

In Silico Analysis of Cellulose Synthase Gene (NcCesA1) in Developing Xylem Tissues of *Neolamarckia Cadamba*

¹Tiong Shing Yiing, ¹Ho Wei Seng, ²Pang Shek Ling and ¹Ismail Jusoh

¹Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak

²Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, 93250 Kuching, Sarawak

Received 05-08-2014

Revised 06-08-2014

Accepted 15-10-2014

Corresponding Author:

Ho Wei Seng,

Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak
Email: wsho@frst.unimas.my

Abstract: This study reported the isolation and *in silico* characterization of full-length cellulose synthase (CesA) cDNA from *Neolamarckia cadamba*, an important tropical plantation tree species. CesA is a member of processive glycosyltransferases that involved in cellulose biosynthesis of plants. CesA acts as a central catalyst in the generation of plant cell wall biomass or cellulose. It also plays an important role in regulating wood formation. The hypothetical full-length CesA cDNA (NcCesA1; JX134621) was assembled by contig mapping approach using the CesA cDNA sequences from NcdbEST and the amplified 5'-and 3'-RACE PCR sequences. The NcCesA1 cDNA has a length of 3,472 bp with 3,126 bp open reading frame encoding a 1,042 amino acid sequence. The predicted NcCesA1 protein contained N-terminal cysteine rich zinc binding domain, 7 putative Transmembrane Helices (TMH), 4 U-motifs that contain a signature D, D, D, QxxRW motif, an alternating Conserved Region (CR-P) and 2 Hypervariable Regions (HVR). These entire shared domain structures suggest the functional role of NcCesA1 is involved in glycosyltransferation of the secondary cell wall cellulose biosynthesis of *N. cadamba*. Sequence comparison also revealed the high similarity (90%) among NcCesA1 and PtrCesA2 of *Populus tremuloides*. This further implies the involvement of NcCesA1 that catalyzes the cellulose biosynthesis of secondary cell wall rather than primary cell wall. This full-length NcCesA1 cDNA can serve as good candidate gene in association genetics study which leads to Gene-Assisted Selection (GAS) in the *N. cadamba* tree breeding programme. Furthermore, the isolation of new CesA genes from tropical tree genomes is essential for enhancing knowledge of cellulose biosynthesis in trees that has many fundamental and commercial implications.

Keywords: *Neolamarckia Cadamba*, Cellulose Synthase (CesA), Wood Formation, Contig Mapping, Gene-Assisted Selection (GAS), Candidate Gene, Association Genetics

Introduction

Wood is made up of secondary xylem tissues and has a chemical complex of cellulose, lignin, hemicellulose and extractives. Cellulose, a homopolymer consisting of β -1,4-glucan chains, is the most abundant form of living terrestrial biomass makes up the major cell wall biopolymer in plants (Crawford, 1981; Kumar *et al.*, 2009). Polymerization of cellulose chains is catalysed and synthesised by specific plasma membrane-bound Cellulose Synthase complexes (CeLS) (Festucci-Buselli *et al.*, 2007). CeLS is postulated to be composed of six hexameric rosette subunits where each of the rosette subunit consists six CesA catalytic subunits. Therefore, a total of 36 β -1,

4-glucan chains are produced by a CeLS in most of the higher plants. These chains will then linked by hydrogen bonds to form microfibrils which will then further bundled to form macrofibrils. CeLS is proved to be encoded by cellulose synthase (CesA) or CesA-related genes. CesA family is a member of Glycosyltransferases (GTs) superfamily under CAZyme family (Ross *et al.*, 2001).

Cellulose synthases are the only identified components of rosettes. Since woody plants are unique in their cellulose biosynthesis, CesA sequence information is useful to serve as the basis for investigating molecular regulation and mechanism of cellulose biosynthesis in tree species (Lu *et al.*, 2008). The first gene discovered to encode for CesA was in a bacteria species, *Acetobacter*

xylimum (Saxena *et al.*, 1990; Wong *et al.*, 1990). Few years later, cotton (*Gossypium hirsutum*) was the first higher plant species found to have homologs of the bacterial *CelA* gene (Pear *et al.*, 1996). *CelA* refers to the cellulose synthase catalytic subunit which is a specific conserved region common to glycosyltransferases found in both bacterial and plants.

Over the years, many Cesa genes were continuously discovered, including at least: 10 Genes from *Arabidopsis thaliana* (Richmond and Somerville, 2000); 12 genes from maize (*Zea mays*) (Appenzeller *et al.*, 2004); 10 genes from rice (*Oryza sativa*) (Tanaka *et al.*, 2003); seven genes from hybrid poplar (*Populus tremula* x *Populus alba*) (Joshi *et al.*, 2004) and six genes from *Eucalyptus grandis* (Ranik and Myburg, 2006). Previous studies have been done by different groups of researchers showing that three different Cesa genes found to be co-expressed in *A. thaliana*: *AtCesA4*, *AtCesA7*, *AtCesA8* (Taylor *et al.*, 2003); poplar (*P. tremula* x *P. alba*): *PtrCesA1*, *PtrCesA2*, *PtrCesA3* (Joshi *et al.*, 2004); rice (*O. sativa*): *OsCesA4*, *OsCesA7*, *OsCesA9* (Tanaka *et al.*, 2003); barley (*Hordeum vulgare*): *HvCesA1*, *HvCesA2* and *HvCesA6*, *HvCesA4* and *HvCesA7*, *HvCesA8* (Burton *et al.*, 2004).

To date, there are considerable amounts of full-length Cesa cDNA being published in NCBI but no such information available for *Neolamarckia cadamba* trees. *N. cadamba* or locally known as kelampayan belongs to the family of Rubiaceae. It has been selected as one of the fast growing plantation species for planted forest development in Sarawak (Tchin *et al.*, 2012; Lai *et al.*, 2013; Tiong *et al.*, 2014a; 2014b; Ho *et al.*, 2014; Phui *et al.*, 2014). The state government of Sarawak has introduced the Forest (Planted Forest) Rules (1997) to encourage the development of commercial planted forests and has set a target of 1.0 million hectares for forest plantations to be established by 2020. It is estimated that 42 million of high quality seedlings are required for the annual planting programme. *N. cadamba* is a large, deciduous and fast growing tree that gives early economic returns within 8-10 years. Under normal conditions, it attains a height of 17 m and diameter of 25 cm at breast height (dbh) within 9 years. It is a lightweight hardwood with a density of 290-560 kg/m³ at 15% moisture content (Joker, 2000). It is one of the best sources of raw material for the plywood industry, besides pulp and paper production. *N. cadamba* can also be used as a shade tree for dipterocarp line planting, whilst its leaves and bark have medical applications (WAC, 2004). *N. cadamba* also has high potential to be utilized as one of the renewable resource of raw materials for bioenergy production such as cellulosic biofuels in the near future.

Hence, the objectives of this study were: (i) To obtain the full-length Cesa cDNA sequences through contig

mapping approach by using Cesa singletons from the kelampayan tree transcriptome database (NcdbEST) and RACE PCR sequences and (ii) to *in silico* characterize the Cesa gene from *N. cadamba*. The full-length Cesa cDNA discovered can serve as good candidate gene for association genetics study in *N. cadamba* to detect the potential genetic variants underlying the common and complex adaptive traits.

Materials and Methods

RNA Isolation and First Strand cDNA Preparation

Total RNA was isolated from the developing xylem tissues of a 4-year old *N. cadamba* tree using RNeasy® Midi Kit (QIAGEN GmbH, Germany) with modification. 5'-and 3'-RACE ready cDNA were prepared using SMARTer™ RACE cDNA Amplification Kit (Clontech, USA).

Rapid Amplification of 5'-and 3'-cDNA Ends (RACE)

Two Gene-Specific Primers (GSPs) for Cesa gene were designed for 5'-RACE PCR (Cesa5'GSP, 5'-TGTAACCGTACCAGGCAGGGCTATGTC-3') and 3'-RACE PCR (Cesa3'GSP, 5'-GTGCTTCGATGGGCTCTTGGTTCTG-3') using Primer Premier 5 software based on the parameters given in the protocol of SMARTer™ RACE cDNA Amplification Kit (Clontech, USA). 5'-and 3'-RACE PCR were prepared in separate tubes by mixing 34.5 µL of PCR-grade water, 5.0 µL of 10× Advantage 2 PCR Buffer, 1.0 µL of dNTP Mix (10 mM) and 1.0 µL of 50× Advantage 2 Polymerase Mix. The master mix was then added into a 0.2 mL PCR tubes containing 2.5 µL of RACE-Ready cDNA, 5.0 µL of 10× Universal Primer Mix (UPM) and 1.0 µL of GSP to make a final volume of 50.0 µL. The contents were then mixed gently and subjected to thermal cycling. Touchdown PCR amplification was used: 5 Cycles of incubation at 94°C for 30 sec and 72°C for 3 min; 5 cycles of 94°C for 30 sec, 70°C for 30 sec and 72°C for 3 min; 25 cycles of 94°C for 30 sec, 68°C for 30 sec and 72°C for 3 min. The RACE PCR amplicons were purified from agarose gel by using QIAquick® Gel Extraction Kit (QIAGEN, Germany). Purified PCR product was ligated into pGEM®-T Easy Vector System (Promega, USA) and transformed into competent cells, *Escherichia coli* JM 109. The recombinant plasmids were isolated and purified using Wizard® Plus SV Minipreps DNA Purification System (Promega, USA) according to the manufacturer's protocol. After verification, the purified plasmids were sent for sequencing in both forward and reverse direction. The sequencing reactions were performed by using ABI Prism™ BigDye™ terminator cycle sequencing Ready reaction kit V.3.1 (Applied Biosystems, USA) and analysed on a ABI 3730XL