Sequence polymorphism of *arabinogalactan* and *beta tubulin* gene in kelampayan (*Neolamarckia cadamba*).

TANG LEH EE (22510)

A project progress report submitted in partial fulfillment of the Final Year Project (STF 3015) course.

Supervisor: Dr. Ho Wei Seng

Co-supervisor: Dr Pang Shek Ling

Resource Biotechnology Programme
Department of Molecular Biology

Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

2011

ACKNOWLEDGMENTS

Above all, I would like to thank God for His blessing and guidance upon the completion of this project. My sincere gratitude to my supervisor, Dr. Ho Wei Seng for his guidance, advice and perseverance throughout the completion of this project. Besides that, I would like to thank my co-supervisor, Dr. Pang Shek Ling for her valuable advice and comments on this project. I would also like to dedicate my appreciation to Mr. Liew Kit Siong, Ms. Tchin Boon Ling and others master students in Genomic Forestry Laboratory for patiently guiding me through every experimental process and assistance. My special thanks go to all my beloved lab members for providing suitable and supportive environment for completing my project. Finally, thank you to my family and friends for the encouragement and support throughout the completion of my project.

DECLARATION

I 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 submitted for any other degree at UNIMAS or any other institutions.

_____

Tang Leh Ee

Resource Biotechnology Programme

Department of Molecular Biology

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Table of Contents

List of Abbreviations

AGPs	Arabinogalactan proteins
β-TUB	Beta- tubulin
bp	Base pair
cDNA	Complementary DNA
CIA	Chloroform-Isoamyl Alcohol
CTAB	Cetyltrimethylammonium Bromide
DDH_2O	Deionized water
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
EtBr	Ethidium bromide
GPI	Glycosylphosphatidylinositol
HRGPs	Hydroxyproline-rich glycoproteins
LB	Luria both
LD	Linkage Disequilibrium
NaCl	Sodium chloride
PCD	Programmed cell death
PCR	Polymerase chain reaction
PVP	Polyvinylpyrrolidone
RNase A	Ribonuclease A
SNP	Single nucleotide polymorphism
UV	Ultraviolet

List of Figures

List of Tables

SEQUENCE POLYMORPHISM OF THE *ARABINOGALACTAN* AND *BETA TUBULIN* GENE OF KELAMPAYAN (*NEOLAMARKIA CADAMBA)*

TANG LEH EE

Resource Biotechnology
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Neolamarckia cadamba or locally known as Kelampayan is an evergreen and tropical tree native to South Asia under the family of Rubiaceae. Kelampayan is the best raw material for plywood industry and also used as raw materials for pulp and paper industry. Arabinogalactan Proteins (AGPs) are extracellular proteoglycans that are implicated in plant growth and development. Tubulin is a major cytoskeleton protein and as a building block of microtubules. The cytoskeleton is important in plant cell shapes, cell division and cell transport. The main objective of this study was to identify the single nucleotide variation in the *AGP* and *beta-tubulin* gene of kelampayan. Genomic DNA was firstly extracted from six *N. cadamba* trees. Targeted DNA sequence of *arabinogalactan (AGP)* gene and *beta tubulin* (*β-TUB*) were amplified by Polymerase Chain Reaction (PCR) by using primer set. The 340 bp of *AGP* and 491 bp of *β-TUB* amplicons were subjected to BLASTn analysis to carry out the sequence homology search in the NCBI. Sequence alignment then was carried out by using ClustalW software for single nucleotide polymorphism (SNP) detection. One sequence variation within *arabinogalactan* gene which was caused by single nucleotide substitution was detected. Five sequence variations and one InDel polymorphism within *beta tubulin* were detected. The finding of SNPs in *AGP* and *β-TUB* was very important as a valuable marker in molecular selective breeding.

Keywords: *Neolamarckia cadamba*, *arabinogalactan* gene (*AGP*), *Beta tubulin* (*β-TUB*), polymerase chain reaction (PCR), Single nucleotide polymorphism (SNP)

ABSTRAK

Neolamarckia cadamba atau dikenali sebagai kelampayan merupakan spesies pokok tropika di Asia Selatan yang dikategori di bawah keluarga Rubiaceae. Kelampayan ialah bahan yang terbaik untuk kegunaan industri kayu lapis dan digunakan sebagai bahan mentah untuk industri pulp dan kertas. Proteins Arabinogalaktan (AGPs) ialah proteoglikan ekstraseluler yang terlibat dalam pertumbuhan dan perkembangan tumbuhan. Tubulin beta (β-TUB) ialah proteins sitoskeleton utama dan digunakan sebagai asas pembuatan mikrotubul. Sitoskeleton adalah sangat penting dalam sel pembentukan, sel pemecahan dan sel pergerakan bagi tumbuh-tumbuhan. Tujuan kajian ini adalah untuk mengenal pasti SNP di gen AGP dan β-TUB di kelampayan. DNA diekstrak daripada enam pokok ibu N. cadamba. Jujukan DNA bagi AGP dan β-TUB diamplifikasi dengan menggunakan teknik tindakbalas berantai polymerase (PCR) dengan bantuan oleh pasangan pencetus. Produk PCR bagi AGP yang bersaiz 340 bp dan β-TUB yang bersaiz 491 bp dianalisa dengan BLASTn untuk mencari urutan homologi dalam pangkalan data NCBI. Penjajaran urutan kemudiannya dianalisa oleh ClustalW bagi pengesanan polimorfisme nukleotida tunggal (SNP). Satu SNP bagi gen arabinogalaktan yang disebabkan substitusi nukleotida tunggal telah dikesan. Tambahan pula, penemuan 5 SNP dan satu polimorfisme InDel bagi gen tubulin beta telah dikesan. Penemuan SNP dalam gen AGP dan β-TUB adalah sangat penting penghasilan penanda molekul dalam pembiakan selektif molekul.

Kata kunci: Neolamarckia cadamba, gen arabinogalaktan (AGP), tubulin beta (β-TUB), teknik tindakbalas berantai polymerase (PCR), polimorfisme nukleotida tunggal (SNP)

SECTION I

INTRODUCTION

Nowadays, a high mass production of genetically improved seeds is required to meet the growing demand of planted forests. Therefore, a number of seed production areas for *N. cadamba* on natural stands are established to help in production of quality seeds for large-scale plantation programs. Besides that, the plantation of fast growing species such as *N. cadamba* have the capability to supply the bulk of wood needs on a long-term basis thus reduce the harvest pressure on natural forest.

Neolamarckia cadamba or locally known as Kelampayan is an evergreen and tropical tree under the family of Rubiaceae. Kelampayan is the best material for plywood industry. It also used as raw materials for pulp and paper industry. Besides that, Kelampayan is characterized as a large, deciduous and fast growing that gives economic returns within 8 to 10 years (Joker, 2000). Kelampayan grows up to 40 m to 45 m high and the trunk has a diameter of 100 cm to 160 cm. Furthermore, its leaves are about 13 cm to 32 cm long and the flowering will begin when the tree is 4 to 5 years old. Kelampayan timber is white with coarse surface due to the presence of large vessels and it is used for light-weight purposes such as picture frames, disposable chopstick, wooden sandals, general utility furniture and plywood (Lim *et al*., 2005).

Arabinogalactans are a biopolymer consisting of arabinose and galactose monosaccharides and are attached to proteins and forming arabinogalactan proteins (AGPs). AGPs have important roles in plant growth and development (Zhang *et al.,* 2000), plant defense (Showalter and Varner, 1989), cell differentiation (Pennell and Roberts, 1990), somatic embryogenesis (Chapman *et al.,* 2000) and pollen tube growth (Wu *et al.,* 2000). AGPs are normally found in the plant cell wall and are heavily glycosylated with

only 2% to 10% protein. Secondary cell wall thickening is due to an interaction between cellulose biosynthesis with the transient appearance of AGPs epitope which will influence the specific gravity of early wood (Sewell *et al.*, 2000). AGPs exist in phloem sieve elements and other tissues (Gao and Showalter, 2000) but are more often associated with xylem development and differentiation by marking the initiation of secondary cell wall thickenings that are caused by the deposition of cellulose microfibrils.

Tubulin is one of the small family of globular proteins. Molecular weight for tubulin is approximately 55 kilo Daltons. The common family of tubulin are α and β *tubulin* which are essential as a building block of microtubules. The cytoskeleton of eukaryotic cells is composed of three distinct types of microfilaments which are microtubules, actin filaments and intermediate filaments. So, microtubules in cytoskeleton help to provide the plant cell with shape, structure and cell movement. Besides that, β *tubulin* will influence the formation, organization and function of microtubules. In an addition, β *tubulin* is also used to determine the orientation of cellulose microfibrils which is important in secondary cell wall formation of woody plants (Spokevicius *et al.*, 2007).

Molecular selective breeding is important in increasing the plant productivity and enhancement of wood quality whereas the traditional selection is expensive and time consuming by only observing the phenotypic characteristics on older tree. Hence, Single nucleotide polymorphisms (SNPs) are useful tool for the selection of germplasm with desired traits with the help of marker-assisted selection (MAS). SNPs are DNA sequence variation occurring when a single nucleotide namely A, T, C or G in the genome differs among individuals. SNPs can divide into coding sequences of genes and non-coding regions of genes. SNPs within a coding sequence will not change the amino acid sequence of the protein that is produced due to degeneracy of the genetic code. A SNP that lead to

the same polypeptide sequence is termed synonymous whereas a different polypeptide sequence is produced are nonsynonymous. A nonsynonymous change can lead to missense which change result in different amino acid or nonsense which change in premature stop codon (Gupta *et al*., 2001).

Thus, the objectives of this study are to design primers for amplifying *arabinogalactan* gene and *beta tubulin* gene from *N. cadamba.* Besides that, this study is also involved in identifying the presence of single nucleotide polymorphisms (SNPs) in *arabinogalactan* and *beta tubulin* genes from *N. cadamba* trees.

SECTION II

LITERACTURE REVIEW

2.1 *Neolamarckia cadamba*

Neolamarckia cadamba (Roxb.) Bosser or locally known as kelampayan is under family Rubiaceace (Naithani and Sahni, 1997; Yoganarsimhan, 2000). The bark and leaves of *N. cadamba* is reported to posses various medicinal uses such as astringent anti-hepatotoxic (Kapil *et al.,* 1995), antidiuretic, wound healing, antiseptic and anthelmintic (Gunasekharan *et al.,* 2006). The dried bark can be used to relieve fever and as a tonic while an extract of the leaves can serve as a mouth wash.

According to Patel and Kumal (2008) in the pharmacognostical studies of *N. cadamba*, the shape of *N. cadamba* leaves is broadly ovate, elliptic-oblong with entire margin, pulvinus base, mucronate apex, length ranged from 7.5 to 18 cm and breath is 4.5 to 16 cm. Through the microscopic studies, the leaf is dorsiventral with thick prominent midrib and uniformly thin lamina. Ground tissue of the midrib contains a broad outer zone of collenchymas and inner and middle zone of parenchyma. Vascular bundle built up of a horseshoe shape inverted vascular strands and two small top-shaped strands placed at the ends of the ‘U’-shaped strands. The epidermis of the midrib is distinct with squarish thick-walled cells and prominent cuticle.

(a) (b)

Figure 2.1 (a) Mature tree and (b) Seeds of *Neolamarckia cadamba*. (Adapted from Source: http://prothom-aloblog.com/users/base/computer/146).

2.2 Arabinogalactan protein

Arabinogalactan proteins (AGPs) are a family of highly glycosylated hydroxyproline-rich glycoproteins (HRGPs) that are analogous to animal proteoglycans (Nothnagel, 1997). AGPs are widely distributed in plasma membrane, cell wall and extracellular matrix (Gao *et al.*, 1999). AGPs are composed of 90% of carbohydrate which is predominantly galactose and arabinose and 10% of protein.

AGPs can be encoded into classical or non-classical due to the structural in the polypeptide backbones (Chen *et al.*, 1994). Both classical and non-classical AGPs have the common structure of N-terminal hydrophobic secretion signal sequence and a central domain rich in hydroxyproline, serine, alanine and threonine. Classical AGPs not only have N-terminal hydrophobic and also have hydrophobic C-terminal which is cleaved from mature protein (Du *et al.*, 1994). The hydrophobic C-terminal are replaced by a glycosylphosphatidylinositol (GPI) lipid anchor to attach the proteins to the plasma membrane as functioning in signal transduction (Oxley and Bacic, 1999).

AGPs have important roles in plant growth and development (Zhang *et al.*, 2000), plant defense (Showalter and Varner, 1989), cell proliferation (Serpe and Nothnagel, 1994), cell expansion (Willats and Knox, 1996), cell differentiation (Pennell and Roberts, 1990), somatic embryogenesis (Chapman *et al.*, 2000) and pollen tube growth (Wu *et al.*, 2000). Secondary cell wall thickening is due to an interaction between cellulose biosynthesis with the transient appearance of AGPs epitope which will influence the specific gravity of early wood (Sewell *et al.*, 2000). AGPs exist in phloem sieve elements and other tissues (Gao and Showalter, 2000) but are more often associated with xylem development and differentiation. For example, JIM13 monoclonal antibody shows species specific labeling during pattern formation in vascular tissue (Samaj *et al.*, 1998). JIM13 antibody reacts with the carbohydrate epitopes of AGPs in different protein backbones and will lead to thickening secondary cell wall and secondary xylem tracheary elements in roots and stems will completely differentiate (Gao *et al.*, 1999).

AGPs are used as markers for cells committed to programmed cell death (PCD) (Fukuda *et al.*, 1998). Secondary thickening of xylem cells marks the initiation of PCD that leads to the formation of mature wood. Addition of *β*-glucosyl Yariv reagent to suspension cells induces cells to undergo PCD (Gao and Showalter, 1999). The involvement of AGPs in xylem development and differentiation is marking the initiation of secondary cell wall thickenings that are caused by the deposition of cellulose microfibrils and are cross linked by hemicelluloses and pectins that embedded in a gel-like matrix.

2.3 Arabinogalactan proteins (AGPs) degrading enzyme

The carbohydrate moieties of arabinogalactan-proteins (AGPs) which are mainly composed of β-1,3-galactosyl (Gal), L-Arabinose (L-Ara), glucuronic acid (GlcA) and 4-methyl-glucuronic acid (4-Me-GlcA) residues are essential for the physiological functions of these proteoglycans in higher plants. L-Arabinose and lesser amounts of other auxiliary sugars such as glucuronic acid, 4-methyl-glucuronic acid, L-rhamnose and L-fucose are attached to the side chains usually at non-reducing terminals. It is important to study carbohydrate degrading enzymes of AGPs because hydrolytic enzymes specific to particular sugar residues and type of glycosidic linkage provide useful tools for structural analysis of the sugar moieties of AGPs (Konishi *et al*., 2008).

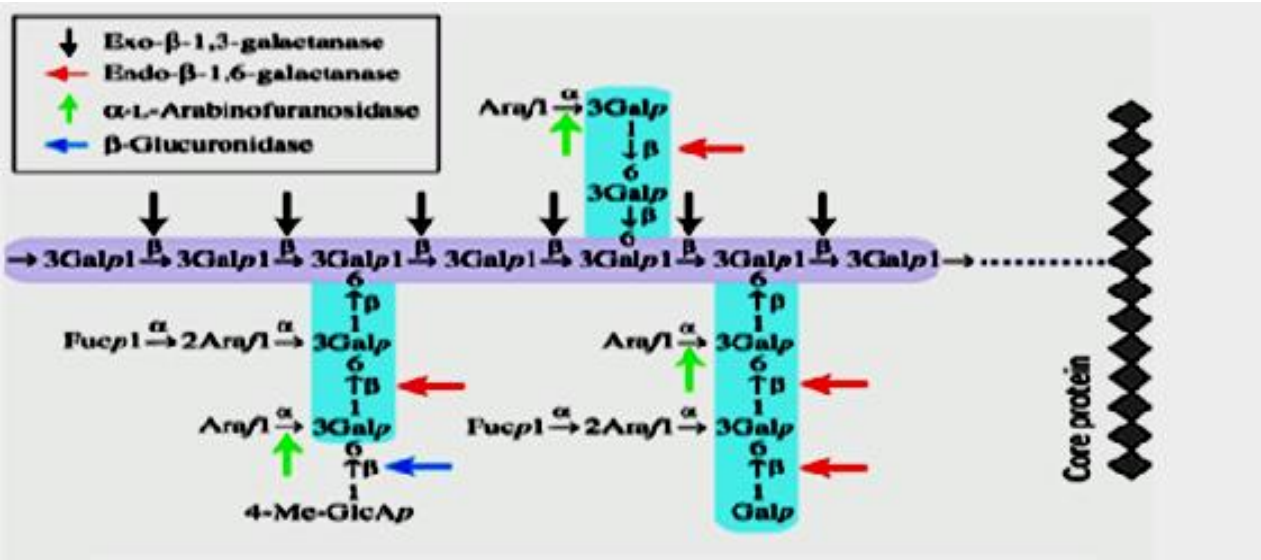

Figure 2.2 Enzyme that act on carbohydrate moieties of *AGP*s. (Adapted from: Konishi *et al*., 2008).

The several hydrolytic enzymes such as exo-β-1,3-galactanase, endo-β-1,6-galactanase, β-galactosidase, α-L-arabinofuranosidase and β-glucuronidase. Two genes encoding family beta-glucuronidases are designated as AnGlcAase and NcGlcAase based on the amino acid sequence. Although the deduced protein sequences of AnGlcAase and NcGlcAase are highly similar, the recombinant AnGlcAase (rAnGlcAase) exhibit distinct

substrate specificity toward 4-Me-GlcA residues of AGPs. Enzyme rAnGlcAase was able to catalyze the transglycosylation of GlcA residues from p-nitrophenyl (PNP) beta-GlcA to various monosaccharide acceptors. Both AnGlcAase and NcGlcAase are instances of a novel type of beta-glucuronidase with the capacity to hydrolyze beta-GlcA and 4-Me-beta-GlcA residues of AGPs (Konishi *et al*., 2008).

2.4 *Beta tubulin (β-TUB)* gene

Tubulin is one of the small family of globular proteins and its molecular weight is approximately 55 kilo Daltons. The common family of tubulin are α and β tubulin which are essential as a building block of microtubules (Starr, 2003). In plant cells, the cytoskeleton comprises distinct and highly dynamic arrays of microtubules and actin microfilaments. Katanin is a protein complex at β tubulin subunits for rapid microtubule transport in higher plant. In the structural unit of microtubules, the alpha and beta tubulin heterodimer are homologous. Each monomer structure can be divided into three functional domains that include the amino-terminal domain containing the nucleotide-binding region, an intermediate domain consists of taxol-binding site and the carboxy- terminal domain constitutes the binding surface for motor proteins which are kinesins and dyneins (Nogales et al., 1998). Free α and β tubulin subunits bind to GTP (guanosine triphosphate) at the plus end of a growing microtubule and forming GTP cap. After incorporated in the microtubule, the GTP will be hydrolyzed into GDP (guanosine diphosphate). This indicates that *β-tubulin* influence the stability of the dimer in the microtubules.

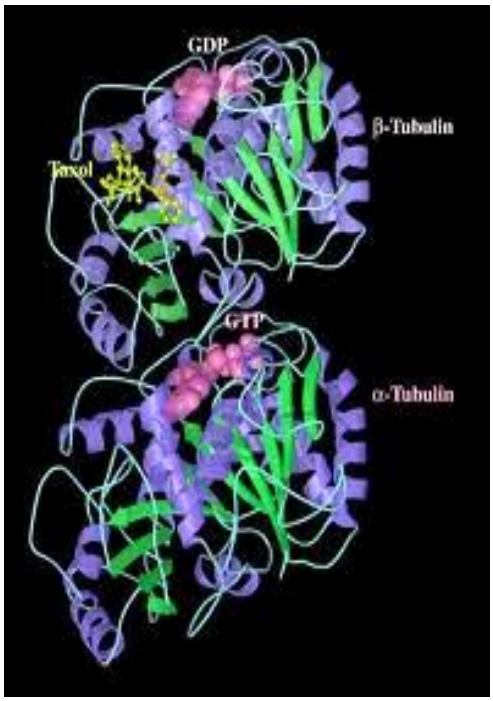

Figure 2.3 Structural of *alpha* and *beta tubulin* gene. (Adapted from source: http://www.snaggledworks.com/em_for_dummies/about.html)

Microtubules are straight, hollow cylinders which wall is made up of a ring of 13 "protofilaments". Microtubules can find in both animal and plant cells. In plant cells, microtubules are created at many sites scattered through the cell. In animal cells, the microtubules originate at the centrosome. Microtubules in cytoskeleton help to provide the plant cell with shape, structure and cell movement. Beta tubulin is essential as a building block of microtubules can direct orientation of cellulose microfibrils which is important in secondary cell wall of woody plants. So, beta tubulin gene is involved in determining the orientation of cellulose microfibrils and thus affects the cell wall formation (Spokevicius *et al.*, 2007). Cellulose synthesis peripheral proteins associated with the inner monolayer of the cell membrane are moved along the under-lying microtubule and extrude cellulose fibrils toward the outside through the outer monolayer of the cell membrane and cause the parallel orientation of the cellulose and microtubules.

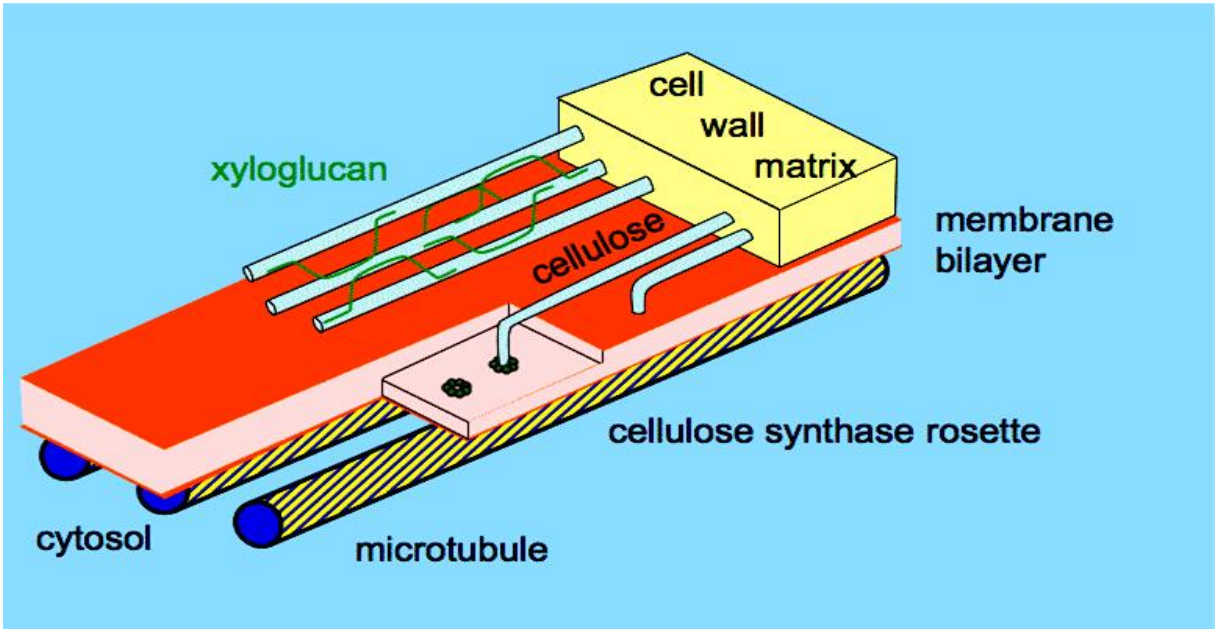

Figure 2.4 Cellulose synthesis is aligned with microtubule. (Adapted from source: http://plantphys.info/plant_physiology/cytoskeleton.shtml)

2.5 Single nucleotide polymorphisms (SNP)

Single nucleotide polymorphisms (SNPs) are single base differences in DNA sequence among individuals of species. SNPs can be used as DNA markers due to display have high information content and depict extremely high level of polymorphism. Besides that, SNP detection technologies have evolved from labor intensive, time consuming and expensive processes to the most highly automated, efficient and relatively inexpensive methods. SNPs detection consist two broad areas which are scanning DNA sequences for previously unknown polymorphisms and screening genotyping of individuals for known polymorphisms (Tabassum and Suman, 2006).

2.6 Application of single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are the most abundant form of genetic variation in many organisms. As a new marker system, SNP has been applied in genetic and breeding studies in some tree species and the related genetic information such as nucleotide diversity and linkage disequilibrium (LD). Unlike humans, the linkage disequilibrium (LD) rapidly decays within candidate genes in forest trees. Thus, SNPs-based candidate gene association studies are considered to dissect the complex quantitative traits in forest trees. The present study demonstrates that LD mapping can be used to identify alleles associated with quantitative traits and useful for performing breeding programs in forest trees.

The concept of Linkage Disequilibrium (LD) is associated with the use of SNPs. The non-random segregation of SNP alleles at different loci is referred to LD. LD is a complex phenomenon and is great interest to population geneticists. The existence of LD enables an allele of one polymorphic marker to be used as surrogate for a specific allele. LD attempts to correlate SNP patterns with phenotypes to directly associate SNPs acting as markers. For example, the elite germplasm of some crops have been subjected to bottleneck relatively and increasing the amount of linkage disequilibrium (LD) present and facilitating the association of SNP haplotypes at candidate gene loci with phenotypes. Whole-genome scans help to identify genome regions that are associated with interesting phenotypes when sufficient LD is present (Goldstein, 2001).

Quantitative trait loci (QTL) mapping is used for identifying the genetic basis of quantitative traits. QTL are regions of DNA that are closely linked within the genes involved in specifying heritable quantitative traits. Dissecting natural genetic variation requires a quantitative trait locus (QTL) analysis and then followed by identification of the particular gene and the type of polymorphism underlying QTL. Therefore, association

genetic studies via SNP discovery by sequencing can identify variation to the single-nucleotide substitutions that are responsible for quantitative trait nucleotides (Koornneef *et al*., 2003).

SECTION III

MATERIALS AND METHOD

3.1 Collection of Plant Material and DNA samples

The leaf samples of eleven *N. cadamba* were collected from Landeh Natural Reserve, Semongok, Sarawak.

3.2 DNA Extration

3.2.1 Chemical and Reagents

Liquid nitrogen, CTAB extraction buffer contains 100 mM Tris HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1.4 M NaCl, 2% Cetyl trimethyl ammonium bromide (CTAB), 1% PVP (Polyvinylpyrrolidone); 2% (v/v) β-mercaptoethanol, chloroform/isoamyl alcohol (24:1 v/v), isopropanol and 70% ethanol.

3.2.2 Total Genomic DNA Isolation protocol

The total genomic DNA of *N. cadamba* was isolated with a modified CTAB method based on Doyle and Doyle (1990). Six mililitres of extraction buffer and 120 µl of β-mercaptoethanol were added into a 50 ml Falcon tube and preheated for 30 minutes at 65˚C water bath. The leaf sample was rinsed with ddH_2O and wiped with 70% ethanol. Then, it was cut into small pieces. Vines of the leaf was removed and then put into the pre-chilled with liquid nitrogen before grinding. Then, liquid nitrogen was added and ground into fine powder. The fine powder was transferred into a 50 ml Falcon tube that containing preheated extraction buffer and β-mercaptoethanol. Then, the falcon tube was incubated for 40 minutes with shaking to mix the content. Then, 750 µl of the mixture was transferred into each of 1.5 ml microcentrifuge tubes with the cut tip pipette. Next, an equal volume of Chloroform-Isoamyl alcohol (CIA) (24:1) was added. The tubes were