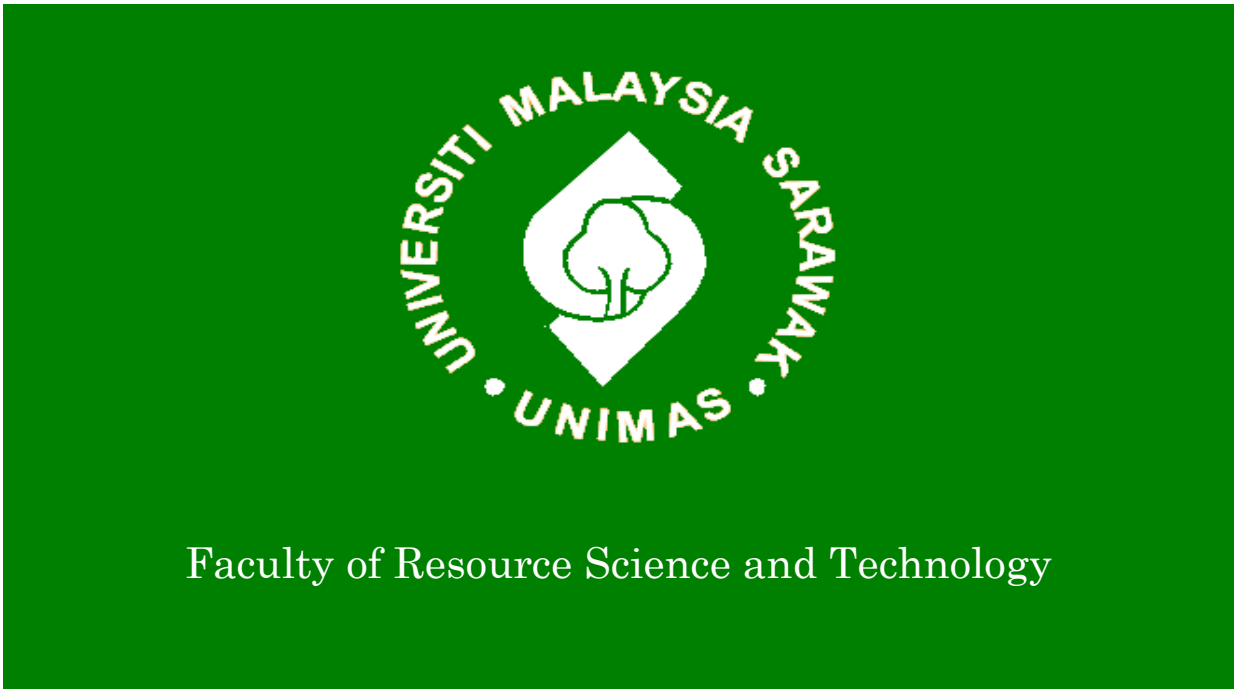

SEQUENCE POLYMORPHISM OF HEAT SHOCK PROTEIN 70 (*HSP70*) GENE IN KELAMPAYAN (*Neolamarckia cadamba)*

Choo Hui Yik

Bachelor of Science with Honours

(Resource Biotechnology)

2011

SEQUENCE POLYMORPHISM OF HEAT SHOCK PROTEIN 70 (*HSP70*) GENE IN KELAMPAYAN (*Neolamarckia cadamba)*

CHOO HUI YIK

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

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2011

DECLARATION

I declare that this thesis entitled "Sequence Polymorphism of Heat Shock Protein 70 (*HSP70*) in Kelampayan (*Neolamarckia cadamba)*" is the result of my own research paper except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ……..

Name : …….

Date : ……..

ACKNOWLEDGEMENTS

Above all, I would like to thank God for His blessings and guidance in the completion of this project. My sincere gratitude goes to my supervisor, Dr. Ho Wei Seng for giving me an opportunity to work on this project under his guidance and encouragement. Besides, I would also like to dedicate my appreciation to my co-supervisor Dr. Pang Shek Ling who has given me a lot of precious opinions in this project. My special thanks go to Mr. Liew Kit Siong (Msc. student), Miss. Tchin Boon Ling (Msc. student) and all my beloved lab members for having provided a most congenial and supportive atmosphere throughout my project. Finally, not forgetting my family and friends for the support and encouragement throughout the completion of my project, I thank them all.

TABLE OF CONTENTS

		Page
DECLARATION		i
ACKNOWLEDGEMENTS		ii
TABLE OF CONTENTS		iii
LIST OF TABLES		vi
LIST OF FIGURES		vii
LIST OF ABBREVIATIONS		viii
ABSTRACT/ ABSTRAK		x
CHAPTER I	**INTRODUCTION**	1
CHAPTER II	**LITERATURE REVIEW**	4
2.1	*Neolamarckia cadamba*	4
2.2	Plant responses to environmental stresses	5
2.3	Heat shock protein 70 (*Hsp70*) gene	6
2.4	Structure and interactions of heat shock protein 70 (Hsp70)	8
2.5	Single nucleotide polymorphisms (SNPs)	10
2.6	Applications of SNPs in genome analysis	11
CHAPTER III	**MATERIALS AND METHODS**	13
3.1	Collection of plant materials	13
3.2	DNA isolation and purification	13
	3.2.1 Chemicals and reagents	13
	3.2.2 Modified DNA isolation protocol	13
	3.2.3 DNA purification	14
3.3	Agarose gel electrophoresis	15
3.4	DNA quantification	16
	3.4.1 Spectrophotometric quantification of DNA	16

	3.4.2 Agarose gel electrophoresis quantification of DNA	16
3.5	Primer design	17
3.6	Polymerase chain reaction (PCR)	18
3.7	Purification of PCR products from agarose gel	18
3.8	Cloning of PCR products	19
	3.8.1 DNA ligation	19
	3.8.2 Transformation	19
	3.8.3 Blue/White colony screening	20
3.9	Plasmid isolation	20
3.10	DNA sequencing and data analysis	21
CHAPTER IV	**RESULTS AND DISCUSSION**	23
4.1	Plant materials	23
4.2	DNA isolation and purification	24
4.3	DNA quantification	26
4.4	Primer design	29
4.5	Polymerase chain reaction (PCR)	30
4.6	PCR optimization	31
4.7	PCR product purification	32
4.8	Cloning of PCR product	33
	4.8.1 Confirmation of desired insert of colony PCR	34
4.9	Plasmid isolation and purification	35
4.10	DNA sequencing and data Analysis	37
	4.10.1 Evaluation of sequence variation in *Hsp70* gene among the *N. cadamba* trees	37
	4.10.2 Determination of nonsynonymous and synonymous mutations caused by single nucleotide variation in *Hsp70* gene among the *N. cadamba* trees	43

4.10.3 Identification of possible restriction enzymes for SNP site 45

CHAPTER V CONCLUSIONS AND RECOMMENDATIONS 46

REFERENCES 48

APPENDICES 52

Appendix A 52

Appendix B 53

LIST OF TABLES

Table		Page
3.1	Mass of λ *Hin*d III DNA in a selected fragment	17
3.2	Ligation reaction mixture and volume	19
4.1	Plot of the collected *N. cadamba* trees	23
4.2	Characteristics of collected *N. cadamba* trees	24
4.3	Estimated DNA concentration and DNA purity of the twelve purified DNA samples	27
4.4	Estimated purified DNA concentration of *N. cadamba* trees based on band intensity of samples and marker on 0.8% (w/v) agarose gel	28
4.5	Primers designed to amplify the *Hsp70* gene	30
4.6	Optimized thermal cycling profile	32
4.7	BLASTn output for amplified *Hsp70* DNA of 527 bp	37
4.8	Nonsynonymous mutations in partial sequences of *Hsp70*	44

LIST OF FIGURES

Figure		Page
2.1	*Neolamarckia cadamba*	5
2.2	Hsp70 interactions	10
4.1	Gel electrophoresis of genomic DNA samples on a 0.8% (w/v) agarose gel	25
4.2	Gel electrophoresis of purified *N. cadamba* genomic DNA samples on a 0.8% (w/v) agarose gel	26
4.3	Annealing temperature (T_a) optimization for the designed primer	31
4.4	Gel electrophoresis of optimized PCR product on a 1.5% (w/v) agarose gel	32
4.5	Agarose gel electrophoresis of PCR product of *Hsp70* amplified by the designed primer	33
4.6	LB agar plate showing the result of a blue white screening for desired insert DNA of *Hsp70* gene from *N. cadamba*.	35
4.7	Agarose gel electrophoresis of colony PCR products of the *Hsp70* amplified by using primers M13 forward and reverse sequence primers.	35
4.8	Agarose electrophoresis of purified recombinant DNA plasmid of *Hsp70* clone of *N. cadamba*	36
4.9	Alignment of nucleotide sequence for *Hsp70* DNA samples using CLC Free Workbench 4	41
4.10	Alignment of amino acid sequence for *Hsp70* DNA samples using CLC Free Workbench 4	44

LIST OF ABBREVIATIONS

ATP	Adenosine triphosphate
BiP	Hsp70 resident protein in the endoplasmic reticulum
bp	Base pair
CAPS	Cleaved-amplified polymorphic sequence
cp	Chloroplast
dbh	Diameter of breast height
DNA	Deoxyribonucleic acid
Dnak	Prokaryotic heat shock protein 70
EcoRV	Type II restriction endonuclease isolated from certain strains of *Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
GAS	Gene assisted selection
Hsc70	Heat shock protein cognate 70
HSE	Heat shock element
HSF	Heat shock factor
Hsp40	Heat shock protein 40
Hsp70	Heat shock protein 70
Hsp110	Heat shock protein 110
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
MAS	Marker-assisted selection
mt	Mitochondrial
NBD	N-terminal nucleotide binding domain
NEFs	Nucleotide exchange factors

ORF	Open reading frame
PCR	Polymerase Chain Reaction
QTN	Quantitative trait nucleotide
SBD	C-terminal substrate binding domain
SNPs	Single nucleotide polymorphisms
SSE	Subfamily of (heat and osmotic) stress proteins
Tris HCL	Tris(hydroxymethyl)aminomethane Hydrochloride
T_m	Melting temperature

SEQUENCE POLYMORPHISM OF HEAT SHOCK PROTEIN 70 (*HSP70*) GENE IN KELAMPAYAN (*Neolamarckia cadamba)*

CHOO HUI YIK

Resource Biotechnology
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Neolamarckia cadamba or locally known as kelampayan, is one of the valuable timber tree species. In forest industry, genetic improvement programs based on Single Nucleotide Polymorphisms (SNPs) at the molecular level has become an important tool in the selection of germplasm with desired traits. In this study, DNA sequence of *Hsp70* which is gene responsible for heat stress was amplified by polymerase chain reaction (PCR) technique with the specific designed primers based on the cDNA sequence of *Hsp70* in *N. cadamba*. The 527 bp *Hsp70* amplicons were subjected to BLASTn analysis to perform the sequence homology search through all known template sequence available in the NCBI. Each sequence was then aligned using CLC Free Workbench 4 software for manual detection of SNPs. A total of ninety SNPs were detected only in the exons of the *Hsp70* gene with 12 non-synonymous mutations and 78 synonymous mutations. The sequence of the twelve Hsp70 DNA samples was then subjected to *in silico* restriction analysis. The exclusiveness of the restriction enzymes *BssS1*, *BstBI*, *HpyCH4IV*, *SacI*, *BsrGI*, and *Bfa1* obtained for the SNPs at nucleotide 113, 124, 217, 293, 341, and 463 respectively could be useful for genetic markers development.
Key words: *Neolamarckia cadamba*, single nucleotide polymorphism (SNP), heat shock protein 70 (*Hsp70)* gene, heat stress, polymerase chain reaction (PCR).

ABSTRAK

Neolamarckia cadamba atau dikenali sebagai kelampayan merupakan salah satu spesies pokok kayu yang berharga. Dalam industri perhutanan, program pembiakbaikan genetik berdasarkan polimorfisme nukleotida tunggal (SNP) pada peringkat molekul telah menjadi satu alat penting dalam pemilihan plasma germa dengan ciri-ciri yang dikehendaki. Dalam kajian ini, jujukan DNA bagi gen 'heat shock protein 70' (Hsp70) yang dikenali sebagai gen yang bertanggungjawab atas tekanan haba telah diamplifikasi dengan menggunakan teknik tindakbalas berantai polimerase (PCR) bersama dengan pencetus yang direka khas berdasarkan jujukan cDNA daripada Hsp70 dalam N. cadamba. Produk PCR bersaiz 527 bp dianalisa dengan BLASTn untuk mencari urutan homologi dalam pangkalan data NCBI. Seterusnya, setiap urutan dijajarkan dengan menggunakan CLC Free Workbench 4.0 bagi pengesanan polimorfisme nukleotida tunggal (SNP) secara manual. Sebanyak sembilan puluh SNPs telah dikesan di bahagian ekson sahaja bagi gen Hsp70 dengan 12 non-sinonim mutasi and 78 sinonim mutasi. Jujukan bagi dua belas Hsp70 DNA sampel dianalisa dengan menggunakan enzim penyekatan secara in silico. Pengkhususan enzim penyekatan BssS1, BstBI, HpyCH4IV, SacI, BsrGI, dan Bfa1 bagi SNPs pada nukleotida 113, 124, 217, 293, 341, dan 463masing-masing adalah amat berguna bagi penghasilan penanda molekul.
Kata kunci: Neolamarckia cadamba, polimorfisme nukleotida tunggal (SNPs), gen heat shock protein 70 (Hsp70), tekanan haba, tindakbalas berantai polimerase (PCR).

CHAPTER I

INTRODUCTION

Forest trees as vital component of biodiversity represent the main source of terrestrial biomass production. However, forest ecosystems have been under intense pressure from commercial harvest and mainly for supplying the timber. In the 1960′s and 1970′s, there was enormous increase in demand for tropical hardwoods in Malaysia due to a decline in availability of wood from logged-out temperate forests. Hence, Malaysia's forests began to fall rapidly. Logging concessions exceeding the total estimated forested area have been seen in Peninsular Malaysia (Rainforest Conservation Fund, 2011).

In Borneo, most of the deforestation has been carried out for the purpose of timber extraction. By 1989, 60% of the land had been licensed for timber logging and this area supplied almost one-third of the world's hardwood timber (Rainforest Conservation Fund, 2011). This results in drop of proportion as these slow growing trees are unable to meet current global demand for wood, resulting in the loss and degradation of forest. Recently, plantation forests have become increasingly important in order to rehabilitate the degraded areas to productive forests. Fast growing species in the plantation forests such as *Neolamarckia cadamba* have the capability to supply the bulk of wood needs on a long-term basis thus reduce the harvest pressure on natural forest for wood production.

Neolamarckia cadamba or locally known as kelampayan is one of the fast growing species for planted forest development in Sarawak. Kelampayan, under the family of Rubiaceae is characterized as a large, deciduous, and evergreen endemic tree species that provides early economic returns within 8 to 10 years. Kelampayan is one of the light-

colored timbers with low density which serves as raw material for pulp and paper industry and it is one of the best materials for plywood industry. The timber can be processed as picture frame, moulding, skirting, wooden sandals, disposable chopstick, general utility furniture, veneer and plywood (Lim *et al*., 2005). Besides that, it is also suitable for use as a shade tree for dipterocarps line planting. Leaves and bark of this plant are reported to possess various medicinal values (Kapil *et al*., 1995).

Plants as sessile organisms are often exposed to different environmental conditions. Plants subjected to a number of external stresses that adversely affect their growth and development. Abiotic stresses such as drought, salinity, extreme temperatures, chemical toxicity and oxidative stress are serious threats to agriculture and are the primary cause of crop loss worldwide (Wang *et al.*, 2003). Plant productivity is limited more by temperature than other environmental factor. Rapid temperature changes, particularly toward the upper limit of the adaptation range for plant species can produce dramatic changes in the gene expression. Severe heat stresses leads to cellular damage and cell death. However, sublethal doses of heat stress induce heat-shock response and express Hsp70 that mediate protection against environmental stresses.

The 70 kilodalton heat shock proteins is a member of a multigene family which expressed in response to environmental stress conditions such as heat, cold and drought, as well as to chemical and other stresses. Hsps70 chaperones have high affinity to bind ATP and are important cellular machinery that assists a wide range of protein folding processes in almost all cellular compartments (*Wang et al.*, 2004). Hsps70 also have essential functions in folding of non-native proteins and stress-associated gene expression.

In forest industry, selective breeding has become an important tool to select plant that possess beneficial trait, continuously breeding with other species to produce a hybrid, and results in increase plant productivity and enhancement of wood quality. However, conventional plant breeding is costly and time consuming as it can only be done on mature trees to observe their phenotypic characteristics. Therefore, selection based on Single Nucleotide Polymorphisms (SNPs) at the molecular level can become a useful tool in the selection of germplasm with desired traits in a relatively short period of time assisted by marker-assisted selection (MAS). These markers are highly abundant in the genomes of plants and thus potentially useful for identifying the desired genotypes to plant stress tolerance.

Currently, there is no study has been done on the sequence variation of *Hsp70* gene from *N. cadamba*. Hence, a study with the objective of determines the feasibility of finding single nucleotide polymorphisms (SNPs) from heat shock protein 70 (*Hsp70*) gene in *N. cadamba* was conducted. Primer is designed for amplifying *Hsp70* gene from *N. cadamba* and genotype across the *Hsp70* samples was compared. Presence of single base differences in the gene among the individuals of the same species can cause non-synonymous or synonymous mutations which will result in the changes of plant stress tolerance.

CHAPTER II

LITERATURE REVIEW

2.1 *Neolamarckia cadamba*

Neolamarckia cadamba belongs to family Rubiaceae is commonly known as kelampayan. Kelampayan is distributed in lowland to mountain forests up to 1000 m altitude, often by streams and rivers and in open sites in the forest. *N. cadamba* is found more commonly in the Asia-Temperate like China and Asia-Tropical country such as India, Nepal, Sri Lanka, Cambodia, Laos, Myanmar, Thailand, Vietnam, Indonesia, Malaysia and Papua New Guinea but also on Cape York Peninsula. It is a large, evergreen endemic tree up to 37.5 m high and 2.4 m in girth with straight cylindrical bole. The bark is gray, leaves opposite, simple, elliptic-oblong, flowers in solitary globose head, orange or yellow, and fruits pseudo carps (Acharyya *et al*., 2010).

N. cadamba is an ornamental plant that is one of the best raw materials for plywood industry. The barks and leaves are reported to possess various medicinal values such as astringent anti-hepatotoxic (Kapil *et al*., 1995), antidiuretic, antiseptic, wound healing and anthelmintic (Gunasekharan and Divyakant, 2006). Its stem has been used for the treatment of diabetes by tribal people and decoction of this stem bark given for diarrhea. The phytochemical screening of *N. cadamba* stem bark revealed the presence of flavonoids, phenolic acids, sterols, alkaloids, tannins and anthocyanins that are known to be bioactive antidiabetic principles (Bussa and Pinnapareddy, 2010). This reveals the importance of stem bark of *N. cadamba* as an economical antidiabetic agent. Figure 2.1 shows the structures of *Neolamarckia cadamba* tree's bark, leaves and flowers.

(a) **(b)**

Figure 2.1 *Neolamarckia cadamba*. (a) Tree's bark (Adapted from http://picasaweb.google.com/manglayang /HanjaJabonSamama#5488813820177245106). (b) Leaves and flowers structures (Adapted from http://www. somewhereinblog.net/blog/mohdfiendblog/28838296/).

2.2 Plant responses to environmental stresses

Abiotic stresses are serious threats to agriculture and the natural status of the environment, leads to a series of morphological, physiological, biochemical and molecular changes that adversely affect plant growth and productivity (Wang *et al.*, 2001). Some abiotic stresses are often interconnected and may cause similar cellular damage. For example, oxidative stress such as high temperature, salinity, or drought stress, may cause denaturation of functional and structural proteins (Smirnoff, 1998). Therefore, these diverse environmental stresses often activate similar cell signaling pathways (Zhu, 2002) and cellular responses, such as the production of stress proteins, up-regulation of anti-oxidants and accumulation of compatible solutes (Cushman and Bohnert, 2000).

Heat stress is a major abiotic stress limiting plant growth and productivity. Each plant has its optimum temperature at which each plant grows most efficiently, with upper and lower limits. For example, optimum temperature for maize (*Zea mays*) is 30-35 ℃ and as temperature approaches these limits, growth diminishes (Hopkins and Huner, 2004).

Molecular control mechanisms for heat stress tolerance are based on the activation and regulation of specific stress-related genes which are involved in the whole sequence of stress responses (Wang *et al.,* 2001). Plant adaptation to heat stress may be affected by changes in the level and expression pattern of some proteins. Heat stress may induce protein metabolism involving protein synthesis and degradation which is one of the most sensitive processes to heat stress (Huang and Xu, 2008). The induction of heat-responsive proteins, particularly heat shock proteins (HSPs), plays a key role in plant tolerance to heat stress. Hence, plants must adapt to stress conditions and exercise specific tolerance mechanisms to maintain growth and productivity.

2.3 Heat shock protein 70 (*Hsp70*) gene

High temperature leads to denaturation of proteins, thus causing their loss of cellular functions in plant physiological process and eventual death of plant. The 70 kDa heat shock proteins is a member of a multigene family which express heat shock protein ubiquitously under stress conditions. Heat shock proteins 70 (Hsp70s) are molecular chaperones found in the cytosol and in other compartments of the cell that play an essential role in the life cycle of proteins under both normal and stressful conditions. Hsp70s assist a wide range of folding processes, including in folding and assists newly synthesized polypeptides to obtain correct conformations and assisting stress denatured proteins to assume their original folded conformations (Dong, 2001). Hsp70s also play a role in folding of non-native proteins that include prevention of aggregation, promote folding to the native state and assist refolding of aggregated protein (Mayer and Bukau, 2005). In addition, they have housekeeping functions in transport of proteins between cellular compartments, degradation of unstable and misfolded proteins, prevention and dissolution of protein complexes, uncoating of clathrincoated vesicles, and control of regulatory

proteins. All of these activities happened when Hsp70s interact with hydrophobic peptide segments of proteins in an ATP-controlled fashion (Mayer and Bukau, 2005).

Hsp70s have essential function in stress-associated gene expression. The interaction between Hsp70 and heat shock factor (HSF) is a negative regulatory mechanism for HSF-mediated transcriptional activation in the heat shock response (Morimoto, 1998). Kim and Schoffl (2002) suggested that the interaction between Hsp70 and HSF prevents trimerization and the binding of HSF to heat shock element (HSE), thereby blocking the transcriptional activation of heat-shock genes by their HSFs. However, interaction of Hsp70 with HSF does not correspond with the activation level of HSF. The overexpression of *Hsp70* genes correlates positively with the acquisition of thermotolerance (Lee and Schoffl, 1996) and results in enhanced tolerance to salt, water and high-temperature stress in plants (Sung and Guy, 2003).

Hsp70 is the major protein synthesized during heat-shock response. The heat-shock response is a programmed change in gene expression carried out by cells in response to environmental stress, such as heat. Heat-shock response has essential function in protects cells and organisms from severe damage, allows resumption of normal cellular and physiological activities, and leads to a higher level of thermotolerance response (Schoffl *et al*., 1998). Deletion alleles of the *Hsp70* genes in *Drosophila melanogaster* have been generated and indicate flies with *Hsp70* deletions have reduced thermotolerance. Although the synthesis of Hsp70 is undetectable in *Drosophila* cells at the normal growth temperature of 25℃, its expression is rapidly induced at least 1000-fold at the temperature of 37℃ (Brody, 2009). The prominent expression of Hsp70 suggested that it plays an essential role in thermotolerance.

Arabidopsis genome contains at least 18 genes encoding members of the *Hsp70* family. Most *Arabidopsis Hsp70s* reached peak induction within 30 minutes of heat shock exposure whereas induction by low temperature condition was limited to cytosolic and mitochondrial members of *Hsp70s* (Dong *et al.*, 2001). The temperature responses suggest that cytosolic and endoplasmic reticulum of *Hsp70* genes are responsible for molecular chaperone activity under heat stress, and mainly cytosolic *Hsp70s* are required under low temperature stress in *Arabidopsis* (Dong *et al.*, 2001).

2.4 Structure and interactions of heat shock protein 70 (Hsp70)

Structurally, Hsp70 composes of a highly conserved N-terminal nucleotide binding domain (NBD) of 45 kDa and a C-terminal substrate binding domain (SBD) of 25 kDa which is further subdivided into a *b*-sandwich subdomain of 15 kDa and a C-terminal *a*-helical subdomain (Mayer and Bukau, 2005). NBD consists of an actin-like fold with two globular subdomains separated by a nucleotide binding cleft whereas the SBD has a β sandwich that forms the substrate binding pocket, with an α-helix packed against the sandwich from one side (helix A) and a helical lid (helix B) closing on top of the substrate binding pocket (Bukau *et al.*, 2006). This structure reveals a flexible linker of 10 residues provide interdomain interaction that connects the NBD and SBD (Figure 2.2). Signal transduction from the catalytic centre of the NBD to the interdomain interface is mediated by hydrogen bond network with key residues being E175 as nucleotide sensor, P147 as structural switch, and R155 as surface-exposed relay (Vogel *et al*., 2006).

Under cellular stress conditions, the Hsp70 are diverted to chaperoning denatured proteins, dissociate HSF to activate transcription of the *Hsp70 g*enes (Figure 2.2). Cellular mechanism begin when peptide stretches exposed in client proteins are bound by the co-

chaperone Hsp40, which delivers it to substrate binding domain of Hsp70 in an ATP bound state. The J-domain of Hsp40 then triggers ATP hydrolysis regulated by nucleotide exchange factors (NEFs) and locking substrate into SDB tightly and promotes its folding and prevents aggregation (Fan, 2003; Qiu, 2006). NEFs are critical for the functional cycle of Hsp70s because they promote the release of ADP and rebinding of ATP that triggers unloading of bound substrate and to enable nucleotide exchange (Bukau *et al.*, 2006). Successive cycles of substrate binding and release are coupled to the intrinsic ATPase activity of Hsp70, which requires Hsp70 co-chaperones such as Hsp40 (Bukau and Horwich, 1998).

Activity of client proteins that is controlled through transient association with Hsp70 includes regulatory proteins such as nuclear receptors, kinases, and transcription factors. This interactions cause the ability of Hsp70 chaperones to regulate important physiological processes such as cell cycle, cell differentiation or programmed cell death, and aging (Mayer, 2005; Cobreros, 2008). While under non-stress conditions, Hsp70 bind to HSF in the ATP state and inactivate transcription of the *Hsp70 g*enes. ATP binding to the NBD induces conformational changes in SBD, which opens the substrate binding pocket and its helical lid with low affinity and fast exchange rates for substrates (Bukau *et al.*, 2006).

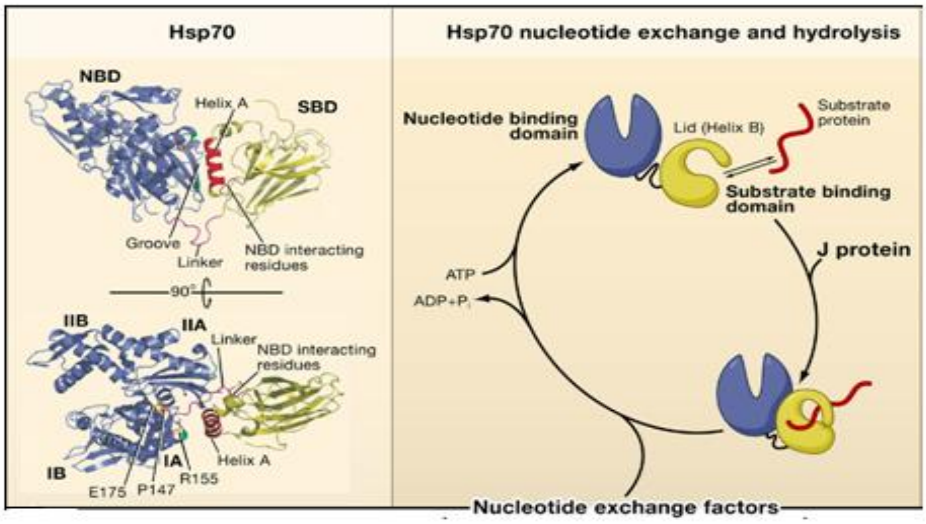

Figure 2.2 Hsp70 interactions. (a) Communication between the nucleotide binding domain (NBD, blue) and the substrate binding domain (SBD, yellow) of an Hsp70, as indicated from the crystal structure of bovine Hsc70 (Adapted from Jiang *et al*., 2005). (b) The features of the Hsp70 cycle. NBD is shown in blue, SBD in yellow, and substrate protein in red (Adapted from Bukau *et al*., 2006).

2.5 Single nucleotide polymorphisms (SNPs)

Single-nucleotide polymorphisms (SNPs) are sequences in the genome of an organism that represents a difference in a single nucleotide between individuals of the same species. SNP is a valuable marker in the studies of the agronomic or adaptive traits in plant species. SNP markers are used in genetic mapping and association genetics studies such as genetic diagnostics, genetic structure of population analysis, phylogenetic analysis and others (Rafalski, 2002).

SNPs can occur in coding, intergenic, and non-coding region of the DNA sequences. However, most of the SNPs located in the non-coding region. SNPs that present in the coding sequence will show a close association with the trait although they may or may not result in the mutant phenotype (Gupta *et al*., 2001). Hence, it is very useful for marker-assisted selection (MAS) and for gene isolation. Protein-coding SNPs that occur in the regulatory regions or exons can alter the function and structure of encoded proteins and

is known as non-synonymous mutation. While SNPs that found in non-coding regions or introns will not change the amino acids composition of the encoded protein due to the degenerative of the genetic code and is known as synonymous mutation.

2.6 Applications of SNPs in genome analysis

Single nucleotide polymorphisms (SNPs) markers have high frequency of occurrence in genomes of the majority of the organisms, including plants. This is proved in a preliminary studies conducted in wheat, one SNP per 20 bp and in the maize genome, one SNP per 70 bp have been recorded in certain regions of these genomes (Gupta *et al*., 2001). Therefore, SNPs is potential to use in the detection of associations between allelic forms of a gene and phenotypes, especially for common diseases that have multifactorial genetics (Jorde, 1995 and 2000).

The study of natural genetic variation in forest trees requires a quantitative trait loci analysis and followed by identification of the particular gene and type of polymorphism underlying QTL (Koornneef *et al.*, 2003). QTL are regions of DNA that are closely linked within the genes involved in identifying heritable quantitative traits. Quantitative inheritance refers to inheritance of a phenotypic characteristic that associated to interactions between two or more genes and their environment. SNPs-based candidate gene association studies are effective approach to identify the complex quantitative traits in forest trees. The linkage disequilibrium (LD) rapidly decays within genes in forest trees and LD mapping can be used to identify alleles associated with quantitative traits and potentially useful for performing breeding programs in forest trees (Zhang and Zhang, 2005).

In plants, most SNPs are not genetic determinants and therefore plants breeders are interest in determine the associations among SNPs and the traits of economic value (Gupta *et al*., 2001). For example, SNP marker for *waxy gene (Wx)* controlling amylase content in rice. Amylose is the principal element that determines the cooking and nutritional properties of cereals. Nevertheless, amylose-free or low amylose starch is desirable for certain food and non-food industries. Therefore, the breeding trait is vital for the development of new cultivars in rice. In most cereals, amylose synthesis is controlled by starch synthase enzyme which is encoded in waxy (*Wx*) gene. In rice, it has been shown that single nucleotide difference which located near the waxy gene can be used to differentiate the high and low amylase types (Gupta *et al*., 2001). Therefore, SNP is utilized in MAS exercised for the selection of low amylose at the seedling stage.

Single nucleotide polymorphism of the *Hsp70* genes of *Drosophila melanogaster* (five genes) and *D. simulans* (four genes) had been used to analyse and characterize the homogenizing and diversifying roles of gene conversion in their evolution. Sequence and organization of *Hsp70* genes in both flies had been examined, firstly, by characterizing nucleotide polymorphism in alleles of *Hsp70* genes to assess intercluster and intracluster divergence and second, the nature of both past and present concerted evolution at the *Hsp70* genes is examined (Bettencourt and Feder, 2002). This study revealed that gene conversion homogenizes the *Hsp70* coding regions in both *D. melanogaster* and *D. simulans* with intracluster are nearly identical, and large intercluster conversion tracts diminish divergence. Comparison of the *Hsp70* genes in both flies indicates rapid propagation of novel mutations among duplicate genes (Bettencourt and Feder, 2002). These results suggested that the homogenizing and diversifying roles of conversion interact to drive evolution of the *Hsp70* genes.