Molecular Cloning of Hypervariable Regions (HVRII) from Cellulose Synthase (*CesA*) Gene in *Neolamarckia Cadamba*

Wee-Yang C. M. Sim[1], Wei-Seng Ho[1*], Shek-Ling Pang[2]

[1] Forest Genomics and Informatics Laboratory (FGIL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak.
[2] Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, 93250 Kuching, Sarawak.

* Corresponding author. Tel.: 60 82-581000 ext. #2975; email: wsho@frst.unimas.my
Manuscript submitted June 9, 2014; accepted August 18, 2014.
doi: 10.7763/ijbbb.2014.v4.387

Abstract: *Neolamarckia cadamba* or locally known as Kelampayan, is one of the fast growing plantation tree species that holds great prospect as a renewable bioresources for plywood, pulp and paper, and biofuel industries. Sufficient information on *cellulose synthase* (*CesA*) gene, especially the hypervariable region II (HVRII) component involved in wood formation of Kelampayan is imperative for future applications. This region is thought to play a role in interaction with other unique cell-type-specific proteins involved in the biosynthesis of cellulose. The aim of this study was to identify and clone the HVRII regions of *cellulose synthase* gene from the developing xylem tissues of Kelampayan. The cDNA of cellulose synthase HVRII regions was amplified using reverse transcription-PCR (RT-PCR) approach using degenerate primers. Three clones, namely *NcCesA1HVRII* (520bp), *NcCesA2HVRII* (580bp) and *NcCesA3HVRII* (620bp) were successfully sequenced and characterized. *NcCesA1HVRII* and *NcCesA3HVRII* were clustered into two distinct clades implicated with secondary cell wall development whereas *NcCesA2HVRII* has renamed to *NcCslD1HVRII* due its high similarity with various plants' *CslD-HVRII*. This study provides an easier and faster access to *NcCesAHVRII* sequences to further understand the role of NcCesA/NcCslD protein for future applications such as selecting trees with optimal cellulose content required for specific industries.

Key words: *Neolamarckia cadamba*, reverse transcription-PCR (RT-PCR), cellulose synthase (*CesA*), cellulose-synthase like (*Csl*), hypervariable region II (*HVRII*)

1. Introduction

Cellulose synthase gene is the gene responsible in the biosynthesis of cellulose, one of the most important components in wood and paper. There are currently more than 20 full-length *CesA* sequences available in the various genebanks which is highly similar to each other across the entire length of the encoded protein with some exception from two small regions of variability [1]. With small introns present, the size of *CesA* genes that ranges from 3.5kb to 5.5kb encodes proteins with 985 to 1,088 amino acids in length. Cellulose synthase shares certain similar domain structures across species. These include a zinc finger, several transmembrane domains, conserved residues and 2 hypervariable regions (HVR), HVRI and HVRII [2], [3].

The HVRII region is one of the plant-specific regions with high sequence divergence, along with the conserved region (CR-P) which instead shows high sequence conservations [3]. These plant-specific domains are thought to be involved in functions unique to plant such as the binding of sucrose synthase, interaction with proteins associated with cytoskeleton or other accessory proteins. These domains are

thought to give specificity for plants cells, tissues and organs [4]. However, until present, researchers have yet to fully understand the regulation and function of *CesA* gene itself and each of its domain structures, including HVRII region. They speculated that HVRII region might be involved in the regulation of the quantity and quality of the cellulose synthesized in plants [5]. Amino acid sequences between the highly conserved motifs of ALYG and VISCG are considered as HVRII regions [6]. With the current information available, HVRII region in plant *CesA* gene has a sequence length ranging from about 500bp to 600bp long. HVRII region was proposed to be renamed as class-specific regions (CSR) instead due to its conservation among CesA orthologs from different plant species [7]. The high sequence similarity of HVRII region among *CesA* orthologs makes it useful in distinguishing individual family members of *CesA* gene and could prove useful for full-length cDNA isolation [8].

Thus, the objective of this research was to isolate and characterize the HVRII regions of *CesA* gene that present in xylem tissues of *Neolamarckia cadamba. N. cadamba* or locally known as kelampayan, has been identified as a promising fast growing species for planted forest development in Malaysia [9]. It is a large, deciduous and fast-growing tree species, thus with characteristics which guarantee early economic return within eight to ten years. Under normal conditions, it reaches a height of 17 m and a diameter of 25 cm at breast height (dbh) within nine years. It is one of the best sources of raw material for the plywood industry, besides pulp and paper production. *N. cadamba* can also be used as a shade tree for dipterocarp line planting, whilst its leaves and bark have medical application [10].

2. Materials and Methods

2.1. Total RNA Isolation and RT-PCR of *HVRII* Regions

Total RNA was extracted and purified from the developing xylem tissues using RNeasy® Midi Kit (Qiagen, Germany). First-strand cDNA was synthesized according to the Ready-To-Go You-Prime First-Strand Beads (GE Healthcare, USA) protocol. Forward degenerate primer, HVR2F, (5′-TGYTATGTYCAGTTYC CWC-3′) and the reverse degenerate primer, HVR2R, (5′-GANCCRTARATCCAYCC-3′) were used in the RT-PCR [6]. First-strand cDNA was synthesised using Ready-To-Go You-Prime First-Strand Beads (GE Healthcare, USA). The RT-PCR reaction was carried out in a Mastercycler Gradient Thermal Cycler (eppendorf, Germany) with the initial denaturation at 95 °C for 10 min, followed by two cycles of non-stringent amplification profile of 94°C for 1 min, 41°C for 1 min 30 s and 72°C for 2 min, followed by another more stringent amplification profile at 28 cycles of 94°C for 1 min, annealing temperatures of 55°C and 45°C for 1 min 30 s and 72°C for 2 min. The total 25 µl amplification reaction volume contained 1× PCR buffer (200 mM Tris-HCl, 500 mM KCl, pH8.4), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 10 pmol primers, 1 U *Taq* DNA polymerase (Promega, USA) and 1 µl template cDNA. The amplified cDNA was subsequently diluted with 50× ultrapure water and amplified again under the same PCR conditions. Gradient PCR was also done with temperatures ranging from 40.0°C – 60.6°C with regards to the more stringent annealing temperature. The RT-PCR amplicons were gel purified using QIAquick Gel Extraction Kit Protocol (QIAGEN, Germany) and cloned into pGEM®-T Easy Vector System (Promega, USA). The desired clone were selected for plasmid extraction using Wizard® *Plus* SV MiniPreps DNA Purification System (Promega, USA) and sequenced using BigDye Version 3.1 (Applied Biosystems, USA) using Applied Biosystems 3730 DNA (Agencourt, USA).

2.2. Sequencing Analysis

All the DNA sequences were checked and edited using Chromas Lite 2.01 prior to checking using BLASTn to search for sequence similarity through all known sequences available in the NCBI database. The edited nucleotide sequences were translated into amino acid sequences using Expert Protein Analysis System (ExPASy) (www.ca.expasy.org) translate tool. All amino acid sequences were multiple-aligned using Molecular Biological Advanced DNA Analysis Version 6.84 (http://www.molbiosoft.de/) and an un-rooted

Neighbor-Joining (NJ) phylogenetic tree with a bootstrap value of 1000 was constructed using MEGA Version 6 to cluster the *Nc*CesAHVRII protein sequences with other available plant HVRII protein sequences retrieved and compiled.

3. Results and Discussion

Total RNA was successfully isolated from developing xylems of *N. Cadamba* from samples obtained by scrapping off the developing xylem wall of felled Kelampayan trees. Fig. 1(a) shows a 0.8% agarose gel electrophoresis photo with two distinct 25S rRNA and 18S rRNA bands. The similarity of band intensity implies similar abundance of both RNA, indicating little or no RNA degradation during the extraction. RT-PCR was conducted by using the degenerate primers and three DNA bands were obtained from different annealing temperatures, 45˚C and 55˚C. After 50× dilution of each respective PCR product, PCR re-amplification was carried out by using the same primers at the same reaction conditions, respectively. Three bands of ~520bp, ~580bp and ~620bp were consistently amplified from *N. cadamba* (Fig. 1(b)). These PCR products were successfully purified, cloned and sequenced in the present study.

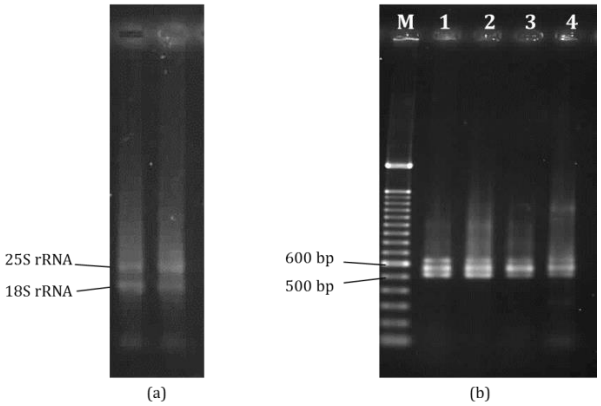

Fig. 1. (a) Agarose gel electrophoresis of total RNA extracted from developing xylem and (b) RT-PCR products on a 1.5% agarose gel. Lanes 1 and 3: PCR product at 45˚C and 55˚C T_a, respectively using cDNA as template. Lanes 2 and 4: PCR product at 45˚C and 55˚C T_a, respectively using respective PCR products diluted 50× as template. Lane M1: 100 bp DNA marker.

The edited nucleotide sequences were analysed by using BLAST tools (www.ncbi.nlm.nih.gov) to search for homologous sequences available in the NCBI nucleotide databases to validate the identity of each of the sequences obtained. Sequences obtained were named as *NcCesA1HVRII* (520 bp), *NcCesA2HVRII* (580 bp) and *NcCesA3HVRII* (620 bp). BLASTn analysis of *NcCesA1HVRII* showed high identities with cellulose synthase (*CesA1*) mRNA of *Shorea parvifolia* ssp. *parvifolia* (79%) and *CesA* of *Eucalyptus globulus* ssp. *globulus* (77%). *NcCesA2HVRII* showed high degree of identities with cellulose-synthase like (*PtrCslD*) of *Populus trichocarpa* (78%) and cellulose-synthase like D2 (CslD2) of *Arabidopsis thaliana* (70%) meanwhile *NcCesA3HVRII* showed high similarities with cellulose synthase (CesA2) of *Eucalyptus grandis* (77%) and cellulose synthase (CesA4) of *Populus tomentosa* (79%). A few of the BLASTn analysis results for all three sequences obtained are shown in Table 1.

The validated nucleotide sequences were then translated into amino acid sequence by using the Expert Protein Analysis System (ExPASy) (www.ca.expasy.org) translate tool. The correct translated frame of amino acid sequence was selected by the similarity of short amino acid sequences 5' CYVQFPQ....GWIYGS 3' flanking the whole HVRII amino acid sequence. This is because the degenerate primer designed was based on conserved regions flanking the HVRII gene. The translated nucleotide sequences were named

NcCesA1HVRII, NcCesA2HVRII and NcCesA3HVRII, respectively and checked again using BLASTp (www.ncbi.nlm.nih.gov). The result of BLASTp also showed that NcCesA1HVRII was highly identical to cellulose synthase (CesA3) of *E. grandis* (86%) and cellulose synthase (CesA) of *S. parvifolia* ssp. *parvifolia* (85%). NcCesA2HVRII showed high identities with cellulose-synthase (CslD) protein of *P. trichocarpa* (86%) and cellulose-synthase like D3 (CslD3) protein of *Populus tremuloides* (51%). NcCesA3HVRII showed high similarities with cellulose synthase (CesA4) of *P. tomentosa* (79%) and cellulose synthase (CesA2) of *Gossypioides kirkii* (88%). The low E-values increases the significance of the matches as it is nearer to "0" and that it also describes the random background noise that exist for matches between sequences. The BLASTp analysis for the three translated sequences is shown in Table 2.

Table 1. BLASTn Output of *NcCesAHVRII* Sequences

Gene Name	Organisms	Accession No.	Identity (%)	E-value
NcCesA1HVRII	*S. parvifolia* ssp. *parvifolia*	GQ338420.1	79	8e-67
	E. globulus ssp. *globulus*	AB527047.1	77	2e-62
NcCesA2HVRII	*P. trichocarpa*	XM_002325781.1	78	9e-87
	A. thaliana	NM_121697.3	70	1e-47
NcCesA3HVRII	*E. grandis*	EU165712.1	77	2e-82
	P. tomentosa	FJ534554.1	79	1e-91

Table 2. BLASTp Output of NcCesAHVRII Amino Acid Sequences

Protein Name	Organisms	Accession No.	Identity (%)	E-value
NcCesA1HVRII	*E. grandis*	AAY60845.1	86	1e-45
	S. parvifolia ssp. *parvifolia*	ACT67415.1	85	3e-43
NcCesA2HVRII	*P. trichorpa*	XP_002325817.1	86	5e-54
	P. tremuloides	AAO03578.1	51	2e-20
NcCesA3HVRII	*P. tomentosa*	ACT78709.1	79	2e-26
	G. kirkii	AAN28294.1	88	7e-24

From this study, only two HVRII regions of *CesA* and one HVRII region of cellulose-synthase like (*Csl*) sequence were obtained successfully from the developing xylem tissues of *N. cadamba.* Phylogenetic tree derived from 62 HVRII regions of various plant species, including *N. cadamba,* suggests that both the HVRII regions of *N. cadamba* represented two of the six distinct CesA classes (Fig. 2). NcCesA2HVRII was not clustered to any of the six distinct clades. Instead, it was clustered together with HVRII sequences of CslD of various plant species. This is expected as NcCesA2HVRII was earlier found to have high similarities with HVRII regions of cellulose-synthase like proteins of other plants.

In Fig. 2, Bootstrap analysis was conducted and bootstrap values of >70 were clustered together as one family. All amino acid sequences were aligned using the BLOSUM matrix. HVRII domains from all CesA and CslD proteins were retrieved from NCBI database, compiled based on the conserved regions of 5' CYVQFPQ

3' and 5' GWIYGS 3' flanking the HVRII regions 5' ALYG 3' and 5' VISCG 3'. All sequences were renamed for the convenience of this study. HVRII regions of *N. cadamba* found were marked by green circles and blue triabgle. Clades I and III are associated with primary cell wall synthesis meanwhile clades II and IV are associated with secondary cell wall synthesis. Clade V indicates clustering of CslD protein between species. Abbreviations: Nc = *Neolamarckia cadamba*; *At* = *Arabidopsis thaliana*; *Ptr* = *Populus tremuloides*; *Ptrem* = *Populus tremula*; *Ptrem x Ptr* = *Populus tremula x Populus tremuloides*; *Ptri* = *Populus trichocarpa*; *Eg* = *Eucalyptus grandis*; *Hcan* = *Hibiscus cannabinus*; *St* = *Solanum tuberosum*; *Spp* = *Shorea parvifolia subsp. parvifolia*; *Gh* = *Gossypium hirsitum*; *Ga* = *Gossypium arboreum*; *Gr* = *Gossypium raimondii*; *Pr* = *Pinus radiata*; *Ptom* = *Populus tomentosa*; *Pu* = *Populus ussuriensis*; *Os* = *Oryza sativa*; *Pe* = *Phyllostachys edulis*; *Am* = *Acacia mangium*; *Na* = *Nicotiana alata*; *Zm* = *Zea mays*; *Smoe* = *Selaginella moellendorffii*: and *Gk* = *Gossypoides kirkii*.

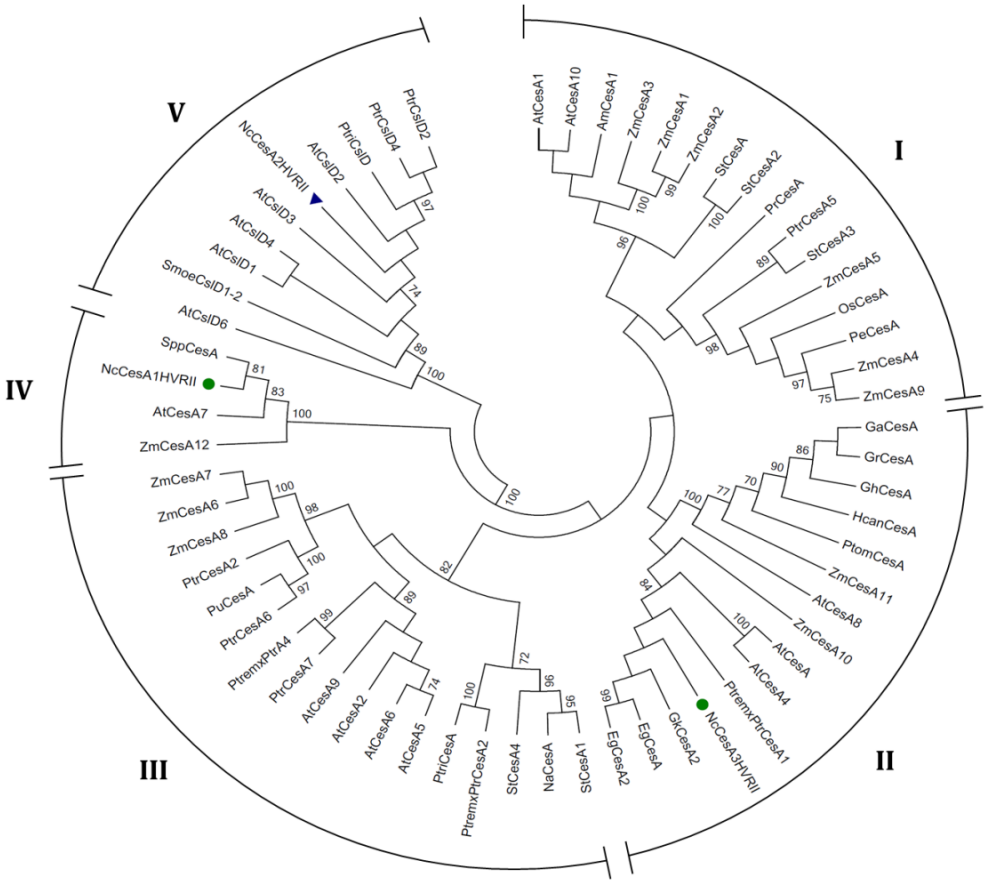

Fig. 2. Unrooted Neighbor-Joining (NJ) phylogenetic tree based on 62 HVRII regions of 21 different plant species.

Sequence analysis via pairwise alignment (Molecular Biological Advanced DNA Analysis Version 6.84 via BLOSUM matrix) based on HVRII regions showed that NcCesA2HVRII protein has very low sequence identity (<50%) with CesAHVRII protein of other plant species. NcCesA2HVRII was found to have only 34% identity with AtCesA1HVRII and 35% identity with NcCesA1HVRII (Table 3). In contrast, NcCesA2HVRII has a high identity (>50%) with CslD-HVRII of other plant species. Sequence pairwise alignment showed that NcCesA2HVRII has 81% identity with AtCslD2 and AtCslD3, and 86% identity with PtrCslD2, PtrCslD4 and PtriCslD suggests the possibility of NcCesA2HVRII being an ortholog of AtCslD2, AtCslD3, PtrCslD2, PtrCslD4, and PtriCslD-HVRII. NcCesA2HVRII should be more appropriately renamed as NcCslD1-HVRII until further research is being done to proof otherwise.

Table 3. Pairwise Alignment of NcCesA2HVRII with Two CesA2 HVRII and Nine CslD-HVRII Proteins using Blosum Matrix

Protein Name	Amino acid sequence identity (%) with NcCesA2HVRII
AtCesA1HVRII	34
NcCesA1HVRII	35
AtCslD1	56
AtCslD2	81
AtCslD3	81
AtCslD4	55
AtCslD6	47
PtrCslD2	86
PtrCslD4	86
PtriCslD	86
SmoeCslD1-2	56

NcCslD1 gene might also probably be involved in the development of xylem as Samuga and Joshi [11] reported of the probable role of *PtrCslD2* in xylem development. They reported that the full length of PtrCslD2 has 76% sequence identity with the full length of AtCslD3. They also found that the full length sequence of both AtCslD2 and AtCslD3 were 83% identical. PtrCslD2-HVRII was also found to be also 83% identical to AtCslD3-HVRII. AtCslD2-HVRII was found to be 83% identical to AtCslD3-HVRII. Although more evidences are needed, but from the comparison of sequence identity between the full length and HVRII amino acid sequence between the same protein, along with the limited information on the specificity of Csl-HVRII region in defining Csl-superfamily, we can use HVRII region of Csl to classify *Csl* gene until more evidences and Csl sequences are available to proof otherwise.

*Nc*CesA1HVRII and *Nc*CesA3HVRII were clustered into different clades. However, both clades were known to be involved in the secondary cell wall synthesis [6]. Thus far only aspen and arabidopsis have representatives of all the six classes of CesA [6] and therefore, it is possible to have only two representative classes of CesA from the developing xylem tissues of *N. cadamba.* However, Haigler and Blanton [12] believed that there should be at least two types of CesA necessary for cellulose synthesis in primary and secondary cell walls of plants. The successful clustering of the new HVRII regions after analyzing 26 HVRII regions of various plant species available at that time further supports the findings that HVRII regions are class-specific regions (CSR) [7]. They proposed that HVRII is not a hypervariable region but is instead a class-specific region as each type of HVRII actually defines a specific class of CesAs in plant species. They further stressed that the CSRs contains conserved motifs which are vital for catalysis. Liang and Joshi [6] also analyzed 56 HVRII regions of various plant species and reaffirmed Vergara and Carpita's [7] findings. In addition, gene expression studies have already confirmed the association of those CesA proteins with primary and secondary cell wall development in a few plant species [2], [11], [13]-[15]. In their phylogenetic tree, they found the presence of only six classes of CesA that are represented in various plant species studied so far.

4. Conclusion

Two *CesA HVRII* and one cellulose-synthase like (*Csl*) HVRII regions were successfully isolated and characterized in the present study by using RT-PCR-mediated amplification of HVRII regions. *Nc*CesA1HVRII and *Nc*CesA3HVRII are found to be involved in the secondary cell wall development. Due to the time constraint of this study, it is not conclusive that *N. cadamba* only have two *CesA* genes and one cellulose-synthase like gene as analysis was done only from total RNA extracted from the young leaves and developing xylem of *N. cadamba*. *CesA* gene expression levels may vary from differing parts of a plant (eg: root, flower, shoots). However, more research should be done in the future to sequence more HVRII regions

and subsequently, cellulose synthase (*CesA*) and cellulose-synthase like (*Csl*) full length genes from different parts of *N. cadamba* (e.g., root, flower, shoots) as there were only two *CesA HVRII* found in this study. Previous studies have shown that there should be at least one representative for primary and secondary cell wall development of plants. Due to class-specificity of HVRII, sequences obtained can be used to design primers based on its conserved regions to isolate the full length *CesA* genes. Functional and mutational studies should also be done to better understand the role of *CesA/Csl* gene in wood development. Upon better understanding of the *CesA/Csl* gene, molecular markers can be developed to genotype seedlings or adult trees with the desired optimal cellulose content required for specific industries, such as plywood, paper and pulp, and biofuel industries.

Acknowledgment

The authors would like to thank all the laboratory assistants and foresters involved in this research programme for their excellent field assistance in sample collection. This work was part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation (SFC) and University of Malaysia Sarawak (UNIMAS) (Grant No.: 02(DPI09)832/2012(1), RACE/a(2)/884/2012(02) & GL(F07)/06/2013/STA-UNIMAS(06)).

References

[1] Richmond, T. (2000). Higher plant cellulose synthases. *Genome Biology*, *1(4)*, 3001.1-3001.6.

[2] Holland, N., Holland, D., Helentjaris, T., Dhugga, K. S., Cazares, B. X., & Delmer, D. P. (2000). A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiology*, *123(4)*, 1313-1324.

[3] Lau, E. T., Ho, W. S., & Julaihi, A. (2009). Molecular cloning of cellulose synthase gene, *SpCesA1* from developing xylem of *Shorea parvifolia* spp. *parvifolia*. *Biotechnology*, *8(4)*, 416-424.

[4] Srivastava, L. M. (2002). *Plant growth and development: Hormones and environment.* California, USA: Academic Press.

[5] Kumar, S., & Fladung, M. (2004). *Molecular genetics and breeding of forest trees: Cloning of new CesA and CSLD gene sequences in Aspen*. Binghamton, New York: The Haworth Press Inc.

[6] Liang, X, & Joshi, C. P. (2004). Molecular cloning of ten distinct hypervariable regions from the cellulose synthase gene superfamily in aspen trees. *Tree Physiology*, *24,* 543-550.

[7] Vergara, C. E., & Carpita, N. C. (2001). β-D-Glycan synthases and the *CesA* gene family: lessons to be learned from the mixed-linkage (1→3), (1→4)β-D-glucan synthase. *Plant Molecular Biology*, *47*, 145-160.

[8] Ranik, M., & Myburg, A. A. (2006). Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree Physiology*, *26*, 545-556.

[9] Ho, W. S., Pang, S. L., Lai, P. S., Tiong, S. Y., Phui, S. L., Liew, K. S., *et al.* (2010). Genomics studies on plantation tree species in Sarawak. *Proceedings of the International Symposium on Forestry and Forest Products 2010: Addressing the Global Concerns and Changing Societal Needs* (pp. 172-182). Malaysia: Kuala Lumpur.

[10] Joker, D. (2000) Seed leaflet *Neolamarckia cadamba* (Roxb.) Bosser (*Anthocephalus chinensis* (Lam.) A. Rich. ex Walp.) Retrieved March 28, 2013, from http://curis.ku.dk/portal-life/files /20648324/ neolamarckia_cadamba_int.pdf.

[11] Samuga, A., & Joshi, C. P. (2004). Cloning and characterization of cellulose synthase-like gene, *PtrCSLD2* from developing xylem of aspen trees. *Physiologia Plantarum*, *120*, 631-641.

[12] Haigler, C., & Blanton, R. L. (1996). New hopes for old dreams: evidence that plant cellulose synthase genes have finally been cloned. *Proceedings of the National Academy of Sciences of the United States of the America* (pp. 93, 12082-2085). USA: Washington D.C..

[13] Pear, J. R., Kawagoe, W. E., Schreckengost, W. E., Delmer, D. P., & Stalker, D. M. (1996). Higher plants contain homologs of the bacterial *CelA* genes encoding the catalytic subunit of cellulose synthase. *Proceedings of the Natural Academy of Sciences of the United States of America* (pp. 93, 12, 637-12642). USA: Washington D.C.

[14] Wu, L., Joshi, C. P., & Chiang, V. L. (2000). A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. *The Plant Journal, 22(6)*, 495-502.

[15] Kalluri, U., & Joshi, C. P. (2003). Isolation and characterization of a new, full length cellulose synthase cDNA from developing xylem of aspen trees. *Journal of Experimental Botany*, *54(390)*, 2187-2188.

Wei-Seng Ho was born in Sarawak. He obtained his PhD degree in genetics at the Department of Genetics, National University of Malaysia, Malaysia (2002). He is a senior lecturer in the Department of Molecular Biology, Faculty of Resource Science and Technology, University of Malaysia Sarawak (UNIMAS). His research interest includes conservation genetics, translational genomics, functional genomics and molecular breeding of commercially important forest tree species, such as Borneo Ironwood, *Shorea* species, Kelampayan, Sawih, Dabai and *Acacia* species.