil bahan api biologi yang dihasilkan. Oleh itu, pelbagai usaha telah diambil untuk mengenal pasti gen biosintesis lignin dalam kelapa sawit dan untuk mengkaji pengawalaturannya pada peringkat molekul. Dalam kajian ini, gen biosintesis lignin, Eg4CL1 *dan promoternya telah dipencilkan daripada kelapa sawit. Eg4CL1 mengandungi motif konsensus enzim pengaktifan asil dan kotak I & II yang terdapat dalam homolog 4CL yang lain. Eg4CL1 berkelompok bersama dengan protein 4CL yang diketahui terlibat dalam biosintesis lignin dalam tumbuhan lain. Analisis pengekspresan gen menunjukkan bahawa* Eg4CL1 *diekspres dengan banyak dalam organ kelapa sawit yang berbeza pada semua peringkat pertumbuhan, mencerminkan penglibatannya dalam biosintesis lignin dalam organ yang berbeza pada semua peringkat pertumbuhan. Kehadiran peti alat lignifikasi - unsur AC dalam promoter* Eg4CL1 *1.5 kb selanjutnya menyokong potensi gen ini yang berperanan dalam biosintesis lignin pada pokok kelapa sawit. Secara keseluruhannya, keputusan kajian ini mencadangkan* Eg4CL1 *sebagai calon gen yang berpotensi terlibat dalam biosintesis lignin pada pokok kelapa sawit.*

Kata kunci: Biofuel; kelapa sawit; lignin; promoter; 4CL

INTRODUCTION

The oil palm (*Elaeis guineensis* Jacq.) is widely cultivated in many countries including Malaysia, Indonesia, Central America and Sri Lanka. It was primarily cultivated as a source of edible oil. Apart from the production of edible oil, the empty fruit brunches have been utilized as feed stock for biofuel production (Ibrahim et al. 2015; Piarpuzán et al. 2011). Production of biofuel from empty oil palm fruit brunches involves the enzymatic saccharification of the lignocellulosic biomass to produce fermentable sugars. However, the hydrolysis of the lignocellulose is hindered by the presence of lignin in the oil palm biomass (Gao et al. 2014).

Lignin is the most abundant polymer in plants following cellulose. It is produced through the polymerization of monolignols including the hydroxycinnamyl alcohols, coniferyl alcohol and sinapyl alcohol (Vanholme et al. 2010). These monolignols are synthesized through the phenylpropanoid pathway which consists of three main enzymes, namely phenylalanine ammonia-lyase (PAL; EC 4.3.1.5), cinnamate 4-hydroxylase (C4H; EC 1.14.13.11) and 4-coumarate: coenzyme A ligase (4CL; EC 6.2.1.12). Being an enzyme located at the branching point of the phenylpropanoid pathway, 4CL regulates the flux of the carbon from the general phenylpropanoid pathway into different branch pathways.

4CL catalyzes the conversion of hydroxycinnamates to its corresponding CoA esters, to produce the precursors required for the biosynthesis of a couple of important secondary metabolites, including lignin, flavonoids and stilbenes. In plants, 4CL is encoded by a gene family with varying numbers of family members as observed in different species, for instance, four copies of the *4CL* gene are present in *Arabidopsis thaliana* (Soltani et al. 2006) and *Physcomitrella patens* (Silber et al. 2008), while five copies are in the *Oryza sativa* and *Populus trichocarpa* genomes (Gui et al. 2011; Hamberger et al. 2007). Within the *4CL* gene family, functionally divergent members have been identified in many species such as *Pueraria lobata* (Li et al. 2014), *Arabidopsis thaliana* (Ehlting et al. 1999) and *Populus tremuloides* (Hu et al. 1998). Basically, there are two types of *4CL* genes in plants, designated as type I and type II. The type I *4CL* genes are responsible for lignification while the type II *4CL* genes are involved in the biosynthesis of flavonoid compounds. Suppression of the type I *4CL* genes led to a significant reduction in lignin content in different plants (Gui et al. 2011; Xu et al. 2011a), while overexpression increased the lignin content (Rao et al. 2015). On the other hand, suppression of the type II *4CL* gene only led to a reduction in eugenol content in *Ocimum sanctum*, without affecting the lignin content (Rastogi et al. 2013). Judging by their peptide sequences, type II 4CLs are different from type I 4CLs owing to the presence of additional amino acid residues at the N-terminal regions of type II 4CLs. Nevertheless, both types of 4CL share common conserved motifs namely Box I 'SSGTTGLPKGV' and Box II 'GEICIRG' (Heath et al. 2002; Kumar & Ellis 2003; Li et al. 2014).

Characterization and study of the lignin biosynthetic genes allow one to identify the lignin regulatory switch, subsequently opening the gateway to manipulate the lignin content in plants. The lignin biosynthetic genes have been well studied in many other species such as rice, poplar and switchgrass (Gui et al. 2011a; Voelker et al. 2010; Xu et al. 2011a), but very little information of lignin biosynthesis in oil palm is available. Hence, the oil palm *4CL*1 gene and its promoter was isolated in the present study and its expression pattern during oil palm development was evaluated. This study provided a gateway for a better understanding of lignin biosynthesis in oil palm.

Materials and Methods

Plant Material

Oil palm (*Elaeis guineensis* Jacq., variety *pisifera*, 367 P) leaf samples obtained from the Malaysian Palm Oil Board (MPOB) Kluang Research Station in Johor, Malaysia were used for the gene and promoter isolation. Various samples of *Elaeis guineensis* Jacq., variety *tenera* (*dura* × *pisifera* hybrid, 0.409) collected from the MPOB/UKM Research Station and Universiti Putra Malaysia in Selangor, Malaysia were used in the gene expression analysis.

Nucleic Acid Extraction and cDNA Synthesis

Genomic DNA was extracted from oil palm root tissues using Carroll's method (Carroll et al. 1995) with minor modifications. Total RNA was isolated from the oil palm tissue according to the method described by Wang et al. (2005). The 5'-RACE cDNA template was synthesized using SMARTer RACE cDNA Amplification Kit (Clontech, USA) according to the manufacturer's instructions. The first strand cDNA used in other work was synthesized using Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, USA).

Gene Isolation

The nucleotide sequences of the *4CL* gene homologues from different plant species were retrieved from GenBank (NCBI, http://www.ncbi.nlm.nih.gov). The gene sequences were aligned by using the Clustal W method to determine the conserved or highly-similar regions of the *4CL* gene. A degenerate primer (4CL-R primer: 5'- CCC TTG TAY TTG ATG AKC TCC TT -3') was designed to bind to a specific conserved or highly similar region of the *4CL* genes. The *Eg4CL1* gene fragment was amplified using the 5'-RACE approach. The PCR was performed in a reaction volume of 50 µL containing 1× *Taq* Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM 4CL-R primer, 1× Universal Primer A Mix (supplied in the SMARTer RACE cDNA Amplification Kit), 250 ng 5'-RACE template, 1.25 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR thermal cycling profile used was 95°C (3 min), followed by 95°C (25 s), 56°C (30 s), 72°C (50 s) for 30 cycles and 72°C (5 min).

Subsequently, the 3'-RACE method was performed to obtain the full-length cDNA sequence of *Eg4CL1*. The 50 µL PCR mixture comprised of 1× *Taq* Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM 4CL1 3'-RACE primer (5'- CTG AGA CTG GAC TAT CAC TGC CT -3'), 0.2 µM Oligo d(T)-adaptor primer (5'- GGC CAC GCG TCG AGT $AC(T)_{18}$ -3'), 250 ng cDNA, 1.25 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR was run at 95°C (3 min); 35 cycles of 95°C (25 s), 60°C (30 s), 72°C (50 s); 72°C (5 min).

The full-length sequence of the *Eg4CL1* cDNA was further verified using a high fidelity DNA polymerase with proofreading activity. The PCR mixture contained 1× Phusion HF Buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM Eg4CL1-F primer (5'- GAG ACA AGA GAA TTG AAC CA -3'), 0.3 µM Eg4CL1-R primer (5'- GGA TGG TCT CAT CCA CTT T -3'), 250 ng cDNA, 1 unit Phusion DNA Polymerase (Thermo Scientific, USA) and dH_2O in a total reaction volume of 50 µL. The PCR thermal cycling profile used was 98°C (30 s), followed by 98°C (10 s), 55°C (20 s), 72°C (40 s) for 35 cycles and 72°C (5 min).

The targeted PCR products were excised and purified from the agarose gel using the QIAquick Gel Extraction kit (QIAGEN, Germany). The purified PCR products were cloned into the pGEM-T Easy vector (Promega, USA) and sequenced by First BASE Laboratories Sdn Bhd (Selangor, Malaysia).

ISOLATION OF *Eg4CL1* PROMOTER

The promoter of *Eg4CL1* was isolated from the oil palm genome by using the inverse-PCR method described by Ochma et al. (1988). The first portion of the promoter was isolated from the self-ligated gDNA template prepared from the gDNA double digested with *Hin*dIII and *Xba*I. The inverse-PCR was carried out in a 20 µL PCR mixture containing 1× *Taq* Buffer, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM 4CL1pro-F1 primer (5'- CGC CTA CGG AGG CGA CCA TCT TC -3'), 0.3 µM 4CL1pro-R1 primer (5'- GGG CCC ATC GCA ATC AAT CGT TTA -3'), 8 ng ligated DNA, 0.5 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR thermal cycling profile was as follow: 95°C (3 min), followed by 95°C (25 s), 62°C (25 s), 72°C (80 s) for 40 cycles and 72°C (5 min).

The second portion of the promoter was isolated by using the self-ligated template derived from the gDNA digested with *Nco*I. The 20-µL PCR mixture comprised of 1× *Taq* Buffer, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM 4CL1pro-F2 primer (5'- ATC AAT CAC CAA CCA AGA CGC C -3'), 0.3 µM 4CL1pro-R2 primer (5'- GCG TGA TCG GAT GGA CAA AGT T -3'), 8 ng ligated DNA, 0.5 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR was performed at 95°C (3 min); 40 cycles of 95°C (25 s), 58°C (25 s), 72°C (60 s); 72°C (5 min).

To verify the sequence of the *Eg4CL1* promoter, the promoter region was amplified with a high fidelity DNA polymerase. The PCR was performed in 50 µL containing 1× Phusion HF Buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM pro4CL1-F primer (5'- CCA TGG TGT GAC CAC GGA A -3'), 0.3 µM pro4CL1-R primer (5'- CGC AAT CAA TCG TTT AGA GAG AAA -3'), 200 ng gDNA, 1 unit Phusion DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR thermal cycling profile used was 98°C (30 s), followed by 98°C (10 s), 65°C (20 s), 72°C (45 s) for 35 cycles and 72°C (5 min).

IN SILICO ANALYSIS

The molecular weight and the theoretical isoelectrical point (pI) of Eg4CL1 were predicted using the ProtParam tool at the ExPASy website (web.expasy.org/protparam). The protein domains in Eg4CL1 were searched for in the NCBI conserved Domain databases (Marchler-Bauer et al. 2010) and PROSITE. The multiple sequence alignment was performed using the ClustalW method in BioEdit version 7.0 (Hall 1999). The 4CL amino acid sequences used in the multiple sequence alignment including Pto4CL1 (AAL02145), Lp4CL2 (AAF37733), Os4CL1 (NP_001061353) and At4CL2 (NP_188761) were retrieved from the NCBI databases. The Eg4CL1 protein structure homology modeling was performed by the SWISS-MODEL using the *Populus tomentosa* 4CL1 (3ni2A) (Hu et al. 2010) as a template. The phylogenetic tree was reconstructed using the Neighbor-Joining method with bootstrap values set to 1000 in the MEGA5 software (Tamura et al. 2011). The 4CL amino acid sequences used for the phylogenetic analysis and the NCBI accession numbers of the sequences are presented in Supplementary Table 1. The *cis*-acting elements present on the promoter sequence of *Eg4CL1* were searched from the PlantCARE online database (Lescot et al. 2002) and other literatures.

GENE EXPRESSION ANALYSIS

A two-step RT-PCR was carried out to investigate the expression behaviors of the *Eg4CL1* gene in several organs including coleoptile and root of the germinating seeds; young leaf and young root of one-year-old palms; and immature fruitlet and young fruit of mature oil palms. The oil palm *GAPDH* gene (accession number: DQ267444) was used as the internal control for the analysis. A 20 µL PCR mixture consisting of 1× *Taq* Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM forward primer, 0.2 µM reverse primer, 100 ng cDNA, 0.5 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O was prepared. The primers for this analysis are listed in Table 1. The RT-PCR was performed as follows: 95°C (3 min); 95°C (20 s), 58°C (25 s), 72°C (25 s) for 28 cycles; 72°C (5 min).

RESULTS

GENE ISOLATION

A full-length cDNA encoding for 4-Coumarate:Coenzyme A Ligase was isolated from the oil palm genome and deposited in GenBank under accession number KM234973. Since this is the first isolation and study of the *4CL* gene in oil palm, this gene was designated as *Eg4CL1*. The *Eg4CL1* cDNA was 1946 bp long and contained a 1623 bp open reading frame, flanked by a 5'-UTR of 55 bp and a 3'-UTR of 268 bp. The putative plant polyadenylation signal (5'-AATAAA-3') was found in the 3' UTR, located 5' upstream of the poly-A tail. The deduced translation product of *Eg4CL1* consisted of 540 amino acids with a predicted molecular weight of 58.64 kDa and a theoretical isoelectric point of 5.67.

TABLE 1. Primers used in gene expression analysis

Gene	Forward primer sequences (5'-3')	Reverse primer sequences (5'-3')	Product size (bp)
Eg4CL1	GGCATTTGTCGTGCGATCAAGT	GCACAACACATAGGCAAAGGCA	291
GAPDH	GTGGGTGTGAACGAGCATGAATA	AGCTTTCCATTTAAGGCAGGAAG	288

IN SILICO ANALYSIS

The *Eg4CL1* gene exhibited high similarity with the *4CL* from angiosperms, especially those from monocots. The BLAST in NCBI showed that *Eg4CL1* shared high identities with several members of the *4CL* gene family from *Phoenix dactylifera* (date palm) and *Musa acuminata* (banana). The highest identity was shown by the date palm *4CL2*-like gene (LOC103718393) at 93%, followed by another date palm *4CL2*-like gene (LOC103707092) at 78% identity. *Eg4CL1* also shared 77% and 76% identities with banana *4CL3* (LOC103998718) and banana *4CL2* like-gene (LOC103973145), respectively, at the nucleic acid level. The multiple sequence alignment of the type I 4CL amino acid sequences showed that Eg4CL1 shared certain conserved regions with the 4CLs from the other plants (Figure 1). Notably, several protein domains including the box I (SSGTTGLPKGV) and box II (GEICIRG) motifs found in other plant 4CLs were also present in the Eg4CL1 (Figure 1). The active site residues (marked with blue dots) previously identified in the 4CL1 of *Populus tomentosa* (described as Pto4CL1 in this paper) based on its crystal structure and mutagenesis experiments are identical with Eg4CL1 (Figure 1). However, there are some variations in the substrate binding pocket residues (marked with green triangles) between Pto4CL1 and Eg4CL1. The Lys^{303} and Gly^{306} residues in the substrate binding pocket of Pto4CL1 appeared as Met^{303} and Ala^{306} in Eg4CL1. Hence, it is speculated that Eg4CL1 might show different preferences of substrates and catalytic efficiencies towards particular substrates compared to Pto4CL1. By using the conserved domain search in the Conserved Domain Database, the active site of 4CL, AMP binding site, putative CoA binding site and acyl-activating enzyme consensus motif were detected in Eg4CL1. Moreover, the putative AMP-binding domain signature with the consensus sequence LPYSSGTTGLPK was detected by PROSITE. Together, the results of the analysis above support that *Eg4CL1* is the putative *4CL1* gene in oil palm.

PROTEIN STRUCTURE

Apart from the amino acid sequence analysis, computer predictions of the secondary structure and the three-dimensional structure of the Eg4CL1 protein were also performed. The Self-Optimized Method with Alignment (SOPMA) website in ExPASy showed that the Eg4CL1 protein predominantly consisted of random coil (42.22%), followed by alpha helix (31.48%) and extended strand (18.70%), while the beta turn only contributed 7.59%. The three-dimensional structure of the Eg4CL1 protein was similar to that of Pto4CL1, which comprised of the larger N-domain and the C-domain (Figure 2). The catalytic residues of the 4CL gene are located within the C-domain (yellow). The N-domain which contains the substrate

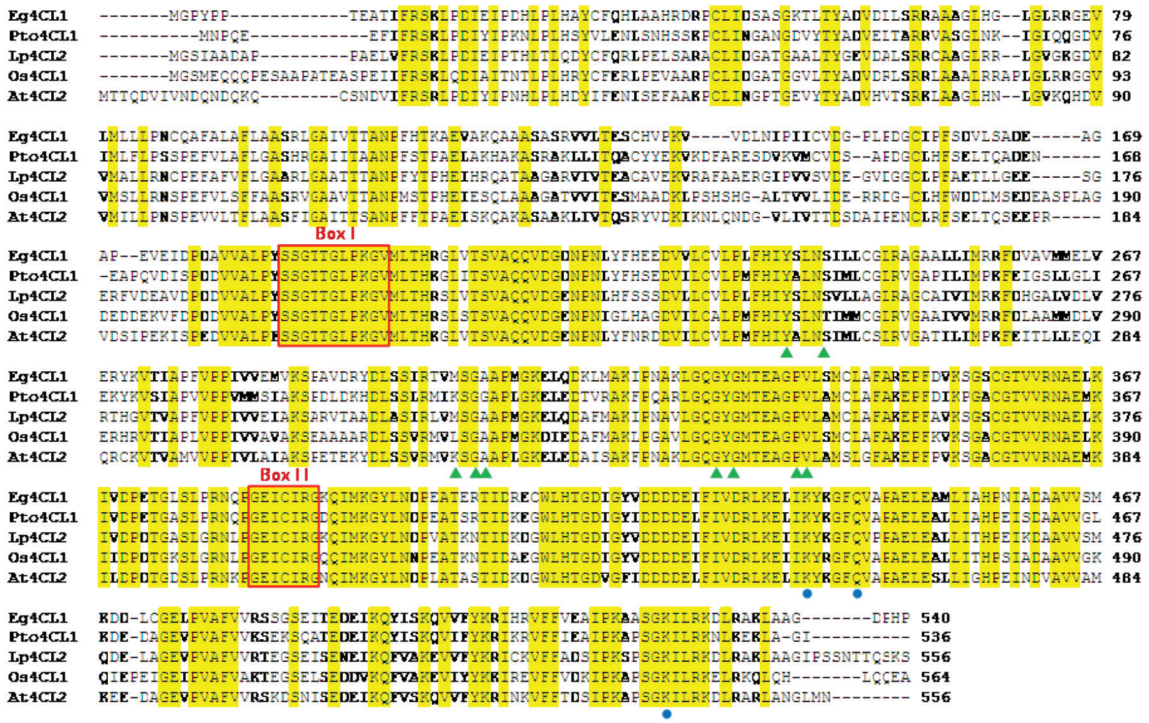


FIGURE 1. Sequence alignment of Eg4CL1 with other type I 4CL amino acid sequences. Sequences used in the alignment were from *Elaeis guineensis* (Eg4CL1), *Populus tomentosa* (Pto4CL1), *Lolium perenne* (Lp4CL2), *Oryza sativa* (Os4CL1) and *Arabidopsis thaliana* (At4CL2). Box I and Box II are two highly conserved motifs for an AMP-binding domain. Amino acid residues involved in substrate binding are marked with triangles (▲) at the bottom, while those required for catalytic activities are indicated with circle (●). The identical amino acid residues are highlighted in yellow

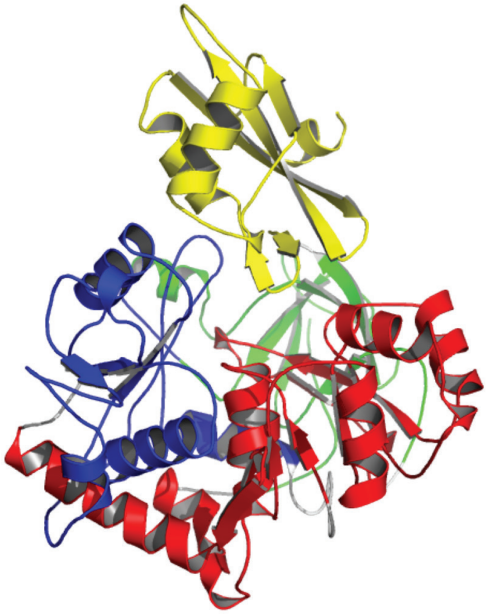


FIGURE 2. 3-Dimensional protein structure of Eg4CL1. The 4CL consists of C-domain (yellow) and a larger N-domain which can be divided into three subdomains namely: N1 (blue), N2 (red) and N3 (green)

binding pocket is further divided into three subdomains which are: N1 (blue), N2 (red) and N3 (green). All of the domains are positioned at the right spots according to the protein structure of Pto4CL1.

PHYLOGENETIC ANALYSIS

The phylogenetic tree divided the 4CLs into two major clades which separated members of the type I 4CL from members of the type II 4CL (Figure 3). Members of type I 4CL were further divided into two distinct subclades representing the dicots and monocots. Two subclades were also observed within members of type II 4CL; in which one clade consists of members from the dicot, while another consist of members from the monocot. Eg4CL1 (indicated with a closed circle in Figure 3) was grouped together with other members of the type I 4CL from the monocots. Members of this clade are suggested to be involved in lignin biosynthesis and some of these *4CL* genes have been characterized by functional studies. For instance, functional studies of the *Pv4CL1* and *Os4CL3* genes showed that they are involved in lignin biosynthesis in switch grass and rice, respectively. Perturbation of these *4CL* genes resulted in reduced lignin content accompanied by profound phenotypic changes in transgenic plants (Gui et al. 2011; Xu et al. 2011a). Hence, the phylogenetic tree provides a hint that Eg4CL1 is the key enzyme involved in the lignin biosynthesis pathway in oil palm.

GENE EXPRESSION ANALYSIS

Despite the phylogenetic analysis providing a clue regarding the function of *Eg4CL1*, gene expression analysis was also performed to further characterize the *Eg4CL1* gene. Since the expression behaviors of a gene reflect its physiological roles, the expression behaviors of *Eg4CL1* were investigated in several major organs of the oil palm including the coleoptile and root of the germinating seed; the young leaf and young root of one-year-old palm; and the immature fruitlet and young fruit. The RT-PCR showed *Eg4CL1* was abundantly expressed in all of the oil palm organs studied at similar expression levels regardless of the developmental stages of the oil palm (Figure 4). The expression pattern of *Eg4CL1* indicated that this gene might play an important role in lignification of the plant throughout the course of the oil palm development. By looking at the information from the phylogenetic analysis and the gene expression analysis, we postulate that *Eg4CL1* is responsible for lignin biosynthesis in oil palm.

PROMOTER SEQUENCE OF *Eg4CL1*

Since the expression behavior of a gene is largely regulated by its promoter, the isolation the *Eg4CL1* promoter could show the identity of the regulating elements which may be involved in coordinating the expression of *Eg4CL1*. A fragment of 534 bp corresponding to the promoter sequence of *Eg4CL1* was amplified in the first attempt. The second attempt produced another fragment (1262 bp) of the *Eg4CL1* promoter sequence, combining these two fragments yielded the 1.521 kb promoter sequence of *Eg4CL1*. The transcription start site (TSS) of *Eg4CL1* was identified based on the result of the 5'-RACE of *Eg4CL1* and defined as '+ 1'. It is an adenine nucleotide located 55 nucleotides upstream of the start codon. There are several motifs analogous to the TATA box found in the 5'- flanking sequence. However, the most probable TATA

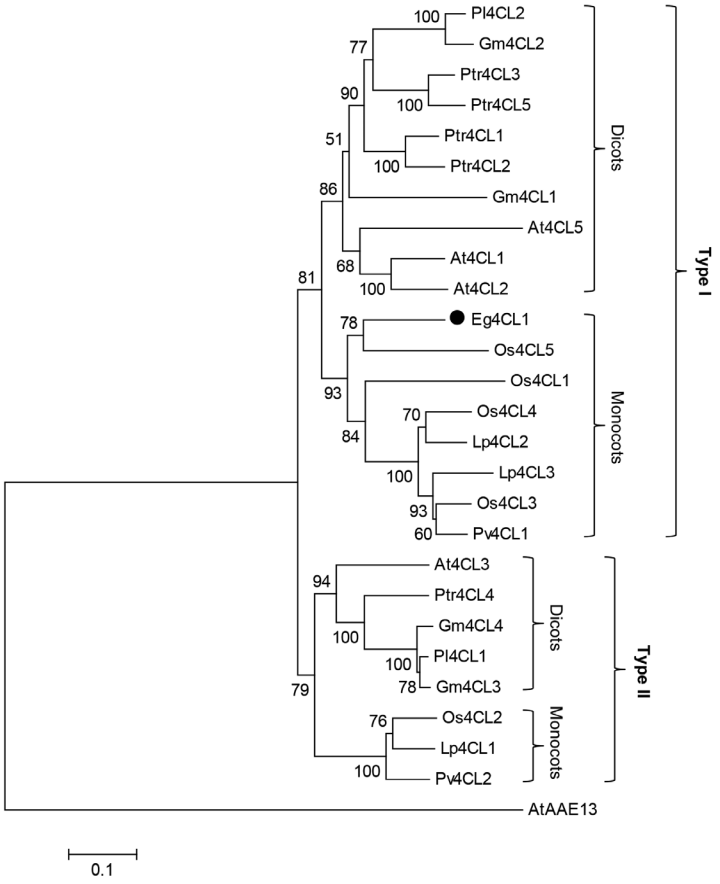


FIGURE 3. Phylogenetic analysis of 4CL proteins from selected angiosperm species. The Eg4CL1 is indicated with a closed circle (●). The phylogenetic tree was constructed using Neighbor-Joining method with 1000 bootstrap replicates in MEGA5 software. The values on each branch represented the bootstrap percentages. The AtAAE (*Arabidopsis thaliana* acyl-activating enzyme 13/ malonate--CoA ligase) was used as the outgroup. The complete scientific name of the organisms and the NCBI accession numbers of the 4CL proteins used in this analysis are presented in Supplementary Table 1

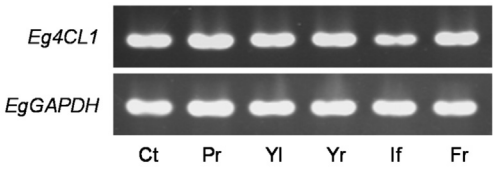


FIGURE 4. Expression profile of the *Eg4CL1* gene in different organs of oil palm. Total RNA from coleoptile (Ct) and primary root (Pr) of germinated seeds, young leaf (Yl) and young root (Yr) of one year old oil palm, immature fruitlet (If) and mesocarp tissues of young fruit (Fr) were converted to cDNA and subjected to RT-PCR

box has the sequence 'TATATTA' located at the position of –31 upstream of the TSS. Several important *cis*-acting elements including phytohormones-responsive elements, light-responsive elements, tissue-specific activation motifs and stress-responsive elements were detected in the promoter of *Eg4CL1* (Table 2). Among these *cis*-acting elements, the AC-II (ACCAACC) element was present twice at the –117 and –326 positions 5' upstream of the *Eg4CL1* gene (Supplemental Figure 1, online resource). The AC elements are the most prominent *cis*-acting element present in the promoter of the lignin biosynthetic genes including the *PAL*, *4CL* and *CAD* genes (Raes et al. 2003; Xu et al. 2014). It serves as the binding site for the MYB transcription factors involved in the regulation of the gene expression. Furthermore, AC elements are also necessary for the xylem specific expression of the lignin biosynthetic genes (Hatton et al. 1995).

DISCUSSION

In this study, a full-length cDNA of *Eg4CL1* which encodes 4-Coumarate:coenzyme A ligase was isolated from the oil palm genome. Phylogenetically, the *Eg4CL1* gene is classified as a type I *4CL* gene which is responsible for the biosynthesis of lignin. The *Eg4CL1* was expressed abundantly in all the oil palm organs studied, indicating it plays an important role in the production of monolignols for lignin biosynthesis in oil palm tissues. The presence of the lignification-regulating *cis*-acting elements in the promoter sequence of *Eg4CL1* further implies the involvement of this gene in lignin biosynthesis. Together

TABLE 2. *Cis*-acting elements present in the promoter of *Eg4CL1*

No.	Motifs	Sequence	Function	Reference
1	ABRE	TACGTG	*Cis*-acting element involved in the abscisic acid responsiveness	plantcare
2	Box I	TTTCAAA	Light responsive element	plantcare
3	CAG-motif	GAAAGGCAGAT	Part of a light response element	plantcare
4	CCAAT-box	CAACGG	MYBHv1 binding site	plantcare
5	ERE	ATTTCAAA	Ethylene-responsive element	plantcare
6	G-Box	CACGTT	*Cis*-acting regulatory element involved in light responsiveness	plantcare
7	GAG-motif	AGAGAGT	Part of a light responsive element	plantcare
8	GARE-motif	AAACAGA	Gibberellin-responsive element	plantcare
9	MNF1	GTGCCC(A/T)(A/T)	Light responsive element	plantcare
10	GT-1 box	GAAAAA	Plays a role in pathogen- and salt-induced SCaM-4 gene expression	(Park et al. 2004)
11	ACGTATERD1	ACGT	Required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis	(Simpson et al. 2003)
12	ARF	TGTCTC	Auxin response factor	(Goda et al. 2004)
13	DRE2	ACCGAC	Drought-responsive element in an RT ABA-dependent pathway	(Kizis & Pagès 2002)
14	I box	GATAAG	Conserved sequence upstream of light-regulated genes	(Rose et al. 1999)
15	LTRE	ACCGACA	Low temperature responsive element	(Nordin et al. 1993)
16	NtBBF1	ACTTTA	Required for tissue-specific expression and auxin induction	(Baumann et al. 1999)
17	POLLEN1LELAT52	AGAAA	Responsible for pollen specific activation	(Filichkin et al. 2004)
18	Pyrimidine box	CCTTTT	Gibberellin-respons cis-element of GARE and pyrimidine box are partially involved in sugar repression	(Mena et al. 2002)
19	SURE	AATAGAAAA	Sucrose Responsive Element	(Grierson et al. 1994)
20	WRKY71OS	TGAC	A core of TGAC-containing W-box	(Zhang et al. 2004)
21	AC-II	ACCAACC	Xylem-specific expression	(Hatton et al. 1995)
22	GATABOX	GATA	Required for high level, light regulated, and tissue specific expression	(Rubio-Somoza et al. 2006)
23	CURECORECR	GTAC	Copper-response element	(Kropat et al. 2005)

the information from the coding region, promoter, phylogeny and expression of the gene suggested that *Eg4CL1* is involved in lignin biosynthesis in the oil palm.

Lignin is a biopolymer which is deposited in the plant secondary cell wall (Neutelings 2011). It provides mechanical support to allow the plant to stand upright and confers defense against pathogen attacks (Xu et al. 2011b). However, the deposition of lignin in the plant cell does not favor industrial applications such as paper making and biofuel production. Removal of lignin from the pulp is costly and leads to the production of chemical wastes that are dangerous to the environment (Vanholme et al. 2010; Zhong & Ye 2009). For biofuel production, the presence of lignin in the lignocellulosic biomass impedes the saccharification process (Gao et al. 2014). Thus, reduces the amount of fermentable sugar produced and lowers the efficiency of biofuel production from the lignocellulosic biomass (Chapple et al. 2007; Chen & Dixon 2007). To overcome these problems, the lignin content of the plant biomass can be manipulated through genetic and molecular approaches (Shen et al. 2013; Van Acker et al. 2014).

The lignin biosynthetic genes such as *PAL*, *4CL*, *COMT* and *CAD* had been identified in many species and their roles were determined by functional studies (Chao et al. 2014; Gui et al. 2011; Huang et al. 2010; Trabucco et al. 2013). Manipulation of the lignin biosynthetic genes has been performed in a few economically important plant species to control the lignin content in the plant tissues (Jung et al. 2013; Sykes et al. 2016). Among the lignin biosynthetic genes in the phenylpropanoid pathway, *4CL* has become one of the targets to manipulate the lignin content of certain plants as it is located at the branching point of the phenylpropanoid pathway; channeling the CoA esters to form either flavonoids or lignin. In switch grass, suppression of the *Pv4CL1* gene resulted in reduced lignin

accumulation in the transgenic plants without affecting the biomass yield. Furthermore, the transgenic plants with reduced lignin content showed higher saccharification efficiency for biofuel production (Xu et al. 2011a). In *Populus tomentosa* Carr., perturbation of the *Ptc4CL1* gene led to changes in lignin content and composition in the transgenic plants. Up- and down-regulation of the *Ptc4CL1* gene shows that there is a positive correlation between the 4CL activity and lignin content in the plant (Tian et al. 2013a). In *Pinus radiata*, silencing of the *4CL* gene resulted in dwarfed plants with severe lignin reductions and changes in lignin composition and structure (Wagner et al. 2009).

The *4CL* genes are present in most of terrestrial plants, ranging from the lower plants such as liverworts and mosses to higher plants (Gao et al. 2015; Hamberger & Hahlbrock 2004; Silber et al. 2008). The gene usually exists in multiple copies which are similar in their sequences (De Azevedo Souza et al. 2008). This occurred as a result of gene-duplication events in the past (Hamberger & Hahlbrock 2004; Hamberger et al. 2007). In certain plants, multiple copies of the gene demonstrated a redundant role. Ehlting et al. (1999) suggested that the *At4CL1* and *At4CL2* genes of Arabidopsis play a redundant role in lignin biosynthesis. Furthermore, Li et al. (2015) showed that the *At4CL1* and *At4CL3* genes of Arabidopsis were both involved in the biosynthesis of sinapoylmalate. In *Populus tomentosa*, its five *Pto4CL* genes also displayed an overlapping function in lignin biosynthesis (Rao et al. 2015). The existence of multiple *4CL* genes could be viewed as a strategy to safe-guide the integrity of important metabolic pathways that serve for plant growth and development like lignification where a loss-of-function mutation in one copy of the gene could be rectified by another copy.

In general, the *4CL* genes of angiosperms can be classified into type I and type II, based on their sequence similarity (Hamberger et al. 2007). Both type I and type II *4CL* genes are sharing the same protein domains. Li et al. (2014) showed that the peptide sequences of Pl4CL1 and Pl4CL2 which represent the type II and type I *4CL* genes of *Pueraria lobata*, respectively, possessed the same Box I 'SSGTTGLPKGV' and Box II 'GEICIRG' conserved motifs of 4CL. Nevertheless, several studies have reported that the peptide sequences of the type II genes displayed an extension of amino acid residues at the N-terminal region compared to the type I 4CL (Heath et al. 2002; Hu et al. 1998; Kumar & Ellis 2003). This is a unique feature displayed by the type II *4CL* genes.

Apart from the *Eg4CL1* gene reported in this study, another three *4CL* genes (designated as *Eg4CL2-4*) were identified from the oil palm genome. *Eg4CL2* is located on chromosome 2 together with *Eg4CL1*. Meanwhile, the *Eg4CL3* and *Eg4CL4* genes are located on chromosome 8 and 11, respectively (Supplementary Table 2). Our analysis also showed that *Eg4CL2* and *Eg4CL3* are clustered together with *Eg4CL1* as type I *4CL* gene, while *Eg4CL4* as type II *4CL* gene (Supplementary Figure 2). Since detailed studies on *Eg4CL2-4* have not been performed, these genes would not be discussed further here.

The different types of *4CL* genes served for different functions in plants. The type I *4CL* genes are responsible for lignin biosynthesis, while the type II genes are involved in the formation of flavonoids and other metabolites (Ehlting et al. 1999; Li et al. 2014). For instance, *At4CL1* (a type I *4CL*) was abundantly expressed in the heavily lignified inflorescence stem in arabidopsis (Ehlting et al. 1999; Lee et al. 1995). The *4cl1* mutant was smaller in size and contained less lignin compared to the wild-type (Li et al. 2015). In contrast, the *At4CL3* gene (a type II *4CL*) was highly expressed in the flowers and siliques but not in the xylem (Ehlting et al. 1999; Li et al. 2015). Mutation of the *At4CL3* gene did not affect the lignin content, but greatly reduced the anthocyanin content of the mutant (Li et al. 2015). In rice, the *Os4CL3* gene (a type I *4CL*) was found to be responsible for lignin biosynthesis, while its homolog, *Os4CL2* (a type II *4CL*) was involved in flavonoid production (Gui et al. 2011; Sun et al. 2013).

The phylogenetic tree reconstructed in this study showed that *Eg4CL1* was clustered together with other type I *4CL* genes from monocots such as *Pv4CL1*, *Lp4CL2/3* and *Os4CL1/3/4/5*, implying that *Eg4CL1* is also a type I *4CL* and may carry out the same function as the other members of this clade. Previously, functional studies have been performed on *Pv4CL1* and *Os4CL3* to dissect their functions. Perturbation of *Pv4CL1* and *Os4CL3* led to lower lignin deposition accompanied by other phenotypic alterations in the transgenic plants (Gui et al. 2011; Xu et al. 2011a). Hence, *Eg4CL1* is very likely involved in lignin biosynthesis in oil palm. Besides the phylogenetic analysis, expression behaviors of the gene also suggested that *Eg4CL1* plays a major role in lignin production in oil palm. The gene expression analysis shows that *Eg4CL1* is highly expressed in all the tissues studied, including vegetative and reproductive organs, regardless of the developmental stage. The expression behaviors of the gene implied that *Eg4CL1* is associated with the onset of the biosynthesis of monolignols in oil palm tissues. In arabidopsis, *At4CL1* was expressed in all of the organs including the leaf, root, inflorescence stem, flower and silique at the seedling and mature stages (Ehlting et al. 1999; Lee et al. 1995). The accumulation of the *At4CL1* transcripts in the cotyledons and roots of the 3-days-old seedlings was correlated with the initiation of lignin biosynthesis after germination (Lee et al. 1995). Gene mutation analysis showed reduced lignin content in the *4cl1* mutant, which indicated that *At4CL1* is responsible for lignin biosynthesis in arabidopsis (Li et al. 2015). The expression of *Eg4CL1* in various tissues would allow the biosynthesis of lignin in various tissues for the development of the normal plant structure as lignin is required for the development of normal organ structure and provide mechanical support to the plant (Hirano et al. 2013; Yan et al. 2013). Therefore, the expression pattern of *Eg4CL1* indicates it plays an important role in lignin biosynthesis in oil palm tissues. Apart from that,

the expression of *Eg4CL1* may also be correlated directly with lignin biosynthesis in oil palm. A direct correlation between the expression of the lignin biosynthetic gene and the lignin content had been observed in previous studies (Fu et al. 2011; Voelker et al. 2010).

The expression behavior of a gene and its activities are mainly regulated by its promoter, although sometimes it may involve the participation of other gene elements such as intron and terminator (Goebels et al. 2013; Nagaya et al. 2009). To show the regulatory elements of the *Eg4CL1* gene, the promoter region was isolated. In line with the gene expression behavior, the type of *cis*-acting elements present in the promoter of *Eg4CL1* also suggests its functional role in lignin biosynthesis. As anticipated, the AC II elements were detected at two locations in the *Eg4CL1* promoter. Previous sequence analysis showed that the AC elements are present in the regulatory region of most of the lignin biosynthetic genes including *PAL, 4CL, COMT* and *CAD* (Hamberger et al. 2007; Raes et al. 2003). The AC elements in the promoter region serve as the binding site for the MYB transcription factors to regulate the expression of the genes (Shen et al. 2012; Tian et al. 2013b; Wang et al. 2014). A study of the bean *PAL2* promoter in transgenic tobacco has shown that the AC element is required for the xylem-specific expression of the lignin biosynthetic genes. Mutation of the AC element led to a reduced or complete loss of xylem specific expression in plants (Hatton et al. 1995). The presence of the AC II elements further supports the involvement of *Eg4CL1* in lignin biosynthesis.

To further confirm the function of the *Eg4CL1*, functional analysis should be performed on the gene and the promoter through the transgenic approach. However, producing transgenic oil palm is technically difficult, inefficient and time consuming (Bahariah et al. 2013; Masani et al. 2014). Hence, this study provides some clues for the identification of the lignin-related *4CL* gene in oil palm. Identification of the lignin production regulatory switch in oil palm will permit the manipulation of lignin content in the palm. Successful down regulation of lignin in oil palm will greatly improve the saccharification process and subsequently enhance biofuel production from oil palm empty fruit bunches.

Conclusion

In this study, the *Eg4CL1* gene and its promoter region have been isolated from oil palm. According to the analysis performed, *Eg4CL1* is potentially involved in lignin biosynthesis in oil palm. Therefore, *Eg4CL1* can be served as a molecular switch to manipulate the lignin content in oil palm biomass. This would allow more efficient production of biofuels from oil palm empty fruit bunches.

Acknowledgements

This study was funded by the Malaysian Palm Oil Board (Grant no.: 6366800). The authors would like to thank Prof. Dr. Ho Chai Ling for providing the 5'-RACE cDNA template and Mdm. Yeoh Keat Ai for her assistance in the preparation of the 5'-RACE cDNA template. Special thanks go to Mr. Rosmidi b. Miswan for his involvement in collection of oil palm samples. Finally, we are indebted to Prof. Dr Tan Soon Guan for proofreading the entire manuscript promptly. MPA, NAAS and IAS conceived and designed the research. YCYL and MNMN conducted the experiments. YCYL and MPA analyzed data. YCYL wrote the manuscript. All authors read and approved the manuscript. The authors declare that they have no conflict of interest.

References

Bahariah, B., Parveez, G.K.A., Masani, M.Y.A., Masura, S.S., Khalid, N. & Othman, R.Y. 2013. Biolistic transformation of oil palm using the phosphomannose isomerase (pmi) gene as a positive selectable marker. *Biocatalysis and Agricultural Biotechnology* 2: 295-304.

Baumann, K., De Paolis, A., Costantino, P. & Gualberti, G. 1999. The DNA binding site of the dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolb oncogene in plants. *The Plant Cell* 11: 323-334.

Carroll, B.J., Klimyuk, V.I., Thomas, C.M., Bishop, G.J., Harrison, K., Scofield, S.R. & Jones, J.D. 1995. Germinal transpositions of the maize element dissociation from T-DNA loci in tomato. *Genetics* 139: 407-420.

Chao, N., Liu, S.X., Liu, B.M., Li, N., Jiang, X.N. & Gai, Y. 2014. Molecular cloning and functional analysis of nine cinnamyl alcohol dehydrogenase family members in *Populus tomentosa*. *Planta* 240: 1097-1112.

Chapple, C., Ladisch, M. & Meilan, R. 2007. Loosening lignin's grip on biofuel production. *Nature Biotechnology* 25: 746-748.

Chen, F. & Dixon, R.A. 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25: 759-761.

Ehlting, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. & Kombrink, E. 1999. Three 4-Coumarate: Coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant Journal* 19: 9-20.

Filichkin, S.A., Leonard, J.M., Monteros, A., Liu, P.P. & Nonogaki, H. 2004. A novel endo-beta-mannanase gene in tomato LeMAN5 is associated with anther and pollen development. *Plant Physiology* 134: 1080-1087.

Fu, C., Xiao, X., Xi, Y., Ge, Y., Chen, F., Bouton, J., Dixon, R.A. & Wang, Z.Y. 2011. Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. *Bioenergy Research* 4: 153-164.

Gao, D., Haarmeyer, C., Balan, V., Whitehead, T.A., Dale, B.E. & Chundawat, S.P. 2014. Lignin triggers irreversible cellulase loss during pretreated lignocellulosic biomass saccharification. *Biotechnology for Biofuels* 7: 175.

Gao, S., Yu, H.N., Xu, R.X., Cheng, A.X. & Lou, H.X. 2015. Cloning and functional characterization of a 4-coumarate COA ligase from liverwort *Plagiochasma appendiculatum*. *Phytochemistry* 111: 48-58.

Goda, H., Sawa, S., Asami, T., Fujioka, S., Shimada, Y. & Yoshida, S. 2004. Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiology* 134: 1555-1573.

Goebels, C., Thonn, A., Gonzalez-Hilarion, S., Rolland, O., Moyrand, F., Beilharz, T.H. & Janbon, G. 2013. Introns

regulate gene expression in *Cryptococcus neoformans* in a Pab2p dependent pathway. *PLoS Genetics* 9(8): e1003686.

Grierson, C., Du, J.S., Zabala, M., Beggs, K., Smith, C., Holdsworth, M. & Bevan, M. 1994. Separate *cis* sequences and *trans* factors direct metabolic and developmental regulation of a potato tuber storage protein gene. *Plant Journal* 5: 815-826.

Gui, J., Shen, J. & Li, L. 2011. Functional characterization of evolutionarily divergent 4-coumarate: Coenzyme A ligases in rice. *Plant Physiology* 157: 574-586.

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B. & Douglas, C.J. 2007. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa, Arabidopsis thaliana,* and *Oryza sativa*: The populus lignin toolbox and conservation and diversification of angiosperm gene families. *Canadian Journal of Botany* 85: 1182-1201.

Hamberger, B. & Hahlbrock, K. 2004. The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proceedings of the National Academy of Sciences of the United States of America* 101: 2209-2214.

Hatton, D., Sablowski, R., Yung, M.H., Smith, C., Schuch, W. & Bevan, M. 1995. Two classes of *cis* sequences contribute to tissue-specific expression of a *pal2* promoter in transgenic tobacco. *The Plant Journal* 7: 859-876.

Heath, R., McInnes, R., Lidgett, A., Huxley, H., Lynch, D., Jones, E., Mahoney, N. & Spangenberg, G. 2002. Isolation and characterisation of three 4-coumarate: Coa-ligase homologue cdnas from Perennial Ryegrass (*Lolium perenne*). *Journal of Plant Physiology* 159: 773-779.

Hirano, K., Kondo, M., Aya, K., Miyao, A., Sato, Y., Antonio, B.A., Namiki, N., Nagamura, Y. & Matsuoka, M. 2013. Identification of transcription factors involved in rice secondary cell wall formation. *Plant and Cell Physiology* 54: 1791-1802.

Hu, W.J., Kawaoka, A., Tsai, C.J., Lung, J., Osakabe, K., Ebinuma, H. & Chiang, V.L. 1998. Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in Aspen (*Populus tremuloides*). *Proceedings of the National Academy of Sciences of the United States of America* 95: 5407-5412.

Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X.N. & Wang, D.C. 2010. Crystal structures of a *populus tomentosa* 4-coumarate: CoA ligase shed light on its enzymatic mechanisms. *The Plant Cell* 22: 3093-3104.

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y.H., Yu, J.Q. & Chen, Z. 2010. Functional analysis of the *Arabidopsis PAL* gene family in plant growth, development, and response to environmental stress. *Plant Physiology* 153: 1526-1538.

Ibrahim, M.F., Abd-Aziz, S., Yusoff, M.E.M., Phang, L.Y. & Hassan, M.A. 2015. Simultaneous enzymatic saccharification and ABE fermentation using pretreated oil palm empty fruit bunch as substrate to produce butanol and hydrogen as biofuel. *Renewable Energy* 77: 447-455.

Jung, J.H., Vermerris, W., Gallo, M., Fedenko, J.R., Erickson, J.E. & Altpeter, F. 2013. RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnology Journal* 11: 709-716.

Kumar, A. & Ellis, B.E. 2003. 4-Coumarate: CoA ligase gene family in Rubus idaeus: cDNA structures, evolution, and expression. *Plant Molecular Biology* 51: 327-340.

Kizis, D. & Pagès, M. 2002. Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. *The Plant Journal* 30: 679-689.

Kropat, J., Tottey, S., Birkenbihl, R.P., Depege, N., Huijser, P. & Merchant, S. 2005. A regulator of nutritional copper signaling in chlamydomonas is an SBP domain protein that recognizes the GTAC core of copper response element. *Proceedings of the National Academy of Sciences of the United States of America* 102: 18730-18735.

Lee, D., Ellard, M., Wanner, L.A., Davis, K.R. & Douglas, C.J. 1995. The *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) gene: Stress and developmentally regulated expression and nucleotide sequence of its cDNA. *Plant Molecular Biology* 28: 871-884.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. & Rombauts, S. 2002. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research* 30: 325-327.

Li, Y., Im Kim, J., Pysh, L. & Chapple, C. 2015. Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiology* 169: 2409-2421.

Li, Z.B., Li, C.F., Li, J. & Zhang, Y.S. 2014. Molecular cloning and functional characterization of two divergent 4-coumarate: coenzyme A ligases from Kudzu (*Pueraria lobata*). *Biological & Pharmaceutical Bulletin* 37: 113-122.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R. & Gwadz, M. 2010. CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Research* 39: 225-229.

Masani, M.Y.A., Noll, G.A., Parveez, G.K.A., Sambanthamurthi, R. & Prüfer, D. 2014. Efficient transformation of oil palm protoplasts by peg-mediated transfection and DNA microinjection. *PloS One* doi. 10.1371/journal.pone.0096831.

Mena, M., Cejudo, F.J., Isabel-Lamoneda, I. & Carbonero, P. 2002. A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in Barley Aleurone. *Plant Physiology* 130: 111-119.

Nagaya, S., Kawamura, K., Shinmyo, A. & Kato, K. 2009. The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells. *Plant and Cell Physiology* 51: 328-332.

Neutelings, G. 2011. Lignin variability in plant cell walls: Contribution of new models. *Plant Science* 181: 379-386.

Nordin, K., Vahala, T. & Palva, E.T. 1993. Differential expression of two related, low-temperature-induced genes in *Arabidopsis thaliana* (L.) Heynh. *Plant Molecular Biology* 21: 641-653.

Ochman, H., Gerber, A.S. & Hartl, D.L. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* 120: 621-623.

Park, H.C., Kim, M.L., Kang, Y.H., Jeon, J.M., Yoo, J.H., Kim, M.C., Park, C.Y., Jeong, J.C., Moon, B.C., Lee, J.H. & Yoon, H.W. 2004. Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiology* 135: 2150-2161.

Piarpuzan, D., Quintero, J.A. & Cardona, C.A. 2011. Empty fruit bunches from oil palm as a potential raw material for fuel ethanol production. *Biomass and Bioenergy* 35: 1130-1137.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y. & Boerjan, W. 2003. Genome-wide characterization of the lignification toolbox in Arabidopsis. *Plant Physiology* 133: 1051-1071.

Rao, G., Pan, X., Xu, F., Zhang, Y., Cao, S., Jiang, X. & Lu, H. 2015. Divergent and overlapping function of five 4-Coumarate/Coenzyme A ligases from *Populus tomentosa*. *Plant Molecular Biology Reporter* 33: 841-854.

Rastogi, S., Kumar, R., Chanotiya, C.S., Shanker, K., Gupta, M.M., Nagegowda, D.A. & Shasany, A.K. 2013. 4-Coumarate: CoA ligase partitions metabolites for eugenol biosynthesis. *Plant and Cell Physiology* 54: 1238-1252.

Rose, A., Meier, I. & Wienand, U. 1999. The tomato i-box binding factor LeMYBI is a member of a novel class of myb-like proteins. *The Plant Journal* 20: 641-652.

Rubio-Somoza, I., Martinez, M., Abraham, Z., Diaz, I. & Carbonero, P. 2006. Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *Plant Journal* 47: 269-281.

Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., Rudis, M.R., Tang, Y., Xiao, X., Jackson, L. & Li, G. 2013. A genomics approach to deciphering lignin biosynthesis in switchgrass. *The Plant Cell* 25: 4342-4361.

Shen, H., He, X., Poovaiah, C.R., Wuddineh, W.A., Ma, J., Mann, D.G., Wang, H., Jackson, L., Tang, Y., Neal Stewart, C. & Chen, F. 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193: 121-136.

Silber, M.V., Meimberg, H. & Ebel, J. 2008. Identification of a 4-Coumarate: CoA ligase gene family in the moss, *Physcomitrella patens* Q. *Phytochemistry* 69: 2449-2456.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. 2003. Two different novel *cis*-acting elements of erd1, a clpA homologous arabidopsis gene function in induction by dehydration stress and dark-induced senescence. *The Plant Journal* 33: 259-270.

Soltani, B.M., Ehlting, J., Hamberger, B. & Douglas, C.J. 2006. Multiple *Cis*-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of *Arabidopsis thaliana* 4CL gene family members. *Planta* 224: 1226-1238.

Souza, A.C., Barbazuk, B., Ralph, S.G., Bohlmann, J., Hamberger, B. & Douglas, C.J. 2008. Genome-wide analysis of a land plant-specific acyl: CoenzymeA synthetase (ACS) gene family in arabidopsis, poplar, rice and physcomitrella. *New Phytologist* 179: 987-1003.

Sun, H., Li, Y., Feng, S., Zou, W., Guo, K., Fan, C., Si, S. & Peng, L. 2013. Analysis of five rice 4-coumarate: Coenzyme a ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochemical and Biophysical Research Communications* 430: 1151-1156.

Sykes, R.W., Gjersing, E.L., Foutz, K., Rottmann, W.H., Kuhn, S.A., Foster, C.E., Ziebell, A., Turner, G.B., Decker, S.R., Hinchee, M.A. & Davis, M.F. 2016. Down-regulation of p-coumaroyl quinate/shikimate 3′-hydroxylase (c3′h) and cinnamate 4-hydroxylase (c4h) genes in the lignin biosynthetic pathway of *Eucalyptus urophylla* × *Eucalyptus grandis* leads to improved sugar release. *Biotechnology for Biofuels* 9: 691-699.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.

Tian, X., Xie, J., Zhao, Y., Lu, H., Liu, S., Qu, L., Li, J., Gai, Y. & Jiang, X. 2013a. Sense-, antisense- and RNAi-4CL1 regulate soluble phenolic acids, cell wall components and growth in transgenic *Populus tomentosa* Carr. *Plant Physiology and Biochemistry* 65: 111-119.

Tian, Q., Wang, X., Li, C., Lu, W., Yang, L., Jiang, Y. & Luo, K. 2013b. Functional characterization of the poplar R2R3-MYB transcription factor PtoMYB216 involved in the regulation of lignin biosynthesis during wood formation. *PLoS ONE* doi. 10.1371/journal.pone.0076369.

Trabucco, G.M., Matos, D.A., Lee, S.J., Saathoff, A.J., Priest, H.D., Mockler, T.C., Sarath, G. & Hazen, S.P. 2013. Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid o-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnology* doi. 10.1186/1472-6750-13-61.

Van Acker, R., Leplé, J.C., Aerts, D., Storme, V., Goeminne, G., Ivens, B., Légée, F., Lapierre, C., Piens, K., Van Montagu, M.C. & Santoro, N. 2014. Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-coa reductase. *Proceedings of the National Academy of Sciences of the United States of America* 111: 845-850.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. & Boerjan, W. 2010. Lignin biosynthesis and structure. *Plant Physiology* 153: 895-905.

Voelker, S.L., Lachenbruch, B., Meinzer, F.C., Jourdes, M., Ki, C., Patten, A.M., Davin, L.B., Lewis, N.G., Tuskan, G.A., Gunter, L. & Decker, S.R. 2010. Antisense down-regulation of 4CL expression alters lignification, tree growth, and saccharification potential of field-grown poplar. *Plant Physiology* 154: 874-886.

Wagner, A., Donaldson, L., Kim, H., Phillips, L., Flint, H., Steward, D., Torr, K., Koch, G., Schmitt, U. & Ralph, J. 2009. Suppression of 4-Coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiology* 149: 370-383.

Wang, S., Li, E., Porth, I., Chen, J.G., Mansfield, S.D. & Douglas, C.J. 2014. Regulation of secondary cell wall biosynthesis by poplar R2R3 MYB transcription factor PtrMYB152 in *Arabidopsis*. *Scientific Reports* 4: 5054.

Wang, T., Zhang, N. & Du, L. 2005. Isolation of RNA of high quality and yield from *Ginkgo biloba* leaves. *Biotechnology Letters* 27: 629-633.

Xu, B., Escamilla-Treviño, L.L., Sathitsuksanoh, N., Shen, Z., Shen, H., Percival Zhang, Y.H., Dixon, R.A. & Zhao, B. 2011. Silencing of 4-coumarate: Coenzyme a ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 192: 611-625.

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L. & Zhang, X. 2011. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-seq-dependent transcriptional analysis and histochemistry. *Journal of Experimental Botany* 62: 5607-5621.

Xu, Q., Yin, X.R., Zeng, J.K., Ge, H., Song, M., Xu, C.J., Li, X., Ferguson, I.B. & Chen, K.S. 2014. Activator-and repressor-type MYB transcription factors are involved in chilling injury

induced flesh lignification in loquat via their interactions with the phenylpropanoid pathway. *Journal of Experimental Botany* 65: 4349-4359.
Yan, L., Xu, C., Kang, Y., Gu, T., Wang, D., Zhao, S. & Xia, G. 2013. The heterologous expression in *Arabidopsis thaliana* of sorghum transcription factor SbbHLH1 downregulates lignin synthesis. *Journal of Experimental Botany* 64: 3021-3032.
Zhang, Z.L., Xie, Z., Zou, X., Casaretto, J., Ho, T.H.D. & Shen, Q.J. 2004. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiology* 134: 1500-1513.
Zhong, R. & Ye, Z.H. 2009. Transcriptional regulation of lignin biosynthesis. *Plant Signaling & Behavior* 4: 1028-1034.

Yusuf Chong Yu Lok
Faculty of Plantation and Agrotechnology
Universiti Teknologi MARA, Kampus Jasin
77300 Merlimau, Melaka
Malaysia

Yusuf Chong Yu Lok
Agricultural Biotechnology Research Group
Faculty of Plantation and Agrotechnology
Universiti Teknologi MARA
40450 Shah Alam, Selangor Darul Ehsan
Malaysia

Idris Abu Seman
Malaysian Palm Oil Board (MPOB)
No 6, Persiaran Institusi, Bandar Baru Bangi
43000 Kajang, Selangor Darul Ehsan
Malaysia

Nor Aini Ab Shukor
Department of Forest Management
Faculty of Forestry
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia

Nor Aini Ab Shukor
Institute of Tropical Forestry and Forest Product
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia

Mohd Norfaizull Mohd Nor & Mohd Puad Abdullah*
Department of Cell and Molecular Biology
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: puad@upm.edu.my

Received: 9 September 2016
Accepted: 26 April 2018

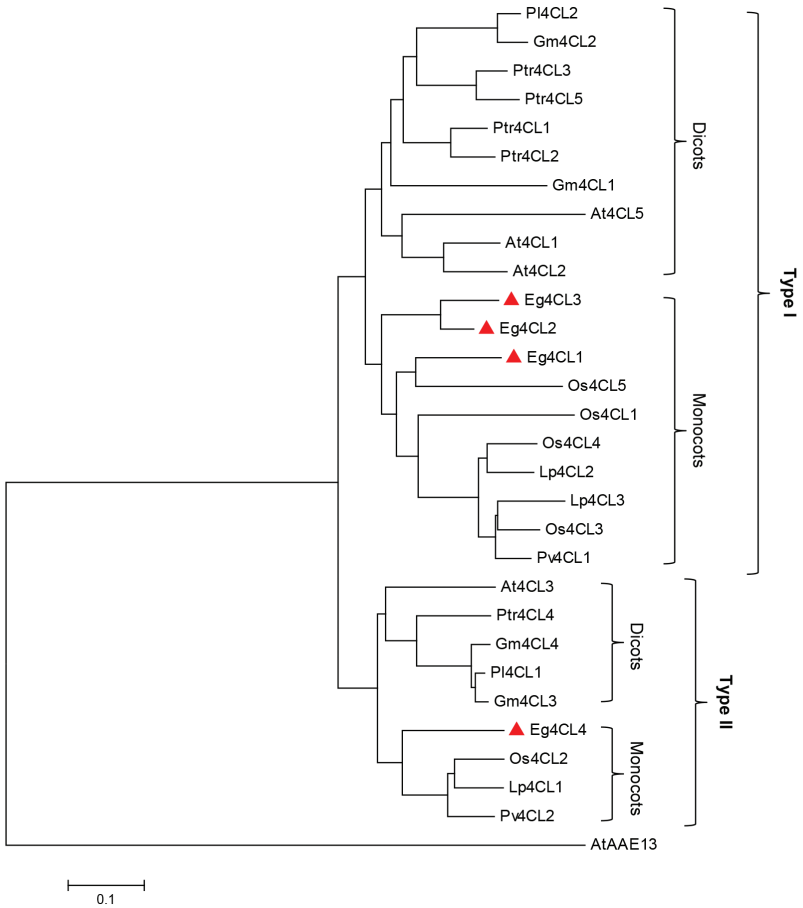
SUPPLEMENTARY TABLE 1. Organism names and the NCBI accession numbers of the 4CL proteins used in the phylogenetic tree (FIGURE 3)

Organism	Protein	Accession number
Elaeis guineensis	Eg4CL1	KM234973
Pueraria lobata	Pl4CL1	AGW16013
	Pl4CL2	AGW16014
Oryza sativa	Os4CL1	NP_001061353
	Os4CL2	NP_001047819
	Os4CL3	NP_001046069
	Os4CL4	NP_001058252
	Os4CL5	Q6ZAC1
Arabidopsis thaliana	At4CL1	NP_175579
	At4CL2	NP_188761
	At4CL3	NP_176686
	At4CL5	NP_188760
Populus trichocarpa	Ptr4CL1	XP_002329649
	Ptr4CL2	XP_00232447
	Ptr4CL3	XP_002297699
	Ptr4CL4	XP_002325815
	Ptr4CL5	XP_002304825
Lolium perenne	Lp4CL1	AAF37732
	Lp4CL2	AAF37733
	Lp4CL3	AAF37734
Panicum virgatum	Pv4CL1	ACD02135
	Pv4CL2	ADZ96250
Glycine max	Gm4CL1	AAL98709
	Gm4CL2	AAC97600
	Gm4CL3	AAC97599
	Gm4CL4	CAC36095
Arabidopsis thaliana	AtAAE13 (Acyl-activating enzyme 13 / malonate--CoA ligase)	NP_566537

SUPPLEMENTARY FIGURE 1. 5'- flanking sequence of *Eg4CL1*. The start codon of the *Eg4CL1* is in bold and underlined (**atg**). The transcription start site (TSS) is indicated with an arrow and marked with +1

SUPPLEMENTARY TABLE 2. Accession numbers and locations of *Eg4CL* genes in oil palm genome

Eg4CL	Accession number of gene sequence	Accession number of peptide sequence	Location (Chromosome)	Locus
Eg4CL1	KM234973 (XM_010915829)	AKC03652 (XP_010914131)	2	LOC105039619
Eg4CL2	XM_010915347	XP_010913649	2	LOC105039259
Eg4CL3	XM_010930350	XP_010928652	8	LOC105050370
Eg4CL4	XM_010935111	XP_010933413	11	LOC105053813



SUPPLEMENTARY FIGURE 2. Phylogenetic analysis of Eg4CL1-4 and other 4CL proteins. The Eg4CL1-4 are indicated with a triangle (▲). The phylogenetic tree was constructed using MEGA5 software in the same way as Figure 3