Resequencing And Nucleotide Variation Of Sucrose Synthase (Nmsusy1) Gene In A Tropical Timber Tree Neolamarckia Macrophylla

Choon-Ju Tan, Wei-Seng Ho, Shek-Ling Pang

Abstract: Sucrose synthase (SuSy) is a key enzyme that catalyses the reversible synthesis and degradation of sucrose. It provides greater impact in regulating the photosynthetic processes and environmental stresses in plants. Thus, the nucleotide variation of partial NmSusy1 genomic DNAs (750 bp) generated through PCR amplification was examined in this study, and this followed by resequencing from 15 selected Neolamarckia macrophylla clones. The consensus sequences were aligned to detect the presence of single nucleotide polymorphisms (SNPs). In total, five SNPs were detected at nucleotide 1, 2, 34, 35 and 397. Of these, four SNPs were located at the predicted coding region while one SNP was located at the predicted non-coding region. Interestingly, one single base pair InDel polymorphism was also detected at nucleotide 17. On average, one SNP at every 150 bp was detected based on the 15 NmSusy1 sequences. There was one non-synonymous mutation detected, whereby amino acid glutamic acid (E) was replaced by arginine (R) in one of the 15 samples tested. This non-synonymous SNP 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. Overall, this study has demonstrated that resequencing is an effective technique for classifying molecular diversity or nucleotide variation in the Susy gene of N. macrophylla. Those SNPs, once validated, could potentially be used as a tool in marker-assisted selection (MAS) that enables more precise and accurate in the selection and prediction of yield or performance at the early developmental stages, such as at the seedling stage.

Index Terms: Neolamarckia macrophylla, red kelampayan; resequencing, sucrose synthase (SuSy); single nucleotide polymorphism (SNP), nucleotide variation, marker-assisted selection

1 INTRODUCTION

Sucrose is important to the plant growth and development. It is the main photosynthate and serves as the main transport sugar, nutrients and potential signal molecule in plants. Sucrose synthase (SuSy) is a key enzyme that catalyses the reversible synthesis and degradation of sucrose. Since sucrose cannot be utilized by plant directly, sucrose and uridine diphosphate (UDP) need to be cleaved by SuSy enzyme into fructose and UDP-glucose at pH 8.0 to pH 8.8 whereby UDP-glucose serves as the precursor for cellulose biosynthesis, which is crucial for wood formation [1]. Hypoxic or anoxic condition on the other hand, will cause the falling of pH in cells (pH 6.0 to pH 6.5), leading to degradation of sucrose by SuSy enzyme into UDP-glucose and fructose [2]. Hence, under reduced oxygen condition, plants are still able to survive as SuSy enzyme will increase degradation of sucrose. *SuSy* gene is regulated by the level of its own enzyme products.

- *Dr Wei-Seng Ho - *Corresponding Author Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak. E-mail: wsho@frst.unimas.my / howeiseng@gmail.com*
- *Dr Shek-Ling Pang - Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, 93250 Kuching, Sarawak*
- *Choon-Ju Tan - Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak.*

It was reported that sucrose, glucose and D-mannose are able to up-regulate Sus1, which is a major *SuSy* gene in *Arabidopsis thaliana* [3]. It is proposed that SuSy enzymes consist of two forms, namely the soluble SuSy (S-Susy) and particulate SuSy (P-SuSy). Soluble SuSy can be found in the cytoplasm while particulate SuSy is bound on the plasma membrane [4]. S-SuSy allocates carbon for metabolic work such as respiration, producing storage and building blocks molecules. It cleaves sucrose into UDP-glucose and fructose whereby fructose is used in glycolysis and starch formation [5]. On the other hand, P-SuSy supplies UDP-glucose from the degradation of sucrose for cellulose and callose synthesis. Polysaccharide is produced by using the high energy bond conserved during the process [6]. The activation of SuSy is thought to be related to adaptation to anaerobic condition. In maize roots, SuSy acts as anaerobic proteins under anaerobic conditions to degrade sucrose. It requires less ATP and thus more energy-efficient [5]. Phosphorylation enabled the conversion of P-SuSy into S-SuSy [1]. Apart from catalysing sucrose, SuSy is suggested to be present in mitochondrial of maize to regulate nutrient or sugar signalling through the opening of mitochondrial permeability transition pores. It is also present in chloroplasts whereby it helps in regulating the photosynthetic processes in plastids. Under abiotic stress conditions such as osmotic stress or draught, excessive synthesis of sucrose occurs in order to maintain membrane function and structure. Thus, *Susy* gene can be used as a potential biomarker and candidate gene for environmental stresses [7]. Despite the high economic value of tropical wood, there is no genetic information available in the public genomic databases about *Susy* gene in *Neolamarckia macrophylla*. In fact, the present study was the first molecular study to isolate the *Susy* genomic sequence and subsequently, to determine the nucleotide variation of *Susy* gene in *N. macrophylla. N. macrophylla* or commonly known as red kelampayan belongs to the Rubiaceae family. It is a fast growing tropical tree native to Eastern Indonesia. It has been cultivated widely in many countries such as in Malaysia, India and Philippines. It is able

to reach diameter at breast height of 30 cm within 5 to 7 years [8]. The leaves of the *N. macrophylla* are hairy and reddish green in colour and it has hard reddish timber. The timber can be used to make all sorts of stuff such as plywood, furniture, canoe, light construction.

2 MATERIALS AND METHODS

2.1 Plant Materials

A total of 15 *N. macrophylla* clones were selected from the breeding stocks maintained by the M-Gen Propagation Sdn Bhd, which is a plant tissue culture company based in Selangor, Malaysia. The fresh young leaves of these 15 samples were collected and DNA extraction was carried out by using a modified CTAB protocol [9]. The isolated DNA was then purified by using the Wizard® Genomic DNA Purification Kit (Promega, USA) according to the manufacture's protocol.

2.2 PCR Amplification and DNA Sequencing

PCR was performed by using a Mastercycler Gradient Thermal Cycler (eppendorf, Germany). The DNA template was mixed with 1 x PCR buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 5 pmol of forward and reverse primers, 1 U *Taq* DNA Polymerase (Invitrogen, USA) and sterile distilled water to a total of 25 µl. The sequence for *SuSy* forward primer was 5' TTG GAA GAG CAG GCA GAG AT 3' and the reverse primer was 5' CCG CAG ACA TCT ACT CAA CAG 3'. This primer pair was designed based on the *Susy* EST sequences from the *Neolamarckia cadamba* EST database [10]. The mixture undergone pre-denaturation for 2 minutes at 95°C; 35 cycles of denaturation for 30 seconds at 94°C; annealing for 30 seconds at 55°C; extension at 72°C for 30 seconds, and lastly a final extension of 8 minutes at 72°C. PCR amplicons were purified by using the GeneJet PCR Purification Kit (Thermo Scientific) and then 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).

2.3 Sequencing Variation (SNP) Analysis

Base calling and weak signal sequencing traces were removed by using the Chromas version 2.33 (Technelysium, AU). Each sequence was verified and checked for their homology by using the BLASTn through BLAST Search Engines (http://blast.ncbi.nlm.nih.gov/). After verification, the DNA sequences of *NmSusy1* were aligned among individuals using CLC Sequence Viewer 6 (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 NmSusy1 were aligned among individuals using CLC Sequence Viewer 6 (CLC Bio, Denmark) in order to detect the nonsynonymous and synonymous mutations.

3 RESULTS AND DISCUSSION

A resequencing approach was employed to determine the SNPs in *Susy* gene from 15 *N. macrophylla* 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 in the present study. Prior to sequence variation analysis, the *NmSusy1* sequences were analysed by using the BLASTn to determine the identity of the sequences obtained. The results showed high degree of similarity with Susy1 *Coffea arabica* (89%), Susy1 *Coffea canephora* (86%) and Susy2 *Coffea canephora* (88%) (Table 1). Multiple alignments were performed to detect the nucleotide polymorphism in *NmSusy1* from 15 *N. macrophylla* trees. Fig. 1 shows the alignment of 15 partial *NmSusy1* genomic sequences by using the CLC Sequence Viewer 6. There were five SNPs detected among the 15 *N. macrophylla* trees. Those SNPs were located at nucleotides 1, 2, 34, 35, and 397. Interestingly, one single base pair insertion-deletion (InDel) located at nucleotide 17 was also identified. InDel is also a type of mutation whereby there are insertions or deletions of bases. InDel has been related to genome size evolution where the possibility of illegitimate recombination explains genomic downsizing [11]. Many small InDels detected in rice hybrids resulted in the formation or disruption of putative *cis*-regulatory elements which are closely associated with the expression of transcription factors and thus, agronomic performance [12]. Plantegenet and his colleagues [13] also reported that the preponderance of 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.

TABLE 1
BLASTN ANALYSIS FOR PARTIAL NMSUSY1 DNA SEQUENCES

Organism	Accession No.	Identity	E-value
Coffea arabica	AM087674.1	89%	8e-138
Coffea canephora	AJ880768.2	86%	0.0
Coffea canephora	DQ834312.1	88%	2e-127
Nicotiana langsdorffii x *Nicotiana sanderae*	EF636817.1	83%	3e-94
Phaseolus vulgaris	AF315375.1	82%	1e-79

The partial *NmSusy1* genomic sequence contains coding and non-coding regions. Intron-exon boundaries were predicted from the ClustalW analysis and a schematic diagram was drawn to represent the predicted intron-exon boundaries (Fig. 2). The SNP position in the coding region and non-coding regions were then determined by using CLC Sequence Viewer 6.0 for all the 15 partial *NmSusy1* genomic sequences (Table 2). Four SNPs were located at the predicted coding region while one SNP was located at the predicted non-coding region. SNPs on the coding regions may cause amino acid substitutions which changes the amino acid of the proteins. If it changes a functionally essential amino acid residue, it may also lead to phenotypic change. On the other hand, SNPs that falls on the non-coding regions will not cause change in the phenotypes or functions [14]. SNP leads to silent mutation in the coding regions whereby the single base mutation does not affect translation of the amino acid sequence and thus produces the same gene product or protein. However, silent mutations are able to change the gene expression and regulation and therefore affect the function of the cells. It affects the structure of mRNA and translation efficiency of tRNA. In some cases, SNP leads to nonsense mutation due to the production of premature stop codon which produces a

truncated, non-functional protein. Missense mutation is the change of single amino acid to another which resulted in the production of different protein [15]. Mutation in the non-coding DNA normally does not contribute to different phenotypic effect. However, if mutation occurs at the non-coding regions such as the promoter sequences or other regulatory regions, it may prevent the binding of transcription factor to the recognition site in the promoter sequences. This will affect the level of gene expression or the gene may not be expressed at all [16].

TABLE 2
SEQUENCE VARIATIONS WITHIN PARTIAL NMSUSY1 GENOMIC DNA AMONG 15 SELECTED N. MACROPHYLLA TREES

Region / Characteristics	*SuSy*
Exons	
No. of SNPs in exon	4
bp in fragment sequenced	720
SNP per bp	1 in every 180 bp
Transition mutations(C/T, G/A)	3
Transversion mutations(C/G, A/T, C/A, T/G)	1
InDel mutation	1
Total no. of synonymous mutation	0
Total no. of non-synonymous mutation	1
Introns	
No. of SNPs in intron	1
bp in fragment sequenced	30
SNP per bp	1 in every 30 bp
Transition mutations(C/T, G/A)	0
Transversion mutations(C/G, A/T, C/A, T/G)	1
Overall	
Total sequence length	750
Total no. of amino acids	127
Total no. of SNPs	5
SNP frequency	1 in every 150 bp

The results obtained from this study revealed that the chances of detecting SNPs in the partial *NmSusy1* genomic sequences are high in the introns, with about one SNP in every 30 bp compared to one SNP in every 180 bp in the exon regions (Table 2). However, this statement was made based on the study of a partial *NmSusy1* genomic DNA of 750 bp and there were only 30 bp identified as intron region. In general, this result was in agreement with other similar studies published elsewhere. For instance, one SNP in every 189 bp for barley [17], one SNP in every 100 bp for grapes [18], one SNP in every 109 bp for *Shorea parvifolia* ssp. *parvifolia* [19], one SNP in every 229 bp for *A. mangium* superbulk [20], one SNP in every 59 bp (*C4H* gene) and 74 bp (*CAD* gene) for *Neolamarckia cadamba* [21] and one SNP in every 259 bp (*CesA* gene) and 38 bp (*XTH* gene) for *N. cadamba* [22]. Majority of the SNPs were resulted from transition mutation (3), two from transversion mutation and one was InDel mutation (Table 2). The predicted coding region was then translated into amino acid sequences and aligned among the 15 samples for determining the synonymous and non-synonymous mutations. It was found out that there was one non-synonymous mutation whereby amino acid glutamic acid (E) was replaced by arginine (R) in sample 1 as shown in Fig. 3. Nonsynonymous SNPs which cause changes to the protein sequence can affect the structural, functional or biochemical properties of the enzyme being produced and this will lead to the changes in phenotypic characteristic of the trees, especially in modification of sucrose biosynthesis [23]. If the same protein is formed, it is known as synonymous [15]. Synonymous mutation may leads to protein misfolding related to translation pausing, RNA splicing, changes in enzyme specificity [24]. Changes in the enzyme specificity may disrupt the active site, producing dysfunction protein that is unable to catalyse its substrate.

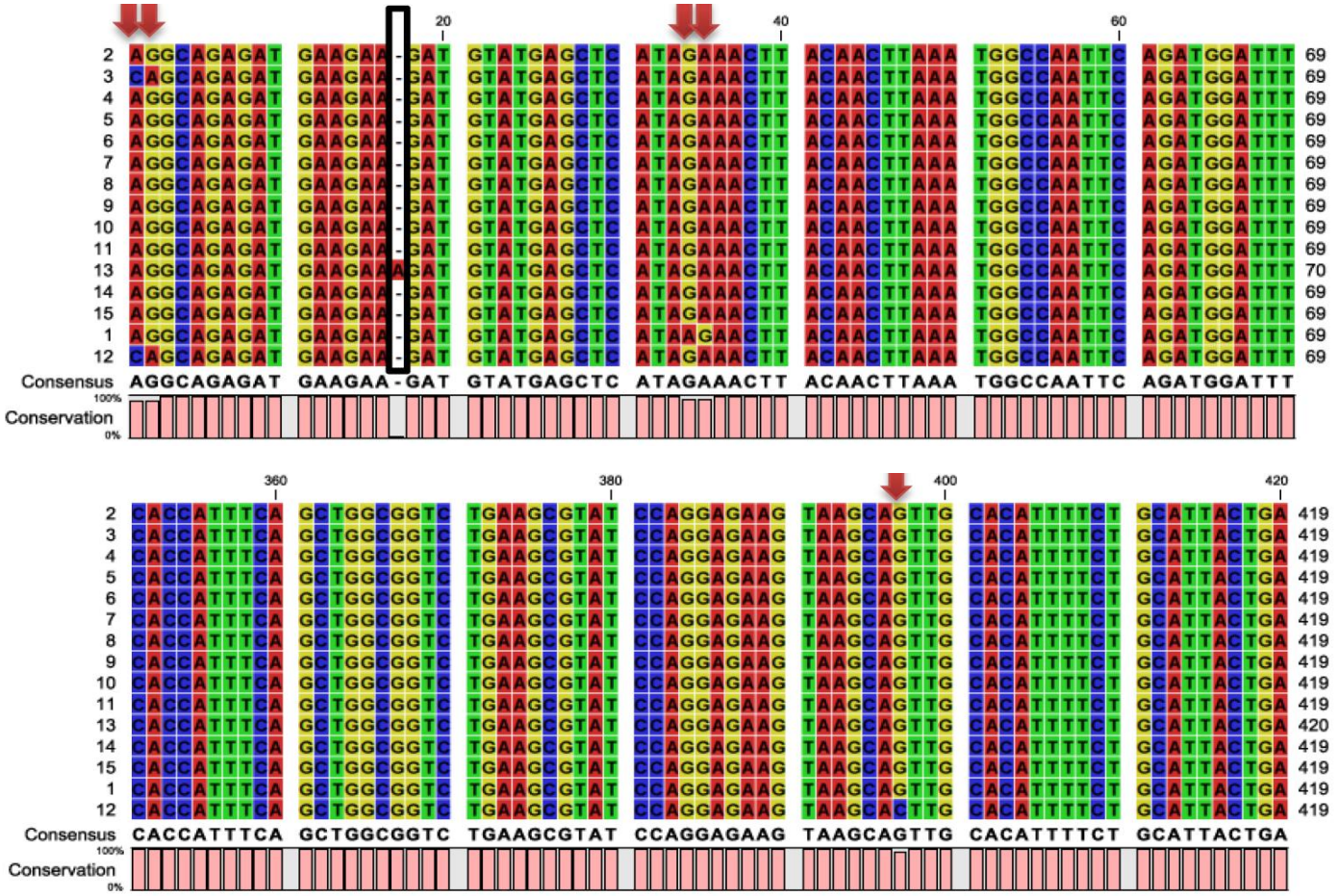

Fig. 1. Alignment of 15 partial sucrose synthase (*NmSusy1*) sequences by using CLC Sequence Viewer 6.

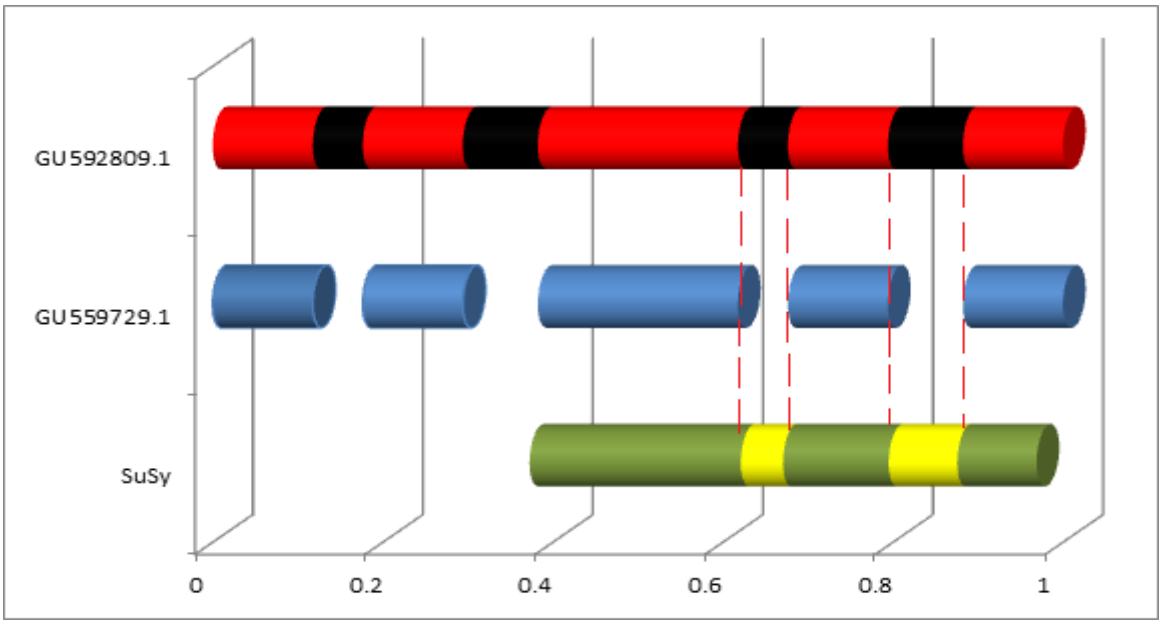

Fig. 2. Comparison of gene structure between Populus trichocarpa SuSy 1 genomic DNA (Accession no: GU592809.1), Populus trichocarpa SuSy mRNA (Accession no: GU559729.1) and the consensus sequence of the 15 sucrose synthase amplicons. Black cylinders represent intron while coloured cylinders represent the exons. Yellow colour cylinders represent intron in the partial sucrose synthase (NmSuSy1) genomic DNA.

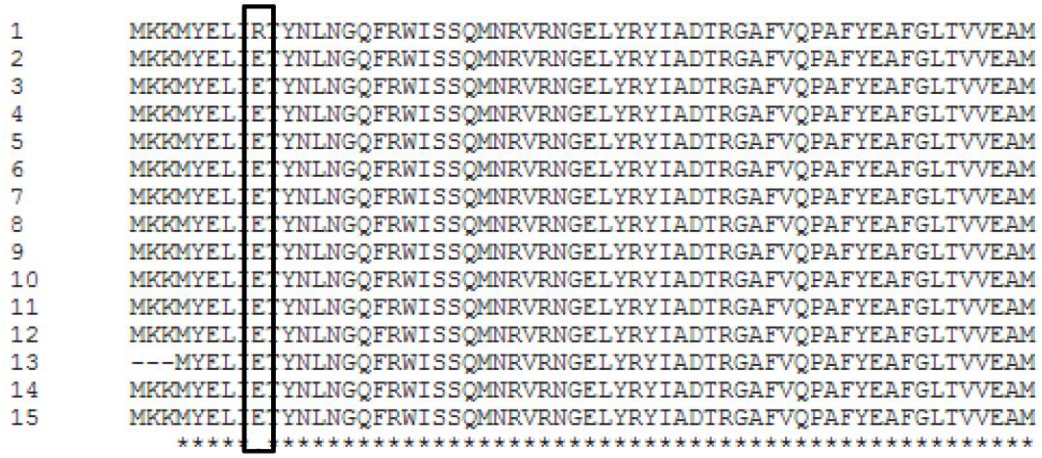

Fig. 3. Alignment of 15 partial sucrose synthase (NmSuSy1) amino acid sequences.

SuSy gene catalyses the reversible degradation of sucrose and UDP into UDP-glucose and fructose, whereby UDP-glucose serves as the precursor for cellulose biosynthesis [1]. Cellulose is then formed when glucose residue from UDP-glucose is transferred to the growing 1, 4-β-glucan chain [25]. Hence, if mutation occurred in the *Susy* gene of *N. macrophylla* and the dysfunction protein unable to catalyse sucrose into UDP-glucose, it will greatly affect the wood formation process and this may probably lead to the production of low quality wood. Besides that, since SNP at non-coding region affect the gene expression level, it controls the production level of sucrose synthase protein. Inadequate production of the protein may lead to production of low quality wood as well.

4 CONCLUSION

This study has demonstrated that the PCR amplification, followed by sequencing using primers designed from the *Susy* EST sequences, is an effective technique for resequencing and classifying molecular diversity or nucleotide variation in the *Susy* gene of *N. macrophylla*. To the best of our knowledge, this is the first report on nucleotide variation of *Susy* gene from *N. macrophylla*. Those SNPs, once validated, could potentially be used as a tool in marker-assisted selection (MAS) that enables more precise and accurate in the selection and prediction of yield or performance at the early developmental stages, such as at the seedling stage. Furthermore, this information is also essential to better understand the molecular diversity of the *Susy* gene in tropical tree genomes, as this could have many fundamental and commercial implications.

ACKNOWLEDGEMENTS

The authors would like to thank all the laboratory assistants and foresters involved in this research programme for their excellent field assistance in sample collection. This work was part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation (SFC), Sarawak Timber Association (STA) and University of Malaysia Sarawak (UNIMAS) under grant no.: RACE/a(2)/884/2012(02) and GL(F07)/06/2013/STA-UNIMAS (06).

REFERENCES

[1]. Winter, H. and Huber, S. C.: “Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes” Critical Reviews in Biochemistry and Molecular Biology, 35(4), 253-289, 2000.

[2]. Plaxton, W. C. and McManus, M.T.: “Control of primary metabolism in plants”, Volume 22.United State of America: Blackwell Publishing Ltd., 2006.

[3]. Ciereszko, I. and Kleczkowski, L. A.: “Glucose and mannose regulate the expression of a major sucrose synthase gene in Arabidopsis via hexokinase-dependent mechanisms” Plant Physiology and Biochemistry, 40, 907-911, 2002.

[4]. Carlson, S. J. and Chourey, P. S.: “Evidence for plasma membrane-associated forms of sucrose synthase in maize” Molecular and General Genetics, 252**(**3), 303-310, 1996.

[5]. Harada, T., Satoh, S., Yoshioka, T. and Ishizawa, K.: “Expression of sucrose synthase genes involved in enhanced elongation of pondweed (Potamogetondistinctus) turions under anoxia” Annals of Botany, 96, 1-2, 2005.

[6]. Haigler, C. H., Datcheva, M. I., Hogan, P. S., Salnikov, V. V., Hwang, S., Martin, K. and Delmer, D. P.: “Carbon partitioning to cellulose synthesis” Plant Molecular Biology, 47**(**1-2), 29-51, 2001.

[7]. Jayashree, B., Pradeep, R., Kumar, A. and Gopal, B.: “Correlation between the sucrose synthase protein subfamilies, variations in structure and expression in stress-derived expressed sequence tag datasets” Journal of Proteomics and Bioinformatics, 1, 408-423, 2008.

[8]. Krisnawati, H., Kallio, M. and Kanninen, M.: “Anthocephalus cadamba Miq. Ecology, silviculture and productivity” Indonesia: CIFOR, 2011.

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

[10]. Ho, W.S., Pang, S.L. and Julaihi, A.: "Identification and analysis of expressed sequence tags present in xylem tissues of kelampayan (Neolamarckia cadamba (Roxb.) Bosser)" Physiology and Molecular Biology of Plants, published online, 2014, DOI 10.1007/s12298-014-0230-x.

[11]. Grover, C.E., Yu, Y.S., Wing, R.A., Paterson, A.H. and Wendel, J.F.: "A phylogenetic analysis of indel dynamics in the cotton genus" Molecular Biology and Evolution, 25(7), 1415-28, 2008.

[12]. Zhang, H.Y., He, H., Chen, L.B., Li, L., Liang, M.Z., Wang, X.F., Liu, X.G., He, G.M., Chen, R.S., Ma, L.G. and Deng, X.W.: "A genome-wide transcription analysis reveals a close correlation of promoter INDEL" Molecular Plant, 1(5), 720-731, 2008.

[13]. Plantegenet, S., Weber, J., Goldstein, D.R., Zeller, G., Nussbaumer, C., Thomas, J., Weigel, D., Harshman, K., and Hardtke, C.S.: "Comprehensive analysis of Arabidopsis expression level polymorphism with simple inheritance" Molecular System Biology, 5, 542, 2009.

[14]. Garg, K., Green, P. and Nickerson, A.: "Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags" Genome Research, 9, 1087-1092, 1999.

[15]. Zhang, Z., Miteva, M. A., Wang, L. and Alexov, E.: "Analyzing effects of naturally occurring missense mutations" Computational and Mathematical Methods in Medicine, 1-15, 2012.

[16]. Turnpenny, P. and Ellard, S.: "Emery's elements of medical genetics" Philadelphia: Churchill Livingstone, 2012.

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

[18]. Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A. and Cestaro, A.: "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety" PLoS ONE, 2, 2007.

[19]. Ho, W.S., Pang, S.L. Lau, P. and Ismail, J.: "Sequence variation in the cellulose synthase (SpCesA1) gene from Shorea parvifolia ssp. parvifolia mother trees" Journal of Tropical Agricultural Science, 34(2), 323-329, 2011.

[20]. Tchin, B.L., Ho, W.S., Pang, S.L. and Ismail, J.: "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, 2011.

[21]. Tchin, B.L., Ho, W.S., Pang, S.L. and Ismail, J.: "Association Genetics of the cinnamyl alcohol dehydrogenase (CAD) and cinnamate 4-hydroxylase (C4H) genes with basic wood density in Neolamarckia cadamba" Biotechnology, 11(6), 307-317, 2012.

[22]. Tiong, S.Y., Ho, W.S., Pang, S.L. and Ismail, J.: "Nucleotide diversity and association genetics of xyloglucan endotransglycosylase/ hydrolase (XTH) and cellulose synthase (CesA) genes in Neolamarckia cadamba" Journal of Biological Sciences, 14(4), 267-375, 2014.

[23]. Bromberg, Y. and Rost, B.: "SNAP: Predict effect of non-synonymous polymorphisms on function" Nucleic Acids Res., 35(11), 3823-3835, 2007.

[24]. Parmley, J. L. and Hurst, L.D.: "How do snynonymous mutations affect fitness?" Bioessays, 29 6), 515-519, 2007.

[25]. Rai, A. and Takabe, T.: "Abiotic stress tolerant in plants" Netherland: Springer, 2006.