ISSN: 1511-3701
Pertanika J. Trop. Agric. Sci. **34 (2)**: 317 - 323 (2011)

Sequence Variation in the Cellulose Synthase (*SpCesA1*) Gene from *Shorea parvifolia* ssp. *parvifolia* Mother Trees

Ho Wei Seng[1*], Pang Shek Ling[2], Pauline Lau[1] and Ismail Jusoh[3]
[1]Forest Genomics and Informatics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak
[2]Applied Forest Science and Industry Development, Sarawak Forest Tree Seed Bank, Sarawak Forestry Corporation
[3]Department of Plant Science and Environmental Ecology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[*]E-mail: wsho@frst.unimas.my

ABSTRACT

Cellulose synthase (CesA) is a key enzyme involved in the regulation of cellulose biosynthesis pathway. It is heritable and important in determining the variability of wood. In particular, it provides greater impact on the design of future genetic improvement strategies in the production of high quality wood. Thus, the molecular diversity of partial *SpCesA1* genomic DNAs (802 bp) generated through PCR amplification was examined in this study, and this followed by sequencing from five selected *Shorea parvifolia* ssp. *parvifolia* mother trees. The consensus sequences were aligned to detect the presence of single nucleotide polymorphisms (SNPs). In total, seven SNPs were detected at nucleotide 58, 66, 69, 194, 224, 376 and 448. Interestingly, one single base pair InDel polymorphism was also detected at nucleotide 67. On average, one SNP at every 109 bp of the sequence data was detected. However, this result was obtained from a study of partial *SpCesA1* genomic DNA of 802 bp. Two possible restriction enzymes were detected on two SNP sites of partial *SpCesA1* genomic DNA. These included *Ear*I (5'-GAAGAG-3') and *Eco*RI (5'-GAATTC-3'), which were recognized and later cut at nucleotides 48 and 370, respectively. The exclusiveness of the restriction enzymes *Ear*I and *Eco*RI obtained for SNPs at nucleotides 58 and 376, respectively, could be useful for the development of cleaved-amplified polymorphic sequence (CAPS) markers which could also be used to understand the molecular diversity of the *CesA* genes in tropical tree genomes.

Keywords: Cellulose synthase, wood formation, *Shorea parvifolia* ssp. *parvifolia*, PCR, single nucleotide polymorphisms, molecular marker

INTRODUCTION

Wood consists of 40 to 50% cellulose. The basic structural units are the crystallized microfibrils (MFs) formed when multiple hydroxyl groups on the glucose residues from one chain of cellulose form hydrogen bonds with the oxygen molecules on the other chain, holding the chains firmly together side-by-side. The water-insoluble cellulose MFs are associated with the mixtures of soluble non-cellulosic polysaccharides, the hemicelluloses, which account for about 20% of the dry weight of wood (Plomion *et al.*, 2001). Xyloglucan is an example of these hemicelluloses. Xyloglucan binds non-covalently to cellulose MFs, thereby

Received: 5 March 2010
Accepted: 15 December 2010
[*]Corresponding Author

creating a strong cellulose-xyloglucan network that accounts for the rigid structure of plant cell walls.

Cellulose synthase (CesA) is the key enzyme involved in the regulation of the cellulose biosynthesis pathway (Campbell *et al.*, 1997). They are heritable and important in determining the variability of the wood. Hence, this presents an opportunity to select for improved wood properties such as superior product quality (Guimaraes *et al.*, 2007). Traditional chemical and technological assays of such selection are costly and the phenotype assessment is a complex process due to the long generation intervals and poor juvenile-mature trait correlation of wood species (Grattapaglia, 2004). The objective of this study was to determine the feasibility of finding single nucleotide polymorphisms (SNPs) from the cellulose synthase (*CesA*) gene in five selected *Shorea parvifolia* ssp. *parvifolia* mother trees. SNPs are the sequences in the genome of an organism that differs by a single nucleotide between the individuals of the same species (Cargill *et al*., 1999; Thumma *et al*., 2005; Guimaraes *et al*., 2007; Ramos-Onsins *et al*., 2008).

Shorea parvifolia ssp. *parvifolia*, which is locally known as meranti sarang punai, is one of the most valuable and sought after commercial timber tree species belonging to Dipterocarpaceae family. It has been identified as one of the potential fast growing indigenous species that grows well in lowland to upper hill land at the altitude of up to 700m (Newman *et al.*, 2006). The trees are important for producing plywood, veneer, furniture, hardboard, and particleboard.

MATERIALS AND METHODS

Total Genomic DNA Isolation

Fresh young leaves from five randomly selected *S. parvifolia* ssp. *parvifolia* mother trees (2a, 3a, 4a, 6a, and 12a) were collected from the Semengoh Forest Reserve, Kuching, Sarawak. Total genomic DNA was extracted using a modified CTAB method from Doyle and Doyle (1990) and purified using the Wizard® Genomic DNA Purification Kit (Promega, USA). The concentration of DNA was quantified using a spectrophotometer (Lambda 25 UV/VIS Spectrometer, Perkin Elmer, USA).

Primer Design and PCR Amplification

Primers SPPT3-F: 5'-CACACGATCGT TATGCCAAC-3' and SPPT3-R:5'-AGC TCTTTTTGGCAT GCAGT-3' were designed based on the full-length cDNA of *SpCesA1* (GenBank accession number: GQ338420) using Primer3.0 software. This primer pair produced an amplicon of size ~800 bp. The PCR reaction was carried out in a Gradient Palm-Cycler™ (Corbett Research, Australia) with 2 minutes of initial denaturation at 95ºC, followed by 35 cycles of 30 seconds of denaturation at 94ºC, 30 seconds of annealing at 61ºC, 30 seconds of elongation at 72ºC, and ending with 8 minutes of the final elongation at 72ºC. The total 25 µl PCR reaction volume contained 1x PCR buffer (200 mM Tris-HCl, 500 mM KCl, pH8.4), 1.5 mM $MgCl_2$, 0.2 mM of dNTPs, 5 pmol of primers, 1 unit *Taq* DNA polymerase (Promega, USA) and 10 ng of template DNA.

Cloning and DNA Sequencing

The PCR amplicons (~800bp) were then gel purified using QIAquick® Gel Extraction kit (QIAGEN, Germany) and cloned into pGEM®-T Easy vector (Promega, USA). Three clones for each mother tree were selected for the plasmid extraction using Wizard® *Plus* SV Minipreps DNA Purification System (Promega, USA) and sequenced using the Big Dye (dRhodamine) terminator cycle sequencing-ready reaction kit (Applied Biosystems, Foster City, CA).

Data Analysis

The raw data of DNA sequences were checked and edited using Chromas Lite 1.0 (http://www.technelysium.com.au/chromas 14x.html). The 3 *SpCesA1* sequences for each mother tree were aligned using CLC Free Workbench 4.0 (CLC Bio, Denmark) to obtain the consensus sequences

and then subjected to BLASTn similarity search against the non-redundant nucleotide database. The consensus sequences for five mother trees were then aligned using CLC Free Workbench 4.0 (CLC Bio, Denmark) and visually inspected for any base difference among the five mother trees. Meanwhile, *in silico* restriction of the consensus *SpCesA1* sequences was done using NEBcutter V2.0.

RESULTS AND DISCUSSION

For each mother tree, a total of three clones were selected, sequenced, and aligned using CLC Free Workbench 4 (CLC Bio, Denmark). Mother tree 12a did not show any nucleotide variation while the others (2a, 3a, 4a, and 6a) showed a variation among the clones from the same tree. In general, all 5 partial *SpCesA1* genomic DNAs for five selected mother trees (2a, 3a, 4a, 6a, and 12a) showed a high degree of similarity with *CesA Betula* spp. (91%), poplar (88%), and eucalypt (85%), as shown in Table 1.

The consensus sequences of 2a, 3a, 4a, 6a, and 12a were aligned together for manual detection of single nucleotide polymorphisms (SNP) (*Fig. 1*). The alignment result revealed that nucleotide variations occurred between mother trees 3a, 4a and 6a and mother trees 2a and 12a. A total of 7 SNPs were detected at nucleotide 58, 66, 69, 194, 224, 376, and 448. One most interesting observation from the alignment result was the discovery of a single base-pair Insertion-Deletion (InDel) polymorphism at nucleotide 67 for mother trees 2a and 12a. A deletion of adenine at nucleotide 67 of the partial *SpCesA1* genomic DNA of mother trees 2a and 12a was confirmed by ClustalW analysis (http://www.ebi.ac.uk/Tools/clustalw2/ index.html, Larkin et al., 2007) on the full-length SpCesA1 cDNA (Genbank accession no.: GQ338420) with *Eucalyptus grandis CesA3* genomic DNA (Genbank accession no.: EU165713), *E. grandis* full-length CesA3 cDNA (Genbank accession no.: EU165710) and the consensus sequence of all the five partial *SpCesA1* genomic DNAs. This explains the notable size differences between the amplicons generated from a single primer set.

SNP differs from InDel, whereby SNP involves the alteration of a single nucleotide at a specific location in the genome, while InDel includes an insertion or deletion of a number of nucleotides relative to the other. InDel has been related to genome size evolution where the possibility of illegitimate recombination explains genomic downsizing (Grover *et al*., 2008). Zhang *et al*. (2008) reported that many of the small InDels detected in rice hybrids resulted in the formation or disruption of putative *cis*-regulatory elements which were closely associated with the expression of transcription factors and thus, agronomic performance. This is also consistent with the finding by Plantegenet *et al*. (2009) who reported that the preponderance of InDels found in the exons of the *Arabidopsis* accessions Eil-0 and Lc-0 caused drastic effects on gene integrity, specifically on the gene representing expression level polymorphisms.

The partial *SpCesA1* genomic DNA contains coding and non-coding regions. Intron-exon boundaries were predicted from the ClustalW analysis carried out to verify the InDel previously. A schematic diagram was drawn to represent the predicted intron-exon boundaries (*Fig. 2*). The predicted partial *SpCesA1* genomic DNA intron-

TABLE 1
BLASTn output for amplified partial *SpCesA1* genomic DNA of 802 bp

Organism	GenBank accession no.	Similarity (%)	E-value
Betula luminifera	FJ410445.1	91	6e-53
Betula platyphylla	EU591531.1	91	6e-53
Populus trichocarpa	XM 002325086.1	88	3e-51
Eucalyptus grandis	DQ014507.1	85	2e-45

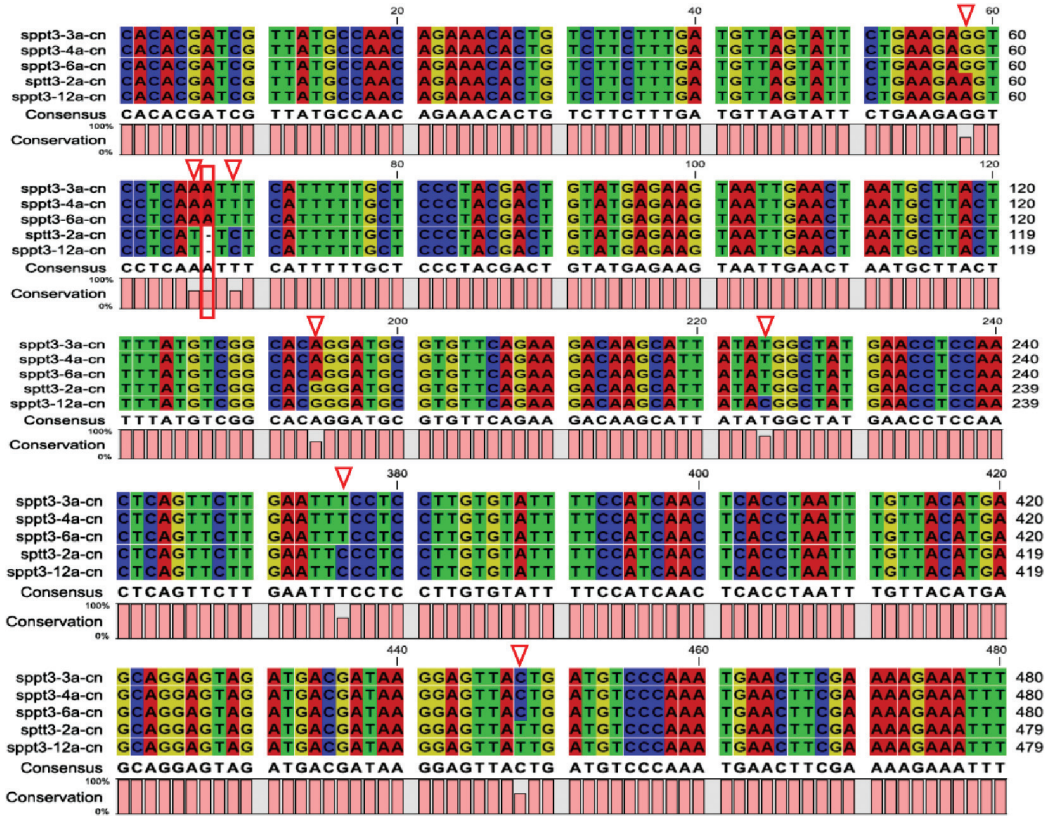

Fig. 1: Alignment of consensus sequences for 2a, 3a, 4a, 6a, and 12a using CLC Free Workbench 4 (CLC Bio, Denmark). Seven SNPs were detected at nucleotides 58, 66, 69, 194, 224, 376, and 448 (red triangle) and one INDEL detected at nucleotide 67 (red box)

exon junctions are conserved as they are also found in *E. grandis CesA3* genomic DNA. There is only one non-coding region detected from the predicted partial *SpCesA1* genomic DNA, with a size of 36 bp from nucleotide 389 to 425.

The data obtained from this study revealed that the chances of detecting SNPs in the partial *SpCesA1* genomic DNA are high in the exons, with about one SNP in every 109 bp compared to zero SNP in the intron region (Table 2). However, this statement was made based on the study of a partial *SpCesA1* genomic DNA of 802 bp. Joshi (2003) reported that even a single base pair mutation in the coding region of the *CesA* gene in *Arabidopsis* would impact the process of cellulose biosynthesis. This is consistent with the findings that the *irx*3 (irregular xylem 3) point mutation in *AtCesA7* shows a defect in the xylem secondary cell wall formation leading to the weakened walls of the treachery elements that later collapse upon themselves (Turner & Somerville, 1997; Taylor *et al.*, 1999). The Indel detected in this study was located in the coding region of the partial *SpCesA1* genomic DNA.

Two possible restriction enzymes were detected for the two SNP sites of the partial *SpCesA1* genomic DNA by using NEBcutter V2.0. They are *Ear*I (5'-GAAGAG-3') and *Eco*RI (5'-GAATTC-3') which recognize and later cut at site numbers 48 and 370, respectively. *Ear*I produces two fragments sized approximately 48 bp and 754 bp, whereas *Eco*RI produces two fragments sized approximately 370bp and 432 bp. The discovery of SNPs is of

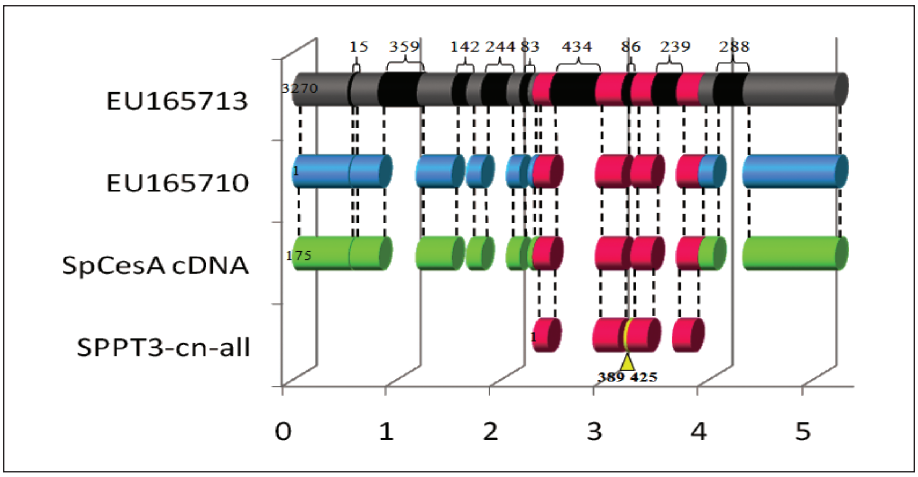

Fig. 2: Comparison of the gene structure of Eucalyptus grandis *CesA3 genomic DNA (Genbank accession no.: EU165713; 8771 nt),* E. grandis *CesA3 mRNA (Genbank accession no.: EU165710; 3331 nt), full-length SpCesA1 cDNA (Genbank accession no.: GQ338420; 3308 nt), and consensus sequence of all the five partial SpCesA1 genomic DNAs. Figures before each sequence denote the start nucleotide in the alignment. Coloured cylinders represent exons and black cylinders represent introns with the number of bases indicated above them. The intron portion as predicted from the partial SpCesA1 genomic DNA is shown in solid yellow. Dotted lines connecting different genes indicate conserved intron-exon junctions.1 unit axis: 1kb*

TABLE 2
Sequence variations within partial *SpCesA1* genomic DNAs among five selected *S. parvifolia* ssp. *parvifolia* mother trees

Region/ Parameter	CesA
Exons	
Total no. of SNPs	7
bp in fragment sequenced	766
SNPs per bp	~ 109
Introns	
Total no. of SNPs	-
bp in fragment sequenced	36
SNPs per bp	-

advantage as many SNPs alter the sites cleaved by restriction enzymes and hence, can be used as cleaved-amplified polymorphic sequence (CAPS) markers (Konieczny & Ausubel, 1993). CAPS markers are amplified by PCR; the amplified DNA will be cleaved with the carefully chosen restriction enzyme, such as that through *in silico* restriction analysis, and the cleaved products are examined on agarose gels. CAPS that posses the property of being co-dominant allows the differentiation of heterogygotes and homozygotes.

CONCLUSIONS

This study has demonstrated that the PCR amplification, followed by sequencing using primers designed from the full-length of the *SpCesA1* cDNA sequence, is an effective technique for obtaining genomic clones and classifying molecular diversity in the cellulose synthase gene (*CesA*) in *S. parvifolia* ssp. *parvifolia*. SNPs in those gene sequences that are significantly associated with the changes of cellulose content, and composition can then be used for early selection of planting material at the seedling stage. In particular, this information is essential to further understand the molecular diversity of the *CesA* genes in tropical tree genomes, as this could have many fundamental and commercial implications.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the foresters from the Forest Research Centre, Sarawak Forestry Corporation (SFC) for establishing the demographic plots at Semengoh Forest Reserve and for species identification. Financial assistance by UNIMAS and SFC is also acknowledged.

REFERENCES

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

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., & Patil, N. (1999). Characterization of single nucleotide polymorphism in coding regions of human genes. *Nature Genetics*, *22*, 231-238.

Doyle, J.J., & Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13-15.

Grattapaglia, D. (2004). Integrating genomics into *Eucalyptus* breeding. *Genetics and Molecular Research*, *3*(3), 369-379.

Grover, C.E., Yu, Y.S., Wing, R.A., Paterson, A.H., & Wendel, J.F. (2008). A phylogenetic analysis of indel dynamics in the cotton genus. *Molecular Biology and Evolution*, *25*(7), 1415-28.

Guimaraes, E.P., Ruane, J., Scherf, B.D., Sonnino, A., & Dargie, J.D. (2007). Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. *Agricultural and Consumer Protection Department* (pp. 283-305).

Joshi, C.P. (2003). Molecular biology of cellulose biosynthesis in plants. In S. Pandalai (Ed.), *Recent research development in plant molecular biology*. India: Research Signpost Press.

Konieczny, A., & Ausubel, F. (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant Journal*, *4*, 403-410.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., & Higgins, D.G. (2007). ClustalW and ClustalX version 2. *Bioinformatics*, *23*, 2947-2948.

Newman, M.F., Burgess, P.F., & Whitmore, T.C. (1996). *Manuals of dipterocarps for foresters: Borneo island light hard woods*. Edinburgh/ Jakarta: Royal Botanic Garden Edinburgh/Centre for International Forestry Research.

Plantegenet, S., Weber, J., Goldstein, D.R., Zeller, G., Nussbaumer, C., Thomas, J., Weigel, D., Harshman, K., & Hardtke, C.S. (2009). Comprehensive analysis of *Arabidopsis* expression level polymorphism with simple inheritance. *Molecular System Biology*, *5*, 542.

Plomion, C., Leprovost, G., & Strokes, A. (2001). Wood formation in trees. *Plant Physiology*, *127*, 1513-1523.

Ramos-Onsins, S.E., Puerma, E., Balana-Alcade, D., Salguero, D., & Aguade, M. (2008). Multilocus analysis of variation using a large empirical data set: Phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Molecular Ecology*, *17*(5), 1211-1223.

Taylor, N.G., Scheible, W-R., Cutler, S., Somerville, C.R., & Turner, S.R. (1999). The irregular xylem3 locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell*, *11*, 769-780.

Thumma, B.R., Nolan, M.F., Evans, R., & Moran, G.F. (2005). Polymorphisms in *Cinnamoyl CoA Reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*, *171*, 1257-1265.

Turner, S.R., & Somerville, C.R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary sell wall. *Plant Cell*, *9*, 689-701.

Zhang, H.Y., He, H., Chen, L.B., Li, L., Liang, M.Z., Wang, X.F., Liu, X.G., He, G.M., Chen, R.S., Ma, L.G., & Deng, X.W. (2008). A genome-wide transcription analysis reveals a close correlation of promoter INDEL. *Molecular Plant*, *1*(5), 720-731.