Differentially expressed transcripts from leaf and root tissue of *Chlorophytum borivilianum*: A plant with high medicinal value

Sunil Kumar a, Shikha Kalra a, Sanjay Kumar b, Jagdeep Kaur a, Kashmir Singh a,⁎

a Department of Biotechnology, Punjab University, Chandigarh-160014, India
b Biotechnology Division, Institute of Himalayan Bioresource Technology, Palampur-176061, HP, India

Abstract

*Chlorophytum borivilianum* is one of the important medicinal plants used for treating different health problems such as diabetes, arthritis, physical weakness, etc. Saponins present in *C. borivilianum* are the primary source of its significant medicinal properties and are synthesized by mevalonate and non-mevalonate pathways in plants. However, the biosynthesis of these compounds at molecular level is not studied in *C. borivilianum*. Cloning and sequencing of genes involved in metabolic processes are prerequisite to study the gene expression, their regulation and genetic engineering experiments. Expressed sequence tags (ESTs) provide a quick insight into various genes and their tissue specific expression. Suppression subtractive hybridization (SSH) libraries were constructed using mRNA from leaf and root tissues of *C. borivilianum*. High quality non-redundant 506 and 303 ESTs were generated from leaf and root specific libraries respectively. These sequences were analyzed using bioinformatics tools and grouped into different categories based on their similarity and cellular functions such as photosynthesis, metabolism, transcription factors, cell signaling, defense stress response. ESTs also showed similarity with genes involved in saponins biosynthesis such as squalene synthase, squalene epoxidase, cytochrome P450, glycosyltransferase, etc. Semi-quantitative analysis of some of the ESTs involved in saponins biosynthesis confirmed their differential regulation in leaves and roots. These ESTs will provide an efficient resource to accelerate gene discovery in *C. borivilianum* and will help in determining promising targets for genetic engineering of saponins pathway.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

*Chlorophytum borivilianum* (family Liliaceae) is among the few medicinal plants witnessing steady growth in phytopharmaceutical and nutraceutical products since 11th century A.D. (Oudhia, 2000; Santapau and Fernandes, 1994). The estimated demand of the species is about 35,000 tons annually (http://www.farmwealth.com/safedmusli; http://www.krdmuslifarm.com/safedmusli). Dried roots per se or its extract are widely used in more than 100 different herbal formulations. The formulations are implicated in curing for physical weakness, diabetes, arthritis, natal and post-natal problems and so on. The species is also referred to as 'Herbal Viagra' (Ranawat et al., 2000). The roots of the species consist of carbohydrates (42%), proteins (5-10%), fiber (3-4%) and saponins (17–20%) (Bordia et al., 1995). Medicinal properties of the plant are attributed to the presence of saponins (Chakraborty and Arri, 2001).

Saponins belong to the terpenoid class of compounds synthesized through isoprenoid pathway in the plants (Fenwick et al., 1991; Herttemann and Marston, 1995; Maga, 1994). These are derived from five-carbon isoprene units and synthesized from two independent routes: mevalonate (MVA) pathway (in cytoplasm) and non-mevalonate pathway (in plastids) (Fig. 1A and B) (Rohdich et al., 1997). Both these pathways synthesize larger class of plant secondary metabolites. MVA pathway starts with the condensation of three molecules of acetyl-CoA while in methylerythritol phosphate (non-mevalonate) pathway, isopentenyl diphosphate (IPP) is formed from pyruvate and glyceraldehyde-3-phosphate (Rohmer, 2003). Both of these pathways yield the activated isoprene units, dimethylallyl diphosphate (DMAPP) and IPP (Davold et al., 1997). IPP is the structural unit of isoprenoid pathway. Saponins are synthesized from IPP via the cyclization of 30 carbon intermediate squalene to synthesize 2,3-oxidosqualene. This cyclization event
followed by various modifications, such as oxidation, substitution, and glycosylation leads to the formation of various types of saponins including steroidal and triterpene saponins.

MVA and non-MVA pathways are central to the synthesis of a large number of important secondary compounds and are important pathways in secondary metabolism research. A lot of emphasis is being paid to clone and analyze the genes involved in these pathways to understand their regulation (Carelli et al., 2011; Kumar et al., 2012; Naoumkina et al., 2010; Singh et al., 2010). However, the detailed information, particularly on the genes involved in late stages of saponins biosynthesis is still lacking (Fig. 1).

Among the species of Chlorophyllum, C. borivilianum yields highest saponin contents vis-à-vis vigorous root system (Bordia et al., 1995). Steroidal saponins, known as borivilianosides, are considered to be the main bioactive compounds present in C. borivilianum (Achariya et al., 2009). However, there is little information on molecular data on this species. Expressed sequence tag (EST) generation and their analysis have proven to be an invaluable tool in the identification of genes involved in various metabolic processes in plants as well as in animals (Ohlrogge and Benning, 2000; Singh et al., 2009a, 2009b). The present work reported differential gene expression profile in leaf and root tissues of the plant using cDNA subtractive hybridization approach and established the gene function with tissue specific expression and the content of pharmaceutically active compounds, saponins.

2. Materials and methods

2.1. Plant material and growth conditions

Plants of C. borivilianum were grown in pots under controlled conditions in the growth chamber with day/night temperature of 22 ± 1 °C, 16 h photoperiod (flux density 200 μmol m⁻² s⁻¹) at Panjab University, Chandigarh (India) [latitude: 30° 44'14 N; longitude 76°47'14E; altitude 350 m above mean sea level].
2.2. RNA isolation and construction of cDNA subtractive library

Three months old plants were removed from soil and cleaned thoroughly. Leaf and root were harvested separately using snap freezing technique and stored at −80 °C till further use. Total RNA was isolated from tissues following the protocol developed by Ghawam et al. (2011) and mRNA was extracted from: total RNA using Oligotex mRNA Kit (Qiagen, Germany).

The forward and reverse suppression subtractive hybridization (SSH) libraries were constructed using the PCR Select CDNA Subtraction Kit (Clontech, USA) following the manufacturer’s instructions. In brief, poly (A)+ mRNA (2 μg) was used for the synthesis of cDNA from leaf and root separately. cDNAs were digested with Rsal at 37 °C for 1.5 h for the creation of driver population for leaf and root. For tester cDNA, adapters were ligated to the digested cDNA. In the forward library, leaf cDNA was used as tester and root cDNA as driver and vice-versa during preparation of reverse library. First and second round of hybridizations using tester and driver populations were carried out following the manufacturer’s instructions. After hybridization, two rounds of PCR amplification were performed to amplify differentially expressed cDNA using the 5’-3’ oligonucleotide complementary to adapter sequences provided by the manufacturer. PCR products were cleaned up using the GenElute PCR purification Kit (Sigma, USA).

The purified subtracted cDNA fragments were cloned in pGEM®-T Easy vector (Promega, USA) to make separate forward and reverse libraries. Ligated products were transformed into E. coli DH5α competent cells and plated onto LB medium containing ampicillin, isopropyl-beta-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal). Recombinant plasmids were isolated using plasmid isolation kit (BIO-RAD, USA).

Table 1

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’→3’)</th>
<th>Tm (°C)</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>26S rRNA</td>
<td>F: CACCATATATGCAGAGAGCAC CAGGAGAC 56 °C 18</td>
<td>R: CAGGCGCCCGCTCCGAGGCACT CGAAG 63 °C</td>
<td></td>
</tr>
<tr>
<td>SQS</td>
<td>F: CTGGGCTTTGAAAGACCAGCT 59 °C 35</td>
<td>R: TAGGGCTACTCTGGATGATAG 59 °C</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>F: GACCTGAGGCGGAGACACTA 62 °C 30</td>
<td>R: GCTAAGCTTCAACCACCA 62 °C</td>
<td></td>
</tr>
<tr>
<td>CYP (K035188)</td>
<td>F: GTCGACATGCGATGCCGAGAGAC 69 °C 27</td>
<td>R: GATCGACATGCGATGCCGAGAGAC 69 °C</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>F: TTCACACTCTCAGTCACACAC 62 °C 30</td>
<td>R: TTTCACACTCTCAGTCACACAC 62 °C</td>
<td></td>
</tr>
<tr>
<td>CYP (K035449)</td>
<td>F: GTCGACATGCGATGCCGAGAGAC 68 °C 28</td>
<td>R: GATCGACATGCGATGCCGAGAGAC 68 °C</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>F: GACACGACGCGCTACCTGTGAT 68 °C 27</td>
<td>R: CTTCACACTCTCAGTCACACAC 68 °C</td>
<td></td>
</tr>
<tr>
<td>3’UTR (J03520)</td>
<td>F: GAGCTCGAGGCGACAGACAC 68 °C 20</td>
<td>R: CTTCACACTCTCAGTCACACAC 68 °C</td>
<td></td>
</tr>
</tbody>
</table>

2.3. Sequencing and sequence analysis

ESTs were generated by sequencing the recombinant plasmids on 3130xl genetic analyzer (Applied Biosystems, USA) using Big Dye Terminator v3.1 Cycle Sequencing Kit. The purified subtracted cDNA fragments were cloned in pGEM®-T Easy vector (Promega, USA) to make separate forward and reverse libraries. Ligated products were transformed into E. coli DH5α competent cells and plated onto LB medium containing ampicillin, isopropyl-beta-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal). Recombinant plasmids were isolated using plasmid isolation kit (BIO-RAD, USA).

The purified subtracted cDNA fragments were cloned in pGEM®-T Easy vector (Promega, USA) to make separate forward and reverse libraries. Ligated products were transformed into E. coli DH5α competent cells and plated onto LB medium containing ampicillin, isopropyl-beta-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal). Recombinant plasmids were isolated using plasmid isolation kit (BIO-RAD, USA). Sequences were filtered for vector and adapter sequences and high quality non-redundant 506 ESTs showing significant similarity with database were further grouped on the basis of their putative function by using various databases and tools such as GO, KEGG, etc. The filtered ESTs were submitted in NCBI EST database.

2.4. Dot-blot analysis

Plasmids isolated from SSH libraries (root and leaf) were denatured with 2 N NaOH containing 2 mM EDTA and dotted onto nylon membranes (Hybond NX, GE). DNA and membranes were crosslinked by baking at 80 °C for 2 hours. Total RNA (20 μg) (from same RNA pool that was isolated from leaves and roots to make SSH library) was used to make cDNA probes using P32 labeled radioactive dATP and reverse primer RNA labeling Kit (Geist Aster Corp., USA). Labeled probes were purified using QIAquick Nucleotide Removal Kit (Qiagen, Germany) for removing un-incorporated nucleotides. Radiolabeled probes were hybridized with blotted DNA on membrane by using ExpressHyb hybridization solution (BD Biosciences, USA). After hybridization, membranes were washed to remove excess of probes and exposed to phosphoimaging screen. Images were captured using BioRad FX phosphor imaging machine (Bio-Rad, USA). The intensity of each dot was estimated by array analysis tools of AlphaEaseFC software (Alpha Innotech, USA).

2.5. Expression analysis of putative genes involved in saponins biosynthesis

Expression analysis of some of the predicted genes involved in saponins biosynthesis was performed by semi-quantitative RT-PCR. Gene specific primers were designed for the selected genes (Table 1). Equal amount of RNA isolated from leaf and root tissue was digested with DNasel (Amplification grade, Invitrogen USA) and reverse transcribed to make cDNA using Oligo(dT)12-18, and superscript III reverse transcriptase (Invitrogen, USA). Primers of 26S rRNA gene served as an internal control (Singh et al., 2004). PCR conditions were standardized for quantitation. PCR was performed at different number of cycles and cycle number that showed the amplification within the exponential phase was chosen for PCR (Table 1). The intensity of the PCR bands was measured with spot densit analyzing tools of AlphaEaseFC software (Alpha Innotech, USA). Experiments were conducted using three biological replicates.

3. Results and discussion

3.1. Library sequencing, sequence assembly and analysis

ESTs provide a quick route for discovering new genes, studying gene expression and regulation, and for constructing genome maps. C. borivilianum is an important medicinal plant and we made an effort to study genes differentially expressing in leaves and roots of the plant. We analyzed the total saponins in both tissues by isolating saponins by soxhletation method and analyzed by thin layer chromatography (TLC) (Brain and Turner, 1975). Based on our observation by TLC (data not shown) and from previous studies, that roots contain high amount of saponins as compared to the leaves, it was hypothesized that the difference may be attributed to differential expression of their respective genes. Thus, leaf (forward) and root (reverse) specific EST libraries were prepared by suppression subtractive hybridization and sequenced. A total of 650 and 450 ESTs were obtained from leaf and root libraries respectively. Sequences were filtered for vector and adapter sequences and high quality non-redundant 506 ESTs selected for gene expression and regulation, and for constructing genome maps.
and 303 sequences were selected for further analysis from leaf and root library respectively with a size range between 150 bp to 1100 bp. The sequences were submitted to NCBI EST database (Forward: JK036481-JK036940 and Reverse: JK036230-JK036480, JK492885-JK492905). ESTs were annotated bioinformatically based on their putative functions using sequence analysis tools. Broadly, ESTs were categorized in three groups viz. significant, non-significant and hypothetical sequences (Fig. 2a). Approximately 55% ESTs in the forward library and 20% in the reverse library had low redundancy and did not show any similarity to NCBI database and were placed in a non-significant match group. Ten percent of the ESTs from forward and 18% from reverse library showed similarity with the hypothetical proteins, which shows the less characterized functional properties but not a specific group of gene. Remaining 35% ESTs for forward and 62% ESTs from reverse library showed significant hits with the database. These sequences were categorized on the basis of their putative functions (Tables 1 and 2) as: photosynthesis genes, metabolism, biosynthesis, cellular organization, nuclear organization, transport proteins, transcriptional regulation, ion carrier, cell signaling, defense response, stress related genes, protein folding, secretory proteins and genes with unknown functions. Fig. 2(b) represents the percentage of gene expression in each category of both libraries. The number of genes varies in both the tissues and some were exclusively present only in one tissue (e.g. photosynthetic genes only in leaves and protein folding and secretary proteins only in roots). Approximately 10% ESTs of known gene were involved in photosynthesis out of which chlorophyll a/b-binding protein, RUBISCO and oxygen evolving complex associated genes were abundant whereas abundance of UV-B and Mg-protoporphyrin IX monomethyl EC1 ESTs were less as compared to other photosynthetic gene. For transcription regulation, 7% and 20% ESTs were identified from forward and reverse library respectively. Transcription factors involved in metabolite regulation such as myb1, zinc knuckle and ethylene-responsive element binding proteins (EREBP) were also identified. ESTs (10%) from leaf specific library showed similarity with ions carrier protein, most of which are involved in photosynthesis regulation pathway such as: ferredoxin, photocytinin, thiodoxin f-type and glutaredoxin. Five percent of transcripts from roots included ions carriers were involved in regulation of metabolic function, for example cytochrome P450 (CYP). Out of these ion carriers, CYPs possess an important role in biosynthesis of secondary metabolites. ESTs related to cell signaling were also sequenced from forward (5%) and reverse (3%) libraries, such as lipid transfer proteins (JK036751), C3HC4 type zinc finger (JK0369635) (expressed in leaf) and calmodulin and phosphatidyl inositol (JK036269 and JK036477) (expressed in root). Defense responsive proteins such as, glyoxalate aminotransferase were found in leaves while nucleotide-binding site-leucine-rich repeat (NBS-LRR), resistance transcripts were found in roots. Protein folding genes such as chaperon and secretary protein, multi-copper oxidase were highly abundant in roots. Metabolism and biosynthetic genes were also present in abundance in both the libraries. SSH library was an efficient approach for enriching the differentially expressed genes but there are still chances of some false positive clones in SSH library. Thus, to further confirm the differential expression of genes, dot blot analysis was performed using cDNA as probe from leaf and root tissues. 192 clones each from forward and reverse libraries were randomly selected for dot blot. The analysis indeed proved that genes were differentially expressed (Supplementary Fig. 1a and b) and their expression behavior (up- or down regulated) is shown in Tables 2 and 3 on the basis of integrated density value.

![Figure 2](image-url)
Table 2
List of candidate ESTs obtained from leaf specific forward library and their assigned groups based of the putative function. Using dot blot analysis (Supplementary Fig. 1a) integrated density values of some of these were measured and their expression was compared in leaf and root (↑ means up-regulated and ↓ means down-regulated).

<table>
<thead>
<tr>
<th>Clone ID</th>
<th>Accession no</th>
<th>Description of Blast X hit in Gene Bank</th>
<th>e-Value</th>
<th>Leaf</th>
<th>Root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ribulose 1.5 bisphosphate 1</td>
<td>2e−43</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LHC a/b-binding protein 1</td>
<td>5e−26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type III chlorophyll a/b-binding protein</td>
<td>4e−33</td>
<td>23.69</td>
<td>17.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LHCA1 type I chlorophyll a/b-binding protein</td>
<td>2e−28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cbf125</td>
<td></td>
<td>23 kDa OEC protein</td>
<td>3e−41</td>
<td>168</td>
<td>32.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS II 22 kDa polypeptide</td>
<td>3e−67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choroplast PS II 19 kDa protein</td>
<td>3e−1</td>
<td>146</td>
<td>59.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSI reaction center subunit N 1</td>
<td>6e−29</td>
<td>118</td>
<td>74.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mg-photosynthetic oxygen-evolving ECT</td>
<td>7e−23</td>
<td>55</td>
<td>65.66</td>
</tr>
<tr>
<td>Cbf76</td>
<td></td>
<td>UDP-glucose-6-phosphate dehydrogenase</td>
<td>7e−10</td>
<td>9</td>
<td>70.46</td>
</tr>
<tr>
<td>Cbf303</td>
<td></td>
<td>δ-glucosidase 1</td>
<td>7e−10</td>
<td>9</td>
<td>70.46</td>
</tr>
</tbody>
</table>

3.2 Identification of putative candidate genes involved in the saponin biosynthesis pathway

Many ESTs were identified showing significant matching with genes involved in biosynthesis of saponins (Table 4). Clone ID cbf1 293 (JK036773) showed similarity with CupR/T2 gene which encodes transcriptase. Transcriptase initiates the synthesis of 1-deoxy-xylulose-5-phosphate by conversion of pyruvate and glyceraldehyde-3-phosphate through the non-mevalonic acid pathway (Rodney et al., 1998), which behaves as the plastidial IPP precursor of the iso-prenoid biosynthesis. Prenylation is the intermediate process of iso-prenoid pathway which is involved in membrane association and specific protein-protein interactions by prenyltransferase family protein (Casey and Seabra, 1996). Prenyltransferase family protein, Rub-
Table 3
List of candidate ESTs obtained from root specific reverse library and their assigned groups based on the putative function. Using dot blot analysis (Supplementary Fig. 1b) integrated density values of some of these were measured and their expression was compared in leaf and root (* means up-regulated and + means down-regulated).

<table>
<thead>
<tr>
<th>Clone I.D.</th>
<th>Accession no</th>
<th>Description of Blast X hit in gene Bank e-Value</th>
<th>Clone id on blot</th>
<th>Integrated density value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcription factor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 40</td>
<td>JK036268</td>
<td>FMR1 (FMR-1 IMPAIRED RESPONSE 1);</td>
<td>4e 0 8</td>
<td>–</td>
</tr>
<tr>
<td>Chrd 247</td>
<td>JK036727</td>
<td>Translation initiation factor 4A2</td>
<td>3e 0 50</td>
<td>–</td>
</tr>
<tr>
<td>Chrd 50</td>
<td>JK036728</td>
<td>Reverse transcriptase</td>
<td>1e 0 16</td>
<td>–</td>
</tr>
<tr>
<td>Chrd 110</td>
<td>JK036313</td>
<td>C2H2L domain class transcription factor</td>
<td>1e 0 28</td>
<td>–</td>
</tr>
<tr>
<td><strong>Transport facilitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 24</td>
<td>JK036273</td>
<td>Pre-mRNA processing; splicing factor</td>
<td>6e 0 09</td>
<td>–</td>
</tr>
<tr>
<td>Chrd 45</td>
<td>JK036274</td>
<td>Multicopper oxidase</td>
<td>2e 0 31</td>
<td>–</td>
</tr>
<tr>
<td><strong>Signal transduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 19</td>
<td>JK036238</td>
<td>Zinc finger C-x3-C-x5-C-x3-H type protein</td>
<td>3e 0 79</td>
<td>50</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 216</td>
<td>JK036477</td>
<td>EFEMP1</td>
<td>5e 0 18</td>
<td>112</td>
</tr>
<tr>
<td><strong>Metabolism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 8</td>
<td>JK036299</td>
<td>Multicopper oxidase L1S</td>
<td>7e 0 12</td>
<td>126</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 2</td>
<td>JK036312</td>
<td>ADP; ATP Carrier protein</td>
<td>2e 0 11</td>
<td>39</td>
</tr>
<tr>
<td><strong>Signal transduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 5</td>
<td>JK036298</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td>Chrd 75</td>
<td>JK036555</td>
<td>NAR2.3</td>
<td>4e 0 22</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 71</td>
<td>JK036598</td>
<td>Reverse transcriptase</td>
<td>4e 0 20</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 13</td>
<td>JK036324</td>
<td>Multicopper oxidase</td>
<td>5e 0 09</td>
<td>–</td>
</tr>
<tr>
<td><strong>Transport facilitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 160</td>
<td>JK036388</td>
<td>Multicopper oxidase</td>
<td>7e 0 12</td>
<td>126</td>
</tr>
<tr>
<td><strong>Signal transduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 226</td>
<td>JK036447</td>
<td>EFEMP1</td>
<td>5e 0 18</td>
<td>112</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 70</td>
<td>JK036298</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036242</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 96</td>
<td>JK036324</td>
<td>Multicopper oxidase</td>
<td>2e 0 50</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 227</td>
<td>JK036447</td>
<td>EFEMP1</td>
<td>5e 0 18</td>
<td>112</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 28</td>
<td>JK036299</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 210</td>
<td>JK036431</td>
<td>Glucosidase S-transferase</td>
<td>4e 0 40</td>
<td>52</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 23</td>
<td>JK036251</td>
<td>ADP; ATP Carrier protein</td>
<td>3e 0 09</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 31</td>
<td>JK036242</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 12</td>
<td>JK036324</td>
<td>Multicopper oxidase</td>
<td>5e 0 09</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 27</td>
<td>JK036425</td>
<td>Multicopper oxidase</td>
<td>5e 0 24</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 112</td>
<td>JK036340</td>
<td>ATPase</td>
<td>3e 0 24</td>
<td>31</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 34</td>
<td>JK036234</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 147</td>
<td>JK036475</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 33</td>
<td>JK036243</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 147</td>
<td>JK036475</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrd 15</td>
<td>JK036244</td>
<td>Multicopper oxidase</td>
<td>4e 0 45</td>
<td>–</td>
</tr>
<tr>
<td><strong>Cellular organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
geranylgeranylantransferase (clone ID chrsl21, JK036250), catalyzes the transfer of isoprenoids from GGPP to the cysteine residues of the protein substrates and form geranylpyrophosphate (GPP-C-10) (Crowell, 2000). Farnesyl synthase is involved in the condensation of GPP to IPP in head-to-tail manner leads to form farnesyl pyrophosphate. Clone chrsl 461 (JC312921) showed significant similarity with squalene synthase (SQS). SQS is the key regulatory enzyme of the isoprenoid pathway which is involved in the condensation of two FPP units in tail to tail manner to form squalene (C30) (Holstein and Hohl, 2004). Oxidation of squalene to 2, 3-oxidosqualene (C30) is regulated by squalene epoxidase (SE, JC312922) (Abe et al., 1993; Haralampidis et al., 2002). 2,3-oxidosqualene is considered to be the last common precursor of triterpenoid saponins, phytosterols and steroidal saponins (Kalinsowska et al., 2005; Phillips et al., 2006). SE mediated oxidation produces 2, 3-oxidosqualene that yield 80 different types of triterpene skeletons (Phillips et al., 2006; Xu et al., 2004). Epoxide hydrolase (KC36686) belongs to a class of proteins that catalyze the hydratation of chemically reactive epoxides to their corresponding diallylalcohol products. Hydration leads to more stable and less reactive intermediates which are converted into cyclic derivatives via pyroteon and produce a carbocation that can undergo several types of cyclization reactions (Freeland and Omiecinski, 2000). The cyclization and rearrangement mechanism are regulated by different cyclase enzymes, cytochrome P450, hydroxysterol synthase, lanosterol synthase, -amyrin synthase that lead to the synthesis of triterpenoid and steroidal saponins (Abe et al., 1993; Haralampidis et al., 2002; Thoma et al., 2004; Wendl, 2005; Wendl et al., 2000). Modification reactions create an enormous diversity of plant natural products by providing new molecules with different biological activities. Plants contain a large number of enzymes that catalyze hydroxylation, epoxidation, arnigination, glycosylation, acylation, prenylation, oxidation and reduction of secondary metabolites. Steroidal sapogenins are biosynthesized from phytosterol via series of oxygenation and hydrogenation steps (Dewick, 2001). Chrsl 100 (J036388) clone expressed in roots showed similarity with CYP of Oryza sativa. CYPs are versatile biocatalysts and their activities include simple hydroxylation or epoxidation, dealkylation, isomerization and arnigination. These enzymes catalyze NADPH-dependent regio- and stereospecific oxygenizations of lipids, phenolics, terpenoids and alkaloids. EST matching with CYP (J036388) was obtained from reverse library with a possible role in catalyzing the introduction of additional oxygen functions into triterpene skeletons. The role of CYP in hydroxylation of triterpenes was presumed by the presence of conserved domain “PCCPRVPCC” (Shimada et al., 2001) and on their similarity with other such sequences in database. Clone chrsl 110 (J036338) showed significant matching with serine proteases of the peptidase family SM in- volved in the amino acid transport and the metabolism of esterases and lipases. The catalytic apparatus involves three residues (catalytic triad): a serine, a glutamate or an aspartate and a histidine. These enzymes act on carboxyl end, Serine carbonylpeptide-like protein (SCP1,SAD7) catalyze coupling of N-methyl anthranilic and benzoic acid to sapogenins in averacin during synthesis of triterpenoid sapogenins in oats (Mugford et al., 2000). Among the reactions for the modification of secondary metabolite scaffolds, glycosylation plays a particularly important role in plants, contributing to the biosynthesis and storage of secondary metabolites, regulation of hormone homoeostatis, detoxification of xenobiotics, enhancement of molecule solubility, and plant defense (Bolwell et al., 1994). Glycosylation is the last step in the biosynthesis of saponins and occurs prior to their entry into the vacuole (Hosettemann and Marston, 1995). The process is regulated by multiple gene family of glycosyltransferases with different domains. Various glycosyltransferases were obtained from root library of C. borivillumum. Out of four, two GTs (JK036259, JK036464) contained the highly conserved consensus sequence, PSPG (plant secondary product glycosyltransferase domain) Glycosyltransferases involved in the amino acid transport and the metabolism of esterases (Raedar and Hughes, 1994). Clone ID Chrsl 248 (JK036469) showed similarity of small molecules (UGTs) catalyze the transfer of sugar residues from uridine diphosphate sugars to an acceptor. The sugar moiety can be transferred to oxygen, nitrogen, or sulfur atoms of different classes of natural products (Dewick, 2001). The transfer of sugar moieties from activated donor molecules to specific acceptor molecules form glycosidic bonds in cytosol before these conjugates are transported and stored in the vacuoles. These molecules are stored in the form of bidesmosidic, called furanoyl-type glycosides (Hughes and Hughes, 1994). Clone ID Chrsl 248 (JK036469) showed similarity glycosyltransferase of Populus trichocarpa and was highly expressed in leaves. Clone ID Chrsl 42 (JK036270) significantly match with glycosyltransferase 47 of Arabidopsis thaliana, involved in secondary cell wall synthesis (Brown et al., 2005). -beta-glucosidase, which specifically cleaves steroid glycosides and hydrolyses furanoyl type saponins at C-26 position leading to the formation of spirotane type saponins were also indentified (Osborne, 1996; Roddick, 1987). EST clone Chrsl 89 and 40 (JK036569, JK036520) showed significant matching with -beta-glucosidase, responsible for the activation of saponins. ABC transporter supports the intracellular transportation of alkaloids in plants (Rea et al., 1998; Tzirakska et al., 2003; Yataki, 2005, 2006). Clone ID Chrsl 210 (J036431) encodes for the enzyme glutathione-S-transferase, which acts as metabolite binding protein that regulates the transportation of secondary metabolites through vesicle trafficking (Collins et al., 2003; Mueller et al., 2000). ATP/EF-hand domain containing genes, a kind of plant transcription factors which regulates the biogenesis of terpenoid indole alkaloids, were also identified (Memelink et al., 2001; Van der Fits and Memelink, 2000). Many of the ESTs were found to be novel with no match to the database, analysis of these novel ESTs may provide...
new information into the mechanism of primary and secondary metabolisms and other metabolic processes in C. borivilianum.

3.3. Validation of expression of putative genes involved in saponin biosynthesis

Putative genes involved in saponin biosynthesis were selected for RT-PCR based semi-quantitative assay to further confirm their differential expression in leaf and root (Fig. 3). PCR was performed for selected number of cycles for each gene to obtain amplification within exponential phase. Based on the integrated density value (IDV), SQS gene expression level in leaf was 43% higher than in root. In a similar trend, the expression level of SE, GT (JK036520) and β-glucosidase (JK036525) was higher in leaves by 8%, 45% and 46% respectively than in roots. Conversely, the expression level of CYP (JK036388) and GT (JK036340, JK036469) was higher in roots by 15%, 40% and 12% respectively as compared to leaves. GT (JK036340, JK036469) has conserved PSPG domain, involved in glycosylation of secondary metabolites. As reported in Achyranthus bidentata, saponins were first synthesized in the leaves and then transported to the roots for storage via the phloem of veins and stem vascular bundles (Jinting and Zhenghai, 2008). Expression data supports that leaf is the main site of isoprenoid pathway in this plant. On the contrary, higher IDV data in roots of CYP and GT (with PSPG domain) indicated that root is the modification site of saponins before storage. On the other hand, GT (JK036207) has higher expression in leaf, and also lacks PSPG domain, thus it may not be involved in modification of saponins and has some other role in the plant. On the basis of IDV data, higher expression of β-glucosidase was observed in leaf but substrate location site is the root. Physical separation of β-glucosidase enzymes from their substrates by differences in subcellular or tissue location are well known in plants. Steroid glucosides, oligofluorostanoids in Dioscorea caucasica are located in the leaves, while the specific β-glucosidase enzyme resides in the roots (Gurielidze et al., 2004). This difference in tissue localization of substrates and enzyme prevents theurosolanum structure from being cleaved during its transport from leaves to roots. However, such a hypothesis requires further studies for verification.

4. Conclusion

This study in C. borivilianum on differentially expressed ESTs will help in understanding the plant metabolism at mo ecular level especially in relation to genes involved in saponins biosynthesis and their regulation in different tissues. Early genes of the pathway are more expressed in leaves while later genes which are involved in saponins modifications such as CYPS and GTs have higher expression at storage site i.e. roots. Such an expression pattern of genes may further help in tracking the specific functions of these genes. These ESTs will be used for full length cloning and functional analysis of the genes involved in saponins biosynthesis and will provide an efficient resource for of gene discovery which may be novel for this plant. This functional genomics approach will accelerate gene discovery in this plant and provide promising targets for genetic engineering of saponins biosynthesis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.09.046.

Acknowledgments

Authors are thankful to CSIR, India for providing financial assistance to the project. Authors would like to thanks Mr. Digvijay Singh, Technical assistant, IHBT, Palampur for his help in sequencing services.

References


Characterization of Squalene synthase Gene from Chlorophytum borivilianum (Sant. and Fernand.)

Shikha Kalra • Sunil Kumar • Neha Lakhanpal • Jagdeep Kaur • Kashmir Singh

© Springer Science+Business Media New York 2013

Abstract Saponins are important group of secondary metabolites known for their pharmacological properties. Chlorophytum borivilianum contains high amount of saponins and is thus, recognized as an important medicinal plant with aphrodisiac properties. Though the plant is well known for its pharmaceutical properties, there is meager information available about the genes and enzymes responsible for biosynthesis of saponins from this plant. Squalene synthase (SqS) is the key enzyme of saponin biosynthesis pathway and here, we report cloning and characterization of SqS gene from C. borivilianum. A full-length CbSqS cDNA consisting of 1,760 bp was cloned which contained an open reading frame (ORF) of 1,233 bp, encoding a protein of 411 amino acids. Analysis of deduced amino acid sequence of CbSqS predicted the presence of conserved isoprenoid family domain and catalytic sites. Phylogenetic analysis revealed that CbSqS is closer to Glycine max and monocotyledonous plants. 3D structure prediction using various programs showed CbSqS structure to be similar to SqS from other species. C-terminus truncated recombinant squalene synthase (TruncCbSqS) was expressed in E. coli M15 cells with optimum expression induced with 1 mM IPTG at 37 °C. The gene expression level was analyzed through semi-quantitative RT-PCR and was found to be higher in leaves as compared to the roots.

Keywords Chlorophytum borivilianum • Aphrodisiac • Saponins • Squalene synthase • Gene expression

Introduction

Plants are sessile organisms that have evolved themselves with the ability to synthesize various chemical compounds referred to as “secondary metabolites” to defend them against insects, fungi, herbivores, and abiotic stresses. These secondary metabolites, while being toxic to plant predators have very specialized functions and are often exploited by humans due to their beneficial effects in terms of nutritional values and treatment of human diseases. The World Health Organization (WHO) estimates that 80 % of the world’s population presently uses herbal medicine for some aspect of primary health care (http://www.who.int/mediacentre/factsheets/fs134/en/).

Chlorophytum borivilianum Sant. and Fernand (family Liliaceae) is a monocot herb witnessing steady growth in phyto-pharmaceutical and nutraceutical products. Tuberous roots of the plant have been in use for aphrodisiac and health promotion purposes since 11th century AD [1]. The plant has drawn huge attention due to its numerous pharmaceutical properties such as aphrodisiac [2], anti-carcinogenic [3], anti-diabetic [4], anti-inflammatory [5], and anti-arthritis [6]. The main constituents of the dried roots of
C. borivilianum are: (a) carbohydrates—42%; (b) proteins—5–10%; (c) fiber—3–4%; and (d) saponins—2–20%. Saponins are primarily responsible for imparting medicinal properties to this plant.

Saponins constitute an important group of secondary metabolites that are widespread throughout the plant kingdom. Saponins are biosynthesized from the isoprenoid pathway through farnesyl pyrophosphate (FPP) by the cyclization of 2,3-oxidosqualene, leading to the formation of steroidal skeletons such as spirostanol and furostanol type [8]. This is subsequently followed by structural modifications such as oxidation, substitution, and glycosylation through various enzymes resulting in the formation of various types of steroidal saponins (Fig. 1).

Squalene synthase (SqS, EC 2.5.1.21) is a membrane bound enzyme that catalyzes the first committed step in sterol and triterpenoid biosynthesis [9]. It is a bifunctional enzyme that catalyzes the condensation of two molecules of FPP to form presqualenediphosphate (PSPP) and then converts the PSPP to squalene in the presence of NADPH and Mg$^{2+}$ [10]. The enzyme plays an important role in the regulation of isoprenoid biosynthesis [11]. In plants, the gene has been isolated from species such as Lotus japonicus, Zea mays, etc. [12–15]. Enhanced expression of SqS genes in plants such as Panax ginseng [16] and Eleutherococcus senticosus [17] resulted in increased levels of phytosterol and triterpene accumulation, thus depicting the important regulatory role of SqS. Kim et al. [18] studied that an increased expression of SqS in Bupleurum falcatum resulted in an increased mRNA accumulation of the downstream genes of the pathway and enhanced production of phytosterol and saikosaponins in the plant.

Chlorophytum borivilianum has been found to be an ideal aphrodisiac with no negative side-effects [19]. Characterization of the genes involved in the saponin biosynthesis pathway and their regulation will help in development of plants with altered saponin contents. Of lately, several reports on the biochemical and molecular studies have been reported from C. borivilianum in lieu to understand the biosynthetic pathway of saponins and its several medicinal properties [5,19]. We have developed 506 and 303 expressed sequence tags (ESTs) from the leaf and root tissues respectively from this plant [19] and to add on to the ongoing investigations, we report full-length cloning and characterization of SqS (designated as CbSqS) from C. borivilianum.

Materials and Methods

Plant Material and Growth Conditions

Chlorophytum borivilianum plants were maintained in pots in the green house at Panjab University, Chandigarh [latitude: 30°44′14N and longitude 76°47′14E and altitude 350 m above sea level]. Young leaves and roots were collected from 8 weeks old plant during the period of its active growth (June, average outside temperature 38 °C), snap frozen in liquid nitrogen and transferred to −80 °C freezer till further use.

RNA Isolation and cDNA Synthesis

Total RNA was isolated from leaves and roots using the method described by Ghawana et al. [20]. RNA was treated with DNaseI (Amplification Grade; Invitrogen, USA) to remove contaminating genomic DNA. First-strand cDNA was synthesized using 2 μg of DNA free RNA, Superscript III First-Strand Synthesis Kit (Invitrogen, USA) and an oligo dT12–18 Primer (Invitrogen, USA) in a total volume of 20 μl [19].

Cloning of CbSqS

A set of degenerate primers as mentioned in Křížík et al. [21] (Table 1) were used for the cloning of CbSqS. Using the leaf cDNA as template, PCR amplification of CbSqS was carried out using the cyclic parameters as; initial denaturation at 94 °C for 3 min followed by 35 cycles of
Table 1  Primer sequences used for various PCR reactions

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Degenerate primers</strong></td>
<td></td>
</tr>
<tr>
<td>SQS F</td>
<td>5'-TA(CT)TC(CT)CA(CT)TA(CT)GTTGCTGGI(CT)TIGT(AGCT)GG-3'</td>
</tr>
<tr>
<td>SQS R</td>
<td>5'-ATIGCCATIAC(CT)TG(AGCT)GG(AG)AT(AGCT)GC(AG)CA(AG)AA-3'</td>
</tr>
<tr>
<td><strong>RACE Primers</strong></td>
<td></td>
</tr>
<tr>
<td>SQS 3'</td>
<td>5'-GGGCTGGTGGGCTAGGATTGTCTAA-3'</td>
</tr>
<tr>
<td>SQS 3' nested</td>
<td>5'-GTATGTTTTGGCCTCGCCAGATCTG-3'</td>
</tr>
<tr>
<td>SQS 5'</td>
<td>5'-TTACCATTATTATAAGCAAAGGTAAAGTTGCCGT-3'</td>
</tr>
<tr>
<td>SQS 5' nested</td>
<td>5'-CTGTAAGGCACTTCTTACATGAACACACGC-3'</td>
</tr>
<tr>
<td><strong>Bacterial expression primers</strong></td>
<td></td>
</tr>
<tr>
<td>CbSqs(ATG) F</td>
<td>5'-ATGGGAACATTGAGGGCGATTTTGAAGA-3'</td>
</tr>
<tr>
<td>CbSqs(ATG) R</td>
<td>5'-CTAAGATCGGTTGCCAGAAAGTTGTG-3'</td>
</tr>
<tr>
<td>Trinc R</td>
<td>5'-ATGCTGGGTCTGGCTTTTATACAGGC-3'</td>
</tr>
<tr>
<td><strong>Expression analysis primers</strong></td>
<td></td>
</tr>
<tr>
<td>SQS F</td>
<td>5'-GTGCAATGCTTGATACGACAT-3'</td>
</tr>
<tr>
<td>SQS R</td>
<td>5'-TAAGCCATGACCTGAGGGATAG-3'</td>
</tr>
<tr>
<td><strong>Internal control primers</strong></td>
<td></td>
</tr>
<tr>
<td>26SrRNA F</td>
<td>5'-CACAATGATAGGAAGAGCCGAC-3'</td>
</tr>
<tr>
<td>26SrRNA R</td>
<td>5'-CAAGGGAACGGGCTTGGCAGAATC-3'</td>
</tr>
<tr>
<td><strong>RACE primers (provided by the supplier)</strong></td>
<td></td>
</tr>
<tr>
<td>UPM F/R</td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>5'-CTAATACGACTCCTATAGGGCAGCAGT-3'</td>
</tr>
<tr>
<td>Long</td>
<td>5'-CTAATACGACTCCTATAGGC-3'</td>
</tr>
<tr>
<td>NUP F/R</td>
<td>5'-AGCGAGTGGTATCAACGCGAG-3'</td>
</tr>
</tbody>
</table>

30 s at 94 °C, 40 s at 56 °C and 1 min at 72 °C and final extension of 10 min at 72 °C. The PCR product was separated on low-melting agarose gel and desired DNA fragment was recovered using GenElute™ gel extraction Kit (Sigma, USA). The eluted fragment was ligated into the pGEM-T® EASY vector (Promega, USA) and propagated in Escherichia coli DH5α cells. Recombinant plasmid was isolated (HiYield™ Plasmid Mini Kit, RBC technologies, Taiwan) and sequenced (BigDye terminator v 3.1 cycle sequencing mix, Applied Biosystems, USA) on an automated DNA sequencer (ABI Prism 310, Genetic Analyzer, Applied Biosystems, USA).

BLAST analysis indicated the sequenced fragment similarity to SqS genes known from other species. Genespecific primers (GSP) and nested GSP (NGSP) (Table 1) were designed based on the partial sequences of CbSqS for the cloning of 5' and 3' ends of the CbSqS gene by rapid amplification of cDNA ends (RACE) PCR. RACE ready cDNAs (5' and 3') were prepared from the total RNA using SMART™ RACE cDNA amplification Kit (Clontech, USA) following the manufacturer’s instructions. Primary and nested PCR was performed using primer pairs; GSP and UPM (universal primer mix; provided with the kit); NGSP and NUP (nested UPM) respectively. Both the primary and the nested PCR amplification procedures were carried out at following conditions: 3 min at 94 °C, 35 cycles (30 s at 94 °C, 35 s at 68 °C, 2 min at 72 °C) and 10 min at 72 °C. The amplified fragments were cloned in pGEM-T® easy vector and sequenced. By aligning 3' and 5' RACE product sequences, the full-length cDNA of CbSqS was obtained. To clone the open reading frame (ORF) of CbSqS, gene-specific primers CbSqS(ATG)F and CbSqS(ATG)R were designed (Table 1) covering the region from the start and stop codons of the gene.

Bioinformatics Analysis

Similarity search for sequences was conducted using the BLAST algorithm (http://www.ncbi.nlm.nih.gov/). The nucleotide sequence was translated using Translate tool.
Expression of CbSqS in E. coli

The CbSqS ORF was amplified and ligated into pQE-30 UA expression vector (Qiagen, Germany). The construct, pQE-30 UA-CbSqS, was transformed into the E. coli M15 strain harboring pREP4 plasmid. In order to amplify the carboxyl-terminal truncated CbSqS cDNA lacking 23 amino acid (389–411 aa) long transmembrane region, two oligonucleotide primers CbSqS(ATG)F and TruncCbSqS(ATG)R were used, and 1,161-bp amplified product was ligated into pQE-30 UA vector. E. coli M15 (pREP4) cells containing pQE-30-CbSqS and pQE-30-TruncCbSqS constructs were grown at 37 °C in Luria Broth medium (LB) containing 25 μg/ml kanamycin until the OD of 0.5–0.6 reached. The expression was induced by IPTG (isopropyl β-D-thiogalactopyranoside) to a final concentration of 1 mM. Induced culture was harvested after 2 h and centrifuged at 5,000×g for 5 min at 4 °C. The cells were resuspended in extraction buffer (50 mmol/l Tris–HCl, pH 8.0, 10 mmol/l MgCl₂, 20 % glycerol, and 5 mmol/l 2-mercaptoethanol) and disrupted by sonication. The lysate was centrifuged at 10,000×g for 20 min at 4 °C, and the supernatant was loaded on 10 % SDS-PAGE gel after denaturing with SDS loading dye at 99 °C for 10 min [27]. The gel was stained with coomassie brilliant blue dye.

Expression Analysis of CbSqS

Tissue specific expression of CbSqS was analyzed in leaf and root tissue by semi-quantitative RT-PCR. Total RNA was isolated from both the tissues and equal amount of DNA free RNA was reverse transcribed using Oligo dT(12-18) (Invitrogen, USA) and Superscript III reverse transcriptase (Invitrogen, USA). Primers for gene expression analysis were designed (Table 1), while 26S rRNA gene was used as an internal control [28]. PCR conditions were optimized for quantitation. PCR amplification was carried out at following conditions: 3 min at 94 °C, different number of cycles (30 s at 94 °C, 35 s at 60 °C, 1 min at 72 °C) and final extension for 10 min at 72 °C. PCR was run at different number of cycles (25, 28, 30, 32, 35) and cycle number yielding amplification within the exponential phase was chosen for expression analysis. The intensity of the PCR bands was measured with spot denso analysis tools of AlphaEaseFC software (Genelinnotech, USA).

Results and Discussion

Cloning and analysis CbSqS

Transcript-derived fragments (TDF) identified by cloning using degenerate primers showed similarity to SqS genes known from other species. A 380 bp TDF of CbSqS shared similarity at nucleotide (77, 78 %) and amino acid (83, 84 %) level with the SqS of Z. mays and Oryza sativa respectively. This sequence information was used to design primers for RACE PCR to amplify full-length cDNA sequence. Amplicons were generated, and sequence analysis of ten different clones obtained from both 5' and 3' RACE PCR revealed the full-length cDNA clone for CbSqS gene. After assembling the 5' and 3' regions, CbSqS contained 1,760 nucleotides, with an ORF of 1,233 nucleotides encoding a protein of 411 amino acid residues. The G+C content of the ORF was found to be 40 %. The sequence contained 126 bp long 5' untranslated region (UTR) and...
Fig. 2 a Nucleotide and the deduced amino acid sequence of CbSqS from C. borivilianum. The TGA stop codon is at position 1357 and the TGA stop codon at position 127 are in italics and regions before start codon and after stop codon represents 5’ UTR and 3’ UTR, respectively. In the 3’UTR, Poly A signal “AATTTA” is in bold and underlined and poly A tail starts from nucleotide 1734. From 5’ end toward 3’ end, six conserved catalytic domains (I–VI) identified using MultiAlin program is highlighted in gray, b Conserved domains of CbSqS protein as determined using NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/) indicating the presence of conserved isoprenoid domain and other substrate binding domains.
However, some exceptions are indeed found only in prokaryotes. The molecular weight of *Staphylococcus aureus* SqS is 56 kDa [29] while SqS from *Botryococcus braunii* is 7.3 kDa [30]. Conserved domain database (CDD) search at NCBI indicated the presence of conserved aspartate-rich regions, substrate binding pocket, substrate-Mg$^{2+}$ binding region suggesting that the protein may encode for functional enzyme (Fig. 2b). Multiple sequence alignment of CbSqS with SqS from other plants, yeast, humans and prokaryotic species (Supplementary Fig. 1) revealed six highly conserved signature domains (I–VI) in CbSqS of 14–23 amino acids (Fig. 2a). These domains are helpful in the catalytic activity of the enzyme. Out of all these domains, domains III, IV and V showed a highly conserved consensus sequence (DDYDEYCHGYVAGLVGLS, GLFLQKTNIIRDYLED and RFCAPQXMAITXLAXCYN respectively) with other SqS sequences while domains I and II were less conserved. Domain II was highly conserved with the aspartate-rich-motif with other SqS amino acid sequences [31]. These aspartate-rich-motifs are proposed binding site for the phosphate groups of the prenyl acceptor (allylic substrate) via a magnesium salt bridge. To further corroborate the importance of aspartate-rich motif, a study was performed by Gu et al. [32] The site-directed mutagenesis of rat hepatic SqS (RSS, EC 2.5.1.21) showed that the first Tyr residue (Tyr171, corresponding to Tyr173 of CbSqS) in the region III is essential for the activity of RSS and is likely involved in the first reaction, and mutation of Tyr171 to the mutants Y171F, Y171S, and Y171W completely abolished formation of PSPP or squalene from FPP [33]. This result also indicated that the phenyl ring is specifically required for activity and cannot be substituted with just either an aromatic or hydroxyl group [33]. The region IV also contains an aspartate-rich motif, 228-DYLED-232 in CbSqS, which overlaps with the 219-DYLED-232 sequence in human SQS. In the RSS, the two Asp residues Asp219 and Asp223 (equivalent to Asp224 and Asp234 of CbSqS, respectively) are essential; neutralization or reversal of charge (D219N, D223 N, and D223 K) cause inactivation [24]. The results were consistent with the participation of Asp219 and Asp224 in the binding of the diphosphates of two substrate FPP molecules via bridging Mg$^{2+}$ ions [34]. Region V is believed to be responsible for the catalysis of the second reaction F/VAIPQXMAIA/GTL (X = V/G/A) sequence is likely the NADPH binding motif [33–35]. The site-directed mutagenesis showed that the RSS mutant F288L (corresponding to Phe298 in the motif FVAIP of CbSqS) caused almost a complete loss of the second activity, but led to accumulation of significant amounts of PSPP even in the presence of NADPH [33]. The aspartate-rich motif (81DTVED85 in CbSqS, relevant to 80DTLED84 in human SQS, data not shown) in region II was also predicted to participate in the binding of the substrates [34]. Most membrane proteins have a hydrophobic region that span the hydrophobic core of the membrane bilayer. Amino acids within the C-terminal region formed domain VI. These were the most variable and hydrophobic compared with the other regions of the sequence. These sequences have a role in anchoring the enzyme to endoplasmic reticulum membrane [36].These transmembrane regions were predicted using TMHMM Server v.2.0 and SPLIT v.4.0 server (Supplementary Figs. 2, 3). The predicted secondary structure by SOPMA program indicated that CbSqS comprises of 68.05 % α-helixes, 24.15 % random coils, 4.88 % extended strands, and 2.93 % β-turns (Supplementary Fig. 4). These percentages are again in range with SqS from other plant species encoding for saponin biosynthesis. For instance, *Withania somnifera* secondary structure prediction showed that it contained 65.69 % α-helixes, 4.38 % β-turns joined by 8.03 % extended strands and 21.90 % random coils [27].

**Phylogenetic Analysis**

Phylogenetic relationships (Fig. 3) and pairwise identities (Fig. 4) among the SqS sequences from different organisms were established. SqS sequences from prokaryotes...
Fig. 4 Pairwise amino acid sequence comparison of CbSqS with SqS from other organisms. Values represent percent identity of sequences. The percent identity values are represented by pink-blue color gradient. Blue color represents high percent identity where as pink represents low percent identity of that sequence to other sequence (Color figure online)

3D Structure Analysis

I-TASSER simulation was used to determine the 3-D structure of CbSqS with human SqS (1EzfB) as a template for comparative modeling (Fig. 5a). The template shared 48 % identity with the CbSqS sequence. The predicted structure of CbSqS consists entirely of α-helices and the protein is folded as a single domain, with a large channel running through the center, surrounded by helices E, F, G, I, M, and N. CbSqS core structure was similar to that of the several class I isoprenoid biosynthetic enzymes, crystal structures of which are known and the conserved feature in all the structures was an α-helical core surrounding a central active site cavity regardless of amino acid sequence identity of which one end was predominantly hydrophobic, and the other end was more hydrophilic and contains a signature “aspartate-rich sequence” [34].

The substrate binding site was present at the core region of the enzyme structure (Fig. 5b). The predicted active site involved Phe 204, Leu 205, Gin 206, Thr 208, Asn 209, Ala 293, and Leu 297 (Fig. 5b). The conserved “aspartate—rich region” in the active site of the compound is consistent with other class I isoprenoid enzymes [36]. The aspartate side chains are involved in binding multiple Mg2+ ions that stabilize binding of diphosphate groups in the substrate. CbSqS 3D structure matches closely with SQS crystal structures from human (Fig. 5c). This can further help in studying the crystal structure of CbSqS. Evolutionary conservation of amino acids at the surface of CbSqS sequence was performed using Consurf program (Fig. 5d). Conservation analysis of positions among members from the same family can often reveal the importance of each position for the protein (or nucleic acid)’s structure or function. The degree to which an amino (or nucleic) acid position is evolutionarily conserved is strongly dependent on its structural and functional importance; rapidly evolving positions are variable while slowly evolving positions are conserved. Based on the Consurf conservation scores (data not shown), it was inferred that amino acids L58, Y70, L73, D82, Y168, D213, P222, R225, P229, and N303 were among the highly conserved amino acid residues in the deduced CbSqS amino acid sequence.
Expression of CbSqS in E. coli

The entire coding region of CbSqS cDNA and the truncated CbSqS cDNA that contains a deletion of the last 23 amino acids from the carboxy terminal were cloned into pQE30-UA vector (Qiagen, Germany), an expression vector with T5 promoter and His tag sequence. The pQE30-CbSqS and pQE30-TruncCbSqS gene constructs were transformed and expression was induced. Low expression levels of SQS polypeptide were observed in E. coli cells that contained the putative full-length CbSqS cDNA (data not shown). On the other hand, due to truncated C-terminus, measurable levels of SQS polypeptide were observed in the extracts of E. coli that expressed the truncated CbSqS (pQE30-TruncCbSqS). This resulted in the formation of a new polypeptide with an expected molecular mass of 43–45 kDa (including 0.84 kDa 6xHis Tag) on the SDS-PAGE gel. High levels of the truncated CbSqS were observed 1 h after the addition of 1 mM IPTG (Fig. 6). The presence of microsomal membrane anchoring structure at the carboxy-terminal of the protein might be the reason of low expression of the polypeptides in the E. coli cells transformed with pQE30-CbSqS. In accordance with this suggestion, Lee et al. [15] reported that the carboxyl-terminal deletion of residues 389–411 in the hot pepper SQS resulted in the accumulation of a functionally soluble SQS protein. Similar kind of result was observed in Withania somnifera [27].

CbSqS Shows Distinct Pattern of Expression in Leaf and Root Tissues

Chlorophytum borivilianum, a traditional medicinal plant was originally grown in thick forest in natural form [37]. It is a monocot and perenates in the soil by fleshy roots attached to a condensed stem disk which remains dormant in the soil for about 7 months. The fleshy roots are economically important, which contain saponins and are used for the preparation of many phyto-tonics. The stem is condensed to a disk from which is produced a whorl of leaves that are long, sessile and somewhat thick. Semi-quantitative expression of CbSqS was, therefore, analyzed
Expression of truncated CbSqS gene (pQE-30-TruncCbSqS) in E. coli M15 (pREP4). Total protein extracts of E. coli cells were used for detection by SDS-PAGE and stained with Coomassie blue. Lane M protein molecular weight marker. Lane 1 induction of recombinant protein after 2 h with 1 mM IPTG at 37 °C. Lane 2 un-induced control after 2 h. The arrow indicates the position and size of the recombinant CbSqS protein (≈ 45 kDa).

Fig. 7. Expression of CbSqS in the leaf (L) and the root (R) tissues of the plant as analyzed through semi-quantitative reverse transcriptase-polymerase chain reaction (RT-PCR). 26S rRNA was used as an internal control. Bar graph represents the comparative expression of CbSqS in the leaf and root tissues of C. borivilianum based on integrated density values (IDVs) measured by AlphaEaseFC software.

in the leaf and root tissues. Based on the IDV calculation, it was seen that the expression of CbSqS was 14.3 % higher in leaves as compared to the roots (Fig. 7). The result supports the expression of SqS from Withania somnifera which predicted that the level of expression of SqS was higher in leaves as compared to the stem and root tissues [27]. Kim et al. [15] conducted the in vitro hybridization studies on various isoforms of SqS in P. ginseng and reported the expression and accumulation of SqS near the conducting tissue of petioles. Jinting and Zhenghai [38] reported that saponins in Achyranthus bidentata were first synthesized in the leaves and then transported to the roots via phloem of viens and stem vascular bundles for storage. The expression pattern of CbSqS supports the fact that initial reactions of isoprenoid biosynthetic pathway occur in leaves, while later step modifications and storage occurs in roots [19].

Conclusion

Due to immense significance of C. borivilianum as a multipurpose medicinal plant, it is instigating to decipher the biosynthetic pathway regulating the saponin biosynthesis in this plant, and also to exploit the full biotechnological potential of this crop. Understanding of various enzymatic steps involved in the pathway is a prerequisite for metabolic engineering. These genes could be used for genetic engineering for beneficial purposes such as increasing the saponins contents. Using degenerate primers and RACE PCR strategy, we have cloned full-length cDNA encoding CbSqS. The gene was analyzed in details, expression analysis was performed in roots and shoots and recombinant protein was used in microbial expression system. SqS, being the first enzyme in the biosynthesis of isoprenoids toward sterols and triterpenes, holds an important key to the regulation of isoprenoid biosynthetic pathway. The cloning and characterization of CbSqS from C. borivilianum not only constitutes an important step unraveling the key committed step of the saponin biosynthetic pathway, but also throws a fresh prospect to understand the regulation of the isoprenoid pathway genes in this medicinal plant which represents a rich repository of bioactive metabolites.

Acknowledgments The Authors are thankful to Council of Scientific and Industrial Research (CSIR), India, for providing the financial assistance (Scheme No. 38(1188)08/EMR-II) to carry out the research work. Authors are also grateful to Dr. Rahul Shibhara Mandal, Scientist—I, NICED, Kolkata for his help in analysis of 3D structure prediction.

References
