Isolation And Characterization Of Hypervariable Region (HVR II) From Cellulose Synthase Gene In Neolamarckia Macrophylla

Chia-Nin Chang, Wei-Seng Ho, Shek-Ling Pang

Abstract: The present study is aimed to isolate and characterize the hypervariable (HVRII) region in cellulose synthase gene from the developing xylem tissues of a tropical timber tree species, Neolamarckia macrophylla. N. macrophylla is locally known as red kelampanyan and it has been selected as one of the important reforestation tree species in Malaysia. RT-PCR was carried out by using the degenerate primers and one of the three amplified DNA bands was successfully sequenced and characterized. The sequence was named as NmCesAHVR II and it was clustered in a distinct clade that is associated with secondary cell wall development. This study has generated a useful genomic resource for a better understanding about the HVRII region of CesA gene in N. macrophylla and its function which is important in future applications, genetic improvement of N. macrophylla. This also facilitates the future selection of trees with optimal cellulose content required for certain specific industries as well as synthesizing of artificial cellulose, hence increasing the economic development and growth in the country.

Index Terms: Neolamarckia macrophylla, reverse transcription-polymerase chain reaction (RT-PCR), cellulose synthase (CesA), hypervariable region II (HVRII)

1 INTRODUCTION

Cellulose is a linear and unbranched polymer which made up of glucose monosaccharide units [1]. It can be found in the cell wall of plants. Cellulose is being synthesized at the plasma membrane by “rosette” complexes through freeze-fracture studies [2], [3]. Rosette complexes are identified as cellulose-synthesizing complexes where there is localization of cellulose synthase. Each cellulose-synthesizing complex contains at least three isofoms of cellulose synthase [4]. Previous studies on different plant species have proved that there are ten or more homologs present in cellulose synthase (CesA) gene family [4]. The cellulose synthase enzymes are believed can be further categorized into two types which function in cellulose synthesis in primary and secondary cell wall of plants [5]. Betancur et al. [6] stated that cellulose synthase enzymes from the primary cell wall phylogenetic clades can also support the secondary cell wall thickening besides cellulose synthase enzymes from the secondary cell wall phylogenetic clades.

In addition, there are at least six classes of CesA proteins exist in plants. In general, CesA protein consists of a zinc finger, two hypervariable regions (HVR I and II), several transmembrane domains and conserved residues [1], [7], [8], [19]. The hypervariable region II (HVRII) is made up of around 500bp to 600bp and the amino acid sequences between highly conserved motifs of ALYG and VISC are associated with this HVRII region. This region might be involved in the regulation of quantity and quality of cellulose synthesized in plants [8]. It also plays a role in interaction with other unique cell-type-specific proteins involved in the cellulose biosynthesis [7] that may regulate and affect the wood quality and properties [4], [9], [10]. Neolamarckia macrophylla has other synonyms names which are Anthocephalus macrophyllus, Bancal palmyryllus and Nauclea macrophylla [11]. It can be found in Indonesia, Malaysia, Vietnam, Filipina, Sri Lanka, Myanmar, Thailand, China and Papua New Guinea. It is a fast growing plantation tree species. The timber is durable and possesses red colour stem. It has fast growing rate where its height and diameter increases by 3 m and 7 cm, respectively per year. It is ready for logging within 4 to 6 years. It has high market price and value due to its good characteristics. In addition, it is resistant to pests and diseases. It is also valuable in the production of furniture and plywood as well as timber trade. It can be produced in large quantity and high productivity with good quality. Hence, this species is a good selection of tropical timber tree for research activities [11]. Despite the high economic value of N. macrophylla as one of the timber resources, little is known about the HVRII region in CesA gene of this important tree species. Therefore, the present study was aimed to isolate and characterize the HVRII region of CesA gene in N. macrophylla. It is hoped that this study could pave the way for a better understanding about the HVRII region of CesA gene in N. macrophylla and its function which is important in future applications.

2 MATERIALS AND METHODS

2.1 Total RNA Extraction and RT-PCR

Total RNA of N. macrophylla was extracted from the developing xylem tissues by using the RNeasy Midi Kit (Qiagen, USA). First-strand cDNA was synthesized according
to Ready-To-Go You-Prime First-Strand Beads protocol (GE Healthcare, USA). The RT-PCR amplification was carried out in a Mastercycler Gradient Thermal Cycler (Eppendorf, Germany). Forward degenerate primer, HVRR2F (5'-TGYTATGTCAGTAYCCGC-3') and reverse degenerate primer, HVRR2R (5'-GANGCRTARATCCAYGC-3') were used in the RT-PCR reaction [8]. The RT-PCR reaction was carried out with the initial denaturation at 95 °C for 10 min, followed by two cycles of non-stringent amplification profile of 94°C for 1 min, 41°C for 1 min 30 s and 72°C for 2 min, followed by another more stringent amplification profile at 28 cycles of 94°C for 1 min, annealing temperatures of 55°C and 45°C for 1 min 30 s and 72°C for 2 min. The total 25 μl amplification reaction volume contained 1× PCR buffer (200 mM Tris-HCl, 500 mM KCl, pH 8.4), 1.5 mM MgCl₂, 0.2 mM dNTPs, 10 pmol primers, 1 U Taq DNA polymerase (Promega, USA) and 1 μl template cDNA. The amplified cDNA was subsequently diluted with 50× ultrapure water and amplified again under the same PCR conditions. The RT-PCR amplicons were gel purified by using the QIAquick Gel Extraction Kit Protocol (QIAGEN, Germany) and sequenced by using the BigDye Version 3.1 (Applied Biosystems, USA) using Applied Biosystems 3730 DNA Sequencer.

2.2 Data Analysis
The DNA sequences were checked and edited using Chromas Lite 2.01. The edited sequences were checked for similarity with other sequences against the NCBI non-redundant database by using the BLASTn (http://www.ncbi.nlm.nih.gov). The nucleotide sequences were then translated into amino acid sequences by using the Expert Protein Analysis System (ExPASy) translate tool (http://www.expasy.org). The amino acid sequences 5’ CYVQFPQ.........GWIYGS 3’ were selected. HVRII region of CesA of *N. macrophylla* was multiple aligned with other HVRII regions from other plant species available in the public genome databases by using the ClustalX 2.0 programme. An un-rooted Neighbor-Joining (NJ) phylogenetic tree was then constructed with bootstrap value of 1000 by using the MEGA 6 software (http://www.megasoftware.net). CesAHVRII protein sequences of various plant species were retrieved, compiled and clustered. In addition, EMBOSS Needle global sequence alignment analysis (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) was used to pairwise align *N. macrophylla* and *N. cadamba*.

3 RESULTS AND DISCUSSION
Total RNA was extracted from the developing xylem tissues of *N. macrophylla* by using the RNeasy Midi Kit (Qiagen, USA) and the first-strand cDNA was successfully synthesized from the isolated RNA. RT-PCR was carried out by using the degenerate primers and three DNA bands were obtained from different annealing temperatures, 45°C and 55°C. After 50× dilution of each respective PCR product, PCR re-amplification was carried out by using the same primers at the same reaction conditions, respectively. One of the three bands (~520bp) was consistently amplified by PCR. The amplicon was successfully purified and sequenced meanwhile the other amplicons were failed to be purified due to the low concentration of the amplified amplicons. The edited nucleotide sequence, named as *NmCesA1HVRII*, was analysed by using the BLAST tools (www.ncbi.nlm.nih.gov) to validate the identity of the sequence obtained. BLASTn analysis of *NmCesA1HVRII* showed high identity with cellulose synthase mRNA of *N. cadamba* (98%), *Shorea parvifolia* spp. *parvifolia* (81%), *Populus tomentosa* (83%), *Eucalyptus globulus* spp. *globulus* (79%), *Gossypium hirsutum* (77%) and *Arabidopsis thaliana* (76%) (Table 1).

### TABLE 1

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Organisms</th>
<th>Accession No.</th>
<th>Identity (%)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NmCesAHVRII</td>
<td>Neotamnus cadamba</td>
<td>JX134621.1</td>
<td>98</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Populus tomentosa</td>
<td>JQ966934.1</td>
<td>83</td>
<td>2e-40</td>
</tr>
<tr>
<td></td>
<td>Shorea parvifolia spp. parvifolia</td>
<td>GQ398420.1</td>
<td>81</td>
<td>2e-103</td>
</tr>
<tr>
<td></td>
<td>Eucalyptus globulus spp. globulus</td>
<td>AB527047.1</td>
<td>79</td>
<td>4e-93</td>
</tr>
<tr>
<td></td>
<td>Gossypium hirsutum</td>
<td>IQ345695.1</td>
<td>77</td>
<td>4e-81</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis thaliana</td>
<td>NM_121748.3</td>
<td>76</td>
<td>2e-78</td>
</tr>
</tbody>
</table>

The validated nucleotide sequence was translated into amino acid sequence by using the Expert Protein Analysis System (ExPASy) translate tool. The degenerate primer was designed based on the conserved regions flanking the HVRII region of *CesA* gene. Thus, the correct translated open reading frame (ORF) of amino acid sequence was selected by the similarity of short amino acid sequences 5’ CYVQFPQ…..GWIYGS 3’ flanking the HVRII amino acid sequence. The protein sequence was checked again using BLASTp for similarity. The result of BLASTp showed that there was high similarity of *NmCesA1HVRII* with *CesA1* *N. cadamba* (96%), *CesA3* *E. grandis* (89%) and *CesA1 S. parvifolia* spp. *parvifolia* (88%). The BLASTp results for the *NmCesAHVRII* are shown in Table 2.

### TABLE 2

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Organisms</th>
<th>Accession No.</th>
<th>Identity (%)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NmCesAHVRII</td>
<td>Neotamnus cadamba</td>
<td>AJP33559.1</td>
<td>96</td>
<td>9e-69</td>
</tr>
<tr>
<td></td>
<td>Eucalyptus grandis</td>
<td>AAY06485.1</td>
<td>89</td>
<td>8e-62</td>
</tr>
<tr>
<td></td>
<td>Shorea parvifolia spp. parvifolia</td>
<td>ACT176415.1</td>
<td>88</td>
<td>9e-59</td>
</tr>
<tr>
<td></td>
<td>Solanum tuberosum</td>
<td>XP_006348620.1</td>
<td>87</td>
<td>9e-58</td>
</tr>
<tr>
<td></td>
<td>Zea mays</td>
<td>NP_001053532.1</td>
<td>79</td>
<td>8e-52</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis thaliana</td>
<td>AAD32031.1</td>
<td>77</td>
<td>2e-51</td>
</tr>
</tbody>
</table>

Global alignment was also conducted to observe the differences and conserved regions among species. To date, many comparative sequence studies have also proved that genes which are similar to each other at the molecular level can be related to one another by evolutionary [12]. Homologous genes or homologs are genes that are derived from a common ancestral gene. The similarity level in their sequences usually reflects the time since they diverged. There are a few factors that may cause the formation of homologous genes. These genes can be generated due to speciation event. Speciation can produce pairs of orthologs carrying similar functions. Furthermore, duplication of a chromosomal segment can also form homologous gene which produces paralogs carrying different functions [12]. In the present study, both HVRII protein sequences from *N. macrophylla* and *N. cadamba* were pairwise aligned by using the EMBOSS Needle global sequence alignment (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). It uses the Needleman-Wunsch dynamic programming algorithm to compare two proteins with similar function. BLOSUM matrices with high numbers like BLOSUM80 are more suitable to compare between closely related sequences whereas low-number BLOSUM matrices like BLOSUM45 are better for...
distantly related relationships [13]. *N. macrophylla* and *N. cadamba* are two different species of the same genus as well as ancestor. It can be said that these two species are genetically closely related. Hence, BLOSUM80 was used as amino acid substitution matrix in the sequence alignment analysis. Both the HVRII protein sequences from *N. macrophylla* and *N. cadamba* were pairwise aligned as shown in Fig. 1. The result showed that there was 91.4% of similarity and identity of 89.2% between these two sequences with 5.4% of gaps between the sequences. The high similarity between the two HVRII protein sequences indicates the high similarity in the function. In other words, *N. macrophylla* and *N. cadamba* are orthologs which carry many evolutionarily conserved sequences as well as similar functions. This further support Vergara and Carpita [14] suggestion of renaming the hypervariable region (HVRII) as class-specific region (CSR) due to the highly conserved hypervariable regions among CesA orthologs from different plant species.

Vergera and Carpita [14] have proposed a new perspective in hypervariable region II (HVRII) regions. In their research, they analysed 26 HVRII regions of various plant species available in the public domains and further identified the conservation of class-specific sequences among the HVRII regions. Hence, they proposed the renaming of HVRII regions of CesA proteins as class-specific regions (CSR). They mentioned that HVRII is a class-specific region due to each type of HVRII actually defines a specific class of CesAs in plant species. This finding was reconfirmed by Liang and Joshi [8] when they analysed 56 HVRII regions of various plant species in their research on aspen trees. In this study, only one HVRII region of CesA sequence was obtained successfully from the developing xylem tissues of *N. macrophylla*. Phylogenetic tree derived from 62 HVRII regions of 23 plant species including *N. macrophylla*, suggests that the HVRII region of *N. macrophylla* represented one of the six distinct CesA classes. Phylogenetic tree constructed as shown in Fig. 2 clearly shows that NmCesA1HVRII sequence was clustered into clade that was known to be implicated in secondary cell wall synthesis [8]. It can be seen that these regions are highly conserved among CesA orthologs from the different species. Hence, the NmCesA1HVRII is associated with the secondary cell wall synthesis and development. During secondary cell wall synthesis, cellulose is synthesized in large quantities [15] and this further confirms that HVRII gene plays an important role in synthesizing the cellulose which is the primary component in secondary cell wall and wood formations. With the gene expression and comparative studies that have been done previously elsewhere, these CesA proteins are confirmed to be associated with the function in the development of primary and secondary cell wall in plant species such as *N. cadamba*, *Zea mays*, *Gossypium hirsutum*, *Oryza sativa*, *Populus tremuloides*, *S. parvifolia* ssp. *parvifolia* and *Eucalyptus* [10], [7], [16], [17], [18], [19]. To date, there are six classes of CesA identified to be present in certain plant species after the analysis and construction of phylogenetic trees. For instance, there were six classes of CesA within *Populus tremuloides* (aspen) and *Arabidopsis thaliana* that have been identified to date [8]. Hence, it is possible that to have only one representative class of CesA in *N. macrophylla*.

![Fig. 1. Pairwise sequence alignment on both HVRII protein sequences from *N. macrophylla* and *N. cadamba* using BLOSUM80 matrix. ‘|’ indicates the matches and ‘.’ indicates the gaps.](image-url)
Fig. 2. Unrooted Neighbour-Joining (NJ) phylogenetic tree derived with ClustalX 2.0 and MEGA 6 software based on 62 HVRII regions of 23 different plant species. Bootstrap analysis was conducted with bootstrap values of 1000. HVRII domains from CesA and CslD proteins of various plant species were retrieved from NCBI website, compiled based on the conserved regions of 5'CYVQFPQ 3' and 5' GWIYGS 3' flanking the HVRII regions 5' ALYG 3' and 5' VISCG 3'. The NJ tree consists of the three clades that are associated with primary (P) cell wall synthesis, secondary (S) cell wall synthesis and CslD (C) protein between plant species. HVRII region of NmCesA1HVRII was marked by a green circle. All sequences were renamed for the convenience of this study.

Abbreviations: Nm: Neolamarckia macrophylla; Nc: Neolamarckia cadamba; At: Arabidopsis thaliana; Ptr: Populus tremuloides; Pp: Populus pumila × Populus tremula; Eg: Eucalyptus globulus; St: Solanum tuberosum; Zm: Zea mays; Ptri: Populus trichocarpa; Hcan: Hibiscus cannabinus; Gk: Gossypiodes kirkii; Spp: Shorea parvifolia spp. parvifolia; Gh: Gossypium hirsutum; Ga: Gossypium arboretum; Gr: Gossypium raimondii, Smoe: Selaginella moellendorffii; Pr: Pinus radiata; Ptom: Populus tomentosa; Pu: Populus ussuriensis; Os: Oryza sativa; Pe: Phyllostachys edulis; Am: Acacia mangium; Na: Nicotiana alata.
4 CONCLUSION
To the best of our knowledge, this is the first report on the isolation and characterization of HVRII region from *N. macrophylla*. One CesA HVRII region has been successfully isolated and characterized in the present study. However, it is not conclusive that *N. macrophylla* only has one CesA gene as our analysis was based on the developing xylem tissues. This is due to the expression level of CesA gene which may differ in different parts of plant such as leaves, roots, flowers and shoots. Hence, further study can be carried out to obtain more information on the expression of CesA gene from various tissues in the species. The characteristic of HVRII region which is class-specific can be used to design primer based on its conserved regions in order to obtain the full length CesA gene from this species via RACE approach. Further, molecular markers can then be developed to identify clone with the desired cellulose content at the early stage of the plant.

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.: 02(DP109)832/2012(1), RACE/a(2)/884/2012(02) and GL(F07)/ 06/2013/STA-UNIMAS(06).

REFERENCES