Molecular Cloning of Cellulose Synthase Gene, \textit{SpCesA1} from Developing Xylem of \textit{Shorea parvifolia} spp. \textit{parvifolia}

E.T. Lau, W.S. Ho and A. Julaihi

Forest Genomics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

\textit{Sarawak Forestry Corporation, Forest Research Centre, 93250 Kuching, Sarawak, Malaysia}

\textbf{Abstract:} This study reported the isolation and \textit{in silico} characterization of full-length cellulose synthase (CesA) cDNA from \textit{Shorea parvifolia} spp. \textit{parvifolia}, an important tropical hardwood tree species. Cellulose synthase (CesA) is a member of processive glycosyltransferases that involved in cellulose biosynthesis of plants. The full-length of \textit{SpCesA1} cDNA with size 3308 and 3120 bp open reading frames encoding a 1040 amino acid was isolated using RT-PCR and RACE-PCR approaches. The predicted \textit{SpCesA1} 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 \textit{SpCesA1} is involved in cellulose biosynthesis in secondary vascular tissues of \textit{S. parvifolia} spp. \textit{parvifolia}. Sequence comparison also revealed the high similarity (87\%) among \textit{SpCesA1} and \textit{PruCesA2} of \textit{Populus tremuloides}. This further implies the involvement of \textit{SpCesA1} in catalyzes the cellulose biosynthesis of secondary cell wall rather than primary cell wall. Thus, identification 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.

\textbf{Key words:} Cellulose synthase (CesA), cell wall biosynthesis, \textit{Shorea parvifolia} spp. \textit{parvifolia}, RACE PCR, wood formation

\textbf{INTRODUCTION}

Cellulose synthase (CesA) is a key enzyme that responsible for the biosynthesis of cellulose (Campbell \textit{et al.}, 1997). More properly designated as "cellulose synthase catalytic subunits", the CesA protein is an integral membrane protein, consisting of approximately 1,000 amino acids. It is imaged by TEM as a rosette consisting of six particles which is termed rosette terminal complexes (RTC) (Brown and Montezinos, 1976). Kimura \textit{et al.} (1999) later confirmed that the RTC are the sites of cellulose synthesis after carrying out immunolocalization of putative cellulose synthase catalytic subunits in the rosette subunits. The rosette portion of the terminal complexes (TC) is approximately 25 nm in diameter when viewed in freeze-fractured plasma membranes. Recently, Saxena and Brown (2005) discovered that the rosette portion and its six subunits are localized to the innermost leaflet of the plasma membrane. They also found that the cytoplasmic portion of the TC contains the globular region of the catalytic subunits and is approximately 40-60 nm in diameter.

Genes encoding CesA proteins in plant were first identified in cotton (\textit{Gossypium hirsutum}) fibers (Pear \textit{et al.}, 1996) and later their roles in cellulose synthesis were confirmed in the \textit{Arabidopsis} \textit{rsw1} mutant by Arioli \textit{et al.} (1998). To date, there are six classes of CesA in higher plants with most of the information coming from \textit{Arabidopsis thaliana}. Multiple CesA genes that have been identified in the \textit{Arabidopsis} genome show high similarity to the cotton CesA cDNAs (Holland \textit{et al.}, 2000; Richmond, 2000). In \textit{Arabidopsis}, it has been found that at least four CesA genes, namely \textit{AtCesA1 (rsw1), AtCesA2, AtCesA3} and \textit{AtCesA6} are involved in the formation of primary cell walls and mutation or antisense repression of these genes cause a reduction in cellulose synthesis which is associated with the decrease in cell elongation (Arioli \textit{et al.}, 1998). Three other CesA genes, \textit{AtCesA4, AtCesA7} and \textit{AtCesA8} have been found to be responsible for the formation of secondary cell walls (Joshi, 2003). Mutation in these genes has brought about a dramatic reduction in cellulose content and secondary cell wall thickness.

\textbf{Corresponding Author:} Ho Wei Seng, Forest Genomics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia Tel: 6082-582975 Fax: 6082-583160

416
causing collapsed xylem phenotype. Although, multiple CesA genes are expressed in the same cell types, mutation of one of them can cause significant reduction in cellulose content (Taylor et al., 2000). However, Scheible et al. (2001) and Desprez et al. (2002) discovered that all reported CesA mutants with a defect in cellulose synthesis have been shown to be recessive except two herbicide-resistant CesA mutants showing semi-dominant to herbicides.

Cellulose synthase (CesA) share the common domain structures in prokaryotes and various plant species like cotton, A. thaliana, maize, rice and poplar. These include putative transmembrane helices (TMH), an N-terminal zinc-binding domain (CXXC-CXXC-FXACX-CXX-PXEX-CxxGXX-CxxC) and a cytoplasmic loop consisting of four conserved U-motifs (U1 to U4) with each containing a D residue of QXXRW sequence predicted to be involved in substrate binding and catalysis (D-D-D-QXXRW motif) (Richmond, 2000). In plant cellulose synthase, there are two plant-species regions, i.e. the plant conserved region (CR-P) that shows the high sequence conservation and the hypervariable region (HVR) that is detected in plant cellulose synthase with apparently high sequence divergence (Pear et al., 1996).

To date, there are limited numbers of full-length CesA genes have been reported from tree species, such as PrcesA1 from hybrid poplar, PrrCesA1, PrrCesA2, PrrCesA3, PrrCesA4, PrrCesA5, PrrCesA6 and PrrCesA7 from Populus tremuloides (aspen) (Wang and Loopstra, 1998; Samuny and Joshi, 2004; Liang and Joshi, 2004). However, none of the CesA gene has been reported from tropical hardwood tree species. In this study, we used a Reverse Transcription-Polymerase Chain Reaction (RT-PCR) with gene specific primers designed based on the conserved regions of 10 complete coding sequences of CesA from NCBI database and rapid amplification of cDNA ends (RACE) PCR based strategies to isolate and subsequently in silico characterise a full-length CesA cDNA isolated from Shorea parvifolia spp. parvifolia. S. parvifolia spp. parvifolia or locally known as meranti sarang punai is one of the main sources of light red meranti timber in Southeast Asia. It has been identified as one of the most important light hardwood species for plantation in the hilly areas. Thus, identification 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.

MATERIALS AND METHODS

This study was conducted in 2005-2008 at the Forest Genomics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak.

Plant materials: Developing xylem tissues of S. parvifolia spp. parvifolia were collected from Semengoh Forest Reserve, Kuching, Sarawak in early 2005. Developing xylem tissues were harvested from the inner barks by peeling the tree barks and scraping the exposed surface of tissues with a sterilized scalpel. The collected developing xylem tissues were labelled, kept in separate plastic bags and fixed in liquid nitrogen prior to store at -80°C.

Total RNA isolation and Reverse Transcription-Polymerase Chain Reaction (RT-PCR): Total RNA was isolated from the developing xylem tissues of S. parvifolia spp. parvifolia using the modified cetyltrimethyl ammonium bromide (CTAB) based extraction method as described by Chang et al. (1993), Asif et al. (2000) and Camacho-Villasana et al. (2002). First-strand cDNA synthesis was performed with 1.0 µg of total RNA from S. parvifolia spp. parvifolia developing xylem tissues and Oligo(dT)₅ primer according to the procedure described in the AMV Reverse Transcription System (Promega, USA). The oligonucleotide primers used for the PCR were (forward 5'-AGTTCCCCAAGTGGAATCC-3'; reverse 5'-ACTGGTCCTGAGATCCATC-3') designed based on the conserved regions of 10 complete coding sequences of CesA from NCBI database.

The PCR reaction was carried out in Mastercycler Gradient thermocycler (Eppendorf, Germany) with 2 min of initial denaturation at 94°C, followed by 35 cycles of 30 sec of denaturation at 94°C, 45 sec of annealing at 49.3°C, 1 min of elongation at 72°C and ending with 7 min of final elongation at 72°C. The PCR amplicons (919 bp) were then gel purified using QIAquick™ Gel Extraction kit (QIAGEN, Germany) and cloned into pGEM®-T Easy vector (Promega, USA). Plasmid DNA was isolated using Wizard® Plus SV Miniprep DNA Purification Systems (Promega, USA) and sequenced using Big Dye (dRhodamine) terminator cycle sequencing-ready reaction kit (Applied Biosystems, Foster City, CA).

Rapid amplification of cDNA ends (RACE) PCR: RACE PCR was performed according to GeneRacer™ Kit instruction manual (Invitrogen, USA). The gene specific primers (forward 5'-GCAAAGCTTAAAGAGGCTATG TG-3'; nested forward 5'-GGTCTGCTATGTAAGCTTCC CTCA-3'; reverse 5'-CCTTCTCATGATCCACTGT ACTGAC-3'; nested reverse 5'-CCCTCTCATACTTGG ACTGGTCCTGAGATCCATC-3') designed from the partial SpCesA1 cDNA were used in conjunction with GeneRacer™ PCR primers.
to amplify the 5' and 3' SpCesAl cDNA ends, PCR was performed in Mastercycler Gradient thermocycler (Eppendorf, Germany) with 2 min of initial denaturation (94°C), followed by 35 cycles of denaturation (94°C) for 30 sec, annealing (67°C) for 45 sec, elongation (68°C) for 1 min and ended by 7 min of final elongation (68°C). The total 50 μL PCR reaction volume contained 1 μL of RACE-ready cDNA templates, 1X PCR buffer, 2 mM MgSO₄, 0.2 mM each dNTP, 0.8 μM GeneRacer™ PCR primers, 0.4 μM SpCesAl gene specific primers and 1 unit of High Fidelity Platinum® Taq (Invitrogen, USA). The resultant PCR products were purified and cloned using TOP Cloning Kit for Sequencing (Invitrogen, USA). Plasmid DNA from three positive clones was isolated and sequenced using Big Dye (dRhodamine) terminator cycle sequencing-ready reaction kit (Applied Biosystems, Foster City, CA).

*In silico analysis of full-length SpCesAl:* The full-length SpCesAl cDNA was manually edited using Chromas programme (Technelysium Proprietary Limited). Sequence homology search of SpCesAl against the GenBank non-redundant nucleotide sequences was performed using BLAST. Amino acid sequence of SpCesAl was predicted using the biological sequence alignment editor (BioEdit) (Hall, 1999). The transmembrane helices of SpCesAl were predicted using HMMTOP (Prediction of transmembrane helices and topology of proteins) software described by Tusnády and Simon (2001) and retrieved from (http://www.enzim.hu/hmmtop/). Multiple alignment analysis was carried out to predict the domain structures of SpCesAl, i.e., zinc binding domain, U-motifs, plant conserved region (CR-P), hypervariable regions (HVR) and D, D, D, QxxRW motif using the default parameters of the ClustalW algorithm. Phylogenetic analysis of SpCesAl with the Genebank CesA amino acid sequence was also carried out using Molecular Evolutionary Genetics Analysis (MEGA) software (Kumar et al., 2001). The phylogenetic tree was constructed based on the neighbour-joining (NJ) method and the bootstrap test (100) is used for evaluating the reliability of the inferred neighbour-joining trees.

**RESULTS AND DISCUSSION**

High quality RNA was isolated from developing xylem tissues of *Shorea parvifolia* spp. *parvifolia* using the modified CTAB-based extraction method (Chang et al., 1993; Asif et al., 2000; Camacho-Villasana et al., 2002). RNA examined by using electrophoresis on 1% agarose gel showed a 25S rRNA band equal to or more abundant than the 18 rRNA (Fig. 1), indicating that little or no RNA degradation occurred during the isolation. The A₅₄₀/A₆₀₀ and A₅₃₀/A₆₃₀ absorbance ratios were 1.916 and 2.099, respectively. This further indicates that less contamination of protein substances and weak contamination of secondary metabolites in the isolated RNA samples (Gehring et al., 2000; Malnoy et al., 2001; Huc et al., 2002). The concentrations of RNA samples ranged from 97-127 μg/µL of fresh weight. In general, the RNA obtained was of high quality and integrity.

The partial SpCesAl cDNA fragment of 919 bp was amplified using RT-PCR (Fig. 2) and cloned into pGEM®-T Easy Vector and sequenced. The RACE PCR amplifications of 5' and 3' SpCesAl cDNA ends produced cDNA fragments with size 1,110 bp and 1,671 bp, respectively using gene specific primers. The 5' SpCesAl cDNA end with the size of 1,033 bp and 3' SpCesAl cDNA end with the size of 1,603 bp were further re-amplified using nested RACE PCR analysis. The cDNA

![Fig. 1: Agarose gel electrophoresis of total RNA isolated from *Shorea parvifolia* spp. *parvifolia* developing xylem tissues using a modified CTAB-based method](image1)

![Fig. 2: 1.5% (w/v) agarose gel electrophoresis analysis of reverse transcription-PCR amplifications. Lane M: 100 bp DNA ladder (Promega, USA); lane 1 and 2: partial SpCesAl cDNA fragments](image2)
Fig. 3. The full-length of SpCesAI cDNA and SpCesAI amino acid sequences. The cDNA sequence starts at the first nucleotide and ends with the polyadenylation tail. The predicted SpCesAI amino acid sequence is provided underneath the corresponding cDNA sequence (open reading frames). The ATG initial codon and TGA consensus stop codon are highlighted in grey.
ends generated by nested RACE PCR analysis of SpCesAI were found to be 77 bp (5'-end) and 68 bp (3'-end) smaller than normal RACE PCR amplified 5' and 3' SpCesAI cDNA ends. An internal EcoRI cutting sequence (5'-GAATTCC-3') at the 3' SpCesAI cDNA end was detected.

A full-length SpCesAI cDNA was 3308 bp long with a 3120 bp open reading frame extending from nucleotides 30-3149. The complete transcription units of SpCesAI gene presented in Fig. 3 shows the presence of an ATG codon that initiated the open reading frames at nucleotides 30 to 32 and a TGA consensus stop codon at nucleotides 3150 to 3152. The full-length SpCesAI gene encoded a predicted peptide of 1040 amino acids. Towards the N-terminal of deduced SpCesAI peptide (amino acids 37 to 82), there is a 46 amino acid long, cysteine (Cys) rich region called zinc binding domain (Delmer, 1999; Roberts et al., 2002; Joshi, 2003; Samuga and Joshi, 2004). The N-terminal zinc binding domain was found to be highly conserved in all CesA proteins known to date (Joshi, 2003). Figure 4 shows the schematic diagram of SpCesAI protein.

The consensus sequence of zinc binding domain in SpCesAI peptide was two tandem repeats of Cx2Cx2FxACx2Cx2PxCx2CxEXx2Gx2Cx2C where x is any amino acid (Samuga and Joshi, 2004). The cysteine rich regions in zinc binding domain of SpCesAI have 4 pairs of cysteine residues separated by two amino acids (Cx2C). It was suggested that the zinc binding domain could determine the proteins interactions in cell microfibril structures. As reported by Kurck et al. (2002), two CesA proteins of Gossypium hirsutum, the GhCesAI and GhCesA2 associated in vitro through their zinc binding domains. The lack of zinc binding domain assembled the CesA proteins as linear terminal complexes and obstructed the cell microfibril structures (Delmer, 1999).

SpCesAI contained 7 putative transmembrane helices (TMH). The first TMH region of SpCesAI is towards the N-terminal of amino acid sequence. The other six TMH regions are toward the C-terminal of SpCesAI sequence (Fig. 4). This phenomenon suggests that the SpCesAI associate as the integral membrane protein (Richmond and Somerville, 2000). Holland et al. (2000) reported many of the glycosyltransferases, including the plant and bacterial CesA proteins are predicted to be anchored in the cell plasma membrane by transmembrane helices. The cytoplasmic loop between the first and second TMH regions of SpCesAI consists of 4 U-motifs (U1 to U4). Each U-motif containing conserved aspartate (Asp) residues or QxxRW motif (D, D, D, QxxRW) sequence that predicted to be involved in substrate binding and catalytic activities of CesA enzymes (Vergara and Carpitu, 2001; Saxena et al., 2001; Beeckman et al., 2002; Roberts et al., 2002; Joshi, 2003; Samuga and Joshi, 2004).

The presence of D, D, D, QxxRW motif in SpCesAI suggests that the SpCesAI associate as glycosyltransferases in catalyzing the biosynthesis of long-chain polysaccharides (Samuga and Joshi, 2004). Three aspartate (Asp) residues within the D, D, D, QxxRW motif were conserved from bacteria to plants. Although, it is not resolved which of the aspartate residues served as bases during the glycosyltransferations, two of the three aspartate residues were required to form two glycosidic linkages simultaneously or sequentially during the synthesis of cellulose (Beeckman et al., 2002). Mutagenesis analysis by Saxena et al. (2001) in bacteria demonstrated the exchange of any of the three aspartate residues results in a reduction of glycosyltransferase activity to less than 1%.

The amino acid sequence of SpCesAI detected two plant-specific regions, i.e. the plant-conserved region (CR-P) showed the high sequence conservation and two

![Schematic diagram of SpCesAI proteins](image)

Fig. 4: A diagrammatic representation of SpCesAI proteins. The diagram indicates the position (numerical number represents the location in amino acid sequence) and identity of the predicted amino acid residues in SpCesAI. The D, D, D, QxxRW motif is indicated by the bold letters at the bottom of the diagram.
hypervariable regions (HVR) with apparently high sequence divergence (Pear et al., 1996) (Fig. 4). The CR-P region of SpCesA1 was situated between U1 and U2 motifs (amino acids 377 to 502), as reported in all known plant CesA proteins (Vergara and Carprita, 2001; Beeckman et al., 2002; Roberts et al., 2002; Joshi, 2003). This region was suggested to be implicated in the cellulose biosynthesis at “rosette” complexes. As reported by Roberts et al. (2002), CR-P region of plant CesA was accompanied with the origin of the rosette terminal complexes that consist of multiple catalytic subunits formed by CesA gene. The “rosette” complexes were reported to be situated at the plasma membrane of plant cells and involved in the biosynthesis of cellulose (Tsekos, 1999; Roberts et al., 2002; Zhong et al., 2003; Liang and Joshi, 2004).

The first highly diverged (hypervariable) region, HVRI of SpCesA1 is towards the N-terminal region (amino acids 113 to 231) of the SpCesA1 peptide. A second hypervariable region, HVRII was situated between U2 and U3 motifs (amino acids 626 to 709). The HVR region was reported to be able to define CesA proteins of Oriza sativa, Zea mays and A. thaliana into different sub-classes (Vergara and Carprita, 2001). This analysis revealed the CesA genes clustered into several distinct sub-classes with the identity of sub-class can be defined by the HVR region. Vergara and Carprita (2001) have recently proposed renaming the HVR domain as class-specific regions (CSR), because although these regions are variable among CesA paralogs (paralogous sequences separated by a gene duplication event) from the same plant species, they appear to be highly conserved among CesA orthologs (orthologous sequences separated by a speciation event) from various plants.

Table 1 shows the sequence comparison of SpCesA1 HVRI region with the hypervariable regions of 8 different CesA proteins. The HVR region of ZmCesA9 shows a low similarity (37 to 45%) when compared with other three members from Zea mays, i.e., ZmCesA1, ZmCesA6 and ZmCesA8 (Table 1). However, the HVR region of ZmCesA9 shows a higher similarity with HVR regions of AtCesA3 from A. thaliana (79%) and SpCesA1 from S. parvifolia sp. parvifolia (52%). This phenomenon indicated that the HVR region of ZmCesA9 was conserved among AtCesA3 and SpCesA1 orthologs from A. thaliana and S. parvifolia sp. parvifolia rather than ZmCesA1 (45%), ZmCesA6 (39%) and ZmCesA8 (37%) paralogs from Zea mays, as proposed by Vergara and Carprita (2001). Sequence comparisons revealed a high degree of sequence identity among the SpCesA1, PtcCesA2 and PtcCesA3 HVR regions. The similarity among SpCesA1 HVRI region to PtcCesA2 and PtcCesA3 hypervariable regions were 79 and 63%, respectively.

Comparison of SpCesA1 sequence with other plant species CesA sequences using the MEGA software resulted in 4 distinct groups as shown in Fig. 5. The phylogenetic analysis supported the clustering structure of multiple alignments of HVR regions discussed earlier.
Sequence comparison revealed three members of CesA proteins originate from the timber species, i.e., *SpCesA1*, *PtCesA2* and *PtCesA3* and were grouped together in a distinct cluster (group 1). The *PtCesA2* of *P. tremuloides* showed 87% similarity at the amino acid level with *SpCesA1* (Table 2) meanwhile *PtCesA3* of *P. taeda* showed 77% similarity with *SpCesA1* at the amino acid level. The highest similarity among *SpCesA1* and *PtCesA2* indicates that *SpCesA1* associate in catalyzing the biosynthesis of secondary cell wall in *S. parvifolia* spp. *parvifolia* rather than primary cell wall (Liang and Joshi, 2004). *PtCesA2* shared a high degree of amino acid sequence similarity (over 91%) with protein encoded by *AtCesA7*, which is associated with secondary wall development in *Arabidopsis* (Joshi et al., 2004). The entire shared domain structures of *SpCesA1* protein as shown in Fig. 4 also further suggests the distinct functional role of *SpCesA1* is in cellulose biosynthesis of secondary vascular tissues of *S. parvifolia* spp. *parvifolia* (Samuga and Joshi, 2004).

In conclusion, the results of our in silico analysis suggest *SpCesA1* gene involved in catalyzes the cellulose biosynthesis of secondary cell wall rather primary cell wall. These secondary wall cellulose qualities are the most desirable wood quality traits for forest product industries. Thus, the detailed understanding on the regulation of CesA1 gene could provide a greater impact on the design of future genetic improvement strategies in the production of wood with better quality cellulose that is typically present in the secondary walls of xylem in *S. parvifolia* spp. *parvifolia*. The full-length *SpCesA1* CDNA can also be used for developing genetic markers to identify economic traits loci (ETL) for wood quality traits via gene-assisted selection (GAS) or candidate gene approach.

ACKNOWLEDGMENTS

The authors would like to thank the foresters from the Forest Research Centre (FRC), Sarawak for granting permission to access the forest reserve and also for their excellent field assistance and species identification.

Financial assistance by UNIMAS and Sarawak Forestry Corporation (SFC) are also acknowledged.

REFERENCES


