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

Bio Technology

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

Biotechnology 10 (4): 303-315, 2011
ISSN 1682-296X / DOI: 10.3923/biotech.2011.303.315

Gene-associated Single Nucleotide Polymorphism (SNP) in *Cinnamate 4-Hydroxylase* (*C4H*) and *Cinnamyl Alcohol Dehydrogenase* (*CAD*) Genes from *Acacia mangium* Superbulk Trees

[1]B.L. Tchin, [1]W.S. Ho, [2]S.L. Pang and [1]J. Ismail
[1]Forest Genomics and Informatics Laboratory (*f*GiL), 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 (AFSID), Sarawak Forestry Corporation,
93250 Kuching, Sarawak, Malaysia

Abstract: Candidate-gene-based association study which involves the identification of causative Single Nucleotide Polymorphisms (SNPs) for excellent traits has been proposed as a promising approach to dissect complex traits in forest trees. Hence, the goal of this study was to identify the genetic association among SNPs from *Cinnamate 4-Hydroxylase* (*C4H*) and *Cinnamyl Alcohol Dehydrogenase* (*CAD*) genes and an array of wood properties namely, specific gravity, wood density, fiber-length, cell wall thickness and microfibril angle from *Acacia mangium* Superbulk trees. Sequence variations within these two genes in 12 *A. mangium* Superbulk trees were examined and wood properties were measured. The data obtained was tested using General Linear Model (GLM) within TASSEL software. Two SNPs were identified in the exon of *C4H*, of which all the SNPs caused nonsynonymous mutations whereas five SNPs were identified in the *CAD* exons along with one deletion mutation. In addition, two SNPs were also identified in the *CAD* introns. Variation in these two lignin biosynthesis genes might change the structural, functional or biochemical properties of the enzyme being produced, and therefore possibly lead to changes in phenotypic characteristic of the trees. The genetic association study also revealed that SNPs in *CAD* gene do associate with the wood density, specific gravity and cell wall thickness ($p<0.05$). However, no significant results were obtained for SNPs in *C4H* gene with wood properties studied. Thickening of cell wall is affected by the arrangement of biopolymer aggregates which comprise of cellulose, hemicellulose and lignin. Results indicated that SNP in *CAD* gene might alter the lignin biosynthesis and thus lead to changes in phenotypic characteristics of the trees. Overall, the study has demonstrated that SNP is very useful in association genetic study to identify Quantitative Trait Nucleotide (QTN) which then leads to Gene-assisted Selection (GAS) in the tree breeding programme.

Key words: Single nucleotide polymorphism (SNP), association genetics, *cinnamate 4-hydroxylase* (*C4H*), *cinnamyl alcohol dehydrogenase* (*CAD*), *Acacia mangium* Superbulk

INTRODUCTION

Acacia mangium Superbulk, as one of the fast growing and with high adaptability species, is introduced to Malaysia for forest plantations and rehabilitation of marginal and degraded lands purposes (Wickneswari *et al*., 2005). To date, *A. mangium* has been extensively planted in Sarawak and the wood derived from this species is widely used for the production of fine furniture, veneers, plywood, construction of building as well as paper and pulp making (Lim *et al*., 2003). Due to the high potential uses of *Acacia* wood, better understanding the structure and composition of the wood is vital for optimal utilization of this wood material.

Wood is one of the most complex materials which composed of polymers of lignin and carbohydrates that are physically and chemically bond together. Lignin, the second most abundant organic compound after cellulose, represents approximately 20-30% of the plant biomass. Lignin is formed through dehydrogenative polymerization of monolignols known as coumaryl alcohol, coniferyl alcohol and sinapyl alcohol which will give rise to ρ-coumaryl units (H), guaiacyl (G) units and sinapyl (S) units, respectively. Functions of lignin involve the mechanical and structural support to the plants as well as a significant protective function again pathogen or decaying fungi (Brett and Waldron, 1990).

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

Many genes are involved in the lignin biosynthesis pathway. In this study, the major focus is on *Cinnamate 4-Hydroxylase* (*C4H*) and *Cinnamyl Alcohol Dehydrogenase* (*CAD*) genes that originate from *A. mangium* Superbulk. 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). Expression of *C4H* and *CAD* genes are very influential in the lignin biosynthesis pathway and any up- or down-regulation of these genes will be resulted in altered lignin production (Baucher *et al*., 2003). Abreu *et al*. (2009) had proposed that the high β-O-4 (Alkyl Aril Ether) bonds in lignin of angiosperms may possibly affect the wood properties. Therefore, extensive study needs to be carried out in order to determine the association between the lignin and the wood properties.

However, it is difficult to study the inherent complexity of these traits through traditional linkage-based approaches since it is time consuming, laborious and often expensive process of establishing the mapping population (Myles *et al*., 2009). On the other hand, candidate-gene-based association study which involves the identification of causative polymorphisms (SNPs) for excellent traits has been proposed as a promising approach to dissect complex traits in forest trees species (Gonzalez-Martinez *et al*., 2006; Ho *et al*., 2011). SNP is a single base mutation where DNA sequences differ only by a single base. SNPs have become marker of choice for many applications in genome analysis because SNPs are abundance, stable, ubiquity and interspersed in nuclear genome (Rafalski, 2002).

Association genetic study is a natural uncontrolled experiment that can give a higher mapping resolution compared with the linkage mapping (Neale and Savolainen, 2004). In the past decades, association study has been widely used not only in human (Abbasi *et al*., 2009), animal (Alashawkany *et al*., 2008) and crop species (Aina *et al*., 2007; Rashwan, 2011) but also become increasingly utilize in forest trees species. This can be proved by the first published association study in forest tree species by Thumma *et al*. (2005) where 2 SNPs found in *CCR* gene were significantly associated with microfibril angle from *Eucalyptus nitens*. The result was then validated in two full-sib families of *E. nitens* and *E. globules*. Yu *et al*. (2006) also detected the association of the *cad-n1* allele with increased stem growth and wood density in 15 years old loblolly pine. Later Gonzalez-Martinez *et al*. (2007) had found that a nonsynonymous substitution for *CAD* in loblolly pine was in strong association with earlywood specific gravity and causing lignin modification in wild trees.

Discovery of SNPs through association genetic study would result in the identification of causative variants (quantitative trait nucleotides, QTNs) which can be used as markers for gene-assisted selection in tree breeding programme to choose for good planting materials based on their DNA sequences (Fusari *et al*., 2008). This strategy is very important since it can increase the productivity of forests and enhance the wood quality or properties. By knowing the exact identity and properties of the timber through gene-assisted selection, one can easily determine and select for the desired trees in short period of time and thus reducing the costs of phenotypic testing.

Hence, the primary goal of this study was to identify the genetic association among single nucleotide polymorphisms (SNPs) from *C4H* and *CAD* genes and an array of wood properties namely, specific gravity, wood density, fiber-length, cell wall thickness and microfibril angle from *A. mangium* Superbulk trees. From the study, association was detected between *CAD* gene and wood properties studied but no significant result was obtained for SNPs in *C4H* gene.

MATERIALS AND METHODS

Plant materials: A total of 12 fresh leaf samples were collected from a 3-year old *A. mangium* Superbulk research plot at the UNIMAS's arboretum in August 2009. The research plot was established by using the second generation seedlings originally from the 49 superior *A. mangium* parent trees planted in the CSIRO first generation seedling seed orchards in north Queensland, Australia. *A. mangium* Superbulk is a plus tree or second generation of *A. mangium* in which the wood properties have been improved through many years of selected planting. Total genomic DNA was extracted from the fresh leaf tissues by using a modified protocol of Doyle and Doyle (1990). The extracted DNA was then purified by using Wizard® Genomic DNA Purification Kit (Promega, USA).

Polymerase chain reaction (PCR) amplification: PCR was performed on a Mastercycler Gradient Thermal Cycler (Eppendorf, Germany). Three specific primer sets designed from the full length *C4H* (EU275980) and *CAD* (EU275981) cDNA of *Acacia* hybrid (Pang *et al*., 2005) were used. PCR reaction mixture consists of 30 ng of DNA template, 5 pmol of forward and reverse primers, 0.2 mM dNTPs, 1× PCR buffer (200 mM Tris-HCI pH 8.8, 100 mM KCI, 1% Triton X-100, 100 mM $(NH_4)_2SO_4$, 1 μg mL^{-1}

BSA), 1.5 mM $MgCI_2$, 1 U Taq DNA Polymerase (Promega, USA) and sterile distilled water to make up 25 μL. *C4H* amplification was performed for 2 min at 95°C, 35 cycles of 45 sec at 94°C, 45 sec at 63°C and 1 min at 72°C, followed by final extension of 10 min at 72°C. *CAD* amplification was performed with PCR parameters same as *C4H* but with annealing temperature at 63°C for primers CAD-PT1 and 61°C for primers CAD-PT2.

Cloning and sequencing: Purified PCR product was ligated into pGEM®-T Easy Vector System (Promega, USA) and transformed into competent cells of *Escherichia coli* JM 109. Positive clones were cultured in 50 mL Falcon tubes containing 5 mL LB Broth supplemented with 100 mg mL^{-1} ampicillin for overnight. The recombinant plasmids 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.

Data 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/). The sequences of all trees were aligned using CLC Free Workbench 4 software (CLC Bio, Denmark) to observe the SNPs. Later, all sequences were translated into open reading frames using ORF finder (http://us.expasy.org/tools/dna.html).The open reading frames were aligned among 12 individuals using CLC Free Workbench 4 software (CLC Bio, Denmark) in order to detect the synonymous and nonsynonymous mutations.

Phenotypic data: Radial wood cores (5 mm) used in this study were sampled at the breast height from 12 *Acacia mangium* Superbulk trees. All trees were at 3 years old and planted in UNIMAS's arboretum. The heartwood was separated from sapwood and only heartwood was used for subsequent measurement. The wood density and specific gravity were firstly determined for each heartwood core.

Fiber-length: Few thin wood pieces were cut from wood cores with scalpel. The wood pieces were immersed into test tube that pre-filled with glacial acid and 30% hydrogen peroxide (1: 2) followed by incubation in water bath at 60-70°C until the wood pieces soft and turned into white color. The fibers then were separated by shaking the tube in the presence of dd H_2O. The fibers were placed on glass slides with cover slips and 50 fiber length measurements were made for each sample by using confocal laser scanning microscope (CLSM).

Microfibril angle (MFA) and cell wall thickness: The softened wood cores were cut with microtome at 16-18 μm thickness and in two dimensions, known as transverse and radial. Microfibril angle was measured at radial section while cell wall thickness was measured at transverse section. Radial sections were stained with 0.1% congo red and sonicated for 3 h. The washed wood slices were mounted on glass slides with cover slips. The microfibril angle in the S_2 layer of the fiber walls was determined using the polarized light microscope. Fifty replicate measurements of the microfibril angle were made for each sample. Transverse sections (Fig. 1) were stained with 1% safranin and then washed with dd H_2O and a series of ethanol at different concentrations. After that, stained sections were mounted permanently on the glass slide by using DPX mounting medium. All the slides were dried and then used for observation. Observations were performed using confocal laser scanning microscope

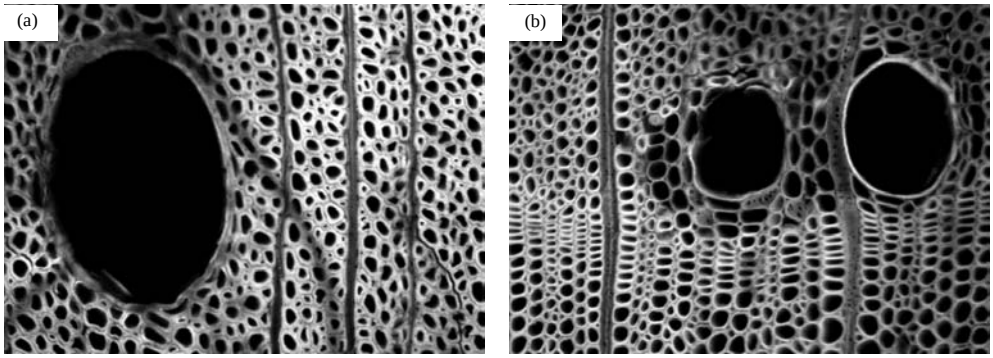

Fig. 1(a-b): Transverse section of *A. mangium* Superbulk. Cell wall thickness was measured from the transverse section of *A. mangium* Superbulk wood by using confocal laser scanning microscope (40X objective). From the figures, the cell wall of SB1 sample (a) was thicker than the cell wall of SB6 sample (b)

(CLSM) and 50 fiber cell wall thickness measurements were made for each sample.

Test statistics for association: Association between the SNPs discovered and phenotypic traits was tested using General Linear Model (GLM) that performs association analysis by a least squares fixed effects model. GLM is available within Trait Analysis by Association, Evolution and Linkage (TASSEL) software (http://www.maizegenetics.net/tassel).

RESULTS AND DISCUSSION

Cloning of partial *C4H* and *CAD genes*: Partial DNA sequences of *C4H* and *CAD* were cloned and sequenced from 12 samples. After trimming, the identity of each nucleotide sequence was verified by using BLASTn. As shown in Table 1, the partial sequence of *C4H* showed high degree of similarity to *C4H* sequence from *Acacia auriculiformis*×*Acacia mangium* hybrid (98%) and *Sorghum bicolor* (86%) but with limited nucleotides coverage. While for partial sequence of *CAD*, high degree of similarity was obtained when compared with *Acacia auriculiformis* × *Acacia mangium* hybrid (98%), *Populus tomentosa* (80%) and *Eucalyptus urophylla* (82%) (Table 2). After verification, the partial DNA sequences of *C4H* and *CAD* were aligned with the *C4H* (ABX75854.1) and *CAD* (EU275982.1) full length cDNA sequence of *Acacia* hybrid to determine the intronic and exonic region. From the alignments, the partial *C4H* sequence obtained was fall within the exonic region (Fig. 2). Meanwhile three exonic regions and two intronic regions were found within the partial *CAD* sequence (Fig. 3).

Table 1: BLASTn output for amplified partial *C4H* DNA sequence

Organisms	GenBank Accession No.	Similarity (%)	E-value
A. auriculiformis×			
A. mangium	EU275980.1	98	0.0
Sorghum bicolor	AY034143.1	86	4e-8e
Salvia miltiorrhiza	EF377337.1	84	3e-100
Rubus coreanus	EU123531.1	82	1e-98

Table 2: BLASTn output for amplified partial *CAD* DNA sequence

Organisms	GenBank Accession No.	Similarity (%)	E-value
A. auriculiformis ×			
A. mangium	EU275982.1	98	0.0
Leucaena leucocephala	EF611250.1	94	0.0
Populus tomentosa	EU760897.1	80	2e-79
Eucalyptus urophylla	GQ387647.1	82	3e-66

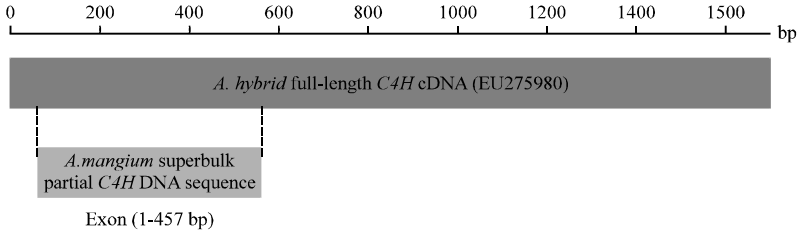

Fig. 2: The exonic region of partial *C4H* DNA sequence. The partial *C4H* DNA sequence (457bp) of *A. mangium* Superbulk was aligned with the full-length *C4H* cDNA from *Acacia* hybrid and the result showed that the partial *C4H* DNA sequence was found within the exonic region of the gene

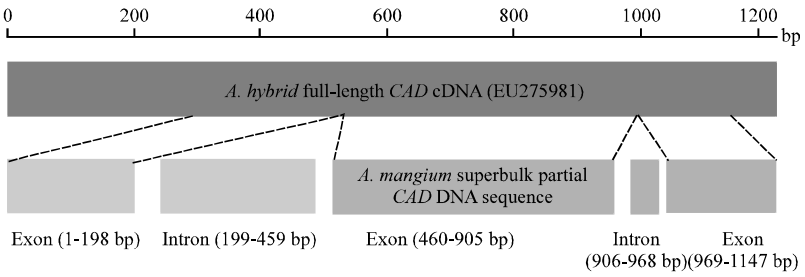

Fig. 3: The exonic and intronic regions of partial *CAD* DNA sequence. The partial *CAD* DNA sequence (1147bp) of *A. mangium* Superbulk was aligned with the full-length *CAD* cDNA from *Acacia* hybrid. The diagram indicates the predicted position of intron and exon region in partial *CAD* DNA sequence. Three exonic regions and two intronic regions were found within the partial *CAD* DNA sequence

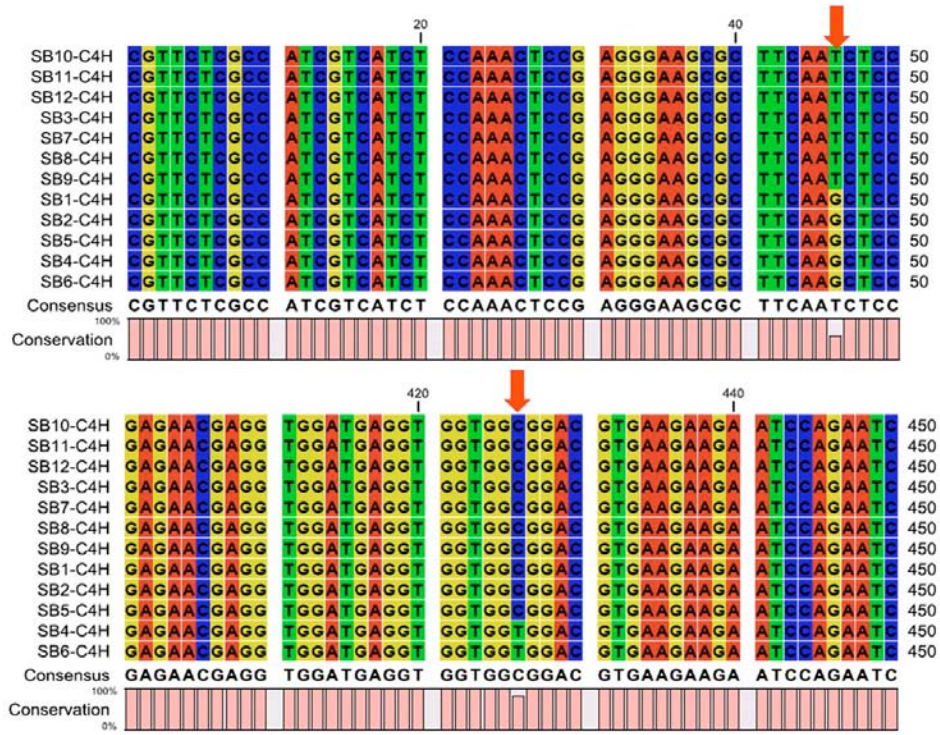

Fig. 4: Sequence variations within partial *C4H* DNA sequences. The arrow indicates the position and identity of SNPs in partial *C4H* DNA sequences. Two SNPs were detected from the multiple alignment of 12 partial *C4H* DNA sequences of *A. mangium* Superbulk (Gene accession number: SB1-C4H: HQ848681, SB2-C4H: HQ848682, SB3-C4H: HQ848683, SB4-C4H: HQ848684, SB5-C4H: HQ848685, SB6-C4H: HQ848686, SB7-C4H: HQ848687, SB8-C4H: HQ848688, SB9-C4H: HQ848689, SB10-C4H: HQ848690, SB11-C4H: HQ848691, SB12-C4H: HQ848692)

Table 3: The distribution of SNPs in partial *C4H* and *CAD* genes

Region	Gene	Position (bp)	Nucleotide changes	Trees
Exon	*C4H*	46	T → G	● SB1, SB2, SB4, SB5 and SB6
		426	C → T	● SB4 and SB6
	CAD	119	A → G	● SB9
		495	A → G	● SB9
		825	A → G	● SB1
		902	A → G	● SB6
		1039	A → G	● SB5
		1083	T → -	● SB1
Intron	*C4H*	-	-	-
	CAD	399	G → T	● SB3, SB4, SB9 and SB12
		416	C → T	● SB4

Table 4: The number of SNPs in partial *C4H* and *CAD* genes

Characteristic	*C4H*	*CAD*
Total Number of SNPs	2	8
Transition Mutation	1	6
Transversion Mutation	1	1
Deletion Mutation	-	1
Total Length of Partial Gene (bp)	457	1147
Occurrence of SNPs in Partial Gene	1 in every 229 bp	1 in every 143 bp

Sequence variations: Partial DNA sequences of *C4H* and *CAD* were aligned together by using CLC Free Main Workbench 4 (CLC Bio, Denmark) software, respectively in order to detect the SNPs. From the partial *C4H* alignment, 2 SNPs were found in the coding region of *C4H* (Fig. 4). Since all the sequences obtained fall within the exonic region, hence no SNP was found within the non-coding region. Among the 2 SNPs

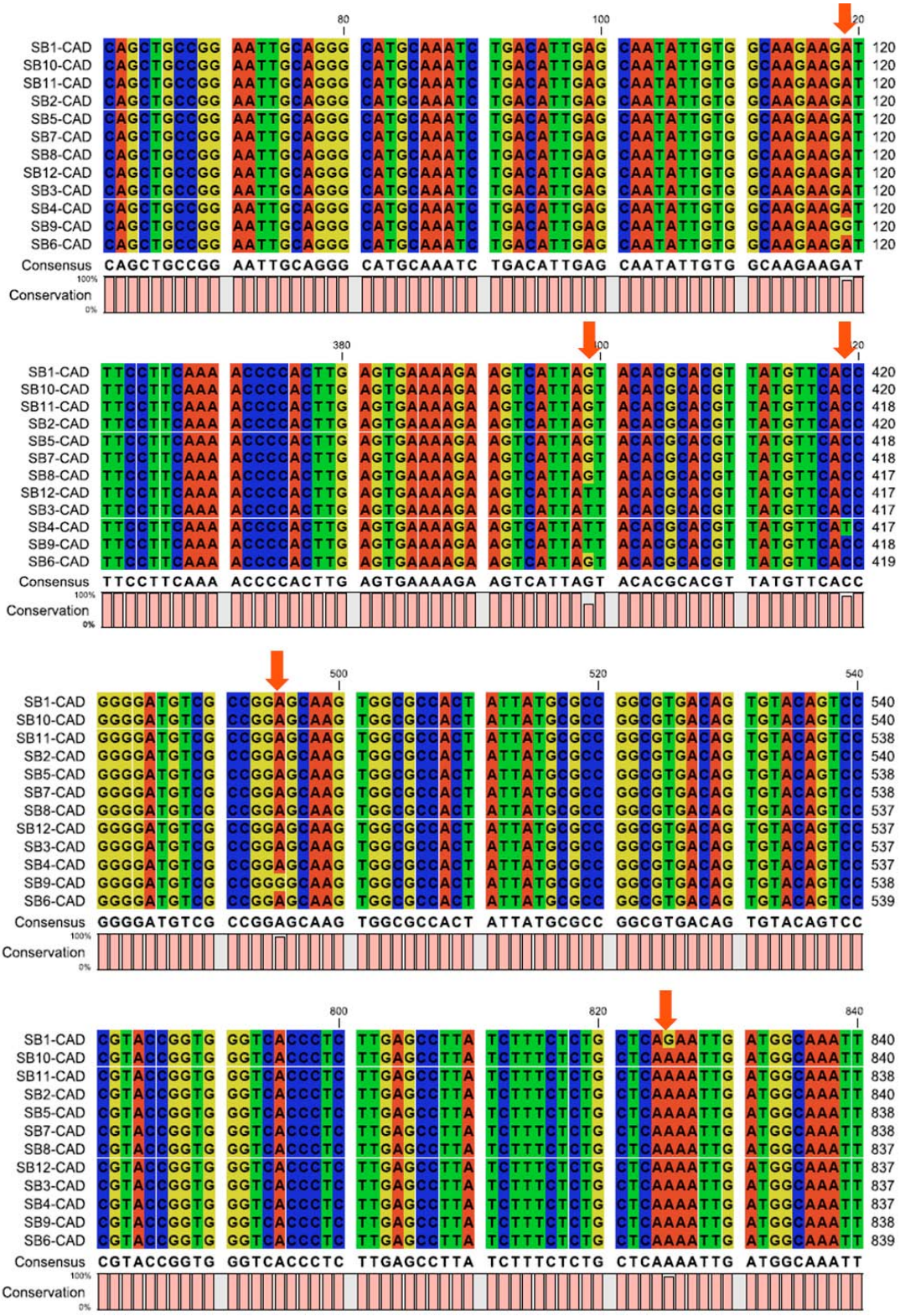

Fig. 5: Continue

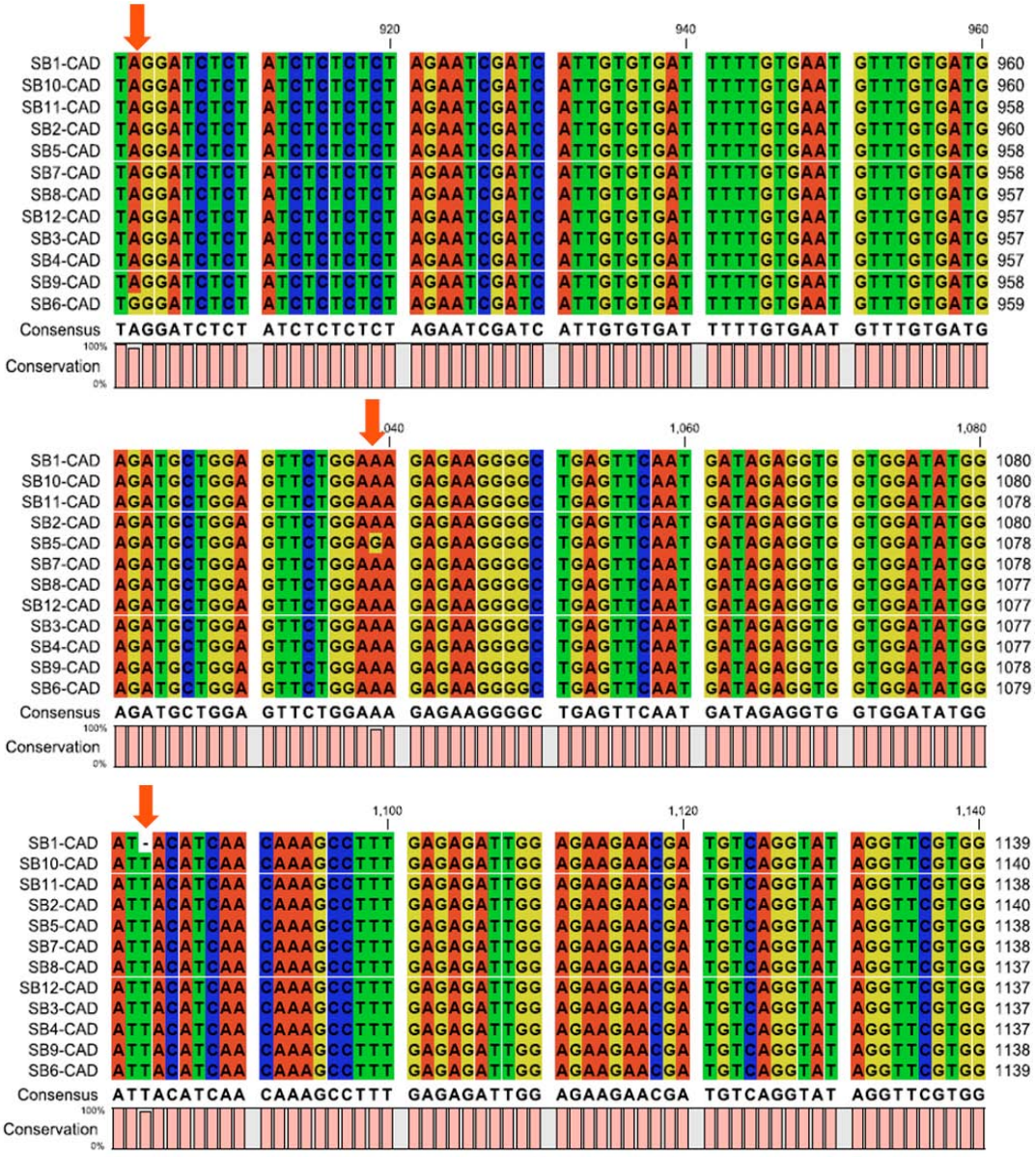

Fig. 5: Sequence variations within partial *CAD* DNA sequences. The arrow indicates the position and identity of SNPs in partial *CAD* DNA sequences. Eight SNPs were detected from the multiple alignment of 12 partial *CAD* DNA sequences of *A. mangium* Superbulk (Gene accession number: SB1-CAD: HQ848669, SB2-CAD: HQ848670, SB3-CAD: HQ848671, SB4-CAD: HQ848672, SB5-CAD: HQ848673, SB6-CAD: HQ848674, SB7-CAD: HQ848675, SB8-CAD: HQ848676, SB9-CAD: HQ848677, SB10-CAD: HQ848678, SB11-CAD: HQ848679, SB12-CAD: HQ848680)

detected, one was resulted from transition mutation while the other one was resulted from transversion mutations.

For partial *CAD* alignment, a total of 8 SNPs were found where 5 of the SNPs as well as an indel mutation were located in the exonic region and another 2 SNPs were located in the intronic region (Fig. 5). Among the SNPs detected, 6 were resulted from transition mutation and one was resulted from transversion mutations. The distribution and occurrence of the SNPs are summarized in Table 3 and 4.

Table 5: Nonsynonymous mutations in partial DNA sequences of *C4H*

Amino acid position	Nucleotide changes	Amino acid alteration
15	T → G	N → K
142	C → T	A → V
Total number of nonsynonymous mutations		2
Total number of amino acid		152
Proportion of nonsynonymous, dN		0.013

Table 6: Synonymous and nonsynonymous mutations in partial DNA sequences of *CAD*

Amino acid position	Nucleotide changes	Amino acid alteration
40	A → G	I → V
78	A → G	E → G
188	A → G	K → R
214	A → G	R → G
238	A → G	E → E
253	T → -	L → Y
Total number of nonsynonymous mutations		5
Total number of synonymous mutations		1
Total number of amino acid		271
Proportion of nonsynonymous, dN		0.018
Proportion of synonymous, dS		0.004

On average, one SNP occurred at every 229 bp in the partial sequence of *C4H* and at every 143 bp in the partial sequence of *CAD* (Table 4). Since the full length of *C4H* and *CAD* genes have not yet discovered, the comparison of SNP frequency to other studies of DNA sequence diversity cannot be done. However, it had been reported that the occurrence of SNPs in *C4H* of *A. mangium*×*A. auriculiformis* is one SNP in every 53 bp (Nur Fariza *et al*., 2008). Meanwhiles, other studies also had reported on the SNP frequency such as in barley (1 SNP in every 189 bp) (Kanazin *et al*., 2002), grapes (1 SNP in every 0.1 kb) (Velasco *et al*., 2007) and soybean (1 SNP in every 273 bp) (Zhu *et al*., 2003). On the other hand, the occurrence of SNP in *CAD* was one SNP in every 147 bp for *E. globules* (Poke *et al*. 2003) and one SNP in every 40 bp for *A. mangium*×*A. auriculiformis* (Nur Fariza *et al*., 2008).

For synonymous and nonsynonymous mutations study, all exonic sequences were translated into open reading frames and then aligned among 12 individuals trees using CLC Free Main Workbench 4 (CLC Bio, Denmark) software. For *C4H*, 2 nonsynonymous mutations were detected within the open reading frame of *C4H* and no synonymous mutation was identified (Fig. 6 and Table 5). Meanwhile, 5 nonsynonymous mutations and one synonymous mutation were detected within the open reading frame of *CAD* (Table 6). The single indel mutation at the exon region of *CAD* sequences had led to the incorporation of 3 stop codons in all amino acid sequences, except SB1 (Fig. 7). This indicates that the indel mutation detected in *CAD* gene may regulate the genetic variation of wood properties or phenotypic differences among trees. For instance, indels found in the exons of the *Arabidopsis* accessions Eil-0 and Lc-0 caused drastic effects on gene integrity, specifically on the gene representing expression level polymorphisms (Plantegenet *et al*., 2009). However, decision on the popularity of synonymous and nonsynonymous mutations can only be made when full length sequences of *C4H* and *CAD* genes are studied. Synonymous mutation does not lead to changes in the amino acid sequence of resulting protein while nonsynonymous mutations do. Nonsynonymous SNPs which cause changes to the protein sequence can affect the structural, functional or biochemical properties of the enzyme being produced (Bromberg and Rost, 2007) and therefore possibly lead to the changes in phenotypic characteristic of the trees, especially in modification of lignin biosynthesis.

Wood properties: Table 7 shows the mean values of fiber-length, cell wall thickness, microfibril angle, specific gravity and wood density. As reviewed by Plomion *et al*. (2000) in *Eucalypytus* and *Pinus*, there are strong relationships between growth strain and wood characteristics, such as wood density, longitudinal and tangential shrinkages, longitudinal modulus of elasticity, anatomical characteristics (microfibril angle) and chemical composition (lignin content, ratio of lignin monomeric units, hemicelluloses and cellulose content). Fibre wall thickness is closely related with wood density which in turn is probably the single most important predictor of wood properties such as strength, hardness, dimensional stability and others. High quality timbers which are heavy, hard and very strong, will have very thick-walled fibres and it is usually not easy to work

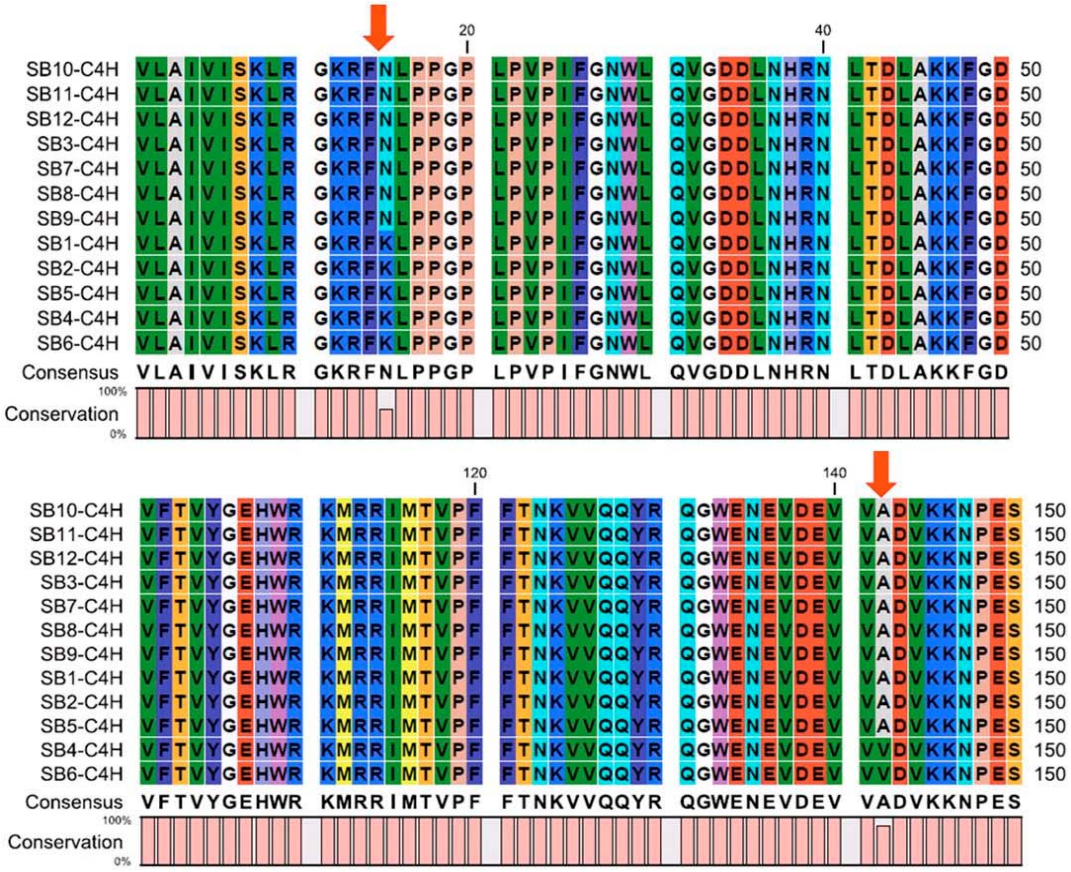

Fig. 6: Synonymous and nonsynonymous mutations in partial *C4H* amino acid sequence. The arrow indicates the position and identity of two nonsynonymous SNPs in predicted partial *C4H* amino acid sequences

Table 7: Wood properties data obtained for each *A. mangium* Superbulk sample

Sample	Fiber-length (μm)	CWT (μm)	MFA (°)	SG	WD (kgm^{-3})
SB1	1006.80	3.595	13.874	0.477	704.43
SB2	1030.06	3.040	12.882	0.384	607.93
SB3	983.88	2.670	13.736	0.314	439.73
SB4	983.18	2.220	14.480	0.357	713.02
SB5	905.60	1.865	14.390	0.364	599.62
SB6	808.55	2.320	13.576	0.205	224.01
SB7	847.69	2.620	15.964	0.383	832.44
SB8	976.03	2.625	13.968	0.306	560.85
SB9	864.96	2.570	14.620	0.279	418.93
SB10	865.62	1.835	14.264	0.434	600.28
SB11	925.62	2.615	14.730	0.310	547.39
SB12	928.89	2.565	15.256	0.305	903.26

SG: Specific gravity, WD: Wood density, CWT: Cell wall thickness, MFA: Microfibril angle

with (Baas and Wheeler, 2000). The thickness of the cell walls is increased by changing the microfibril orientation continually (Higuchi, 1997). Therefore, the microfibril angle is highly correlated with the cell wall thickness. Micofibril angle is also closely related to the mechanical properties of wood and pulp fibers since low microfibril angle will lead to high tensile strength and stiffness (Lima *et al*., 2004). Other than that, Bonham and Barnett (2001) observed that low microfibril angles are also associated with long fiber lengths. However, no significant correlation was observed among the physical wood property traits measured in this study.

Association genetics analysis: The association between the SNPs discovered and phenotypic variations on 12 *A. mangium* Superbulk samples were tested using

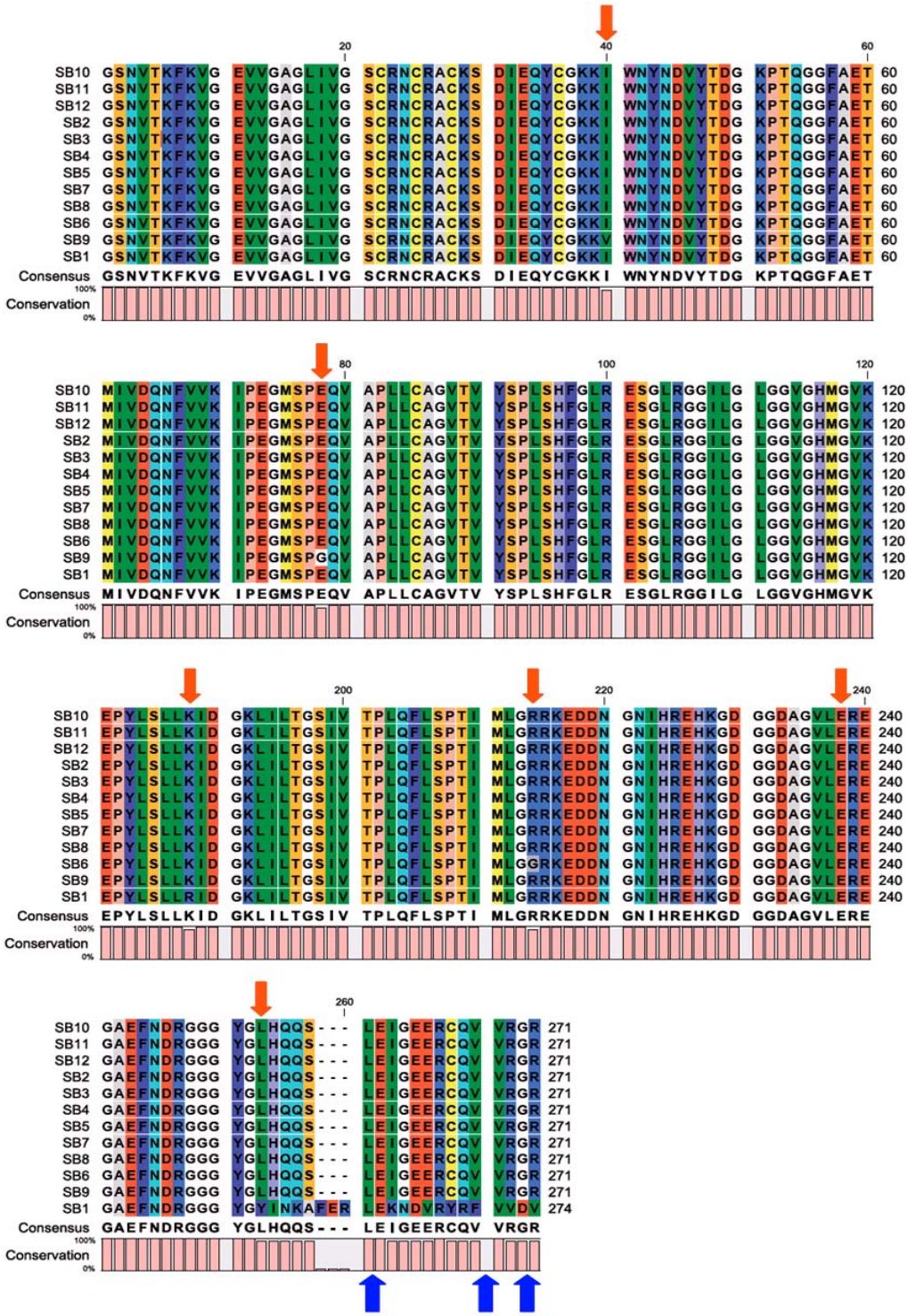

Fig. 7: Synonymous and nonsynonymous mutations in partial *CAD* amino acid sequence. The red arrow indicates the position and identity of nonsynonymous and synonymous SNPs and the blue arrow indicates the positions of stop codons at the end of all predicted amino acid sequences, except SB1

Table 8: Significant association detected between marker in *CAD* gene and wood properties

		Marker effect		
Trait	SNP	F	P	r^2
SG	M5	5.0755	0.0479	0.3367
	M6	5.5964	0.0396	0.3588
	M8	5.0755	0.0479	0.3367
WD	M6	6.7770	0.0263	0.4039
CWT	M5	9.2438	0.0125	0.4804
	M8	9.2438	0.0125	0.4804

SG: Specific gravity, WD: Wood density, CWT: Cell wall thickness

General Linear Model (GLM) provided by TASSEL software. Genetic association study revealed that nonsynonymous SNPs in *CAD* gene do associate with the specific gravity and wood density ($p<0.05$) (Table 8). These findings were similar with the previous studies done by Gonzalez-Martinez *et al.* (2007) and Yu *et al.* (2006), respectively. Specific gravity and wood density are tightly link to most of the mechanical properties of the wood. The strength, stiffness, heat transmission and some other wood properties are increased with specific gravity. The most interesting is that, the yield of the pulp per unit volume is also directly related to specific gravity. This give sense that the specific gravity or wood density should be the first wood property to be sincerely investigated (Haygreen and Bowyer, 1996).

Moreover, significant association ($p<0.05$) was also found for nonsynonymous SNPs in *CAD* gene with fiber cell wall thickness (Table 8). But the associations of SNPs in *CAD* gene with fiber-length and microfibril angle were not significant. Thickening of cell wall is affected by the arrangement of biopolymer aggregates which comprise of cellulose, hemicellulose and lignin. While the concentration of lignin is as a proportion of the cell wall and decrease when come closer to the lumen, therefore single nucleotide mutation in *CAD* gene might alter the lignin biosynthesis and thus, lead to changes in cell wall thickness as well as the wood density and specific gravity.

On the other hand, no significant results were obtained for SNPs in *C4H* gene with all wood properties studied. Although association was found among *CAD* and wood properties, the possibility of the false positive result that may arise due to the structure of population cannot be ignored (Yu *et al.*, 2006), particularly when a small sample size was used (Yu and Buckler, 2006). Therefore, it is very important to test for population structure in the future association study by using an independent set of random markers. Furthermore, only partial genes were screened for the presence of SNP in this study, therefore some of the genetic associations involving the partially screened genes may not detected. Hence, full-length sequencing of the candidate genes, including the promoter, intron, exon and 5'/3'-untranslated regions is advisable for identifying sufficient SNPs that can represent the candidate gene locus (Zhu *et al.*, 2008; Lau *et al.*, 2009).

The sample size of *Acacia* population should be increased to obtain more powerful data for association mapping in future. Long and Langley (1999) had proposed on the order of 500 samples to overcome the problem of low repeatability of an association study. Nowadays, effort has been placed to discover SNP in a small sample size (10-100) and then genotype in large association population by using low cost and high throughput SNP genotyping platform (Neale, 2007; Syvanen, 2001).

In conclusion, the present study on gene-associated single nucleotide polymorphism in *C4H* and *CAD* genes from *Acacia mangium* Superbulk trees revealed an association between SNPs in *CAD* gene with the wood density, specific gravity and cell wall thickness. This indicates that SNP in *CAD* gene might alter the lignin biosynthesis and thus, lead to changes in phenotypic characteristics of the trees. Once these SNPs have been validated, it could potentially be used as a tool in marker-assisted selection that enables more precise and accurate in selection and prediction of yield or performance early in life such as at the seedling stage.

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 work is part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation (SFC) and Universiti Malaysia Sarawak (UNIMAS) and Ministry of Sciences, Technology and Innovation (MOSTI), Malaysia.

REFERENCES

Abbasi, S., C. Azimi, F. Othman, M.R.N. Daloii and Z.O. Ashtiani *et al.*, 2009. Estrogen receptor-gene codon 10 (T392C) polymorphism in Iranian women with breast cancer: A case study. Trends Mol. Sci., 1: 1-10.

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.

Aina, O.O., A.G.O. Dixon and E.A. Akinrinde, 2007. Trait association and path analysis for cassava genotypes in four agroecological zones of Nigeria. J. Boil. Sci., 7: 759-764.

Alashawkany, A.R., F.E. Shahroudi, M.R. Nassiry, A.H. Moussavi, M. Heydarpour and B. Sadeghi, 2008. Association of SNP in the ExonII of leptin gene with milk and reproduction traits in Holstein Iranian cows. Biotechnology, 7: 347-350.

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.

Baas, P. and E.A. Wheeler, 2000. Wood Structure of Southeast Asian timbers-The PROSEA Woods Reviewed. In: New Horizons in Wood Anatomy: Proceedings of the 4th Pacific Regional Wood Anatomy Conference, Kim, Y.S. (Ed.). Chonnam National University Press, Kwangju, Korea, pp: 1-9.

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.

Bonham, V.A. and J.R. Barnett, 2001. Fibre length and microfibril angle in silver birch (*Betula pendula* Roth). Holzforschung, 55: 159-162.

Brett, C. and K. Waldron, 1990. Physiology and Biochemistry of Plant Cell Walls. Unwin Hyman, London.

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

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

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., 8: 7-7.

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.

Haygreen, J.G. and J.L. Bowyer, 1996. Forest Products and Wood Science. 3rd Edn., Iowa State University Press, Iowa State, USA., pp: 102-108.

Higuchi, T., 1997. Biochemistry and Molecular Biology of Wood. Springer-Verlag, Berlin, New York.

Ho, W.S., S.L. Pang, L. Pauline and J. Ismail, 2011. Sequence variation in the *cellulose synthase* (*SpCesA1*) gene from *Shorea parvifolia* ssp. *parvifolia* mother trees. Pertanika J. Trop. Agric. Sci., 34: 323-329.

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.

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.

Lim, S.C., K.S. Gan and K.T. Choo, 2003. The characteristics, properties and uses of plantation timbers-rubberwood and *Acacia mangium*. Timber Technol. Bull., 26: 1-11.

Lima, J.T., M.C. Breese and C.M. Cahalan, 2004. Variation in microfibril angle in *Eucalyptus* clones. Holzforschung, 58: 160-166.

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.

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 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.

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.

Pang, S.L., C.Y. Choong and R. Wickneswari, 2005. Isolation and characterisation of *C4H*, *CCoAOMT* and *CAD* from interspecific hybrid *Acacia mangium* x *Acacia auriculiformis*. Proceeding of *Acacia* research in Malaysia: Seminar on Current Updates on *Acacia* Genomes and Breding, July 12, Marriotte Hotel, Putrajaya, Malaysia, pp: 7-10.

Plantegenet, S., J. Weber, D.R. Goldstein, G. Zeller and C. Nussbaumer *et al*., 2009. Comprehensive analysis of *Arabidopsis* expression level polymorphism with simple inheritance. Mol. Syst. Biol., 5: 252-252.

Plomion, C., C. Pionneau, J. Brach, P. Costa and H. Bailleres, 2000. Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol., 123: 959-969.

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.

Rafalski, J.A., 2002. Novel genetics mapping tools in plants: SNPs and LD-based approaches. Plant Sci., 162: 329-333.

Rashwan, A.M.A., 2011. Study of genotype and phenotypic correlation for some agro-economic traits in Okra (*Abelmoschus esculents* (L.) Moench). Asian J. Crop Sci., (In Press).

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

Thumma, B., R., M.F. Nolan, R. Evans and G.F. Moran, 2005. Polymorphisms in *Cinnamoyl CoA Reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics, 171: 1257-1265.

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, 2: e1326-e1326.

Wickneswari, R., J.A. Harikrishna and C.Y. Choong, 2005. Analyses of ESTs generated from Inner bark tissue of *Acacia* Hybrid. *Acacia* research in Malaysia. Proceeding of the Seminar on Current Updates on *Acacia* Genomes and Breeding, July 12, Putrajaya, Malaysia, pp: 1-6.

Yu, J. and E.S. Buckler, 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol., 17: 155-160.

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.

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.