Asian J Agri & Biol. 2018;6(2):278-286.

AJAB

Original Research Article

Isolation and *in silico* characterization of cinnamate 4-hydroxylase (C4H) gene controlling the early stage of phenylpropanoid biosynthetic pathway in Kelampayan (*Neolamarckia cadamba*, Rubiaceae) developing xylem tissues

Boon Ling - Tchin, Wei Seng – Ho*, Shek Ling – Pang

Forest Genomics and Informatics Laboratory (fGiL), Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia

Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, 93250 Kuching, Sarawak, Malaysia

Received:
July 20, 2017

Accepted:
February 03, 2018

Published:
June 30, 2018

**Corresponding author email:*
wsho@unimas.my

Abstract

Cinnamate 4-hydroxylase (C4H) is one of the enzymes involved at the starting point of the phenylpropanoid and lignin biosynthesis pathway. It involves in the hydroxylation of cinnamate to 4-coumarate. In this paper, we isolated and *in silico* characterized the complete sequence of cinnamate 4-hydroxylase (C4H) gene from *Neolamarckia cadamba* in Malaysia. The C4H singletons obtained from the NcdbEST were used to predict the hypothetical full-length of NcC4H through the contig mapping approach. RT-PCR was used to amplify the full-length C4H cDNA clone and subsequently the PCR amplicons were sequenced and analysed. The NcC4H cDNA was 1,651 bp long with a 505 amino acid sequence, a 18 bp 5'-UTR and a 115 bp 3'-UTR. The predicted NcC4H protein contains P450-featured motifs. These include the heme-binding domain, a threonine-containing binding pocket motif and the proline-rich region. Peptide sequence comparison and phylogenetic analyses revealed that NcC4H was clustered with class I C4H instead of class II C4H, which is preferentially involved in phenylpropanoid and lignin biosynthesis pathway. This full-length NcC4H cDNA can be used for developing genetic marker to identify economic trait loci (ETL) for wood quality traits via genomics-assisted selection (GAS) or candidate gene mapping approach.

Keywords: *Neolamarckia cadamba*, RT-PCR, Lignin biosynthesis, Cinnamate 4-hydroxylase (C4H), Expressed sequence tags (ESTs), Genomics-assisted selection

Introduction

Lignin represents about 20-30% of all the plant stem biomass. It is the second most abundant organic compound found in wood, especially in supporting and conducting tissue of the plants such as fibers and tracheary elements. Lignin is produced by dehydrogenative polymerization of monolignols known as coniferyl alcohol, coumaryl alcohol and sinapyl alcohol. The polymerization of these monolignols will give rise to guaiacyl (G) units, ρ-coumaryl units (H) and sinapyl (S) units of lignin

(Brett and Waldron, 1990). Due to the mechanically rigid nature of the lignin and the deposition on cell wall, lignin provides mechanical and structural supports to the plants, and allows the transportation of water becomes smoother in the tracheids and vessels. According to Brett and Waldron (1990) and Higuchi (1997), lignin is very resistant to degradation in nature and therefore, it has provided a significant role in defending against pathogen or decaying fungi.

In pulp and paper industry, lignin should be separated from cellulose and hemicelluloses by an expensive and polluting process (Sederoff, 1999). In regard to this, study on lignin biosynthesis genes, such as cinnamate 4-hydroxylase (C4H) has gaining attention over the years and any up- or down-regulation of the gene may lead to the changes in lignin production (Baucher et al., 2003). The key role of C4H is to catalyze the hydroxylation of cinnamate to 4-coumarate during the first stage of lignin biosynthesis pathway (Lewis, 1999). To date, a considerable amount of C4H sequences has been isolated and characterized from various plant species. For example, a full-length C4H cDNA was isolated from the Korean black raspberry (*Rubus* sp.) and this gene is present as a single gene (Baek et al., 2008). Chen et al. (2007) also found two isoforms of C4H that are BnC4H-1 and BnC4H-2 from oilseed rape (*Brassica napus*). They had successfully cloned those genes into vectors and both of the genes contained two introns and a 1,518 bp open reading frame encoding a 505 amino acid polypeptide. Other studies such as isolation and characterization of C4H gene from *Parthenocissus henryana* (Liu et al., 2009) and tea (Singh et al., 2009).

Recently, C4H has been used as candidate gene for SNP discovery in *Acacia mangium* (Tchin et al., 2011). A significance genetic association was detected between C4H gene and lignin content in black cottonwood (Wegrzyn et al., 2010). A missense mutation study on C4H gene had also impact metabolism, growth and development in *Arabidopsis* (Schilmiller et al., 2009). A reduction in lignin content and wood density has been reported in *Populus* after the C4H gene being down-regulated (Bjurhager et al., 2010). These results demonstrate the importance of C4H gene towards the phenotype characteristics of plants and therefore, more studies are needed to better understand and identify marker-trait associations of this important gene on other species.

To date, a considerable amount of full-length C4H cDNA sequences has been published and made available in NCBI. Unfortunately, no information about the full-length C4H cDNA of *Neolamarckia cadamba* is available in the online database to date. *N. cadamba* or locally known as Kelampayan is one of the indigenous plantation tree species in Malaysia. Kelampayan has been selected for planted forest establishment in Sarawak. It has been proven as one of the best raw materials for the plywood industry (Lai et al., 2013; Ho et al., 2014; Tiong et al., 2014a,b,c&d; Phui et al., 2014; Sim et al., 2014; Pang et al., 2015). The leaves and barks have been extensively studied and reported to have high medicinal values (Joker, 2000; Patel and Kumar, 2007; Zaky et al., 2014a&b). Hence, the main objective of this study was to isolate and *in silico* characterize the full-length C4H cDNA from *N. cadamba* by using the contig mapping approach based on the Kelampayan expressed sequence tags (ESTs) obtained from the transcriptome database (NcdbEST) (Ho et al., 2014; Pang et al., 2015).

Material and Methods

EST data mining

A full-length C4H gene was predicted through contig mapping approach based on the ESTs obtained from the transcriptome database (NcdbEST) (Ho et al., 2014; Pang et al., 2015). The hypothetical full-length C4H gene (1,777 bp) was constructed by combining five EST singletons (i.e., Ncdx081E11; Ncdx040F06; Ncdx082A01; Ncdx082A01; Ncdx039B11 and Ncdx042B09) which have 100% sequence similarity at the overlapping regions. It contains open reading frame, start and stop codon, 5'-untranslated region (UTR) and 3'-untranslated region (UTR). The 3'-UTR for C4H gene has a long poly (A) tail attached to it at the end of sequence. The respective start and stop codons are located at the same position as other published gene sequences and the translated amino acid sequences also showed high sequence similarity to the C4H genes in NCBI database. A specific primer pair was designed using the Primer Premier 5 (Biosoft International, USA) based on the hypothetical full-length C4H gene. The oligonucleotide primers used for amplifying full-length C4H cDNA were FL-NcC4H2-F (5'-CATTTCCCGCCACCCATCA-3') and FL-NcC4H2-R (5'-CCTTGCGAATACAAAGATTATGG-3').

Amplification, cloning and sequencing of full-length C4H cDNA clone

Developing xylem tissues were collected from a 4-year old Kelampayan tree. Total RNA isolation, cloning and sequencing were based on the procedures as described in Tiong et al. (2014a&b). The PCR amplification was performed using a Veriti™ Thermal Cycler (Applied Biosystems, USA) using the PCR profile as described in Tchin et al. (2012)

***In Silico* sequence analysis of full-length C4H cDNA clone**

The vector sequences were trimmed of by using the Chromas version 2.33 (Technelysium, AU). The edited sequences were subjected to homology search using the BLASTn (Altschul et al., 1990) (http://blast.ncbi.nlm.nih.gov/). The C4H cDNA sequences were then translated into open reading frames using the ORF finder (http://us.expasy.org/tools/dna.html). The motifs of C4H were predicted from the multiple alignment analysis of C4H peptides with the Genbank deduced C4H amino acid sequences by using the ClustalW (Larkin et al., 2007). Phylogenetic trees were also constructed for the full-length C4H gene by using MEGA5 software (Tamura et al., 2011). Moreover, the tertiary structure of C4H was predicted by using Phyre2 software (Kelley and Sternberg, 2009). The graphical representation of tertiary protein structure was performed using the Jmol (http://www.jmol.org/) programme. The predicted structures were compared with the protein crystal structures available in the Protein Data Bank by using the Dali Server (Holm and Rosenstrom, 2010) to search for structure homology.

Results and Discussion

Full-length C4H cDNA clone sequence

The isolated full-length C4H cDNA was 1,651 bp long with a 1,518 bp open reading frame, a 18 bp 5'-UTR and a 115 bp 3'-UTR. The open reading frame of C4H cDNA encoded a 58.28 kDa protein with 505 amino acids and an isoelectric point of 9.42. From the blasting result, the cDNA sequence of C4H was 81% identical to C4H from *Cathanratus roseus*, 81% identical to C4H-2 from *Lithospermum erythrorhizon*, 80% identical to C4H-2 from *Populus trichocarpa* and others. This indicates that the C4H gene was successfully isolated and designated as NcC4H (Genbank NCBI accession number: JQ946327).

***In silico* analysis of NcC4H cDNA sequence**

The deduced amino acid sequence of NcC4H has P450-featured motifs, such as the proline-rich region (PPGPIPVP), the heme-binding domain (FGVGRRSCPG) and a threonine-containing binding pocket motif (AAIETT) (Chapple, 1998) (Figure 1). The proline-rich region is required for optimal orientation of the enzyme and for proper heme incorporation (Yamazaki et al., 1993). Meanwhile, the function of threonine-containing binding pocket is to bind the oxygen molecule required in catalysis (Chapple, 1998). The conserved cysteine amino acid in the heme-binding region serves as the fifth ligand to the heme iron which is essential for the catalysis reactions (Wachenfeldt and Johnson, 1995, cited in Chapple, 1998).

Peptide sequence comparison analysis revealed that the NcC4H gene discovered was categorized into class I C4H rather than class II C4H (Figure 2). It has been reported that the class I C4H genes are widely identified from various plant species than class II C4H genes (Lu et al., 2006). According to Harakava (2005), class I C4H is preferentially involved in phenylpropanoid and lignin biosynthesis pathway. Meanwhile class II C4H is particularly involved in stress responses but with minor role in lignin biosynthesis. This indicates that the NcC4H gene in *N. cadamba* may carry a key role in the lignin biosynthesis.

Phylogenetic analysis for NcC4H

Phylogenetic analysis for NcC4H was conducted by using MEGA5 software. The partial or full-length sequences of C4H gene from different plant species were retrieved from NCBI database to include in the analysis. From the phylogenetic tree constructed, two clusters were observed and NcC4H was grouped together with most of the plant species in one cluster. However, it showed a higher similarity with C4H from *Coffee arabica* (Figure 3), documenting the close evolutionary relationship within the Rubiaceae family. According to Lu et al. (2006), the PtriC4H1 and PtriC4H2 from *Populus trichocarpa* are involved in the lignin biosynthesis pathway, whereas the PtriC4H3 has preferred role in stress responses. This further indicating that the NcC4H gene has a major responsibility in lignin biosynthesis as it is grouped in the same cluster with PtriC4H1 and PtriC4H2.

Tertiary structure of NcC4H protein

The tertiary structure of NcC4H protein (Figure 4) was predicted by the using Phyre2 (Kelley and Sternberg, 2009). To date, the C4H protein crystal structure is still

unavailable in the online database, and therefore direct comparison of the C4H structures cannot be carried out. The structure comparison again PDB database by using the Dali server revealed that the modelled NcC4H protein structure share certain percentage of similarity (≤ 25%) with other members of P450 superfamily, with z-score value up to 64.1 (Table 1). This structure similarity was in consistent with the argument stated by Hasemann et al. (1995) where the members of plant P450 superfamily should possesses conserved tertiary structure.

Table 1. Comparison of NcC4H protein structure again structures in PDB by using Dali server

PDB	Description	Z-score	% Identity
3e4e	Human cytochrome P450 2E1	64.1	24
2fdv	Human Microsomal P450 2A6	54.1	25
2ve3	Cyanobacterial cytochrome P450 CYP120A1	34.8	17
1smi	*Bacillus megaterium* bifunctional P-450:NADPH-P450 reductase	34.1	20
3k9y	Rat mitochondrial P450 24A1	32.6	21
3awm	*Sphingomonas paucimobilis* fatty acid alpha-hydroxylase	31.5	15
3a50	*Pseudonocardia autotrophica* vitamin D hydroxylase	30.2	14

Figure 1. Multiple alignment of NcC4H protein sequence with C4H protein sequences from other species
Highlighted in grey colour regions indicated the conserved domains found within the sequences. (Domain 1: proline-rich region (PPGPIPVP); Domain 2: threonine-containing binding pocket motif (AAIETT); Domain 3: heme-binding domain (FGVGRRSCPG))

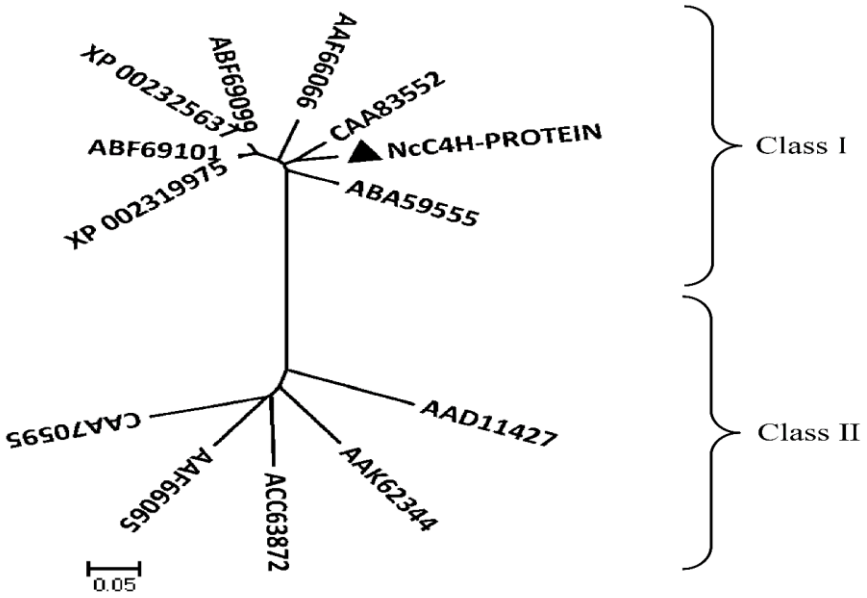

Figure 2. Classification of *C4H* genes from different plant species

(NcC4H-PROTEIN: *Neolamarckia cadamba*; XP_002319975: *Populus trichocarpa*; XP_002325637: *Populus trichocarpa*; ABF69099: *Populus tremuloides*; ABF69101: *Populus tremuloides*; AAF66066: *Citrus sinensis*; ABA59555: *Parthenocissus henryana*; CAA83552: *Catharanthus roseus*; AAF66065: *Citrus sinensis*; CAA70595: *Phaseolus vulgaris*; AAD11427: *Mesembryanthemum crystallinum*; AAK62344: *Nicotiana tabacum*; ACC63872: *Populus trichocarpa*)

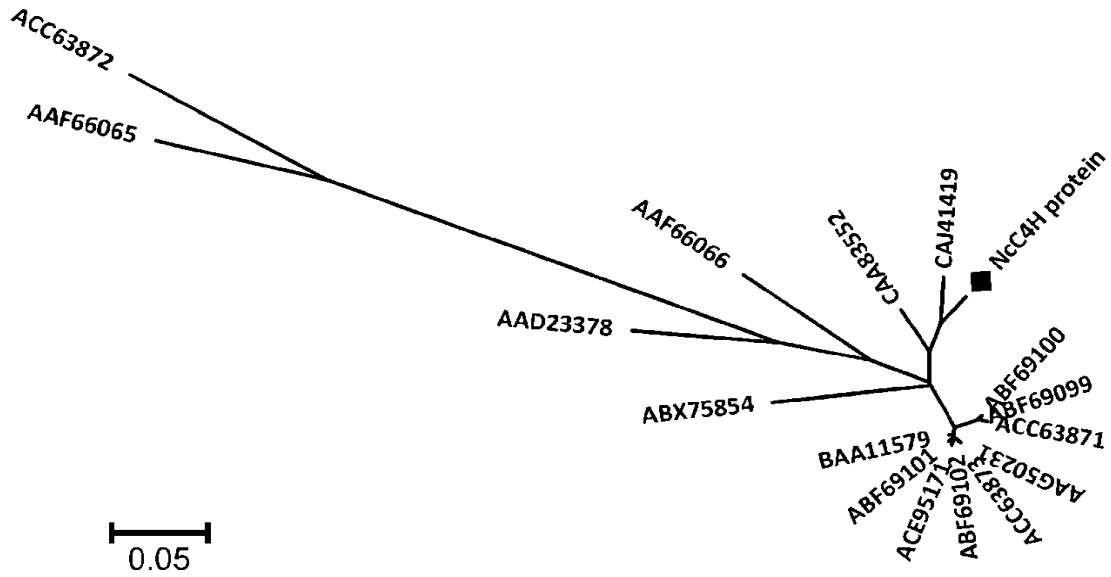

Figure 3. Phylogenetic tree constructed for *NcC4H* gene from Kelampayan by using MEGA5

(ACE95171: *P. tomentosa*; ABX75854: *A. auriculiformis* × *A. mangium*; AAG50231: *P. trichocarpa* × *P. deltoides*; ABF69101: *P. tremuloides* C4H2-1; CAJ41419: *Coffea arabica*; CAA83552: *Catharanthus roseus*; ABF69100: *P. tremuloides* C4H1-2; ABF69102: *P. tremuloides* C4H2-2; BAA11579: *P. kitakamiensis*; ABF69099: *P. tremuloides* C4H1-1; ACC63873: *P. trichocarpa* C4H1; ACC63871: *P. trichocarpa* C4H2; AAF66066: *Citrus sinensis* C4H2; AAF66065: *Citrus sinensis* C4H1; ACC63872: *P. trichocarpa* C4H3; AAD23378: *Pinus taeda*; NcC4H protein: *N. cadamba*)

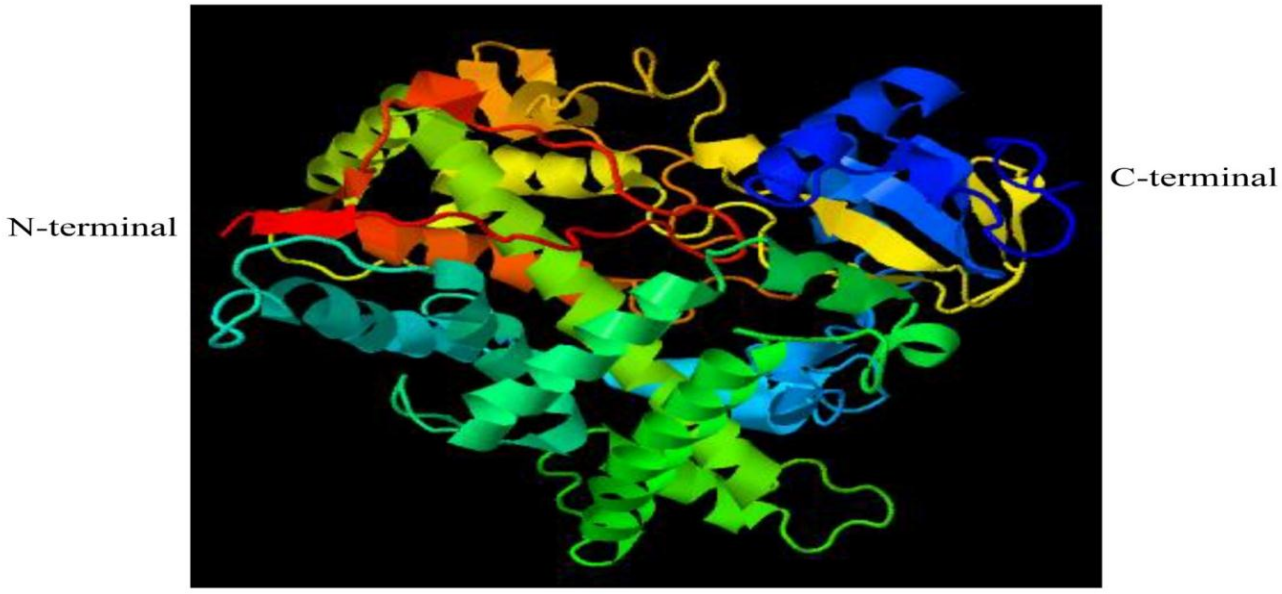

Figure 4. Tertiary structure of NcC4H protein predicted by using Phyre2

Conclusion

The present study clearly indicates that the NcC4H gene is preferentially involved in phenylpropanoid and lignin biosynthesis pathway rather than in stress responses. Further *in silico* analysis also indicates that it may carry a major responsibility in the lignin biosynthesis. In future, this full-length NcC4H cDNA can be used for developing genetic marker to identify economic trait loci (ETL) for wood quality traits via genomics-assisted selection (GAS) or candidate gene mapping approach.

Acknowledgment

This work was supported by the funding from the Sarawak Timber Association and Universiti Malaysia Sarawak (Grant No. GL(F07)/06/2013/STA-UNIMAS(06) and the Ministry of Higher Education, Malaysia (Grant No. RACE/a(2)/884/2012(02). Sarawak Forestry Corporation is also acknowledged for their field assistance in sample collection.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ, 1990. Basic local alignment search tool. J. Mol. Biol. 215(3): 403-410

Baek MH, Chung BY, Kim JH, Kim JS, Lee SS, An BC, Lee IJ and Kim TH, 2008. cDNA cloning and expression pattern of cinnamate 4-hydroxylase in the Korean black raspberry. BMB Report. 41(7): 529-536

Baucher M, Halpin C, Petit-Conil M and Boerjan W, 2003. Lignin: Genetic engineering and impact on pulping. Crit. Rev. Biochem. Mol. Biol. 38: 305-350

Bjurhager I, Olsson AM, Zhang B, Gerber L, Kumar M, Berglund LA, Burgert I, Sundberg B and Salmen L, 2010. Ultrastructure and mechanical properties of *Populus* wood with reduced lignin content caused by transgenic down-regulation of cinnamate 4-hydroxylase. Biomacromolecules. 11(9): 2359-2365

Brett C and Waldron K, 1990. Physiology and Biochemistry of Plant Cell Walls. London: Unwin Hyman

Chapple C, 1998. Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. Ann. Rev. Plant Phys. 49: 311-343

Chen AH, Chai YR, Li JN and Chen L, 2007. Molecular cloning of two genes encoding cinnamate 4-hydroxylase (C4H) from oilseed rape (*Brassica napus*). J. Biochem. Mol. Biol. 40(2): 247-260

Harakava R, 2005. Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*. Genet. Mol. Biol. 28(3): 601-607

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA and Deisenhofer J, 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure. 3: 41-62

Higuchi T, 1997. Biochemistry and molecular biology of wood. New York: Springer

Ho WS, Pang SL and Julaihi A, 2014. Identification and analysis of expressed sequence tags present in xylem tissues of Kelampayan (*Neolamarckia cadamba* (Roxb.) Bosser). Physiol. Mol. Biol. Plants. 20(3): 393-397.

Holm L and Rosenstrom P, 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38: W545-W549

Joker D, 2000. Seed Leaflet *Neolamarckia cadamba* (Roxb.) Bosser (*Anthocephalus chinensis* (Lam.) A. Rich. ex Walp.). Danida Forest Seed Centre, 17. Available from: http://www.dfsc.dk

Kelley LA and Sternberg MJE, 2009. Protein structure prediction on the web: a case study using the Phyre server. Nat. Protoc. 4(3): 363-371

Lai PS, Ho WS and Pang SL, 2013. Development, characterization and cross-species transferability of expressed sequence tag-simple sequence repeat (EST-SSR) markers derived from Kelampayan tree transcriptome. Biotechnology. 12(6): 225-235.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG, 2007. ClustalW and clustalX version 2.0. Bioinformatics. 23(21): 2947-2948

Lewis NG, 1999. A 20th century roller coaster ride: A short account of lignifications. Curr. Opin. Plant Biol. 2(2): 153-162

Liu S, Hu Y, Wang X, Han L, Song S, Cheng H and Lin Z, 2009. Isolation and characterization of a gene encoding cinnamate 4-hydroxylase from *Parthenocissus henryana*. Mol. Biol. Reports. 36(6): 1605-1610

Lu SF, Zhou YH, Li LG and Chiang VL, 2006. Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. Plant Cell Physiol. 47(7): 905-914

Pang SL, Ho WS, Mat-Isa MN and Julaihi A, 2015. Gene discovery in the developing xylem tissue of a tropical timber tree species: *Neolamarckia cadamba* (Roxb.) Bosser (Kelampayan). Tree Genet. Genomes. 11:47

Patel D and Kumar V, 2008. Pharmacognostical studies of Neolamarckia cadamba (roxb.) Bosser leaf. Int. J. Green Pharm. 2(1): 26-27

Phui SL, Ho WS, Pang SL and Julaihi A, 2014. Development and polymorphism of simple sequence repeats (SSRs) in Kelampayan (*Neolamarckia cadamba* – Rubiaceae) using ISSR suppression method. Arch. App. Sci. Res. 6(4): 209-218.

Schilmiller AL, Stout J, Weng JK, Humphreys J, Ruegger MO and Chapple C, 2009. Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. The Plant J. 60: 771-782

Sederoff R, 1999. Building better trees with antisense. Nat. Biotechnol. 17: 750-751

Sim, WYCM, Ho WS and Pang SL, 2014. Molecular cloning of hypervariable regions (HVRII) from cellulose synthase (CesA) gene in *Neolamarckia cadamba*. Int. J. Biosci. Biochem. Bioinformatics. 4(6): 475-482.

Singh K, Kumar S, Rani A, Gulati A and Ahuja PS, 2009. Phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. Funct. Integ. Genomics. 9: 125–134

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S, 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol. Biol. Evol. 28: 2731-9.

Tchin BL, Ho WS, Pang SL and Ismail J, 2011. Gene-associated single nucleotide polymorphism (SNP) in cinnamate 4-hydroxylase (C4H) and cinnamyl alcohol dehydrogenase (CAD) genes from *Acacia mangium* superbulk trees. Biotechnology. 10(4): 303-315

Tiong SY, Ho WS, Pang SL and Ismail J, 2014. Nucleotide diversity and association genetics of xyloglucan endotransglycosylase/hydrolase

(XTH) and cellulose synthase (CesA) genes in *Neolamarckia cadamba*. J. Biol. Sci. 14(4): 267-275.

Tiong SY, Chew SF, Ho WS and Pang SL, 2014. Genetic diversity of *Neolamarckia cadamba* using dominant DNA markers based on inter-simple sequence repeats (ISSRs) in Sarawak. Adv. App. Sci. Res. 5(3): 458-463

Tiong SY, Ho WS, Pang SL and Ismail J, 2014. *In silico* analysis of cellulose synthase gene (NcCesA1) in developing xylem tissues of *Neolamarckia cadamba*. Am. J. Bioinformatics. 3(2): 30-44.

Tiong SY, Ho WS, Pang SL and Ismail J, 2014. Bioinformatics analysis of xyloglucan endotransglycosylase/hydrolase (XTH) gene from developing xylem of a tropical timber tree *Neolamarckia cadamba*. Am. J. Bioinformatics. 3(1): 1-16.

Wachenfeldt CV and Johnson EJ, 1995. Structures of eukaryotic cytochrorne P450 enzymes in cytochrome P450 structure, mechanism and biochemistry (2nd edition). Edited by Ortiz de Montellano PR. Plenum Press, New York, USA.

Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai CJ and Neale DB, 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol. 188: 515-532

Yamazaki S, Sato K, Suhara K, Sakaguchi M, Mihara K and Omura T, 1993. Important of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. J. Biochem. 114(5): 652-657

Zaky ZM, Ho WS, Pang SL and Fasihuddin BA, 2014. EMS-induced mutagenesis and DNA polymorphism assessment through ISSR markers in *Neolamarckia cadamba* (Kelampayan) and *Leucaena leucocephala* (Petai Belalang). Eur. J. Exper. Biol. 4(4): 156-163.

Zaky ZM, Fasihuddin BA, Ho WS and Pang SL, 2014. GC-MS analysis of phytochemical constituents in leaf extracts of *Neolamarckia cadamba* (Rubiaceae) from Malaysia. Int. J. Pharm. Pharm. Sci. 6(9): 123-127.