Pharmacogn. Mag.
A multifaceted peer reviewed journal in the field of Pharmacognosy and Natural Products
www.phcog.com | www.phcog.net

ORIGINAL ARTICLE

Cloning, Identification, and *in silico* Analysis of Terpene Synthases Involved in the Competing Pathways of Artemisinin Biosynthesis Pathway in *Artemisia annua* L

Umara Nissar Rafiqi[1,2], Irum Gul[1], Monica Saifi[1], Nazima Nasrullah[1], Javed Ahmad[1], Prasanta Dash[2], Malik Zainul Abdin[1]

[1]Department of Biotechnology, Centre for Transgenic Plant Development, [2]Indian Agricultural Research Institute, New Delhi, India

Submitted: 15-05-2018 Revised: 22-06-2018 Published: 26-04-2019

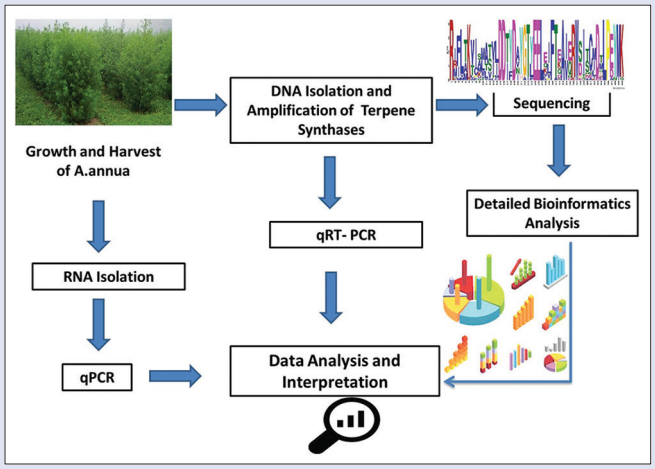
ABSTRACT

Background: Endoperoxide sesquiterpene lactone, artemisinin, is a widely used antimalarial drug. *Artemisia annua* L. synthesizes this terpenoid and is the only source of artemisinin. In plants, the content of artemisinin is low (0.1–0.8% by dry weight). One of the best approaches to increase artemisinin production is metabolic engineering. **Methods:** Both the genes were amplified and cloned in Topo vector. Using computational approach, full gene sequencing and a detailed *in silico* analysis was performed to check the functional and structural properties of these enzymes. Expression patterns of both the genes were assessed at different developmental stages (vegetative, preflowering, flowering, and postflowering stage) of the plant reverse transcription polymerase chain reaction. **Results:** Deduced amino acid sequence of these genes possessed two important and highly conserved aspartate-rich motifs, and lacks an N-terminal signal peptide, a characteristic of sesquiterpene synthases. Physiochemical properties demonstrated are thermostable. Low hydropathy values ascertain them to be hydrophobic and are active at neutral pH. Structural analysis disclosed that both the proteins possess more α-helices followed by random coils. Ramachandran analysis showed a C-score of −0.35, TM-score of 0.67 ± 0.13 for β caryophyllene synthase model while as C-score of −0.21, TM-score of 0.69 ± 0.12 for β-Farnesene synthase model. Both the proteins contain enormous nitrosylation sites suggesting their functional link through nitrosylation. Gene expression pattern of both the enzymes were upregulated during preflowering and flowering stage. **Conclusion:** A thorough analysis of these two putative genes in *A. annua* L paves way to essential insights concerning terpene biosynthesis in general and regulation in artemisinin production in particular. This study also strongly indicates that these two enzymes are developmentally controlled and may have the regulatory effects on the terpene biosynthesis.

Key words: *Artemisia annua,* artemisinin, DDXXD, E-β-caryophyllene synthase, E-β-farnesene synthase, NSE/DTE, RXR, terpenes

SUMMARY

- Endoperoxide sesquiterpene lactone, artemisinin, is a widely used antimalarial drug. *Artemisia annua* L. synthesizes this terpenoid and is the only source of artemisinin. At present, *A. annua* plant is the only commercial source of artemisinin. Its content in the plants is however relatively low (0.1%–0.8% by dry weight) compared to its demand in international market. To increase the content of artemisinin, understanding of its complete biosynthetic pathway as well as competing pathways is required. In the present study, a thorough analysis of the two putative sideways competing pathway genes E-β-Farnesene synthase and E-β-Caryophyllene synthase genes from *A. annua* L were studied using computational approach. This analysis showed several interesting aspects related to their structure and brought novel information-related substrate binding. This data may provide a way forward in understanding their regulatory role in artemisinin biosynthesis. However, the experimental validation for the direct involvement of these enzymes in artemisinin biosynthesis is underway.



Correspondence:

Prof. Malik Zainul Abdin,
Department of Biotechnology, Centre for Transgenic Plant Development, Jamia Hamdard, Hamdard Nagar, New Delhi - 110 062, India.
E-mail: malikzabdin@gmail.com

DOI: 10.4103/pm.pm_244_18

Access this article online
Website: www.phcog.com
Quick Response Code:

INTRODUCTION

Artemisinin, which is a widely used antimalarial drug is obtained from *A. annua* L. (*Asteraceae*).[1] It is effective against malaria, especially the cerebral and chloroquine-resistant forms of this disease. Besides antimalarial activity, artemisinin and its derivatives have been reported to possess antiviral, anticancer, and antischistosomal activities.[2,3] At present, *A. annua* plant is the only commercial source of artemisinin. Compared to its demand in international market, artemisinin content is relatively low in plants (0.8% by dry weight). Lower content leads

Cite this article as: Rafiqi UN, Gul I, Saifi M, Nasrullah N, Ahmad J, Dash P, *et al*. Cloning, identification, and *in silico* analysis of terpene synthases involved in the competing pathways of artemisinin biosynthesis pathway in *Artemisia annua* L. Phcog Mag 2019;15:S38-46.

to meager production of artemisinin that results in increased cost of artemisinin-based treatment, especially in developing countries where malaria is endemic.[4,5] Numerous efforts have been made to improve artemisinin production, to reduce the price of artemisinin-based antimalarial drugs which involves many physiological and cell culture studies.[6] The alternative source of artificial artemisinin involves chemical synthesis in laboratory, but it has met with limited success due to its complexity and poor yield.[7] The importance of metabolic engineering has been recently reported to improve artemisinin production in plants and microbe and is considered as one of the best approaches to increase artemisinin production.[8] The limitation with this approach is that it depends on either the biotransformation using the plant source[9] or semisynthesis[10] for the end results.

Terpenoids are large and diverse class of secondary metabolites synthesized by a special class of enzymes i.e., terpene synthases (TPS). On the basis of distribution of introns and exons, TPSs are classified into 7 clades: TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f, TPS-g, and TPS-h. TPS-a, TPS-b, and TPS-g clades are discretely found in angiosperms, with TPS-a containing mostly sesquiterpenes. Sesquiterpenoids are biosynthesized from farnesyl pyrophosphate (FPP).[11] Mevalonate and 2-C-methyl-D-erythritol 4-phosphate pathways generate a product, isopentenyl pyrophosphate, which is the precursor of the biosynthesis for FPP in *A. annua* L. plants.[12] The first committed step of artemisinin biosynthesis is cyclization of FPP into amorpha-4,11-diene (ADS), to produce carbon skeleton for artemisinin biosynthesis.[13]

Besides artemisinin biosynthesis, FPP is also used as a precursor for the synthesis of various other terpenoids in the sideways competing pathways such as caryophyllene, farnesene, and sterols as shown in Figure 1. The modulation of TPS, involved in competing pathways, may also increase or decrease the artemisinin production in *A. annua* plants.[14] In the present study, we have cloned and characterized two terpene synthase genes, *(E)-β farnesene synthase* and *(E)-β caryophyllene synthase* (*bcs* and *bfs*, respectively) from *A. annua* plants. BFS and BCS are involved in biosynthesis of secondary metabolites farnesene and caryophyllene, respectively. Although secondary metabolites are not required for plant growth and development, they play an important role in plant defense mechanism.[15] An *in silico* analysis was performed to get an insight into the functional and structural properties of TPS, bcs, and bfs. Analysis of these two putative TPSs in *A. annua* plants showed several interesting aspects related to structure and substrate binding. Understanding of the artemisinin biosynthesis pathway along with the competing pathways and their regulation with an aim to improve the artemisinin content of *A. annua* L. employing system biology approach is required. The information available regarding the structural and functional contribution of TPS involved in these pathways is meager.

MATERIALS AND METHODS

Plant materials and tissue culture conditions

Seeds of artemisinin yielding genotype of *A. annua* L. plants were acquired from herbal garden, Jamia Hamdard, New Delhi, India**.** These were immersed in 70% ethanol for 2 min and then surface sterilized by soaking in 5% sodium hypochlorite for 20 min. Thereafter, the seeds were rinsed with distilled water and allowed to germinate under sterile conditions in 50 ml germination medium (½ MS media, Himedia). After germination, the plantlets were grown in a controlled-growth chamber in a light/dark cycle of 16/8 h using fluorescent lamps (with a light intensity of 2800 lx) at 25°C and 70% relative humidity for three weeks. Seedlings were then transferred to greenhouse and allowed to grow for four more weeks. Young green leaves from these plants were collected for genomic DNA isolation.

Database analysis and primer designing

TPS of *A. annua L.*, namely *bfs* and *bcs* were included in the study. The primers for both the genes were designed from 3' and 5' UTR regions using Clone manager suite 7 (Sci-ED software) to clone full-length gene sequences including introns and exons. A four base pair overhang CACC was added to the 5' end of forward primer to allow directional cloning of genes. The primer sequences are summarized in Table 1.

Isolation and cloning of terpene synthases genes

Genomic DNA was extracted from the fresh leaves of *A. annua* L. at preflowering stage using the DNeasy plant mini kit (Qiagen) as per manufacturer's instruction. The primer annealing temperature for optimum amplification of bcs and bfs genes was optimized by gradient polymerase chain reaction (PCR) temperature which ranges from 50°C to 60°C. The maximum amplification temperature for *bcs* was found to be 56.2°C, and for *bfs,* it was 54.3°C. PCR was carried out using *phusion taq* polymerase on a PCR thermal cycler with following temperature program: 98°C for 5 min, followed by 35 cycles of amplification (94°C for 30 s, 56.2°C for bcs; 54.3°C for bfs for 30 s, and 72°C for 2 min 30 s) and 72°C for 10 min.

The PCR products were electrophoresed through 0.7% agarose gel and eluted using QIAquick Gel extraction kit (Qiagen). Before ligation, the PCR products of *bcs* and *bfs* were quantified and diluted as per the cloning system's required insert-vector ratio. PCR products were directionally cloned into pENTR/SD/D/TOPO vector (Thermofisher). Ligation was carried out following the manufacturer's protocol optimized for TOPO cloning. The ligation mixture was used to transform DH5α competent cells provided with the kit. The positive colonies were restreaked and confirmed through colony PCR (94°C for 30 s, 56.2°C for BCS; 54.3°C for BFS for 30 s and 72°C for 2 min 30 s and 72°C for 10 min). The positive colonies on LA (Luria agar) plates were picked and used for plasmid extraction, and restriction digestion was done using *NotI* restriction enzyme.

Sequencing and sequence alignment of E-β-Farnesene synthase and β-Caryophyllene synthase

Along with universal M13 forward and reverse primers, three set of gene-specific primers were used to sequence the two genes on both the strands by primer walking to determine the complete sequence of genes [Supplementary Figure 4a and b]. Protein sequences were deduced using ExPasy translate tool[16] and were converted into *fasta* format for further analysis.

The sequence alignment of DNA, cDNA, and deduced amino acid sequences of both the proteins encoded by bcs and bfs available in NCBI revealed the position and length of introns and exons, respectively. The alignment of cloned sequences to know introns and exons was done with BIOEDIT software using the default parameters (http://www.mbio.ncsu.edu/bioedit/bioedit. html).The alignment of nucleotide sequences and deduced amino acid sequences helped in predicting the conserved regions of both the TPS (*bcs* and *bfs*). The FASTA sequences of BCS and

Table 1: Primer sequences used for directional cloning of *E-β-Caryophyllene synthase* and *E-β-Farnesene synthase*

Gene	Sequence
bcsF	5'-CACC-CAATCCAAACTTCTCATAGACATG-3',
bcsR	5'-ACAAATGCCACACAGAAGAGG-3'
bfsF	5'-CACC-ATGTCGACTCTTCCTATTTCTAG-3',
bfsR	5'-TTAGACAACCATAGGGTGAACGAAG-3'

bcs: E-β-Caryophyllene synthase; bfs: E-β-Farnesene synthase

E-BFS proteins were uploaded in dbSNO 2.0 under the SNO prediction tool to predict cysteine nitrosylation sites in bcs and bfs.

Physiochemical characterization and phylogenetic relationship

Expasy's ProtParam server was used for primary structure analysis of both the genes[16]. The biophysical and biochemical properties such as pI, molecular weight, instability index,[17] aliphatic index,[18] extinction coefficient, and GRAVY[19] were computed using this program. Conserved domain search for functional characterization of proteins was performed using the conserved domain database available at NCBI. MEME Suite (http://meme.ncbr.net | meme | cgi-bin | meme.cgi) was used to predict the motifs of proteins.[20] Phylogenetic relationship was assessed using MEGA5.1 software to draw evolutionary trees for both the genes.

Structure analysis

Amino acid sequences of both the genes were analyzed for the secondary structure prediction.[21] For secondary structure prediction, PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/) was used, which provides a simple and accurate secondary structure prediction method.[22] The relative availability of alpha helix, extended strand, and random coils was determined for both the protein sequences. To predict the role of α-helix, β sheets, and random coil structures at each position based on 17 amino acid sequence windows, the deduced protein sequences were calculated using I-TASSER prediction tool.[23] PROCHECK program was used to check the stereochemical excellence and the overall structural geometry of the homology model.[24]

Analysis of terpene synthases (E-β-Farnesene synthase and β-Caryophyllene synthase) of *Artemisia annua*-L by reverse transcription polymerase chain reaction

We extracted total RNA from 30–50 mg of leaves using RNeasy Plant mini kit (Qiagen), following manufacturer's instructions. cDNA was generated from 5 μg of total RNA using Maxima First Strand cDNA synthesis kit (Thermo Scientific, USA). Expression patterns of both the enzymes were assessed at different developmental stages (vegetative, preflowering, flowering, and postflowering stage) of the plant using gene-specific RT primers, purchased from Applied Biosystems. Using a LightCycler® 480 System (Roche Diagnostics), quantitative PCR was performed. Each reaction for analysis was carried out in triplicates and was normalized using glyceraldehyde 3-phosphate dehydrogenase as a reference gene. The data are represented by $2^{\Delta\Delta CT}$ method to show the relative mRNA expression. The sequences of primers used are listed in Table 2.

RESULTS

Cloning and primary structure analysis

The terpene synthase genes *bcs* and *bfs* were amplified from *A. annua* L. by PCR. The sequence of *bcs* was 2554 bp nucleotides in length and contained 1863 bp translational region encoding 621 amino acids, whereas *bfs* was 2560 bp nucleotides in length and contained 1716 bp translational region encoding 572 amino acids [Supplementary Figure 1a-d]. Analysis of the genomic structure of bcs revealed that it contained 5 introns and 6 exons, whereas bfs contained 7 introns and 8 exons. As computed from the 3' end of mRNA sequence, the position of introns was considerably constant, but the length of introns varied to a great extent [Figure 2a and b]. Both the genes were examined for the presence of conserved domains. The C-terminal domain of both these genes contained two important and highly conserved aspartate-rich motifs, namely DDXXD and NSE/DTE. Positionally, conserved RXR motif was observed 35 amino acids upstream of DDXXD motif [Figure 3a and b]. A phylogenetic tree of E-β–Farnesene (BFS), BCS, and similar proteins from different plant species was constructed using MEGA5.1 to investigate the evolutionary relations. This revealed that both the proteins are originated from common ancestor and share similarities with TPS [Figure 4a and b].

Physiochemical properties

Expasy's ProtParam tool was used for computing the physiochemical properties of both the genes. The calculated parameters are enlisted in Table 3. The calculated molecular weight of BCS is 72kDa and that of BFS is 66kDa. Aliphatic index for the protein sequences of BCS and BFS are 89.81 and 89.25, respectively. The Grand Average Hydropathy (GRAVY) value for BCS and BFS are −0.174 and −0.252, respectively. The pI value for BCS protein is 5.05 and for BFS protein is 5.10. The instability index value for BCS protein is 47.37, and for BFS protein, it is 57.52. The estimated half-life of BCS protein in different cell systems are 30 hrs (mammalian reticulocytes, *in vitro*), >20 hrs (*yeast*, *in vivo*), >10 hrs (*Escherichia coli, in vivo*) and that of BFS are 30 h (mammalian reticulocytes, *in vitro*), >20 h (yeast, *in vivo*), and >10 h (*E. coli, in vivo*).

The extinction coefficient of BCS is 95605; Abs 0.1% (=1 g/l) 1.500, assuming all pairs of cysteine residues form cysteines, extinction coefficient 95230; Abs 0.1% (1 g/l) 1.494, assuming all cysteine residues are reduced (cysteine does not absorb appreciably at wavelength >260 nm, while cysteine does). The extinction coefficient of BFS is Abs 0.1% (=1 g/l) 1.500, assuming all pairs of cys residues form cysteines, extinction coefficient 95230; Abs 0.1% (1 g/l) 1.494, assuming all cys residues are reduced. (Cysteine is the amino acid formed when a pair of cysteine molecules is joined by a disulfide bond).

Amino acid composition of major BCS and BFS was compared with reference, which showed that in BCS leucine (L), glutamic acid (E), and alanine (A) were the most prominent amino acids with leucine being the most variable amino acid noted. There was an increase in frequency of asparagine; tyrosine and serine while a decrease in frequency of histidine, and tryptophan in the BCS as compared to reference [Supplementary Table 1]. Similarly, BFS showed that glutamic acid (E), leucine (L), and valine (V) were the most prominent amino acids with glutamic acid being the most variable amino acid noted. An increase in frequency of serine and lysine was also observed as compared to the reference while there is a decrease in cysteine and tryptophan [Supplementary Table 2].

Table 2: Primer sequences used for reverse transcription-polymerase chain reaction of *E-β-Caryophyllene synthase* and *E-β-Farnesene synthase*

Gene	Sequence
BFScF	5'-CCCAGATACAAGAAGCGCTAAA -3'
BFScR	5'-TGGGACATCAAGACCTTTCC -3'
BCScF	5'-GCAACCCTGCACCATCTA -3'
BCScR	5'-AGACTCATTGTGAGATGCTAGTT-3'

bcs: E-β-Caryophyllene synthase; bfs: E-β-Farnesene synthase

Table 3: The parameters computed using Expasy's ProtParam tool in *β-caryophyllene synthase* and for *β-Farnesene synthase*, respectively

Enzyme	AA	Molecular weight (KDa)	PI	Instability index	Aliphatic index	Gravy
BCS	621	72.05	5.05	57.52	89.25	−0.25
BFS	552	66.36	5.1	47.37	89.81	−0.17

BCS: E-β-Caryophyllene synthase; BFS: E-β-Farnesene synthase

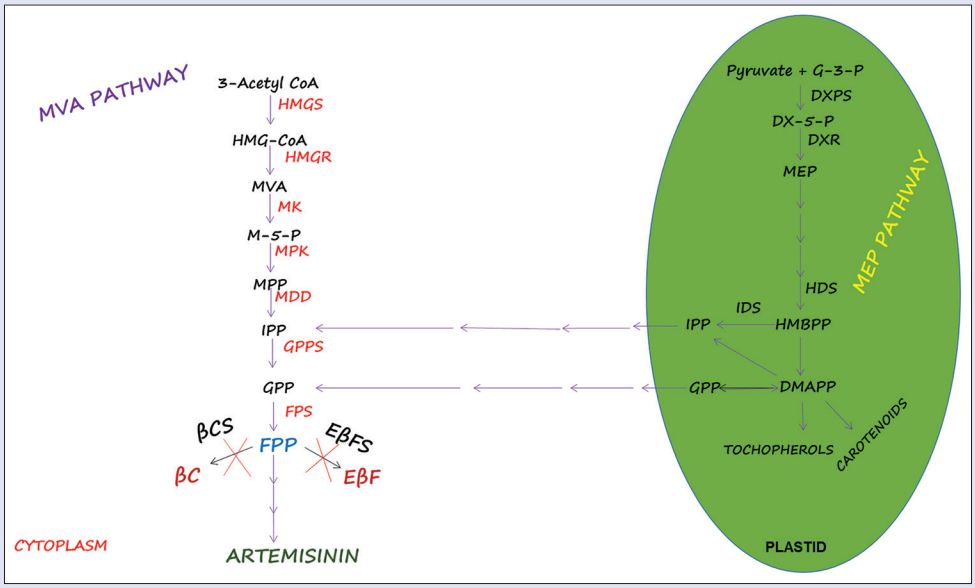


Figure 1: Artemisinin synthesis pathways in *Artemisia annua*. The enzymes shown in red belong to mevalonate pathway. The 2-C-methyl-D-erythritol 4-phosphate pathway enzymes are shown in green oval cartoon. The enzymes for artemisinin biosynthesis start from farnesyl diphosphate (farnesyl pyrophosphate), an intermediate product of terpenoid metabolism. The arrows between mevalonate pathway and 2-C-methyl-D-erythritol 4-phosphate pathway demonstrate the crosstalk between these two pathways during artemisinin biosynthesis. The enzymes starting from the branch point (assigned with a cross sign below them) showing two putative sideway pathway enzymes in terpenoid metabolism

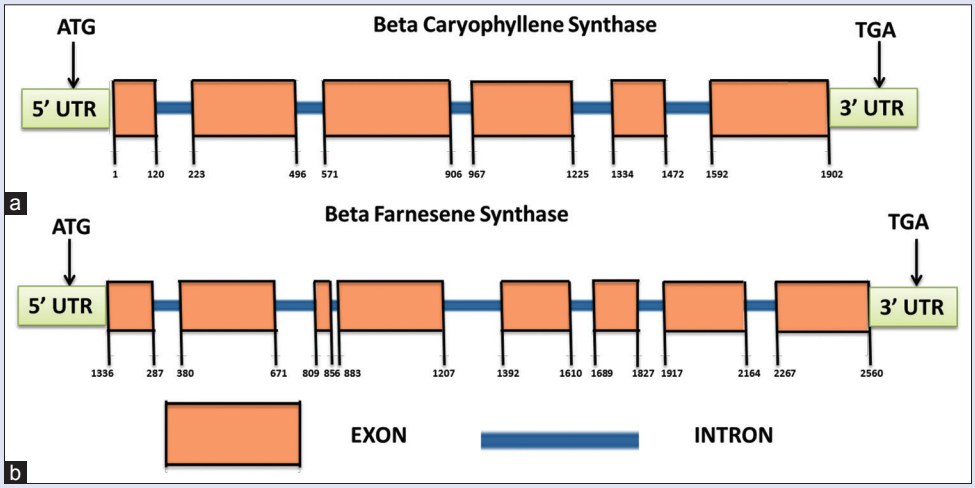


Figure 2: (a and b) Number and position of introns. Genomic organization of β-Caryophyllene synthase. (a) 6 exons in orange and 5 introns in blue and Beta Farnesene synthase gene. (b) 8 exons in orange and 7 introns in blue. Numbers shown below the lines are the start and end positions of respective exons

Structural analysis and model development

Using a stringent cross-validation method to evaluate the method's performance, an average Q3 score of 81.6% was achieved. The secondary structure showed that the BCS sequence consisted of 60.58% (332) α-helix (H), 25.36% (139) random coils (C), and 14.04% (77) β-sheets (E), and the BFS sequence consisted of 40.21% (232) α-helix, 34.32% (198) β-sheets, and 25.47% (147) random coils. The percentage distribution of predicted secondary features, i.e., alpha helix, extended strand, and random coils are represented in Table 4. The secondary structure results revealed that the predicted alpha helix dominated among other features followed by random coils, extended strands for both the protein sequences.

Using the alignment as input, four different structural models were generated for both the genes using I-TASSER server [Supplementary Figure 2a and b]. The structure fulfilling all the structural constraints in accordance with Ramachandran plot was chosen for further analysis ([Figure 5a and b], respectively). In both the modeled proteins, we observed that most of the residues were in the favored regions when compared to reference model, limonene synthase (*Mentha spicata*).

Table 4: The summary of secondary structure elements identified in *β-caryophyllene synthase* and for *β-Farnesene synthase* proteins, respectively

Enzyme	Alpha helix (%)	Extended strand (%)	Random coil (%)
BCS	60.58	14.05	25.36
BFS	40.21	25.48	34.32

BCS: E-β-Caryophyllene synthase; BFS: E-β-Farnesene synthase

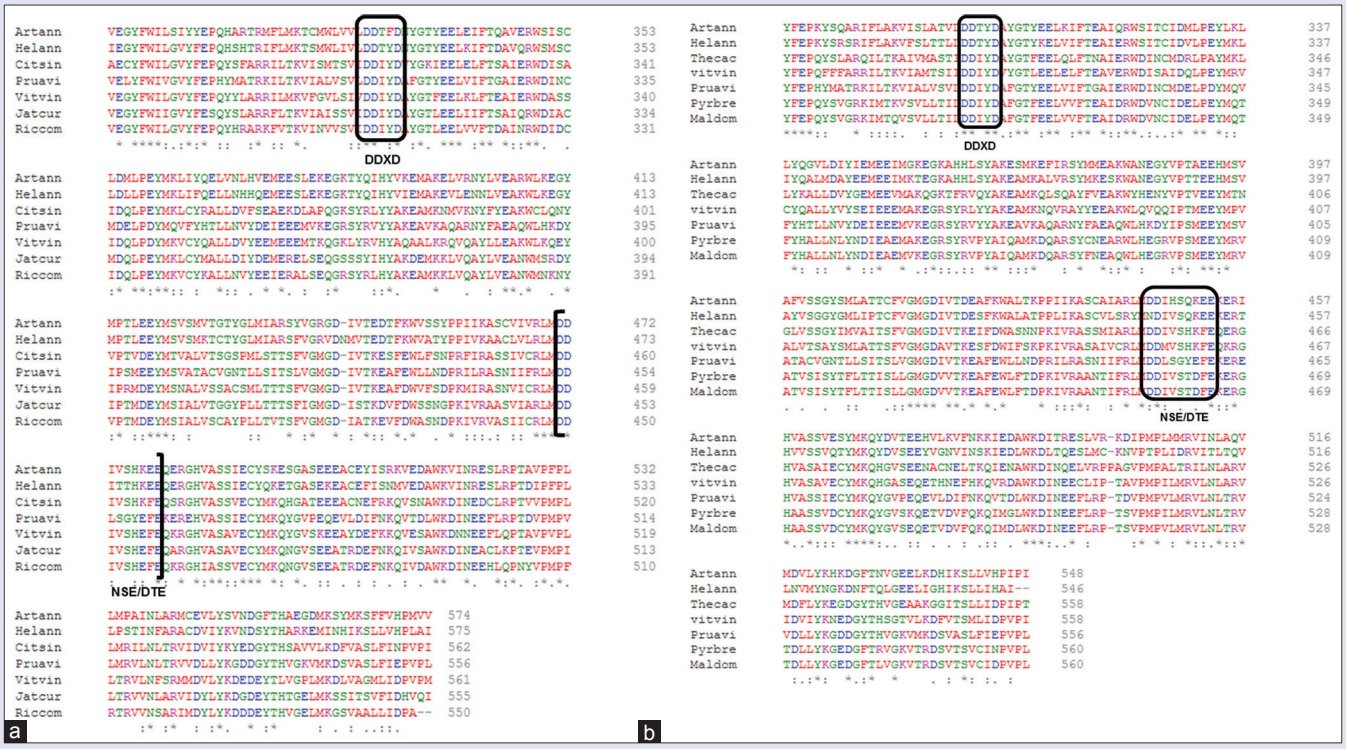


Figure 3: (a and b) Multiple sequence alignment and comparison of the deduced amino acid sequences of β-Farnesene synthase and β-Caryophyllene synthase and related proteins: from the BLASTX analysis, the identified homologs was aligned with the deduced amino acid sequence of E-β-Farnesene synthase and β-Caryophyllene synthase. Sequences highlighted in black indicate identical residues, while those in gray indicate similar residues. The highly conserved motifs DDXXD, RXR, and NSE/DTE are highlighted in black boxes. Artann: beta-caryophyllene synthase QHS1 (*Artemisia annua*); Helan: beta-caryophyllene synthase-like (*Helianthus annuus*); Vitvin: germacrene D synthase isoform X1 (*Vitis vinifera*); Pruavi: alpha-pinene synthase-like (*Pyrus x bretschneideri*); Thecac: delta-cadinene synthase isozyme A (*Theobroma cacao*); Maldom: alpha-pinene synthase-like (*Malus domestica*); Pyrbre: alpha-pinene synthase-like (*Prunus avium*)

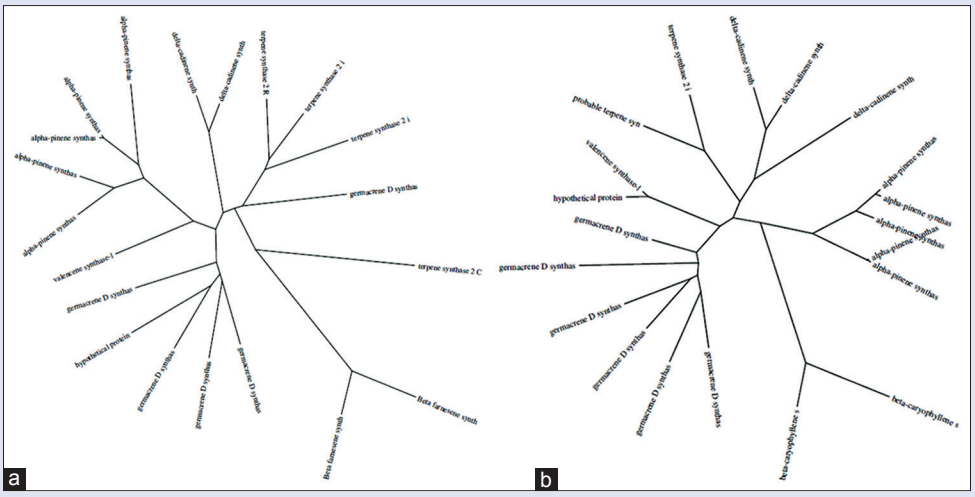


Figure 4: Phylogenetic tree of the amino acid sequences of: (a) β-Farnesene synthase of *Artemisia annua* and (b) β-Caryophyllene synthase of *Artemisia annua* and other closely associated plant species constructed by neighbor-joining method using MEGA5. 1: GenBank accession numbers: for β-Farnesene synthase, AAX39387.1 (*Artemisia annua*), XP_022017387.1 (*Helianthus annuus)*, XP_002282488.1 (*Vitis vinifera)*, XP_020535626.1 (*Jatropha curcas*), XP_002523635.1 (*Ricinus communis*), XP_021801780.1 (*Prunus avium)*, XP_006475286.1 (*Citrus sinensis)* and for β-Caryophyllene synthase, AAL79181.1 (*Artemisia annua)*, XP_021988936.1 (*Helianthus annuus*), XP_019072406.1 (*Vitis vinifera*), XP_009355684.1 (*Pyrus x bretschneideri*), XP_017979694.1 (*Theobroma cacao)* NP_001281061.1 (*Malus domestica*), XP_021801780.1 (*Prunus avium)*

While BCS model has a C-score of − 0.35 and BFS model has a C-score of − 0.21. An estimated TM-score of 0.67 ± 0.13 and 0.69 ± 0.12 was obtained for BCS and BFS, respectively. In both the modeled proteins (BCS and BFS), 96.3% and 96.1% of residues were observed in the favored regions, respectively, whereas 2.4% and 3.0% residues were observed in the allowed regions, respectively, as compared to the reference [Table 5].

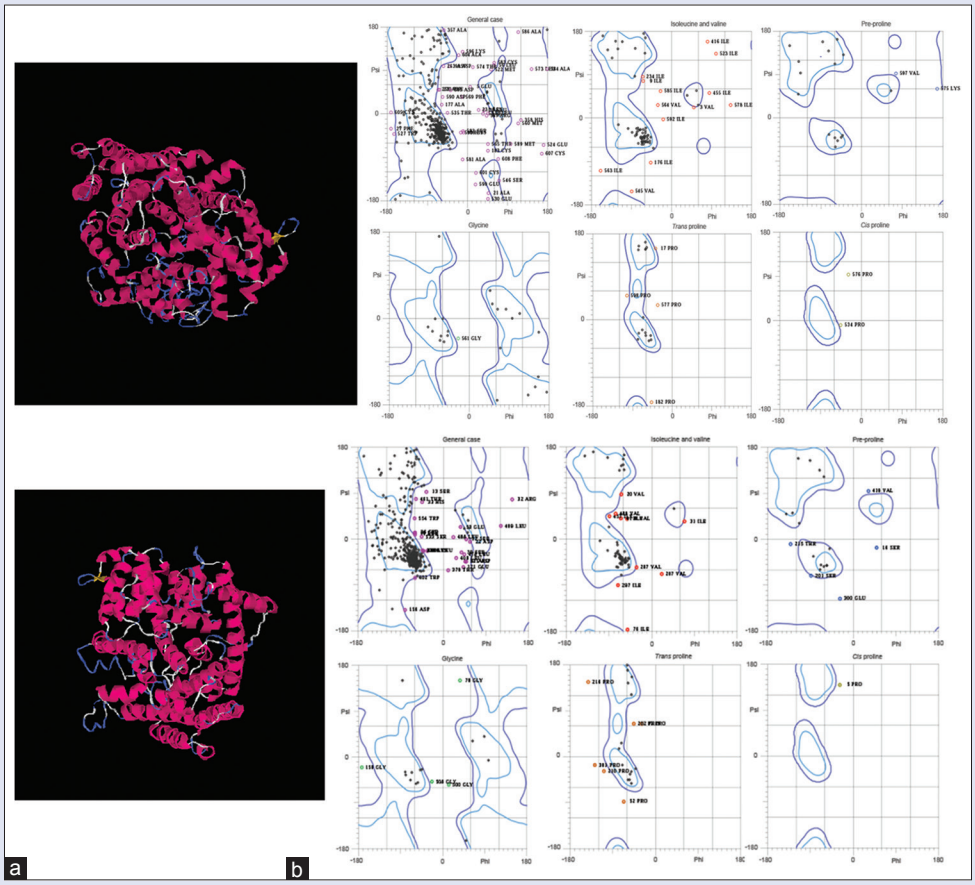


Figure 5: Homology modeling and evaluation of tertiary structure: Prediction of tertiary structure of the modeled (a) β-Caryophyllene synthase and (b) E-β-Farnesene synthase generated by the I-TASSER server. The different domains of the protein are color coded

Table 5: The summary of Ramachandran plot analysis for *β-caryophyllene synthase* and for *β-Farnesene synthase* proteins, respectively

Enzyme	Residues in most favored regions (%)	Residues in additionally allowed regions (%)	Residues in outlier regions (%)
BCS	−96.3	−2.4	−1.2
BFS	−96.1	−3	−0.9

BCS: E-β-Caryophyllene synthase; BFS: E-β-Farnesene synthase

BCS and BFS synthase proteins harbor predicted sites for nitrosylation of cysteine residues. β-Caryophyllene synthase has a total of 6 sites of s-nitrosylation [Figure 6a] whereas E-BFS has 10 predicted sites for s-nitrosylation [Figure 6b].

Reverse transcription polymerase chain reaction analysis

We made an estimation of relative mRNA expression of *bcs* and *bfs* enzymes at different developmental stages of the plant. We observed that the mRNA expression patterns of both the biosynthetic enzymes were highest at preflowering and flowering stage in the plant leaves, followed by vegetative and least expression was observed in postflowering stage of the plant [Figure 7a and b]. This implies that their primary function may be mainly confined to the preflowering and flowering stage.

DISCUSSION

Artemisinin is a medicinal compound obtained from the marginalized medicinal crop, *A. annua* L. A key constituent of *A. annua* is artemisia. Many studies have been so far conducted, to improve qualitative and quantitative production of artemisinin. Earlier approaches of genetic engineering carried for increased production of artemisinin were focused on the overexpression of genes involved in the artemisinin biosynthesis such as FPP, ADS, and CYP.[25-27] However, the overexpression of genes had little impact on artemisinin content;[28] therefore, a broader insight is required for enhancing its content in plant. In recent reports, the role of enzymes involved in competing pathways was highlighted to have a regulatory effect on the artemisinin biosynthesis.[29] However, little knowledge is available about the chemical and structural aspects of these enzymes. Aimed at providing a better understanding of these enzymes involved in competing pathway sideways of artemisinin biosynthesis. In the present study, we examined the sequential and structural aspects of BCS and BFS, two bona fide enzymes of this pathway in detail. We chose to directionally clone these genes into pENTR/SD/D/TOPO cloning vector. Clones were confirmed through colony PCR and restriction digestion. The full-length sequences of BCS and BFS served as an input sequence for further analysis.

Trapp and Croteau[30] classified these TPS on the basis of number of introns, which were reduced by the time of evolution. As computed from the 3' end of mRNA sequence, the position of introns is considerably constant but the length of introns varies to a great extent. Since BCS has 5 introns, we classified it to TPS-a subfamily, whereas BFS having 7 introns was classified into TPS-b subfamily, of class III terpenoid synthases. Unlike other members of TPS-a subfamily, BCS lacked sixth and final intron a characteristic of TPS-a subfamily. Similar results in terms of loss and gain of introns in TPS gene family was reported earlier in Arabidopsis and tomato.[11,31-33]

The presence or absence of conserved domains in sesqui-TPS has been reported and documented in earlier studies.[34] DDXXD is involved in the binding of water molecules and stabilization of the active site. It is also considered to be the binding site for the substrate divalent cation (Mn^{2+}, Mg^{2+}) complex.[35] RXR motif is thought to direct diphosphate ion away from the carbocation on cleavage of the substrate complex.[31] NSE/DTE is also reported to be consensus sequence (L, V) (V, L, A) (N, D) D (L, I, V) X (S, T) XXXE and a modified version LM (N, D) D (I, M) X (S, G, T) XXXE is found in both the TPS and forms a second divalent cation (Mg^{2+}) binding site in terpenoid synthases. We also demonstrated by paired alignment the presence of these consensus sequences in both the genes. These data are also in line with the absolute requirement for divalent metal ion as a cofactor for the substantial activity of sesqui-TPS.[36-38] Both these genes lack an N-terminal signal peptide, which is responsible for the transportation of mono and di-terpene synthase proteins to plastid, suggesting that both these genes encode a sesquiterpene synthase.

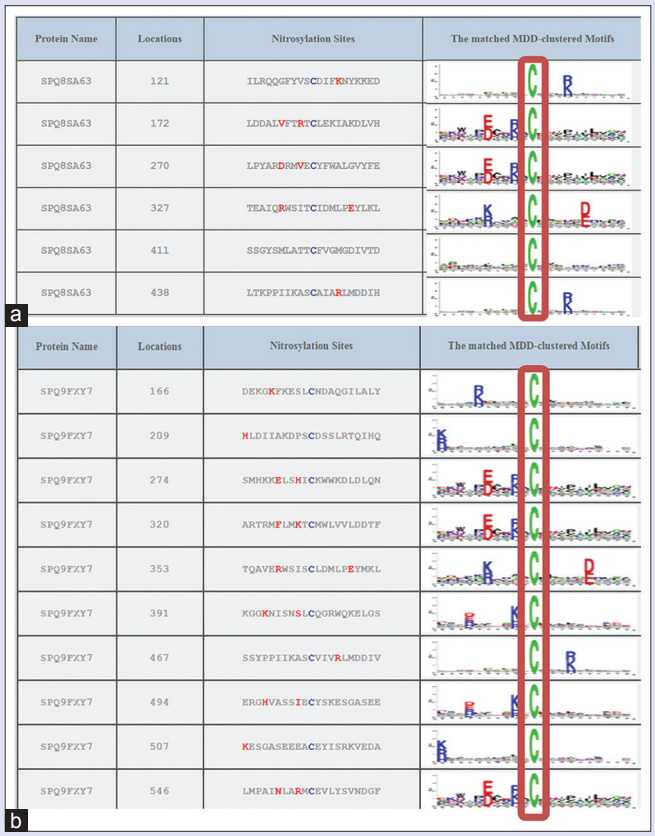


Figure 6: Predicted sites of cysteine nitrosylation: The predicted cysteine nitrosylation sites in (a) β-Farnesene synthase and (b) β-Caryophyllene synthase are represented by green bold letters highlighted in red box. The numbers in the table show the location of these predicted sites in the protein

Previous studies demonstrated that sesqui-TPS differ in amino acid composition and frequency.[37,39] Comparison of the nucleotide sequence of BCS and BFS revealed that deduced amino acid sequences differ in variability and frequency by three amino acids. Nevertheless, it was also noted that all the amino acids showing these differences were also different between the two sequences [Supplementary Tables 1 and 2]. In this study, we also observed that the alignment of the two deduced amino acid sequences was shorter than the reference sequences.[36-38] This is probably because of the better alignment and higher homology. The present study also demonstrated the alteration of the hydrophobicity of amino acid residues when compared to hydrophobicity index of each amino acid caused by variation [Table 3].

Depending on the physiochemical properties and location inside the cell, the proteins are assigned various functions.[16] The activity of a protein depends on the packing of its domains. Analysis of physiochemical properties of both the genes demonstrated novel information regarding the molecular weight, aliphatic index, GRAVY, pI value, instability index, estimated half-life and extinction coefficient were calculated and are summarized in Table 3. Both these TPS BCS (72kDa) and BFS (66kDa) are high molecular weight proteins. Since aliphatic index determines the thermostability of proteins, Both the protein sequences BCS and BFS showed higher values of aliphatic index, 89.81 and 89.25, respectively, implicating their thermostability within wider temperature range. The GRAVY value of a protein ascertains the interaction of a particular protein with water. The lower values of GRAVY for BCS and BFS, −0.174

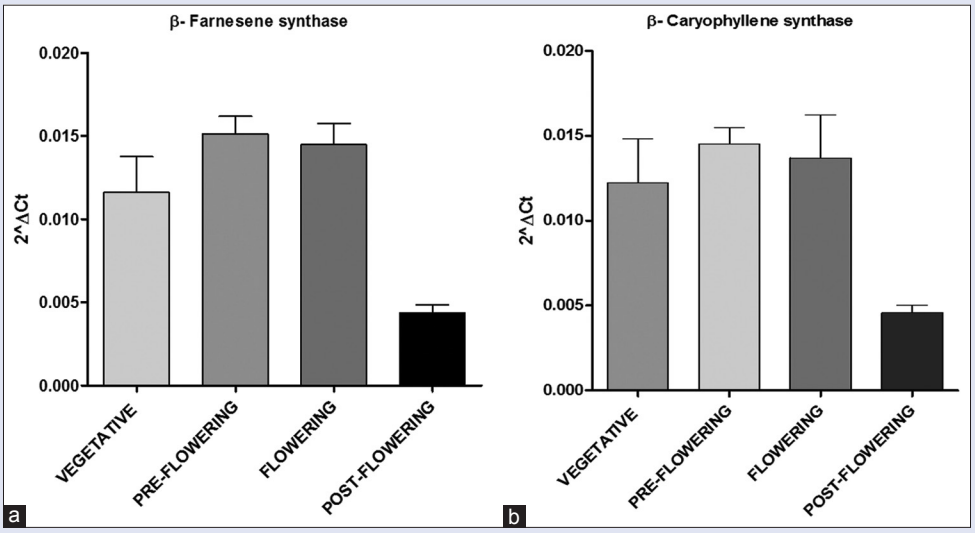


Figure 7: Differential gene expression pattern at different developmental stages of the plant: Increased expression of (a) β-Farnesene synthase and (b) β-Caryophyllene synthase was observed in preflowering and flowering stages followed by vegetative and post flowering stage. The data are expressed as mean ± SEM ($n = 5$) at every stage

and -−0.252, respectively, indicate the possibility of better interaction with water. Similarly, at pI, mobility of a protein in an electrofocusing system is zero. Both the proteins bear zero net charge at acidic pH. These data are also in line with the previous studies suggesting sesquiterpenes are active at neutral or basic pH confirming the observed results.[33]

The instability index evaluates the stability of a protein *in vitro*. Guruprasad *et al.* related the stability of a protein to its dipeptide composition.[17] The computed instability index, 47.37 for BCS and 57.52 for BFS protein sequences, fall in a range of highly unstable proteins. This is in agreement with the earlier studies suggesting the nonstatic and dynamic nature of sesqui-TPS.[33]

Protein cysteine nitrosylation (P-SNO) is physiologically important posttranslational modification that affects a wide variety of proteins and their activity. The SNO site prediction analysis showed that both the proteins E-BFS and β-caryophyllene synthase have multiple sites for cysteine nitrosylation. The presence of multiple SNO site indicates that their activity might be regulated through nitrosylation; however, its precise role needs to be elucidated. This assumption is also supported by previous reports suggesting various ecological or commercial roles of terpenes obtained after additional modification.[40-42] It has also been reported to be used for chemical identification and classification of TPSs.[43]

All protein functions are dependent on their structures. Structure analysis and model development help in the prediction of its folding and its secondary and tertiary structure from its primary structure. The secondary structures of both the genes were obtained at 1.95-A° resolution [Table 4]. These results revealed that the structure of both the proteins comprised dominating features of alpha helices, followed by random coils and extended strands. The secondary structure results of sesqui-TPS reported earlier also predicted that the percentage distribution of alpha helix and random coils is analogously higher. Three-dimensional structures have been predicted for TPSs, but such assembled data for *A. annua* L. TPs are unavailable. The experimental structures for these proteins are inadequate. The structures of both the TPSs are based on principle template crystal structure of limonene synthase (Mentha spicata) deposited in PDB. BCS model has a C-score of −0.35 and BFS model −0.21. The C-score is a value of standard of predicted model and its value ranges from −5 to 2. A higher C-score signifies a high quality model and C-score >−1.5 has a correct fold.[44] The C-score for both the proteins obtained indicates a good quality structure and correct folding with an estimated TM-score of 0.67 ± 0.13 and 0.69 ± 0.12 for BCS and BFS, respectively. A TM-score of >0.5 indicates a model of correct topology. Thus, the TM-score of both the proteins indicate a model of correct topology. Analysis of stereochemical quality and accuracy of refined protein model using PROCHECK[24] revealed that dihedral angles of all the residues were located in the most favored region of the Ramachandran Plot [Table 5]. In both the modeled proteins (BCS and BFS), 96.3% and 96.1 of residues were observed in the favored regions, whereas 2.4% and 3.0% residues were observed in the allowed regions, respectively, as compared to limonene synthase (Mentha spicata). Occurrence of 90% or more than 90% residues in the most favored region of Ramachandran plot classifies the refined models to be of good quality and within the values statistically expected for proteins with a resolution of at least 2.0 Angstroms and R-factor no greater than 20. These structures provide a basis for understanding the stereochemical selectivity displayed by the TPSs and provide templates for the prediction of other TPS structures.

A differential expression pattern was observed by these sesqui-TPS at different developmental stages. The RT PCR data revealed clearly that the expression levels of bcs and bfs enzymes were highest in the leaves of artemisia at preflowering and flowering stage. This was followed by the vegetative and postflowering stage. This strongly indicates that these two enzymes are developmentally controlled and may have the regulatory effects on the terpene biosynthesis. This assumption is supported by previous reports showing sesquiterpenes were expressed higher in younger leaves compared to older leaves. Varying expression of different terpene synthase enzymes at different developmental stages and source of origin is also well documented. These differences have also been correlated with the chemotypic variation.[45,46]

CONCLUSION

A thorough analysis of these two putative genes involved in terpene biosynthesis showed several interesting aspects related to structure and brought novel information related to their structures and substrate binding. Domain analysis determined the conserved motifs of both the proteins and these conserved domains were found to be involved in active site stabilization. Structural analysis revealed the important structural aspects of both the proteins. The structural characterization of these genes would pave the way to essential insights concerning terpene biosynthesis and regulation in the production of artemisinin in *A. annua* L. The gene expression patterns also strongly indicate that these two enzymes are developmentally controlled and may have the regulatory effects on the terpene biosynthesis.

Financial support and sponsorship

This work was financially supported by the University Grant Commission (UGC), New Delhi under SAP-DRS II Program, UGC-BSR fellowship and the facilities provided by the department of Biotechnology, Jamia Hamdard, New Delhi-110062. The funding agencies have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Klayman DL. Qinghaosu (artemisinin): An antimalarial drug from China. Science 1985;228:1049-55.
2. McGovern PE, Christofidou-Solomidou M, Wang W, Dukes F, Davidson T, El-Deiry WS, *et al.* Anticancer activity of botanical compounds in ancient fermented beverages (review). Int J Oncol 2010;37:5-14.
3. Crespo-Ortiz MP, Wei MQ. Antitumor activity of artemisinin and its derivatives: From a well-known antimalarial agent to a potential anticancer drug. J Biomed Biotechnol 2012;2012:247597.
4. Abdin MZ, Israr M, Rehman RU, Jain SK. Artemisinin, a novel antimalarial drug: Biochemical and molecular approaches for enhanced production. Planta Med 2003;69:289-99.
5. Roth RJ, Acton N. Isolation of arteannuic acid from *Artemisia annua*. Planta Med 1987;53:501-2.
6. Wen W, Yu R. Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective. Pharmacogn Rev 2011;5:189-94.
7. Bouwmeester HJ, Wallaart TE, Janssen MH, van Loo B, Jansen BJ, Posthumus MA, *et al.* Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. Phytochemistry 1999;52:843-54.
8. Liu B, Wang H, Du Z, Li G, Ye H. Metabolic engineering of artemisinin biosynthesis in *Artemisia annua* L. Plant Cell Rep 2011;30:689-94.
9. Chakrabarti R, Klibanov AM, Friesner RA. Computational prediction of native protein ligand-binding and enzyme active site sequences. Proc Natl Acad Sci U S A 2005;102:10153-8.
10. Chen CC, Hwang JK, Yang JM. (PS) 2: Protein structure prediction server. Nucleic Acids Res 2006;34:W152-7.
11. Chen F, Tholl D, Bohlmann J, Pichersky E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the Kingdom. Plant J 2011;66:212-29.

12. Chang WC, Song H, Liu HW, Liu P. Current development in isoprenoid precursor biosynthesis and regulation. Curr Opin Chem Biol 2013;17:571-9.
13. Brown GD. The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao). Molecules 2010;15:7603-98.
14. Wang H, Ye HC, Liu BY, Li ZQ, Li GF. Advances in molecular regulation of artemisinin biosynthesis. Sheng Wu Gong Cheng Xue Bao 2003;19:646-50.
15. Harborne JB. Role of secondary metabolites in chemical defence mechanisms in plants. Ciba Found Symp 1990;154:126-34.
16. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, *et al.* Protein identification and analysis tools in the exPASy server. Methods Mol Biol 1999;112:531-52.
17. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Eng 1990;4:155-61.
18. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem 1980;88:1895-8.
19. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105-32.
20. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, *et al.* MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res 2009;37:W202-8.
21. Rost B, Yachdav G, Liu J. The predictProtein server. Nucleic Acids Res 2004;32:W321-6.
22. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED protein analysis workbench. Nucleic Acids Res 2013;41:W349-57.
23. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y, *et al.* The I-TASSER suite: Protein structure and function prediction. Nat Methods 2015;12:7-8.
24. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283-91.
25. Alam P, Kiran U, Ahmad MM, Kamaluddin, Khan MA, Jhanwar S, *et al.* Isolation, characterization and structural studies of amorpha-4, 11-diene synthase (ADS (3963)) from *Artemisia annua* L. Bioinformation 2010;4:421-9.
26. Zeng Q, Qiu F, Yuan L. Production of artemisinin by genetically-modified microbes. Biotechnol Lett 2008;30:581-92.
27. Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. FEBS Lett 2006;580:1411-6.
28. Alam P, Abdin MZ. Over-expression of HMG-coA reductase and amorpha-4,11-diene synthase genes in *Artemisia annua* L. And its influence on artemisinin content. Plant Cell Rep 2011;30:1919-28.
29. Lv Z, Zhang F, Pan Q, Fu X, Jiang W, Shen Q, *et al.* Branch pathway blocking in *Artemisia annua* is a useful method for obtaining high yield artemisinin. Plant Cell Physiol 2016;57:588-602.
30. Trapp SC, Croteau RB. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 2001;158:811-32.
31. Falara V, Akhtar TA, Nguyen TT, Spyropoulou EA, Bleeker PM, Schauvinhold I, *et al.* The tomato terpene synthase gene family. Plant Physiol 2011;157:770-89.
32. Aubourg S, Lecharny A, Bohlmann J. Genomic analysis of the terpenoid synthase (atTPS) gene family of *Arabidopsis thaliana*. Mol Genet Genomics 2002;267:730-45.
33. Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, *et al.* Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol 2010;10:226.
34. McAndrew RP, Peralta-Yahya PP, DeGiovanni A, Pereira JH, Hadi MZ, Keasling JD, *et al.* Structure of a three-domain sesquiterpene synthase: A prospective target for advanced biofuels production. Structure 2011;19:1876-84.
35. Davis EM, Croteau R. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. In: Leeper FJ, Vederas JC, editors. Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000. p. 53-95.
36. Crock J, Wildung M, Croteau R. Isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (*Menthaxpiperita*, L.) that produces the aphid alarm pheromone (E)-beta-farnesene. Proc Natl Acad Sci U S A 1997;94:12833-8.
37. Picaud S, Brodelius M, Brodelius PE. Expression, purification and characterization of recombinant (E)-beta-farnesene synthase from *Artemisia annua*. Phytochemistry 2005;66:961-7.
38. Picaud S, Olsson ME, Brodelius M, Brodelius PE. Cloning, expression, purification and characterization of recombinant (+)-germacrene D synthase from *Zingiber officinale*. Arch Biochem Biophys 2006;452:17-28.
39. Van Geldre E, De Pauw I, Inzé D, Van Montagu M, Van den Eeckhout E. Cloning and molecular analysis of two new sesquiterpene cyclases from *Artemisia annua* L. Plant Sci 2000;158:163-71.
40. Singh B, Sharma RA. Plant terpenes: Defense responses, phylogenetic analysis, regulation and clinical applications. 3 Biotech 2015;5:129-51.
41. Miller DJ, Allemann RK. Sesquiterpene synthases: Passive catalysts or active players? Nat Prod Rep 2012;29:60-71.
42. Pichersky E, Gershenzon J. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. Curr Opin Plant Biol 2002;5:237-43.
43. Tilden WA, Shenstone WA. XIX.-Isomeric nitroso-terpenes. J Chem Soc 1877;31:554-61.
44. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725-38.
45. Maes L, Van Nieuwerburgh FC, Zhang Y, Reed DW, Pollier J, Vande Casteele SR, *et al.* Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants. New Phytol 2011;189:176-89.
46. Cai Y, Jia JW, Crock J, Lin ZX, Chen XY, Croteau R, *et al.* A cDNA clone for beta-caryophyllene synthase from *Artemisia annua*. Phytochemistry 2002;61:523-9.