ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 11 (6): 307-317, 2012
ISSN 1682-296X / DOI: 10.3923/biotech.2012.307.317
© 2012 Asian Network for Scientific Information

Association Genetics of the *Cinnamyl Alcohol Dehydrogenase* (*CAD*) and *Cinnamate 4-hydroxylase* (*C4H*) Genes with Basic Wood Density in *Neolamarckia cadamba*

[1]B.L. Tchin, [1]W.S. Ho, [2]S.L. Pang and [1]J. Ismail
[1]Forest Genomics and Informatics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia
[2]Applied Forest Science and Industry Development, Sarawak Forestry Corporation, 93250 Kuching, Sarawak, Malaysia

Abstract: Association genetics study is a powerful approach to detect the potential genetic variants (i.e., SNPs) underlying the common and complex adaptive traits. Once the quantitative trait nucleotides are identified, such powerful approach provides significant advantages to the forest industry. Hence, attempts were made to discover SNPs from *Neolamarckia cadamba* partial *C4H* (3,538 bp) and *CAD* (2,354 bp) DNA sequences and further associate those SNPs with basic wood density. Overlapping primers were designed in flanking the partial *C4H* and *CAD* DNA from 12 *N. cadamba* trees. The amplified DNA fragments were sequenced and the basic wood density measurements were determined for each tree. The sequence variation analyses revealed that there were 60 and 32 SNPs detected in the partial *C4H* and *CAD* DNA sequences, respectively. Those SNPs were distributed throughout the exon, intron, 5'-UTR and 3'-UTR regions. The total nucleotide diversities were $\pi = 0.00302$ and $\theta_W = 0.00412$. The synonymous mutations ($\pi = 0.00983$; $\theta_W = 0.01210$) were more common than nonsynonymous mutations ($\pi = 0.00045$; $\theta_w = 0.00089$) for both *C4H* and *CAD* genes. LD declined linearly over short distance at the loci examined. Association genetics study also revealed that 4 and 6 SNPs from *C4H* and *CAD* genes, respectively were in significant associations with basic wood density of *N. cadamba* ($p<0.05$). The genetic variation identified by the SNP markers, once validated, will facilitate the selection of *N. cadamba* parental lines or seedlings with optimal quality through Gene-assisted Selection (GAS) approach.

Key words: *Neolamarckia cadamba*, *cinnamate 4-hydroxylase*, *cinnamyl alcohol dehydrogenase*, single nucleotide polymorphism, association genetics

INTRODUCTION

Neolamarckia cadamba, or locally known as Kelampayan is one of the Lesser-known Commercial Timbers (LKCT) which possess various benefits for wood-based industry. It is used for making picture frame, moulding, skirting, wooden sandals, disposable chopstick, general utility furniture, veneer, plywood as well as pulp and paper (Lim *et al*., 2005). The leaves, roots and bark of Kelampayan also have been reported to have high pharmacological values (Joker, 2000; Patel and Kumar, 2008; Acharyya *et al*., 2010). To date, studies on Kelampayan at molecular level are still limited. Recently, a Kelampayan tree transcriptome database (NcdbEST) had been developed (Ho *et al*., 2010). It provides useful genomics information and resources for researchers to deeply explore the genomics basic of the Kelampayan.

Association genetics or mapping study is an alternative approach that enables researcher to use modern genomic approaches to identify marker-trait associations at higher resolution compared to the widely used family-based Quantitative Trait Locus (QTL) mapping. It has emerged as a new tool in dissecting the genetic basis of complex traits such as physical and chemical wood property traits variation down to sequence level by harnessing the genetic variation at the population level. In just a few years, association genetics study has been broadly embraced in forest tree species, such as *Pinus*, *Pseudotsuga*, *Populus* and *Eucalyptus* (Neale and Kremer, 2011). Association genetics study is a powerful approach in identifying genes or loci that contribute to variation in complex traits (Gonzalez-Martinez *et al*., 2006; Hall *et al*., 2010). It has long been argued that the phenotypic variations among

Corresponding Author: Ho Wei Seng, Forest Genomics and Informatics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia Tel: 60 82-581000 Ext: 2975 Fax: 60 82-583160

individuals may due to the cumulative effect of a number of genes and/or the environmental influences (Bentz *et al*., 2011; Flatscher *et al*., 2012). The basic genetic architecture of such complex adaptive traits is very difficult to discover through traditional linkage-based approaches as it is laborious, time consuming and expensive due to the requirement of mapping population establishment (Myles *et al*., 2009). However, association genetics study is able to eliminate such drawbacks of traditional breeding method and give a higher mapping resolution (Ingvarsson and Street, 2011).

Single Nucleotide Polymorphism (SNP), where the sequences differ only in a single nucleotide, has become marker of choice for association genetics study due to the abundance, stable, ubiquity and interspersed characteristics in nuclear genome (Fusari *et al*., 2008). SNPs are less polymorphic than other genetic markers, e.g., Simple Sequence Repeats (SSRs), but still can provide more information regarding the genetic constituent of a living organism (Rafalski, 2002). SNPs are far easier to be detected since only one single base in specific sequences become target (Prince *et al*., 2001). Hayashi *et al*. (2004) also argued that SNP markers are more efficient, cost effective and suitable for the germplasm selection at early seedling stage due to the amount of genomic DNA required for detection is relatively low.

Candidate gene-based association mapping approach is most suitable for organism lacking a genome sequence or reference genomic resources (Nichols and Neale, 2010). Meanwhile, genes underlying quantitative traits that have been extensively studied in model organisms such as *cinnamyl alcohol dehydrogenase* (*CAD*) and *cinnamate 4-hydroxylase* (*C4H*) (Baucher *et al*., 2003) may provide the best candidates for association genetics study in non-model organisms (Fitzpatrick *et al*., 2005). The main function of *C4H* is to catalyze the hydroxylation of cinnamate to 4-coumarate at the early stage of lignin biosynthesis pathway while *CAD* catalyzes the reduction of cinnamaldehydes to ρ-coumaryl, coniferyl and sinapyl alcohols during the final stage of lignin biosynthesis pathway (Lewis, 1999).

Moreover, *C4H* and *CAD* genes are known to be correlated with some other phenotypic variations (specific gravity, wood density, microfibril angle, stem growth, etc.) in forest tree species rather than lignin production only (Yu *et al*., 2006; Gonzalez-Martinez *et al*., 2007; Tchin *et al*., 2011; Schilmiller *et al*., 2009; Bjurhager *et al*., 2010; Wegrzyn *et al*., 2010). Abreu *et al*. (2009) also proposed that the high β-O-4 (Alkyl Aryl Ether) bonds in lignins of angiosperms may possibly affect the wood properties. Therefore, further exploration on the *C4H* and *CAD* genes-traits association is crucial for Kelampayan tree breeding programme.

Wood density, a measure of wood mass relative to wood volume, is a convenient indicator of mechanical properties. The stiffness, bending strength and compression strength of wood are increased with density of wood (Van Gelder *et al*., 2006). Wood density will also determine the yield of pulp per unit volume and also some of the characteristic of the pulp (Bowyer and Haygreen, 1996). Moreover, the wood density is able to affect the growth rates (King *et al*., 2005), tree architecture (Lida *et al*., 2012), stem biomass, lifespan and maximum size of an individual plants (King *et al*., 2006), which eventually lead to the determination of function and structure of the forest (Suzuki, 1999).

The identification of gene-associated SNPs that correlated with Kelampayan wood density and subsequently applied to Gene-assisted Selection (GAS) will become an efficient and feasible tool for forest tree breeding programme (Neale and Savolainen, 2004; Boerjan, 2005; Lau *et al*., 2009). Gene-assisted selective breeding method in forest industry by using SNP markers are expected to increase the selection efficiency and reduce the time and cost associated with measuring the traits. Hence, the objectives of this study were to determine the SNPs from *CAD* and *C4H* genes and to identify the genetic association among SNPs from *CAD* and *C4H* genes with the basic wood density of Kelampayan.

MATERIALS AND METHODS

Plant materials: A total of 12 Kelampayan trees were selected randomly from the wild stands in Kota Samarahan, Sarawak, Malaysia in October 2011. The inner bark tissues were collected and DNA extraction was carried out by using a modified Doyle and Doyle (1990) protocol. The isolated DNA then was purified by using Wizard® Genomic DNA Purification Kit (Promega, USA) according to the manufacture's protocol.

PCR amplifications: Three and four overlapping primers which flanked the partial *CAD* and *C4H* genomic DNA sequences, respectively were designed based on the *NcC4H* (GenBank accession number: JQ946327) and *NcCAD* (GenBank accession number: JQ946326) cDNA by using Primer Premier 5 software (PREMIER Biosoft International, USA) (Table 1). PCR reaction mixture consists of 10 pmol of forward and reverse primers, 50 ng template DNA, 0.2 mM dNTPs, 1×Advantage 2 PCR buffer (400 mM Tricine-KOH, 150 mM KOAc, 35 mM

Table 1: Primers designed for partial *CAD* and *C4H* DNA sequences

Gene	Primer sequence	Amplicon size (bp)
C4H1	FL-NcC4H2-F: 5'-CATTTCCCGCCACCCATCA-3'	675
	PS-NcC4H-R1: 5'-GCTTTGAGCCAGCCTACT-3'	
C4H2	PS-NcC4H-F2: 5'-GAGTAGGCTGGCTCAAAG-3'	1450
	PS-NcC4H-R2a: 5'-CGGCGACATTGATGTTTTCCA-3'	
C4H3	PS-NcC4H-F3a: 5'-GAAGCCCAGCAGAAGGGAG-3'	899
	PS-NcC4H-R3a: 5'-CACCAGGTCCAAGCACAG-3'	
C4H4	PS-NcC4H-F4a: 5'-GGCATTGCGGAATTGGTT-3'	682
	FL-NcC4H2-R: 5'-CCTTGCGAATACAAAGATTATGG-3'	
CAD1	FL-NcCAD-F: 5'-TTTTCCCTCTGCTCCTTGC-3'	688
	PS-NcCAD-R1a: 5'-AACCAACAGCCTTAGACAA-3'	
CAD3	PS-NcCAD-F3a: 5'-AGGGCAACACTGGAGGATG-3'	656
	PS-NcCAD-R2: 5'-TGTGGCTTCCTGGTTTGG-3'	
CAD4	PS-NcCAD-F: 5'-TTGCCAAACCAGGAAGCC-3'	1030
	FL-NcCAD-R: 5'-GCCACAGGCATACGAGACAC -3'	

$Mg(OAc)_2$, 37.5 μg mL^{-1} BSA, 0.05% Tween 20 and 0.05% Nonidet-P40) and 1×Advantage 2 Polymerase Mix (Clontech, US). Amplification was performed in Veriti™ Thermal Cycler (Applied Biosystems, USA) for 2 min at 95°C, 35 cycles of 45 sec at 94°C, 45 sec at 61.2 or 63.1°C and 1 min at 72°C, followed by final extension of 10 min at 72°C.

Cloning and sequencing: PCR amplicons were purified using QIAquick®Gel Extraction Kit (QIAGEN, Germany) and then ligated into pGEM®-T Easy Vector System (Promega, USA). Colony PCR with M13 forward and reverse sequence primers were performed to identify the presence of positive clones. After that, plasmids DNA from positive clones were isolated and purified using Wizard® *Plus* SV Minipreps DNA Purification System (Promega, USA) according to the manufacture's protocol. The purified plasmids were sent for sequencing either in single forward or both forward and reverse directions. The sequencing was conducted by using 3730XL DNA Analyzer (Applied Biosystems, USA) and BigDye version 3.1 (Applied Biosystems, USA).

Sequence variation analysis: Base calling and vector sequences were removed using Chromas version 2.33 (Technelysium, AU). Each sequence was verified and checked for their homology using BLASTn (Altschul *et al*., 1990) through BLAST Search Engines (http://blast.ncbi.nlm.nih.gov/). After verification, the DNA sequences for each gene were aligned among individuals using CLC Free Workbench 4 software (CLC Bio, Denmark) to observe the single nucleotide polymorphisms. All sequence polymorphisms detected then were visually rechecked from chromatograms. Later, the open reading frame sequences for each gene were translated into amino acid sequences by using ORF finder (http://us.expasy.org/tools/dna.html). Then the amino acid sequences of *C4H* and *CAD* genes were aligned among individuals using CLC Free Workbench 4 software (CLC Bio, Denmark) in order to detect the nonsynonymous and synonymous mutations.

Basic wood density measurement: Two radial wood cores with 5 mm diameter and 8 cm long were sampled from 12 Kelampayan trees at breast height (about 1.3 m from the ground) and at opposite direction by using increment borer. The wood cores were immediately transferred into 15 mL microcentrifuge tubes, labelled and stored in a cooler box. The hole created on the tree was covered with grease. The volume of each wet wood core was taken by using water immersion method. The mass of water displaced in grams was taken to correspond closely to the volume of wet wood cores in cm^3. Later the cores were oven dried at 105°C for 48 h. The oven-dried weight of each wood core was then measured. After that, the basic wood density was calculated as oven dry mass per fresh volume.

Nucleotide diversity analysis: Haplotypes were directly inferred from *C4H* and *CAD* sequencing PCR products amplified in Kelampayan. Haplotype diversity (H_d) was computed according to Nei (1987), except that n is used instead of 2n. Levels of genetic variation were estimated as average per site pairwise nucleotide diversity, π (Nei, 1987) and as the relationship between segregating sites and alleles sampled, θ_w (Watterson, 1975) without considering InDels, at three different levels: (i) the whole sequenced region; (ii) noncoding regions (including introns, 3' and 5' UTRs) and (iii) coding regions. To test whether the folded site frequency spectrum was consistent with expectations derived from neutral evolution, Tajima's D-test (Tajima, 1989) was carried out. The estimates of haplotype diversity, nucleotide diversity and neutrality were calculated using DnaSP Version 5.10 software (Rozas *et al*., 2003). InDels were excluded from all estimations. Moreover, Linkage Disequilibrium (LD)

descriptive statistic, r^2, was also calculated by using TASSEL Version 3.0 software (Bradbury *et al*., 2007) in order to infer the LD decay within genes over distance.

Association genetics analysis: Genetic association among SNPs from *CAD* and *C4H* genes and the basic wood density were statistically tested using General Linear Model (GLM) which available in the TASSEL Version 3.0 software (http://www.maizegenetics.net/tassel) (Bradbury *et al*., 2007). Single-marker models were utilized for all SNP-trait combinations, with SNP markers as fixed effects. *P*-values were generated for each test using 1,000 permutations of genotypes with respect to phenotypic trait values. The p-values lower than 0.05 was considered as significant SNP-trait association.

RESULTS AND DISCUSSION

Single nucleotide polymorphism discovery: A simple re-sequencing approach was used to detect the SNPs in Kelampayan *C4H* and *CAD* genes. In total, 48 *C4H* and 36 *CAD* DNA fragments were cloned and sequenced from 12 Kelampayan trees. Ambiguous sequences were removed before subsequent analysis. Only nucleotide sequences that showed clear signals without overlay were aligned together to detect the sequence variations.

From the *C4H* DNA sequences alignment, 60 SNPs were detected. Of these, 24 were in exon regions, 35 were in intron regions and one in 3'-UTR (Table 2). Among the SNPs detected, 34 were resulted from transition mutation, 17 were resulted from transversion mutation and 9 were InDel mutation. On average, one SNP occurred at every 59 bp in *C4H* DNA sequence, which is similar to the SNP frequency reported for *C4H* from *Acacia mangium*×*A. auriculiformis* (1 SNP in every 53 bp) (Nur Fariza *et al*., 2008). Meanwhile, others also reported that the SNP frequency was one SNP in every 189 bp for barley (Kanazin *et al*., 2002), one SNP in every

Table 2: SNPs detected within partial *C4H* and *CAD* DNA sequences and the resulted synonymous and nonsynonymous mutations

Characteristic	*C4H*	*CAD*
No. of SNPs exon	24	7
Intron	35	20
5'-UTR	-	4
3'-UTR	1	1
Total no. of SNPs	60	32
Transition mutation	34	17
Transversion mutation	17	9
InDel mutation	9	6
Total sequence length	3,538	2,354
Occurrence of SNPs	1 SNP in 59 bp	1 SNP in 74 bp
Total no. of synonymous mutation	16	6
Total no. of nonsynonymous mutation	8	1
Total no. of amino acid	505	361

273 bp for soybean (Zhu *et al*., 2003), one SNP in every 100 bp for grapes (Velasco, *et al*., 2007), one SNP in every 109 bp for *Shorea parvifolia* ssp. *parvifolia* (Seng *et al*., 2011) and one SNP in every 229 bp for *A. mangium* superbulk (Tchin *et al*., 2011).

For *CAD* DNA sequence variations analysis, 32 SNPs was discovered (Table 2). Of these, 7 were in exon regions, 20 were in intron regions, four in 5'-UTR and one in 3'-UTR. Majority of the SNPs were resulted from transition mutation (17), 9 from transversion mutation and 6 were InDel mutation. The SNP frequency in *CAD* DNA sequence was one SNP in every 74 bp, which is higher than the SNP occurrence reported for *A. mangium* superbulk *CAD* (1 SNP in every 143 bp) (Tchin *et al*., 2011) and *Eucalyptus globules CAD* (1 SNP in every 147 bp) (Poke *et al*., 2003). But is lower than the occurrence reported for *A. mangium*×*A. auriculiformis* (1 SNP in every 40 bp) (Nur Fariza *et al*., 2008) and *E. nitens* (1 SNP in every 34 bp) (Southerton *et al*., 2010).

Overall, the transition mutations (55.0%) were frequently found in *C4H* and *CAD* as compared with transversion mutations (28.0%) (Table 2). This phenomenon also was seen in rice (Feltus *et al*., 2004), maize (Batley *et al*., 2003) and oil palm (Riju *et al*., 2007). In contrast, high proportion of transversion mutation over transition mutation was found in ginger (Chandrasekar *et al*., 2009) and eucalyptus (Singh *et al*., 2011) transcriptome.

InDel, which code for insertion and deletion mutation, is a special form of SNP caused by addition or removal of nucleotide bases from the DNA sequences. There were 15 InDels being discovered in this study which represent 16.0% of the total number of sequence variations detected (Table 2). The size of the InDels was ranged from single-bp to double-bp. Majority of the InDel were detected in intron regions and there was only one case found in coding region. Gene structure is often modified by small InDel and SNP (Zhao *et al*., 2007). InDel not only resulted in length variation but also sequence polymorphisms. At the position 894 bp of C4H3 fragments, the deletion of G nucleotide in NcMT2 sample had totally changed the amino acid sequence starting from the mutation site (Fig. 1). This frameshifting InDel may probably have important consequences toward the phenotypic variations of Kelampayan since it had altered the amino acid sequence of *C4H* in NcMT2 sample. Sweeney *et al*. (2006) reported that the frameshifting InDel in exon 6 of *basic helix-loop-helix (bHLH)* gene had resulted in two premature stop codons which truncated the protein sequence. They also found that this InDel was the main reason for the lack of red pigment in the pericarp of cv. Jefferson and cv. Nipponbare seeds. In rice, an InDel

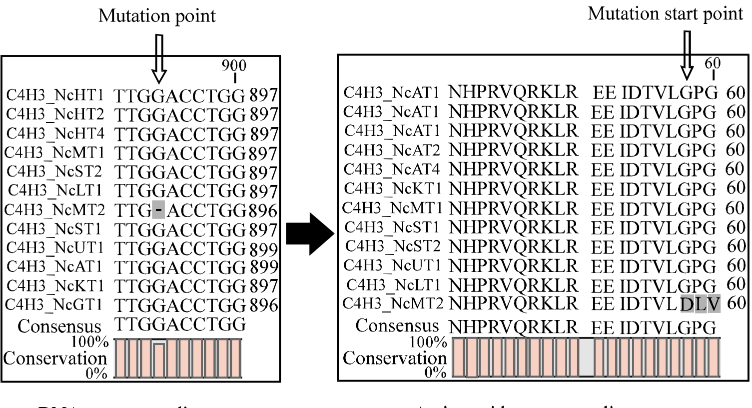

Fig. 1: The deletion of G nucleotide at position 894 bp of C4H3_NcMT2 fragment had changed the amino acid sequence starting from the mutation site

Table 3: Haplotype (H_d) diversity, nucleotide diversity (θ and π) and neutrality test statistics (D) in *C4H* and *CAD* candidate genes

Region	S	H	H_d	θ_w	π_{tot}	π_{sil}	π_{syn}	π_{nonsyn}	θ_{silent}	θ_{syn}	θ_{nonsyn}	Tajima's D test
C4H												
C4H1	12	7.0	0.77300	0.00589	0.00373	0.01332	0.01496	0.00065	0.02021	0.02271	0.00130	-1.54193
C4H2	7	4.0	0.45500	0.00344	0.00193	0.00244	0.00000	0.00000	0.00433	0.00000	0.00000	-1.71347
C4H3	17	8.0	0.84800	0.00628	0.00585	0.00672	0.01636	0.00122	0.00745	0.01663	0.00241	-0.52511
C4H4	7	7.0	0.83300	0.00340	0.00227	0.00418	0.00482	0.00116	0.00530	0.00753	0.00230	-1.30420
Mean		6.5	0.72725	0.00475	0.00345	0.00667	0.00904	0.00076	0.00932	0.01172	0.00150	-1.27118
CAD												
CAD1	13	7.0	0.87900	0.00628	0.00490	0.00553	0.02457	0.00000	0.00707	0.02685	0.00000	-0.92609
CAD3	0	1.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CAD4	13	7.0	0.77300	0.00418	0.00288	0.00457	0.00728	0.00040	0.00648	0.01061	0.00080	-1.31371
Mean		5.0	0.55067	0.00349	0.00259	0.00337	0.01062	0.00013	0.00452	0.01249	0.00027	-0.74660

InDels and sites with missing data were excluded for all measurement, S: Segregating sites, H: No. of haplotype, θ_w: Watterson estimator, π: Average pairwise nucleotide diversity, tot: Total (nonsynonymous, synonymous and noncoding), sil: Silent (synonymous and noncoding), syn: Synonymous, nonsyn: Nonsynonymous

in the exonic region of the blast disease resistance gene (*Pi54*) was also found to be associated with the blast disease resistance and this polymorphism was further developed as marker for use in rice breeding programme (Ramkumar *et al*., 2011).

For synonymous and nonsynonymous study, all coding regions of *C4H* and *CAD* were translated into amino acid sequences and then aligned among the 12 samples. Synonymous mutation does not lead to changes in the amino acid sequence of resulting protein while nonsynonymous mutations do. In total, 16 synonymous mutations were detected within the open reading frame of *C4H* and 8 nonsynonymous mutations were identified (Table 2). For *CAD*, 6 synonymous mutations and one nonsynonymous mutation were discovered (Table 2). Nonsynonymous SNPs which cause changes to the protein sequence can affect the structural, functional or biochemical properties of the enzyme being produced and therefore possibly lead to the changes in phenotypic characteristic of the trees, especially in modification of lignin biosynthesis (Bromberg and Rost, 2007). The synonymous SNPs in turn may possibly modify an RNA splicing site or may lead to different mRNA translation kinetics, thus yielding a protein with different final structure and function (Edwards *et al*., 2007).

Nucleotide diversity analysis: An average number of 5 different haplotypes per gene and relatively high haplotype diversity was observed for *CAD* (0.55067) and *C4H* (0.72725) genes (Table 3). Level of genetic variation as estimated by segregating sites (S), θ_w, π and H_d in *C4H* gene was greater than in *CAD* gene (Table 3). The estimated nucleotide diversity, π_{tot} and θ_w with values as low as zero was found in the CAD3 region and several times higher in the C4H3 region (π_{tot} = 0.00585; θ_w = 0.00628). Zero nucleotide diversity was due to the absent of nucleotide substitution in CAD3 region and an InDel mutation in that region had been excluded from the analysis.

The negative value of the neutrality test statistic (Tajima's D) in all gene fragments may indicate that the population are mainly under negative selection or reflect a recent population expansion (Krutovsky and Neale,

Table 4: Mean nucleotide diversity (θ_w and π) in different nucleotide sites or gene regions for *C4H* and *CAD* candidate genes

Sites	π	θ_w
All (coding+noncoding)	0.00302	0.00412
Coding	0.00277	0.00362
All coding		
Nonsynonymous	0.00045	0.00089
Synonymous	0.00983	0.01210
Noncoding	0.00361	0.00464
Silent (synonymous+noncoding)	0.00502	0.00692

InDels were excluded for all measurement, π: Average pairwise nucleotide diversity, θ_w: Watterson estimator

2005). In our study, all gene regions showed negative neutrality test statistic values but none of the value was significant (Table 3). This indicates that no evidence for departure from the neutral evolution in the Kelampayan populations for both *C4H* and *CAD* genes. Nevertheless, the negative value in the neutrality test may due to an excess of low- and high- frequency alleles detected in the loci, but the definition of low frequency is depends on the sample size (Ferretti *et al*., 2010).

The total nucleotide diversities were π = 0.00302 and θ_w = 0.00412 on average (Table 4). The nucleotide diversities at coding region (π = 0.00277; θ_w = 0.00362) were relatively lower than the noncoding region (π = 0.00361; θ_w = 0.00464). Meanwhile, the synonymous mutations (π = 0.00983; θ_w = 0.01210) were more common than nonsynonymous mutations (π = 0.00045; θ_w = 0.00089) for both *C4H* and *CAD* genes. In this study, the sites with missing data for C4H2_NcST1, C4H2_NcHT4 and C4H2_NcGT1 samples as well as InDels detected within the *C4H* and *CAD* gene fragments were not included in nucleotide diversity analysis. However, according to Olson *et al*. (2010), the overall level of nucleotide diversity may be underestimated if the regions with length variations or InDels were excluded from the analysis. Therefore, the nucleotide diversity obtained in this study could be lower than the exact value.

Linkage disequilibrium: The prerequisite for an effective use of association genetics is the need to understand the structure of Linkage Disequilibrium (LD) in the genome. Therefore, the LD decay was estimated by plotting r^2 value against the pairwise distances between sites (in bp) within all genes and fitted with a linear and logarithmic trend line (Fig. 2). LD declined linearly over short distance at the loci examined, to values of r^2 approaching 0.15 within several hundred base pairs, but a fair amount of LD were remained even for pairs separated by >1 kbp. However, it should be noted that this study was designed to address the LD within *C4H* and *CAD* candidate genes that have relatively short studied sequences. Therefore, the extent of LD across the entire Kelampayan genome is remained to be established. As reviewed by Abdurakhmonov and Abdukarimov (2008), the genome-wide LD in many forest trees species

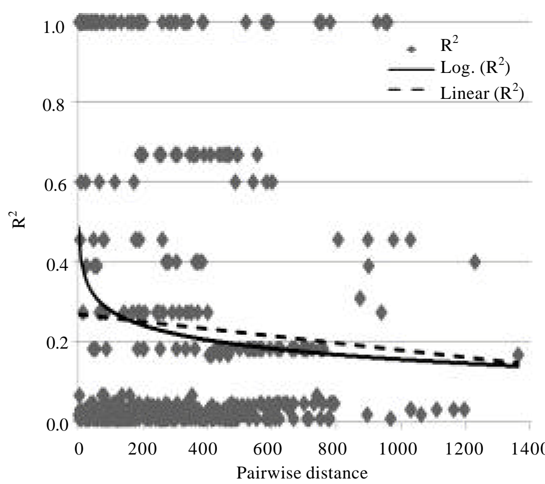

Fig. 2: LD plot for all paired polymorphic sites in *C4H* and *CAD* genes and fitted with a linear and logarithmic trend line

were extended up to 16-34 kbp in poplar (*P. trichocarpa*), <500 bp in European aspen (*P. tremula*), 2,000 bp in loblolly pine (*P. taeda*), 1,000 bp in Douglas-fir (*Pseudotsuga menziesii*) and 100-200 bp in Norway Spruce (*Picea abies*).

The rate of decay of LD with distance is a crucial factor that affects the success of candidate genes-based association mapping by using SNP markers. Genetic-traits association detection would become easier if the LD extends over longer distances in genome, but there is a drawback whereby the causal gene or Quantitative Trait Nucleotide (QTN) will remain unknown. In contrast, if LD decays over short distance, the significant SNP-traits association detected would be more likely is the causative rather than due to linkage with other unknown genes (Krutovsky and Neale, 2005). Although high mapping resolution can be achieved through the decline of LD, but a large number of markers is also needed for high resolution mapping. On the other hand, the number of markers required will greatly reduce if the LD extends over long distance across the genome which eventually leads to a low mapping resolution (Zhu *et al*., 2008). Nevertheless, this preliminary analysis suggests a relatively short range of LD in the sample of Kelampayan.

Table 5: Basic wood density measured for 12 Kelampayan samples

		Basic wood density (kg m^{-3})		
Sample	DBH (cm)	R1	R2	Average
NcLT1	25.0	411	404	407
NcKT1	41.0	359	362	360
NcST1	28.9	424	337	381
NcST2	39.6	338	349	343
NcAT1	32.3	414	377	395
NcHT1	59.7	359	423	391
NcHT2	42.5	305	387	346
NcHT4	24.0	335	264	300
NcMT1	102.0	440	447	444
NcMT2	34.5	362	335	348
NcUT1	40.1	381	413	397
NcGT1	25.8	372	306	339

Basic wood density: Basic wood density was measured for 12 Kelampayan samples (Table 5). On average, the basic wood density of Kelampayan was 371.452 kg m^{-3}. A tree that has low wood density usually will have a fast growth rate (Roderick, 2000; Muller-Landau, 2004). In contrast, high density wood generally correlated with superior mechanical performance (Macdonald and Hubert, 2002). However, the advantages of high wood density are recently being reclaimed (Larjavaara and Muller-Landau, 2010). Further revaluation has revealed that the rupture and elastic modulus are positively correlated with wood density meanwhile the bending resistance and wood strength are negatively correlated with wood density and square root of wood density, respectively. This indicates that at a given construction cost, the tree with lower wood density also will have high strength and high resistance to bending (Larjavaara and Muller-Landau, 2012). But this finding was later been challenged by Niklas and Spatz (2010, 2012) who found that the construction cost cannot be constant over an entire ensemble of stems composed of different wood species.

Association genetics study: Single marker association tests were carried out for SNPs data and basic wood density measured from the selected 12 Kelampayan trees. Significant marker-trait associations ($p<0.05$) were detected for four SNPs in *C4H* gene examined in this study (Table 6). The synonymous C4H_M4 SNP and nonsynonymous C4H_M5 SNP were significantly associated with the basic wood density of NcHT2 and NcHT4 samples. The C allele at C4H_M5 was the minor allele and caused a V→A amino acid substitution in *C4H*. Meanwhile, the noncoding *C4H*_M26 and C4H_M30 SNPs were significantly associated with the basic wood density of NcMT1 tree. Among the 12 samples measured, NcMT1 was the sample that having the highest values of basic wood density (444 kg m^{-3}). All the four SNP markers explained a proportion of the phenotypic variance, with individual effects ranging from 33.62 to 35.0%. According to a study carried out by Bjurhager *et al.* (2010), the down regulation of *C4H* gene in *Populus* had resulted in reduced lignin content and wood density as well as change in wood stiffness and morphology. Hence, along with the significant associations detected in this study, it highlighted the importance of *C4H* gene towards the wood properties of forest tree species.

Table 6: Significant associations for SNPs from *C4H* and *CAD* genes with basic wood density

		Marker effect		
Gene	SNP	F-value	p-value	r^2
C4H	C4H_M4	5.0642	0.0481	0.3362
	C4H_M5	5.0778	0.0479	0.3368
	C4H_M26	5.3842	0.0427	0.3500
	C4H_M30	5.3842	0.0427	0.3500
CAD	CAD_M14	5.0642	0.0481	0.3362
	CAD_M15	5.0642	0.0481	0.3362
	CAD_M17	5.0642	0.0481	0.3362
	CAD_M18	7.6617	0.0199	0.4338
	CAD_M24	5.0778	0.0479	0.3368
	CAD_M32	5.3842	0.0427	0.3500

Besides that six SNPs from *CAD* gene were also in significant associations ($p<0.05$) with basic wood density of Kelampayan (Table 6). This result was similar with the findings reported for *CAD* of loblolly pine (Yu *et al.*, 2006), *P. taeda* (Gonzalez-Martinez *et al.*, 2007) and *A. mangium* Superbulk (Tchin *et al.*, 2011). There were one synonymous SNP (CAD_M24), two noncoding SNPs (CAD_M14 and CAD_M17) and an InDel (CAD_M15) that significantly associated with the basic wood density of NcHT2 and NcHT4 samples. Meanwhile, one of the InDel (CAD_M18) and one SNP at 3'-UTR (CAD_M32) were significantly associated with the basic wood density of NcLT1 and NcMT1 samples. Individually, each of the six SNP markers explained a proportion of the phenotypic variance, with effects ranging from 33.62 to 43.38%.

Candidate gene-based association mapping study in Kelampayan is a useful tool in detecting the linkage between genetic variants and basic wood density. As reviewed by Macdonald and Hubert (2002), high density timber is usually correlated with stronger mechanical properties and high pulp yield per unit volume. Lower wood density will benefit the sawn timber and panel manufacturing industry as well as yielding a better pulp quality. Wood density is known to be depending on genetic effects and has a high heritability (Foster *et al.*, 1995; Louzada and Fonseca, 2002). Therefore, the development of SNP markers for high or low density Kelampayan wood would be able to facilitate in the selection of seedlings with traits of interest at the early stage. This approach not only reduces the cost and time for the plantation development but also extend to the

wood-based industry. However, inferences based on wood density alone are insufficient for Kelampayan that may have complex life histories. Moreover, SNPs that have negative correlation with wood density may have significance association with other desirable characters of Kelampayan. Therefore, more economically and ecologically important trait data such as wood quality, stress or drought tolerance and diseases resistant traits are needed (Neale and Kremer, 2011).

Furthermore, a genome-wide SNP data is necessary for population structure determination in order to capture the complex pattern of relatedness that may have in the Kelampayan samples. Population structure determination is essential to reduce the false positive and false negative rate in association genetics study (Yu *et al.*, 2006; Zhao *et al.*, 2007). Krutovsky and Neale (2005) had proposed on the order of 1,300 trees to obtain more data for population structure determination. Meanwhile, Long and Langley (1999) had proposed on the order of 500 samples to overcome the low repeatability problem of an association study. Thus, SNP marker density shall be increased by increasing the number of candidate genes studied as well as the sample size of Kelampayan in future studies. Recently, effort has been placed to discover SNP in a small sample size (10-100) and then genotype in a large population by using low cost and high throughput SNP genotyping platform (Neale, 2007; Syvanen, 2001).

In conclusion, the present study has demonstrated that association genetics study is a powerful approach in dissecting the potential genetic variants (i.e. SNPs) underlying the common and complex adaptive traits by harnessing the genetic variation at the population level. The gene-associated SNP identified in both *C4H* and *CAD* genes of *N. cadamba* could be potentially used as a tool in Gene-assisted Selection (GAS) of *N. cadamba* once there are validated in the near future. These validated SNPs will greatly facilitate the selection of quality planting materials that are of faster growth, high-yield and high wood quality and also adapted to local conditions, so that we may achieve economic benefits of great significance.

ACKNOWLEDGMENTS

The authors would like to thank all the lab assistants and foresters involved in this research project for their excellent field assistance in sample collection. This study is part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation (SFC) and Universiti Malaysia Sarawak (UNIMAS).

REFERENCES

Abdurakhmonov, I.Y. and A. Abdukarimov, 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics, 10.1155/2008/574927

Abreu, H.S., J.V.F. Latorraca, R.P.W. Pereira, M.B.O. Monteiro, F.A. Abreu and K.F. Amparado, 2009. A supramolecular proposal of lignin structure and its relation with the wood properties. An. Acad. Bras. Cienc., 81: 137-142.

Acharyya, S., G.K. Dash, S. Mondal and S.K. Dash, 2010. Studies on glucose lowering efficacy of the *Anthocephalus cadamba* (Roxb.) Miq. roots. Int. J. Pharm. Biosci., 1: 1-9.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool. J. Mol. Biol., 215: 403-410.

Batley, J., G. Barker, H. O'Sullivan, K.J. Edwards and D. Edwards, 2003. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol., 132: 84-91.

Baucher, M., C. Halpin, M. Petit-Conil and W. Boerjan, 2003. Lignin: Genetic engineering and impact on pulping. Crit. Rev. Biochem. Mol. Biol., 38: 305-350.

Bentz, B.J., R.R. Bracewell, K.E. Mock and M.E. Pfrender, 2011. Genetic architecture and phenotypic plasticity of thermally-regulated traits in an eruptive species, *Dendroctonus ponderosae*. Evol. Ecol., 25: 1269-1288.

Bjurhager, I., A.M. Olsson, B. Zhang, L. Gerber and M. Kumar *et al.*, 2010. Ultrastructure and mechanical properties of *Populus* wood with reduced lignin content caused by transgenic down-regulation of *cinnamate 4-hydroxylase*. Biomacromolecules, 11: 2359-2365.

Boerjan, W., 2005. Biotechnology and the domestication of forest trees. Curr. Opin. Plant Biol., 16: 159-166.

Bowyer, J.L. and J.G. Haygreen, 1996. Forest Products and Wood Science. 3rd Edn., John Wiley and Sons, Ames, IA., USA., ISBN-13: 9780813822563, Pages: 484.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler, 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23: 2633-2635.

Bromberg, Y. and B. Rost, 2007. SNAP: Predict effect of non-synonymous polymorphisms on function. Nucl. Acids Res., 35: 3823-3835.

Chandrasekar, A., A. Riju, K. Sithara, S. Anoop and S.J. Eapen, 2009. Identification of single nucleotide polymorphism in ginger using expressed sequence tags. Bioinformation, 4: 119-122.

Doyle, J.J. and J.L. Doyle, 1990. Isolation of plant DNA from fresh tissue. Focus, 12: 13-15.

Edwards, D., J.W. Forster, D. Chagne and J. Batley, 2007. What are SNPs? In: Association Mapping in Plants, Oraguzie, N.C., E.H.A. Rikkerink, S.E. Gardiner and H.N.D. Silva (Eds.). Springer, New York, USA., pp: 41-52.

Feltus, F.A., J. Wan, S.R. Schulze, J.C. Estill, N. Jiang and A.H. Paterson, 2004. An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. Genome Res., 14: 1812-1819.

Ferretti, L., M. Perez-Enciso and S. Ramos-Onsins, 2010. Optimal neutrality tests based on the frequency spectrum. Genetics, 186: 353-365.

Fitzpatrick, M.J., Y. Ben-Shahar, H.M. Smid, L.E. Vet, G.E. Robinson and M.B. Sokolowski, 2005. Candidate genes for behavioural ecology. Trends Ecol. Evol., 20: 96-104.

Flatscher, R., B. Frajman, P. Schonswetter and O. Paun, 2012. Environmental heterogeneity and phenotypic divergence: Can heritable epigenetic variation aid speciation? Genet. Res. Int., 10.1155/2012/698421.

Foster, G.S., N. Jones and E.D. Kjaer, 1995. Economics of Tree Improvement in Development Projects in the Tropics. In: Environmental and Economic Issues in Forestry: Selected Case Studies in Asia, Shen, S. and A. Contreras-Hermosilla (Eds.). The World Bank, Washington, DC., USA., ISBN-13: 9780821332337, pp: 99.

Fusari, C.M., V.V. Lia, H.E. Hopp, R.A. Heinz and N.B. Paniego, 2008. Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol., Vol. 8. 10.1186/1471-2229-8-7

Gonzalez-Martinez, S.C., K.V. Krutovsky and D.B. Neale, 2006. Forest-tree population genomics and adaptive evolution. New Phytol., 170: 227-238.

Gonzalez-Martinez, S.C., N.C. Wheeler, E. Ersoz, C.D. Nelson and D.B. Neale, 2007. Association genetics in *Pinus taeda* L. I. wood property traits. Genetics, 175: 399-409.

Hall, D., C. Tegstrom and P.K. Ingvarsson, 2010. Using association mapping to dissect the genetic basis of complex traits in plants. Briefings Funct. Genomics, 9: 157-165.

Hayashi, K., N. Hashimoto, M. Daigen and I. Ashikawa, 2004. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. Theor. Applied Genet., 108: 1212-1220.

Ho, W.S., S.L. Pang, P.S. Lai, S.Y. Tiong and S.L. Phui *et al*., 2010. Genomics studies on plantation tree species in Sarawak. Proceedings of the International Symposium on Forestry and Forest Products 2010: Addressing the Global Concerns and Changing Societal Needs, October 5-7, 2010, The Legend Hotel, Kuala Lumpur, Malaysia, pp: 172-182.

Ingvarsson, P.K. and N.R. Street, 2011. Association genetics of complex traits in plants. New Phytol., 189: 909-922.

Joker, D., 2000. *Neolamarckia cadamba* (Roxb.) Bosser (*Anthocephalus chinensis* (Lam.) A. Rich. ex Walp.). Seed Leaflet No. 17, Danida Forest Seed Centre. http://curis.ku.dk/portal-life/files/20648324/neolamarckia_cadamba_int.pdf

Kanazin, V., H. Talbert, D. See, F. DeCamp, E. Nevo and T. Blake, 2002. Discovery and assay of single nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol. Biol., 48: 529-537.

King, D.A., S.J. Davies, M.N. Nur Supardi and S. Tan, 2005. Tree growth is related to light interception and wood density in two mixed dipterocarp forests of Malaysia. Funct. Ecol., 19: 445-453.

King, D.A., S.J. Davies, S. Tan and M.N. Nur Supardi, 2006. The role of wood density and stem support costs in the growth and mortality of tropical trees. J. Ecol., 94: 670-680.

Krutovsky, K.V. and D.B. Neale, 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas fir. Genetics, 171: 2029-2041.

Larjavaara, M. and H.C. Muller-Landau, 2010. Rethinking the value of high wood density. Funct. Ecol., 24: 701-705.

Larjavaara, M. and H.C. Muller-Landau, 2012. Still rethinking the value of high wood density. Am. J. Bot., 99: 165-168.

Lau, E.T., W.S. Ho and A. Julaihi, 2009. Molecular cloning of *cellulose synthase* gene, *SpCesA1* from developing xylem of *Shorea parvifolia* spp. *parvifolia*. Biotechnology, 8: 416-424.

Lewis, N.G., 1999. A 20th century roller coaster ride: A short account of lignification. Curr. Opin. Plant Biol., 2: 153-162.

Lida, Y., L. Poorter, F.J. Sterck, A.R. Kassim, T. Kubo, M.D. Potts and T.S. Kohyama, 2012. Wood density explains architectural differentiation across 145 co-occurring tropical tree species. Funct. Ecol., 26: 274-282.

Lim, S.C., K.S. Gan and B.K. Thi, 2005. Identification and utilization of lesser-known commercial timbers in Peninsular Malaysia. 4. Kelempayan, Melembu, Membuloh and Mempari. Timber Technology Bulletin No. 32, Forest Research Institute Malaysia, pp: 1-8.

Long, A.D. and C.H. Langley, 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res., 9: 720-731.

Louzada, J.L.P.C. and F.M.A. Fonseca, 2002. The heritability of wood density components in *Pinus pinaster* Ait. and the implications for tree breeding. Ann. For. Sci., 59: 867-873.

Macdonald, E. and J. Hubert, 2002. A review of the effects of silviculture on timber quality of Sitka spruce. Forestry, 75: 107-138.

Muller-Landau, H.C., 2004. Interspecific and inter-site variation in wood specific gravity of tropical trees. Biotropica, 36: 20-32.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich and E.S. Buckler, 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell, 21: 2194-2202.

Neale, D.B. and A. Kremer, 2011. Forest tree genomics: Growing resources and applications. Genetics, 12: 111-122.

Neale, D.B. and O. Savolainen, 2004. Association genetics of complex traits in conifers. Trends Plant Sci., 9: 325-330.

Neale, D.B., 2007. Genomics to tree breeding and forest health. Curr. Opin. Genet. Dev., 17: 539-544.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, USA.

Nichols, K.M. and D.B. Neale, 2010. Association Genetics, Population Genomics and Conservation: Revealing the Genes Underlying Adaptation in Natural Populations of Plants and Animals. In: Molecular Approaches in Natural Resource Conservation and Management, DeWoody, J.A., J.W. Bickham, C.H. Michler, K.M. Nichols, O.E. Rhodes Jr. and K.E. Woeste (Eds.). Cambridge University Press, New York, USA., ISBN-13: 9780521731348, pp: 123-168.

Niklas, K.J. and H.C. Spatz, 2010. Worldwide correlations of mechanical properties and green wood density. Am. J. Bot., 97: 1587-1594.

Niklas, K.J. and H.C. Spatz, 2012. Mechanical properties of wood disproportionately increase with increasing density. Am. J. Bot., 99: 169-170.

Nur Fariza, M.S., S.L. Pang, C.Y. Choong and R. Wickneswari, 2008. Extensive DNA sequence variations in two lignin genes, *cinnamate 4-hydroxylase* and *cinnamyl alcohol dehydrogenase* from *Acacia mangium* and *Acacia auriculiformis*. J. Boil. Sci., 8: 687-690.

Olson, M.S., A.L. Robertson, N. Takebayashi, S. Silim, W.R. Schroeder and P. Tiffin, 2010. Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). New Phytol., 186: 526-536.

Patel, D. and V. Kumar, 2008. Pharmacognostical studies of *Neolamarckia cadamba* (Roxb.) Bosser leaf. Int. J. Green Pharm., 2: 26-27.

Poke, F.S., R.E. Vaillancourt, R.C. Elliot and J.B. Reid, 2003. Sequence variation in two lignin biosynthesis genes, *cinnamoyl CoA reductase* (*CCR*) and *cinnamyl alcohol dehydrogenase 2* (*CAD2*). Mol. Breed., 12: 107-118.

Prince, J.A., L. Feuk, W.M. Howell, M. Jobs, T. Emahazion, K. Blennow and A.J. Brookes, 2001. Robust and accurate single nucleotide polymorphism genotyping by Dynamic Allele-Specific Hybridization (DASH): Design criteria and assay validation. Genome Res., 11: 152-162.

Rafalski, J.A., 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol., 5: 94-100.

Ramkumar, G., K. Srinivasarao, K.M. Mohan, I. Sudarshan and A.K.P. Sivaranjani *et al*., 2011. Development and validation of functional marker targeting an InDel in the major rice blast disease resistance gene Pi54 (Pik^h). Mol. Breed., 27: 129-135.

Riju, A., A. Chandrasekar and V. Arunachalam, 2007. Mining for single nucleotide polymorphisms and insertions/deletions in expressed sequence tag libraries of oil palm. Bioinformation, 2: 128-131.

Roderick, M.L., 2000. On the measurement of growth with applications to the modelling and analysis of plant growth. Funct. Ecol., 14: 244-251.

Rozas, J., J.C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19: 2496-2497.

Schilmiller, A.L., J. Stout, J.K. Weng, J. Humphreys, M.O. Ruegger and C. Chapple, 2009. Mutations in the *cinnamate 4-hydroxylase* gene impact metabolism, growth and development in *Arabidopsis*. Plant J., 60: 771-782.

Seng, H.W., P.S. Ling, P. Lau and I. Jusoh, 2011. Sequence variation in the *Cellulose synthase* (SpCesA1) gene from *Shorea parvifolia* ssp. *parvifolia* mother trees. Pertanika J. Trop. Agric. Sci., 34: 317-323.

Singh, T.R., A. Gupta, A. Riju, M. Mahalaxmi, A. Seal and V. Arunachalam, 2011. Computational identification and analysis of single-nucleotide polymorphisms and insertions/deletions in expressed sequence tag data of *Eucalyptus*. J. Genet., 90: e34-e38.

Southerton, S.G., C.P. MacMillan, J.C. Bell, N. Bhuiyan and G. Downes *et al*., 2010. Association of allelic variation in xylem genes with wood properties in *Eucalyptus nitens*. Aust. For., 73: 259-264.

Suzuki, E., 1999. Diversity in specific gravity and water content of wood among Bornean tropical rainforest trees. Ecol. Res., 14: 211-224.

Sweeney, M.T., M.J. Thomson, B.E. Pfeil and S. McCouch, 2006. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell, 18: 283-294.

Syvanen, A.C., 2001. Accessing genetic variation: Genotyping single nucleotide polymorphisms. Nat. Rev. Genet., 2: 930-942.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123: 585-595.

Tchin, B.L., W.S. Ho, S.L. Pang and J. Ismail, 2011. Gene-associated single nucleotide polymorphism (SNP) in *cinnamate 4-hydroxylase* (*C4H*) and *cinnamyl alcohol dehydrogenase* (*CAD*) genes from *Acacia mangium* superbulk trees. Biotechnology, 10: 303-315.

Van Gelder, H.A., L. Poorter and F.J. Sterck, 2006. Wood mechanics, allometry and life-history variation in a tropical rain forest tree community. New Phytol., 171: 367-378.

Velasco, R., A. Zharkikh, M. Troggio, D.A. Cartwright and A. Cestaro *et al*., 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE, Vol. 2.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Population Biol., 7: 256-276.

Wegrzyn, J.L., A.J. Eckert, M. Choi, J.M. Lee and B.J. Stanton *et al*., 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol., 188: 515-532.

Yu, Q., B. Li, C.D. Nelson, S.E. McKeand, V.B. Batista and T.J. Mullin, 2006. Association of the *cad-n1* allele with increased stem growth and wood density in full-sib families of loblolly pine. Tree Genet. Genomes, 2: 98-108.

Zhao, K., M.J. Aranzana, S. Kim, C. Lister and C. Shindo *et al*., 2007. An *Arabidopsis* example of association mapping in structured samples. PLoS Genet., Vol. 3. 10.1371/journal.pgen.0030004

Zhu, C., M. Gore, E.S. Buckler and J. Yu, 2008. Status and prospects of association mapping in plants. Plant Genome, 1: 5-20.

Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell and L.K. Matukumalli *et al*., 2003. Single-nucleotide polymorphisms in soybean. Genetics, 163: 1123-1134.