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

1B.L. Tchin, 2W.S. Ho, 3S.L. Pang and 1J. Ismail

1Forest Genomics and Informatics Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia
2Applied 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 significantly associated 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 (NedbEST) 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.
individuals may due to the cumulative effect of a number of genes and/or the environmental influences (Benz 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 a marker of choice for association genetics study due to the abundance, stable, ubiquity and interspersed characteristics in the 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 constitutive of a living organism (Rafalski, 2002). SNPs are far easier to be detected since only one single base in specific sequences become target (Pringle 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., 2005) 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 o-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; Tehin et al., 2011; Schulmiller et al., 2009; Bjurhager et al., 2010; Wegryn et al., 2010). Abreu et al. (2009) also proposed that the high β-O-4 (Alkyl Aeryl Ether) bonds in lignins of angiosperms may possibly affect the wood properties. Therefore, further exploration on the C4H and CAD genes-trait 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 manufacturer’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 NccC4H (GenBank accession number: JQ946327) and NccCAD (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
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, USA). 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 Miniprep 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 C4D 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 cm³. 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 C4D sequencing PCR products amplified in Kelampayan. Haplotype diversity (Hₑ) 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, θₑ (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 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 (Tehn 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) (Tehn et al., 2011) and *Eucalyptus globules* \( CAD \) (1 SNP in every 147 bp) (Pokk 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 (Rijal 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 InDel being discovered in this study which represent 16.0% of the total number of sequence variations detected (Table 2). The size of the InDel 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 framingshift 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 framingshifting 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

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>( C4H )</th>
<th>( CAD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of SNPs exon</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>Intron</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>5'-UTR</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3'-UTR</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total no. of SNPs</td>
<td>60</td>
<td>32</td>
</tr>
<tr>
<td>Transition mutation</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Transversion mutation</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>InDel mutation</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Total sequence length</td>
<td>3,538</td>
<td>2,354</td>
</tr>
<tr>
<td>Occurrence of SNPs</td>
<td>1 SNP in 50 bp</td>
<td>1 SNP in 74 bp</td>
</tr>
<tr>
<td>Total no. of synonymous mutation</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Total no. of nonsynonymous mutation</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Total no. of amino acid</td>
<td>565</td>
<td>361</td>
</tr>
</tbody>
</table>
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_2 in C4H gene was greater than in CAD gene (Table 3). The estimated nucleotide diversity, π, and θ_w with values as low as zero was found in the CAD3 region and several times higher in the C4H3 region (π = 0.00085; θ_w = 0.00628). Nucleotide diversity was due to the absence 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, 2011).
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 $\pi = 0.00302$ and $\theta_w = 0.00412$ on average (Table 4). The nucleotide diversities at coding region ($\pi = 0.00277$; $\theta_w = 0.00362$) were relatively lower than the noncoding region ($\pi = 0.00361$; $\theta_w = 0.00464$). Meanwhile, 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. 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 kb. 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 were extended up to 16-34 kb 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-trait 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-trait 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.
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 correlates 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 reevaluation has revealed that the rupture and elastic modulus are positively correlated with wood density while 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 NeHT2 and NeHT4 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 Bjurfager et al. (2010), the down regulation of C4H gene in Populus has 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.

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., 2005), P. taeda (Gonzalez-Martinez et al., 2007) and A. mangium Superbolk (Teihin 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 Forseca, 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


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
