CHAPTER: 3

ISOLATION, CLONING AND CHARACTERIZATION OF TERPENOIDS BIOSYNTHETIC PATHWAY GENE (S)
Chapter 3. Isolation, Cloning and Characterization of Terpenoids Biosynthetic pathway gene (s)

This chapter deals with different strategies used for isolation of full-length cDNA clones of terpenoid biosynthetic pathway genes, 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR), farnesyl diphosphate synthase (FPS) and squalene synthase (SQS) from \textit{B. monniera} and characterization of isolated genes.

3.1 Introduction

Plants synthesize various kinds of secondary metabolites like terpenoids, carotenoids, flavonoids, phytosterols, saponins, phytoestrogens, glucosinolates and much more. Plant secondary metabolites can boost the immune system, protect the body from free radicals, and kill pathogenic germs. In plants several important classes of terpenoid compounds are synthesized via the mevalonate pathway. The terpenoids sometimes referred to as isoprenoids represent the largest group of biologically active metabolites with at least 30000 chemical compounds. Plant isoprenoids are essential for the numerous physiological and developmental processes in plants and many plant isoprenoids are of considerable commercial interest as pharmaceuticals, flavors and fragrances or food colorants. Understanding the regulation of isoprenoid synthesis is therefore of immense scientific and commercial interest. In today’s world medicinal plants are very important, as they contribute to about 80% in global economy in the form of plant or plant extracts (Dhyani and Kala, 2005).

As discussed in chapter one, \textit{B. monniera} is an important drug in Ayurvedic medicine system for the improvement of intelligence, memory and revitalization of sensory organs. The active medicinal principle compounds of this plant are terpenoid saponins collectively known as bacosides and different types of alkaloids. Biosynthesis of these compounds takes place via isoprenoid pathway. Some important genes such as HMG-CoA reductase, FPP synthase, squalene synthase etc., play an important role in biosynthesis of secondary metabolites and considered as key steps. The details of these genes have been already discussed in chapter one.

The present chapter deals in detail regarding isolation, cloning and characterization of three key steps genes namely HMG-CoA reductase, FPP synthase and squalene synthase from \textit{B. monniera}. There is considerable interest in understanding role of these
genes in bacosides biosynthesis which will facilitate future metabolic pathway engineering for development of elite plant species with improved bacoside contents.

Isolation of a gene is the pre-requisite for its thorough study or characterization. To isolate any gene of interest generally there are two major approaches as follows:

1. Genomic/cDNA library screening- It is the classical method of gene isolation in which genomic/cDNA libraries are screened with homologous or heterologous probe. After 2-3 rounds of screenings, the plaques showing strong positive signals are excised and the DNA fragment sequenced.

2. PCR- It is one of the most popular approaches for gene isolation because of its simplicity and rapidity. Forward and reverse primers are designed on the basis of available sequences in the database. PCR is performed using these primers and genomic/cDNA as template. The amplicon is sequenced to confirm its identity.

In the present study, PCR based approach was followed to fish out the terpenoid biosynthetic pathway gene(s) from *B. monniera*.

3.2 Materials and methods

3.2.1 Bacterial strains and plasmids used in the study

*Escherichia coli* XL-10 Gold (Stratagene, USA)
pGEM-T Easy Vector Cloning vector (Promega, USA)

3.2.2 RNA isolation and cDNA first strand synthesis

Total RNA was isolated from *B. monniera* leaves and 1 µg of total RNA was used for first-strand cDNA synthesis using oligo (dT)15 primer and AMV-RT (as mentioned in Chapter 2). cDNA was stored at -20 °C and further used for PCR reactions.

3.2.3 Polymerase Chain Reaction (PCR)

PCR amplification was done using cDNA 1st strand as template. Amplified PCR products were eluted from agarose gel and used for ligation into T/A cloning vector pGEM-T Easy. The PCR program is as follows: 1 cycle of 94 °C for 5 min; 35 cycles of 94 °C 30 s, 50-58 °C for 30 s, 72 °C for 1 min 30 s and a final extension of 72 °C for 7 min.
3.2.4 Transformation and selection
The ligation mixture was used for transformation of *E. coli* XL10 GOLD, host cell strain. Putative transformants were selected using antibiotic ampicillin and Blue-white screening by adding X-gal and IPTG. Colony PCR of white colonies with primers T7 (5’-TAATACGACTCACTATAGGG-3’) and SP6 (5’-ATTAGGTGACACTATAGAA-3’) was done to screen the putative transformants. Plasmids were isolated from clones showing positive colony PCR results and subjected to restriction digestion. The plasmids with expected insert release were further sequenced to confirm its identity with reported sequences available at NCBI GenBank database.

3.2.5 Bioinformatic and phylogenetic tree analysis
The nucleotide sequences available in the NCBI GenBank database were aligned and multiple sets of primers were designed from the conserved regions. Nucleotide and amino acid sequence analysis was done using software pDRAW 32, ClustalX 2.0 and online bioinformatics analysis facility available at www.justbio.com, www.expasy.org and www.ncbi.nlm.nih.gov. Various physical and chemical parameters of proteins were predicted by using ExPasy tool ProtParam (http://web.expasy.org/protparam/). Multiple sequence alignments of the amino acid sequences were carried out with the ClustalW 2.0 program (http://www.ebi.ac.uk/clustalw/). The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analysis was conducted in MEGA4.0.2 (Tamura et al., 2007).
3.2.6 Rapid amplification of cDNA ends (RACE)

RACE was done to isolate the full-length genes with its 5' and 3' UTRs. The PCR conditions are as follows:

### Primary PCR

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>94 °C</td>
<td>3 min</td>
<td>1 cycle</td>
</tr>
<tr>
<td>94 °C</td>
<td>30 s</td>
<td>10 cycles</td>
</tr>
<tr>
<td>70 °C</td>
<td>*1 min 30 s</td>
<td></td>
</tr>
<tr>
<td>94 °C</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>60 °C</td>
<td>30 s</td>
<td>25 cycles</td>
</tr>
<tr>
<td>72 °C</td>
<td>*1 min 30 s</td>
<td></td>
</tr>
<tr>
<td>72 °C</td>
<td>10 min</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>

# Annealing temperature may vary according to T_m of primers. *1 min/kb DNA extension.

### Secondary (Nested) PCR

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>94 °C</td>
<td>5 min</td>
<td>1 cycle</td>
</tr>
<tr>
<td>94 °C</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>65 °C</td>
<td>30 s</td>
<td>35 cycles</td>
</tr>
<tr>
<td>72 °C</td>
<td>*1 min 30 s</td>
<td></td>
</tr>
<tr>
<td>72 °C</td>
<td>10 min</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>

*1 min/kb DNA extension.

3.2.7 Sequencing

DNA sequencing was performed with the ABI Prism Big-Dye Terminator Cycle Sequencing Kit on the ABI Prism 3730 DNA analyzer (Applied Biosystems) at GenomeBio Biotech Pvt. Ltd., Pune, Maharashtra, India.
3.3 Results and discussion

3.3.1 Total RNA isolation and cDNA synthesis
Total RNA was isolated from aerial part of B. monniera plant and cDNA synthesis was done using protocol as mentioned in materials and methods chapter.

![Figure 3.1](image)

**Fig. 3.1:** Total RNA isolated from aerial tissue (Lanes 1 & 2) of B. monniera, electrophoresed on 1% agarose gel.

3.4 PCR based approach for the isolation of terpenoid biosynthetic pathway gene(s)

3.4.1 HMG-CoA reductase: Its isolation, cloning and characterization

3.4.1.1 Multiple sequence alignment of HMG-CoA reductase nucleotide sequences from different plants

HMG-CoA reductase gene nucleotide sequences available at NCBI GenBank database were aligned using Clustal W program. Different set of primers were designed from conserved regions, as shown in figure in red and green colours (Fig. 3.2).

<table>
<thead>
<tr>
<th>Plant</th>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. miltiorrhiza</td>
<td>ATGACGACGAAAATGGAGAAAGAGGCGGCGGCCAAG</td>
<td>GCCTCGGACGCACTCCCCCTCCCTGGCTGACCCGCCGCTGGCTCTCTTCCTTCCTTCCTTCCTGCTGGCTGAGAAGAT</td>
</tr>
<tr>
<td>A. majus</td>
<td>ATGATGAGGATGAGGATGGAAGGGCAGCGGACGAGGAAGGGCGATTCTGATTCTGATTC</td>
<td>GGGGCTGATTATTAAAGCATCGGACGCCCTGCCTCTCTGGTGACCCGCCGCTGGCTGAGAAGAT</td>
</tr>
</tbody>
</table>

**HMG F1**

<table>
<thead>
<tr>
<th>Plant</th>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. miltiorrhiza</td>
<td>ATGACGACGAAAATGGAGAAAGAGGCGGCGGCCAAG</td>
<td>GCCTCGGACGCACTCCCCCTCCCTGGCTGACCCGCCGCTGGCTCTCTTCCTTCCTTCCTTCCTGCTGGCTGAGAAGAT</td>
</tr>
<tr>
<td>A. majus</td>
<td>ATGATGAGGATGAGGATGGAAGGGCAGCGGACGAGGAAGGGCGATTCTGATTCTGATTC</td>
<td>GGGGCTGATTATTAAAGCATCGGACGCCCTGCCTCTCTGGTGACCCGCCGCTGGCTGAGAAGAT</td>
</tr>
</tbody>
</table>

**HMG F2**

<table>
<thead>
<tr>
<th>Plant</th>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. miltiorrhiza</td>
<td>ATGACGACGAAAATGGAGAAAGAGGCGGCGGCCAAG</td>
<td>GCCTCGGACGCACTCCCCCTCCCTGGCTGACCCGCCGCTGGCTCTCTTCCTTCCTTCCTTCCTGCTGGCTGAGAAGAT</td>
</tr>
<tr>
<td>A. majus</td>
<td>ATGATGAGGATGAGGATGGAAGGGCAGCGGACGAGGAAGGGCGATTCTGATTCTGATTC</td>
<td>GGGGCTGATTATTAAAGCATCGGACGCCCTGCCTCTCTGGTGACCCGCCGCTGGCTGAGAAGAT</td>
</tr>
</tbody>
</table>
Chapter 3

S. miltiorrhiza
CCGCAATTCCGTCCTCCCCATCGGCATCTCGCACCGCCATCGGGCTTGGTCTGAGAT

A. majus
CGGAAATCCCAACCTTGCATCGGCATCTCGCACCGCCATCGGGCTTGGTCTGAGAT

* \* ** \* ** \* ** \* ** ** \* ** ** ** ** ** ** **

S. miltiorrhiza
CTCGCACTGCATCACCATGATGGAAGCCGTCAACGGCGGCAAGGACCTCCACGTGTCAGT

A. majus
TTCCATACGCCCGAGCGGCGCCGAGCGCCGCGCCCGGGCCGCGCCGCGGCCGCGCCGC

** \* ** \* ** ** \* ** ** ** ** ** ** **

S. miltiorrhiza
CAATATTGTCTCCGCCATCTTCATCGCTACGGGTCAGGATCCCGCACAAAATGTCGAGAG

A. majus
CGCGGTCCATTGGATCGAAGGGCGGGCAGTCCGTGGTGTGCGAGGCCGTCATCAAAGA

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAACATCGTCTCCGCCGTCTACATCGCCACGGGCCAGGACCCCGCACAGAACGTGGAGAG

A. majus
GGAGGTGGTGCACAAGGTGCTCAAAACTAACGTCGCCGCCCTCGTGGAGCTCAACATGCT

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
******** ** **.***** ** *****  * ** ******** ** ** *****

A. majus
TGAGGAGTTCACATTACAGGCGGCATACACTGGCGACTGTCGCTGAGTCCGAGTGATCC

** \* ** ** \* ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAAGAACCTCACTGGTTCAGCCATGGCTGGAGCTTTGGGTGGCTTCAATGCTCATGCCAG

A. majus
TAGGTTCGCCACCGCCAAGCGGGCTGCTGACTTGAAGTTCTTCCTCGAGGACCCTCTCAA

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
GGAGGGGGAAGATCCCTTCCTACGCCCTGGAGTCCAAGCTGGGAGACTGCCGCCGCGCCGC

A. majus
S.miltiorrhiza
-------TTACCCGCTGAGCCGCTCCCGGAGGAAGACGAGGAGATCGTCAAGGCCGTGGT

** \* ** ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CTTTGTCCGCCGAGGGCCCTCGACGACACACXCGGAGAAAATCCTCGGAGGCGGCTTCCCT

A. majus
TTCCATACGCCCGAGCGGCGCCGAGCGCCGCGCCCGGGCCGCGCCGCGGCCGCGCCGC

HMG F3

** \* ** ** ** ** ** ** ** ** ** ** **

A. majus
S.miltiorrhiza
CCGCAATTCCGTCCTCCCCATCGGCATCTCGCACCGCCATCGGGCTTGGTCTGAGAT

A. majus
CGGAAATCCCAACCTTGCATCGGCATCTCGCACCGCCATCGGGCTTGGTCTGAGAT

** \* ** \* ** ** \* ** ** ** ** ** ** **

S. miltiorrhiza
CTCGCACTGCATCACCATGATGGAAGCCGTCAACGGCGGCAAGGACCTCCACGTGTCAGT

A. majus
TTCCATACGCCCGAGCGGCGCCGAGCGCCGCGCCCGGGCCGCGCCGCGGCCGCGCCGC

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAATATTGTCTCCGCCATCTTCATCGCTACGGGTCAGGATCCCGCACAAAATGTCGAGAG

A. majus
CGCGGTCCATTGGATCGAAGGGCGGGCAGTCCGTGGTGTGCGAGGCCGTCATCAAAGA

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAACATCGTCTCCGCCGTCTACATCGCCACGGGCCAGGACCCCGCACAGAACGTGGAGAG

A. majus
** ** ** ** ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
******** ** **.***** ** *****  * ** ******** ** ** *****

A. majus
TGAGGAGTTCACATTACAGGCGGCATACACTGGCGACTGTCGCTGAGTCCGAGTGATCC

** \* ** ** \* ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAAGAACCTCACTGGTTCAGCCATGGCTGGAGCTTTGGGTGGCTTCAATGCTCATGCCAG

A. majus
TAGGTTCGCCACCGCCAAGCGGGCTGCTGACTTGAAGTTCTTCCTCGAGGACCCTCTCAA

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
GGAGGGGGAAGATCCCTTCCTACGCCCTGGAGTCCAAGCTGGGAGACTGCCGCCGCGCCGC

A. majus
S.miltiorrhiza
-------TTACCCGCTGAGCCGCTCCCGGAGGAAGACGAGGAGATCGTCAAGGCCGTGGT

** \* ** ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CTTTGTCCGCCGAGGGCCCTCGACGACACACXCGGAGAAAATCCTCGGAGGCGGCTTCCCT

A. majus
TTCCATACGCCCGAGCGGCGCCGAGCGCCGCGCCCGGGCCGCGCCGCGGCCGCGCCGC

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAATATTGTCTCCGCCATCTTCATCGCTACGGGTCAGGATCCCGCACAAAATGTCGAGAG

A. majus
CGCGGTCCATTGGATCGAAGGGCGGGCAGTCCGTGGTGTGCGAGGCCGTCATCAAAGA

** \* ** \* ** ** ** ** ** ** ** ** ** **

S. miltiorrhiza
CAACATCGTCTCCGCCGTCTACATCGCCACGGGCCAGGACCCCGCACAGAACGTGGAGAG

A. majus
** ** ** ** ** ** ** ** ** ** ** ** ** **
Fig. 3.2: Multiple sequence alignment of HMGR nucleotide sequences of *Salvia miltiorrhiza* (Acc. No. FJ747636) and *Antirrhinum majus* (Acc. No. EF666139) Highlighted regions are considered for primer synthesis (Red colored-Forward and Blue colored- Reverse primers).

Table 3.1: Primers used for isolation of partial HMGR cDNA

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Forward and Reverse primers (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMG CoA-Reductase (HMGR)</td>
<td></td>
</tr>
<tr>
<td>HMG F1</td>
<td>CTCTCTACCTCACCACGG(A/C)GT(C/G)TTC</td>
</tr>
<tr>
<td>HMG F2</td>
<td>CG(A/C)TGGCAGAGAAGATCCGAAAC</td>
</tr>
<tr>
<td>HMG F3</td>
<td>GGCGACGCCATGGCAGCATGAACATG</td>
</tr>
<tr>
<td>HMG R1</td>
<td>TCTGTGGTGGCCATCCGGGAC</td>
</tr>
<tr>
<td>HMG R2</td>
<td>GACCTATTGTACTTCATATGGCTATTG</td>
</tr>
</tbody>
</table>

3.4.1.2 PCR amplification of partial cDNA of *HMGR* gene from *B. monniera*

Two primer sets i.e. HMG F2-HMG R1 and HMG F3-HMG R2 (as mentioned in table 3.1) gave amplifications of ~600 bp fragments of *HMGR* gene using cDNA 1st strand as a template (Fig. 3.3 A & B). The fragments were cloned in pGEM-T Easy vector (Fig. 3.4) and confirmed by restriction analysis (Fig. 3.5A & B) and sequencing (Fig. 3.6A & B). These partial clones were named as HMG F2/R1 and HMG F3/R2 respectively.
**Fig. 3.3:** A. PCR amplification product of ~600bp (HMG F2/R1) fragment of HMGR gene separated on 1% agarose gel. Lane 1-3: Amplified product and lane M- Medium range DNA ruler. B. PCR Amplicon of ~600bp (HMG F3/R2) of HMGR gene on 1% agarose gel. Lane 1-3 amplified product and lane M- Low range DNA ruler.

**Fig. 3.4:** Map of pGEM-T Easy vector.
Fig. 3.5: A. Partial clone of HMGR (HMG F2/R1) gene in pGEM-T Easy vector releasing ~600 bp insert when digested with EcoRI, lane 1 & 2: cloned ~600bp insert.
B. Partial clone of HMGR (HMG F3/R2) gene digested with EcoRI, lane 1 & 2-
Digested product, lane M- Low range DNA ruler.
After confirmation by restriction digestion, 6-8 plasmids from both clones were sequenced. Sequencing result revealed that fragment HMGR F2/R1 and HMGR F3/R2 were 598 bp and 657 bp respectively (Fig. 3.6).

### A. HMGR F2/R1

<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CGATGCGCTG AGAAGATCAAG AAACCTCACC CCCCTCCACG TCGTCAACCT</td>
</tr>
<tr>
<td>51 TTCTGAAATGC GCCGCCATCG TCACCTTCGT TGCTCTCTTT ATTTATCTCA</td>
</tr>
<tr>
<td>101 TTGGTTTCTCT TGGCGTTTGA TTTGTCTCCTC CCCCTAACAT TCCCCGTGCT</td>
</tr>
<tr>
<td>151 TCCCCACGAG ATATCCCTTG GGAAGATGAC GAAATGTGTG ATGGAGATCT</td>
</tr>
<tr>
<td>201 AGATCCGCTG ATGCTGAAGG AGGATTCCCG CTCCTCTG</td>
</tr>
<tr>
<td>251 CGCTTCAACTC GTGGAGATCT CTCCTTCCAG ATCCATGGTC</td>
</tr>
<tr>
<td>301 CACTGAGTGGG GTAATCCGGT CGCGCCGCCG TACGCTTCTG</td>
</tr>
<tr>
<td>351 GTAATCCTCA ATATGGCCGT ATGTCTCCAAT CACGCTTCTG</td>
</tr>
<tr>
<td>401 CGCCTAATCA CTTGAGTGGG CTCCTCAGATC GCAGGAGTCC</td>
</tr>
<tr>
<td>451 AGATCCGCTT GGGCATCGCC GCCCTCTGTG AGGCAACAGA</td>
</tr>
</tbody>
</table>

### B. HMGR F3/R2

<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 GGGCAACGCCA TGCGCCATGAA CATGGTCCTCA AAGGGAGTAC AGAAGCTCAT</td>
</tr>
<tr>
<td>51 GAACATTTCAG ATATCTTTCGA AAGACCCTGC GGTGACCTG GATGGAGGCA</td>
</tr>
<tr>
<td>101 AGGAGGAAATGT CAGTTTGTGTT TGGAGCATA ATCTGAAGAT GTGAGTGTCA</td>
</tr>
<tr>
<td>151 GAAGACAAATCG CACCGGCTGG ATGCTGAGCTA GATGGACCTG AAAAAATAC</td>
</tr>
<tr>
<td>201 AGAGAGCTTCC AGGCCTGAGC ATGGAGGTCG TCTGCGCTG AGATCTTATG</td>
</tr>
<tr>
<td>251 GAAGACTGCCA CAGTGCCTCTT TGGCGCCCCG GCCCTCAAGC GACAGGACTG</td>
</tr>
<tr>
<td>301 CTCCTCCTCTT CGCTGGACGC ATGGCCGTCG TACGCTTCCAG</td>
</tr>
<tr>
<td>351 CGCAGAGAAG ATGAGAGAGC ATCCTGTCCTA CACATCGATG GAGCCACTCC</td>
</tr>
<tr>
<td>401 ATGCTTACCAT CAGTGTCTTC CAGGCTTACG TGCTTGGTCTG AACGACTTCA</td>
</tr>
<tr>
<td>451 GGGAGCTTCC TGGGAAGGTC GCAGGCTTCTC GACAGAGGCA GAGGGGAGCC</td>
</tr>
<tr>
<td>501 CTTGCTTGGA GTGGAGAGGA CAGGTGCTCTC GAGGCAATCC AAGGGAGGAG</td>
</tr>
</tbody>
</table>

Rishi Kishore Vishwakarma  Ph.D Thesis  University of Pune
Fig. 3.6: A & B. Nucleotide sequences of partial HMGR gene from *B. monniera*. Primers used for 5’ (HMG F2/R1 sequence) & 3’ (HMG F3/R2) RACE are underlined.

BLAST analysis of both the partial sequences showed significant similarity with reported *HMGR* nucleotide sequences at NCBI GenBank database.

### 3.4.1.3 Rapid amplification of cDNA ends (RACE)

RACE is an important tool to obtain the UTRs (Un-Translated Regions) of a particular gene and is also useful to obtain full-length gene. The limitation of RACE PCR is partial sequence of the gene must be known to design gene specific primers.

### Primers for 5’ and 3’ RACE

Based on the sequence of partial HMGR gene fragments, primers for 5’ RACE and 3’RACE were designed from HMGF2/R1 and HMG F3/R2 clone sequences respectively (Underlined in Fig. 3.6 A & B).

- **RACE HMG R**
  5’-GAGCGGGAATCCTCCTTCAGCATC-3’

- **RACE HMG NestedR**
  5’-CTAGATCTCCATCAACAATTTCGTC-3’

Primers supplied with kit:

- **5’GeneRacer primer (5’GRP)**
  5’-CGACTGGAGCACGAGGACACTGA-3’

- **5’GeneRacer nested primer (5’NGRP)**
  5’-GGACACTGACATGGACTGAAGGAAGTA-3’

- **RACE HMG F**
  5’-CTCACTGTATCAATGATGGAAGCC-3’

- **RACE HMG NestedF**
  5’-GGGAAGGATCTTCATATCTCTGTC-3’

Primers supplied with kit:

- **3’GeneRacer primer (3’GRP)**
  5’-GCTGTCAACGATACGCTACGTAACG-3’

- **3’GeneRacer nested primer (3’NGRP)**
  5’-CGCTACGTAACGGCATGACAGTG-3’

### 3.4.1.3.1 5’ RACE PCR

In order to get the 5’ region of *HMGR* gene, 5’ Rapid amplification of cDNA ends (RACE) PCR was performed as per manual instructions. Primary PCR was done with RACE HMG R & 5’GRP primers and 5’RACE ready cDNA as a template. The
primary PCR product was diluted (1:50) and used as a template for secondary PCR with primers RACE HMG NestedR and 5’ NGRP. Approximately 550 bp secondary PCR product (Fig. 3.7) was cloned in pGEM-T Easy vector and sequenced.

![550 bp secondary PCR product](image1)

**Fig. 3.7:** A. 1 % agarose gel showing 5’ RACE product of *HMGR* gene. Lane M- Medium range DNA ruler, lane 1-3- ~550 bp 5’ RACE product. B. Restriction digestion with *EcoRI*. Lane 1 & 2- Showing vector backbone (~3 kb) and release of insert (~550 bp), Lane M- Low range DNA ruler.

The analysis of the sequenced 5’ RACE product of *HMGR* gene revealed 132 bp 5’ UTR regions, which is shown highlighted in red and the start codon (ATG) is highlighted in green (Fig. 3.8).

![Sequence analysis of 5’ RACE PCR product of *HMGR* gene](image2)

**Fig. 3.8:** Sequence analysis of 5’ RACE PCR product of *HMGR* gene. Underlined- 5’ NGRP sequence; 5’ UTR (132 bp) highlighted with red colour; Start codon (ATG) highlighted with green and underlined.
3.4.1.3.2 3' RACE PCR

To isolate 3’ end of the HMGR gene from Bacopa, 3’ RACE PCR was performed. Primary PCR was done using 3’ RACE ready cDNA and primers RACE HMG F & 3’GRP. Diluted (1:50) primary PCR product was used as a template for secondary PCR with RACE HMG Nested F and 3’ NGRP primers. Approximately 600 bp amplicon (Fig. 3.9) was cloned and sequenced.

![Fig. 3.9: 1 % agarose gel showing 3’ RACE product of HMGR gene. Lane M- Medium range DNA ruler, lane 1-3- ~600 bp 3’ RACE product.](image)

Sequence analysis of 3’ RACE clone showed stop codon (TAA), 3’ UTR (224 bp), putative polyadenylation site and poly ‘A’ tail (Fig. 3.10).

```
1  GGGAAAGGATC TTCAATATCT TGTCAC ATG TCCTCAGGATG ACCTGGTGAT
51  AGTCGGAATG GAGGACCATG TGGGTCATCA AGTGCAATGT CTAACCTGTC
101  TGGAGTGGAA GGGAGCAGTA AAGATGTCCT CAGGTCGAAA CTCCGGGAGC
151  CGGACAGAAT ACCTGCTTTC AGCGGAGCAG TGTCGTGAGT
201  GTCGGCTTTTG GCCGCAGAGC AACTGTGATC CAACTACTGTA
251  GGTCCAGTAAA GGATGTGTCA AAGGTTTCGT CTCCA TAA GCT CTCTTTTCTCC
301  TTTATTTTTAT TTATTTTTAT TTTTGTGTGT TCCTATTTTTT TACTCTCCTT
351  AGGAAACCAG ATTCCGCAGG GCCGAAGAAG GAGGATAGGT AATAACTGTA
401  TTGAAACAAC AGTGTGTGTA CCAAGCCAGT GTAAATGTGT AAGGATGTAT
451  CGTCTATTTTA TTGATGTGTA ATTTATGCA ATAAAATACC
501  TTCATTATTTA TCAAATCTA AAAAAAAAAAA ACTGTCATAC CGTTACGATAG
551  CG
```

![Fig. 3.10: Sequence analysis of 3’ RACE PCR product of HMGR gene. Underlined-Black (RACE HMG Nestd F) and blue (3’ NGRP); Stop codon TAA (Red underlined); 3’ UTR (224 bp) highlighted with green colour; Putative polyadenylation site (AATAAAA) orange underlined.](image)
3.4.1.4 Amplification of full-length HMGR cDNA from *B. monniera*

On the basis of 5' and 3' RACE sequence information, primers were designed from start and stop codon to amplify the ORF of HMGR gene as a single PCR product.

HMG Full F 5' - ATGGAGGCTAAGGGACGATCCATG-3'
HMG Full R 5' - TTATTGAGACGAAACCTTGGAGACATC-3'

PCR was performed with primers HMG Full F/R and cDNA as a template with high fidelity Taq DNA polymerase. Approximately 1.7 kb amplicon was cloned in pGEM-T Easy vector and confirmed by restriction digestion and sequencing (Fig. 3.11 A & B).

![Fig. 3.11: A. 1% agarose gel showing 1.7 kb amplification (Lane 1-3), Lane M- Low range DNA ruler. B. Restriction digestion with EcoRI. Lane 1- Undigested plasmid, Lane 2 & 3- Showing vector backbone (~3 kb) and release of insert (~1.7 kb), Lane M- Medium range DNA ruler](image)

Sequence analysis of 6-8 clones showed the exact size of 1770 bp and two types of clones probably isoforms of the HMGR gene. Both isoforms were designated as *BmHMGR1* and *BmHMGR6* and submitted to NCBI GenBank data base having accession number HM222606 and HM222607 respectively (Fig. 3.12).

**Nucleotide Sequence (BmHMGR1; Accession no.:HM222606)**

```
ATGGAGGCTA AGGGACGATC CATGAAATTG AAAGGGCGGTG CCCTTGATCG
GAAGCTTCAA CGCAAATCCT CCGATTCCGA TCAGCCCAAG GCTTCCGACG
ACTTCCCCT TCCATTAGCC TTAGGTGGC TGTATTATCT CCTCCGCCGT TGGCGTGAGA
GCCATCGTCA CCTTCGTTGC CTCCTTTATT TATCTCATTG GTTTCTTTGG CGTTGGATTT
```

Rishi Kishore Vishwakarma  Ph.D Thesis  University of Pune  68
Deduced amino acid sequence

MEAKGRSMKLKGGALDRKLQRKSSDSDQPKASDALPLPLAFTNAVFITLFFSVVYYLLRR
WREKIRNSTPLHVTLSEIAAVTFFVSFPVFYLGFEVGFVQSLIIPIARASDDILGEDDE
IVGDRLQMLKEDRIPACPAALAEAEVESAKSFVDEHRAGLNEHLLRRI
VEKGKPAVESLPKCGLRRALPAREALQRRTTGKSLDGPLDNYSILQCCMEMPV
YQFQGAVGQPLLNGKYESYVMTETEGLAVASTNRCGAFAEAGGAGSLVILKDGMRAIP
VVRFGTAKRAALKEEFLDEPLNFEPLSFLNFSSRFRLQSSIKCNVAGSNLNYRFCTGTG
DDSSMVNSVRKVQNMVDFKLQKEFFPDMDVIGISGNVCSDKPDFPAVWGTEGKTGKSVCEAI
KEDVVKVVKLTVDASLVELMLNKNTSGAMAGALGGFNAHASNIVSAYIATQDPAQNV
ESSHCITMMEAVNDKDLISHVTPMSEEGTVGGQGLASQSAANLNNLVGLKGASKDAPGS
NSQCLATIVAGAVLAGELSALAGQLVKSHMKNYRSSKDVSVQVSAQ

Nucleotide Sequence (BmHMGR6; Accession no.:HM222607)

1 ATGGAGGCTA AGGGACGATC CATGAAATTG AAAGGCGGTG CCCTTGATCG
51 GAGCGCTCAA CGCAAATCCT CCGGATTCCGA TCAGCCCAAG GCTTCCGACG
101 TTTAAGAGTGGG TTTATACAGA GCTGTTATTC CAGGACCTTG CTG
151 TTTCTCGCTG ATGTTCTTCA CTGTTCTCTG CAGGACCTTG AGG
201 TTTAGCCGCTT TTTGCATTGG AATGTTCTGG AGTTCTCTCT CAC
251 TTTGACGGCC TAATCCTTTT CCTGTTCTTG AATGAGGCTA AGG
301 GTTCAATCCC TAATCATCCC CCGGTCTTCC CACGACGATA TCCTTGAGGA
351 AGATGACGAA ATTGTGATGAG TGAGTCGCGA GCTGCGAAGA ATGGACAGA
401 ATTCGCCGCTC GCCGCTTCCG GCCGCGCTCG CTGCATCTGC GTGTTCTTCC
451 AGCGTGCTTG ATCCAGGCTC AGAGGCTGTC TATGGTTTTT CAC
501 GCCGATTCGCA AGATGCTCGA GGTACGGATG CATGTTCTTCC CACGACGATA
551 GGTGACGAA ATTGTGATTG GAGATCTAGA TCGCCAGATG CTGAAGGAGG
601 ATTCCCGCTC CGCCCCTTGC GCCGCCGCGC TTCCATCCGT GGATTCTGGC
651 TGACGCGGTAC ATCCAGCTTG GATGTTCTTG AATGAGGCTA AGG
701 GTTCGAGGAT GGGGTGCGAA CTGGTCCGAGCT ATGGGATCTGG CACGACGATA
751 CCTCTGTGTT GAAGGCGGAA GGAATCTCTT TGGCCAGATG CACGACGATA
801 GGCTCCGGGG GTGCAGATGC TGGAGGCTGA GTGTTCTTCC CACGACGATA
851 CCGGCGGGCC TCCACGCTGC ATACCTAAGT CAAAGATTG CACGACGATA
901 GTTCGTAAGAT TGTTGTAAGA GAAAGCTGCC GCCGAGCTCA AATTCTTCCT
951 GAGAAGCCCT CTAATTTGGC AACCCCTTCC TCTATTTTCA AAGGAGGCTA
1001 GCCCGATCCG AAGGAGGCTA GCCGCGCGCA CGG
1051 CTGTACATGA GATTGTTGAC AATGAGGCTA AGG
1101 GGTCTCAAGA GATGTTTCTT TTTTCTCTCA AAGGAGGCTA
1151 CCCGACGCTG AGGAGGCTA GCCGCGCGCA CGG
1201 AGCCGATTAG GGTACGGATG CATGTTCTTCC CACGACGATA
1251 CTGCTACATC CTGCTACATC CTGCTACATC CTGCTACATC
1301 CCTGCACAGA CTCGACGACG CTCGACGACG CTCGACGACG
1351 GCCGACGCTG AGGAGGCTA GCCGCGCGCA CGG
1401 GTGCAAGCTG AGGAGGCTA GCCGCGCGCA CGG
1451 ACCGACGCTG AGGAGGCTA GCCGCGCGCA CGG
1501 ATGGGATCTGG CAGGACCGAG CTGCTACATC CTGCTACATC
1551 ATGGGATCTGG CAGGACCGAG CTGCTACATC CTGCTACATC
1601 ATGGGATCTGG CAGGACCGAG CTGCTACATC CTGCTACATC
1651 ATGGGATCTGG CAGGACCGAG CTGCTACATC CTGCTACATC
1701 ACAACTGCTG AGAAGCCCA TAATGCTCTG TAGATCTTG AGG
1751 CCAAGGGTGC TCTCATTAAC

Chapter 3
Chapter 3

70

Nucleotide and deduced amino acid sequences of BmHMGR1 and BmHMGR6.

3.4.1.5 Characterization of BmHMGR1 and BmHMGR6

The full-length cDNA sequences having ORF of 1770 bp (including stop codon), encode polypeptide of 589 amino acid residues. The presence of internal EcoRI restriction site was observed in BmHMGR6 (Fig. 3.12, Blue underlined). Deduced amino acid sequences show identity and similarity of 98.1% and 99% respectively with each other. Differences are highlighted with green colour in alignment (Fig. 3.13).
Fig. 3.13: Deduced amino acid sequence alignment of BmHMGR1 and BmHMGR6. Differences are highlighted with green.

3.4.1.5.1 Multiple sequence alignment of BmHMGRs with other plant HMGRs and CDD search

Multiple sequence alignment showed substrate binding motifs (highlighted in green), NAD (P) H binding motifs (Red highlighted) and different catalytic residues (Yellow highlighted). Conserved domain database search on NCBI server showed specific hits with HMG-CoA_reductase_class1 and super family of HMG-CoA_reductase (Fig. 3.14 A & B).
Fig. 3.14: A. ClustalW alignment of BmHMGRs with other plant HMGR amino acid sequences. Different motifs and catalytic residues are highlighted with different colours. B. CDD search on NCBI server.
3.4.1.5.2 Amino acid sequence analysis

The deduced amino acid sequence of BmHMGR shared 86.8%, 85.0%, 83.7% and 82.4% similarity and identities of 77.8%, 76.6%, 72.8% and 71.7% with *Salvia miltiorrhiza*, *A. majus*, *N. tabaccum* and *C. annuum* respectively. The calculated molecular mass and predicted pI value of BmHMGRs were 62.63 kDa and 8.33 respectively. The instability index was computed to be 37.11, indicates that protein is stable and atomic formula of BmHMGR was C_{2749}H_{4462}N_{758}O_{843}S_{31}. The total number of positively (Arg+Lys) and negatively (Asp+Glu) charged amino acids were 64 and 60 respectively.

3.4.1.5.3 Hydropathy plot and transmembrane domain analysis

The hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain (Kyte and Doolittle, 1982). Each amino acid is given a hydrophobicity score between -4.5 and 4.5. A score of 4.5 is the most hydrophobic and a score of -4.5 is the most hydrophilic. The amino acid sequences of BmHMGRs were analyzed using Kyte-Doolittle Hydropathy plot of JustBio hosted tool (http://www.justbio.com/index.php?page=plots) at window size 9. Strong negative peaks indicate possible surface regions of globular proteins and score greater than 1.8 indicate possible transmembrane regions. In BmHMGRs two transmembrane regions at N-terminal and one at C-terminal were observed (Fig. 3.15 A). Here, only one isoform (BmHMGR6) was taken an account because both sequences are 99% similar.

Presence of transmembrane domains was also predicted on TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/). Three domains (I- 39-58 amino acid; II-79-101 amino acid & III- 545-566 amino acid residues) were found (Fig. 3.15 B).
Fig. 3.15: A. Kyte-Doolittle Hydropathy plot B. Transmembrane domains (Domain I 39-58 AA; Domain II 79-101 AA; Domain III-545-566 AA).
3.4.1.5.4 Phylogenetic tree analysis

A phylogenetic tree was constructed by using known HMGR sequences from broad range of organisms including dicots, monocots, gymnosperms and bacteria. The dicot HMGRs were clustered into one group, in which BmHMGR was closely related to the *S. miltiorrhiza* and *A. majus* (Fig. 3.16).

**Fig. 3.16:** Phylogenetic Relationship of BmHMGR with other HMGRs. The evolutionary tree was formed using the ClustalX2 program and MEGA 4.0.2 software with neighbor-joining method. Accessions Nos. are given in brackets.
3.4.2 Farnesyl pyrophosphate (FPP) synthase: Its isolation, cloning and characterization

3.4.2.1 Multiple sequence alignment for designing of primers

Nucleotide sequences of FPP synthase available at NCBI GenBank database were aligned using Clustal W program. Different set of primers were designed with red and blue respectively (Fig. 3.17).

**L. albus**

ATGG---CAGATCT---------------------AAGGTCAACTTTCTTGAATGTGTAT 36

**G. lutea**

ATGG---CAAATCTGAACGGAACTACATCCGATCTGAGAACTACATTCTTGCAAGTCTAT 57

**A. thaliana**

ATGGAGACCGGATCT---------------------CAAGTCAACCTTTCTCAACGTTTAT 39

**L. albus**

TCGGTTCTGAAATCAGAGCTCCTTCATGACCCAGCTTTTGAATTCTCTCCTGATTCTCGT 96

**G. lutea**

TCCGTACTCAAATCTGAGCTTTTGAATGACCTTGCTTTCGAGTGGTCTCCTGATTCTCGT 117

**A. thaliana**

TCTGTTCTCAAGTCTGACCTTCTTCATGACCCTTCCTTCGAATTCACCAATGAATCTCGT 99

**L. albus**

FPS F1

CAATGGCTCGACCGGATGCTGGACTACAATGTGCCTGGAGGAAAGCTAAACCGTGGACTG 156

**G. lutea**

CAATGGGTCGAACGGATGCTGGATTACAATGTTCCTGGAGGAAAGCTCAACAGAGGCCTT 177

**A. thaliana**

CTCTGGGTTGATCGGATGCTGGACTACAATGTACGTGGAGGGAAACTCAATCGGGGTCTC 159

**L. albus**

TCAGTTATTGACAGCTACAGATTGTTGAAAGATGGACATGAATTAAACGATGATGAAATT 216

**G. lutea**

TCTGTCATTGATAGCTATGAACAGCTAAAAGAAGGAAAGGAGTTAACAAAGGATGAGGTC 237

**A. thaliana**

TCTGTTGTTGACAGTTTCAAACTTTTGAAGCAAGGCAATGATTTGACTGAGCAAGAGGTC 219

**L. albus**

FPS F2

GATGACATTATGGATAACTCCCACACCGCGTGGTCAGCCATGTTGGTTCAGAGTACCC 336

**G. lutea**

GATGACATTATGGATGGCTCTCACACTAGGCGTAGCCAACCTTGTTGGTTCAGATTGCCA 357

**A. thaliana**

GATGACATTATGGATAACTCTGTCACCCGCCGTGGTCAACCTTGCTGGTTCAGAGTTCCT 339

**L. albus**

AAGGTTGGAATGATTGCAGCAAATGATGGGGTGCTGCTACGGAACCATATTCCTCGTATC 396

**G. lutea**

AAGGTTGGAATGATAGCTGTAAATGACGGTGTTCTCCTTCGCAACCACATACCAAGATCC 417

**A. thaliana**

CAGGTTGGTATGGTTGCCATCAATGATGGGATTCTACTTCGCAATCACATCCACAGGAT 399

**L. albus**

CTTAAGAAACACTTCAGGGGAAAACCTTATTATGCTGATCTTCTTGATCTGTTTAATGAG 456

**G. lutea**

CTCAAGAAGCCTTAGAGAAAAAGCTTATTATGCTGATCTTCTTGATCTGTTTAATGAG 477

**A. thaliana**

CTCAAAAAGCANTTCTCGTAGAAGCTCTACTATGAGATCTTCTTCTTGAGCTTCACACC 459

**L. albus**

AACCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 516

**G. lutea**

GATTATTTGGATTGCTTTGGTGCTCCTGAAACAATTGGAAAGATAGGTACAGATATTGAA 537

**A. thaliana**

AAPCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 519

**L. albus**

AACCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 537

**G. lutea**

AAPCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 519

**A. thaliana**

AAPCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 519

**L. albus**

TATTACTCATTTTACCTCCGACATGCAGTCTTACCTTATGAGATCTTCTTCTTGAGCTTC 636

**G. lutea**

TATTACTCATTTTACCTCCGACATGCAGTCTTACCTTATGAGATCTTCTTCTTGAGCTTC 657

**A. thaliana**

TATTACTCATTTTACCTCCGACATGCAGTCTTACCTTATGAGATCTTCTTCTTGAGCTTC 639

**L. albus**

AACCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 696

**G. lutea**

AACCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 699

**A. thaliana**

AACCATATTGACGTGAAAAACATTCTTGTTGATATGGGAACGTACTTTCAAGTACAGGAT 717

**L. albus**

GATTATTTGGATTGCTTTGGTGCTCCTGAAACAATTGGAAAGATAGGTACAGATATTGAA 756

**G. lutea**

GATTATTTGGATTGCTTTGGTGCTCCTGAAACAATTGGAAAGATAGGTACAGATATTGAA 777

**A. thaliana**

GATTATTTGGATTGCTTTGGTGCTCCTGAAACAATTGGAAAGATAGGTACAGATATTGAA 777
Fig. 3.17: Multiple sequence alignment of FPS nucleotide sequences of *Lupinus albus* (Acc. No. LAU15777), *Gentiana lutea* (AB017371) and *Arabidopsis thaliana* (Acc. No. X75789). Forward (Red) and reverse (Blue) primer sequences are highlighted.

Table 3.2: List of primers used for isolation of BmFPS partial gene.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Forward and Reverse primers (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farnesyl pyrophosphate</td>
<td></td>
</tr>
<tr>
<td>synthase (FPS)</td>
<td></td>
</tr>
<tr>
<td>FPSBM F1</td>
<td>GACTACAATGTACG(T/C)GGAG</td>
</tr>
<tr>
<td>FPSBM F2</td>
<td>CTCGGTTGGTGCATTGAATGG</td>
</tr>
<tr>
<td>FPSBM F3</td>
<td>CTTGTGCTTGATGACATTATG</td>
</tr>
<tr>
<td>FPSBM R1</td>
<td>CAATCCAGATAATCATCCTGACTTG</td>
</tr>
<tr>
<td>FPSBM R2</td>
<td>ACTCCATGAACACTCCTCCTAAG</td>
</tr>
<tr>
<td>FPSBM R3</td>
<td>GCCTCTTGTAGATCTTGCAAC</td>
</tr>
</tbody>
</table>

3.4.2.2 Amplification of partial cDNA fragment of *FPS* gene from *B. monniera* by PCR

Approximately 900 bp amplicon was obtained in PCR with primer combination (FPSBM F1 and FPSBM R3) and cDNA as a template (Fig. 3.18 A). This amplicon was cloned in pGEM-T Easy vector and insert release was confirmed by restriction digestion with *EcoRI* (Fig. 3.18 B). Sequence analysis and restriction digestion with *EcoRI* showed presence of *EcoRI* restriction site inside the gene (Fig. 3.19). BLAST analysis showed more than 80% homology with reported FPS sequences at NCBI. Blast
search of deduced amino acid sequence predicted the presence of putative conserved domains specific to Isoprenoid_Biosyn_C1 super family. This confirmed that isolated partial sequence was FPP synthase from *B. monniera*.

**Fig. 3.18:** A. 1% agarose gel showing amplification of ~900 bp (Lane1-3), Lane M- Low range DNA ruler. B. Restriction digestion with *Eco*RI. Lane 1- Showing vector backbone (~3 kb) and release of inserts (~600 bp and 300 bp), Lane M- Medium range DNA ruler

**Partial BmFPS nucleotide sequence**

```
1 GACTACAATG TGCCGCGAGG GAAGTTAAAC CGAGTTCTGT CAGTAATTGA
51 TAGCTATAAG TTACTAAGAG AAGGAAAGGA CCTTACTGAA GAAGAAGTGT
101 TTCTTGGATAG GGCTCTGTG TGGTGTATCG AATGGC TTCA GGCATATTTT
151 CTTGCCCTTG TACGATATAT GGTAACTCT CACACTCGAC GATGGCAATT
201 GAGATTGCAC TATAGCTGAA TATGTAATGT GTGATGAAAG TGTGAAGGAT
251 TACATGTGCT CACTACTATT TTACCTTCCA GACAGATGAT GACTACCTAG
301 AGGTGTCCTG CTGTCCTAGT CTGGCAGAGT TAGGATGGTC GATCTACGAG
351 AAATGCGCTA ATCAGTGGAAG AGATTGGAAA GATTGGAACT GATATTGAAG
401 TAGAATGATC TATGCTGATG TACAGGATG ACTACCTAGA TTACCTGTCG
451 TCATCGCTCG CGACGACGATG CCTGACGCG GCCTGACGCG ACTGACGACG
501 GAGGGCGAAT CTGTACGCGT GTACGCGTAC AGACGATGAT GACTACCTAG
551 CAAGCTATCG CAGTACGATG TACAGGATG ACTACCTAGA TTACCTGTCG
601 GAGGTGTGAG AATGCGCTA ATCAGTGGAAG AGATTGGAAA GATTGGAACT
651 TACATGACGT CGTTGGCAA CAGTACGATG TACAGGATG ACTACCTAGA TTACCTGTCG
701 AAATGCGCTA ATCAGTGGAAG AGATTGGAAA GATTGGAACT GATATTGAAG
751 TAGAATGATC TATGCTGATG TACAGGATG ACTACCTAGA TTACCTGTCG
801 GAGGTGTGAG AATGCGCTA ATCAGTGGAAG AGATTGGAAA GATTGGAACT
851 TACATGACGT CGTTGGCAA CAGTACGATG TACAGGATG ACTACCTAGA TTACCTGTCG
901 AGA
```

**Fig. 3.19:** Nucleotide sequence of partial BmFPS. Presence of *Eco*RI site is highlighted. Primers used for 5’ RACE (Red, Reverse and Nested reverse) and 3’ RACE (Blue, Forward and Nested forward) are underlined.
3.4.2.3 RACE PCR (5’ and 3’) of BmFPS

To obtain the 5’ and 3’ end of the gene two steps RACE PCR was done. Gene Specific Primers were designed from above partial sequence of BmFPS (Fig. 3.19, Underlined).

3.4.2.3.1 5’ RACE

Primers
RACE FPSBM R 5’- GATCATACCAACCTTGGGGACCCT-3’ (For primary PCR)
RACE FPSBM NR 5’- GCCATTCGATACACCAGCCAGG-3’ (For Nested PCR)

The primary PCR product with RACEFPSBM R/ 5’GRP was diluted (1:50) and used as a template for secondary PCR. Nested PCR with primers RACEFPS NR and 5’ NGRP yielded approximately 350 bp fragment (Fig. 3.20 A). It was cloned in pGEM-T Easy vector, confirmed by restriction digestion (Fig. 3.20 B) and sequenced.

Fig. 3.20: A. 1.5% agarose gel showing ~ 350 bp amplification (Lane1-3), Lane M-100bp DNA ruler. B. Restriction digestion with EcoRI. Lane 1and 2 Showing vector backbone (~3 kb) and of inserts (~ 350 bp), Lane M- Low range DNA ruler.

Sequence analysis of 5’RACE clone revealed start codon (green), 9 bp 5’UTR (Red) and 5’ NGRP sequens is underlined (Fig. 3.21).

1 GGACACTGAC ATGGACTGAA GGAGTAGAAA AAGAGATGCC GAATCTATAAT
51 GGACAGACGA CGGATCTGAG GGAAGGATTT CTAGGTGCTCT ACTCTGTCCCT
101 TAAATCTGAG CTGTTGGAAG ACCAGCTTT TGAATGAGCC GACGACTCTC
151 GCGAATGGGT CGAGCGGATG CTGGACTACA ATGTGCCTGG AGGGAAGTTA
201 AAACGAGGTC TGTCAGCTAT TGAATGCTAT AAGTTACTAA AAGAAGGAA
251 GGACCTTACT GAAGAAGAAG TGTTTCTTGC TAGTGCTCTC GGCTGGTGTA
301 TCAGATGCC

Fig. 3.21: Nucleotide sequence of 5’ RACE clone. Start codon (ATG, underlined) in green and 5’ UTR in red. 5’ NGRP black underlined.
3.4.2.3.2 3’ RACE

**Primers**

RACE FPSBM F 5’- GTGCAAGCCCTGGGAACACTGCAA-3’

RACE FPSBM NF 5’- ATCCTGCGACGTGCTGCAAGATC-3’

To isolate 3’ end and UTR of the gene two steps RACE PCR was done. Nested PCR with RACEFPSBM NF/ 3’NGRP yielded an amplicon of ~400 bp (Fig. 3.22 A) which was cloned in pGEM-T Easy vector, confirmed by restriction digestion (Fig.3. 22 B) and sequencing.

![Fig. 3.22: A. 1% agarose gel showing amplification of ~ 400 bp 3’ RACE PCR product. B. Restriction digestion with EcoRI. Lane 1- Showing vector backbone (~3 kb) and release of insert (~400 bp), Lane M- Low range DNA ruler](image)

Sequencing of 3’RACE clone showed stop codon (TAA, red) 3’ UTR (204 bp), putative polyadenylation site (Orange) and poly A tail (Fig. 3.23).

![Fig. 3.23: Nucleotide sequence of 3’ RACE clone. Stop codon (TAA) in red, 3’UTR in green and polyadenylation site in orange underlined. Primer 3’ NGRP is underlined.](image)
3.4.2.4 Amplification of full-length cDNA of FPS

Primers were designed from start and stop codon to amplify full-length ORF of BmFPS as a single PCR product on the basis of 5’ and 3’ RACE sequences.

FPS Full F 5’- ATGGCGAATCATAATGGACAGAC-3’
FPS Full R 5’- TTACTTCTGCTCTTTGTATATCTTACC-3’

PCR was performed with above full-length primers and cDNA as a template using high fidelity Taq DNA polymerase. Approximately 1.1 kb amplicon (Fig. 3.24 A) was cloned in pGEM-T Easy vector and confirmed by restriction digestion (Fig. 3.24 B) and sequencing.

![Image](image1)

**Fig. 3.24:** A. 1% agarose gel showing ~ 1.1 kb amplification band (Lane 1-3), Lane M-Medium range DNA ruler B. Gel showing restriction digestion with EcoRI. Lane 1&2-Showing vector backbone (~3 kb) and release of inserts (~650 bp and 450 bp), Lane M-Low range DNA ruler.

Sequence analysis of 6-8 clones showed the exact size of 1050 bp (Fig. 3.25) and full-length ORF with 5’ and 3’ UTRs was submitted to NCBI GenBank data base having accession number GU385740.

**Nucleotide sequence of ORF (BmFPS)**

```
1 ATGGCGAATC ATAATGGACA GACGACGGAT CTGAGGGAGG CATTTCTAGG
51 TGTCTACTCT GTCCCTAAAT CTGAGCTGTT GAACGACCCA GCTTTTGAAT
101 GGACCGACGC ATCTCGCGAA TGGGTCGAGC GGATGCTGGA CTACAATGTG
151 CCTGGAGGGA AGTTAAACCG AGGTCTGTCA GTAATTGATA GCTATAAGTT
201 ACTAAAGAA GGAAGGACC TTAATGAAGA ATAGCTGTGA GGTCAACCTT GTTGGTTTAG
251 GGTCCCCAAG GGTGGTATGA TCGCGGTCGA TGATGGAATC ATCCTTCGGA
```
Deduced amino acid sequence

MANHNGQTTDLREAFLGVYSVLKSELLNDPAFEWTDASREWVERMLDYNVP GGKLNRGLS VIDSYKLLKEKGDLTEEEFLALGMWCEWLQAYFLVLDDIMDNSHRTRGQPCRW FPK VGMIAVNGDIIHPRILKHKHFRNKPYYVDLDFNEVEFQTASQGMIDLLTIEGK DLKSYSLPLHRIRYQKTYASSYPLIPACCALLLMMAGENLNDHNVKVLIDMGI FYQVDD YLDGFGEPEKIKGIGTDIEDFKSCWLVKAEHCNEEQKIKLFEHGKDPAVAKIKAL YNEINLQISFTDYESKYEKLNNSSIESHPSKSVQAVLKSFLGKIYQKQ

Fig. 3.25: Nucleotide and deduced amino acid sequence of full-length BmFPS

3.4.2.5 Characterization of BmFPS

The full-length cDNA sequence having ORF of 1050 bp (including stop codon), encoding a polypeptide of 349 amino acid residues.

3.4.2.5.1 Protein sequence analysis and its identity with other plant FPS

The calculated molecular mass and predicted pl value of BmFPS were 40.26 kDa and 5.57 respectively. The total number of positively (Arg+Lys) and negatively (Asp+Glu) charged amino acids were 42 and 51 respectively. The instability index was computed to be 36.06, indicates that protein is stable and atomic formula of BmFPS was C_{1825}H_{2824}N_{470}O_{531}S_{13}. The aliphatic index is 93.27 which indicate thermo stability of proteins. The deduced amino acid sequence of BmFPS shared 95.7, 95.1, 92.3, 90.5 & 90.3% similarity and 90.8, 90.8, 85.1, 82.2 & 81.7% identities with M. piperita, S. miltiorrhiza, C. roseus, H. brasiliensis and P. ginseng respectively.
3.4.2.5.2 Multiple sequence alignment of BmFPS with other plant FPS and CDD search

Multiple sequence alignment of BmFPS with other plant FPSs showed substrate binding domains (red colored), Mg$^{2+}$ binding sites (Green colored) and active site lid residues (Indigo colored, underlined) (Fig. 3.26 A). Active site lid residues shield highly reactive carbocationic intermediates from solvents. The importance of domain I and II (Red colored Fig. 3.26 A) in catalysis has been demonstrated by mutagenesis studies. Mutation of aspartate and arginine in region I and, to a lesser extent, of the first aspartate and lysine in region II, reduced the FPP synthase activity (Joly and Edwards, 1993; Marrero et al., 1992; Song and Poulter, 1994). NCBI conserved domain database (CDD) search showed specific hits with Trans_IPPS_HT and super family of Isoprenoid_Biosyn_C1. Different domains and catalytic motifs are also observed (Fig. 3.26 B).
Fig. 3.26: A. ClustalW alignment of BmFPS with other plant FPSs showing different domains and catalytic motifs in different color. B. CDD search results on NCBI server.
3.4.2.5.3 Hydropathy plot analysis of BmFPS
The amino acid sequence of BmFPS was analyzed using Kyte-Doolittle Hydropathy plot of JustBio hosted tool (http://www.justbio.com/index.php?page=plots) at window size 9. Plot analysis revealed even distribution of hydrophilic and hydrophobic residues (Fig. 3.27).

![Kyte & Doolittle Hydropathy Plot](www.justbio.com)

**Fig. 3.27:** Kyte-Doolittle Hydropathy plot

3.4.2.5.4 Phylogenetic tree analysis
A phylogenetic tree was constructed by using known FPS sequences from broad range of organisms including dicots, monocots, gymnosperms, yeasts and bacteria. The dicot FPSs were clustered in to one group, in which BmFPS was grouped with *S. miltiorrhiza* and *M. peperita* and *C. roseus* (Fig. 3.28).
Fig. 3.28: Phylogenetic tree analysis of BmFPS with other FPS amino acid sequences. The tree was formed using MEGA 4.0.2 software with neighbor-joining method. Protein accession nos. is given in brackets.
3.4.3 Squalene synthase (SQS): Its isolation, cloning and characterization from *B. monniera*

3.4.3.1 Multiple sequence alignment for designing of primers

Nucleotide sequences of squalene synthase available at NCBI GenBank database were aligned using Clustal W program. Different set of primers were designed from conserved regions. Forward and reverse primers are highlighted with red and blue respectively (Fig. 3.29).

<table>
<thead>
<tr>
<th></th>
<th>60</th>
<th>120</th>
<th>180</th>
<th>240</th>
<th>300</th>
<th>360</th>
<th>420</th>
<th>480</th>
<th>540</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. japonicus</td>
<td>ATGGGGAGTTTGGGAGCGATTGTGAAGCATCCAGATGATTTGTACCCGCTGCTGAAGCTC</td>
<td>AAAATGGCGGCGAGGCACGCCGAGAAGCAGATCTCCTCGGCGCGCGATGCGGTTCTGC</td>
<td>TACTCCATGCTCCACAAGGTTTCTAGAAGCTTCGCCCTCGTCATTCAGCAACTCGACACT</td>
<td>GAAGACATTACTGACAGAATGGGTGCTGGAATGGCCAAATTTATTTGCAAGGAGGTAGAA</td>
<td>ACAATTGATGACTACGACGAATATTGTCACTATGTGGCAGGACTTGTTGGGCTGGGTTTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. asiatica</td>
<td>ATGGGAAGTTTAGGGGCGATTCTGAAGCATCCGGACGATTTCTATCCGTTGTTGAAGCTG</td>
<td>AAAATGGCGGCGAGGCATGCTGAGAAACAGATCCCGCCGGAGCCGCACTGGGCATTTTGT</td>
<td>GATCTTCGCAACGCCGTTTGCATATTCTACTTGGTTCTTCGAGCTCTGGATACCGTTGAG</td>
<td>GAAGATATTACCATGAGAATGGGTGCAGGAATGGCAAAATTTATATGCAAGGAGGTAGAG</td>
<td>ACAATTGACGATTATGATGAATATTGTCACTATGTGGCAGGTCTGGTTGGGTTAGGGTTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. thaliana</td>
<td>ATGGGAAGCTTGAGTACGATTTTGAGACATCCGGATGAGTTATATCCGCTCTTGAAGTTG</td>
<td>AAACTTGCGATTACGAAAGCTCAGAAGCAGATTCCACTTGAGCCACACTTGCAGGCCAC</td>
<td>GATGATACGAGCATAGCTACAGAAGTAAAGGTCCCCATACTGAAAGCTTTTCACCGTCAC</td>
<td>GATGACACTAGCATATCTACAGAGGTTAAAGTACCAATATTGAAGGCTTTTCATCGTCAC</td>
<td>ACAATTGATGACTATGATGAATACTGCCATTATGCTGCAGGACTTGTTGGTTTAGGTTTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SQS F1**

<table>
<thead>
<tr>
<th></th>
<th>713</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. japonicus</td>
<td>AAACGAGATCCCCAAATCACGCATGTTTTGGCCACGGCAGATCTGGAGTAAATATGTTAA</td>
</tr>
<tr>
<td>C. asiatica</td>
<td>AAATGAGATACCCAAGTCACGGATGTTTTGGCCCTCGTGAGATTTGGGGAAAATATGTTGA</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>TAATGAGAGACCAAATGTCGCGCATGTTTTGGCCTCGTGAGATTTGGGGAAAATATGTTGA</td>
</tr>
</tbody>
</table>

**SQS R1**

<table>
<thead>
<tr>
<th></th>
<th>719</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. japonicus</td>
<td>AAACGAGATCCCCAAATCACGCATGTTTTGGCCACGGCAGATCTGGAGTAAATATGTTAA</td>
</tr>
<tr>
<td>C. asiatica</td>
<td>AAATGAGATACCCAAGTCACGGATGTTTTGGCCCTCGTGAGATTTGGGGAAAATATGTTGA</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>TAATGAGAGACCAAATGTCGCGCATGTTTTGGCCTCGTGAGATTTGGGGAAAATATGTTGA</td>
</tr>
</tbody>
</table>

---

Rishi Kishore Vishwakarma  Ph.D Thesis  University of Pune  87
Chapter 3

Fig. 3.29: Multiple sequence alignment of SQS nucleotide sequences of *Lotus japonicus* (Acc. No. AB102688), *Centella asiatica* (AY787628) and *Arabidopsis thaliana* (Acc. No. NM_119631). Forward and reverse primer sequences are highlighted.

Table 3.3: Primers used for isolation of partial BmSQS gene

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Forward and Reverse primers (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squalene synthase (SQS)</td>
<td></td>
</tr>
<tr>
<td>SQS F1</td>
<td>TACTCCATGTCCTCCACAAGGTGTTTCT</td>
</tr>
<tr>
<td>SQS F2</td>
<td>GCTCTGGATACCGTTGAGGATGATAC</td>
</tr>
<tr>
<td>SQS R1</td>
<td>TCAATTTGTTTACCTCTCTTGCA</td>
</tr>
<tr>
<td>SQS R2</td>
<td>GTTCCAATAGCCATAATCTGTGGAAATTTG</td>
</tr>
</tbody>
</table>

Rishi Kishore Vishwakarma  Ph.D Thesis  University of Pune  88
3.4.3.2 Amplification of partial cDNA fragment of \textit{SQS} gene from \textit{B. monniera} by PCR

Approximately 700 bp amplicon was obtained in PCR with primer combination (SQS F2 and SQS R2) and cDNA as a template (\textbf{Fig. 3.30 A}). This amplicon was cloned in pGEM-T Easy vector and insert release was confirmed by restriction digestion with \textit{EcoRI} (\textbf{Fig. 3.30 B}) and sequenced. Sequencing result showed nucleotides of 669 bp (\textbf{Fig.3.31}). BLAST analysis showed more than 80\% homology with reported \textit{SQS} sequences at NCBI. Blast search of deduced amino acid sequence predicted the presence of putative conserved domains specific to Isoprenoid_Biosyn_C1 super family. This confirmed that isolated partial sequence was squalene synthase from \textit{B. monniera}.

\textbf{Fig. 3.30: A.} 1\% agarose gel showing amplification of ~ 700 bp (Lane1-3) of partial BmSQS, Lane M- Low range DNA ruler. \textbf{B.} Restriction digestion with \textit{EcoRI}. Lane 1 & 2- Showing vector backbone (~3 kb) and release of inserts (~700 bp), Lane M- Medium range DNA ruler.

\textbf{Partial BmSQSS nucleotide sequence}

\begin{verbatim}
1  GCTCTGGATA CCGTTGAGGA TGATACGAGC ATATCTGCAG AGGTCAAAGT
51  GCCTATTCTG AAGGCTTTCC ATCGTCACAT ATATGGCCAT GAATGGCATT
101 TTGCATGTGG TACGAAAGAC TATAAAGTTC TCATGGACGA GTTCCATCAT
151 GTTTCAACCG CTTTTCTGGA GCTTGGAAGT GGTTATCAGG AGGCAATTGA
201 GGAATTTACC ATGAGAATGG GTGCTGGAAT GGCACAAATTT ATTTGCAAGG
251 AGGTAGAAAC TGTTGATG ATATGATGAAT ATTGCCATTA TTGTGCTGGA
301 CTAGTTGGAT TAGGGTTGTC AAAGCTGTTC CATGCTCTG GAAAGGAGGA
351 TCTGGCTTCC GATTCACCTT CGAACTCAAAT GGGTTTPATTC CTTCCAGAAA
401 CAAATAATAT AAGAGACTAT TTGGAGGACGA TTATATGGAT ACCAAAATCG
451 CGTATGTTTT GCCCTCGTCA AATTTGGAGC AAAATGTTA ACAAACTTGA
501 GGACTTGAAA TATGAGGAAA ACTCTGTTAA GGCTGTGCAG TGCTCTGAATG
\end{verbatim}
551 ACATGGTAAC AAATGCTTTA ATACATGTCG AAAGATTGCA TCAAATACAT
601 GTCTGCATTG CGGGATCCAG CGATTCTTTC GTTTTTTGTT GCAATTCCAC
651 AGATTATGGC TATTGGGAAC

Fig. 3.31: Nucleotide sequences of partial BmSQS. Primers for RACE are underlined. Red coloured for 5’ RACE and blue coloured for 3’ RACE.

3.4.3.3 RACE PCR (5’ and 3’) of BmSQS
To obtain the 5’ and 3’ end of the gene, two steps RACE PCR was done. Gene Specific Primers were designed from above partial sequence of BmSQS (Fig. 3.31, Underlined).

3.4.3.3.1 5’ RACE
RaceSQS R 5’-CATCAACAGTTTCTACCTCCTTGC-3’
RaceSQS NR 5’-CAGCACCCATTCTCATGGTAATA-3’

Nested PCR with primers RaceSQS NR and 5’ NGRP yielded approximately 500 bp fragment (Fig. 3.32 A). It was cloned in pGEM-T Easy vector, confirmed by restriction digestion (Fig. 3.32 B) and sequenced.

Fig. 3.32: A.1% agarose gel showing amplification of ~ 500 bp (Lane 1 & 2) of 5’RACE PCR product, Lane M- Medium range DNA ruler. B. Restriction digestion with EcoRI. Lane 1 & 2- Showing vector backbone (~3 kb) and release of inserts (~500 bp), Lane M- Low range DNA ruler.

Sequence analysis of 5’ RACE clone revealed 5’ UTR of 62 bp (Red) and start codon (ATG, Green underlined) (Fig. 3.33). BLAST analysis of sequence showed significant similarity with reported SQS from other plants.
Fig. 3.33: Nucleotide sequence of \textit{BmSQS} 5’ RACE PCR clone. Start codon (ATG) green underlined and 5’UTR (Red). Primer, 5’ NGRP is underlined.

3.4.3.3.2 3’RACE

RaceSQS F 5’- GCAGTGCTCTGAATGACATGGTAAC-3’
RaceSQS NF 5’- GCATCAAATACATGTCTGCATTGC-3’

Approximately 750 bp amplicon was obtained in nested PCR with RaceSQS NF and 3’ NGRP (Fig. 3.34 A). This was then cloned in pGEM-T Easy vector and confirmed by restriction digestion (Fig. 3.34 B) and sequencing.

Fig. 3.34: A. 1% agarose gel showing amplification of ~ 750 bp (Lane1-3) of 3’RACE PCR product, Lane M- Low range DNA ruler. B. Restriction digestion with \textit{EcoRI}. Lane 1 & 2- Showing vector backbone (~3 kb) and release of inserts (~750 bp), Lane M- Low range DNA ruler.

Sequence analysis of 3’ RACE clone showed stop codon (TGA, red coloured), 3’ UTR of 263 bp (Green coloured) and putative polyadenylation site (Orange coloured, underline) followed by poly A tail (Fig. 3.35).
Fig. 3.35: Nucleotide sequence of 3’ RACE clone of BmSQS. Stop codon (Red), 3’ UTR (Green) and polyadenylation site (orange underlined) are shown. Primer 3’NGRP is underlined.

3.4.3.4 Amplification of full-length BmFPS ORF

Primers were designed from start and stop codon to amplify full-length ORF of BmSQS as a single PCR product on the basis of 5’ and 3’ RACE sequences.

SQS Full F 5’- ATGGGTAGTTTGCGTGCG-3’
SQS Full R 5’- TCACATGTGAGTGGGCCGAG-3’

PCR was performed with above primers SQS Full F/R and cDNA as a template using high fidelity Taq DNA polymerase. Approximately 1.2 kb amplicon (Fig. 3.36 A) was cloned in pGEM-T Easy vector and confirmed by restriction digestion (Fig. 3.36 B) and sequencing.

Fig. 3.36: A. 1% agarose gel showing amplification of ~ 1.2 kb (Lane1&2) of full-length BmSQS PCR product, Lane M- Medium range DNA ruler. B. Restriction digestion with EcoRI. Lane 1-3 - Showing vector backbone (~3 kb) and release of inserts (~1.2 kb), Lane M- Medium range DNA ruler.
Sequence analysis of 8-10 clones showed the exact size of 1245 bp (Fig. 3.37) and full-length ORF with 5’ and 3’ UTRs was submitted to NCBI GenBank data base having accession number GU734711.

1 ATGGTATGCCT CCGTTCGTTT CCGTATGACG TGGATGATGT TGATACGTAG
51 GCTTGGCGGA GGAGATGTCC AAGTCATTGC TGGATGATGT TCGTATCTGC
101 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
151 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
201 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
251 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
301 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
351 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
401 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
451 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
501 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
551 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
601 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
651 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
701 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
751 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
801 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
851 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
901 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
951 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
1001 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
1051 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
1101 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
1151 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT
1201 TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT TGGATGATGT

Deduced amino acid sequence

MGSLRAILKHPDLPVMKVLKAVRRAEIQPFEFHVGFCYTMLHKVSRSFALVIQQLDTLRDAVRIFYLVLR
ALDTVEEDTSISAELVKPVILKAFHRHYDERHWFACGTGDKYVMDEFHSTAFFLESGYGEAQIEDMRM
AGMAKFICKEVETVVDYECHRVCVLGLGKLFLHSGKEDLADISLNSMLFLQKTNIRDYLEDINEIF
KSRMFWPRQIWSKYNVKELEDYKVEENVSVAQVCLNMVNAHLHVED accountable ICYI
LALCYNNIQFVFRGVMRGLITAKIDRTKSMDVGGAFYDFSRLMKSVEDDSDPNARCTKIDLEAILKIC
RNLNGQRSYIQSKFQNYATSIVIIIIFIIIAILLAYLSATRKPHM

Fig. 3.37: Nucleotide and deduced amino acid sequence of full-length BmSQS.
3.4.3.5 Characterization of BmSQS

3.4.3.5.1 Protein sequence analysis of BmSQS

The deduced amino acid sequence of BmSQS shared 84.3%, 83.6%, 82.9% and 78.9% identities with *Salvia miltiorrhiza*, *Diospyros kaki*, *Capsicum annuum* and *Artemisia annua* respectively. The calculated molecular mass and predicted pI value of BmSQS were 47.37 kDa and 7.97 respectively. The total number of positively and negatively charged amino acids was 53 and 51 respectively. The computed instability index and aliphatic index were 36.77 and 95.17 respectively, indicates thermo stability of the BmSQS. According to NCBI conserved domain search, BmSQS showed specific hits with “Trans_IPPS_HH” and super family Ispprenoid_Biosyn_C1. Six (I-VI, **Fig. 3.38 B**) highly conserved SQS domains (Important for the catalytic activity of squalene synthase) were found after multiple alignment of BmSQS deduced amino acid sequence with other SQS proteins. Three domains (III, IV and V) showed highly conserved consensus sequence with other SQS proteins, whereas domains I and II are much less conserved (**Fig. 3.38 A**). Domain II showed highly conserved aspartate rich motif among the other SQS enzymes ((Devarenne et al., 1998). Domain IV was also aspartate rich motif and these motifs were proposed to coordinate/facilitate FPP binding through a magnesium ion requirement (Huang et al., 2007). Even though domain VI exhibits a low level of sequence homology with other SQS in alignment, the domain in the carboxy terminus of the protein contained highly hydrophobic residues, very similar to other SQS sequences and consistent with the putative endoplasmic reticulum membrane anchoring function.
Fig. 3.38: A. ClustalW alignment of BmSQS with other plant SQSs showing different domains (I-VI) and catalytic motifs in different color. B. CDD search results on NCBI server.
3.4.3.5.2 Hydropathy plot analysis of BmSQS

The amino acid sequences of BmSQS was analyzed using Kyte-Doolittle Hydropathy plot of JustBio hosted tool (http://www.justbio.com/index.php?page=plots) at window size 9. Plot analysis revealed even distribution of hydrophilic and hydrophobic residues (Fig. 3.39). Presence of two transmembrane domains was found in BmSQS on TMHMM bioinformatics program (Fig.3.40).

![Kyte & Doolittle Hydropathy Plot](image)

**Fig. 3.39:** Kyte-Doolittle Hydropathy plot. Arrow indicates C-terminus highly hydrophobic region.

![TMHMM analysis of BmSQS](image)

**Fig. 3.40:** TMHMM analysis of BmSQS. Trans membrane domain I (281-303 AA), domain II (387-409 AA).
3.4.3.5.3 Phylogenetic tree analysis

A phylogenetic tree was constructed by Neighbor-Joining method with MEGA 4.0.2 program from the ClustalX2 alignment of the BmSQS with number of SQSs from different organisms which were retrieved from the NCBI GenBank database. Phylogenetic analysis demonstrated that BmSQS had closer relationship with dicots SQS (Cluster I) including *S. miltiorrhiza*, *D. kaki*, *Lotus japonicas* and *A. annua* (Fig. 3.41).

Fig. 3.41: Phylogenetic tree analysis of BmSQS. Cluster I- Dicots, II- Monocots, III- Gymnosperm, IV- Algae, V- Yeasts, VI- Bird, VII- Mammals, VIII- Bacterium. Tree was formed by MEGA 4.0.2 using neighbor-joining method.
3.5 Conclusions

- Three isoprenoids biosynthetic pathway genes encoding HMG-CoA reductase ($BmHMGR$), FPP synthase ($BmFPS$) and Squalene synthase ($BmSQS$) have been isolated from $B. monniera$ by PCR based approach.
- Two isoforms of $BmHMGR$ ($BmHMGR1$ and $BmHMGR6$) have been identified sharing 98% identity with each other and shared more than 73% sequence homology in BLAST analysis.
- $BmFPS$ and $BmSQS$ showed more than 80% sequence homology with other reported sequences at NCBI.
- RACE PCR was used to identify 5’ and 3’UTRs of all three genes.
- Analysis of deduced amino acid of $BmHMGR$ showed presence of different domains involved in catalytic activities. Presence of two transmembrane domains at N-terminus and one at C-terminus were predicted.
- In case of $BmFPS$ presence of two domains (domain I & II) involved in substrate binding, Mg$^{2+}$ binding sites and active site lid residues shield highly reactive carbocationic intermediates from solvents, were predicted.
- Hydropathy plot analysis of $BmSQS$ showed presence of highly hydrophobic region at C-terminus and two transmembrane domains at C-terminus were also detected.
- Phylogenetic analysis revealed, isolated genes are closely related with other dicot plant species and $B. monniera$ genes are close to $S. miltiorrhiza$. 