Original Research Paper

In Silico Analysis of Cellulose Synthase Gene (NcCesA1) in Developing Xylem Tissues of *Neolamarckia Cadamba*

[1]Tiong Shing Yiing, [1]Ho Wei Seng, [2]Pang Shek Ling and [1]Ismail Jusoh

[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*

Abstract: This study reported the isolation and *in silico* characterization of full-length cellulose synthase (CesA) cDNA from *Neolamarckia cadamba*, an important tropical plantation tree species. CesA is a member of processive glycosyltransferases that involved in cellulose biosynthesis of plants. CesA acts as a central catalyst in the generation of plant cell wall biomass or cellulose. It also plays an important role in regulating wood formation. The hypothetical full-length CesA cDNA (NcCesA1; JX134621) was assembled by contig mapping approach using the CesA cDNA sequences from NcdbEST and the amplified 5'-and 3'-RACE PCR sequences. The NcCesA1 cDNA has a length of 3,472 bp with 3,126 bp open reading frame encoding a 1,042 amino acid sequence. The predicted NcCesA1 protein contained N-terminal cysteine rich zinc binding domain, 7 putative Transmembrane Helices (TMH), 4 U-motifs that contain a signature D, D, D, QxxRW motif, an alternating Conserved Region (CR-P) and 2 Hypervariable Regions (HVR). These entire shared domain structures suggest the functional role of NcCesA1 is involved in glycosyltransferation of the secondary cell wall cellulose biosynthesis of *N. cadamba*. Sequence comparison also revealed the high similarity (90%) among NcCesA1 and PtrCesA2 of *Populus tremuloides*. This further implies the involvement of NcCesA1 that catalyzes the cellulose biosynthesis of secondary cell wall rather than primary cell wall. This full-length NcCesA1 cDNA can serve as good candidate gene in association genetics study which leads to Gene-Assisted Selection (GAS) in the *N. cadamba* tree breeding programme. Furthermore, the isolation of new CesA genes from tropical tree genomes is essential for enhancing knowledge of cellulose biosynthesis in trees that has many fundamental and commercial implications.

Keywords: *Neolamarckia Cadamba*, Cellulose Synthase (CesA), Wood Formation, Contig Mapping, Gene-Assisted Selection (GAS), Candidate Gene, Association Genetics

Introduction

Wood is made up of secondary xylem tissues and has a chemical complex of cellulose, lignin, hemicellulose and extractives. Cellulose, a homopolymer consisting of β-1,4-glucan chains, is the most abundant form of living terrestrial biomass makes up the major cell wall biopolymer in plants (Crawford, 1981; Kumar *et al.*, 2009). Polymerization of cellulose chains is catalysed and synthesised by specific plasma membrane-bound Cellulose Synthase complexes (CelS) (Festucci-Buselli *et al.*, 2007). CelS is postulated to be composed of six hexametric rosette subunits where each of the rosette subunit consists six CesA catalytic subunits. Therefore, a total of 36 β-1,

4-glucan chains are produced by a CelS in most of the higher plants. These chains will then linked by hydrogen bonds to form microfibrils which will then further bundled to form macrofibrils. CelS is proved to be encoded by cellulose synthase (CesA) or CesA-related genes. CesA family is a member of Glycosyltransferases (GTs) superfamily under CAZyme family (Ross *et al.*, 2001).

Cellulose synthases are the only identified components of rosettes. Since woody plants are unique in their cellulose biosynthesis, CesA sequence information is useful to serve as the basis for investigating molecular regulation and mechanism of cellulose biosynthesis in tree species (Lu *et al.*, 2008). The first gene discovered to encode for CesA was in a bacteria species, *Acetobacter*

xylinum (Saxena *et al*., 1990; Wong *et al*., 1990). Few years later, cotton (*Gossypium hirsutum*) was the first higher plant species found to have homologs of the bacterial *CelA* gene (Pear *et al*., 1996). *CelA* refers to the cellulose synthase catalytic subunit which is a specific conserved region common to glygosyltransferases found in both bacterial and plants.

Over the years, many CesA genes were continuously discovered, including at least: 10 Genes from *Arabidopsis thaliana* (Richmond and Somerville, 2000); 12 genes from maize (*Zea mays*) (Appenzeller *et al*., 2004); 10 genes from rice (*Oryza sativa*) (Tanaka *et al*., 2003); seven genes from hybrid poplar (*Populus tremula* x *Populus alba*) (Joshi *et al*., 2004) and six genes from *Eucalyptus grandis* (Ranik and Myburg, 2006). Previous studies have been done by different groups of researchers showing that three different CesA genes found to be co-expressed in *A. thaliana*: *At*CesA*4*, *At*CesA*7*, *At*CesA*8* (Taylor *et al*., 2003); poplar (*P. tremula* x *P. alba*): *Ptr*CesA*1*, *Ptr*CesA*2*, *Ptr*CesA*3* (Joshi *et al*., 2004); rice (*O. sativa*): *Os*CesA*4*, *Os*CesA*7*, *Os*CesA*9* (Tanaka *et al*., 2003); barley (*Hordeum vulgare*): *Hv*CesA*1*, *Hv*CesA*2* and *Hv*CesA*6*, *Hv*CesA*4* and *Hv*CesA*7*, *Hv*CesA*8* (Burton *et al*., 2004).

To date, there are considerable amounts of full-length CesA cDNA being published in NCBI but no such information available for *Neolamarckia cadamba* trees. *N. cadamba* or locally known as kelampayan belongs to the family of Rubiaceae. It has been selected as one of the fast growing plantation species for planted forest development in Sarawak (Tchin *et al*., 2012; Lai *et al*., 2013; Tiong *et al*., 2014a; 2014b; Ho *et al*., 2014; Phui *et al*., 2014). The state government of Sarawak has introduced the Forest (Planted Forest) Rules (1997) to encourage the development of commercial planted forests and has set a target of 1.0 million hectares for forest plantations to be established by 2020. It is estimated that 42 million of high quality seedlings are required for the annual planting programme. *N. cadamba* is a large, deciduous and fast growing tree that gives early economic returns within 8-10 years. Under normal conditions, it attains a height of 17 m and diameter of 25 cm at breast height (dbh) within 9 years. It is a lightweight hardwood with a density of 290-560 kg/m^3 at 15% moisture content (Joker, 2000). 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 applications (WAC, 2004). *N. cadamba* also has high potential to be utilized as one of the renewable resource of raw materials for bioenergy production such as cellulosic biofuels in the near future.

Hence, the objectives of this study were: (i) To obtain the full-length CesA cDNA sequences through contig mapping approach by using CesA singletons from the kelampayan tree transcriptome database (NcdbEST) and RACE PCR sequences and (ii) to *in silico* characterize the CesA gene from *N. cadamba*. The full-length CesA cDNA discovered can serve as good candidate gene for association genetics study in *N. cadamba* to detect the potential genetic variants underlying the common and complex adaptive traits.

Materials and Methods

RNA Isolation and First Strand cDNA Preparation

Total RNA was isolated from the developing xylem tissues of a 4-year old *N. cadamba* tree using RNeasy® Midi Kit (QIAGEN GmbH, Germany) with modification. 5'-and 3'-RACE ready cDNA were prepared using SMARTer^{TM} RACE cDNA Amplification Kit (Clontech, USA).

Rapid Amplification of 5'-and 3'-cDNA Ends (RACE)

Two Gene-Specific Primers (GSPs) for CesA gene were designed for 5'-RACE PCR (CesA5'GSP, 5'-TGTAACCGTACCAGGCAGGGCTATGTC-3') and 3'-RACE PCR (CesA3'GSP, 5'-GTGCTTCGATGGGCTCTTGGTTCTG-3') using Primer Premier 5 software based on the parameters given in the protocol of SMARTer^{TM} RACE cDNA Amplification Kit (Clontech, USA). 5'-and 3'-RACE PCR were prepared in separate tubes by mixing 34.5 µL of PCR-grade water, 5.0 µL of 10× Advantage 2 PCR Buffer, 1.0 µL of dNTP Mix (10 mM) and 1.0 µL of 50× Advantage 2 Polymerase Mix. The master mix was then added into a 0.2 mL PCR tubes containing 2.5 µL of RACE-Ready cDNA, 5.0 µL of 10× Universal Primer Mix (UPM) and 1.0 µL of GSP to make a final volume of 50.0 µL. The contents were then mixed gently and subjected to thermal cycling. Touchdown PCR amplification was used: 5 Cycles of incubation at 94°C for 30 sec and 72°C for 3 min; 5 cycles of 94°C for 30 sec, 70°C for 30 sec and 72°C for 3 min; 25 cycles of 94°C for 30 sec, 68°C for 30 sec and 72°C for 3 min. The RACE PCR amplicons were purified from agarose gel by using QIAquick® Gel Extraction Kit (QIAGEN, Germany). Purified PCR product was ligated into pGEM®-T Easy Vector System (Promega, USA) and transformed into competent cells, Escherichia coli JM 109. The recombinant plasmids were isolated and purified using Wizard® Plus SV Minipreps DNA Purification System (Promega, USA) according to the manufacture's protocol. After verification, the purified plasmids were sent for sequencing in both forward and reverse direction. The sequencing reactions were performed by using ABI Prism^{TM} Bigdye^{TM} terminator cycle sequencing Ready reaction kit V.3.1 (Applied Biosystems, USA) and analysed on a ABI 3730XL

capillary DNA sequencer (Applied Biosystems, USA). The 5'-and 3'-RACE cDNA sequences were manually edited using Chromas Lite version 2.01 programme to remove the vector sequence.

Hypothetical Full-Length CesA *cDNA Sequences Assembly*

Singletons of CesA gene were selected from the kelampayan tree transcriptome database (NcdbEST) (Ho *et al*., 2014). These singletons were blast again NCBI database to search for sequence homology and binding position on the respective gene. Subsequently, the singletons were grouped according to the alignment score and position on gene. Singletons which have overlapping fragment were then identified and jointed together with 5'-and 3'-RACE cDNA sequences to form a full-length sequence via contig mapping approach for *N. cadamba* (NcCesA1).

Sequence Analysis of Full-Length CesA *Gene*

Sequence homology search for NcCesA1 cDNA were performed against GenBank non-redundant nucleotide sequence using the NCBI Basic Local Alignment Search Tool (BLAST) server. ExPASy (Expert Protion Analysis System) translate tool (Gasteiger *et al*., 2003) was use to change the nucleotide sequence into amino acid sequence and to identify the Open Reading Frame (ORF). Amino acid sequence was used to predict the domains, motifs and hypervariable regions using EBI-ClustalW2 multiple alignment tool and CLC Main Workbench version 5.0 software (CLC bio, Denmark). The transmembrane helices in protein were predicted using TMHMM Server version 2.0 (Krogh *et al*., 2001). Signal peptide was predicted using SignalP 4.0 Server (Petersen *et al*., 2011).

Three-Dimensional (3-D) Protein Structure Prediction

There is no homologous protein structure was found in the Protein Data Bank (PDB). Thus, threading protein or ab initio protein modeling using I-TASSER server (Zhang, 2008; Roy *et al*., 2010) was used for protein structure and function prediction. Predicted 3-D structures of NcCesA1 were viewed using Jmol, an open-source Java viewer for chemical structures in 3-D with features for chemicals, crystals, materials and biomolecules (Jmol, 2012). Vector Alignment Search Tool (VAST) provided by NCBI was used to identify similar 3-D structures in the Molecular Modeling Database (MMDM) and compared with predicted CesA protein structures. Sequence similarity (% Id) for the parts of the protein that have been superimposed and SCORE (the VAST structure-similarity score) that

reflects the quality of superimposed elements were recorded (Panchenko and Madej, 2004).

Phylogenetic Analysis of Full-Length CesA *Genes*

The full-length CesA genes were obtained from NCBI nucleotide database and the GenBank accession numbers were recorded. All of the CesA genes of *Arabidopsis thaliana* were obtained from The Arabidopsis Information Resource (TAIR) database (http://arabidopsis.org) using the gene models accession. CesA genes of rice (*Oryza sativa*) were obtained from OryGenesDB database (http://orygenesdb.cirad.fr/) with locus name accession (Droc *et al*., 2005). Protein sequence of selected genes was aligned using EBI-ClustalW2 multiple alignment tool and neighbour-joining tree was generated using MEGA version 5 software (Tamura *et al*., 2011).

Results

Assembly of CesA *cDNA from Singletons*

A total of 13 CesA singletons were obtained from NcdbEST (Ho *et al*., 2014). These singletons showed high maximum identity value scores of more than 80% as shown in Table 1. The best hits chosen were the complete CesA coding sequences from other species, such as *E. camaldulensis* (HQ864585), *B. luminifera* (FJ410444 and FJ410444), *P. tremula* x *P. tremuloides* (AY573571 and AY573572), *P. tremuloides* (AY055724), *P. tomentesa* (HQ585870, HQ585871 and HQ585872) and *S. parvifolia* (GQ338420). Pairwise and multiple alignment of 13 singletons were carried out to find the suitable singletons for contig-mapping of a partial hypothetical CesA gene. Five singletons (Ncdx015G11, Ncdx017H02, Ncdx029C01, Ncdx092C07 and Ncdx007A11) were selected for contig mapping to produce a partial consensus CesA cDNA sequence, namely CesA*1*, with a length of 1,387 bp. The BLAST result of CesA*1* against NCBI nucleotide database showed that this gene has high similarity with CesA nucleotide sequences of other tree species, such as *Shorea*, *Populus*, *Betula* and *Eucalyptus*. Table 2 shows that CesA*1* covers more than 80% of other CesA sequences with high nucleotide similarity and as much as 83% identity.

Alignment of CesA*1* with closely related full-length CesA cDNAs of other species showed that this hypothetical partial sequence was located at the 3'-end with a predicted stop codon (TGA) at nucleotides 1,167 to 1,169 as shown in Fig. 1. The stop codon was confirmed when the translated protein (Fig. 2) stopped at the predicted nucleotide (nucleotide 1,166). The full-length NcCesA1 is predicted to be approximately 3,100 bp, so about 2,000 nucleotides at the 5'-end yet to be identified.
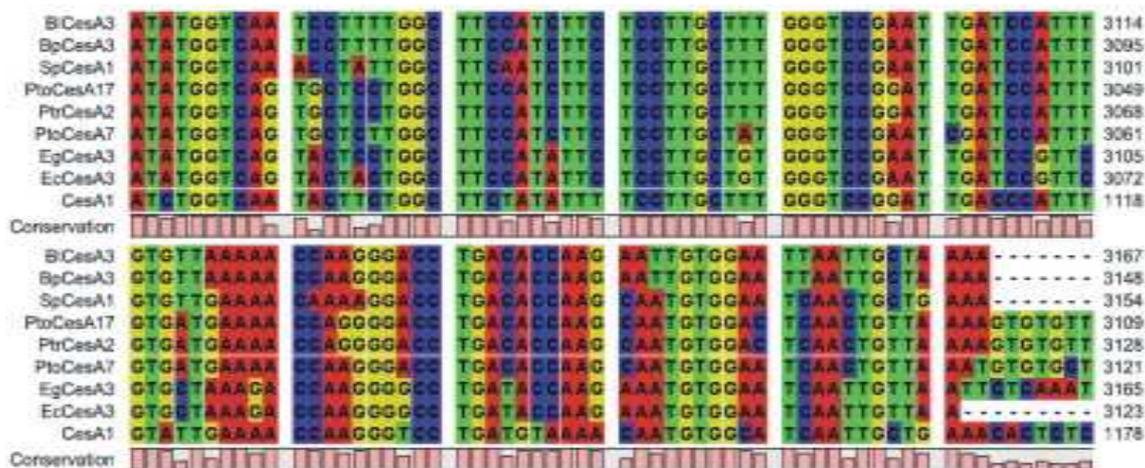
Fig. 1. A part of the alignment result of CesA1 with full-length CesA cDNA from *Betula luminifera* (BlCesA3; FJ410445), *Betula platyphylla* (BpCesA3; EU591531), *Eucalyptus camaldulensis* (EcCesA3; HQ864585), *Eucalyptus grandis* (EgCesA3; DQ014507), *Populus tomentosa* (PtoCesA7; HQ585870 and PtoCesA17; HQ585871), *Populus tremuloides* (PtrCesA2; AY095297) and *Shorea parvifolia* (SpCesA1; GQ338420) showing the stop codons as indicated in the black square box

MSQMNFEKKFGQSAIFVTSTLMIEGGVPPSSSPAALLKEAIHVISCGYEDKTEWGTELGWIYG-
SITEDILTGFKMHCRGWRSIYCMPKLAAFKGSA-
PINLSDRLNQVLRWALGSVEIFFSRHSPAWYGYKGGNLKWLERFAYVNTTVYPFTSIP-
LLAYCTLPAICLLTGKFIMPEISTFASLFFIALFLSIFATGI-
LELRWSGVSIEEWWRNEQFWVIGGISAHLFAVIQGLLKVLAGIDTNFTVTSKAADDEDFGE-
LYAFKWTTLLIPPTTILIINLVGVVAGISDAINNGYQSWGPLFGKLFFAFWVIIH-
LYPFLKGLMGKQNRTPTIVIIWSILLASIFSLLWVRIDPFVLKTKGPDVKQCGINC*

Fig. 2. Translated protein sequence of CesA1 contained 368 amino acids. Stop codon was represented by a "*"

Table 1. Singletons from NcdbEST and the BLAST results with CesA genes of other species from NCBI nucleotide database

| No. | Singleton | Length (bp) | BLAST results | | |
			Highest identity (%)	E value	Subject accession no.
1	Ncdx015G11	609	83	$4e^{-125}$	HQ585871
2	Ncdx017H02	666	86	$1e^{-56}$	HQ864585
3	Ncdx018G11	507	82	$8e^{-108}$	FJ410444
4	Ncdx029C01	569	83	$3e^{-140}$	FJ410445
5	Ncdx033F05	499	87	$5e^{-47}$	AY573571
6	Ncdx058G12	609	81	$4e^{-157}$	AY573572
7	Ncdx066F10	723	83	$2e^{-155}$	FJ410444
8	Ncdx075G12	760	84	0	HQ864585
9	Ncdx077H10	412	85	$7e^{-57}$	HQ585872
10	Ncdx092C07	625	82	$3e^{-178}$	HQ585870
11	Ncdx098E04	433	82	$3e^{-30}$	GQ338420
12	Ncdx105F07	544	83	$2e^{-45}$	FJ410444
13	Nxdc007A11	697	83	$3e^{-140}$	FJ410445

Table 2. BLAST sequence homology analysis of CesA1 sequence with GenBank database nucleotide collection (nr/nt)

Organism	Accession number	Coverage (%)	Identity (%)	E value
Shorea parvifolia ssp. *parvifolia*	GQ338420	81	83	0
Populus tomentosa	HQ585870	81	83	0
Populus tomentosa	HQ585871	81	83	0
Betula platyphylla	EU591531	83	82	0
Betula luminifera	FJ410445	83	82	0
Populus tremuloides	AY095297	81	82	0
Eucalyptus camaldulensis	HQ864585	81	82	0
Eucalyptus grandis	DQ014507	81	81	0

Amplification of 5'-and 3'-cDNA Ends Sequences and Full-Length NcCesA1 cDNA Assembly

Since the CesA1 cDNA was incomplete, Gene Specific Primers (GSPs) were designed to find the 5'-and 3'-termini. Overlapping forward and reverse primers were designed for 5'-and 3'-RACE PCR amplification. The product size of 5'-ends and 3'-ends was predicted from the alignment with full-length CesA cDNA of other species and was expected to be about 2,400 bp and 1,060 bp, respectively. Figure 3 shows the diagrammatic representative of the location of the Gene Specific Primers (GSPs) with the predicted product size. The 5'-and 3'-RACE products were sent for sequencing in both Forward (F) and Reverse (R) direction to get the maximum nucleotide sequence reading. Full sequences of three 3'-cDNA ends, 3'CesA04, 3'CesA05 and 3'CesA18 were successfully obtained with the lengths of 1,030 bp, 1,063 bp and 1,078 bp, respectively. Alignment of these three 3'-cDNA ends showed 99% similarity in nucleotide sequence and the only variation was at the poly-A tail region. Since the nucleotide differences did not affect the protein translation, 3'CesA18 was chosen for the full-length contig mapping because the sequence length (1,063 bp) was closest to the expected size (1,060 bp).

For 5'-cDNA ends, the complete sequences could not be generated from the forward and reverse sequencing readings due to the longer expected nucleotide sequences of 5'CesA13 and 5'CesA17, which were 1,900 bp and 2,600 bp, respectively. 5'CesA13 and 5'CesA17 were considered amplified from the same gene because both of their forward sequences showed 99% similarity when aligned together. The reverse sequences of 5'CesA13 and 5'CesA17 were not similar but they overlapped at both ends (Fig. 4). 5'CesA13 was predicted to be the 'shortened' version of 5'CesA17 at the middle part of the sequence region due to PCR errors. According to Pienaar *et al.* (2006), editing errors might happen during DNA polymerase-catalysed enzymatic copying or DNA fragments were subjected to thermal damage during PCR, especially for long amplification.

All the 5'-RACE sequences and 3'CesA18 were used in contig mapping of the hypothetical full-length NcCesA1 cDNA as shown in Fig. 4. Full-length cDNA of closely related CesA gene from *B. luminifera* (BlCesA3; FJ410445), which had high sequence coverage (83%) and high sequence similarity (82%) was used as the reference gene in the contig mapping. The 5'-cDNA ends (5'CesA13 and 5'CesA17 forward and reverse sequences) and 3'-cDNA ends (3'CesA18) were successfully mapped to produce a hypothetical full-length NcCesA1 cDNA sequence with a length of 3,784 bp.

The BLAST results of the hypothetical full-length NcCesA1 (GenBank accession number: JX134621) cDNA against NCBI nucleotide database supported this gene as cellulose synthase with the identity and coverage value up to 81 and 91%, respectively as shown in Table 3. The common genetic features in CesA protein were identified in NcCesA1 as shown in the protein alignment (Fig. 5) and the diagrammatic representation of the protein structure is shown in Fig. 6.

3-D Structure Prediction of NcCesA1

The full-length protein sequence of NcCesA1 with 1,042 amino acids was used to predict 3-D secondary protein structure using I-TASSER server (Zhang, 2008; Roy *et al.*, 2010). The predicted protein model of NcCesA1 is shown in Fig. 7. The predicted 3-D structure of ring-finger or zinc binding domain for NcCesA1 is shown in Fig. 8.

Phylogenetic Analysis of NcCesA1

A total of 56 CesA genes from various species were obtained from NCBI database, *Arabidopsis* TAIR database (http://arabidopsis.org) and rice OryGenesDB (http://orygenesdb.cirad.fr/). The genes were named and the designated abbreviations were: *Arabidopsis thaliana* (At), *Betula luminifera* (Bl); *Betula platyphylla* (Bp); *Eucalyptus camaldulensis* (Ec); *Eucalyptus grandis* (Eg); *Oryza sativa* (Os); *Populus tomentosa* (Pto); *Populus tremuloides* (Ptr); *Populus tremula* x *Populus tremuloides* (Ptt); *Shorea parvifolia* (Sp); *Zea mays* (Zm). Complete protein sequences of these CesA genes were aligned with NcCesA1 amino acid sequence. Neighbour-joining phylogenetic tree was generated to show the relationship among all the selected proteins (Fig. 9).
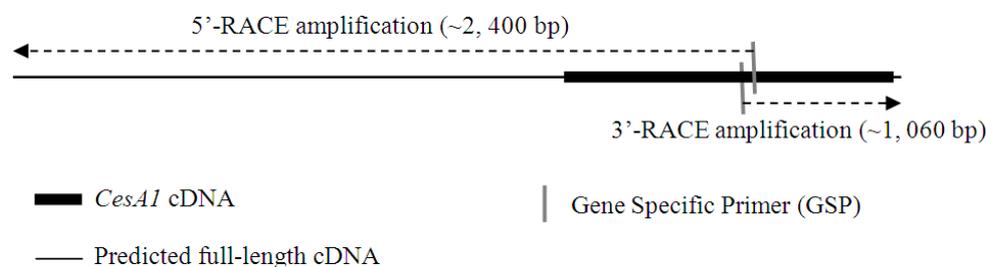
Fig. 3. Diagrammatic representative showing the location of the GSPs for CesA1 5'-and 3'-ends amplification with the predicted amplicon size of 2,400 bp and 1,060 bp, respectively
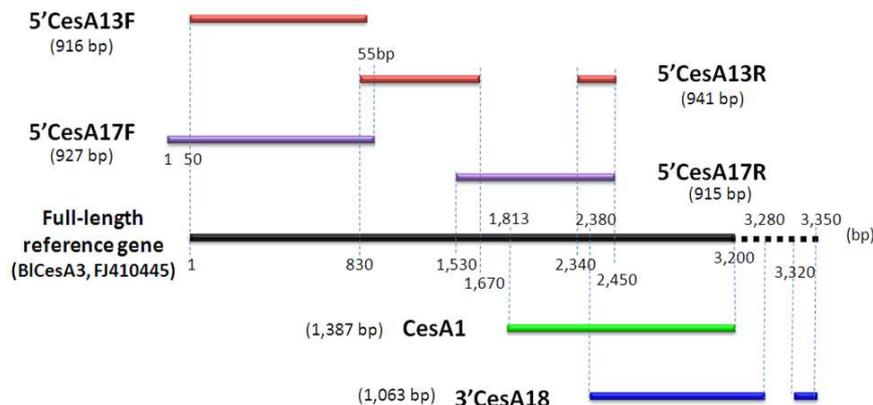
Fig. 4. Diagrammatic representative of contig mapping of hypothetical full-length NcCesA1 cDNA
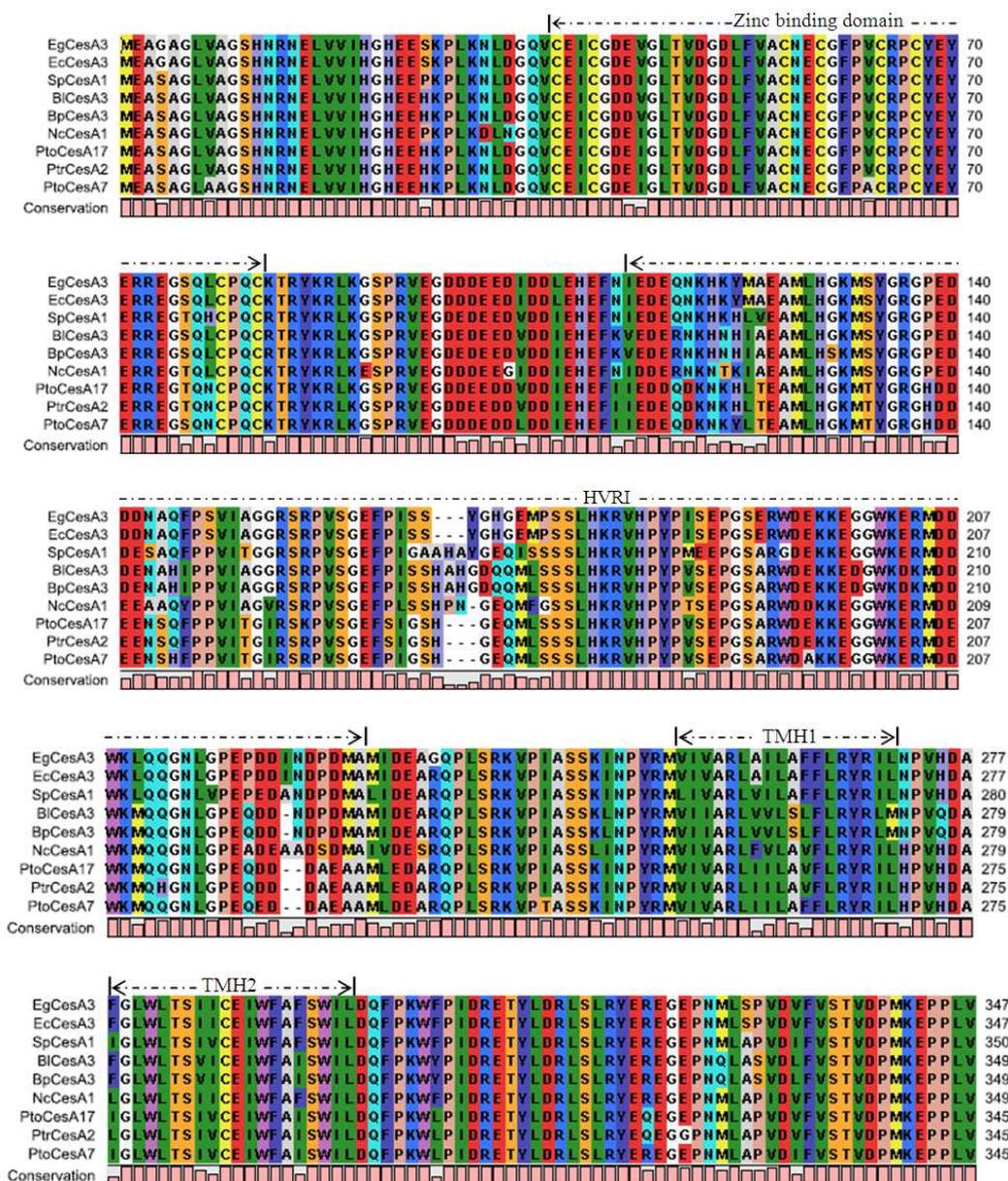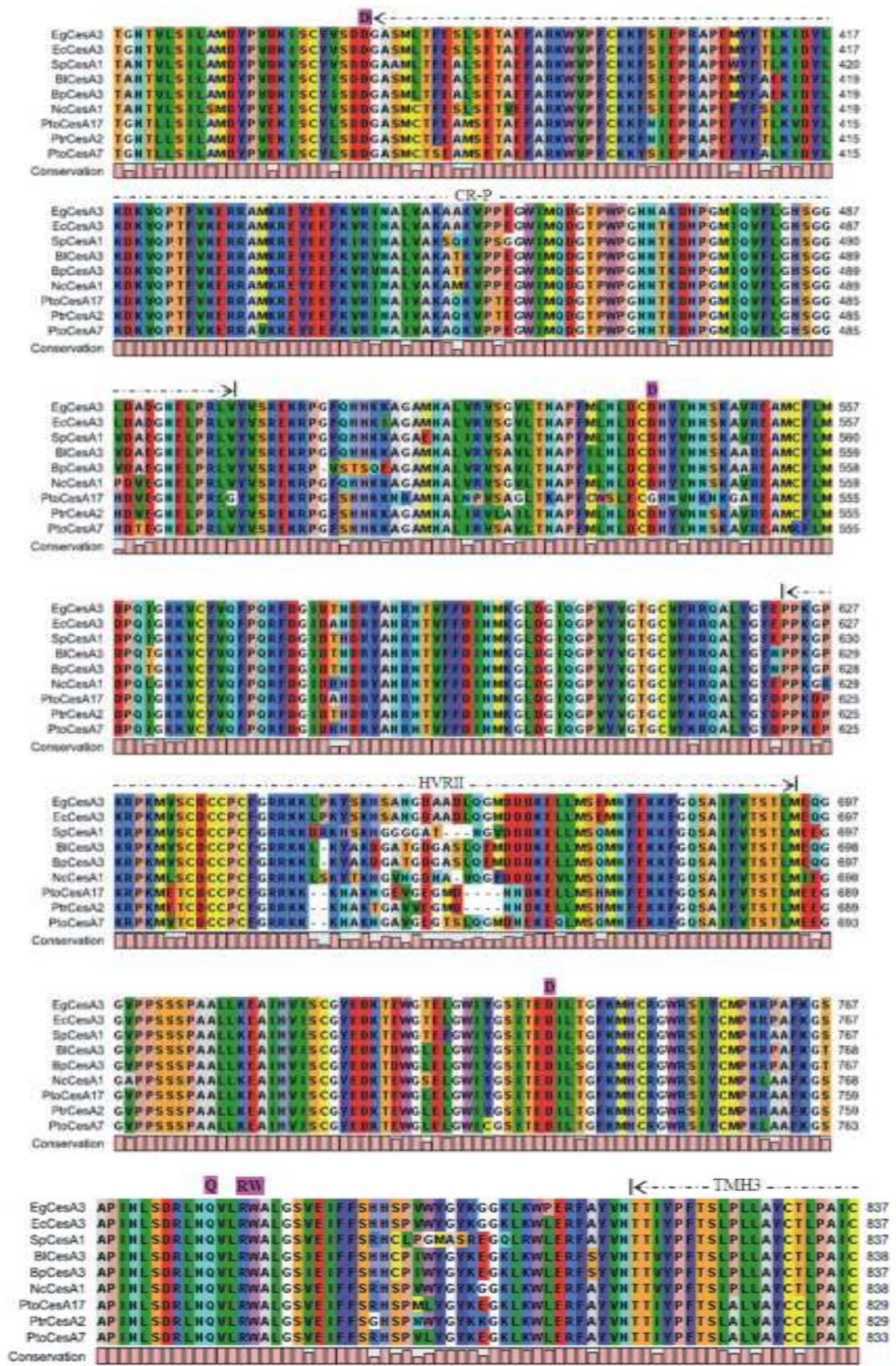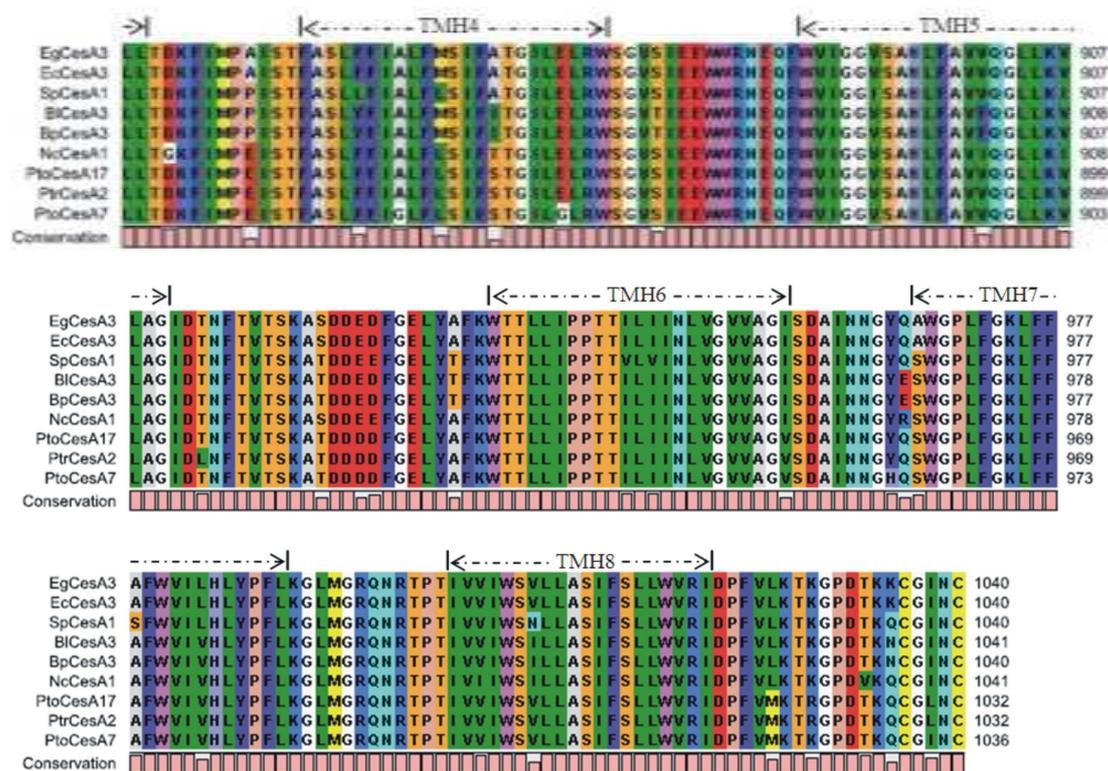
Fig. 5. Amino acid sequence alignment of NcCesA1 with the CesA protein sequences of *B. luminifera* (BlCesA3; FJ410445), *B. platyphylla* (BpCesA3; EU591531), *S. parvifolia* (SpCesA1; GQ338420), *P. tomentosa* (PtoCesA7; HQ585870), *E. grandis* (EgCesA3; DQ014507) HQ585871) and *P. tremuloides* (PtrCesA2; AY095297). The conserved regions, domains and motifs are shown by the dotted arrows and alphabets above the alignment sequences. HVR: Hyperviariable Region; TMH: Transmenbrane Helic; CR-P: Plant-specific conserved region; D, D, D, QxxRW: U-motif
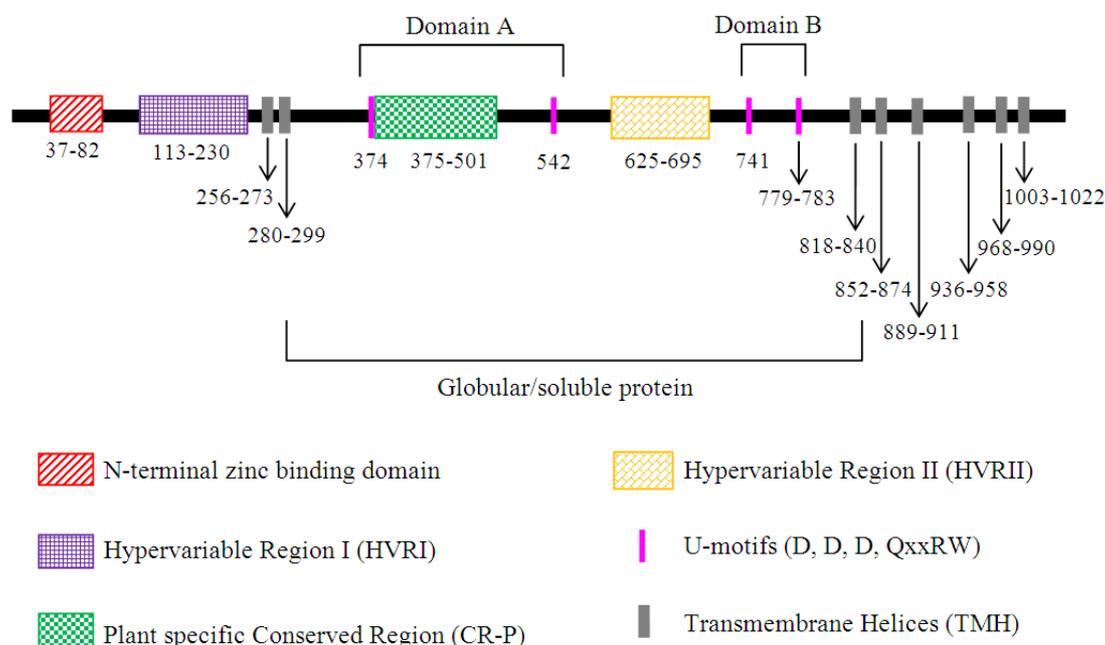
Fig. 6. A diagrammatic representation of NcCesA1 protein. The amino acid sequence position of domains, motifs and targeted region are indicated by the numbers

Fig. 7. NcCesA1 predicted 3-D protein structure using protein threading and ab initio modelling approach. α-Helix (pink) and β-strand (yellow) were observed

| (a) | (b) |
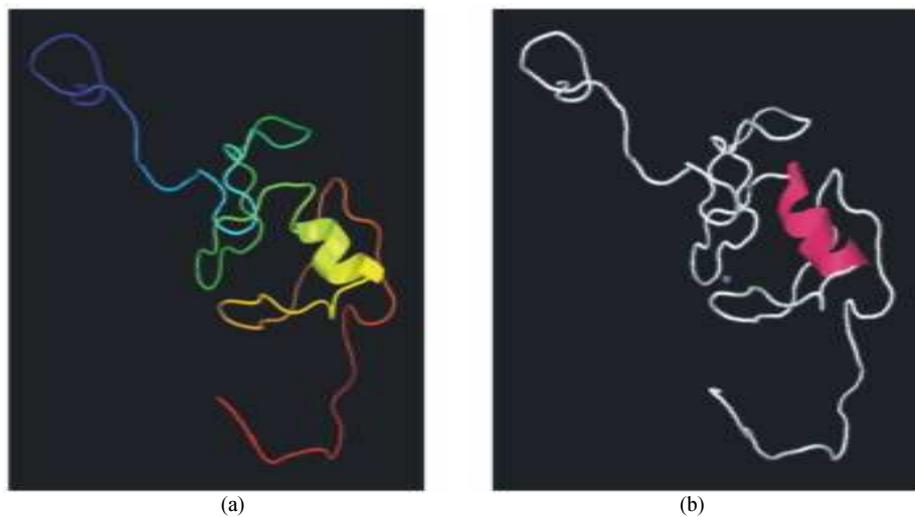
Fig. 8. Ring-finger structure of NcCesA1 (a) predicted using Phyre2 server (Kelley and Sternberg, 2009) has high similarity to the ring-finger structure of Arabidopsis IRX3 (AtCesA7) (b)

Table 3. BLAST sequence homology analysis of full-length NcCesA1 cDNA against GenBank database nucleotide collection (nr/nt)

Organism	Accession number	Coverage (%)	Identity (%)	E value
Betula luminifera	FJ410445	90	81	0
Betula platyphylla	EU591531	90	81	0
Shorea parvifolia	GQ338420	90	81	0
Populus tomentosa	HQ585870	89	80	0
Eucalyptus grandis	DQ014507	91	80	0
Eucalyptus camaldulensis	HQ864585	89	80	0
Populus tomentosa	HQ585871	90	80	0
Populus tremuloides	AY095297	90	80	0

Discussion

The hypothetical full-length CesA cDNA (NcCesA1; JX134621) was assembled by contig mapping approach using five CesA singletons (Ncdx015G11, Ncdx017H02, Ncdx029C01, Ncdx092C07 and Ncdx007A11) from NcdbEST and the amplified 5'-and 3'-RACE PCR sequences (Fig. 4). The NcCesA1 showed high sequence similarity with CesA gene of other tree species, such as *Betula*, *Shorea*, poplar or aspen (*Populus*) and *Eucalyptus*. The full-length CesA cDNA of *Betula* and *Shorea* species showed the highest sequence similarity (81%) with NcCesA1, with the coverage scored up to 90%. Alignment of the NcCesA1 sequence with other closely related full-length CesA genes from tree species had predicted the start (ATG) and stop codon (TGA) codons at nucleotides 73-75 and nucleotides 3,196-3,198, respectively. NcCesA1 has a total length of 3,472 bp, which consists of 3,126 bp of coding sequence (cds) with 72 bp of 5'-UTR and 277 bp of 3'-UTR.

The full-length NcCesA1 with a 3,123 bp of ORF was translated into a 1,042 amino acid sequence. The length of the NcCesA1 protein sequence was about the same as those closely-related CesA genes as listed in Table 3. These protein sequences were in the range of 1,032-1,041 amino acids. CesA3 protein of two *Eucalyptus* species, *E. camaldulensis* (EcCesA3; HD864585) and *E. grandis* (EgCesA3; DQ014507) showed the highest protein sequence similarity (90%) to NcCesA1. Other CesA proteins also showed that the amino acid sequences are highly conserved among them and scored in between 85 to 94% similarity with NcCesA1 or among each other.

CesA family genes are characterised by a conserved U-motif D, D, D, QxxRW (x represents any amino acid), which can be divided into two sub-domains: Domain A (D, DxD) and Domain B (D, QxxRW). NcCesA1 protein which contains this motif was predicted to be involved in CesA enzyme substrate binding and catalytic activities (Beeckman *et al*., 2002; Saxena *et al*., 2001). Domain A consists of three aspartic acids (D). The first D found in amino acid 374 was widely spaced from the second D (amino acid 542) in NcCesA1. This domain was predicted to be involved in both processive and non-processive UDP-glucose substrate binding (Richmond, 2000). Addition of a few or more sugar residues to a growing chain is catalysed by processive enzyme while non-processive enzyme only catalyses one to one addition reaction. Domain B of NcCesA1 was found at amino acid 741 for D residue, QxxRW was found at amino acid 779 to 783 (Fig. 6). This second motif was predicted to be the catalytic site in CesA and possesses only processive catalytic activity (Richmond, 2000).

A highly conserved zinc binding domain ($Cx_2Cx_{12}FxACx_2Cx_2PxCx_2CxEx_5Gx_3Cx_2C$, where x is any amino acid) was found at amino acid 37 to 82. This N-termini domain was predicted to correspond to the protein interactions in determining cell microfibril structures (Delmer, 1999). This CxxC motif which is only found in CesA proteins can distinguish cellulose synthase family from Cellulose Synthase-Like (CLS) family. Therefore, NcCesA1 was confirmed to be cellulose synthase and not CLS protein.

Two Hypervariable Regions (HVRI and HVRII) were found within the NcCesA1 open reading frame. HVRI encodes 118 amino acids at position 113 to 230 near to the N-terminus. HVRII which is positioned in between two sub-domains have shorter length (70 amino acids) at position 625-695 (Fig. 6). The hypervariable region was reported to be able to classify CesA protein into subclasses (Vergara and Carpita, 2001). HVRII variable in CesA paralogous sequences but are highly conserved between CesA orthologs (Nairn and Haselkorn, 2005; Tan *et al*., 2014). This suggested that HVRII can be termed as Class-Specific Region (CSR) (Vergara and Carpita, 2001). To date, lots of CesA genes have been isolated and identified, for example twelve distinct CesA genes were identified in Arabidopsis and nine in maize (Holland *et al*., 2000).

Eight transmembrane helices were predicted throughout the whole protein sequence of NcCesA1 and the output generated by TMHMM 2.0 server (Krogh *et al*., 2001) Two transmembrance helices were predicted to be formed after the plant-specific conserved region, within the amino acid 256 to 299 (Fig. 6). Another six transmembrane helices were predicted at the carboxyl terminus (near to 3'-end) starting from amino acid 779. The region between sets of transmembrane helices is known as the globular domain or soluble domain (Richmond, 2000). These six transmembrane helices will then form a loop that extends into the cytoplasm. This suggests that CesA is an integral membrane protein. Plant-Conserved Region (CR-P) is the plant-specific region that was suggested to be implicated in the cellulose biosynthesis at "rosette" complexes (Roberts *et al*., 2002).

The homologous modeling cannot be used to predict 3-D protein structure of NcCesA1 as to date, there is no full-length 3D structure being deposited in Protein Data Bank (PDB). Thus, native NcCesA1 secondary protein structure was predicted using threading and *ab initio* modeling approach using I-TASSER server (Zhang, 2008; Roy *et al*., 2010). Protein threading or also known as fold recognition, models proteins that have the same fold as proteins of known structures, but do not match any homologous protein structures in PDB. *Ab initio* or *de novo* protein modeling involves the prediction of 3-D protein structure based on physical principles without referring to previous solved structures.

The predicted protein model of NcCesA1 has a confidence level scored at-0.85 (-5< C-score <2). This indicated that the quality of the predicted model was

acceptable. A TM-score of more than 0.5 (0.61±0.14) suggested that the topology of predicted NcCesA1 structure was correct. TM-score was proposed to measure the structural similarity between two structures (Zhang and Skolnick, 2004). NcCesA1 model as shown in Fig. 7 has the highest cluster density (0.1665) among other models, with a total of 177 decoys (low temperature replica). The more the number of structural decoys used, the higher the cluster density, which means the quality of prediction improves.

The overall 3-D structure showed that NcCesA1 contains more α-helix structures than β-strands. Residue 19 to 110 showed the highest sequence identity (75%) with ring-finger domain of *Arabidopsis* IXR3 (AtCesA7) (PDB ID: 1WEO). The predicted 3D structure of this domain for NcCesA1 and AtCessA7 was almost the same as shown in Fig. 8. Ring-finger or zinc binding domain was conserved at the N-terminal among CesA proteins of different species and therefore they showed high similarity in sequence and structure. In NcCesA1, this domain was located between amino acid 20 to 110 as shown in the diagrammatic representative of NcCesA1 protein in Fig. 5. This structure was suggested to be involved in determining the cell microfibril structures during protein interactions (Delmer, 1999).

Six main clades can be distinguished from the un-rooted neighbour-joining tree of CesA protein sequences of various species (Fig. 9). Previous studies had shown that CesA from higher plant can be grouped into six distinct clades, such as in *Populus* (Liang and Joshi, 2004; Djerbi *et al.*, 2005; Nairn and Haselkorn, 2005) and *Eucalyptus* (Ranik and Myburg, 2006). Vergara and Carpita (2001) had discovered that the appearance of different CesA sub-classes was primarily determined by HVRII and therefore, this region was termed as Class-Specific Region (CSR). They also showed that the topology of phylogenetic tree is mainly determined by HVRII regions, where such topology remained when either HVRII regions or full-length protein sequences were used in the phylogenetic analysis. All of the six clades contain CesA members associated with either primary or secondary cell wall development. In this study, NcCesA1 was clustered in group III, the clade that is associated with secondary cell wall formation (Fig. 9). Another two clades that are involved in secondary cell wall development are groups V and VI. Clade I, II and IV consisted of CesA proteins associated with the developmental process in primary cell walls. Ranik and Myburg (2006) also grouped six *Eucalyptus grandis* CesA and other 52 CesAs from dicot, monocot and gymnosperm species into three clades of primary cell wall-associated CesAs and three clades of secondary cell wall-associated CesAs. The similar phylogenetic topology also can be observed from 18 CesA genes of *P. Trichocarpa* (Djerbi *et al.*, 2005).

The involvement of CesA proteins in clade I, II and IV in primary cellulose synthesis had been discussed and proven by some expression and functional studies (Arioli *et al.*, 1998; Beeckman *et al.*, 2002; Desprez *et al.*, 2002; Ellis *et al.*, 2002; Caño-Delgado *et al.*, 2003). A study carried out by Arioli *et al.* (1998) showed that clade II CesA might carry out primary cellulose synthesis. This study showed that mutated *Arabidopsis* AtCesA1 (radial sewlling1, rsw1) gene reduced cellulose deposition and crystallisation in young, expanding cell of seedlings. AtCesA3 or cev1 in clade I which expressed predominantly in root tissues that actively carry out primary cell wall synthesis, was shown to decrease the cellulose content in root tissues when the gene was mutated (Ellis *et al.*, 2002). Fagard *et al.* (2000) also confirmed the role of AtCesA6 or PROCUSTE1 (PRC1) (clustered in clade IV of Fig. 9) in primary cellulose synthesis of roots and dark-grown hypocotyls in *Arabidopsis*.

CesAs that associate with cellulose biosynthesis in secondary cell walls (clades III, V and VI) also have been predicted through functional and expression studies (Taylor *et al.*, 2002; Appenzeller *et al.*, 2004; Djerbi *et al.*, 2005; Nairn and Haselkorn, 2005; Ranik and Myburg, 2006). Recently, three CesA genes (ZmCesA10, ZmCesA11 and ZmCesA12) were found to be highly expressed in the stalk, where the tissues are rich in secondary cell walls (Appenzeller *et al.*, 2004), however these three CesA genes were clustered separately in clades V, VI and III, respectively. A study of CesA gene in *E. grandis* also found three proteins, EgCesA1, EgCesA2 and EgCesA3 grouped in three different clades of secondary cell wall-associated cellulose synthesis, similar with clades VI, V and III, respectively as shown in Fig. 9 (Ranik and Myburg, 2006).

The NcCesA1 which was grouped under clade III is predicted to associate with cellulose biosynthesis in secondary cell wall. The close relationship of NcCesA1 with *Arabidopsis* AtCesA7, which share 86% protein sequence similarity, supports this assumption. AtCesA7 or *Arabidopsis* irregular xylem3 (irx3) was confirmed to be involved in secondary cell wall development since the mutation of this gene led to the collapse of xylem cells due to a great deficiency of cellulose content (Turner and Somerville, 1997; Taylor *et al.*, 1999). Besides that, EgCesA3 also showed high sequence similarity (90%) with NcCesA1 and clustered in the same group in the phylogenetic analysis. EgCesA3 has been proven to be highly expressed in xylem tissues that carried out secondary cellulose biosynthesis (Ranik and Myburg, 2006). *S. parvifolia* CesA1 that is closely related to NcCesA1 also showed high similarity to PtrCesA2 of *P. tremuloides* that catalysed cellulose biosynthesis in secondary cell walls (Lau *et al.*, 2009).
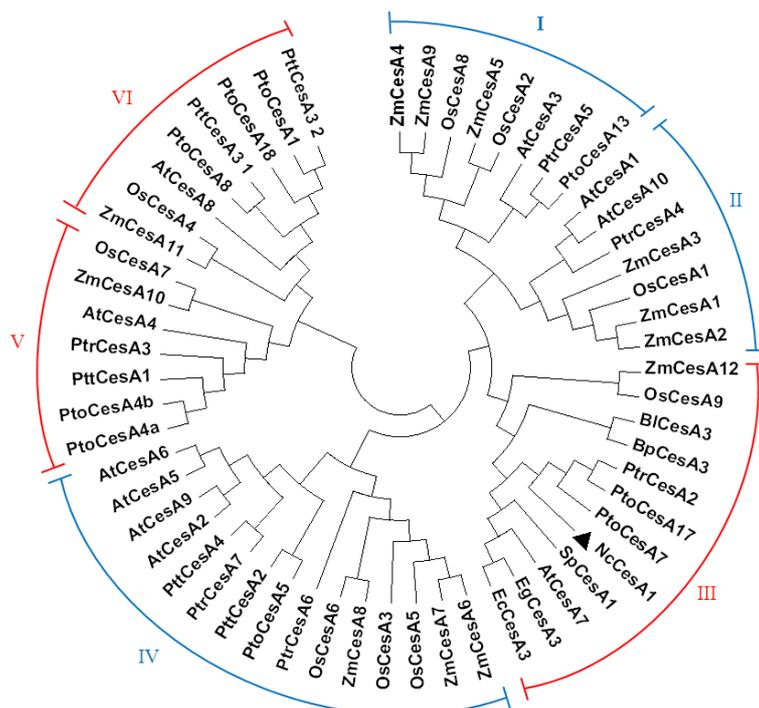
Fig. 9. Neighbour-joining phylogenetic tree generated from multiple alignments of full-length amino acid sequences of NcCesA1 and selected species. Clades I to VI indicate the six main clades that containing the CesAs associated with primary cell wall synthesis (non-symmetry blue circle) and secondary cell wall synthesis (non-symmetry red circle). The accession numbers of CesA used are: NcCesA1 (JX134621); AtCesA1 (AT4G32410); AtCesA2 (AT4G39350); AtCesA3 (AT5G05170); AtCesA4 (AT5G44030); AtCesA5 (AT5G09870); AtCesA6 (AT5G64740); AtCesA7 (AT5G17420); AtCesA8 (AT4G18780); AtCesA9 (AT2G21770); AtCesA10 (AT2G25540); BlCesA3 (FJ410445); BpCesA3 (EU591531); EcCesA3 (HQ864585); EgCesA3 (DQ014507); PtoCesA1 (DQ004570); PtoCesA4a (FJ534554); PtoCesA4b (HQ585869); PtoCesA5 (HQ585873); PtoCesA7 (HQ585870); PtoCesA8 (HQ585867); Pto-CesA13 (HQ585872); PtoCesA17 (HQ585871); PtoCesA18 (HQ585868); PtrCesA2 (AY095297); PtrCesA3 (AF527387); PtrCesA4 (AY162181); PtrCesA5 (AY055724); PtrCesA6 (AY196961); PtrCesA7 (AY162180); PttCesA1 (AY573571); PttCesA2 (AY573572); PttCesA3_1 (AY573573); PttCesA3_2 (AY573574); PttCesA4 (AY573575); SpCesA1 (QG338420); ZmCesA1 (AF200525); ZmCesA2 (AF200526); ZmCesA3 (AF200527); ZmCesA4 (AF200528); ZmCesA5 (AF200529); ZmCesA6 (AF200530); ZmCesA7 (AF200531); ZmCesA8 (AF200532); ZmCesA9 (AF200533); ZmCesA10 (AY372244); ZmCesA11 (AY372245); ZmCesA12 (AY372246); OsCesA1 (LOC_OS05G08370); OsCesA2 (LOC_OS03G59340); OsCesA3 (LOC_OS07G24190); OsCesA4 (LOC_OS01G54620); OsCesA5 (LOC_OS03G62090); OsCesA6 (LOC_ OS07G14850); OsCesA7 (LOC_OS10G32980); OsCesA8 (LOC_OS07G10770); OsCesA9 (LOC_OS09G25490)

NcCesA1 is predicted to be important and most abundant among CesA protein members that are involved in secondary cell wall development in xylem tissues. In a study of CesA genes in *E. grandis*, EgCesA3 which was grouped in the same clade as NcCesA1 had proven to be the most highly expressed CesA transcript in the expression profile (Ranik and Myburg, 2006). PtrCesA2 which was clustered in the same clade with NcCesA1 also supported this assumption in a study by Samuga and Joshi (2002). In their study, PtrCesA2 was shown to be highly expressed in xylem tissues of *P. tremuloides* compared to PtrCesA1 that was predicted to have the similar role.

Conclusion

The full-length NcCesA1 (JX134621) gene with the total length of 3,472 bp, encoding a 1,042 amino acid, was successfully isolated and characterized from *N. cadamba* by using singletons of CesA from the kelampayan tree transcriptome database (NcdbEST) and sequences generated from the 5'-and 3'-RACE to form a hypothetical full-length CesA sequence. To the best of our knowledge, this is the first report on the assembly of full-length CesA gene in *N. cadamba* through a contig mapping approach. NcCesA1 was successfully characterised by its common genetic features that are conserved among CesA proteins. One of the important features, D,D,D,QxxRW, that was found in NcCesA1 has confirmed that this protein is a cellulose synthase protein and not a cellulose synthase-like protein. Phylogenetic analysis also revealed that NcCesA1 protein is involved in cellulose biosynthesis of the secondary cell walls. We hope this newly isolated CesA gene could pave the way for identifying molecular

mechanisms controlling wood formation in future and will also be candidate for association genetic studies aiming at the production of high value forests (Thumma *et al.*, 2005; Ho *et al.*, 2011; Tchin *et al.*, 2012; Tiong *et al.*, 2014a; Tan *et al.*, 2014). Furthermore, the detailed understanding on the regulation of CesA gene could provide a greater impact on the design of future genetic improvement strategies in the production of better quality wood that is typically present in the secondary walls of xylem in *N. cadamba*.

Acknowledgement

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Appenzeller, L., M. Doblin, R. Barreiro, H.Y. Wang and K.S. Dhugga *et al.*, 2004. Cellulose synthesis in maize: Isolation and expression analysis of the cellulose synthase (CesA) gene family. Cellulose, 11: 287-299. DOI: 10.1023/B:CELL.0000046417.84715.27

Arioli, T., L. Peng, A.S. Betzner, J. Burn and R.E. Williamson *et al.*, 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science, 279: 717-720. DOI: 10.1126/science.279.5351.717

Beeckman, T., G.K.H. Przemeck, G. Stamatiou, R. Lau and T. Berleth *et al.*, 2002. Genetic complexity of cellulose synthase a gene function in *Arabidopsis* embryogenesis. Plant Physiol., 130: 1883-1893. DOI: 10.1104/pp.102.010603

Burton, R.A., N.J. Shirley, B.J. King, A.J. Harvey and G.B. Fincher, 2004. The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol., 134: 224-236. DOI: 10.1104/pp.103.032904

Caño-Delgado, A., S. Penfield, C. Smith, M. Catley and M. Bevan, 2003. Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. Plant J., 34: 351-362. DOI: 10.1046/j.1365-313X.2003.01729.x

Crawford, R.L., 1981. Lignin Biodegradation and Transformation. 1st Edn., John Wiley and Sons Australia, New York, ISBN-10: 0471057436, pp: 154.

Delmer, D.P., 1999. Cellulose biosynthesis: Exciting times for a difficult field of study. Ann. Rev. Plant Physiol. Plant Molecular Biol., 50: 245-276. DOI: 10.1146/annurev.arplant.50.1.245

Desprez, T., S. Vernhettes, M. Fagard, G. Refrégier and H. Höfte *et al.*, 2002. Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CesA*6*. Plant Physiol., 128: 482-490. DOI: 10.1104/pp.010822

Djerbi, S., M. Lindskog, L. Arvestad, F. Sterky and T.T. Teeri, 2005. The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. Planta, 221: 793-746. DOI: 10.1007/s00425-005-1498-4

Droc, G., M. Ruiz, P. Larmande, A. Pereira and C. Périn, 2005. OryGenesDB: A database for rice reverse genetics. Nucleic Acids Res., 34: 736-740. DOI: 10.1093/nar/gkj012

Ellis, C., I. Karafyllidis, C. Wasternack and J.G. Turner, 2002. The Arabidopsis Mutant cev$_1$ Links Cell Wall Signaling to Jasmonate and Ethylene Responses. Plant Cell, 14: 1557-1566. DOI: 10.1105/tpc.002022

Festucci-Buselli, R.A., W.C. Otoni and C.P. Joshi, 2007. Structure, organization and functions of cellulose synthase complexes in higher plants. Brazil. J. Plant Physiol., 19: 1-13. DOI: 10.1590/S1677-04202007000100001

Fagard, M., T. Desnos, T. Desprez, F. Goubet and F. Goubet *et al.*, 2000. PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of arabidopsis. Plant Cell, 12: 2409-2423. DOI: 10.1105/tpc.12.12.2409

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi and A. Bairoch *et al.*, 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31: 3784-3788. DOI: 10.1093/nar/gkg563

Ho, W.S., S.L. Pang and A. Julaihi, 2014. Identification and analysis of expressed sequence tags present in xylem tissues of kelampayan (*Neolamarckia cadamba* (Roxb.) Bosser). Physiol. Molecular Biol. Plants, 20: 393-397. DOI: 10.1007/s12298-014-0230-x

Ho, W.S., S.L. Pang, P. Lau and J. Ismail, 2011. Sequence variation in the cellulose synthase (SpCesA1) gene from *Shorea parvifolia* ssp. *parvifolia* mother trees. J. Tropical Agric. Sci., 34: 323-329.

Holland, N., D. Holland, T. Helentjaris, K.S. Dhugga and B. Xoconostle-Cazares *et al.*, 2000. A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiol., 123: 1313-1323. DOI: 10.1104/pp.123.4.1313

Jmol, 2012. Jmol: An open-source Java viewer for chemical structures in 3D.

Joker, D., 2000. Neolamarckia Cadamba (Roxb.) Bosser. Danida Forest Seed Centre, Humlebaek, pp: 2.

Joshi, C.P., S.B. Bhandari, P. Ranjan, U.C. Kalluri and A. Smuga *et al.*, 2004. Genomics of cellulose biosynthesis in poplars. New Phytologist, 164: 53-61. DOI: 10.1111/j.1469-8137.2004.01155.x

Kelley, L.A. and M.J.E. Sternberg, 2009. Protein structure prediction on the Web: A case study using the Phyre server. Nature Protocols, 4: 363-371. DOI: 10.1038/nprot.2009.2

Krogh, A., B. Larsson, G. Von Heijne and E.L.L. Sonnhammer, 2001. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. J. Molecular Biol., 305: 567-580. DOI: 10.1006/jmbi.2000.4315

Kumar, M., S. Thammannagowda, V. Bulone, V. Chiang and B.E. Ellis *et al.*, 2009. An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci., 14: 248-254. DOI: 10.1016/j.tplants.2009.02.004

Lai, P.S., W.S. Ho and S.L. Pang, 2013. Development, characterization and cross-species transferability of expressed sequence tag-simple sequence repeat (EST-SSR) markers derived from kelampayan tree transcriptome. Biotechnology, 12: 225-235. DOI: 10.3923/biotech.2013.225.235

Lau, E.T., W.S. Ho and A. Julaihi, 2009. Molecular cloning of cellulose synthase gene, *Sp*CesA*1* from developing xylem of *Shorea parvifolia* spp. *parvifolia*. Biotechnology, 8: 416-424. DOI: 10.3923/biotech.2009.416.424

Liang, X.E. and C.P. Joshi, 2004. Molecular cloning of ten distinct hypervariable regions from the cellulose synthase gene superfamily in aspen trees. Tree Physiol., 24: 543-550. DOI: 10.1093/treephys/24.5.543

Lu, S., L. Li, X. Yi, C.P. Joshi and V.L. Chiang, 2008. Differential expression of three eucalyptus secondary cell wall-related cellulose synthase genes in response to tension stress. J. Exp. Botany, 59: 681-695. DOI: 10.1093/jxb/erm350

Nairn, C.J. and T. Haselkorn, 2005. Three loblolly pine CesA genes expressed in developing xylem are orthologous to secondary cell wall CesA genes of angiosperms. New Phytol., 116: 907-915. DOI: 10.1111/j.1469-8137.2005.01372.x

Panchenko, A.R. and T. Madej, 2004. Analysis of protein homology by assessing the (dis)similarity in protein loop regions. Proteins, 57: 539-547. DOI: 10.1002/prot.20237

Pear, J.R., Y. Kawagoe, W.E. Schreckengost, D.P. Delmer and D.M. Stalker, 1996. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proceedings of the National Academy of Science of USA, (NAS' 96), Science Sessions, 12637-12642. DOI: 10.1073/pnas.93.22.12637

Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nature Methods, 8: 785-786. DOI: 10.1038/nmeth.1701

Phui, S.L., W.S. Ho, S.L. Pang and A. Julaihi, 2014. Development and polymorphism of simple sequence repeats (SSRs) in kelampayan (*Neolamarckia cadamba*-Rubiaceae) using ISSR suppression method. Archives Applied Sci. Res., 6: 209-209.

Pienaar, E., M. Theron, M. Nelson and H.J. Viljoen, 2006. A quantitative model of error accumulation during PCR amplification. Comput. Biol. Chem., 30: 102-111. DOI: 10.1016/j.compbiolchem.2005.11.002

Ranik, M. and A.A. Myburg, 2006. Six new cellulose synthase genes from Eucalyptus are associated with primary and secondary cell wall biosynthesis. Tree Physiol., 26: 545-556. DOI: 10.1093/treephys/26.5.545

Richmond, T., 2000. Higher plant cellulose synthases. Genome Biol., 1: 3001.1-3001.6. DOI: 10.1186/gb-2000-1-4-reviews3001

Richmond, T.A. and C.R. Somerville, 2000. The cellulose synthase superfamily. Plant Physiol., 124: 495-498. DOI: 10.1104/pp.124.2.495

Roberts, A.W., E.M. Roberts and D.P. Delmer, 2002. Cellulose synthase (CesA) genes in the green alga *Mesotaenium caldariorum*. Eukaryotic Cell, 1: 847-855. DOI: 10.1128/EC.1.6.847-855.2002

Ross, J., Y. Li, E.K. Lim and D.J. Bowles, 2001. Higher plant glycosyltransferases. Genome Biol., 2: 3004.1-3004.6. DOI: 10.1186/gb-2001-2-2-reviews3004

Roy, A., A. Kucukural and Y. Zhang, 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738. DOI: 10.1038/nprot.2010.5

Samuga, A. and C.P. Joshi, 2002. A new cellulose synthase gene (*Ptr*CesA*2*) from aspen xylem is orthologous to *Arabidopsis At*CesA*7* (irx3) gene associated with secondary cel wall synthesis. Gene, 296: 37-44. DOI: 10.1016/S0378-1119(02)00864-8

Saxena, I.M., F.C. Lin and J.R.M. Brown, 1990. Cloning and sequencing of the cellulose synthase catalytic sub-unit gene of *Acetobacter xylinum*. Plant Molecular Biol., 13: 673-683. DOI: 10.1007/BF00016118

Saxena, I.M., J.R.M. Brown and T. Dandekar, 2001. Structure-function characterization of cellulose synthase: Relationship to other glycosyltransferases. Phytochem., 57: 1135-1148. DOI: 10.1016/S0031-9422(01)00048-6

Tamura, K., D. Peterson, N. Peterson, G. Stecher and S. Kumar *et al.*, 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Molecular Biol. Evolut., 28: 2731-2739. DOI: 10.1093/molbev/msr121

Tan, C.J., W.S. Ho and S.L. Pang, 2014. Resequencing and nucleotide variation of sucrose synthase (*SuSy*) gene in a tropical timber tree *Neolamarckia macrophylla*. Int. J. Sci. Technol. Res., 3: 135-140.

Tanaka, K., K. Murata, M. Yamazaki, K. Onosato and H. Hirochika *et al.*, 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol., 133: 73-83. DOI: 10.1104/pp.103.022442

Taylor, A.M., B.L. Gartner and J.J. Morrell, 2002. Heartwood formation and natural durability-a review. Wood Fiber Sci., 34: 587-611.

Taylor, N.G., R.M. Howells, A.K. Huttly and S. Turner *et al.*, 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proceedings of the National Academy of Science, (NAS' 03), Science Sessions, pp: 1450-1455. DOI: 10.1073/pnas.0337628100

Taylor, N.G., W.R. Scheible, S. Cutler, C.R. Somerville and S.R. Turner, 1999. The irregular xylem3 locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell, 11: 769-780. DOI: 10.1105/tpc.11.5.769

Tchin, B.L., W.S. Ho, S.L. Pang and J. Ismail, 2012. Association genetics of the cinnamyl alcohol dehydrogenase (*CAD*) and cinnamate 4-hydrolase (*C4H*) genes with basic wood density in *Neolamarckia cadamba*. Biotechnology, 11: 307-317. DOI: 10.3923/biotech.2012.307.317

Thumma, B.R., M.F. Nolan, R. Evans and G.F. Moran, 2005. Polymorphisms in cinnamoyl CoA reductase (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics, 171: 1257-1265. DOI: 10.1534/genetics.105.042028

Tiong, S.Y., S.F. Chew, W.S. Ho and S.L. Pang, 2014a. Genetic diversity of *Neolamarckia cadamba* using dominant DNA markers based on Inter-Simple Sequence Repeats (ISSRs) in Sarawak. Adv. Applied Sci. Res., 5: 458-463.

Tiong, S.Y., W.S. Ho, S.L. Pang and J. Ismail, 2014b. Nucleotide diversity and association genetics of xyloglucan endotransglycosylase/hydrolase (*XTH*) and cellulose synthase (CesA) genes in *Neolamarckia cadamba*. J. Biological Sci., 14: 267-375. DOI: 10.3923/jbs.2014.267.275

Turner, S.R. and C.R. Somerville, 1997. Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell, 9: 689-701. DOI: 10.2307/3870425

Vergara, C.E. and N.C. Carpita, 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 Biol., 47: 145-160. DOI: 10.1023/A:1010631431620

WAC, 2004. A tree species reference and selection guide: *Anthocephalus cadamba*. World Agroforestry Centre.

Wong, J.T., E. Schott, E.M. Sabga, G.G. Kielpinski and R.B. Colvin, 1990. Immunogenic epitopes of the p55 chain of the IL-2 receptor. Relationships to high-affinity IL-2 binding and modulation of the p55 chain. Transplantation, 49: 587-96. PMID: 1690471

Zhang, Y. and J. Skolnick, 2004. Scoring function for automated assessment of protein structure template quality. Proteins: Structure Funct. Bioinformat., 57: 702-710. DOI: 10.1002/prot.20264

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9: 40-40. DOI: 10.1186/1471-2105-9-40