Faculty of Resource Science and Technology

MOLECULAR CLONING OF HYPERVARIABLE REGION (HVR II) FROM *CELLULOSE SYNTHASE* GENE IN KELAMPAYAN
(Neolamarckia cadamba)

Meldon Calvin Sim Wee Yang

Bachelor of Science with Honours
(Resource Biotechnology)
2011

Universiti Malaysia Sarawak

Faculty of Resource Science and Technology

Molecular Cloning of Hypervariable Region (HVR II) from *Cellulose Synthase* Gene in Kelampayan (*Neolamarckia cadamba*)

Meldon Calvin Sim Wee Yang

(21436)

Bachelor of Science with Honours

Resource Biotechnology

2011

ACKNOWLEDGEMENTS

DECLARATION

I hereby declare that this thesis is of my original work except for quotations and citations, all of which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UNIMAS or any other institutions.

Meldon Calvin Sim Wee Yang

Resource Biotechnology Programme

Department of Molecular Biology

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Table of Contents

Materials and Methods

Results and Discussion

List of Abbreviations

A	Adenine
bp	Base pair(s)
C	Cytosine
CesA	Cellulose synthase protein
CesA	Cellulose synthase gene
CSR	Class-specific region
cDNA	Complementary deoxyribonucleic acid
ddH₂O	Double-deionized distilled water
DEPC	Diethylpyrocarbonate
dNTP	Deoxyribonucleotide triphosphate
G	Guanine
HVR	Hypervariable region
HVR2F	Hypervariable region II forward primer
HVR2R	Hypervariable region II reverse primer
HVRII	Hypervariable region II
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LAIX	LB/Ampicillin/IPTG/X-Gal
LB	Luria broth
MgCl₂	Magnesium chloride
mRNA	Messenger ribonucleic acid
*Nc*CesA	*Neolamarckia cadamba* cellulose synthase
RT-PCR	Reverse transcriptase-polymerase chain reaction
T	Thymine
Tₘ	Melting temperature
X-Gal	Bromo-chloro-indolyl-galactopyranoside

List of Tables

List of Figures

MOLECULAR CLONING OF HYPERVARIABLE REGIONS (HVR II) FROM CELLULOSE
Synthase Gene In Kelampayan *(Neolamarckia cadamba)*

MELDON CALVIN SIM WEE YANG

Resource Biotechnology
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Neolamarckia cadamba or locally known as Kelampayan, is one of the light hardwood trees that holds great prospects as a source of raw materials for the paper, pulp, wood and biofuel industries. Sufficient information on *cellulose synthase* gene, especially the hypervariable region II component involved in wood formation of Kelampayan is imperative for future applications. The aim of this study is to identify and clone the hypervariable region (HVRII) from *cellulose synthase* gene of *N. cadamba*. Total RNA was isolated from young leaves and developing xylem tissues of *N. cadamba*. The cDNA of *N. cadamba* cellulose synthase HVRII (NcCesAHVRII) region was amplified using reverse transcription-PCR (RT-PCR) approach with forward degenerate primers, HVR2F (5'-TGYTATGTYCAGTTYCCWC-3') and reverse degenerate primer, HVR2R (5'-GANCCRTARATCCAYCC-3'). *NcCesA1HVRII*, *NcCesA2HVRII* and *NcCesA3HVRII* were successfully sequenced. *NcCesA1HVRII* and *NcCesA3HVRII* were clustered to 2 distinct clades implicated with secondary cell wall development. Analysis on *NcCesA2HVRII* suggests renaming it as *NcCslD1HVRII* due to high similarity with various plants' *CslD-HVRII*. This study will provide an easier and faster access to *NcCesAHVRII* sequences to further understand the role of NcCesA/NcCslD protein for future application such as in selecting trees with optimal cellulose content required for specific industries.

Key words: *Neolamarckia cadamba*, reverse transcription-PCR (RT-PCR), cellulose synthase (CesA), cellulose-synthase like (Csl), hypervariable region II (HVRII),

ABSTRAK

Neolamarckia cadamba ataupun lebih dikenali sebagai Kelampayan merupakan salah satu pokok kayu keras ringan yang memiliki potensi yang tinggi sebagai bahan mentah untuk industri pembuatan kertas, pulpa, kayu dan biofuel. Informasi yang mencukupi terhadap selulosa sintase, terutamanya terhadap komponen hypervariable region II yang terlibat dalam formasi kayu amat penting untuk aplikasi masa hadapan. Objektif kajian ini adalah untuk menganal pasti dan mengklon hypervariable region (HVRII) dari gen selulosa sintase N. cadamba. RNA keseluruhan diekstrak dari daun muda dan tisu xylem N. cadamba yang sedang berkembang. cDNA selulosa sintase HVRII (NcCesAHVRII) N. cadamba diamplifikasikan dengan menggunakan transkripsi berbalik-tindak balas berantai polymerase (RT-PCR) dengan primer degenerate hadapan HVR2F (5'-TGYTATGTYCAGTTYCCWC-3') dan primer degenerate belakang, HVR2R (5'-GANCCRTARATCCAYCC-3'). NcCesA1HVRII, NcCesA2HVRII dan NcCesA3HVRII berjaya dijujuk. NcCesA1HVRII dan NcCesA3HVRII dikumpul kepada dua kelas berbeza yang terlibat dalam perkembangan dinding sell sekunder. Analisa ke atas NcCesA2HVRII mencadangkan pengubahan nama kepada NcCslD1HVRII atas persamaan yang tinggi dengan pelbagai CslD-HVRII tumbuhan. Kajian ini akan memudahkan serta mempercepat akses kepada jujukan-jujukan NcCesAHVRII bagi pemahaman yang lebih mendalam ke atas peranan protein NcCesA/NcCslD bagi aplikasi masa hadapan seperti dalam pemilihan pokok yang mengandungi kandungan selulosa optima yang diperlukan bagi industri terbabit.

Kata kunci: Neolamarckia cadamba, transkripsi berbalik-tindak balas berantai polymerase (RT-PCR), selulosa sintase (CesA), selulosa-sintase mirip (Csl), hypervariable region II (HVRII),

SECTION I

INTRODUCTION

Currently, an effort has been taken by a group of researchers from the Forest Genomic Laboratory in Universiti Malaysia Sarawak (UNIMAS) in generating a large scale genomic resource for wood formation of *Neolamarckia cadamba* or locally known as Kelampayan. Kelampayan from the Rubiaceae family is a large and fast-growing tree species with straight cylindrical bole (Acharyya *et al.*, 2010). According to Ismail *et al.* (1995), Kelampayan self-prunes and grows well in exploited and denuded areas. Besides being easy to grow and manage, Kelampayan also has a highly dense fibre characteristic, an important factor in determining wood quality (Panshin and De Zeeuw, 1980). This makes Kelampayan one of the best raw materials for sawn timber, veneer, chips, pulp and composites (Monsalud and Lopez, 1967; Peh, 1970; Phillips *et al.*, 1979). Kelampayan has yet to be used in the wood, paper and pulp industries in Malaysia. However, the future of this tree as a source of raw material in the industry is great (Ismail *et al.*, 1995).

Cellulose is one of the most abundant biopolymer. Cellulose is synthesised in plants, a few bacteria and fungi, most algae and some animals (Brown *et al.*, 1996). Cellulose is a homopolymer of β-1,4-linked glucan chains which have alternate glucose molecule flipped by 180° (Brown *et al.*, 1996). Brown *et al.* (1996) further described that the degree of polymerization of the glucan chain can range from about 2,000 to more than 25,000 glucose residues. This natural biopolymer is the major constituent in plant cells, primary and secondary plant cell walls, serving to maintain plant structure and rigidity (O'Sullivan, 1997). In fact, cellulose makes up about 40-50% of dry wood (Kumar and Fladung, 2004). This makes cellulose a highly demanded natural biopolymer in the current

1

market. To date, it has been of our knowledge that cellulose synthase is a key enzyme responsible in the biosynthesis of cellulose in plants (Campbell *et al.*, 1997; Taylor, *et al.*, 2000). However, the regulation of the deposition of cellulose and the parameters that control the quantity and quality of the deposited cellulose remains unclear (Joshi and Mansfield, 2007).

Cellulose synthase share certain similar domain structures across species. These include a zinc finger, several transmembrane domains, conserved residues and also two hypervariable regions (HVR), HVRI and HVRII (Holland *et al.*, 2000). Researchers have not fully understood on the regulation and function of *CesA* gene itself and each of its domain structures, including HVRII region. According to Kumar and Fladung (2004), they speculated that HVRII region might be involved in the regulation of the quantity and quality of the cellulose synthesised in plants. Holland *et al.* (2000) instead, writes on the probable role of the region in interacting with other unique cell-type-specific proteins involved in the biosynthesis of cellulose. Clearly, an accurate understanding on the function and regulation of cellulose synthase, specifically the role of HVRII region is still unclear.

Thus, the research objective was to clone HVRII region of *cellulose synthase* gene (*NcCesA*) from *N. cadamba* trees. This was done by reverse transcription-PCR (RT-PCR) using degenerate primers designed based on the conserved sequence flanking the HVRII region. The amplified *Nc*CesA HVRII region was sequenced and subsequently, *in-silico* analysis was performed to better understand its phylogenetic relationship with other tropical woody plants. The findings may provide invaluable information known thus far on *NcCesA HVRII* region for future research and applications probably in synthesising artificial cellulose for the paper, pulp, wood and biofuel industries.

2

SECTION II

LITERATURE REVIEW

2.1   Selection of Species Studied

2.1.1   Family Rubiaceae

The Rubiaceae comprises of about 450 genera and over 6,500 species of trees, shrubs and some herbs, including several lianous forms. It grows in the tropical and sub-tropical regions. This family of plants are usually characterized by the inferior ovary, opposite leaves with stipules and absence of internal phloem, with the exception of a few aberrant taxa (Bremer, 1992).

2.1.2   *Neolamarckia cadamba*

Neolamarckia cadamba (Roxb.) Bosser, *Anthocephalus cadamba* (syn. *A. chinensis*) or more commonly known as Kelampayan, belongs to the Rubiaceae family and is widely distributed throughout Bangladesh, Nepal, India, Myanmar, Sri Lanka, the Philippines, Indonesia, Malaysia and Papua New Guinea (Willis, 1973; Banerji, 1977, 1978; Sahu, *et al.*, 2000). According to Manandhar and Manandhar (2002), Kelampayan propagates through seeds or shoot cuttings. Kelampayan thrive in tropical secondary forest such as the openings after logging and also along logging trails (Meijier, 1970). Otherwise, Kelampayan can be found growing by streams and rivers, in open sites and deep moist alluvial soils in the lowlands to mountain forest of about 1000 m tall as a pioneer species (Choo *et al.*, 1999). In India, Sri Lanka, Myanmar, Indonesia, Malaysia and the Philippines, Kelampayan plantations are raised for industrial purposes. Kelampayan is

3

planted in Java, Indonesia to replace poor teak plantations after harvest and in Northern Sumatra and Central Kalimantan for their pulpwood industries (Nair and Sumardi, 2000; Nair, 2007). In his book, Nair (2007) also wrote that Kelampayan was introduced to and planted in other tropical and subtropical country such as South Africa, Puerto Rico and Taiwan.

Kelampayan is a medium to large-sized tree growing up to about 37.5 to 40 m tall and about 2.4 m in girth with a straight cylindrical bole (Acharyya *et al.*, 2010). According to Acharyya *et al.* (2010), the bark of Kelampayan tree is gray, smooth in young trees, rough and longitudinally fissured in old trees (Figure 2.1b). The leaves are simple and elliptic-oblong in shape with orange or yellow flowers in solitary globose head and pseudo carp fruits easily found all over India (Figure 2.1a) (Acharyya *et al.*, 2010).

(a)   (b)

Figure 2.1 *Neolamarckia cadamba*. *(a) Simple, elliptic-oblong shaped leaves and orange solitary globose flowers and (b) Young tree with smooth and gray-colored bark. (*Adapted from http://www.flowersofindia.net/catalog/slides/Kadam.jpg)

Kelampayan holds great prospects as the source of raw materials for the paper and wood industries in the future. Kelampayan has a soft and light timber, white heartwood with a yellow tinge which darkens to a creamy yellow on exposure. Apart from that, Kelampayan timber has straight grains with texture that is moderately fine and even. Wood density of Kelampayan ranges from 370 to 465 kg m^{-3}. Most importantly, Kelampayan is easy to treat, re-saw and cross-cut. Moreover, planning of the wood is easy, producing a smooth planed surface (Choo *et al.*, 1999). Choo *et al.* (1999) further stated that Kelampayan wood is suitable for plywood manufacture, packing case, wooden sandal, toys, disposable chopsticks and possibly as a short-fibred pulp.

Kelampayan is also important in treating illnesses. Various parts of Kelampayan tree can be used as an antidiuretic, antimalarial, antidiarrhoeal (Kitagawa *et al.*, 1996; Alam *et al.*, 2008), antimicrobial, antioxidant, laxative (Acharrya *et al.*, 2010), in fever treatment, lowering blood sugar in diabetes mellitus patients (Bussa and Pinnapareddy, 2010), as a mouth wash, tonic, wounds, uterine complaints, menorrhigia, colitis, eye inflammation, leprosy, dysentery, stomatitis, skin diseases, anemia, tumor, for the improvement of semen quality (Umachigi *et al.*, 2007; Acharrya *et al.*, 2010; Gurjar *et al.*, 2010), an astringent anti-hepatoxic (Kapil *et al.*, 1995), a galactagogue, aphrodisiac (Chandrashekar *et al.*, 2010), analgesic and anti-inflammatory (Mondal *et al.*, 2009).

Global warming has become one of the most significant catastrophes which is gradually getting from bad to worse as years passes by. In a joint project between the Netherlands and Malaysia, efforts have been taken to rehabilitate logged forests and also to reduce the amount of CO_2 gases in the atmosphere which is causing the greenhouse effect. In Sabah, Malaysia, Kelampayan has already been trial-planted on log landing areas and

skid trails. Kelampayan was chosen as one of the species of trees to be planted as it possesses potential commercial use whilst able to reduce the greenhouse effect (Pinso and Costa, 1993).

2.2   Cellulose Synthase *(CesA)* Gene

Cellulose synthase gene is the genes responsible in the biosynthesis of cellulose, one of the most important components in wood and paper. Interestingly, cellulose is synthesised not only in plants, some algae (Roberts and Roberts, 2009), bacteria and fungi but it is also synthesised in some animals (Richmond, 2000). According to Richmond (2000), there are currently more than 20 full-length *CesA* genes sequenced available in the various genebanks which is highly similar to each other across the entire length of the encoded protein with some exception from two small regions of variability.

With small introns present, the size of *CesA* genes that ranges from 3.5 to 5.5 Kbp encodes proteins with 985 to 1088 amino acids in length (Richmond, 2000). In general, CesA protein includes a zinc finger, several transmembrane domain, conserved residues and also 2 hypervariable region (HVR), HVRI and HVRII (Figure 2.2) (Holland *et al.*, 2000; Lau *et al.*, 2009; Richmond, 2000).
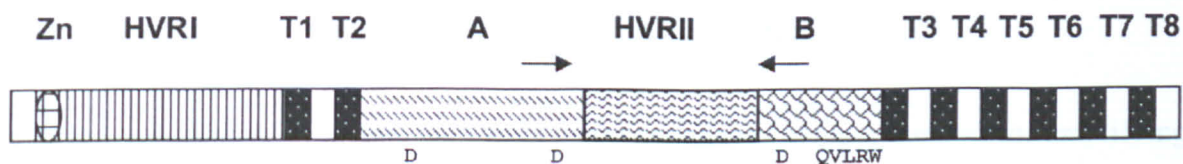
Figure 2.2 Diagrammatic representation of general components present in a plant *cellulose synthase (CesA)* gene showing the presence of a putative zinc-binding motif (Zn) followed by hypervariable region I (HVRI). T1 to T8 blocks represents the predicted transmembrane domains. The second hypervariable region (HVRII) is flanked by highly conserved sub-domains A and B. (Modified from Liang and Joshi, 2004)

6

2.3   Hypervariable Region II (HVRII)

Hypervariable region II can be found in the *CesA* gene (Figure 2.2). Amino acid sequences between the highly conserved motifs of ALYG and VISCG are considered as HVRII regions (Liang and Joshi, 2004). With the current information available, HVRII region in plant *CesA* gene has a sequence length ranging from about 500-600 bp long. HVRII region was proposed to be renamed as class-specific regions (CSR) instead (Vergara & Carpita, 2001) due to its conservation among CesA orthologs from different plant species and limited conservation among CesA family members of the same species. This region is thought to play a role in interaction with other unique cell-type-specific proteins involved in the biosynthesis of cellulose (Holland *et al.*, 2000). Srivastava (2002) speculated that these plant-specific domains are thought to be involved in functions unique to plant such as the binding of sucrose synthase and interaction with proteins associated with cytoskeleton or other accessory proteins. These domains are also thought to give specificity for plant cells, tissues and organs (Srivastava, 2002). According to Liang and Joshi (2004), the high sequence similarity of HVRII region among CesA orthologs makes it useful in distinguishing individual family members of *CesA* gene and could prove useful for full-length cDNA isolation (Ranik and Myburg, 2006).

2.4   Reverse Transcription - Polymerase Chain Reaction

Reverse transcription - Polymerase Chain Reaction (RT-PCR) is a method for amplifying defined sequences of RNA using enzymes *in vitro* (Rappolee *et al.*, 1988) into complementary DNA (cDNA) strands and allows for the analysis of minute or limited amounts of RNA samples. RT-PCR is suitable to compare the levels of mRNAs in

different sample populations, to characterize patterns of mRNA expression, to discriminate between closely related mRNAs and also to analyse RNA structure (Bustin, 2000). According to Chelly *et al.* (1989), Kwok and Higuchi (1989) and Henke *et al.* (1997), although a highly sensitive technique, RT-PCR compromises the specificity of the reaction by increasing the likelihood of contamination and detection of false positives with the inability to distinguish genuine low-level transcription from illegitimate transcription.

Generally, RT-PCR reaction vessel contains a mixture of buffers, nucleotides, primers, thermostable polymerase and RNA from the specimen of interest. Successfulness of RT-PCR involves a few simple principles which are denaturation of the double-stranded DNA, the annealing of primers and primer extensions (Schochetman *et al.*, 1988). RNA sample must be reverse transcribed enzymatically with suitable primers into cDNA. The cDNA generated is subsequently used as template for replication.

To amplify the cDNA, denaturation is usually done by simply heating the cDNA to 95-100°C to separate the duplex held together by the hydrogen bonds although there are other chemical and physical methods. In the annealing process, the specific primers, single-stranded sequence of oligonucleotides which are each complementary to one of the original DNA strands, to either the 5' or 3' sides of the sequence of interest, will anneal to the cDNA. The presence of the primers in a high excess concentration makes the annealing of primers with single-stranded cDNA to be much higher than self annealing of the cDNA. Once the annealing process starts, DNA polymerase will synthesise new strands of DNA, adding nucleotides complementary to those in the unpaired DNA strands in which the primer has annealed. The number of DNA strand doubles with each cycle. According to Schochetman *et al.* (1988), a single copy of DNA can be increased up to 1 million copies after just thirty cycles. Subsequent analysis of PCR product can be done by the

incorporation of a radiolabel, staining the agarose gel with ethidium bromide or by probing the product post-PCR (Schmittgen *et al.*, 2000).

2.5 Degenerate Primers

A Polymerase Chain Reaction (PCR) primer is called a degenerate primer when there are several possible bases on one or more positions (Linhart and Shamir, 2002). For example, in the primer GA[C,T]AT[A,T,C], the third position is C or T and the 6th position can be A, T or C. To calculate the degeneracy of the primer, multiply the number of possible bases contained in a sequence. Based on the example given, the degeneracy of the primer is 6. Degenerate primers have been used in various applications as they are easy, cheap to produce and useful for amplifying several related genomic sequences (Linhart and Shamir, 2002). Degenerate primers are useful when only the gene sequence of related species is known but the gene sequence of interest of the organism in question is unknown. The forward and reverse degenerate primers can be designed based on the conserved regions flanking the gene of interest. However, it should be noted that the further related the organism is, the harder it is to design a degenerate primer.

Primer sites selection is a compromise between placing primers at the codons of amino acids which are most conserved and also placing them at codons with the lowest possible degeneracy. When degeneracy is too high, amplification of unrelated sequence may occur and thus losing specificity (Linhart and Shamir, 2005). Certain amino acids are being coded by more than one codon. Some organisms have strong biases in using particular codons to encode for certain amino acids (Anderson and Sharp, 1996; Akashi, 1996; Gustaffson *et al.*, 2004; Martindale, 1989). Theoretically, by only including those

few common codons the degeneracy of the primer pool can be decreased. However, such assumptions risk the pool of primers to become useless (Ogino *et al.*, 2005).

Primers should have a primer-length window of about 20-30 nucleotides in length (Linhart and Shamir, 2005). The base composition should contain about 50% guanine (G) and cytosine (C), especially at the 3' ends to prevent the ends from "breathing" besides increasing the efficiency of priming. However, due to the stability of annealing during PCR analysis, runs of three or more Cs or Gs at the end of 3' ends of primers may promote mispriming at the G or C rich sequences. The primer(s) designed should also have a melting temperature (T_m) of about 55-80°C. Primer dimers will be formed preferentially to any other product during PCR if the 3' ends of the primers are complementary to each other. Hairpins formation probability can also happen due to primer self-complementarity

According to Kwok, *et al.* (1994), the success rate of degenerate PCR is highly dependent on the PCR thermal profile. The amplification profile of degenerate PCR should start with a non-stringent annealing temperature ranging from 35-45°C for 2-5 cycles, followed by 25-40 cycles at a more stringent annealing temperature. The non-stringent condition allows the hybridization of the short complementary primer portion to the target (Kwok, *et al.*, 1994). After the second cycle, the 5' extension will also be amplified, serving as the template for subsequent rounds of amplification. After that, increased specificity can be achieved by using the more stringent annealing temperature.

SECTION III

Materials and Methods

3.1   Degenerate Primer

Degenerate primers HVRIIF 5'-TGYTATGTYCAGTTYCCWC-3' of 16× degeneracy and

HVRIIR 5'-GANCCRTARATCCAYCC-3' of 32× degeneracy from Liang and Joshi's

(2004) research on HVRII regions of Aspen was used. Degenerate primers were designed

based on the conserved region in CesA protein sequence CYVQFPQ and GWIYGS.

3.2   Plant Materials

Developing xylem and fresh young leaves tissues from *Neolamarkia cadamba* (Roxb.)

Bosser were used for total RNA extraction.

3.3   Apparatus Treatment

All apparatus required during total RNA extraction such as 1.5ml microcentrifuge tube,

pipette tips, beakers, spatula, mortar and pestle and 15ml Falcon tubes were treated with

0.1% DEPC water for overnight prior to autoclaving. PCR tubes and microcentrifuge tubes

were autoclaved prior to use.