ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 8 (4): 416-424, 2009
ISSN 1682-296X

Molecular Cloning of Cellulose Synthase Gene, *SpCesA1* from Developing Xylem of *Shorea parvifolia* spp. *parvifolia*

[1]E.T. Lau, [1]W.S. Ho and [2]A. Julaihi
[1]Forest Genomics Laboratory, Department of Molecular Biology,
Faculty of Resource Science and Technology, Universiti Malaysia Sarawak,
94300 Kota Samarahan, Sarawak, Malaysia
[2]Sarawak Forestry Corporation, Forest Research Centre, 93250 Kuching, Sarawak, Malaysia

Abstract: This study reported the isolation and *in silico* characterization of full-length cellulose synthase (*CesA*) cDNA from *Shorea parvifolia* spp. *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 *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 *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 *SpCesA1* is involved in cellulose biosynthesis in secondary vascular tissues of *S. parvifolia* spp. *parvifolia*. Sequence comparison also revealed the high similarity (87%) among *SpCesA1* and PtrCesA2 of *Populus tremuloides*. This further implies the involvement of *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.

Key words: Cellulose synthase (CesA), cell wall biosynthesis, *Shorea parvifolia* spp. *parvifolia*, RACE PCR, wood formation

INTRODUCTION

Cellulose synthase (CesA) is a key enzyme that responsible for the biosynthesis of cellulose (Campbell *et al*., 1997). More properly designated as "ellulose 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 *et al*. (1999) later confirmed that the RTC are the sites of cellulose synthesis after carrying out immunolocation 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 leaftlet 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 (*Gossypium hirsutum*) fibers (Pear *et al*., 1996) and later their roles in cellulose synthesis were confirmed in the *Arabidopsis rsw1* mutant by Arioli *et al*. (1998). To date, there are six classes of CesA in higher plants with most of the information coming from *Arabidopsis thaliana*. Multiple *CesA* genes that have been identified in the *Arabidopsis* genome show high similarity to the cotton *CesA* cDNAs (Holland *et al*., 2000; Richmond, 2000). In *Arabidopsis*, it has been found that at least four *CesA* genes, namely *AtCesA1* (*rsw1*), *AtCesA2*, *AtCesA3* and *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 *et al*., 1998). Three other *CesA* genes, *AtCesA4*, *AtCesA7* and *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,

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

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 ($Cx_2Cx_{12}FxAC x_2Cx_2PxCx_2CxEx_5Gx_3Cx_2C$) and a cytoplasmic loop consisting of four conserved U-motifs (U1 to U4) with each containing a D residue or 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 *PcCesA1* from hybrid poplar, *PtrCesA1*, *PtrCesA2*, *PtrCesA3*, *PtrCesA4*, *PtrCesA5*, *PtrCesA6* and *PtrCesA7* from Populus tremuloides (aspen) (Wang and Loopstra, 1998; Samuga 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 cetytrimethy 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$(dT)_{15}$ primer according to the procedure described in the AMV Reverse Transcription System (Promega, USA). The oligonucleotide primers used for the PCR were (forward 5'-AGTTCCCCAAGTGGAATCCT-3'; reverse 5'-ACTGGTCCTTGAATGCCATC-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 49.3ºC, 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 Minipreps 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'-GCAAAGCTGTAAGAGAGGCTATG TG-3'; nested forward 5'-GGTCTGCTATGTACAGTTCC CTCA-3'; reverse 5'-CCTTCATAGGATCCACTGT ACTGAC-3'; nested reverse 5'-CCC TCTCATACCTGAG GGAAAGC-3') designed from the partial *SpCesA1* cDNA were used in conjunction with GeneRacer™ PCR primers

to amplify the 5' and 3' *SpCesA1* 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 $MgSO_4$, 0.2 mM each dNTP, 0.8 µM GeneRacer™ PCR primers, 0.4 µM *SpCesA1* gene specific primers and 1 unit of High Fidelity Platinum® Taq (Invitrogen, USA). The resultant PCR products were purified and cloned using TOPO TA 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 *SpCesA1*:** The full-length *SpCesA1* cDNA was manually edited using Chromas programme (Technelysium Proprietary Limited). Sequence homology search of *SpCesA1* against the GenBank non-redundant nucleotide sequences was performed using BLAST. Amino acid sequence of *SpCesA1* was predicted using biological sequence alignment editor (BioEdit) (Hall, 1999). The transmembrane helices of *SpCesA1* 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 *SpCesA1*, 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 *SpCesA1* 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 *S. 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 A_{260}/A_{280} and A_{260}/A_{230} 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^{-1} of fresh weight. In general, the RNA obtained was of high quality and integrity.

The partial *SpCesA1* 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' *SpCesA1* cDNA ends produced cDNA fragments with size 1,110 bp and 1,671 bp, respectively using gene specific primers. The 5' *SpCesA1* cDNA end with the size of 1,033 bp and 3' *SpCesA1* cDNA end with the size of 1,603 bp were further re-amplified using nested RACE PCR analysis. The cDNA

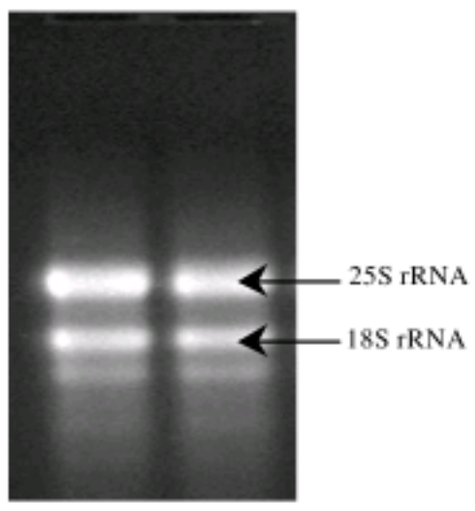

Fig. 1: Agarose gel electrophoresis of total RNA isolated from *Shorea parvifolia* spp. *parvifolia* developing xylem tissues using a modified CTAB-based method

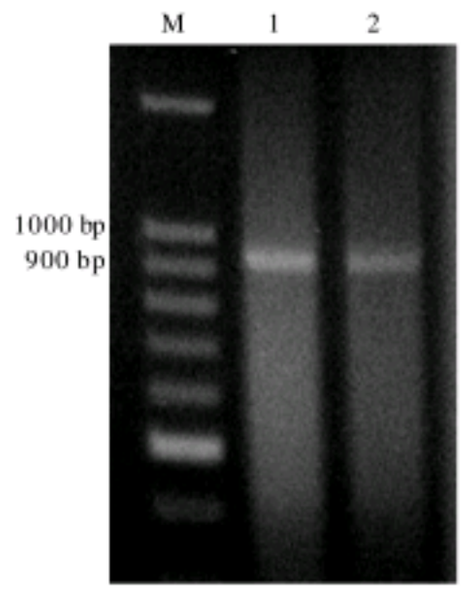

Fig. 2: 1.5% (w/v) agarose gel electrophoresis analysis of reverse transcription-PCR amplifications. Lane M: 100 bp DNA ladder (Promega, USA); lane 1 and 2: partial *SpCesA1* cDNA fragments

```
1   AC ATC ATC CCT CGC GGC CCC CAT TTC CTC ATG GAA GCC AGT GCA GGA CTT GTT GCC GGC TCC    62
1                                          M   E   A   S   A   G   L   V   A   G   S    11

63     CAC AAC CGG AAT GAG CTG GTC GTT ATT CAT GGC CAC GAA GAG CCC AAG CCT TTG AAG AAT    122
12      H   N   R   N   E   L   V   V   I   H   G   H   E   E   P   K   P   L   K   N    31

123    CTG GAT GGT CAG GTC TGT GAG ATT TGC GGG GAC GAC ATT GGC CTG ACG GTG GAC GGA GAT    182
32      L   D   G   Q   V   C   E   I   C   G   D   D   I   G   L   T   V   D   G   D    51

183    CTT TTC GTG GCC TGT AAT GAG TGT GGG TTT CCG GTT TGT AGG CCC TGC TAT GAG TAT GAA    242
52      L   F   V   A   C   N   E   C   G   F   P   V   C   R   P   C   Y   E   Y   E    71

243    AGG AGG GAA GGG ACT CAG CAC TGC CCT CAA TGC AGG ACT AGA TAC AAG CGT CTT AAA GGG    302
72      R   R   E   G   T   Q   H   C   P   Q   C   R   T   R   Y   K   R   L   K   G    91

303    AGT CCT AGG GTT GAG GGA GAC GAA GAC GAA GAG GAT GTG GAT GAT ATT GAA CAC GAA TTT    362
92      S   P   R   V   E   G   D   E   D   E   E   D   V   D   D   I   E   H   E   F   111

363    AAC ATT GAA GAC GAG CAA AAC AAG CAC AAG CAT CTC GTT GAA GCG ATG CTT CAT GGG AAG    422
112     N   I   E   D   E   Q   N   K   H   K   H   L   V   E   A   M   L   H   G   K   131

423    ATG AGC TAT GGA AGA GGG CCT GAA GAC GAT GAG AGT GCT CAG TTC CCA CCT GTT ATA ACT    482
132     M   S   Y   G   R   G   P   E   D   D   E   S   A   Q   F   P   P   V   I   T   151

483    GGT GGC AGA TCC AGA CCA GTA AGT GGT GAG TTT CCG ATT GGG GCG GCT CAT GCT TAT GGA    542
152     G   G   R   S   R   P   V   S   G   E   F   P   I   G   A   A   H   A   Y   G   171

543    GAA CAG ATA TCA TCT TCT TCT CTT CAT AAA CGA GTG CAC CCA TAT CCC ATG GAA GAA CCT    602
172     E   Q   I   S   S   S   S   L   H   K   R   V   H   P   Y   P   M   E   E   P   191

603    GGA AGT GCA AGA GGG GAT GAA AAG AAA GAG GGA GGG TGG AAA GAG AGG ATG GAT GAC TGG    662
192     G   S   A   R   G   D   E   K   K   E   G   G   W   K   E   R   M   D   D   W   211

663    AAA CTG CAG CAA GGA AAC CTT GTT CCT GAA CCA GAA GAC GCC AAT GAT CCT GAC ATG GCA    722
212     K   L   Q   Q   G   N   L   V   P   E   P   E   D   A   N   D   P   D   M   A   231

723    CTG ATT GAT GAA GCT AGG CAG CCA CTC TCA AGG AAA GTT CCA ATT GCA TCT AGC AAG ATC    782
232     L   I   D   E   A   R   Q   P   L   S   R   K   V   P   I   A   S   S   K   I   251

783    AAT CCT TAT CGT ATG TTG ATT GTG GCT CGG CTG GTT ATT CTT GCT TTC TTC CTC CGG TAC    842
252     N   P   Y   R   M   L   I   V   A   R   L   V   I   L   A   F   F   L   R   Y   271

843    AGA ATT TTG AAC CCG GTG CAT GAT GCA ATT GGG CTC TGG CTG ACC TCT ATT GTC TGT GAA    902
272     R   I   L   N   P   V   H   D   A   I   G   L   W   L   T   S   I   V   C   E   291

903    ATC TGG TTT GCA TTT TCA TGG ATC CTT GAT CAG TTC CCC AAG TGG TTC CCT ATT GAT CGT    962
292     I   W   F   A   F   S   W   I   L   D   Q   F   P   K   W   F   P   I   D   R   311

963    GAA ACA TAC CTA GAT CGG CTT TCC CTC AGG TAT GAG AGG GAG GGG GAG CCC AAT ATG CTT   1022
312     E   T   Y   L   D   R   L   S   L   R   Y   E   R   E   G   E   P   N   M   L   331

1023   GCT CCA GTG GAT ATA TTT GTC AGT ACA GTG GAT CCT ATG AAG GAA CCT CCT CTT GTT ACA   1082
332     A   P   V   D   I   F   V   S   T   V   D   P   M   K   E   P   P   L   V   T   351

1083   GCC AAC ACA GTT TTA TCA ATC TTG GCT ATG GAC TAC CCA GTG GAC AAG GTC TCA TGC TAC   1142
352     A   N   T   V   L   S   I   L   A   M   D   Y   P   V   D   K   V   S   C   Y   371

1143   GTC TCT GAT GAT GGT GCT GCC ATG TTA ACC TTT GAG GCT TTA TCT GAA ACT GCA GAA TTT   1202
372     V   S   D   D   G   A   A   M   L   T   F   E   A   L   S   E   T   A   E   F   391

1203   GCT CGG AAA TGG GTA CCG TTC TGC AAG AAG TTC TCT ATA GAG CCT AGA GCC CCT GAG TGG   1262
392     A   R   K   W   V   P   F   C   K   K   F   S   I   E   P   R   A   P   E   W   411

1263   TAC TTC ACC TTA AAG ATC GAT TAT CTT AAG GAC AAA GTC CAA CCC ACA TTT GTT AAG GAA   1322
412     Y   F   T   L   K   I   D   Y   L   K   D   K   V   Q   P   T   F   V   K   E   431

1323   CGT CGT GCC ATG AAG AGA GAA TAT GAA GAA TTC AAG ATT AGG ATA AAT GCG CTT GTA GCA   1382
432     R   R   A   M   K   R   E   Y   E   E   F   K   I   R   I   N   A   L   V   A   451

1383   AAA TCT CAA AAG GTT CCA TCC GGG GGT TGG ATC ATG CAA GAT GGG ACC CCA TGG CCA GGG   1442
452     K   S   Q   K   V   P   S   G   G   W   I   M   Q   D   G   T   P   W   P   G   471

1443   AAC AAT ACT AAG GAT CAC CCT GGT ATG ATT CAA GTT TTT CTT GGC CAC AGT GGA GGC GTT   1502
472     N   N   T   K   D   H   P   G   M   I   Q   V   F   L   G   H   S   G   G   V   491

1503   GAT GCT GAA GGA AAT GAG CTC CCT CGC CTT GTT TAT GTA TCT CGT GAG AAA AGA CCG GGT   1562
492     D   A   E   G   N   E   L   P   R   L   V   Y   V   S   R   E   K   R   P   G   511

1563   TTT CAG CAT CAC AAG AAA GCT GGT GCC GAG AAT GCC TTG ATT CGT GTT TCT GCA GTG CTT   1622
512     F   Q   H   H   K   K   A   G   A   E   N   A   L   I   R   V   S   A   V   L   531

1623   ACA AAT GCT CCG TTT ATG CTG AAC TTG GAC TGT GAC CAT TAT GTG AAT AAT AGC AAA GCT   1682
532     T   N   A   P   F   M   L   N   L   D   C   D   H   Y   V   N   N   S   K   A   551
```

Fig. 3: Continued

```
1683  GTA AGA GAG GCT ATG TGT TTC TTA ATG GAC CCC CAA ATT GGA AAG AAG GTC TGC TAT GTA   1742
552    V   R   E   A   M   C   F   L   M   D   P   Q   I   G   K   K   V   C   Y   V    571

1743  CAG TTC CCT CAA AGA TTT GAT GGA ATA GAT ACA CAC GAT CGT TAT GCC AAC AGA AAC ACT   1802
572    Q   F   P   Q   R   F   D   G   I   D   T   H   D   R   Y   A   N   R   N   T    591

1803  GTC TTC TTT GAT ATT AAT ATG AAA GGT TTA GAT GGT ATT CAG GGT CCT GTT TAT GTC GGT   1862
592    V   F   F   D   I   N   M   K   G   L   D   G   I   Q   G   P   V   Y   V   G    611

1863  ACA GGA TGC GTG TTC AGA AGA CAA GCA TTA TAT GGC TAT GAA CCT CCA AAG GGT CCT AAA   1922
612    T   G   C   V   F   R   R   Q   A   L   Y   G   Y   E   P   P   K   G   P   K    631

1923  CGC CCT AAG ATG GTA AGC TGT GAT TGC TGC CCT TGC TTT GGA CGT CGC AAA AAG GAT AGA   1982
632    R   P   K   M   V   S   C   D   C   C   P   C   F   G   R   R   K   K   D   R    651

1983  AAA CAC TCT AAG CAC GGT GGA GGT GGT GCA ACC AAT GGA GTA GAT GAC GAT AAG GAG TTA   2042
652    K   H   S   K   H   G   G   G   G   A   T   N   G   V   D   D   D   K   E   L    671

2043  TTG ATG TCC CAA ATG AAC TTC GAA AAG AAA TTT GGG CAG TCA GCC ATC TTT GTG ACT TCA   2102
672    L   M   S   Q   M   N   F   E   K   K   F   G   Q   S   A   I   F   V   T   S    691

2103  ACT TTG ATG GAA GAA GGT GGT GTC CCT CCT TCC TCA AGC CCT GCA GCC CTG CTT AAA GAA   2162
692    T   L   M   E   E   G   G   V   P   P   S   S   S   P   A   A   L   L   K   E    711

2163  GCC ATT CAT GTA ATT AGC TGC GGT TAT GAA GAC AAA ACT GAA TGG GGA ACA GAG TTT GGC   2222
712    A   I   H   V   I   S   C   G   Y   E   D   K   T   E   W   G   T   E   F   G    731

2223  TGG ATC TAT GGC TCA ATT ACA GAG GAT ATC TTA ACA GGC TTC AAG ATG CAT TGC CGT GGT   2282
732    W   I   Y   G   S   I   T   E   D   I   L   T   G   F   K   M   H   C   R   G    751

2283  TGG AGA TCT ATC TAC TGC ATG CCA AAA AGA GCT GCC TTT AAG GGT TCA GCT CCT ATC AAC   2342
752    W   R   S   I   Y   C   M   P   K   R   A   A   F   K   G   S   A   P   I   N    771

2343  CTG TCA GAC CGT CTC AAC CAA GTG CTC AGG TGG GCA CTT GGC TCC GTG GAG ATC TTC TTC   2402
772    L   S   D   R   L   N   Q   V   L   R   W   A   L   G   S   V   E   I   F   F    791

2403  AGT CGT CAC TGC CTG CCT GGT ATG GCT TCA AGG GAG GGA CAG CTC AGG TGG CTT GAA AGA   2462
792    S   R   H   C   L   P   G   M   A   S   R   E   G   Q   L   R   W   L   E   R    811

2463  TTT GCT TAT GTC AAC ACA ACT ATC TAC CCC TTC ACC TCC TTA CCT CTC CTT GCC TAT TGC   2522
812    F   A   Y   V   N   T   T   I   Y   P   F   T   S   L   P   L   L   A   Y   C    831

2523  ACT CTT CCA GCA ATT TGC TTG CTC ACA GAT AAA TTC ATC ATG CCG CCA ATA AGC ACC TTT   2582
832    T   L   P   A   I   C   L   L   T   D   K   F   I   M   P   P   I   S   T   F    851

2583  GCA AGT TTA TTG TTC ATT GCT CTC TTC CTT TCA ATC TTT GCA ACG GGA ATT CTT GAG CTA   2642
852    A   S   L   L   F   I   A   L   F   L   S   I   F   A   T   G   I   L   E   L    871

2643  AGA TGG AGT GGA GTG AGC ATT GAA GAA TGG TGG AGG AAT GAG CAG TTC TGG GTC ATC GGT   2702
872    R   W   S   G   V   S   I   E   E   W   W   R   N   E   Q   F   W   V   I   G    891

2703  GGC ATT TCA GCC CAC CTC TTT GCT GTG GTA CAA GGT CTC CTC AAG ATT TTA GCC GGA ATT   2762
892    G   I   S   A   H   L   F   A   V   V   Q   G   L   L   K   I   L   A   G   I    911

2763  GAC ACC AAT TTC ACT GTG ACA TCA AAG GCA ACG GAT GAT GAG GAG TTT GGA GAA TTG TAT   2822
912    D   T   N   F   T   V   T   S   K   A   T   D   D   E   E   F   G   E   L   Y    931

2823  ACC TTC AAA TGG ACA ACT CTC CTC ATT CCT CCA ACC ACT GTC CTG GTT ATC AAC CTT GTT   2882
932    T   F   K   W   T   T   L   L   I   P   P   T   T   V   L   V   I   N   L   V    951

2883  GGA GTT GTT GCC GGC ATC TCT GAT GCC ATA AAC AAC GGA TAC CAA TCA TGG GGA CCT CTT   2942
952    G   V   V   A   G   I   S   D   A   I   N   N   G   Y   Q   S   W   G   P   L    971

2943  TTT GGA AAG CTC TTC TTT TCC TTC TGG GTG ATT CTC CAT CTC TAT CCA TTC CTT AAA GGG   3002
972    F   G   K   L   F   F   S   F   W   V   I   L   H   L   Y   P   F   L   K   G    991

3003  CTG ATG GGG AGG CAG AAC CGG ACA CCC ACC ATT GTT GTC ATA TGG TCA AAC CTA TTG GCT   3062
992    L   M   G   R   Q   N   R   T   P   T   I   V   V   I   W   S   N   L   L   A   1011

3063  TCA ATC TTC TCC TTG CTT TGG GTC CGA ATT GAT CCA TTT GTG TTG AAA ACA AAA GGA CCT   3122
1012   S   I   F   S   L   L   W   V   R   I   D   P   F   V   L   K   T   K   G   P   1031

3123  GAC ACC AAG CAA TGT GGA ATC AAC TGC TGA AAC AAT TGT TTT TAA CCT TTC TCC TCT GAT   3182
1032   D   T   K   Q   C   G   I   N   C                                               1040

3183  TGT GTT CTT CTC CAT GTT ATA CAT GAT ATG TGA TGT ATG AAA AAG AAA ACT GGA GAT ACA   3242

3243  AAA CAA GAA ATA AAT TTA GAC CGA AAA ATT TTG TAC GGT AAA AAA AAA AAA AAA AAA AAA   3302

3303  AAA AAA                                                                         3308
```

Fig. 3: The full-length of *SpCesA1* cDNA and SpCesA1 amino acid sequences. The cDNA sequence starts at the first nucleotide and ends with the polyadenylation tail. The predicted SpCesA1 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 *SpCesA1* were found to be 77 bp (5'-end) and 68 bp (3'-end) smaller than normal RACE PCR amplified 5' and 3' *SpCesA1* cDNA ends. An internal *EcoRI* cutting sequence (5'-GAATTC-3') at the 3' *SpCesA1* cDNA end was detected.

A full-length *SpCesA1* cDNA was 3308 bp long with a 3120 bp open reading frame extending from nucleotides 30-3149. The complete transcription units of *SpCesA1* 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 *SpCesA1* gene encoded a predicted peptide of 1040 amino acids. Towards the N-terminal of deduced *SpCesA1* 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 *SpCesA1* protein.

The consensus sequence of zinc binding domain in *SpCesA1* peptide was two tandem repeats of $Cx_2\ Cx_{12}$ $FxAC x_2\ Cx_2\ PxC x_2\ CxE x_5\ Gx_3\ Cx_2\ C$ where x is any amino acid (Samuga and Joshi, 2004). The cysteine rich regions in zinc binding domain of *SpCesA1* have 4 pairs of cysteine residues separated by two amino acids (Cx_2C). It was suggested that the zinc binding domain could determine the proteins interactions in cell microfibril structures. As reported by Kurek *et al*. (2002), two CesA proteins of *Gossypium hirsutum*, the GhCesA1 and GhCesA2 associated *in vitro* through their zinc binding domains. The lacks of zinc binding domain assembled the CesA proteins as linear terminal complexes and obstructed the cell microfibril structures (Delmer, 1999).

SpCesA1 contained 7 putative transmembrane helices (TMH). The first TMH region of *SpCesA1* is towards the N-terminal of amino acid sequence. The other six TMH regions are toward the C-terminal of *SpCesA1* sequence (Fig. 4). This phenomenon suggests that the *SpCesA1* 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 *SpCesA1* 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 Carpita, 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 *SpCesA1* suggests that the *SpCesA1* 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 *SpCesA1* detected two plant-specific regions, i.e. the plant-conserved region (CR-P) showed the high sequence conservation and two

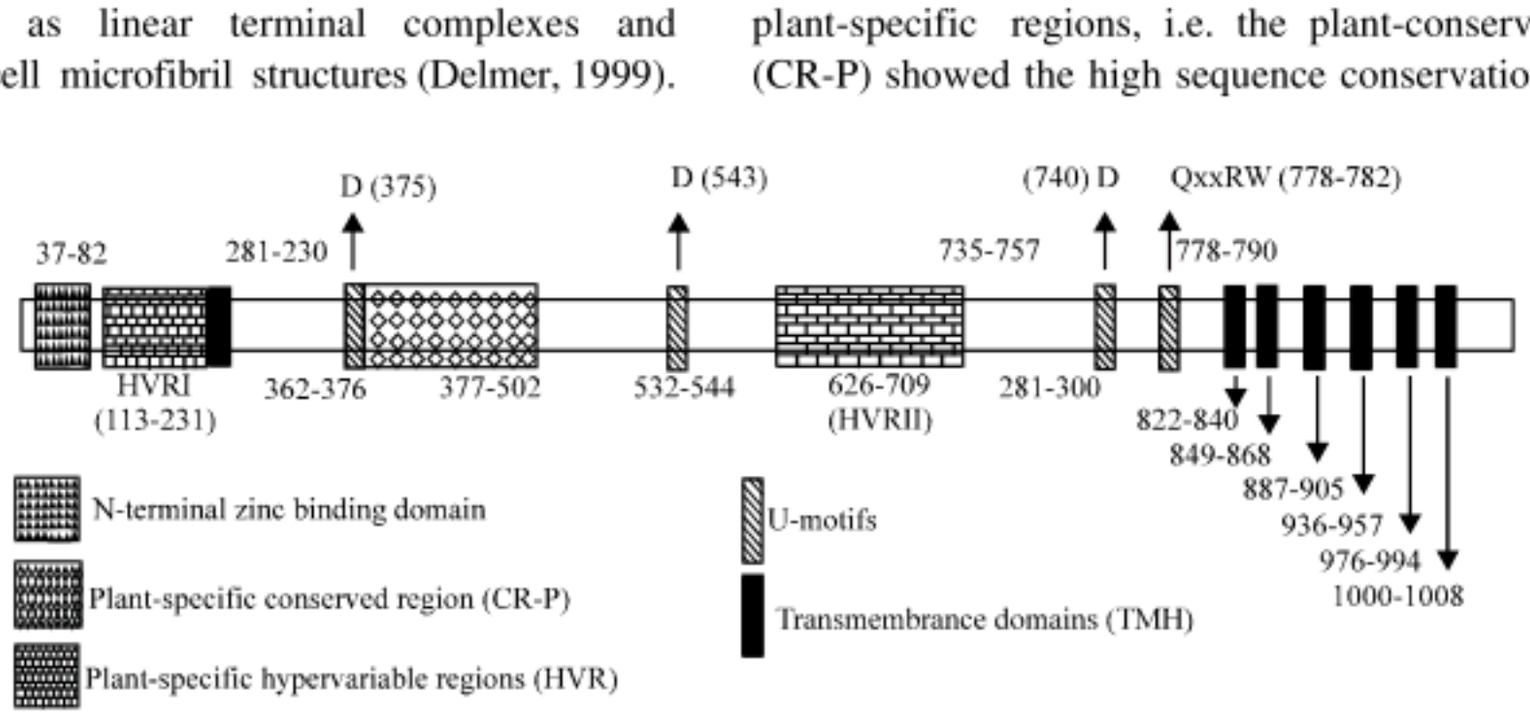

Fig. 4: A diagrammatic representation of *SpCesA1* proteins. The diagram indicates the position (numerical number represents the location in amino acid sequence) and identity of the predicted amino acid residues in *SpCesA1*. The D, D, D, QxxRW motif is indicated by the bold letters at the bottom of the diagram

Table 1: Sequence comparison analysis of *SpCesA1* (GenBank accession No. GQ338420) HVRII region with hypervariable regions from 8 different CesA proteins of PtrCesA2 (AAM26299) from *Populus tremuloides*, PtCesA3 (AAX18649) from *Pinus taeda*, AtCesA1 (AAC39334.1) and AtCesA3 (AAC39336.1) from *Arabidopsis thaliana*, ZmCesA1 (AAF89961), ZmCesA6 (AAF89966), ZmCesA8 (AAF89968) and ZmCesA9 (AAF89969) from *Zea mays* generated using the ClustalW algorithm

Sequences	*SpCesA1*	PtrCesA2	PtCesA3	AtCesA1	AtCesA3	ZmCesA1	ZmCesA6	ZmCesA8	ZmCesA9
SpCesA1	100								
PtrCesA2	79	100							
PtCesA3	63	60	100						
AtCesA1	36	32	43	100					
AtCesA3	48	44	51	50	100				
ZmCesA1	39	40	43	70	49	100			
ZmCesA6	40	41	47	31	40	36	100		
ZmCesA8	41	41	50	32	38	36	78	100	
ZmCesA9	52	44	47	49	79	45	39	37	100

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 Carpita, 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 *Oryza sativa*, *Zea mays* and *A. thaliana* into different sub-classes (Vergara and Carpita, 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 Carpita (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* HVRII 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

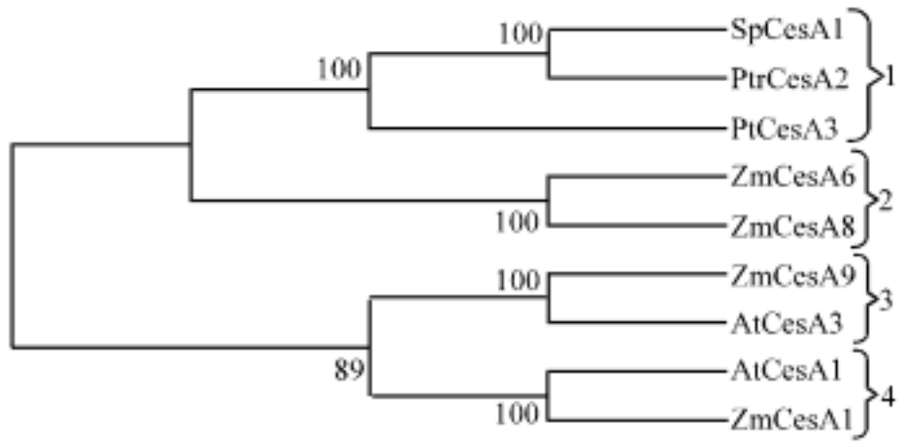

Fig. 5: Neighbour phylogenetic tree generated from the alignment of deduced *SpCesA1* (GQ338420) amino acid sequence with 8 published CesA amino acid sequences of PtrCesA2 (AAM26299) from *Populus tremuloides*, PtCesA3 (AAX18649) from *Pinus taeda*, AtCesA1 (AAC39334.1) and AtCesA3 (AAC39336.1) from *Arabidopsis thaliana*, ZmCesA1 (AAF89961), ZmCesA6 (AAF89966), ZmCesA8 (AAF89968) and ZmCesA9 (AAF89969) from *Zea mays* using MEGA software. Alignments were made using the default parameters of the ClustalW algorithm. Bootstrap values are indicated next to the relevant node

AtCesA3 from *A. thaliana* (79%) and *SpCesA1* from *S. parvifolia* spp. *parvifolia* (52%). This phenomenon indicated that the HVR region of ZmCesA9 was conserved among AtCesA3 and *SpCesA1* orthologs from *A. thaliana* and *S. parvifolia* spp. *parvifolia* rather than ZmCesA1 (45%), ZmCesA6 (39%) and ZmCesA8 (37%) paralogs from *Zea mays*, as proposed by Vergara and Carpita (2001). Sequence comparisons revealed a high degree of sequence identity among the *SpCesA1*, PtrCesA2 and PtCesA3 HVR regions. The similarity among *SpCesA1* HVRII region to PtrCesA2 and PtCesA3 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.

Table 2: Sequences comparison analysis of predicted *SpCesA1* (GenBank accession No. GQ338420) amino acid sequence with 8 published CesA amino acid sequences of PtrCesA2 (AAM26299) from *Populus tremuloides*, PtCesA3 (AAX18649) from *Pinus taeda*, AtCesA1 (AAC39334.1) and AtCesA3 (AAC39336.1) from *Arabidopsis thaliana*, ZmCesA1 (AAF89961), ZmCesA6 (AAF89966), ZmCesA8 (AAF89968) and ZmCesA9 (AAF89969) from *Zea mays* generated using the ClustalW algorithm

Sequences	*SpCesA1*	PtrCesA2	PtCesA3	AtCesA1	AtCesA3	ZmCesA1	ZmCesA6	ZmCesA8	ZmCesA9
SpCesA1	100								
PtrCesA2	87	100							
PtCesA3	77	77	100						
AtCesA1	67	66	67	100					
AtCesA3	68	67	68	67	100				
ZmCesA1	68	66	68	80	68	100			
ZmCesA6	67	67	69	65	66	67	100		
ZmCesA8	69	69	69	65	65	66	88	100	
ZmCesA9	69	67	68	67	79	68	66	65	100

Sequence comparison revealed three members of CesA proteins origin from the timber species, i.e., *SpCesA1*, PtrCesA2 and PtCesA3 were grouped together in a distinct cluster (group 1). The PtrCesA2 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 PtrCesA2 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). PtrCesA2 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 trait 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

Arioli, T., L. Peng, A.S. Betzner, J. Burn and W. Wittke *et al*., 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science, 279: 717-720.

Asif, M.H., P. Dhawan, and P. Nath, 2000. A simple procedure for the isolation of high quality RNA from ripening banana fruit. Plant Mol. Biol. Rep., 18: 109-115.

Beeckman, T., G.K.H. Przemeck, G. Stamatiou, R. Lau and N. Terryn *et al*., 2002. Genetic complexity of cellulose synthase A gene function in *Arabidopsis* embryogenesis. Plant Physiol., 130: 1883-1893.

Brown, Jr. R.M. and D. Montezinos, 1976. Cellulose microfibrils: Visualization of biosynthetic and orienting complexes in association with the plasma membrane. PNAS, 73: 143-147.

Camacho-Villasana, Y.M., N, Ochoa-Alejo, L. Walling and E. Bray, 2002.. An improved method for isolating RNA from dehydrated and nondehydrated chili pepper (*Capsicum annuum* L.) plant tissues. Plant Mol. Biol. Rep., 20: 407-414.

Campbell, J.A., G.J. Davies, V. Bulone and B. Henrissat, 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochemistry, 326: 929-939.

Chang, S., J. Pryear and J. Cairney, 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep., 11: 113-117.

Delmer, D.P., 1999. Cellulose biosynthesis: Exciting times for a difficult field of study. Annu. Rev. Plant Physiol. Plant Mol. Biol., 50: 245-276.

Desprez, T., S. Vernhettes, M. Fagard, G. Refregier and T. Desnos *et al*., 2002. Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CesA6. Plant Physiol., 128: 482-490.

Gehring, H.H., K. Winter, J. Cushman, A. Borland and A. Taybi, 2000. An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. Plant Mol. Biol. Rep., 17: 369-376.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid. Symp. Ser., 41: 95-98.

Holland, N., D. Holland, T. Helentjaris, K.S. Dhugga, B. Xoconostle-Cazares and D.P. Delmer, 2000. A comparative analysis of the plant cellulose synthase (*CesA*) gene family. Plant Physiol., 123: 1313-1324.

Huc, G., C. Honda, B. Kita, Z. Zhang, T. Tsuda and T. Moriguchi, 2002. A Simple protocol for RNA isolation from fruit trees containing high levels of polysaccharides and polyphenol compounds. Plant Mol. Biol. Reporter, 20: 69a-69g.

Joshi, C.P., 2003. Molecular Biology of Cellulose Biosynthesis in Plants. In: Recent Research Developments in Plant Molecular Biology, Pandalai, R. (Ed.). Research Signpost Press, India, pp: 19-38.

Joshi, C.P., B. Suchita, R. Priya, C.K. Udaya, X. Liang, F. Takeshi and S. Anita, 2004. Genomics of cellulose biosynthesis in poplars. New Phytologist, 164: 53-61.

Kimura, S., W. Laosinchai, T. Itoh, X. Cui, C.R. Linder and R.M. Jr. Brown, 1999. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. Plant Cell, 11: 2075-2086.

Kumar, S., K. Tamura, I.B. Jakobsen and N. Masatoshi, 2001. MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics, 17: 1244-1245.

Kurek, I., Y. Kawagoe, D. Jacob-Wilk, M. Doblin and D.P. Delmer, 2002. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. PNAS, 99: 11109-11114.

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.

Malnoy, M., J.P. Reynoird, F. Mourgues, E. Chevreau and P. Simoneau, 2001. A method for isolating total RNA from pear leaves. Plant Mol. Biol. Rep., 19: 69-69.

Pear, J.R., Y. Kawagoe, W.E. Schreckengost, D.P. Delmer and D.M. Stalker, 1996. Higher plants contain horologes of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. PNAS, 93: 12637-12642.

Richmond, T., 2000. Higher plant cellulose synthases. Genome Biol., 1: 3001.1-3001.6.

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

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.

Samuga, A. and C.P. Joshi, 2004. Cloning and characterization of cellulose synthase-like gene, *PtrCSLD2* from developing xylem of aspen trees. Plant Physiol., 120: 631-641.

Saxena, I.M., R.M. Brown and T. Dandekar, 2001. Structure-function characterization of cellulose synthase: Relationship to other glycosyltransferases. Phytochemistry, 57: 1135-1148.

Saxena, I.M. and R.M. Jr. Brown, 2005. Cellulose biosynthesis: Current view and evolving concepts. Ann. Bot., 96: 9-21.

Scheible, W.R, R. Eshed, T. Richmond, D. Delmer and C. Somerville, 2001. Modification of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis ixr* 1 mutants. PNAS, 98: 10079-10084.

Taylor, N.G., S. Laurie and S.R. Turner, 2000. Multiple cellulose synthesis catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell, 12: 2529-2539.

Tsekos, I., 1999. The sites of cellulose synthesis in algae: diversity and evolution of cellulose-synthesizing enzyme complexes. Phycology, 35: 635-655.

Tusnády, G.E. and I. Simon, 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics, 17: 849-850.

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 Mol. Biol. Rep., 47: 145-160.

Wang, H.Y. and C.A. Loopstra, 1998. Cloning and characterization of a cellulose synthase cDNA (Accession No. AF081534) from xylem of hybrid poplar (*Populus tremula*×*Populus alba*). Plant Physiol., 118: 1101-1102.

Zhong, R.Q., W. Herbert Morrison III, G.D. Freshour, M.G. Hahn and Z.H. Ye, 2003. Expression of a mutant form of cellulose synthase *AtCesA7* causes dominant negative effect on cellulose biosynthesis. Plant Physiol., 132: 786-795.