Isolation and Characterization of Full-Length Cellulose Synthase Gene (HsCesA1) from Roselle (Hibiscus sabdariffa L. var. UMKL)

SEYEHDEH SAREH SEYEDI, SOON GUAN TAN, PARAMESWARI NAMASIVAYAM & CHRISTINA SEOK YIEN YONG*

ABSTRACT

The Hibiscus sabdariffa var. UMKL (Roselle) investigated here may potentially be used as an alternative fibre source. To the best of our knowledge, there was no study focusing on the genetics underlying the cellulose biosynthesis machinery in Roselle thus far. This paper presents the results of the first isolation of the cellulose synthase gene, HsCesA1 from this plant, which is fundamental for working towards understanding the functions of CesA genes in the cellulose biosynthesis of Roselle. A full-length HsCesA1 cdNA of 3528 bp in length (accession no: KJ608192) encoding a polypeptide of 974 amino acid was isolated. The full-length HsCesA1 gene of 5489 bp length (accession no: KJ661223) with 11-introns and a promoter region of 737 bp was further isolated. Important and conserved characteristics of a CesA protein were identified in the HsCesA1 deduced amino acid sequence, which strengthened the prediction that the isolated gene being a cellulose synthase belonging to the processive class of the 2-glycosyltransferase family 2A. Relative gene expression analysis by semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) on young leaf and stem tissues found that HsCesA1 had similar levels of gene expression in both tissues. Phylogenetic and Blast analyses also supported the prediction that the isolated HsCesA1 may play roles in the cell wall depositions in both leaf and stem tissues.

Keywords: Cellulose; cellulose synthase gene (CesA); fibre; HsCesA1; promoter; semi-quantitative RT-PCR

INTRODUCTION

Cellulose is a main component of the cell wall and it is the most abundant biopolymer in the world (Hayashi 1989; Kimura & Itoh 1995; Lin & Aronson 1970). The cellulose synthase enzyme complexes consisting of multiple CesA protein subunits play important roles in the production of cellulose in many organisms. Mueller and Brown Jr. (1980) reported that cellulose synthase complex (CSC) is a rosette or hexameric like membrane-bound protein complex located in the inner surface of the plasma membrane. The first plant gene encoding the catalytic subunit of the CesA enzyme was isolated from cotton Gossypium hirsutum, reported by Pear et al. (1996). Since then, many cellulose synthase genes were isolated from other plants such as Arabidopsis (Arioli et al. 1998), Populus (Wu et al. 2000), rice (Tanaka et al. 2003), barley (Burton et al. 2004), Eucalyptus (Ranik & Myburg 2006) and Acacia hybrid (Yong & Wickneswari 2013). Different CesA isoforms are needed for the deposition of cellulose into primary and secondary cell walls, which explained the presence of multiple CesA isoforms found in most plant species such as Arabidopsis, maize, poplar (Appenzeller et al. 2004; Kumar & Turner 2014; Somerville 2006).

Sains Malaysiana 45(5)(2016): 717–727
Hibiscus sabdariffa L. or commonly known as Roselle is an erect short-day annual, bushy herbaceous shrub belongs to the family of Malvaceae (Coble 1956). Roselle has great economical potential as it is used in the medicinal, beverage and food and pulping industries (Da-Costa-Rocha et al. 2014; Hopkins et al. 2013; Juliani et al. 2009; Rodrigues 2010; Sáyago-Ayerdi et al. 2007). Though its origin is unsure, the plant is cultivated in many tropical and subtropical countries currently including India, Malaysia, Saudi Arabia, China, Indonesia, The Philippines, Sudan, Egypt, Nigeria and Mexico (Da-Costa-Rocha et al. 2014). In Malaysia, there are two main varieties of Roselle available for commercial planting and one of them is the Hibiscus sabdariffa L. var. UMKL. The Hibiscus sabdariffa L. var. UMKL were studied here is mainly cultivated for its calyces in Malaysia but it has potential to be used as a non-wood based fibre source. In a study conducted by Singha and Thakur (2008), they discovered that the Hibiscus sabdariffa has the potential as an alternative fibre source. Agricultural residues from non-wood annual plants are a promising alternative to wood fibre as an industrial feedstock. These agricultural wastes are abundant, cheap and their use will yield economic and environmental dividends. Some of the agricultural residues that have been utilized in making pulp and particleboard including almond shells (Gürü et al. 2006), sunflower stalks (Bektas et al. 2012), eggplant stalks (Guntekin & Karakus 2008), rice straw (Yasina et al. 2010) and bagasse (Xu et al. 2009). We believe that the Roselle stalk that is readily available as agricultural residues after harvesting can be turned into useful biomass. To the best of our knowledge, there was no study focusing on the genetics underlying the cellulose biosynthesis machinery in Roselle thus far. Here, we report the first cellulose synthase gene, HsCesA1 isolated from the commercially valuable Roselle.

MATERIALS AND METHODS

ISOLATION OF FULL-LENGTH HSCESA1 CDNA

The seeds of H. sabdariffa L. var. UMKL used in this study were obtained from the Department of Agriculture, Terengganu, Malaysia. Total RNA was extracted from young leaf tissues using the RNAeasy Plant mini kit (Qiagen, Germany). First strand cDNA was generated using the SMARTer™ RACE cDNA Amplification Kit (Clontech, USA) according to the manufacturer’s protocols. Four primers were designed for the RACE PCR amplifications using the Primer 3 software (Untergrasser et al. 2012), two 3’ RACE primers: SP1 (5’ CATTGGCCCCCTGTTAGGTGGCCTTGGAG3’) and HsCesA2F (5’ GTTGCTGTCGGCTTGGGCTTTGCTC3’), and two 5’ RACE primers: Hs_Race_R1 (5’ CGCCCTCCGATGACCCAAAATCTGGTC3’) and SR2 (5’ TGGGCAAGCTGGGCAATTTTAGGGT3’). RACE PCR reactions were performed to isolate the full-length cDNA of HsCesAl using the Advantage 2 polymerase Mix (Clontech, USA). The purified PCR products were cloned using the pGEM-T Easy Vector (Promega, USA) following the manufacturer’s instructions. Purified plasmids from three positive clones were sequenced using an ABI PRISM 3730X1 Genetic analyzer at 1st BASE Sdn. Bhd., Malaysia.

ISOLATION OF HSCESA1 FULL-LENGTH GENE

Genomic DNA was extracted from the young leaf tissue using the CTAB method (Doyle & Doyle 1990) with minor modifications. Five pairs of primers were designed based on the DNA sequence obtained from RACE PCR using the Primer 3 software for the isolation of the intronic regions. The primer pairs were SDF3 (5’ TCTCCCTACCCGTTATGC3’) and SDR3 (5’ CTCTTCCGACCTAACTGG3’); SDF4 (5’ GGTGAGATGGGAGAGTG3’) and SDR4 (5’ TGTGGACGATGAGGTG3’); SDF5 (5’ GCTTGCTGGAGAATGTG3’) and SDR5 (5’ CTGTCAGCCCTTTAGTTG3’); SDF6 (5’ AACTAAGGCTGACACAGG3’) and SDR7 (5’ GAAGCAAGTTGGAGAGGG3’); SDF2 (5’ CAGTTATTGTGGGAGAACAGG3’) and SDF2A (5’ GAGAAGACAGATTGTGCTAGT3’). PCR amplification was started with a single cycle of pre-denaturation at 94°C for 1 min, followed by 30 cycles of 94°C for 30 s, 52°C-57°C for 30 s and 72°C for 90 s and ending with a single cycle of final extension at 72°C for 5 min. Gel-purified PCR products were sequenced from both ends using an ABI PRISM 3730X1 genetic analyzer at 1st BASE Sdn. Bhd., Malaysia. The promoter region of HsCesA1 was isolated using genomic DNA according to the protocol described by Seibert et al. (1995). Three primers GWGSP1: (5’ ATGAACTCCCAACTTTCTGTAAACCAAC3’), G W S A P 2 : (5’ CATTCAACCCCACATTCTCACCACAAG3’), and G W G S P 3 : (5’ TGAAGAAAAAGTGGTTAGGGGAGTTTGGG3’) were designed using Primer 3 software. 0.5 μL of EcoR V digested genomic DNA was employed for primary PCR and the primary PCR product was used as the template for secondary PCR reaction. For both primary and secondary PCR amplifications, the following cycling profile was used: 7 cycles (94°C, 25 s; 72°C, 3 min), followed by 32 cycles (94°C, 25 s; 67°C, 3 min) and then a final elongation step at 67°C for 4 min. Gel-purified secondary PCR products were cloned using pGEM-T Easy Vector (Promega, USA). Three positive clones were sequenced using an ABI PRISM 3730X1 Genetic analyzer at 1st BASE Sdn. Bhd., Malaysia.

RELATIVE GENE EXPRESSION ANALYSIS BY SEMI-QUANTITATIVE RT-PCR

Total RNA was extracted from 100 mg of fresh young leaf and young stem tissues using the RNAeasy Plant Mini kit. A pair of gene specific primer for HsCesA1 was designed near to the 3’ end of the HsCesAl cDNA sequence using the Primer3 software. These primers were RTF2 (5’ GTAACGAGCAGTTTTTGGGTATC3’) and RTR2 (5’ TGGGCAAGCTGGGGAATTTG3’).
(5’GTTGCTTCCTGTTTGTCTTTC3’). Actin and the catalytic subunit of protein phosphatase 2A genes, both designed and normalized on Gossypium hirsutum (Artico et al. 2010), were used as internal controls. The expression level of these two housekeeping genes (HKGs) was tested prior to use. The relative expression pattern of the CesA gene in leaf and stalk tissue was conducted using the QIAGEN OneStep RT-PCR Kit (Qiagen, Germany). For the preparation of reaction mixture for HsCesA1, a volume of 10 μL of QIAGEN one step RT-PCR buffer (5×), 2 μL of dNTP mix (10 mM), 3 μL of each forward and reverse primers (10 μM), 25 μL of RNase-free water, 2 μL of QIAGEN one step RT-PCR enzyme mix and 5 μL of RNA template (200 ng) were mixed in a total volume of 50 μL in a 0.2 mL microcentrifuge tube. For the preparation of reaction mixture for the amplification of housekeeping genes, similar reaction mixture was used with the exception of primers where 1.5 μL of each forward and reverse (10 μM) and 28 μL of RNase-free water were used. For the RT-PCR amplifications of both the HKGs and HsCesA1 gene, the following cycle profile was applied: 50°C for 30 min, 95°C for 16 min, 40 cycles (94°C, 30 s; 54°C, 30 s; 72°C, 45 s), followed by 72°C for 10 min as final elongation. PCR products for all the tissues were analyzed on 8% (W/V) PAGE at 108 V for 2 h in 1× TBE buffer. The gel-purified PCR products were sequenced at 1st BASE Sdn. Bhd. (Seri Kembangan, Selangor) to validate the PCR product specifically.

DATA ANALYSIS

ANALYSIS OF FULL-LENGTH HSCESA1 CDNA SEQUENCE

CAP contig assembly of full-length cdNA sequence, prediction of the molecular weight and nucleotide composition of full-length cdNA were carried by using the Bioedit software (Version 5.0.6). The full-length cdNA sequence was subjected to similarity search using Blast analyses (Altschul et al. 1990) and analyze with Plazá 3.0 software (Proost et al. 2015). Multiple sequence alignment between the sequences retrieved from NCBI database (http://www.ncbi.gov.com) and the HsCesA1 cdNA sequence were performed using ClustalW2 (Version 2.0) (http://www.ebi.ac.uk/Tools/msa/clustalw2/). ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used for ORF prediction. The HsCesA1 cdNA sequence was analyzed using the Genescan program accessible from (http://genes.mit.edu/GENSCAN.html) and Eukaryotic GeneMark (Lomsadze et al. 2005) software. The UTRScan program (http://ittbiotools.ba.rit.cn.it/utrsan) and the RgRNA software (http://regma2.mbc.nctu.edu.tw) were applied for 5’ and 3’ UTR predictions. TAIR Blast analysis was carried out using the TAIR program (Lamesch et al. 2012). In-silico translation of the cdNA sequence was performed using the transeq program available from (http://www.ebi.ac.uk/Tools/st/). Deduced amino acid was also analyzed using motif search site based on both ProDom and Pfam (http://www.genome.jp/tools/motif/) and Pfam software (http://pfam.janelia.org/search). Conserved domains and motifs were also predicted for the deduced amino acid sequence of HsCesA1 using Conserved Domains prediction software (NCBI) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Phylogenetic analysis was performed using neighbor joining method (Saitou & Nei 1987) using MEGA version 5 (Tamura et al. 2011). The phylogenetic tree was constructed with 1000 bootstrap value based upon the neighbor joining method by aligning the deduced amino acid sequence of HsCesA1 and 21 deduced proteins from known CesAs of Gossypium, Eucalyptus, Populus and Arabidopsis. Acytostelium subglobosum was used as out-group. The CesA sequences retrieved from the NCBI database as follow: Gossypium hirsutum CesA1 (ADY68800.1), GhCesA3 (AAD39534.2), GhCesA4 (AAL37718.1), GhCesA5 (AFB18634.1), GhCesA6 (AFB18635.1), GhCesA7 (AFB18636.1), Eucalyptus grandis CesA1 (AAY60843.1), EgCesA3 (AAY60845.1), EgCesA4 (AAY60846.1), EgCesA5 (AAY60847.1), Populus tomentosa CesA1 (AAY21910.1), PtoCesA4 (AGW52080.1), PtoCesA5 (AEE60899.1), PtoCesA7 (AEE60896.1), PtoCesA8 (AEE60893.1) and Arabidopsis thaliana CesA1 (NP-194967.1), AtCesA3 (NP-19613.1), AtCesA4 (NP-199216.2), AtCesA5 (NP-196549.1), AtCesA7 (NP-197244) and AtCesA8 (NP-567564.1). Acytostelium subglobosum CesA (BAK54004.1).

FULL-LENGTH GENE ANALYSIS

The HsCesA1 cdNA and the intron sequences amplified were assembled using CAP contig assembly (Huang & Madan 1999). Blast (http://blast.ncbi.nlm.nih.gov/) was used for similarity search. Multiple sequence alignment between the sequences retrieved from the NCBI database (http://www.ncbi.gov.com) and full-length gene of HsCesA1 nucleotide sequence was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Version 2.0). The eukaryotic GeneMark software (Lomsadze et al. 2005) was applied to predict the exon-intron boundaries.

PROMOTER ANALYSIS

The promoter sequence was analyzed using prediction plant promoter (http://linux1.softberry.com/berry.phtml), neural network promoter prediction (http://www.fruitfly.org/seq_tools/promoter.html) and plant promoter analysis navigator (http://plantpan.mbc.nctu.edu.tw) software.

THE RELATIVE GENE EXPRESSION ANALYSIS

The intensities of the PCR products on the gel were used as a measurement to compare the relative expression levels of the cellulose synthase gene in the leaf and stem tissues of H. sabdariffa semi-quantitatively. The relative expression levels of HsCesA1 in the leaf and stem tissues were analyzed using GelAnalyzer software (developed by Dr. Istvan Lazar) and GelQuant.NET software provided by biochemlabsolutions.com.
RESULTS AND DISCUSSION

ISOLATION AND IN-SILICO ANALYSIS OF FULL-LENGTH HsCESA1 cDNA

This study is the first report on the cellulose synthase gene isolated from the *H. sabdariffa* var. UMKL. The full-length *HsCESA1* cDNA is 3528 bp (accession no: KJ608192) with an open reading frame of 2925 bp flanked by 379 bp of 5′untranslated region (UTR) and 168 bp of 3′ UTR. A putative polyadenylation signal (AATAAA) was found at position 3472 bp in the *HsCESA1* cDNA sequence, 16 bp upstream of the poly A site. *In-silico* analyses of *H. sabdariffa* *CesA* cDNA suggested that it encoded a putative CesA protein. Several important conserved domains were predicted in the amino acid sequence of *HsCESA1* including cellulose synthase domain and bcsA domain. The cellulose synthase domain, a key characteristic of CesA, was identified at positions 1-974 residues in the *HsCESA1* protein sequence. These conserved features of the CesA proteins are critical in facilitating cellulose assembly, membrane association and glycosyltransferase activity.

Cellulose synthase domain is responsible for cell wall synthesis, essential for growth and tissue development (Arioli et al. 1998; Scheible et al. 2001). Apart from the cellulose synthase domain, another key feature predicted in the HsCESA1 protein sequence was the bcsA catalytic subunit. Cellulose synthase catalytic subunits utilize UDP-glucose as substrate for glucan chain elongation and UDP-Glc is also thought to bind to an active site on the cytoplasmic surface of the plasma membrane with the polysaccharide being extruded through the membrane, presumably through a pore-type structure, into the wall (Brown Jr. & Saxena 2000; Kumar & Turner 2014; Sethaphong et al. 2013).

The Zn-binding domain/ Ring finger domain, two conserved domains (Domain A and B) and two Hypervariable regions (Figure 1) were also identified in the deduced amino acid sequence of *HsCESA1*. Zn-binding domain/ Ring finger domain containing 70 residues predicted at the amino terminus of the HsCESA1 deduced amino acid sequence has a significant conserved motif CxxC. Only the CesA proteins have the conserved motif CxxC, (Cx,Cx, Cx, Cx) within their zinc finger (LIM)/Ring domains (Richmond 2000; Wu et al. 2000). A study conducted by Timmers et al. (2009) demonstrated that each CesA protein contains a double zinc-finger motif that is highly homologous to the ring-finger domain. Ring-fingers are known to play a role in protein-protein interactions in the CesA complex during the cellulose biosynthesis process in a redox regulated bridging between cysteine residues, Kurek et al. (2002) showed that the ring-fingers domain, which is located on the N-terminus of the protein, only present in CesA proteins but not in cellulose synthase like (Csl) proteins. The ring-fingers domain play role in the ubiquitin-mediated proteolysis where CesA proteins aggregate into rosette subunits via the Zn\(^{2+}\) binding Ring finger-like motif (Joshi et al. 2004; Wu et al. 2000).

Domain A was 274 residues in length starting at residue 219 and included two conserved D residues positioned at 293 and 461 residues, respectively. Domain B predicted in the deduced amino acid sequence of HsCESA1 is 52 residues in length; positioned from residues 663 to 714, which included a single conserved D residue positioned at 672 and a conserved QXXRW (QXXRW) motif at 710-714 residues. The first two conserved D residues in Domain A existed in both processive and the non-processive glycosyltransferases, but the third conserved D residue in the Domain B is only present in processive glycosyltransferases. Charnock and Davies (1999) use crystallographic evidence to describe a model that the three D residues were involved in the binding of UDP-sugar substrate and in catalysis of glycosyltransfer, in conjunction with a divalent cation. The QXXRW motif was also found in the CesA proteins of many plants. Saxena and Brown (1997) showed that the conserved motif QXXRW identified in the Domain B of higher plants (*Arabidopsis thaliana*, *Brassica campestris* and *Oryza sativa*) was part of the catalytic site of CesA and is only found in processive type of CesAs. The specific third D residue and QXXRW motif found in the deduced amino acid of HsCESA1 classified it as a processive type of CesAs.

The Hypervariable Region I (HVRI) was predicted at position 86 until 169 residues whereas the Hypervariable Region II (HVRII) was identified at 494 until 662 residues in the HsCESA1. HVRII predicted at the downstream of Domain A might be important in distinct cell- and tissue-specific patterns of expression. HVRII can be used to examine structural and functional relations among CesA genes and might be involved in the regulation of quality and quantity of cellulose synthesized in plants (Holland et al. 2000; Liang & Joshi 2004). It is also worth mentioning that the HVRI in the HsCESA1 was flanked by two conserved sequences CVVQFPQ and G1WYG5. These two sequences are highly conserved in all known plants CesAs (Ranik & Myburg 2006). Multiple sequence alignment between HsCESA1 protein and CesA protein sequences of the other species indicated these two highly conserved sequences across species.

Nine transmembrane domains (TMDs) were predicted in the HsCESA1, in which two were identified in the N-terminus (169-218 residues) and seven others in the C-terminus (residues 714-954). Transmembrane domains (TMDs) can be found in the CesAs of most plant species ranging from eight to ten TMDs, which two of them usually are located at the N-terminus and the rest placed near the C-terminus regions (Ranik & Myburg 2006). This was also observed in the HsCESA1 isolated here where two TMDs are located at the N-terminus and seven at the C-terminus. These TMDs formed pores in the plasma membrane and allowed the growing glucan chain to reach the newly formed cell wall through the pores (Richmond 2000). All these specific sequence characteristics predicted in the HsCESA1 (Figure 2) strongly supported that the isolated gene in this study encodes a CesA protein in *Hibiscus sabdariffa*.
Figure 1. Multiple alignments of deduced amino acids of HsCesA1 with five most significant similarities of cellulose synthase protein sequences. Identical conserved residues at a particular site are shown in color bracket. The purple box at the top shows a significant conserved motif $C \times C_2 \times F \times C_2 \times P \times C_2 \times E \times C_5 \times G \times C_3 \times C$ (CxxC) within the Zn-binding domain/ Ring finger domain. Domain A and B are shown with red boxes, respectively. The blue boxes indicate three D conserved residues. The yellow boxes show CYVQFPQ and GWIYGS at the end of sub-domain A and beginning of sub-domain B, respectively. QxxRW ($QVLRW$) motif is indicated with bright purple box. The accession numbers are: GhCesA1, ADY68800.1; GtCesA, AEN70827.1; GbCesA1, ADY68796.1; GaCesA, ACD56660.1; GdCesA, AEN70825.1
Multiple sequence alignment of the full-length cDNA showed significant sequence similarity to the CesA genes of other species. Blast n and p analyses showed significant matches with more than 80% identity between HsCesA1 cDNA with other known CesA of higher plants, for instance: Arabidopsis (AtCesA8), Gossypium species (GhCesA1, GbCesA1), Eucalyptus species (EgCesA1, EgCesA3) and Populus species (PtCesA1, PtCesA8). In both Blast n and p analyses, the HsCesA1 cDNA matched most significantly to GhCesA1 with 90% identity and E-value 0.0. GhCesA1 is highly expressed at the initiation and during secondary cell-wall formation (Pear et al. 1996). Moreover, Li et al. (2013) stated that GhCesA1 expressed predominantly in all tissues almost like a housekeeping gene.

A homology search was done based on protein levels. The alignments of the sequences were done with several closely related genes using clustalW2. For further clarification of the gene of interest (HsCesA1) phylogenetic analysis was constructed based on amino acid sequences and using the neighbor joining method with bootstrapped (n = 1000 replicates) in MEGA version 5. This allowed the illustration of the evolutionary relationship between HsCesA1 and CesA genes from Gossypium, Eucalyptus, Populus and Arabidopsis (Figure 3). Phylogenetic analysis groups the CesA genes into two major clades with 23 operational taxonomic units (OTUs) and 21 internal nodes. Bootstrap values ranged from 71 to 100 provided a measure of strong supports of the inferred tree. Based on the phylogenetic analysis HsCesA1 was grouped with GhCesA1, GhCesA4, EgCesA1, AtCesA8, PtCesA1 and PtoCesA8 and high degree of amino acid sequence similarity suggested that HsCesA1 might be orthologous to these CesA genes. Although most of them are suspected to be involved in secondary cell wall synthesis (Endler & Persson 2011; Kim et al. 2011; Ranik & Myburg 2006; Wu et al. 2000), HsCesA1 showed the closest genetic relationship with GhCesA1 that is expressed in all tissues in Gossypium hirsutum. Therefore, further analyses are needed to corroborate the function of HsCesA1 in roselle.

**ISOLATION AND IN-SILICO ANALYSIS OF HSCESA1 FULL-LENGTH GENE**

Full length HsCesA1 gene of 5489 bp (accession no: KJ661223) was successfully isolated, which consisted of 12 exons and 11 introns and a promoter region of 737 bp.

The nucleotide sequence of the full-length HsCesA1 gene showed 82% identity and E-value 0.0 with Gossypium herbaceum var. africanum (HQ143021.1) based on Blastn analysis. Multiple sequence alignments between the full-length gene and the full-length cDNA of HsCesA1 gene with full-length gene and full-length cDNA of Gossypium hirsutum CesA1, Gossypium hirsutum CesA4, Gossypium herbaceum CesA1 and Gossypium barbadense CesA showed significant similarity in the exon-intron organization of cellulose synthase genes across species. The transcriptional start site (TSS) of HsCesA1 in the promoter region was identified using using Prediction Plant promoter positioned at 379 bp upstream of the ATG. The TATA, CAAT and GC boxes were also predicted in the promoter sequence of the HsCesA1, positioned at -36, -61 and -188 bp, respectively, from the TSS. Through promoter sequence analysis using the PlantPAN software (Chang et al. 2008), MYB, Athb-1 and AGL3 were identified as the regulatory elements in the isolated HsCesA1. The genes encoding the putative transcription factors, MYB family members are enriched during early stages of fibre development (Lee et al. 2007). Athb-1 and AGl3 were key regulators for genes expressed during leaf development and expression of Athb-1 affected the development of palpide parenchyma under normal growth conditions, resulting in light green sectors in leaves and cotyledons (Aoyama et al. 1995).

**SEMI-QUANTITATIVE RT-PCR GENE EXPRESSION STUDY**

Semi-quantitative RT-PCR expression studies of the HsCesA1 gene isolated from Hibiscus sabdariffa were carried out in both young leaf and young stem tissues. Three biological replicates were used for each tissue. Extracted total RNAs with OD260/280 ratios between 1.8 and 2 were used for each sample (Figure 4). In order to ensure the accuracy of the assay, two housekeeping genes (HKGs) were used as the reference genes in all the RT-PCR reactions. A combination of HKGs was used to reduce the possibility of risk in the use of a single HKG. Actin gene family (GhACT4) and catalytic subunit of protein phosphatase 2A (GhPP2A1) were used as internal control for RT-PCR to verify that the same amounts of RNAs were used for analysis. PCR amplifications of the HKGS and HsCesA1 gene in both leaf and stem tissues resulted in a single clear DNA fragment and sequencing of the PCR.
FIGURE 3. Bootstrapped neighbour-joining tree of the HsCesA1 protein sequence that is shown in a box and 21 CesA protein sequences from *Gossypium*, *Eucalyptus*, *Arabidopsis* and *Populus* (n = 1000 replicates). *Acystostelium subglobosum* (AsCesA) was used as out-group for this analysis. The GenBank accession numbers are in the parenthesis GhCesA1 (ADY68800.1), GhCesA3 (AAD39534.2), GhCesA4 (AAL37718.1) GhCesA5 (AFB18634.1), GhCesA6 (AFB18635.1), GhCesA7 (AFB18636.1), EgCesA1 (AAY60843.1), EgCesA3 (AAY60845.1), EgCesA4 (AAY60846.1), EgCesA5 (AAY60847.1), AtCesA1 (NP-194967), AtCesA3 (NP-19613.1), AtCesA4 (NP-199216.2), AtCesA5 (NP-196549.1), AtCesA7 (NP197244), AtCesA8 (NP_567564.1), PtoCesA1 (AAY21910.1), PtoCesA4 (AGW52080.1), PtoCesA5 (AEE60899.1), PtoCesA7 (AEE60896.1), PtoCesA8 (AEE60893.1) and AsCesA (BAK54004.1)

FIGURE 4. Gel electrophoresis of isolated total RNA from young leaf (A) and young stem (B) of *Hibiscus sabdariffa* on 1% (W/V) agarose gel. Lane M: 2- log DNA ladder
products confirmed amplification specificity. The band intensity fractions of all reverse transcription products of the three biological replicates from each tissue were analyzed using GelAnalyzer and GelQuantNET software and the expressions of the HKGs were similar in all the biological replicates across the tissues. The consistency and reproducibility of the HKGs amplifications showed that the two selected HKGs were suitable in this study. The band intensity fractions of GhPP2A1 and GhACT4 were similar in the two tissues used in this study and estimated as 0.49 and 0.46, respectively. In comparison, the HsCesA1 gene showed slight variations in the PCR product intensities between the two tissues though not significant with only a 1.3 fold of expression differences. An estimated band intensity fractions were 0.58 in the three biological replicates of young leaf tissue and 0.42 in three biological replicates of young stem tissue (Figure 5). Sequencing of the RT-PCR product and Blast analysis of HsCesA1 sequence showed that the amplification was specific. This analysis showed that the HsCesA1 expressed both in the young leaf and stem tissues.

Phylogenetic and Blast analyses supported that the isolated HsCesA1 may play roles in the cell wall depositions in both leaf and stem tissues. Based on the results from phylogenetic analysis, the HsCesA1 might be orthologous to Gossypium hirsutum CesA1 and CesA4, Eucalyptus grandis CesA1, Populus tomentosa CesA1 and CesA8 and Arabidopsis thaliana CesA8. Although AtCesA8, EgCesA1 and PtoCesA1 were orthologous genes and found to be responsible for secondary

FIGURE 5. Electrophoretic characterization of reverse transcription PCR products of HsCesA1 (A), GhCAT4 (B) and GhPP2A1 (C) on 8% (W/V) PAGE. Three biological replicates from leaf tissues (lanes 1, 2 and 3) and three biological replicates from stem tissues (lanes 4, 5 and 6)
cell wall cellulose synthesis, but EgCesA1 was also expressed in tissues undergoing primary growth (Ranik & Myburg 2006; Wu et al. 2000; Zhou 2005). According to Li et al. (2013), although GhCesA1 expressed during secondary wall synthesis, they were also found expressed predominantly in all tissues almost like a housekeeping gene. HsCesA1 showed the closest genetic relationship with GhCesA1. The results of Blast analyses also provided evidence that *Gossypium barbadense* CesA1 (GbCesA1) and GhCesA1 sequences were significantly similar to the HsCesA1. The results of the Blast analyses may suggest possible roles of HsCesA1 in different tissues of Roselle, which coincide with the results of RT-PCR.

**CONCLUSION**

We have isolated the first cellulose synthase gene from the *Hibiscus sabdariffa*, designated as HsCesA1 (full-length HsCesA1 cDNA, accession no: KJ608192; full-length HsCesA1 gene, accession no: KJ661223). Several conserved and important characteristics such as cellulose synthase domain, Zn-binding domain/ Ring finger domain and two highly conserved regions (Domains A and B) were predicted in the HsCesA1 and these characteristics strongly supported the prediction that the HsCesA1 isolated here is a true CesA belonging to the processive class of the 2-glycosyltransfrase family A. The results based on Blastn, Blastx and Plza3.0 analyses showed high sequence similarities between the HsCesA1 gene and other plant CesAs such as *Gossypium, Eucalyptus, Populus* and *Arabidopsis*. Both Blast and phylogenetic analyses suggested that the HsCesA1 might play roles in cell wall deposition in different tissues. Semi-quantitative RT-PCR expression of the HsCesA1 gene isolated also showed that it was expressed in both leaf and stem tissues. However, further investigations must be carried out to confirm the function of this HsCesA1. This study provides information on the primary structure of the HsCesA1 gene, which is fundamental for understanding the function of the gene in Roselle plant in the future. Future work should focus on isolating the different CesA isoforms from *H. sabdariffa* and the HVRII region that can be used to examine structural and functional relationships among CesA genes and define them into distinct subclasses. The current study provides the foundation for future studies on dissecting the mechanism of cellulose biosynthesis in Roselle specifically and in other non-woody plants generally.

**ACKNOWLEDGMENTS**

This study was funded by the Research University Grant Scheme- Universiti Putra Malaysia, project numbers 05-05-10-1066RU and 05-02-12-1716RU.

**REFERENCES**


