CrossMark
click for updates

Isolation and in *silico* characterization of full-length *cinnamyl alcohol dehydrogenase* gene involved in lignin biosynthesis in *Neolamarckia cadamba*

Boon-Ling Tchin[1], Wei-Seng Ho[1*], Shek-Ling Pang[2]

[1]Forest Genomics and Informatics Laboratory, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia, [2]Applied Forest Science and Industry Development Unit, Sarawak Forestry Corporation, 93250 Kuching, Sarawak, Malaysia.

ARTICLE INFO

ABSTRACT

Cinnamyl alcohol dehydrogenase (CAD) catalyzes the reduction of cinnamaldehyde to ρ-coumaryl, coniferyl, and sinapyl alcohols during the last stage of lignin biosynthesis pathway. The *CAD* gene expression is believed to be important toward the phenotypic characteristics of plants. In the present study, a full-length *CAD* gene was successfully inferred from EST database (NcdbESTs) of *Neolamarckia cadamba* through a contig mapping approach. Reverse transcription polymerase chain reaction was conducted to validate the identity of the isolated *CAD* gene. The full-length *CAD* gene, designated as *NcCAD*, is 1,240 bp long with a 1,086 bp open reading frame encoding a protein of 361 amino acids, a 68 bp 5'-UTR, and a 86 bp 3'-UTR. Phylogenetic analysis showed that *NcCAD* was grouped in the cluster containing both *CAD* and *sinapyl alcohol dehydrogenase (SAD)* genes, in which both genes are involved in lignin biosynthesis. This result also demonstrated that the *NcCAD* gene may pose intermediate characteristics of both *CAD* and *SAD* genes. This *NcCAD* gene can serve as a good candidate gene for further insight into the wood properties of *N. cadamba* through association genetics study.

1. INTRODUCTION

Lignin 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. It represents approximately 20–30% of the plant biomass. Cinnamyl alcohol dehydrogenase (CAD) is one of the lignin biosynthesis genes with a major function in catalyzing the reduction of cinnamaldehyde to ρ-coumaryl, coniferyl, and sinapyl alcohols during the last stage of lignin biosynthesis [1]. Later, the dehydrogenative polymerization of these monolignols will give rise to the formation of lignin molecule in plant [2]. Lignin provides mechanical and structural supports to the plants. It allows transportation of water become smoother in tracheids and vessels. Moreover, lignin is very resistant to degradation in nature, and thus, it plays a significant protective function again pathogen or decaying fungi [3].

CAD is recognized as one of the regulating enzymes which control the formation of guaiacyl and syringyl lignin. According to a study carried out by Kutsuki *et al*. [4], angiosperm CADs reduce both coniferyl

and sinapyl aldehydes to their corresponding alcohols almost equally, but the gymnosperm CADs were extraordinarily specific for the reduction of coniferyl aldehyde. CAD displays distinct characteristics between gymnosperms and angiosperms [5]. CAD in gymnosperm is encoded by a single gene which is responsible for the biosynthesis of mainly guaiacyl lignin, and it has been characterized from various gymnosperms species [6,7]. In contrast, multiple CAD isoforms as well as the putative CAD sequences have been purified and isolated from many angiosperms [8-10].

Many studies had shown that any up- or down-regulation of *CAD* gene resulted in altered lignin production [11]. *CAD* gene has been widely used for the association genetic study, and it showed significant correlations with lignin composition, C6 sugar, and S: G ratio in black cottonwood [12]. Meanwhile, single nucleotide variation detected in *CAD* gene also showed significant associations with several wood properties traits such as wood density of loblolly pine [13], *Acacia mangium* [14], and *Neolamarckia cadamba* [15] as well as earlywood specific gravity and lignin composition in *Pinus taeda* [16]. Such significant genetic association reflects the importance of *CAD* gene toward phenotypic characteristics of plants.

N. cadamba or locally known as Kelampayan is one of the indigenous plantation tree species with high productivity and short rotation time [17-20]. It poses various purposes to the timber users including certain pharmacological values [21-24]. Here, we present the newly

Corresponding Author
Wei-Seng Ho
Forest Genomics and Informatics Laboratory (fGiL),
Faculty of Resource Science and Technology,
Universiti Malaysia Sarawak, 94300,
Kota Samarahan, Sarawak, Malaysian.
Email: wsho@unimas.my

isolated full-length cDNA sequence of *NcCAD* gene from *N. cadamba* with the aid of *N. cadamba* EST database [17,18]. This full-length *NcCAD* gene can serve as a good candidate gene for further insight into the wood properties of *N. cadamba* through association genetics study.

2. MATERIALS AND METHODS

2.1. *CAD* EST Data Analysis

A full-length *CAD* gene was predicted through contig mapping approach based on the ESTs obtained from the transcriptome database (NcdbEST) [17,18]. The database is generated by sequencing of 5' end of cDNA clones derived from developing xylem tissues of a 2-year-old *N. cadamba* tree. The hypothetical full-length *CAD* gene was constructed by combining four EST singletons (i.e., Ncdx040G11; Ncdx086F07; Ncdx036G07; and Ncdx049H04) which have 100% sequence similarity at the overlapping regions. It contains open reading frame, start and stop codon, 5'-untranslated region (UTR), 3'-UTR, and a poly (A) tail at the end of 3' sequence. A specific primer pair was designed using the Primer Premier 5 (Biosoft International, USA) based on the hypothetical full-length *CAD* gene. The oligonucleotide primers used for amplifying full-length *CAD* cDNA were FL-NcCAD-F (5'-TTTTCCCTCTGCTCCTTGC-3') and FL-NcCAD-R (5'-GCCACAGGCATACGAGACAC-3').

2.2. RNA Isolation, PCR, Cloning, and Sequencing of Full-length *CAD* cDNA

Total RNA isolation, cloning, and sequencing were based on the procedures as described in Tiong *et al*. [25,26]. Total RNA was isolated from the developing xylem tissues of a 4-year old *N. cadamba* tree. The PCR amplification was performed using a Veriti™ Thermal Cycler (Applied Biosystems, USA) using the PCR profile as described in Tchin *et al*. [15].

2.3. *In Silico* Sequence Analysis of Full-length *CAD* cDNA

The vector sequences were trimmed off using the Chromas version 2.33 (Technelysium, AU). The edited sequences were subjected to homology search using the BLASTn [27]. The *CAD* cDNA sequences were then translated into open reading frames using the ORF finder (http://us.expasy.org/tools/dna.html). The domain motifs of *CAD* were predicted through PROSITE [28] and conserved domain database (CDD) [29] search engines.

2.4. Phylogenetic Analysis and Three-dimensional (3D) Protein Structure Prediction of Full-Length *CAD* Genes

Phylogenetic trees were also constructed for the full-length *CAD* gene using MEGA 5 software [30]. The tertiary structures of *CAD* were predicted using Phyre2 software [31]. The Jmol (http://www.jmol.org/) program was used for the graphical representation of tertiary protein structure. The predicted tertiary structures were compared with the protein crystal structures available in the Protein Data Bank using Dali Server [32] for searching the structure homology.

3. RESULTS AND DISCUSSION

3.1. PCR Amplification and Cloning of *CAD* cDNA

A full-length *CAD* gene was successfully predicted from the *N. cadamba* EST database through a contig mapping approach. The hypothetical full-length *CAD* gene (1,478 bp) contains open reading frame, start

and stop codon, 5'-UTR, and 3'-UTR. According to the hypothetical full-length *CAD* sequence, full-length primer pair was synthesized and then used for PCR amplifications (Fig. 1). An expected bright band was observed after analyzed on a 1.5% agarose gel. The purified PCR product was then cloned and sequenced.

3.2. Full-length *NcCAD* cDNA Sequence

Full-length *CAD* cDNA is 1240 bp long with a 1086 bp open reading frame, a 68 bp 5'-UTR, and a 86 bp 3'-UTR. The NCBI BLASTn result indicated that the full-length *CAD* cDNA shared 68-72% of identities with other known *CAD/sinapyl alcohol dehydrogenase (SAD)* genes from *Populus trichocarpa*, *Populus tremuloides*, *Fragaria × ananassa,* and *Arabidopsis thaliana* (Table 1). This result indicated that the isolated gene was encoded for CAD. The annotated sequence was then designated as *NcCAD* (GenBank accession number: JQ946326). The *NcCAD* cDNA encodes a 38.563 kDa protein with 361 amino acids and an isoelectric point of 7.14.

3.3. Sequence Analysis of *NcCAD* Gene

The motif domains of *NcCAD* gene were detected using two independent programs, namely, PROSITE (http://prosite.expasy.org/) and CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). From the analysis, a zinc-containing alcohol dehydrogenase signature (GHEIVGEVTEVGSKV) was detected in the deduced *NcCAD* amino acid sequence from position 72 to 86 using PROSITE search engine. Meanwhile, an NADP-binding domain motif (GLGGLG) was identified at amino acid position 192 to 197 using CDD search engine. In addition, three catalytic zinc binding sites (His-73, Cys-51, and Cys-167) and four structural zinc binding sites (Cys-104, Cys-107, Cys-110, and Cys-118) were also detected in *NcCAD* amino acid sequence (Fig. 2).

3.4. 3D Structure Prediction of *NcCAD* Protein

The 3D secondary protein structures of NcCAD protein (Fig. 3) were predicted using the Phyre2 [31]. The result showed that NcCAD is a globular protein which contains alpha-domain and beta-domain. Furthermore, the structure comparison again PDB database using the

Figure 1: The hypothetical full-length *CAD* predicted through a contig mapping approach. The highlighted sequences indicate start and stop codon. The boxed sequences indicate the position where the full-length primers being designed.

Table 1: The BLASTn output for full-length *NcCAD* cDNA sequence discovered from *N. cadamba*.

Accession No.	Species	Query coverage (%)	E value	Maximum identity (%)
XM 002322786.1	*P. trichocarpa SAD*	86	1e-172	72
EU603305.1	*P. trichocarpa CAD2*	86	1e-172	72
AY850131.1	*P. tremula×P. tremuloides SAD*	86	5e-172	72
AF273256.1	*P. tremuloides SAD*	86	2e-170	72
XM 002299914.1	*P. trichocarpa CAD* like	87	4e-154	71
U63534.1	*Fragaria×ananassa CAD*	84	1e-135	70
AY050931.1	*A. thaliana CAD*	84	7e-113	68

N. cadamba: Neolamarckia cadamba, A. thaliana: Arabidopsis thaliana, P. trichocarpa: Populus trichocarpa, P. tremuloides: Populus tremuloides, P. tremula: Populus tremula.

Table 2: Comparison of NcCAD protein structure again structures in PDB using Dali server.

PDB	Description	Z-score	% Identity
1yqx	*P. tremuloides* sinapyl alcohol dehydrogenase	66.3	77
2cf5	*Arabidopsis* cinnamyl alcohol dehydrogenase	49.8	51
1piw	*S. cerevisiae* cinnamyl alcohol dehydrogenase	48.6	36

P. tremuloides: Populus tremuloides, S. cerevisiae: Saccharomyces cerevisiae.

```
>NcCAD Protein Sequence

                                                          ********
MAGKSPEEEHPVKAYGWAARDSSGVLSPFKFSRRATLEDDVRFKVLFCGICHTDIHFLKNEWGFSLYPLVPGHEIVGEV
*******
TEVGSKVTKVKVGDKVGVGCLVGSCRTCENCSANLENYCPKMVLTYAAPNFDGTITYGGYSNEMVCNEHFIIRFPENM
                          ######
PLAGGAPLLCAGVTVYSPMKYYGLAKPGSHIGINGLGGLGHVAVKFAKALGAKVTVISTSDRKKEEALKRLGADAFLISR

NADEMQAAAGTLDGILDCVSAKHALVPLLSLLKYHGKLITVGAPAEPLELPVAPLIIGRKLVGGSNIGGLAETQEMIDLAA

KHNITADIEVVSMEDVNTALERLAKGDVRYRFVIDVANTLKAP
```

**Fig. 2:** The motif domains detected within NcCAD amino acid sequence. (*: Zinc-containing alcohol dehydrogenases signature, #: NADP-binding domain motif, catalytic zinc binding sites, and C: Structural zinc binding sites).

Dali server revealed that the modelled NcCAD protein structure shares similarity to SAD of *P. tremuloides* (77%), CAD of *Arabidopsis* (51%), and CAD of *Saccharomyces cerevisiae* (36%), with z-score values in between 48.6 and 66.3 (Table 2).

3.5. Phylogenetic Analysis of *NcCAD* Gene

A phylogenetic analysis was performed for deduced NcCAD amino acid sequence to investigate the evolutionary relationships of the *NcCAD* gene with other plant species. The partial or full-length sequences of the *CAD* gene from different plant species were retrieved from NCBI database to include in the analysis. From the neighbor joining tree generated using MEGA 5 software [30], two clusters were observed. *NcCAD* was grouped in the cluster containing both *CAD* and *SAD* genes, but with more close distribution to *Populus SAD* (Fig. 4).

As indicated by Barakat *et al*. [10], both *CAD* and *SAD* genes were involved in lignin biosynthesis in the xylem of *P. trichocarpa* and *P. tremuloides*. *SAD* is essential for the biosynthesis of syringyl lignin in angiosperms [33]. Although *SAD* maintains the highest specificity for the substrate sinapaldehyde, it also catalyzes the reduction of coniferaldehyde [34]. Therefore, the *NcCAD* cDNA discovered in this study may pose intermediate characteristics of both *CAD* and *SAD*
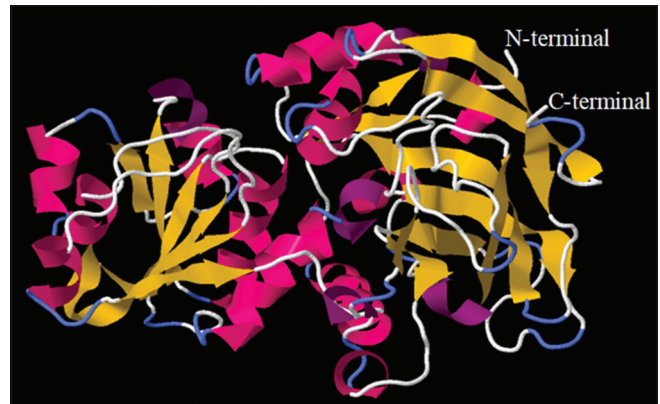
Figure 3: Predicted three-dimensional structure of NcCAD protein modeled using Phyre2 (color by secondary structure).
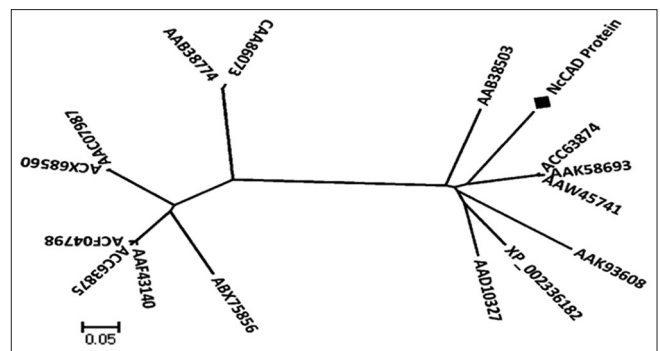
Figure 4: Phylogenetic tree constructed for *NcCAD* gene from *Neolamarckia cadamba* using MEGA 5 software [30]. *NcCAD* was grouped in cluster containing both *CAD* and *SAD* genes (NcC4H protein: *N. cadamba*; ACC63875: *Populus trichocarpa* CAD1; ACC63874: *P. trichocarpa* CAD2/ SAD; XP_002336182: *P. trichocarpa* CADL; AAK93608: *Arabidopsis thaliana*; AAB38503: *Mesembryanthemum crystallinum*; AAD10327: *Fragaria × ananassa*; AAK58693: *Populus tremuloides* SAD; ACF04798: *P. tomentosa*; AAW45741: *Populus tremula × P. tremuloides* SADL; AAF43140: *P. tremuloides*; AAC07987: *Eucalyptus globulus*; ACX68560: *E. camaldulensis*; ABX75856: *Acacia auriculiformis × Acacia mangium*; AAB38774: *Pinus radiata*; CAA86073: *Pinus taeda*).

genes. However, further structural and biochemical studies are needed to identify the specific function for *NcCAD* gene.

CAD proteins are encoded by a gene family in plants such as *A. thaliana* [35], *Oryza sativa* [36], and *Populus* [10]. According to Barakat *et al*. [10], the *CAD* gene family in woody plants could

be classified into three main classes based on the differences in gene structure and function. The Class I is *CAD* sequences from gymnosperms and angiosperms. Meanwhile, the Class II and III are dominated by sequences only from angiosperms. They further suggested that the Class I and II *CAD* genes are involved in wood development, and some other *CAD* genes from Class II and Class III may function in plant tissues under biotic stresses. In this classification, *NcCAD* was grouped into Class II *CAD* (Fig. 5). Based on the close distribution of *NcCAD* to *P. tremuloides SAD* and *P. trichocarpa CAD*10, it is further suggested that the *NcCAD* gene is involved in lignin biosynthesis.

We hope that this newly isolated and characterized *CAD* gene in *N. cadamba* could be used as one of the candidate genes for association mapping study aiming at the production of high-value planted forests in Malaysia [14,15,19,37,38]. For example, a significant association was detected in two lignin biosynthesis genes of *A. mangium* superbulk [14] and *N. cadamba* [15,38] with the basic wood density ($P < 0.05$). Furthermore, the detailed understanding on the regulation

of *CAD* gene could pave the way for a better understanding of lignin biosynthesis mechanism in this species. This will provide a greater impact on the design of advanced tree improvement programs of *N. cadamba*.
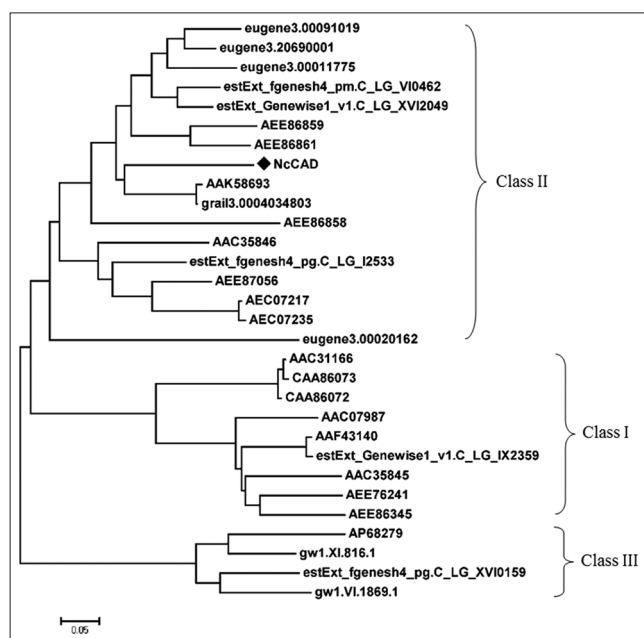
4. CONCLUSION

To the best of our knowledge, this is the first report on the assembly of a full-length *CAD* sequence (NCBI accession number: JQ946326) from N. *cadamba* using singletons of *CAD* from the kelampayan tree transcriptome database (NcdbEST) through a contig mapping approach. *In silico* analyses showed that *NcCAD* may pose intermediate characteristics of both *CAD* and *SAD* genes, in which both genes are involved in lignin biosynthesis. Further, phylogenetic analysis also predicted that *NcCAD* gene is involved in lignin biosynthesis.

5. ACKNOWLEDGMENTS

Figure 5: Phylogenetic tree constructed using MEGA 5 software [30] showing the classification of *CAD*s in *CAD* gene family (AP68279: *Arabidopsis thaliana* CAD1; AEC07217: *A. thaliana* CAD2; AEC07235: *A. thaliana* CAD3; AEE76241: *A. thaliana* CAD4; AEE86345: *A. thaliana* CAD5; AEE86858: *A. thaliana* CAD6; AEE86859: *A. thaliana* CAD7; AEE86861: *A. thaliana* CAD8; AEE87056: *A. thaliana* CAD9; AAC07987: *Eucalyptus globulus*; AAC35845: *M. sativa* CAD1; AAC35846: *M. sativa* CAD2; AAC31166: *Pinus radiata*; CAA86072: *P. taeda* CAD1; CAA86073: *Pinus taeda* CAD3; AAK58693: *Populus tremuloides* SAD/CAD1; AAF43140: *P. tremuloides* CAD2; estExt_fgenesh4_pg.C_LG_I2533: *Populus trichocarpa* CAD1; estExt_fgenesh4_pg.C_LG_ XVI0159: *P. trichocarpa* CAD2; estExt_ fgenesh4_pm.C_LG_VI0462: *P. trichocarpa* CAD3; estExt_Genewise1_ v1.C_LG_IX2359: *P. trichocarpa* CAD4; estExt_Genewise1_v1.C_LG_ XVI2049: *P. trichocarpa* CAD5; eugene3.00011775: *P. trichocarpa* CAD6; eugene3. 00020162: *P. trichocarpa* CAD7; eugene3.00091019: *P. trichocarpa* CAD8; Eugene 3.20690001: eugene3.20690001: *P. trichocarpa* CAD9; grail3.0004034803: *P. trichocarpa* CAD10; gw1.VI.1869.1: *P. trichocarpa* CAD11; gw1.XI.816.1: *P. trichocarpa* CAD12; NcCAD: *N. cadamba* CAD).

6. REFERENCES

1. Lewis NG. A 20th century roller coaster ride: A short account of lignification. Curr Opin Plant Biol 1999;2:153-62.
2. Brett C, Waldron K. Physiology and Biochemistry of Plant Cell Walls. London: Unwin Hyman; 1990.
3. Higuchi T. Biochemistry and Molecular Biology of Wood. New York: Springer; 1997.
4. Kutsuki H, Shimada M, Higuchi T. Regulatory role of cinnamyl alcohol dehydrogenase in the formation of guaiacyl and syringyl lignins. Phytochemistry 1982;21:19-23.
5. Ma QH. Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. J Exp Bot 2010;61:2735-44.
6. O'malley DM, Porter S, Sederoff RR. Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L.). Plant Physiol 1992;98:1364-71.
7. Galliano H, Cabané M, Eckerskorn C, Lottspeich F, Sandermann H Jr., Ernst D, *et al*. Molecular cloning, sequence analysis and elicitor-/ ozone-induced accumulation of cinnamyl alcohol dehydrogenase from norway spruce (*Picea abies* L.). Plant Mol Biol 1993;23:145-56.
8. Goffner D, Joffroy I, Grima-Pettenati J, Halpin C, Knight ME, Schuch W, *et al*. Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase from eucalyptus xylem. Planta 1992;188:48-53.
9. Kim SJ, Kim MR, Bedgar DL, Moinuddin SG, Cardenas CL, Davin LB, *et al*. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in arabidopsis. Proc Natl Acad Sci U S A 2004;101:1455-60.
10. Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto DS, Yellanki P, *et al*. The cinnamyl alcohol dehydrogenase gene family in populus: Phylogeny, organization, and expression. BMC Plant Biol 2009;9:26.
11. Baucher M, Halpin C, Petit-Conil M, Boerjan W. Lignin: Genetic engineering and impact on pulping. Crit Rev Biochem Mol Biol 2003;38:305-50.
12. Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, *et al*. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa, salicaceae*) secondary xylem. New Phytol 2010;188:515-32.
13. Yu Q, Li B, Nelson CD, McKeand SE, Batista VB, Mullin TJ.

Association of the cad-n1 allele with increased stem growth and wood density in full-sib families of loblolly pine. Tree Gen Genomes 2006;2:98-108.

14. Tchin BL, Ho WS, Pang SL, Ismail J. Gene-associated single nucleotide polymorphism (SNP) in cinnamate 4-hydroxylase (C4H) and cinnamyl alcohol dehydrogenase (CAD) genes from *Acacia mangium* superbulk trees. Biotechnology 2011;10:303-15.

15. Tchin BL, Ho WS, Pang SL, Ismail J. Association genetics of the cinnamyl alcohol dehydrogenase (CAD) and cinnamate 4-hydroxylase (C4H) genes with basic wood density in *Neolamarckia cadamba*. Biotechnology 2012;11:307-17.

16. González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. Association genetics in *Pinus taeda* L. I. Wood property traits. Genetics 2007;175:399-409.

17. Ho WS, Pang SL, Abdullah J. Identification and analysis of expressed sequence tags present in xylem tissues of kelampayan (*Neolamarckia cadamba (*Roxb.) bosser). Physiol Mol Biol Plants 2014;20:393-7.

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

19. Tiong SY, Chew SF, Ho WS, Pang SL. Genetic diversity of *Neolamarckia cadamba* using dominant DNA markers based on inter-simple sequence repeats (ISSRs) in Sarawak. Adv Appl Sci Res 2014;5:458-63.

20. Phui SL, Ho WS, Pang SL, Julaihi A. Development and polymorphism of simple sequence repeats (SSRs) in Kelampayan (*Neolamarckia cadamba-Rubiaceae*) using ISSR suppression method. Arch Appl Sci Res 2014;6:209-18.

21. Zaky ZM, Zaki MA, Fasihuddin BA, Ho WS, Pang SL. Comparison of mimosine content and nutritive values of *Neolamarckia cadamba* and *Leucaena leucocephala* with *Medicago sativa* as forage quality index. Int J Sci Technol Res 2014;3:146-50.

22. Jøker D. Neolamarckia Cadamba. Seed Leaflet No. 17. Danida Forest Seed Centre; 2000. Available from: http://www.curis.ku.dk/ws/files/20648324/neolamarckia_cadamba_int.pdf. [Last accessed on 2017 Mar 18].

23. Patel D, Kumar V. Pharmacognostical studies of *Neolamarckia cadamba* (roxb.) Bosser leaf. Int J Green Pharm 2008;2:26-7.

24. Zaky ZM, Fasihuddin BA, Ho WS, Pang SL. GC-MS analysis of phytochemical constituents in leaf extracts of *Neolamarckia cadamba* (*Rubiaceae*) from Malaysia. Int J Pharm Pharm Sci 2014;6:123-7.

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

26. Tiong SY, Ho WS, Pang SL, Ismail J. Bioinformatics analysis of xyloglucan endotransglycosylase/hydrolase (XTH) gene from

developing xylem of a tropical timber tree *Neolamarckia cadamba*. Am J Bioinform 2014;3:1-16.

27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.

28. Sigrist CJ, Cerutti L, Castro ED, Langendijk-Genevaux PS, Bulliard V, Bairoch A, *et al*. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 2010;38:D161-6.

29. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, *et al*. CDD: A Conserved domain database for the functional annotation of proteins. Nucleic Acids Res 2011;39:D225-9.

30. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, *et al*. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731-9.

31. Kelley LA, Sternberg MJ. Protein structure prediction on the web: A case study using the Phyre server. Nat Protoc 2009;4:363-71.

32. Holm L, Rosenström P. Dali server: Conservation mapping in 3D. Nucleic Acids Res 2010;38:W545-9.

33. Li L, Cheng XF, Leshkevich J, Umezawa T, Harding SA, Chiang VL. The last step of syringyl monolignol biosynthesis in angiosperm is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. Plant Cell 2001;13:1567-85.

34. Bomati EK, Noel JP. Structural and kinetic basis for substrate selectivity in populus tremuloides sinapyl alcohol dehydrogenase. Plant Cell 2005;17:1598-611.

35. Raes J, Rohde A, Christensen JH, Peer YV, Boerjan W. Genome-wide characterization of the lignifications toolbox in *Arabidopsis*. Plant Physiol 2003;133:1051-71.

36. Tobias CM, Chow EK. Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. Planta 2005;220:678-88.

37. Ho WS, Pang SL, Lau P, Ismail J. Sequence variation in the *cellulose synthase* (*SpCesA1*) gene from *Shorea parvifolia* ssp. *Parvifolia* mother trees. J Trop Agric Sci 2011;34:323-9.

38. Tiong SY, Ho WS, Pang SL, Ismail J. Nucleotide diversity and association genetics of xyloglucan endotransglycosylase/hydrolase (XTH) and cellulose synthase (CesA) genes in *Neolamarckia cadamba*. J Biol Sci 2014;14:267-75.