Available online at www.ijrp.org

International Journal of Research Publications

Isolation and nucleotide variation of COBRA gene in a tropical timber tree Neolamarckia cadamba

Sing-Woei Bong[a], Wei-Seng Ho[a]*, Shek-Ling Pang[b]

[a]Forest Genomics and Informatics Laboratory (fGiL), Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak

[b]Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, 93250 Kuching, Sarawak

*Corresponding Authors: e-mail: wsho@unimas.my

Abstract

COBRA gene is involved in the regulation of cellulose deposition and orientation of cell expansion. Any single nucleotide differences may influence the functions of this gene. Hence, this study was carried out to isolate the partial COBRA genomic sequence and subsequently, to determine the nucleotide variation of COBRA gene in Neolamarckia cadamba. N. cadamba, also locally known as kelampayan is one of the fast-growing deciduous and indigenous tree species with great commercial values in Malaysia. The targeted DNA sequence of COBRA gene was amplified with the designed primer pair by using Polymerase Chain Reaction (PCR) technique. The partial COBRA genomic sequence (517 bp) was subjected to BLASTn analysis to search for homology sequence and validate the identity of the sequence through NCBI. Multiple alignment was carried out by ClustalW for manual detection of single nucleotide polymorphisms (SNPs). Five SNPs were detected in the exon region and two SNPs in the intron region of COBRA partial genomic sequences. Of these five detected SNPs, four non-synonymous mutations and one synonymous mutation were discovered in the COBRA amino acid sequences. Based on the in silico restriction analysis, one possible restriction enzyme, HpyCH4III was detected to restrict at a SNP site (384 bp) which could be useful for genetic marker development, such as CAPS marker development in the efforts of genotyping project.

Keywords: Neolamarckia cadamba, polymerase chain reaction (PCR), COBRA gene, single nucleotide polymorphisms (SNPs), molecular markers

1. Introduction

Plant cell wall plays a vital role in human life as it provides large scale of dietary fibers and raw materials for textile, lumber, pulping, and potentially for biofuels. In economic point of view, understanding the regulatory mechanisms involved for biosynthesis of cell wall and identifying regulatory genes such as cellulose, hemicelluloses and lignin are of obvious important (Zhang et al., 2009). Cellulose that composes 40% to 50% of wood is an essential component of the load-bearing network by providing strength required in cell walls through cellulose deposition. The physical properties of cellulose also determine the orientation of cell expansion (Brown, 2003). The presence of COBRA gene in the woody plant species is essential as it plays the role in the regulation of cellulose deposition and orientation of cell enlargement.

Identification of COBRA gene and sequence polymorphism controlling wood quality traits are vital especially in forest tree breeding programmes. An improvement of economically important traits such as pulp yield can produce large gains for a pulp and paper mill industry. Therefore, the present study was the first molecular study to isolate the COBRA genomic sequence and to determine the nucleotide variation of COBRA gene in Neolamarckia cadamba. N. cadamba, also locally known as kelampayan is one of the fast-growing deciduous and indigenous tree species for areas from India through Southeast Asia to Papua New Guinea. Due to the fast growth rate and great economic value of its timber, plantations of N. cadamba have been established in India, Sri Lanka, Myanmar, Indonesia, Malaysia and the Philippines. It has been proven as one of the best raw materials for the plywood industry (Lai et al., 2013; Ho et al., 2014; Tiong et al., 2014a,b,c&d; Phui et al., 2014; Sim et al., 2014; Pang et al., 2015). The leaves and barks have been extensively studied and reported to have high medicinal values (Joker, 2000; Patel and Kumar, 2007; Zaky et al., 2014a&b). It grows well in freshwater swamps and it is commonly found in logging area of lowland dipterocarp forests (Nair, 2007).

2.0 Materials and Methods

2.1 Plant materials and DNA isolation

Fresh young leaves were collected from six randomly selected N. cadamba trees from the Kelampayan Trial Plot at Landeh Nature Reserve, Semengok, Sarawak. Total genomic DNA of N. cadamba was extracted by using the modified CTAB method from Doyle and Doyle (1990). The isolated DNA was purified by using Wizard® Genomic DNA Purification Kit (Promega, USA) based on the manufacture's protocol. The concentration of purified DNA was quantified using Lambda 25 UV/VIS Spectrophotometer (Perkin Elmer, USA).

2.2 Primer Design

One primer pair for COBRA gene was designed based on the EST sequence of N. cadamba (536 bp) by using Primer Premier 6.0 (PREMIER Biosoft International, USA). Based on the primer pair designed, the pair with highest pair rating was chosen. The sequence for forward primer was 5'-GGCGTGTTAAGATTGCTGTTAC-3' and the reverse primer was 5' GTATGAATCAGGAGGCGGAAG-3'.

2.3 Polymerase Chain Reaction (PCR)

PCR amplification was done in a Gradient Palm-Cycler™ (Corbett Research, Australia) for 2 minutes at

95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 50°C ± 5°C and 1 minutes at 72°C. It is then followed by final extension of 10 minutes at 72°C. PCR reaction mixture was 30 ng of DNA template, 5 pmol of forward and reverse primers, 0.2 mM dNTPs, 1 x PCR buffer, 1.5 mM $MgCl_2$ and 1 U Taq DNA Polymerase (Invitrogen, USA). Ultra-pure water was used to top up mixture to 25 µl. The PCR products were checked using 1.5% agarose gel and run simultaneously with 100bp DNA ladder (Promega, USA). The verified PCR products were cloned into the pGEM®-T Easy Vector System (Promega, USA) and the recombinant plasmids were isolated and purified using the Wizard® Plus SV Minipreps DNA Purification System (Promega, USA) according to the manufacture's protocol.

2.4 DNA sequencing and analysis

The purified plasmids were sent for sequencing in both forward and reverse directions. The sequencing was performed using an ABI 3730XL capillary DNA sequencer (Applied Biosystems, USA). The DNA sequence was checked using BLASTn (www.ncbi.nlm.nih.gov) to search for homology sequence in GenBank (www.ncbi.nlm.nih.gov) and to validate the identity of the DNA sequence. Then the DNA sequences were subjected to alignment using ClustalW software (http://www.ebi.ac.uk/Tools/clustalw2/index.html) to check for consensus sequence. From the alignment, differences in nucleotide among the sequences were manually checked. Further analysis of the DNA sequences was carried out with bl2seq (www.ncbi.nlm.nih.gov) by comparing the DNA sequence with other species to check for intron-exon boundaries. The identified intron region was then removed manually and checked again for validation using BLASTp (www.ncbi.nlm.nih.gov). Detection of non-synonymous and synonymous mutations was also conducted by translating the consensus sequences into protein sequences by using ExPASy translate tool (http://expasy.org/tools/dna.html). The translated protein sequences were aligned by using ClustalW software. In silico restriction of the consensus SpCesA1 sequences was done using NEBcutter V2.0.

3.0 Results and Discussion

The partial COBRA genomic sequence was successfully isolated from the N. cadamba trees. From the BLASTn analysis, the 517 bp COBRA gene sequence showed moderate degree of similarity with COBRA and COBRA-like genes from Populus tomentosa (68%, GU053565), Eucalyptus nitens (67%, FJ213604) and Ricinus communis (65%, XM_002514582). It was observed that the amplicons have coding and non-coding regions. Further analysis on the intron-exon boundaries was carried out with bl2seq (www.ncbi.nlm.nih.gov) by comparing the sequenced COBRA gene with other species such as Populus tomentosa. A gap around 162 bp between two consensus sequences was observed from the alignment result with Populus tomentosa DNA (GU053565) and mRNA (GQ410777) (Figure 1).

In Arabidopsis, AtCOBL4 is identified to be implicated in the cellulose biosynthesis in the secondary wall through mutant analysis (Brown et al., 2005). The expression of AtCOBL4 is found to be similar with Arabidopsis cellulose synthase (AtCesA) genes that are involved in secondary cell wall synthesis (Persson et al., 2005). In a study done by Zhang et al. (2009), a COBRA-like protein (PtCOBL4) was successfully isolated and identified in poplar when comparison of nucleotide sequence was conducted with the known full-length of Arabidopsis COB gene sequences. It is confirmed as one of the members of COBL family, PtCOBL4 shared 72.7%, 64.8%, and 65.3% similarity with Arabidopsis COBL members, rice BC1 members and maize BK2 proteins respectively. PtCOBL4 is suggested to be an ortholog of AtCOBL4 as they shared high nucleotide homology (70.9%) and amino acid identity (72.7%). Other than that, PtCOBL4 gene is found to be expressed dominantly in the mature xylem fiber cells during the late stage of cell wall thickening. This further shows that PtCOBL4 is involved in the secondary cell wall deposition in the stem of trees (Zhang et al.,

2009). According to Roudier et al. (2005), Arabidopsis mutants of COBRA gene has shown to reduce cellulose content and cause the cells to swell due to disorganize cellulose microfibril deposition. However, the disorganization of cellulose occurs before the depletion of cellulose content in the cell wall. This shows that COBRA is involved in the orientation of deposition of cellulose microfibrils while the reduction in cellulose content is likely caused by a feedback mechanism that reflects the already disordered deposition of cellulose microfibrils.

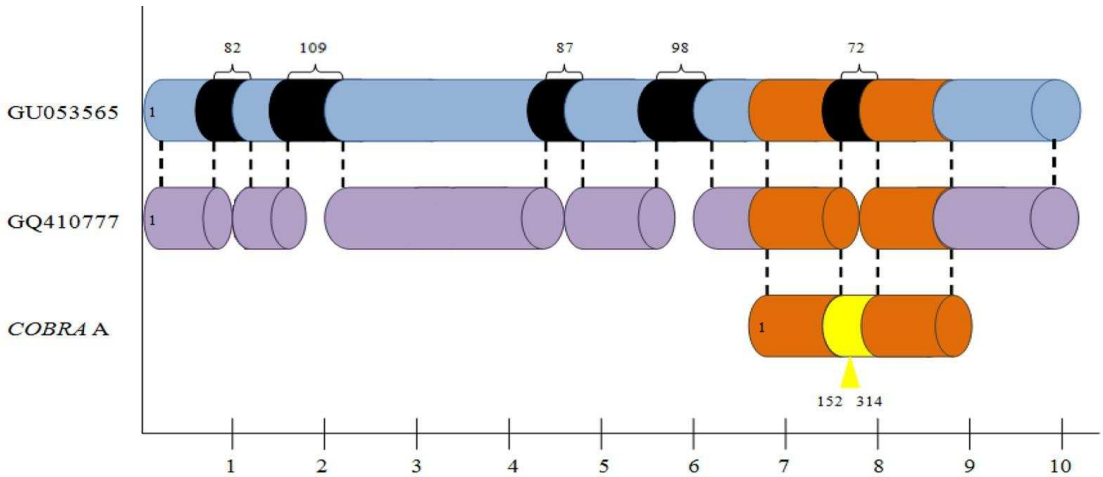

Figure 1 Comparison of sequence structure of Populus tomentosa genotype 38 COBRA-like 4 protein DNA (GU053565), Populus tomentosa COBRA-like 4 protein mRNA (GQ410777) and COBRA partial genomic sequence from sample A (COBRA A). Coloured cylinders represent the exon region and black coloured cylinders represent the intron region with number of bases indicated below them. Intron portion as predicted from 517 bp COBRA sequence is shown in solid yellow. The dotted lines connecting different genes are the conserve intron-exon junctions. 1 unit axis: 200 bp.

ClustalW software was used to detect the single nucleotide polymorphism (SNP) from the six partial genomic sequences of COBRA gene. From the sequence alignment result, a total number of seven putative SNPs were detected. Based on the intron-exon boundaries, two putative SNPs were in the intron region while the other SNPs were in the exon regions (Figure 2).

The identified intron region was manually removed, and the predicted exon region was translated into amino acid sequence by using ExPASy translate tool. The resulted amino acid sequence was then subjected to Protein BLAST for further validation. The protein sequence showed high and moderate degree of similarity with COBRA and COBRA-like protein with Oryza sativa (100%), Arabidopsis thaliana (100%), Rhizobium leguminosarum (84%) and Phoneutria keyserlingi (82%). Out of the five SNPs located in the exon regions, four of them showed the occurrence of non-synonymous mutations (R (A**G**A) → K (A**A**A); T(**A**CA) → A (**G**CA); L (**C**TT) → F (**T**TT); P (C**C**T) → H (C**A**T)) meanwhile a synonymous mutation (P (CC**A**) → P (CC**G**)) occurred in one of the SNPs (Figure 3). According to Sloan and Taylor (2010), synonymous mutations are sites where changes of base occur, but amino acid sequence remains the same due to the redundancy of the genetic code. It is often assumed that this type of mutation is relatively neutral, and it is used as a basis of estimating mutation rates. As for non-synonymous mutation of SNP, it occurs in the coding region that changes amino acid sequence in the corresponding protein product and affects the phenotype of the host organism (Kumar et al., 2009).

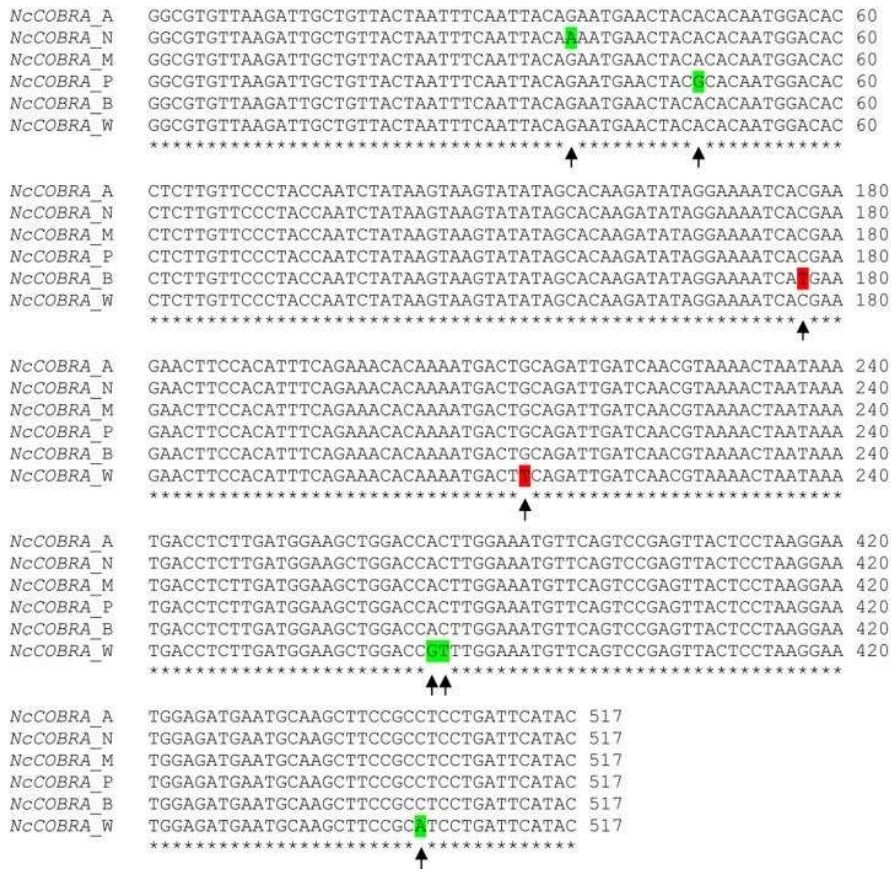

Figure 2 Alignment result of six partial sequences of COBRA gene using ClustalW software. Five SNPs were detected in the exon regions at nucleotide numbers 37, 48, 385 and 386 (green). Two putative SNPs were in the intron region at nucleotide numbers 177 and 212 (red).

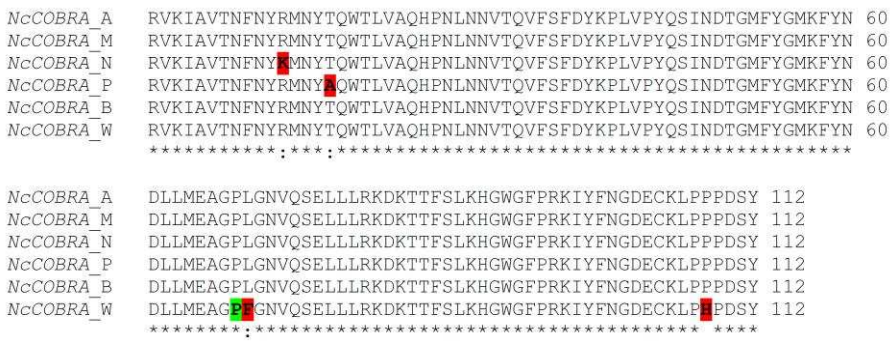

Figure 3 Alignment of the protein sequences using ClustalW software. Four non-synonymous mutations were detected at number 12 (R/K), 16 (T/A), 69 (L/F) and 108 (P/H) (red) and one synonymous mutation detected at number 68 (P/P) (green) of the COBRA amino acid sequence.

With synonymous (Ks) and non-synonymous (Ka) mutation ratio, the type of selection that acts on the gene can be determined (Stukenbrock and Mcdonald, 2007; Berglund et al., 2005). According to Berglund et al. (2005), the ratio of synonymous and non-synonymous mutation has been established as a valuable method to assay selective pressures on protein encoding gene. The neutral theory of molecular evolution proposed that neutral evolution happens when Ks = Ka. Positive diversifying selection occurs when Ks < Ka while purifying selection happens when Ks > Ka. Positive diversifying is a natural selection process that fixes adaptive variants and enhances the frequency of the adaptive allele relative to others while purifying selection is a natural selection process involving the removal of deleterious mutations from the population (Nei et al., 2010). Based on Figure 3, it is likely that positive diversifying occurs as the ratio of synonymous mutation and non-synonymous mutation is Ks < Ka. This is due to the function of COBRA gene in the deposition of cellulose microfibrils is crucial in the development of secondary cell wall. However, no association can be made in accordance to the discovered SNP in COBRA gene and physical characteristics of the kelampayan trees. More polymorphic studies are needed to be carried out to identify whether COBRA gene is subjected to positive selection.

One possible restriction enzyme was detected for one SNP site of the partial COBRA genomic sequence by using NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/index.php). The restriction enzyme found was HpyCH4III (5'-ACNGT-3') which recognize and restrict at site number 384 bp of NcCOBRA-W sample (Figure 4). HpyCH4III produces two fragments at approximately 133 bp and 384 bp. According to Morgan and Xu, (2000), this restriction endonuclease is obtainable from Helicobacter pylori CH4 which is then referred to as HpyCH4III. With the identification of restriction enzyme for the detected SNP, it can be utilized in the development of cleaved-amplified polymorphic sequence (CAPS) marker.

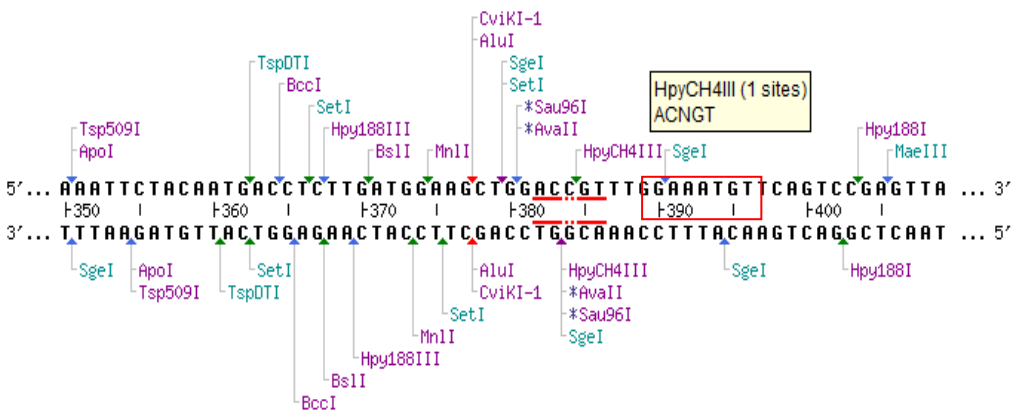

Figure 4 A possible restriction enzyme for SNP site from NcCOBRA_W using NEBcutter V2.0. HpyCH4III has been identified to recognize and restrict a SNP site (ACNGT) of NcCOBRA-W at 384 bp.

4.0 Conclusion

To best of our knowledge, this is the first report on isolation and sequence polymorphism of COBRA gene in Neolamarckia cadamba through PCR amplification and DNA sequencing. A total of seven SNPs was detected in the partial genomic sequence of COBRA gene, and four non-synonymous mutations and one

synonymous mutation were discovered in the COBRA amino acid sequence. Other than that, a restriction enzyme, HpyCH4III was also identified to restrict at a SNP site. PCR-based marker is one of the robust and inexpensive methods in detecting SNP site in other samples of same species. In contrast to detecting DNA polymorphism by sequencing, SNPs can also be detected through the cleavage of amplified DNA with the chosen restriction endonuclease. The development of CAPS marker through detection of SNPs is essential in the efforts of genotyping project. Further studies on polymorphism of full length of COBRA gene are still required as this study compromises partial length of COBRA gene only. Besides that, study on sequence polymorphism of COBRA gene on larger sample can also be carried out. Association study of the phenotype with the SNPs can also be done with the identified SNPs. Through the identification of SNPs, genetic markers such as CAPS marker can be developed to identify advantageous alleles that assist in genetic improvement activities and selective breeding.

Acknowledgements

The authors would like to thank all the lab and research 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), Sarawak Timber Association (STA) and Universiti Malaysia Sarawak (UNIMAS) with Grant No. RACE/a(2)/884/2012(02) and GL(F07)/06/2013/STA-UNIMAS(06)).

References

Berglund, A. C., Wallner, B., Elofsson, A. and Liberles, D. A., 2005. Tertiary windowing to detect positive diversifying selection. Journal of Molecular Evolution. 60, 499-504.

Brown, D. M., Zeef, L. A. H., Ellis, J., Goodacre, R. and Turner, S. R., 2005. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. The Plant Cell. 17, 2281-2295.

Brown, R. M. Jr., 2003. Cellulose structure and biosynthesis: What is in store for the 21st century? Journal of Polymer Science. 42, 487-495.

Doyle, J. J. and Doyle, J. L., 1990. Isolation of plant DNA from fresh tissue. Focus. 12, 1241-1251.

Ho, W. S., Pang, S. L. and Julaihi, A., 2014. Identification and analysis of expressed sequence tags present in xylem tissues of Kelampayan (Neolamarckia cadamba (Roxb.) Bosser). Physiology and Molecular Biology of Plants. 20(3), 393-397.

Joker, D., 2000. Seed Leaflet Neolamarckia cadamba (Roxb.) Bosser (Anthocephalus chinensis (Lam.) A. Rich. ex Walp.). Danida Forest Seed Centre, 17. Available from: http://www.dfsc.dk

Kumar, P. Henikoff, S. and Ng, P. C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols. 4(8), 1073-1082.

Lai, P. S., Ho, W. S. and Pang, S. L., 2013. Development, characterization and cross-species transferability of expressed sequence tag-simple sequence repeat (EST-SSR) markers derived from Kelampayan tree transcriptome. Biotechnology. 12(6), 225-235.

Morgan, R. D. and Xu, Q., 2000. Type II restriction endonuclease, HpyCH4III, obtainable from Helicobacter pylori CH4 and a process for producing the same. United States patent number US 6133009.

Nair, K. S. S., 2007. Tropical forest of insect pests: Ecology, impact, and management. New York, USA: Cambridge University Press.

Nei, M., Suzuki, Y. and Nozawa, M., 2010. The neutral theory of molecular evolution in the genomic era. Annual Review of Genomics and Human Genetics. 11, 265-289.

Pang, S. L., Ho, W. S., Mat-Isa, M. N. and Julaihi, A., 2015. Gene discovery in the developing xylem tissue of a tropical timber tree species: Neolamarckia cadamba (Roxb.) Bosser (Kelampayan). Tree Genetics & Genomes. 11:47.

Patel, D. and Kumar, V., 2008. Pharmacognostical studies of Neolamarckia cadamba (roxb.) Bosser leaf. Int. J. Green Pharm. 2(1), 26-27

Persson, S., Wei, H., Milne, J., Page, G. P. and Somerville, C. R., 2005. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proceedings of the National Academy of Sciences. 102(24), 8633-8638.

Phui, S. L., Ho, W. S., Pang, S. L. and Julaihi, A., 2014. Development and polymorphism of simple sequence repeats (SSRs) in Kelampayan (Neolamarckia cadamba – Rubiaceae) using ISSR suppression method. Archives of Applied Science Research. 6(4), 209-218.

Roudier, F., Fernandez, A. G., Fujita, M., Himmelspach, R., Borner, G. H. H., Schindelman, G., Song, S., Baskin, T. I., Dupree, P., Wasteneys, G. O. and Benfey, P. N., 2005. COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. The Plant Cell. 17, 1749-1763.

Sim, W. Y. C. M., Ho, W. S. and Pang, S. L., 2014. Molecular cloning of hypervariable regions (HVRII) from cellulose synthase (CesA) gene in Neolamarckia cadamba. International Journal of Bioscience, Biochemistry and Bioinformatics. 4(6), 475-482.

Sloan, D. B. and Taylor, D. R., 2010. Testing for selection on synonymous sites in plant mitochondrial DNA: The role of codon bias and RNA editing. Journal of Molecular Evolution. 70, 479-491.

Stukenbrock, E. H. and Mcdonald, B. A., 2007. Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. Molecular Plant Pathology. 8(3), 321-332.

Tchin, B. L., Ho, W.S., Pang, S. L. and Ismail, J., 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(4), 303-315.

Tiong, S. Y., Chew, S. F., Ho, W. S. and Pang, S. L., 2014a. Genetic diversity of Neolamarckia cadamba using dominant DNA markers based on inter-simple sequence repeats (ISSRs) in Sarawak. Advances in Applied Science Research. 5(3), 458-463

Tiong, S. Y, Ho, W. S., Pang, S. L. and Ismail, J., 2014b. Bioinformatics analysis of xyloglucan endotransglycosylase/ hydrolase (XTH) gene from developing xylem of a tropical timber tree Neolamarckia cadamba. American Journal of Bioinformatics. 3(1), 1-16.

Tiong, S. Y., Ho, W. S., Pang, S. L. and Ismail, J., 2014c. In silico analysis of cellulose synthase gene (NcCesA1) in developing xylem tissues of Neolamarckia cadamba. American Journal of Bioinformatics. 3(2), 30-44.

Tiong, S.Y., Ho, W. S., Pang, S. L. and Ismail, J., 2014d. Nucleotide diversity and association genetics of xyloglucan endotransglycosylase/hydrolase (XTH) and cellulose synthase (CesA) genes in Neolamarckia cadamba. Journal of Biological Sciences. 14(4), 267-275.

Zaky, Z. M., Fasihuddin, B. A., Ho, W. S. and Pang, S. L., 2014a. GC-MS analysis of phytochemical constituents in leaf extracts of Neolamarckia cadamba (Rubiaceae) from Malaysia. Int J Pharm Pharm Sci. 6(9), 123-127.

Zaky, Z. M., Ho, W.S., Pang, S. L. and Fasihuddin, B. A., 2014b. EMS-induced mutagenesis and DNA polymorphism assessment through ISSR markers in Neolamarckia cadamba (Kelampayan) and Leucaena leucocephala (Petai Belalang). European Journal of Experimental Biology. 4(4), 156-163.

Zhang, D., Yang, X., Zhang, Z. and Li, B. (2009). Expression and nucleotide diversity of the poplar COBL gene. Tree Genetics and Genomics. 6, 331-344.