Detection of DNA Sequence Variation in Cellulose Synthase (*CesA1*) Gene from *Shorea parvifolia* ssp. *parvifolia*

Teo Chiew Phin

This project is submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honours

(Resource Biotechnology)

Department of Molecular Biology
Faculty of Resource Science and Technology
UNIVERSITY MALAYSIA SARAWAK
2010

ACKNOWLEDGEMENT

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Ho Wei Seng for giving me an opportunity to work on this project under his careful guidance and supervision. Besides, I would like to thank my co-supervisor, Dr. Pang Shek Ling for her advices and suggestions in this project. Then, my special thanks go to all the postgraduates in Forest Genomic and Informatics laboratory, UNIMAS especially Mr. Phui Seng Loi, Mr Liew Kit Siong and Ms Tiong Shin Yiing for providing the information required and patiently guiding me through every experimental stage. The valuable information and encouragement given are highly appreciated. Following that, I wish to thank all my laboratory team mates especially Linda Jong, Kho Tieng Tieng, Sia Jennie and Mohamad Khairil Bin Mohd. Asyawandie for sharing their knowledge and experiences with me. Last but not least, I would like to take this opportunity to thank my lovely family members for caring and supporting me throughout my study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS		ii
TABLE OF CONTENTS		iii
LIST OF TABLES		v
LIST OF FIGURES		vi
LIST OF ABBREVIATIONS		vii
ABSTRACT/ABSTRAK		viii
CHAPTER I	**INTRODUCTION**	1
CHAPTER II	**LITERATURE REVIEW**	4
2.1	*Shorea parvifolia* ssp. *parvifolia*	4
	2.1.1 Taxonomic Classificaiton	4
	2.1.2 Family Dipterocarpaceae	4
	2.1.3Genus *Shorea*	5
2.2	Wood Formation	7
2.3	Cellulose Synthase (CesA)	7
2.4	Cleaved Amplified Polymorphic Sequence (CAPS)	11
2.5	Restriction Endonuclease	12
2.6	Previous Study	14
CHAPTER III	**MATERIALS AND METHODS**	16
3.1	Total Genomic DNA Extraction	16
3.2	DNA Purification	17
3.3	Agarose Gel Electrophoresis	17
3.4	Spectrophotometer Quantification of DNA	18
3.5	CAPS	18
	3.5.1 Polymerase Chain Reaction (PCR)	18
	3.5.2 PCR Product Precipitation	19

3.5.3 Quantification of DNA 20

3.5.4 CAPS Analysis 21

3.5.4.1 *Eco*RI Digestion of Purified PCR Product 21

3.5.4.1 *Ear*I Digestion of Purified PCR Product 22

3.5.5 Data Analysis 22

CHAPTER IV **RESULTS AND DISCUSSION** 23

4.1 DNA Extraction and Purification 23

4.2 Estimation of DNA concentration 26

4.3 PCR optimization 28

4.4 PCR Product Purification 30

4.5 Agarose Gel Electrophoresis Quantification of DNA 32

4.6 Restriction Enzyme Digestion 33

4.6.1 *Eco*RI Digestion 33

4.6.2 *Ear*I Digestion 37

CHAPTER V **CONCLUSION AND RECOMMENDATIONS** 40

REFERENCES 42

LIST OF TABLES

Table		Page
2.1	The comparison of *Shorea parvifolia* ssp. *parvifolia* and *Shorea parvifolia* ssp. *velutina*.	6
2.2	Twenty-two full-length *CesA* genes from three plant species with GenBank accession number.	10
3.1	PCR reaction mixture	19
3.2	PCR cycles	19
3.3	Amount of DNA in each λ *Hind* III bands	20
3.4	*Eco*RI Restriction digests condition	21
3.5	*Ear*I Restriction digests condition	22
4.1	Estimated DNA concentration and DNA purity of the purified *S. parvifolia* ssp. *parvifolia* DNA samples.	26
4.2	PCR reaction mixture	29
4.3	PCR cycles	29
4.4	DNA mass of purified PCR product of Sp11 and Sp12	31
4.5	DNA concentration of purified PCR product of Sp1-Sp12	32

LIST OF FIGURES

Figure		Page
2.1	The cellulose synthase (CESA) complex in cellular context.	8
2.2	The predicted amino acid residues in *SpCesA1*.	9
2.3	Comparison of the *CesA* gene structure.	15
2.4	Alignment of consensus sequences for *S. parvifolia* ssp. *parvifolia*.	15
4.1	Gel electrophoresis of genomic DNA samples on a 0.8% (w/v) agarose gel.	25
4.2	Gel electrophoresis of PCR product after $MgCl_2$ optimization on 1.5% (w/v) agarose gel.	28
4.3	Gel electrophoresis of PCR product on a 1.5% (w/v) agarose gel.	30
4.5	Gel electrophoresis of purified PCR product using ethanol precipitation method on a 1.5% (w/v) agarose gel.	31
4.6	Gel electrophoresis of digested DNA using *Eco*RI on 1.5% agarose gel.	36
4.7	Gel electrophoresis of digested DNA using *Ear*I on 1.5% agarose gel.	39

LIST OF ABBREVIATIONS

bp	Base pair
CAPS	Cleaved Amplified Polymorphic Sequence
CesA	Cellulose Synthase
cm	Centimeter
CIA	Chloroform-Isoamyl Alcohol
CTAB	Cetyltrimethylammonium Bromide
^{o}C	Degree Celsius
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediamine Tetraacetic Acid
L	Liter
µl	Microliter
ml	Mililiter
mm	Milimeter
%	Percentage
PCR	Polymerase Chain Reaction
RTC	Rosette Terminal Complex
rpm	Rotation Per Minute
SNP	Single Nucleotide Polymorphism
TAE	Tris-Acetate EDTA
UV	Ultraviolet

Detection of DNA Sequence Variation in Cellulose Synthase (*CesA1*) Gene from *Shorea parvifolia* ssp. *parvifolia*

Teo Chiew Phin

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

Abstract

Shorea parvifolia ssp. *parvifolia* is belongs to the Dipterocarpaceae family which constitutes the major valuable timber in the tropical forest of Asia. The resistance and hard wood properties of this species are mainly contributed by the formation of secondary cell wall made up by a large portion of cellulose. Wood consists of 40 to 50 % of cellulose and the percentage varies among species and within species. This variation can be detected and it enables the selection of desired wood properties. In this study, Cleaved Amplified Polymorphic Sequence (CAPS) technique was used to detect and verify the 2 SNP sites detected in cellulose synthase (*CesA1*) gene which is responsible in encoding the cellulose material in12 *S. parvifolia* ssp. *parvifolia* samples. Primer SPPT3-F and SPPT3-R were used to amplify the 802 bp *CesA1* gene prior to restriction enzyme (*Eco*RI and *Ear*I) digestion. As *Eco*RI cutting site was presented in all of the 12 samples, it shows that no mutation was occurred at the SNP site detected at nucleotide number 376 in the 802 bp *CesA1* gene. On the other hand, the partial *CesA1* gene from all of the 12 samples was not restricted by the *Ear*I. Due to the absence of positive control and the possibility of non optimal reaction condition and the inactivity of restriction ezyme, the mutation occurred at the SNP site detected at nucleotide number 58 cannot be confirmed in this study. Although both of the SNPs detected in previous study cannot be verified in this study, the CAPS technique showed cost effectiveness and the result can be obtained in shorter period.

Key words: *Shorea parvifolia* ssp. *parvifolia*, Cleaved Amplified Polymorphic Sequence (CAPS), Restriction enzyme, cellulose synthase gene (*CesA*), Single nucleotide polymorphism (SNP).

Abstrak

Shorea parvifolia ssp. parvifolia dari famili Dipterocarpaceae merupakan spesies pokok kayu yang mempunyai nilai komersial yang tinggi dalam hutan tropika di Asia. Kayu balak spesies ini mempunyai ketahanan yang tinggi kerana ia mempunyai lapisan dinding sel kedua yang diperbuat daripada selulosa dalam kuantiti yang banyak. Biasanya, kayu balak mengandungi 40-50 % selulosa dan peratus ini berlainan di antara dan dalam spesies. Kandungan selulosa yang berlainan ini membolehkan pemilihan pokok kayu yang mempunyai ciri-ciri yang diingini. Dalam kajian ini, teknik Cleaved Amplified Polymorphic Sequence (CAPS) telah digunakan untuk mengesan dua polimorfisme nukleotida tunggal (SNP) di dalam gen selulosa sintase (CesA1) daripada 12 pokok. Pasangan pencetus SPPT3-F dan SPPT3-R telah digunakan untuk mengamplifikasikan 802 bp gen selulosa sintase sebelum dicerna dengan enzim EcoRI dan EarI. Daripada hasil kajian yang diperolehi, enzim EcoRI berjaya mencerna gen CesA1 dalam semua 12 pokok tersebut dan ini menunjukkan mutasi tidak berlaku di bahagian polimorfisme nukleotida tunggal (SNP) yang dikesan di nombor nukleotida ke-376. Selain itu, pencernaan enzim EarI tidak berlaku ke atas gen CesA1 dalam semua 12 pokok. Kekurangan kawalan positif dan kebarangkalian berlakunya ketidakaktifan enzim EarI serta keadaan pencernaan yang tidak optimum menyebabkan mutasi yang berlaku pada SNP di nombor nukleotida ke-58 tidak boleh dibuktikan. Walaupun kedua-dua SNPs yang telah berjaya dikesan dalam projek dahulu tidak dapat dibuktikan dalam projek ini, namun penggunaan teknik CAPS menunjukkan pengurangan masa dan wang berbanding dengan teknik-teknik yang lain.

Kata kunci: Shorea parvifolia ssp. parvifolia, Cleaved Amplified Polymorphic Sequence (CAPS), enzim restriksi, gen selulosa sintase (CesA), polimorfisme nukleotida tunggal (SNP).

CHAPTER I

INTRODUCTION

Shorea parvifolia ssp. *parvifolia* is a light red meranti belongs to the Dipterocarpaceae family which constitutes the major species for valuable timber in the tropical forest of Asia. It is locally known as meranti sarang punai and is greatly used in producing furniture, plywood, and veneer. Besides, non-timber products such as butter fat, tannin and dammar can be produced from this species (Pooma, 2003). According to Ashton (1998), *S. parvifolia* ssp. *parvifolia* is native to Indonesia, Malaysia, Thailand and Singapore. This species has been overexploited in the South East Asia. In order to reduce the pressure on the natural tropical forest, efforts have been done to select, grow and genetically manipulate the plus trees of some economically important hardwood, mainly the species of Dipterocarpaceae (Bosman, 1997).

According to Pooma (2003), some *Shorea* timbers are more resistant to pest due to their hard wood properties. These properties are mainly contributed by the formation of secondary cell wall which is made up by the different arrangements of cellulose microfibrils, mixed with the hemicelluloses and lignin in the wood. Based on Plomion *et al.* (2001), wood consists of 40 to 50% of cellulose. Cellulose is important as it is used as major biopolymer in textile, forest product and chemical industries. The wood properties and the content of cellulose vary among species and within species. This variability is heritable and thus, provides an opportunity to determine the variability in woods following by the selection of desired wood properties.

In plant cell, cellulose is synthesized by a complex consists of catalytic subunit of cellulose synthase (CesA) (Wang *et al*., 2001). These cellulose synthase are encoded by *CesA* genes and are located at the plasma membrane of cells (Zhong *et al*., 2003). In 1997, the study

done by Bosman (1997) on five superior red meranti (*S. leprosula, S. parvifolia* and *S. pauciflora)* had revealed the large variation in fibre wall percentage between the trees of the same species and he proposed that this variation will offer possibilities for the selection of superior clones in the future. Following that, molecular studies in arabidopsis and maize had been conducted and variable *CesA* genes which caused the production of different amounts of cellulose deposited on the cell wall had been detected (Zhong *et al*., 2003).

Recently, molecular variation including 7 single nucleotide polymorphisms (SNP) and 1 deletion in partial 802 bp *CesA* genes from *S. parvifolia* ssp. *parvifolia* had also been detected using DNA sequencing method (Pauline, 2009). SNP is generally referred as the single base mutation. As the single base mutation in cellulose biosynthesis gene may caused different quantity and quality of cellulose in an individuals from the same species, the association between genotype and the phenotype can then be studied by comparing the SNPs with wood properties. By using the DNA sequencing method, direct identification of variation at the molecular level can be performed and this provides opportunity to develop a new DNA marker which can be used to select tree species with desired traits over a short period of time.

According to Thiel *et al*. (2004), SNPs are the most abundant form of DNA variation in a genome. Besides, they are stable and found interspersed in the nuclear genome (Fusari *et al*., 2008). Therefore, it can be used for marker assays and to address some biological and genetic problems. Previously, different selection methods have been applied in order to select the desired or improved wood properties. The traditional chemical and technological methods are not suitable due to the high initial cost and the long duration needed to assess the phenotype of the wood species (Plomion *et al*., 2001). Besides, the DNA sequencing method used in previous study showed moderate throughput and moderate start up cost (Pauline,

2009). Therefore, Cleaved Amplified Polymorphic Sequence (CAPS) will be used in this study.

CAPS is also named as PCR-RFLP. In this method, the restriction site will be amplified before digested with properly chosen restriction enzyme. By performing gel electrophoresis, the presence of certain SNP in a gene can be determined and verified based on the bands observed on gel. According to Pauline (2009), restriction endonuclease whose recognition sequence has been introduced by the SNPs in *CesA* gene from *S. parvifolia* ssp. *parvifolia* have been detected. In combination with PCR assays, the SNPs can then be analyzed as CAPS marker. In this study, CAPS is chosen as it is considered as high reproducibility, cost effective and a rapid method to analyze the result. Moreover, sophisticated equipments are unneeded in the analysis (Thiel *et al*., 2004).

The objective of this study is to detect the DNA sequence variation in the partial 802 bp *CesA* gene from *S. parvifolia* ssp. *parvifolia* using CAPS method. This is done by comparing the fragments of DNA cut by the restriction enzymes. Two restriction enzymes (*Eco*RI and *Ear*I) detected using NEBcutter V2.0 by Pauline (2009) for the two single nucleotide polymorphism (SNPs) sites in the partial 802 bp *CesA gene* in *S. parvifolia* ssp. *parvifolia* were used in the CAPS analysis in order to verify the presence of the SNPs in the gene. Following the verification of the SNPs, CAPS marker can be developed and can be used to detect other *S. parvifolia* ssp. *parvifolia* with desired properties in the future.

CHAPTER II

LITERATURE REVIEW

2.1 *Shorea parvifolia* ssp. *parvifolia*

2.1.1 Taxonomic Classification (Ashton, 1998)

Kingdom : Plantae

Phylum : Tracheophyta

Class : Magnoliopsida

Order : Theales

Family : Dipterocarpaceae

Genus : *Shorea*

Species : *Shorea parvifolia*

Sub-species : *Shorea parvifolia* ssp. *parvifolia*

2.1.2 Family Dipterocarpaceae

Family Dipterocarpaceae consists of 16 genera and 470 to 580 species (Ng, 1991; Maury-Lechon and Curtet, 1998). This family plays an important role in contributing high value products such as camphor, resins, and timbers. According to Ng (1991), there are about 9 genera and 155 of the species available in Malay Peninsula. The Dipterocarps dominate

throughout the Southeast Asia and are widely distributed in Malesian, mainland Southeast Asia, South Asia, Sri Lanka, and Seychelles. They can be found in variety habitats ranging from coastal to inland, riverine to swampy, well-drained to poorly drained, and rich to poor soils (Maury-Lechon and Curtet, 1998). The special characteristic of this family is their winged fruits which are developed from persistent sepals, fleshy bilobed unequal cotyledons, simple stipulate leaves and the dimorphic shoot system (Ng, 1991 cited in Pooma, 2003).

Based on Maury-Lechon and Curtet(1998), Dipterocarpaceae can be divided into three subfamily which are Pakaraimoideae, Monotoideae, and Dipterocarpoideae. The Dipterocarpoideae is the largest subfamily and it can be further divided into two more groups: *Valvate-Dipterocarpi* and *Imbricate-Shoreae.* From these two groups, genus *Shorea* falls into *Imbricate-shoreae* group which is characterized by their imbricated sepals in fruit, resin canals in tangential band, grouped vessels and most importantly, by their basic chromosome numbers (n=7) (Yuwa-Amornpitak *et al*., 2006).

2.1.3 Genus *Shorea*

Genus *Shorea* is the largest genera classified under *Imbricate-shoreae* group. Based on Yuwa-Amornpitak *et al*. (2006), genus *Shorea* consists of 194 species and 163 of the species can be found in Malesia. According to Pooma (2003), genus *Shorea* is widely distributed in Sri Lanka, India, Myanmar, and Malesia. Species in this genus are differentiated into four major groups including White Meranti, Red Meranti, Yellow Meranti, and Balau. White Meranti refers to light yellowish with moderate hard timber, Red Meranti refers to the soft to rather hard wood with red tinged inner bark, Yellow Meranti refers to the rather soft timber with thin inner bark

while Balau refers to the very hard timber. Among the four timber groups, *Shorea parvifolia* falls into the Red Meranti group. *S. parvifolia* are further divided into two subspecies and are differentiated in Table 2.1.

Table 2.1: The comparison of two subspecies of *Shorea parvifolia* (Ashton, 1998 & Pooma, 2003).

Subspecies	***Shorea parvifolia* ssp. *parvifolia***	***Shorea parvifolia* ssp. *velutina***
Native	Indonesia, Malaysia, Singapore, Thailand.	Indonesia, Malaysia.
Distribution	Widespread in Peninsular Malaysia, Singapore, Borneo and Sumatra from sea level to 700m altitude.	Confined to Pahang, Perak and Negeri Sembilan in Peninsular Malaysia, distributed in Sumatra, and Borneo near the coast.
Diagnostic characters: **I. Leaves**	The leaves are broadly ovate in shape with 4-9 x 3-5 cm. They appeared reddish brown in colour when dry. The midrib is depressed above with 10-13 pairs of secondary nerves and few intermediate nerves.	The leaves are ovate-oblong in shape with 3-9 x 1.5-3.5cm. There are pale creamy fine scales beneath the leaves with scattered hairs. They appeared pale brown in colour when dry. Besides, they have long acuminate apex, rounded base, slightly rolled margin and stout secondary nerves without intermediate nerves.
II. Domatia	Pubescent with one or two pairs at the base.	This species do not possess domatia.
III. Fruiting calyx	The species has 3 longer calyx lobes with 5.5-7 x 1-1.5 cm and 2 short lobes with 2-2.5 x 0.3-0.5 cm. It is Sac like and enclose half of the nut.	The species has 3 longer calyx lobes with 5-5.5 x 0.7-1 cm and 2 short lobes with 2.5-3 x 0.2 cm.
IV. Nut	Ovoid, pubescent with 1.5 x 1 cm.	Nearly spherical with 7 x 7 mm.
Flowering period	January to November	April to October
Fruiting Period	January to December	February to December

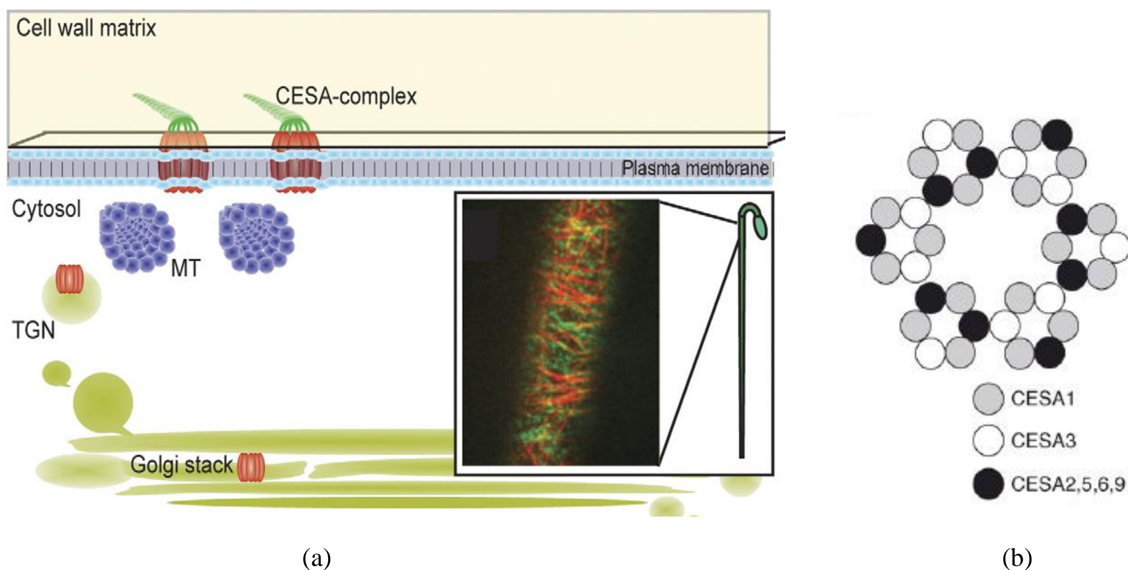
2.2 Wood Formation

Wood formation involves both primary and secondary growth. Cell division, cell expansion, cell wall thickening, programmed cell death, and heartwood formation are the main stages involved in wood formation (Plomion *et al*., 2001). In primary growth, the cells at apical meristem divide and increase their size. Once these processes are accomplished, the secondary growth is followed by the formation of secondary phloem and xylem at the vascular cambium. Following that, secondary xylem cell wall is formed by the deposition of cellulose, hemicelluloses, lignin and other secondary metabolites to the cell wall. Lastly, the xylem vessel elements undergo programmed cell death and the heartwood is formed by the deposition of various chemical compounds such as oils, gums, resins and other extractives. Based on Spicer (2005), the cells undergo active cell metabolism in synthesizing secondary metabolites and blockage such as tyloses and gels within the conducting elements before death. The ability of water conduction will lose while the mechanical strength of the tree will increase by the formation of wood.

2.3 Cellulose Synthase (CesA)

Cellulose is the most abundant macromolecule on earth and there is 40-50% of cellulose in wood. It is the aggregate of linear polymers of β-1, 4 linked glucose residues (Richmond, 2000) and can be found as microfibrils in plant. According to Taylor *et al*. (1999), cellulose comprises 20 to 30 % of the dry weight of primary cell wall while increase up to 90 % in the secondary cell wall. This important component is synthesized by the cellulose synthase (CesA) which is also known as catalytic subunit of cellulose synthase (Richmond, 2000; Holland *et al*.,

2000; Zhong *et al*., 2003). Based on Richmond (2000), cellulose synthase is a member of protein complex that is located at the plasma membrane. It exists in a structure of 'rosettes' termed as Terminal Rosette Complex (TRC) (Brown and Montezinos, 1976). Figure 2.1a shows the cellulose synthase (CESA) complex in cellular context. In Mutwil *et al*. (2008), it is assumed that a transmembrane complex consist of 36 individual cellulose synthase (CESA) proteins. These 36 CESA proteins form 6 subunits within the complex and are shown in Figure 2.1b. As microfibrils are known to consist of 36 glucan chains, it is suggested that a subunit within the complex is responsible in synthesis 6 glucan chains. Before the cellulose is synthesized in the cell wall, it is assumed that the CESA complexes are assembled in the Golgi and exported to the plasma membrane through exocytosis process. Following that, the cortical microtubules are thought to facilitate the cellulose deposition in the primary cell wall.



(a) (b)

Figure 2.1: The cellulose synthase (CESA) complex in cellular context. Figure 2.1a: The assembly of CESA complex in plasma membrane. CESA complex is represented in red colour, golgi is represented in green colour and microtubules are represented in blue colour. Figure 2.1b: The schematic view of putative CESA rosette complex found in primary cell wall. The position of CESA1 and 3 are fixed while the CESA2, 5, 6, 9 are interchangeable for their positions (Mutwil *et al*., 2008).

Recently, a study has been conducted in order to identify new *CesA* gene from tropical tree genomes. Following that, a full length *SpCesA1* cDNA with size 3308 bp and 3120 bp open reading frame encoding 1040 amino acids has been identified using RT-PCR and RACE-PCR methods (Lau *et al.*, 2009). The *SpCesA1* protein was then estimated and it is shown in Figure 2.2. This protein plays an important role in *S. parvifolia* ssp. *parvifolia* as it helps in the biosynthesis of cellulose in the secondary vascular tissues. The predicted *SpCesA1* protein is known to contain a N-terminal cysteine rich zinc binding domain, 7 putative transmembrane helices (TMH), 4 U-motifs containing a signature D, D, D, QxxRW motif , an alternating conserved region and 2 hypervariable regions.

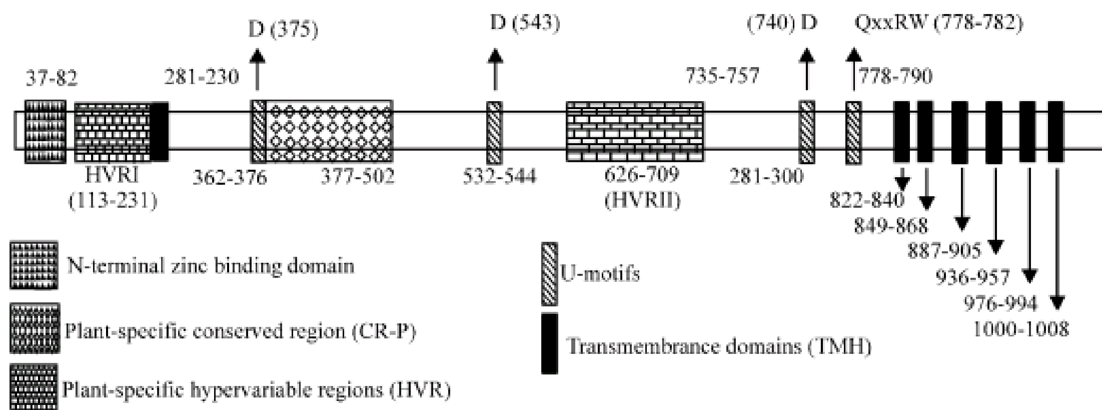


Figure 2.2: SpCesA1 protein. The diagram indicates the position and identity of the predicted amino acid residues in SpCesA1. The numerical number represents the location in amino acid sequence (Lau *et al.*, 2009).

The gene encoding the CesA protein has been extensively studied in cotton, maize, and also arabidopsis. Nowadays, there are more than 20 full-length *CesA* genes available from different plant species. Twenty-two fully sequenced *CesA* genes from three plants are shown in Table 2.2. The proteins encoded by these full-length *CesA* genes showed high similarity to each other as there are a number of highly conserved residues (Richmond, 2000). However, there are few small regions of variability among the protein encoded. As the genome in

Arabidopsis thaliana is almost sequenced, the *CesA* genes in this species are well-studied. There are 12 distinct *CesA* genes identified and they are expressed at different regions and stages (Holland *et al*., 2000). In expressing these genes, the mutation of a dominant or a semidominant *CesA* gene will cause the reduction of cellulose content in plant cells.

Table 2.2: Twenty-two full-length *CesA* genes from three plants with GenBank accession number (Holland *et al*., 2000).

Gene Name	**GenBank Accession Number**
Arabidopsis thaliana	
AtCesA-1	AF027172
AtCesA-2	AF027173
AtCesA-3	AF027174
AtCesA-4	AB006703
AtCesA-5	AB016893
AtCesA-6	AF062485
AtCesA-7	AF088917
AtCesA-8	AL035526
AtCesA-9	AC007019
AtCesA-13	AC006300
Zea mays	
ZmCesA-1	AF200525
ZmCesA-2	AF200526
ZmCesA-3	AF200527
ZmCesA-4	AF200528
ZmCesA-5	AF200529
ZmCesA-6	AF200530
ZmCesA-7	AF200531
ZmCesA-8	AF200532
ZmCesA-9	AF200533
Gossypium hirsutum	
GhCesA-1	U58283
GhCesA-2	U58284
Shorea parvifolia* ssp. *Parvifolia	
SpCesA-1	GQ338420

2.4 Cleaved Amplified Polymorphic Sequence (CAPS)

In 1980th, genome analysis was highly facilitated with the development of Polymerase Chain Reaction (PCR). Following that, newly emerging techniques based on PCR have been developed and one of the examples is the Cleaved Amplified Polymorphic Sequence (CAPS). Generally, CAPS is also known as PCR-Restriction Fragment Length Polymorphism (PCR-RFLP). According to Konieczny and Ausubel (1993), CAPS polymorphism is the differences in fragments due to single nucleotide polymorphism (SNP) or insert/deletion (INDELs).

CAPS method is a technique used to detect the genetic difference between individuals. Based on Edwards (1998), CAPS is the alternative to both Single Strand Conformation Polymorphism (SSCP) and direct sequencing in order to detect genetic variation at molecular level. Different from RFLP, PCR included in CAPS analysis can shorten the process needed to analyze the result as Southern blotting is excluded in CAPS analysis. In this method, the restriction site will be amplified before digested with properly chosen restriction enzyme. By performing gel electrophoresis, the digested PCR product will give particular bands in which the heterozygotes can be distinguished from the homozygotes.

As CAPS required only agarose gel and ethidium bromide staining, it can be carried out under limited laboratories equipment and researcher does not need to deal with hazardous radioactive (Edwards, 1998). By using CAPS, disease-associated gene can be analyzed by studying the SNP in the genome (Xiao *et al*., 2005). Previously, the *Seiridium cardinale*, *S. cupressi* and *S. unicorne* that cause cankers on *Cupressus* species cannot be distinguished using morphological characteristic. However, the application of CAPS method using *Hae*III

to digest the *β-tubulin* amplicons has successfully identified the *Seiridium* species that caused cypress canker (Krokene *et al*., 2004).

Besides, the CAPS method had also been applied in *A. thaliana* in order to determine the chromosomal location of three out of the four genes identified in the *RBCS* gene family (*RBCS-1B, RBCS-2B, and RBCS-3B*) (Niwa *et al*., 1997). These genes are responsible in encoding the Ribulose-1, 5-biphosphate carboxylase/oxygenase (Rubisco). By using *SspI*, the *RBCS-B* genes were successfully detected on the chromosome 5. Other than that, genetic diversity in *A. thaliana* and *Cryptomeria japonica* had also been successfully studied using the CAPS marker (Barth *et al*., 2002).

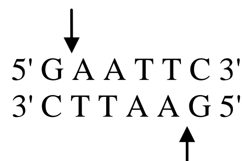
2.5 Restriction Endonuclease

Restriction endonuclease was recovered at the 1960s and was commercially available in the early of 1970s. Restriction endonuclease was named for their ability to cleave the invading foreign DNA into bacterial cells (Pingoud and Jeltsch, 2001). In order to protect themselves from the restriction digest, DNA methyltransferases occur as methyl groups are transferred to adenine or cytosine to produce N6-methlyadenine or 5-methylcytosine. As invading foreign DNA is not modified, they will be degraded by the host restriction enzyme. From the past 40 years, over 3500 restriction enzymes have been isolated from different strains of bacteria (Martin *et al*., 2006). The restriction endonuclease can recognize specific DNA sequence and cleave double-stranded DNA by digesting the two phosphodiester bonds formed within each strand of the DNA (Williams, 2001). Generally, restriction digestion involve the recognition

of binding site, binding of the restriction enzyme to the DNA, cleavage of the DNA following by the release of the enzyme.

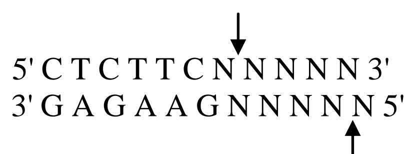
Based on Raymond and Williams (2001), restriction enzyme can be classified into four classes depending on their structure, recognition site, cleavage site, activators and also the co-factors. The majority of the enzymes recognize palindromic sequence with 4, 6 or 8 bases long. As class II enzymes have been studied extensively, they are proved to be important and are widely used in the molecular biology study involving genomic mapping, restriction fragment length polymorphism (RFLP), Cleaved Amplified Polymorphic Sequence (CAPS), cloning and others (Goppelt *et al.*, 1980).

*Eco*RI is one of the restriction enzyme which has been studied extensively. This restriction enzyme can be found in *Escherichia coli* RY13 and it is belonging to the class II enzyme. *Eco*RI recognizes the specific sequence shown as below:



When the *Eco*RI encounter this sequence, it will cleave between the G and A base residues in both DNA strand, producing sticky ends in both strands.

Besides, *Ear*I is another restriction enzyme that belongs to the class II enzyme. This enzyme has been isolated from *Enterobacter aerogene*. Based on Polisson and Morgan (1988), *Ear*I recognizes 6 base non-palindromic sequence 5’CTCTTC3’ and it cleaves at the sites shown by the arrows as below:



In restriction digestion, star activity is one of the major problems concerned. A restriction enzyme will show star activity under nonstandard condition by cleaving at sites that are similar, but not identical with its normal recognition sites. These nonstandard conditions include low ionic strength; presence of organic solvents in the reaction; high concentration of glycerol, extremely high concentration of enzyme is used, prolonged incubation time with enzyme and the use of incorrect cofactor (Williams, 2001). Therefore, optimal condition has to be used in order to obtain a good result from restriction endonuclease digestion.

2.6 Previous Study

In Pauline (2009), DNA from 5 *S. parvifolia* ssp. *parvifolia* mother trees were extracted and subjected to PCR analysis using primer SPPT3-F and SPPT3-R designated based on the full length cDNA of S*pCesA1*. The full length cDNA of *SpCesA1* can be accessed with GenBank accession number GQ338420. In the study, sequence homology search was performed using BLASTn and the result showed that the partial *CesA1* genes from all of the 5 samples have high degree of similarity (91%) with the *CesA* gene from *Betula luminifera* and *Betula platyphylla*. Following that, comparison of gene structure between *CesA*3 genomic DNA from *Eucalyptus grandis, CesA*3 mRNA from *E. grandis,* full-length *CesA1* cDNA from *S. parvifolia* ssp. *parvifolia* and the consensus sequence of all of the five 802 bp *CesA1* amplicons were performed. The outcome of the comparison is shown in Figure 2.3. From the 802 bp *CesA1* gene, an intron was detected from nucleotide 389 to 425. The intron-exon junction found in the 802 bp *CesA1* gene is conserved as it is also found in the *E. grandis CesA*3 genomic DNA.

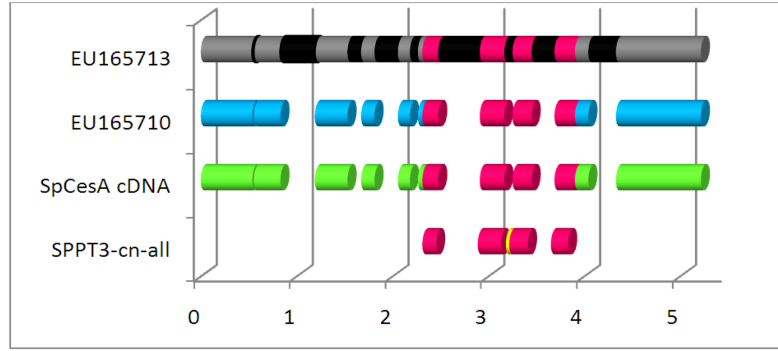


Figure 2.3: Comparison of the gene structure of *Eucalyptus grandis CesA*3 genomic DNA (accession no.: EU165713), *E. grandis CesA*3 mRNA (accession no.: EU165710), full-length *CesA1* cDNA from *S. parvifolia* ssp. *parvifolia* (Lau, 2008), and consensus sequence from all of the five *CesA1* amplicons. Colored cylinders represent exons and black cylinders represent introns. The intron portion as predicted from *CesA1* amplicons is shown in solid yellow.

In the study, a total of 7 SNPs and an insertion-deletion (INDEL) were detected. Since all of the SNPs were found in exons, it is predicted that the chances to detect SNPs in *S. parvifolia* ssp. *parvifolia* is higher in the exons compared to the introns. Then, two restriction enzymes (*EcoR*I and *Ear*I) were detected using NEBcutter V2.0 software for the two SNPs sites in the partial *CesA1* amplicons. The recognition site and the cutting site of the two restriction enzymes are shown in Figure 2.4. The *Ear*I will restrict at the site number 48 producing two bands with 48 bp and 754 bp while *EcoR*I will restrict at the site number 370 producing two bands with 370 bp and 432 bp. By using both of these restriction enzymes, CAPS marker can be developed in future study.

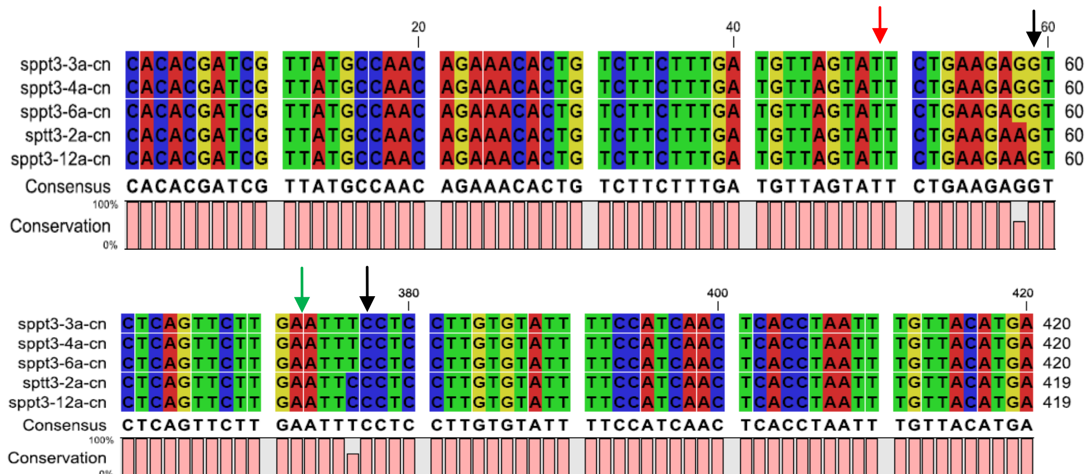


Figure 2.4: Alignment of consensus sequences for sample 2a, 3a, 4a, 6a and 12a. The black arrow indicates the SNPs site, the red arrow indicates the *Ear*I cutting site and the green arrow indicates the *EcoR*I cutting site (Pauline, 2009).

CHAPTER III

MATERIALS AND METHODS

3.1 Total Genomic DNA Extraction

A total of 12 leaf samples of *Shorea parvifolia* ssp. *parvifolia* were collected from Seed Bank, Sarawak Forestry Corpotation (SFC), Kuching. Then, the total genomic DNA was extracted using modified CTAB method from Doyle and Doyle (1990). 4 ml of CTAB extraction buffer [100mM Tris-Cl pH 8.0; 1.4M NaCl; 20mM EDTA; 2% CTAB; 1% polyvinlylpyrrolidone (PVP); 2% (v/v) β-mercaptoethanol] was added into a 15 ml Falcon tube and preheated for 30 minutes at 65°C. Then, liquid nitrogen was added to approximately 0.5 g of the leaf and the plant material was ground into fine powder using mortar and pestle. After that, the powder was added into the pre-heated buffer, swirled gently and incubated at 65°C for 30 minutes.

Following that, 600 µl of the mixture in Falcon tube was transferred into 1.5 ml microcentrifuge tubes. Then, an equal volume of chloroform/isoamyl alcohol (24:1; v/v) was added and mixed gently. The sample was centrifuged at 13,000 rpm for 10 minutes and the aqueous phase was then transferred into clean 1.5 ml microcentrifuge tubes. The chloroform/isoamyl alcohol (24:1; v/v) extraction and centrifugation steps were repeated to remove carbohydrate impurities and the final aqueous phase was transferred into new 1.5 ml microcentrifuge tubes. Then, 2/3 volume of cold isopropanol (-20°C) was added in order to precipitate the nucleic acids. The sample was stored at -20°C freezer for one hour.

After that, the sample was centrifuged at 13,000 rpm for 10 minutes and the supernatant was poured off. The pellet was washed using 500 µl of cold 70% ethanol and air-dried for 15 minutes. The pellet was then resuspended in 40 µl of ddH_2O and stored at -20°C.