

RESEARCH

Open Access

Characterization of squalene synthase gene from *Gymnema sylvestre* R. Br.



Kuldeepsingh A. Kalariya* , Ram Prasanna Meena, Lipi Poojara, Deepa Shahi and Sandip Patel

Abstract

Background: Squalene synthase (SQS) is a rate-limiting enzyme necessary to produce pentacyclic triterpenes in plants. It is an important enzyme producing squalene molecules required to run steroidal and triterpenoid biosynthesis pathways working in competitive inhibition mode. Reports are available on information pertaining to SQS gene in several plants, but detailed information on SQS gene in *Gymnema sylvestre* R. Br. is not available. *G. sylvestre* is a priceless rare vine of central eco-region known for its medicinally important triterpenoids. Our work aims to characterize the GS-SQS gene in this high-value medicinal plant.

Results: Coding DNA sequences (CDS) with 1245 bp length representing GS-SQS gene predicted from transcriptome data in *G. sylvestre* was used for further characterization. The SWISS protein structure modeled for the GS-SQS amino acid sequence data had MolProbity Score of 1.44 and the Clash Score 3.86. The quality estimates and statistical score of Ramachandran plots analysis indicated that the homology model was reliable. For full-length amplification of the gene, primers designed from flanking regions of CDS encoding GS-SQS were used to get amplification against genomic DNA as template which resulted in approximately 6.2-kb sized single-band product. The sequencing of this product through NGS was carried out generating 2.32 Gb data and 3347 number of scaffolds with N50 value of 457 bp. These scaffolds were compared to identify similarity with other SQS genes as well as the GS-SQSs of the transcriptome. Scaffold_3347 representing the GS-SQS gene harbored two introns of 101 and 164 bp size. Both these intronic regions were validated by primers designed from adjoining outside regions of the introns on the scaffold representing GS-SQS gene. The amplification took place when the template was genomic DNA and failed when the template was cDNA confirmed the presence of two introns in GS-SQS gene in *Gymnema sylvestre* R. Br.

Conclusion: This study shows GS-SQS gene was very closely related to *Coffea arabica* and *Gardenia jasminoides* and this gene harbored two introns of 101 and 164 bp size.

Keywords: *Gymnema sylvestre*, Intron, Long-range PCR, Next-generation sequencing, Squalene synthase

1 Background

Gymnema sylvestre R. Br., traditionally known as *madhunashini*, is one of the important medicinal plants of India. It belongs to the family Asclepiadaceae—milkweed. The economic plant part is the leaf. The phytochemicals present in the leaf causes loss of taste of sweetness. Because of this reason, it is called *madhuna-shini* meaning “killer of sugar.” It grows in the tropical

forests of India and has been used for more than 2000 years in the traditional system of medicines [1]. This plant has also found application in pharmaceuticals. The whole plant is rich in secondary metabolites, which impart medicinal uses to this plant. Ethanolic extract of leaves contains tannins, gum, flavonoids, proteins, and saponins. The principal constituent, gymnemic acid, is found in the aqueous leaf extract of *Gymnema*. It is one of the common plants used in the Indian system of medicine [2]. Various parts of the plant are used in the treatment of different diseases like skin problems,

* Correspondence: Kuldeep.Kalariya@icar.gov.in; Kuldeep_ka@yahoo.co.in
ICAR-Directorate of Medicinal and Aromatic Plants Research, Boriavi, Anand
387310, India

bronchitis, eye disease, cancer, and diabetes. It possesses medicinal properties like digestive, diuretic, emetic, expectorant, laxative, stimulant, and stomachic. It is also used for its antifungal properties and in urinogenital infection [3]. It has anti-diabetic, anti-sweetener, and anti-inflammatory activity [4]. Despite the potential medical importance, little is known about the molecular biology of triterpene biosynthesis in *G. sylvestre*. Recently, biosynthetic pathway of gymnemic acid [5] and polyoxy-pregnane glycoside [6] and putative lncRNA and genes regulating terpenoid biosynthesis pathway [7] along with 13 potential miRNA [8] have been reported in *G. sylvestre*. In all the plants, triterpenes are synthesized via the mevalonate pathway, which involves the sequential conversion of farnesyl diphosphate (FPP) to squalene and then to 2,3-oxidosqualene, followed by a series of cyclization, oxidation, and reduction reactions [9]. Squalene synthase (SQS) and squalene epoxidase (SQE), both rate-limiting enzymes, are necessary to produce pentacyclic triterpenes. There co-exist a positive correlation among level of expression SQS with the quantity of triterpenoids that are produced [10]. SQS is a bifunctional enzyme which is membrane bound and it undergoes condensation of two molecules of C_{15} allylic farnesyl pyrophosphate to form a 30-C precursor which is linear called squalene, which acts like a precursor for both sterol and triterpenoid. The process occurring in two stages involves the first step to be the formation of pre-squalene diphosphate by head-to-head condensation reaction of two FPP molecules. In the second step, there occurs subsequent reduction to squalene in NADPH-dependent manner and this step requires divalent cations [11]. Thus, SQS is an especially important enzyme producing squalene molecules required to run the steroidal and triterpenoid biosynthesis pathways working in competitive inhibition mode [12]. Squalene molecules are precursor to many important secondary metabolites known for their medicinal, chemical, and pharmaceutical values. Squalene synthase is a key enzyme responsible in producing squalene molecules. The secondary products formed from squalene molecules include saponins, triterpenoids, and polyoxy-pregnans in *G. sylvestre*. The triterpenoids include oleanane and dammarene in leaves of *G. sylvestre*. Oleanane saponins include gymnemic acids as well as gymnemasaponins. In case of dammarene, saponins possess gymnemasides [13, 14]. These terpenoids, besides a role in plant defense, are also involved in various clinical properties like anti-viral, anti-tumor, anti-inflammatory, immune activation, and cholesterol lowering. Thus, SQS is a crucial enzyme in regulating triterpenoid biosynthesis [12]. Reports are available on information pertaining to SQS gene in several plants,

namely, *Arabidopsis thaliana*, *Glycine max*, *Magnolia officinalis*, *Panax ginseng*, *Panax notoginseng*, *Salvia miltiorrhiza* Bunge, *Chimonanthus zhejiangensis*, *Trypterygium wilfordii*, and *Taraxacum officinale* [15, 16]. However, information is not available on the SQS gene in *G. sylvestre*. Availability of such information at gene sequence level may provide scope of cloning of this gene and further overexpression of the enzyme squalene synthase intended to harvest higher quantity of phytochemicals whose precursor molecule is squalene. Therefore, attempt was made to characterize the GS-SQS gene in this high-value medicinal plant.

2 Methods

2.1 Experimental material and growth conditions

G. sylvestre accession DGS 22 was originally collected on July 03, 2009, from the Central Region of Eastern Ghats, Vishakhapatnam, Andhra Pradesh, India. It was maintained with standard package of practices in the experimental field of ICAR-Directorate of Medicinal Plants Research, Boriavi, Anand, Gujarat, India. This location is situated at an altitude of 40.63 m above mean sea level having an average rainfall of 800 mm. The plant used in this study was identified from the layout of the germplasm block, available with the curator of *Gymnema* germplasm at ICAR-Directorate of Medicinal Plants Research, Boriavi, Anand, Gujarat, India. A voucher specimen of this germplasm (Fig. 1) is available in the *Gymnema* breeding block properly maintained at ICAR-Directorate of Medicinal Plants Research, Boriavi, Anand, Gujarat, India. For transcriptome study, leaf (L2) and flower (F1) samples of this genotype were collected during the last week of November, while developing fruits (F2) were collected during the second week of December in the year 2016.

2.2 Transcriptome profiling and validation of SQS

The entire process of transcriptome analysis was followed as per recent transcriptome study in *Gymnema sylvestre* [3]. The consensus CDS representing GS-SQS predicted from this transcriptome analysis of leaf, flower, and fruit was selected for further characterization.

2.3 Prediction of protein model and phylogenetic analysis

Protein model was generated through the SWISS-MODEL online module and MolProbity score, QMEAN and $C\beta$, and Ramachandran plot analysis were recorded [17]. GS-SQS protein sequences were obtained from published reports. After sequences were aligned and configured for the highest accuracy, phylogenetic trees were constructed by PHILP method [18]. The bootstrapping method was used to assess the reliability of internal branches.



Fig. 1 a *Gymnema sylvestre* R. Br. breeding block and b voucher specimen of the germplasm

2.4 Validation of GS-SQS gene

To ascertain information at sequence level in GS-SQS gene at genomic DNA level, primers were designed from flanking regions of the CDSs sequences representing GS-SQS gene (Table 1). The amplification was carried out through a long-range PCR master mix (Biolabs, Inc.) using genomic DNA as template. Genomic DNA was extracted from fresh leaves of DGS 22 genotypes of *G. sylvestre*. A total of 0.1 g leaf material was used for DNA was isolated using DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. The DNA concentration was estimated by 0.8% agarose gel electrophoresis using DNA standard. Quantification of DNA was done by Nanodrop-2000 spectrophotometer.

2.5 Long-range PCR to amplify full-length SQS gene

For PCR amplification, 10- μ l reaction mixtures containing 20 ng of template DNA, long-range PCR master mix (Biolabs, Inc.), and 0.25 μ M of each primer was used. Thermal cycling was carried out on Bio-Rad make Thermal Cycler. The PCR steps used were a pre-denaturing (95 °C for 5 min) followed by denaturing (95 °C for 30 s), annealing (55-60 °C for 45 s), extension (72 °C for 45 s) for 35 cycles, and a final extension at 72 °C for 10 min.

Amplified PCR products were initially visualized to check the product size was confirmed on 0.8% agarose gel. The amplified product of GS-SQS gene gave a single band which was compared with long-range DNA ladder, and the approximate size of the product was guessed to be 6.2 kb. This PCR product was purified using 1X Agencourt AMPure XP (Beckman Coulter Genomics: A63882) DNA beads to remove dimers and enzymes of PCR reactions. Purified PCR product was analyzed on 0.8% agarose gel (loaded 3 μ l) for the single intact band. The voltage and time to run gel were 110 V and 30 min, respectively. One microliter of the sample was used for determining concentration using a Qubit® 2.0 fluorometer.

2.6 Sequencing, library, and de novo assembly preparation

The paired-end sequencing library was prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina. The PCR product sample was mechanically sheared into smaller fragments by Covaris followed by a continuous step of end-repair where an 'A' is added to the 3' ends making the DNA fragments ready for adapter ligation. Both ends of the DNA fragments were ligated with adapters suitable to the Illumina platform. With a high-

Table 1 Details of primers (5'→ 3') used in the study

Designed from	Name of the primer	Sequences	Product size	GC%	Tm	Corresponding lane on gel
GS-SQS CDS 10191	GS-SQSLR	F ATGGGGAGCTTGGGAACCAT	6.2 k	55	53.8	Fig. 4, lane 2
	GS-SQSLR	R TCACATATCAATTGGTCGATTGGC		42	54.0	
GS-SQS scaffold_3347	GS-SQSINT1	F ATGGGTTTGTCTACAG	144 bp	39	49.3	Fig. 7, lane 2
	GS-SQSINT1	R GTCTCCAGATA ATCCCTAATGATG		40	62.0	
	GS-SQSINT2	F GGAGCAAATATGTCAATAAACTTGAGG	217 bp	37	63.7	Fig. 7, lane 6
	GS-SQSINT2	R GACAGAATTCTCCTCGTATTTAAGT		35	61.7	
GS-SQSINT1	GS-SQSINT1	F ATGGGTTTGTCTACAG	415 bp	39	49.3	Fig. 7, lane 10
	GS-SQSINT2	R GACAGAATTCTCCTCGTATTTAAGT				
Housekeeping gene	EFTU F	F AGGTCCACCAACCTTGACTG	200 bp	55	60.00	Fig. 7, lane 14, 16
	EFTU R	R GGAGCAAAGGTCACAACCAT		50	60.00	

fidelity amplification step employing HiFi PCR Master Mix, maximum yields were ensured from initially limited product. Bioanalyzer 2100 (Agilent Technologies) was used to analyze an amplified library employing High Sensitivity (HS) DNA chip. After obtaining the Qubit concentration for the library and the mean peak size from the Bioanalyzer profile, the library was loaded onto the Illumina platform for cluster generation and paired-end sequencing.

High-quality paired-end data was assembled with CLC genomics workbench-6 with reads map back option restricting minimum contig length of 200 bp and mismatch cost, insertion cost, and deletion cost as 2, 3, and 3, respectively. The length fraction and similarity fraction were 0.5 and 0.8, respectively. Further gap closure was used for filling the gaps existing in CLC assembly which resulted in improvement of assembly.

2.7 Identification of scaffold representing squalene synthase and confirmation

To identify the scaffolds representing the GS-SQS gene, a similarity search was carried out for scaffolds representing SQS against NCBI's non-redundant (NR) protein database using blastX algorithm. For the identification of intronic gaps, the scaffold representing the GS-SQS gene was aligned with the CDS representing GS-SQS. For the confirmation and validation of the introns, two sets of specific primers are designed as per details in Table 1. Genomic DNA and cDNA were used as templates to amplify the introns. Total RNA was isolated using a total RNA purification kit (Sigma-Aldrich) from the leaves. cDNA was synthesized by using a first-strand c-DNA synthesis kit (Thermo scientific). The first set of primers was designed to amplify the first intron with a product size of 152 bp. Likewise, the second intron was amplified with the corresponding product size of 217 bp. Both the intron together was also amplified with the corresponding products of 415 bp. Thermal cycling was carried out on Bio-Rad make thermal cycler. The PCR steps used

were a pre-denaturing (95 °C for 5 min) followed by denaturing (95 °C for 20 s), annealing (55 °C for 30 s), extension (72 °C for 40 s) for 35 cycles, and a final extension at 72 °C for 10 min. Amplified PCR products were initially visualized to check the product size was confirmed on 0.8% agarose gel.

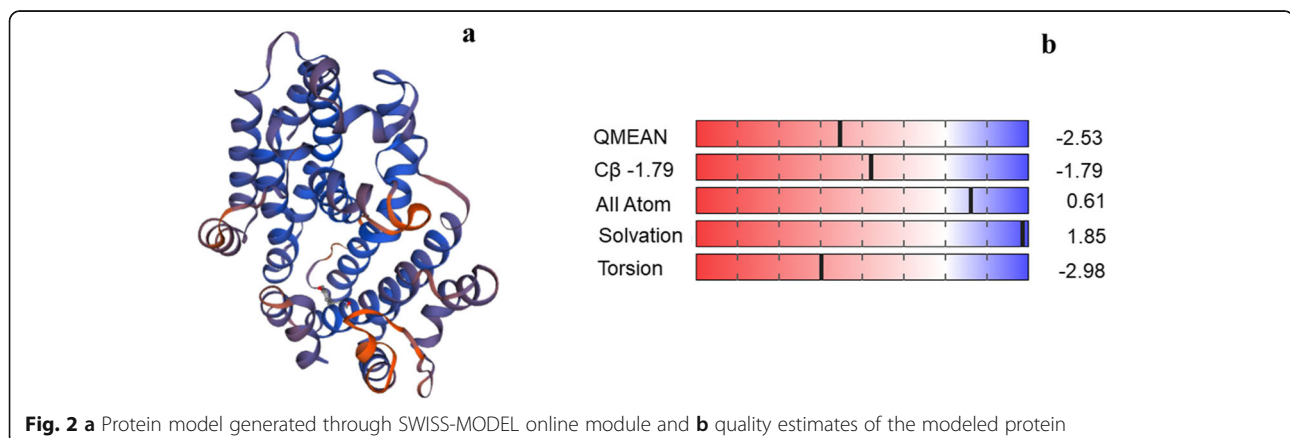
3 Results

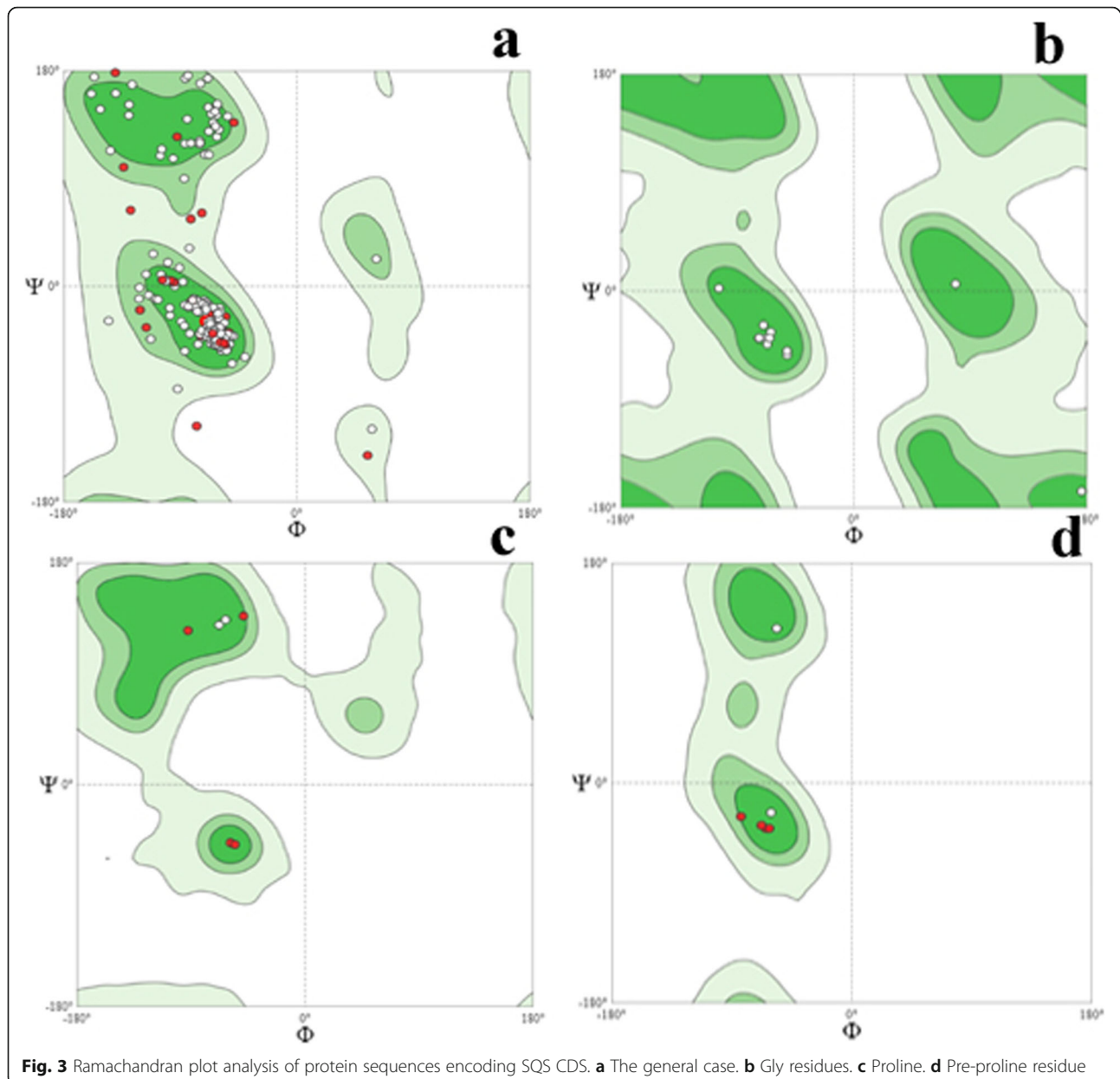
3.1 Protein model

Total three CDS representing GS-SQS predicted from this transcriptome analysis of leaf (L2), flower (F1), and fruit (F2) were deposited to the NCBI GeneBank with the CDS names as GS-SQSCDS_10191-L2 (leaf sample), GS-SQSCDS_64527-F1 (flower sample), and GS-SQSCDS_35012-F2 (fruit sample) with GeneBank accession numbers as MT812194, MT812195, and MT812196, respectively. All these CDSs had the same sequences of 1245 bp length including start and stop codons. The consensus CDS encoding protein sequence of the GS-SQS was used to predict the protein model through SWISS-MODEL online window (Fig. 2). The MolProbity is a structure-validation web service that provides broad-spectrum solidly based evaluation of model quality at both the global and local levels for both proteins and nucleic acids. MolProbity Score for the generated model was 1.44 and the Clash Score 3.86. The quality estimates revealed that QMEAN and C β value were -2.53 and -1.79, respectively. Statistical score of Ramachandran plots analysis (Fig. 3) showed that the percentage of residue within the most favored ϕ , ψ regions were 96.27%, whereas the Ramachandran outliers were 0.31% which indicate that the homology model is reliable.

3.2 Comparison of GS-SQS with SQS sequences from other sources

Innovations in molecular biology and protein sequencing techniques have enabled to characterize the proteome of numerous organisms. Similarly, computational biology methods are also being used routinely to analyze protein





sequences and structures in detail at the molecular level [19–22]. In the present study, an attempt has been made to investigate sequence similarities between *Gymnema sylvestris* and other plant species using bioinformatics methods. Phylogenetic studies of the protein sequences of *Gymnema sylvestris* have provided valuable evidence about their taxonomy, protein makeup, plant systematics, DNA barcoding, and common ancestor. For SQS proteins, the phylogenetic trees were constructed, and their reliability was checked by assessing the bootstrap values. We display the Maximum Likelihood (ML) method consensus trees in Fig. 4. The phylogenetic analysis result of SQ Sgene shows common similarity to *Coffea arabica* and *Gardenia jasminoides*.

3.3 Quality check and quantification of PCR product

Quality of the PCR purified product was checked on 0.8% agarose gel. The size of the product was approximately 6.2 kb (Fig. 5). Quantification of the sample was done using a Qubit fluorometer. Concentration of the product was 96.6 ng/μl.

3.4 NGS data generation and de novo assembly preparation

The library was prepared from the culture sample by NEBNext Ultra DNA Library Prep Kit. The average size of the library is 355 bp. The library was sequenced on the Illumina platform (2 × 150 bp chemistry) and 2.32 GB data was generated. Generated raw data is available

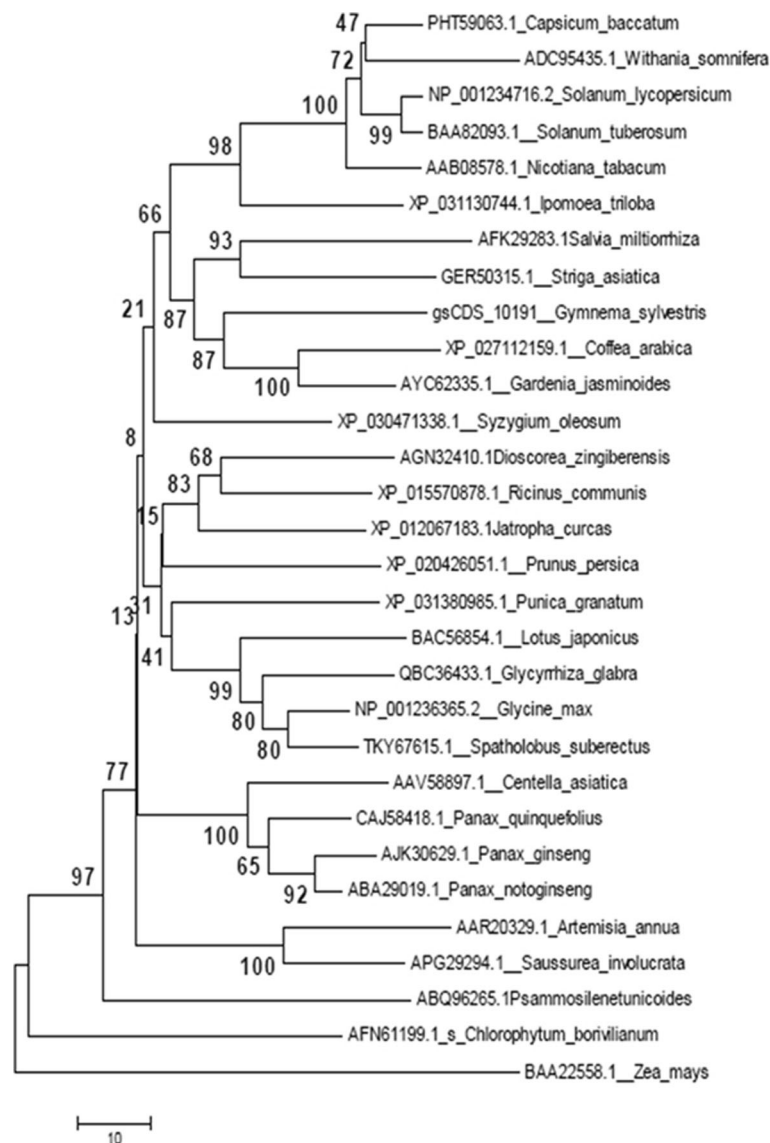


Fig. 4 Phylogenetic analysis SQS gene based on Maximum Likelihood (ML) method

at NCBI data repository with SRA accession number SRR10829862 under project ID PRJNA599051. The statistics of data and assembly elements derived using in-house Perl script is provided in Table 2. The number of scaffolds identified through NGS was 3347. The mean scaffold size and N50 value of the elements were 461 and 497, respectively. The maximum size of the scaffold was 23,583 bp and the total scaffold size in terms of bp was 1,543,977 bp. The scaffold identified representing SQS gene was searched for similarity at domain level using the CDD database. During the blast search, the *E* value threshold was fixed at 0.01, and the maximum number of hits was 500. The results along with interval and *E* value are given in Fig. 6. The sequences of scaffold_3347 representing the GS-SQS

gene is deposited at the NCBI gene repository with submission ID as BlankIt2375151 with accession MT892813.

3.5 Alignment of scaffold and CDS representing GS-SQS and analysis of the data

After aligning the scaffold and the CDS representing GS-SQS, the UTR sequences on both the sites of the scaffold were eliminated, and the remaining sequences of the scaffold were compared and analyzed with the CDS encoding GS-SQS. The alignment and analysis showed that there were a total of 06 sites (four with one and two with two nucleotide) mismatches on the scaffold side which seems to be the erroneous reading during the sequencing. At the same time, it is also revealed

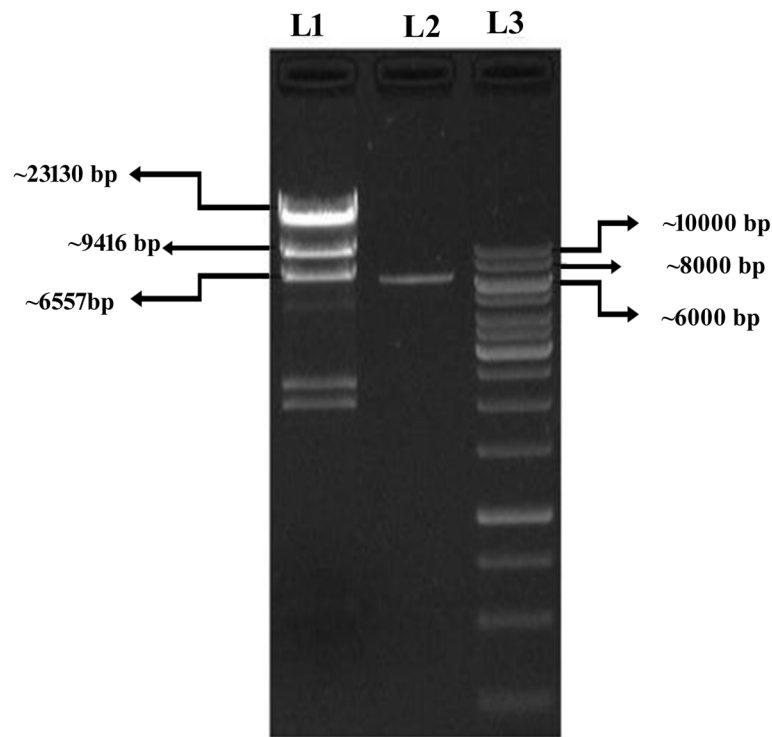


Fig. 5 Long-range PCR product quality check on 0.8% agarose gel. Lane 1, Lamda DNA/HindIII DNA ladder; lane 2, long-range PCR product; and lane 3, 1 kb DNA ladder

that the CDS encoding GS-SQS had two predicted intronic regions bigger than 100 bp size. The first intron was with 101 bp length (657 to 758 bp of CDS) and the second with 164 bp size (864 to 1028 bp of CDS) (Fig. 7). The amplification occurred for the genomic DNA as template and not for the cDNA as template, which confirmed the presence of two introns of 121 and 171 bp size, respectively in the GS-SQS gene in *Gymnema sylvestre* R. Br (Fig. 8).

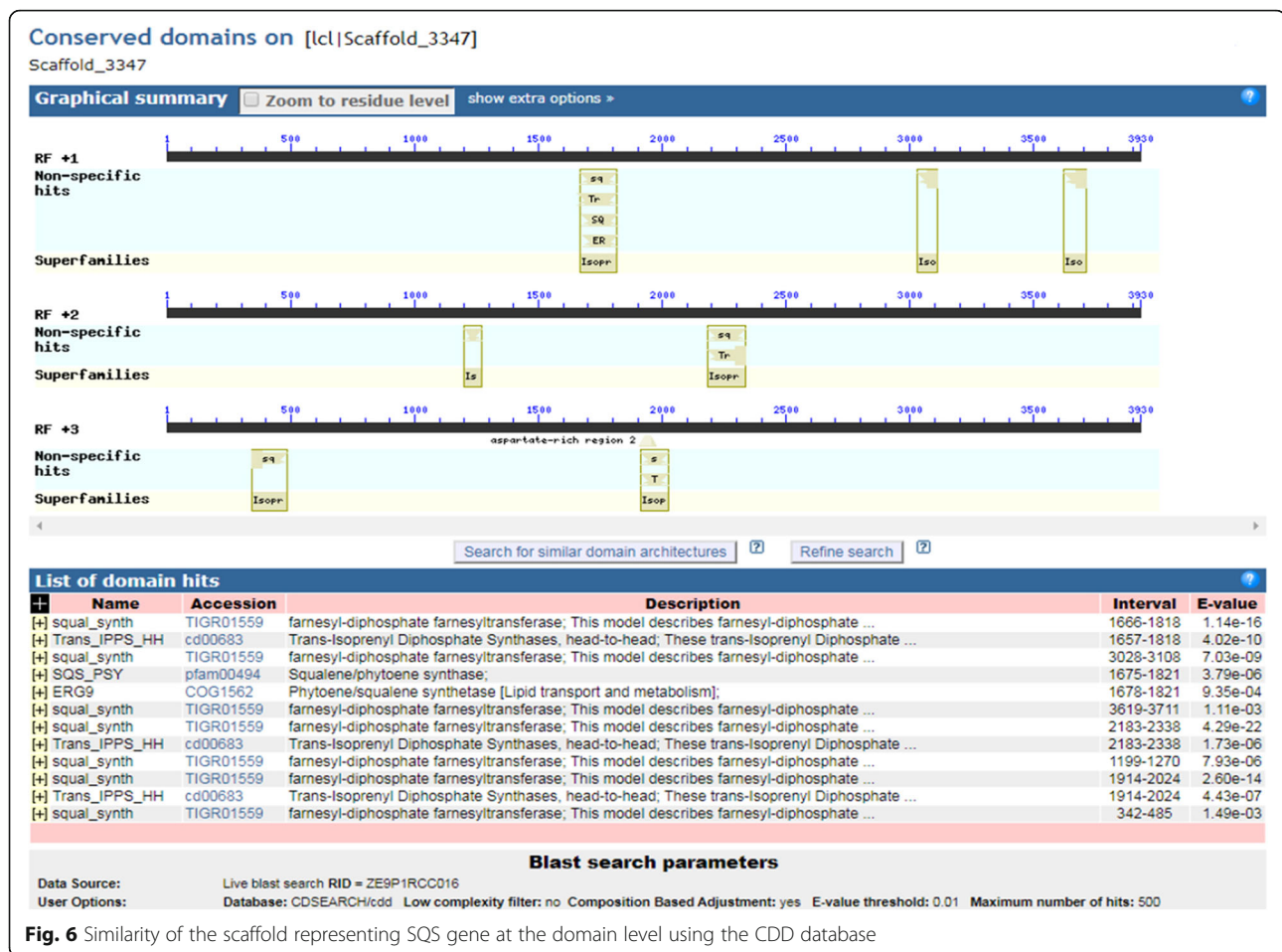
Table 2 Data statistics of NGS of PCR product and de novo assembly

Data statistics of NGS sequences	
No. of reads	7,853,888
Total reads (R1 + R2)	15,707,776
Total bases (R1 + R2)	2,328,437,394
Total data (Gb)	2.32
Assembly elements	
No. of scaffolds	3347
Total scaffold size (bp)	1543977
Max scaffold size (bp)	23583
Mean scaffold size (bp)	461
N50 value (bp)	497

4 Discussion

From the result, it was observed that scaffold_3347 of length 3926 bp was showing similarity against *Panax ginseng*, AJK30629.1 of 444 amino acid length. Simultaneously, blastN of de novo assembled scaffolds was carried out against NCBI's non-redundant nucleotide (NT) database. It was noted that scaffold_3346 of length 1244 bp was showing similarity against *Olea europaea* var. *sylvestris* squalene synthase-like (LOC111412627), mRNA of 1204 bp length. The CLC gap closed de novo assembly was searched for similarity against the transcriptome data. From 3347 scaffolds, there were a total of three scaffolds found significant alignments against the CDS representation GS-SQS gene. Scaffold_3347 and scaffold_3346 had 281 and 3e-76, 171 and 6e-43-bit scores, and *E* value respectively. Although the scaffold did not cover complete CDS, the middle portion of the CDS was covered by scaffold_3347.

Domain search revealed that sequences of scaffold_3347 representing GS-SQS were having farnesyl-diphosphate farnesyltransferase domain. There were a total of twelve conserved sites representing five different genes. The maximum number of conserved sites was covered under the farnesyl-diphosphate farnesyltransferase gene which was represented by accession TIGR01559. This family is related to phytoene synthases. The C-terminal predicted transmembrane



region is absent in archaeal homologs, not included in this model [23]. The scaffold had three conserved sites for the gene Trans-Isoprenyl Diphosphate Synthases (Trans_IPPS). It was represented by accession cd00683. The head-to-head (HH) (1'-1) condensation is carried out by Trans_IPPS. This conserved domain encompasses two genes, viz., squalene synthases and phytoene synthases [23]. These residues mediate binding of prenyl phosphates. The enzymatic process of squalene production is a two-step reaction. A stable intermediate, cyclopropylcarbinyl diphosphate, is formed by squalene synthase with the help of two molecules of FPP. The squalene molecules are produced from this intermediate product by biochemical processes like heterolysis, isomerization, and reduction with NADPH. Therefore, it is a two-step reaction. Phytoene, a precursor of beta-carotene is produced by phytoene synthase (CrtB) causing condensation of two molecules of geranylgeranyl diphosphate. These enzymes, having a wide spectrum presence across eukaryote, bacteria, and archaea, are responsible for biosynthesis of many triterpene and tetraterpene precursors. Chain of these enzymes

produce the triterpene and tetraterpenes in plants. Triterpenoid alkaloids and steroids are further produced from these triterpenes and tetraterpene. Another two conserved sites belong to squalene/phytoene synthase represented by pfam00494 and a phytoene/squalene synthetase represented by ERG9 COG1562 each.

After aligning the scaffold and the CDS representing GS-SQS, the analysis of the nucleotide composition of the predicted introns revealed that A+T content was more than 63%. Both the introns had AG dinucleotides sequences at 3' splice site to facilitate the second step of the splicing event. The initial nucleotide sequences of the intron GT in the first intron whereas it was TT in case of the second intron. Thus, both the introns had common conserved branch point fitting the requirement of the spliceosome to act upon. To amplify the intronic region, primers designed from flanking regions of the introns on the scaffold encoding GS-SQS. The house-keeping gene EFTU was successfully amplified with the use of cDNA as well as the genomic DNA as template. However, the amplification of intronic region took place with the primers designed from adjoining

during evolution, they are important in genomic studies [24, 28]. They may function in the cells like regulation of gene expression and the increase of protein diversity by alternative splicing [25, 29]. Sequences of the whole intronic region are not conserved, and therefore, accumulation of mutations in such region becomes easier [30]. A wide variation in size of introns is reported. It may be as longer than dozens of kilobase pairs (kbp)—to as shorter than 10 bp. In *Arabidopsis*, as revealed in Arabidopsis Genome Initiative 2000, the majority of introns are small with size of a few hundred bp. The smallest exon in *Arabidopsis* was found to be 1 bp [31].

Chlorophytum borivilium, *Euphorbia tirucalli*, *Euphorbia pekinensis*, *Lotus japonicus*, *Oryza sativa*, and *Taxus cuspidate* are the plants in which single SQS genes exist. Two paralogs exist in case of *Arabidopsis thaliana*, *Glycyrrhiza glabra*, *Glycine max*, *Malus domestica*, *Nicotiana tabacum*, *Salvia miltiorrhiza* Bunge, and *Withania somnifera*. There are two SQSs, SQS1 and SQS2, reported in *A. thaliana*. The SQS 1 was found to be broadly expressed in every tissues that are involved in the development of plant whereas the SQS2 was profoundly expressed in hypocotyl of seedlings as well as vascular tissue of cotyledon and leaf petiole. Squalene was not synthesized from recombinant SQS2 from FPP even in the presence of NADPH and Mg^{2+} or Mn^{2+} , whereas in the presence of SQS1, under the same conditions and equivalent preparation, it was able to generate SQ; hence, we can say SQS1 is the ultimate functional SQS present in *Arabidopsis thaliana*. Three SQS paralogs exist in case of *Panax ginseng* [32, 33]. Three SQS genes found in *P. ginseng*, SS1, SS2, and SS3, were found to be capable of converting yeast *erg9* mutant to ergosterol prototrophy despite the divergence in sequence yeast, and similarly, in the case of *Glycine max*, which possesses two SQS, GmSQS1 and GmSQS2 were capable of converting yeast sterol auxotrophy *erg9* mutant to sterol prototrophy. The product sterols were also found to be raised in *Arabidopsis* seed, due to overexpression of *Glycine max* GmSQS1. A similar observation was found in *W. somnifera* SQS that possesses 2 SQS, WsSQS1 and WsSQS2, in which cDNA investigation was performed, and finally, preliminary enzyme activity as well as recombinant expression was reported [16, 33]. It is also noted that the accumulated phytosterol and triterpenoid compounds in *Bupleurum falcatum*, *Eleutherococcus senticosus*, *Panax ginseng*, *Solanum cha-coense*, and *Withania somnifera* were elevated with the overexpression of SQS genes [16].

5 Conclusion

G. sylvestre is one of the most important medicinal plants producing triterpenes. CDS encoding GS-SQS gene with 1245 bp length was predicted from transcriptome data and was analyzed for reliability for quality

scores through modeling SWISS protein structure. Long-range PCR performed with primers designed from flanking regions of this CDS successfully amplified 6.2-kb-sized product against genomic DNA as template. Total 2.32 Gb data and 3347 number of scaffolds with N50 value of 457 bp were generated from NGS of the purified PCR product. The alignment of scaffold_3347 representing GS-SQS gene revealed that it harbors two introns of 101 and 164 bp size in GS-SQS gene. Primers designed from adjoining outside regions of the introns on the scaffold gave amplification when the template was genomic DNA and failed when the template was cDNA which confirms the presence of these two introns in GS-SQS gene in *Gymnema sylvestre* R. Br.

Abbreviations

CDD: Conserved Domain Database; cDNA: Complementary DNA; CDS: Coding DNA sequences; DNA: Deoxyribonucleic acid; EFTU: Elongation factor thermo unstable; FPP: Farnesyl pyrophosphate; Gb: Gigabyte; GS-SQS: *Gymnema sylvestre* squalene synthase; Kbp: Kilobase pair; Mg: Magnesium; Mn: Manganese; mRNA: Messenger RNA; NADP H: Nicotinamide adenine dinucleotide phosphate; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; SQE: Squalene epoxidase; SQS: Squalene synthase

Acknowledgements

The authors are grateful to the Indian Council of Agricultural Research, New Delhi, and the Director, ICAR-DMAPR, Anand, for providing all basic facilities and the Gujarat State Biotechnology Mission, Government of Gujarat, India, for funding assistance. We are also thankful to the germplasm collector and curator of the germplasm used in this study.

Authors' contributions

KAK conceived the project, designed the experiment, analyzed the data, and drafted the manuscript RPM designed primers and analyzed the data. LP standardized PCR reactions, analyzed the data, and help in drafting the manuscript DS standardized PCR reaction conditions. SP standardized PCR reaction conditions. All authors read and approved the final manuscript.

Funding

This research was financially supported by the FAP Scheme, Gujarat State Biotechnology Mission (GSBTM), Government of Gujarat, Gandhinagar, Gujarat, India. The funds were utilized for purchase of consumables including chemicals and glassware and outsourcing services for primer synthesis and sequencing including bioinformatics analysis required in this study.

Availability of data and materials

The data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The transcriptome raw data are deposited at NCBI with accession number SRR5965323 (leaf), SRR5965320 (flower), and SRR5965321 (fruit) and amplicon sequencing (SRR10829862), and the CDS sequences are available with submission ID BankIt 2368826 with accession numbers MT812194, MT812195, and MT812196. The sequences of scaffold_3347 representing the GS-SQS gene is available with submission IDBankIt 2375151 with accession MT892813.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that there is no competing interest.

Received: 14 August 2020 Accepted: 21 December 2020

Published online: 15 January 2021

References

- Gupta P, Sujata Ganguly S, Pratibha Singh P (2012) A miracle fruit plant—*Gymnema sylvestris* R. Br. (Retz) Pharmacoglobale. Int J Comprehens Pharm 03(12):1–12
- Pragya T, Mishra BN, Neelam S (2014) Phytochemical and pharmacological properties of *Gymnema sylvestris*: an important medicinal plant. Biomed Res Int 2014:83028518 pages. <https://doi.org/10.1155/2014/830285>
- Venkatakishore T, Rao MP, Thulasi B, Durga PK, Harikrishna B, Prasanth SS (2016) Hepatoprotective activity of ethanolic extract of aerial part of *Gymnema sylvestris* against CCl₄ and paracetamol induced hepatotoxicity in rats. World J Biomed Pharm Sci 5(12):1007–1016
- Thakur GS, Sharma R, Sanodiya BS, Pandey M, Prasad GB, Bisen PS (2012) *Gymnema sylvestris*: an alternative therapeutic agent for management of diabetes. J Appl Pharmaceut Sci 2(12):1–6
- Kalariya KA, Gajbhiye N, Minipara D, Meena RP, Kumar S, Saha A, Trivedi A, Manivel P (2019) Deep sequencing-based de novo transcriptome analysis reveals biosynthesis of gymnemic acid in *Gymnema sylvestris* (Retz) Schult. Eco Gen Genom 13:100047
- Kalariya KA, Minipara DB, Manivel P (2018) De novo transcriptome analysis deciphered polyoxypregnane glycoside biosynthesis pathway in *Gymnema sylvestris*. 3 Biotech 8(9):381
- Ayachit G, Shaikh I, Sharma P, Jani B, Shukla L, Sharma P, Bhairappanavar SB, Joshi C, Das J (2019) De novo transcriptome of *Gymnema sylvestris* identified putative lncRNA and genes regulating terpenoid biosynthesis pathway. Sci Rep 9(1):1–13
- Kalariya KA, Meena RP, Saran PL, Manivel P (2019) Identification of microRNAs from transcriptome data in gurmar (*Gymnema sylvestris*). Hort Env Biotech 60(3):383–397
- Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. Ann Rev Plant Biol 65:225–257
- Zhao M, Liang W, Zhang D, Wang N, Wang C, Pan Y (2007) Cloning and characterization of squalene synthase (SQS) gene from *Ganoderma lucidum*. J Microbiol Biotechnol 17(7):1106
- Kalariya KA, Minipara D (2018) An overview of triterpenoid biosynthesis in plants and structural depiction of gymnemasides and gymnemosides from *Gymnema sylvestris*. J Plant PhysiolPathol 6(1):1–10. <https://doi.org/10.4172/2329-955X.1000174>
- Qian J, Liu Y, Ma C, Chao N, Chen Q, Zhang Y, Luo Y, Cai D, Wu Y (2019) Positive selection of squalene synthase in cucurbitaceae plants. Int J Genom 19:1–15. <https://doi.org/10.1155/2019/5913491>
- Kalariya KA, Gajbhiye N, Minipara D, Meena RP, Kumar S, Saha A, Trivedi A, Manivel P (2019) Deep sequencing-based de novo transcriptome analysis reveals biosynthesis of gymnemic acid in *Gymnema sylvestris* (Retz) Schult. Ecol Genet Genom 13:100047
- Sarker P, Rahman MM, Khan F, Ming LC, Mohamed IN, Zhao C, Rashid MA (2019) Comprehensive review on phytochemicals, pharmacological and clinical potentials of *Gymnema sylvestris*. Front Pharmacol 10:1223
- Zhang B, Liu Y, Chen M, Feng J, Ma Z, Zhang X, Zhu C (2018) Cloning, expression analysis and functional characterization of squalene synthase (sq) from *Tripterygium wilfordii*. Molecules 23(2):269
- Liu G, Fu J (2018) Squalene synthase cloning and functional identification in wintersweet plant (*Chimonanthus zhejiangensis*). Bot Stud 59(1):1–10
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R (2018) SWISS-MODEL: homology modelling of protein structures and complexes. NuclAcids Res 46(W1):W296–W303
- Hugenholtz P, Pace NR (1996) Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotech 14(6):190–197
- Sonawane KD, Barage SH (2015) Structural analysis of membrane-bound hECE-1 dimer using molecular modeling techniques: insights into conformational changes and Aβ 1–42 peptide binding. Amino Acids 47(3):543–559
- Dhanavade MJ, Sonawane KD (2014) Insights into the molecular interactions between aminopeptidase and amyloid beta peptide using molecular modeling techniques. Amino Acids 46(8):1853–1866
- Parulekar RS, Barage SH, Jalkute CB, Dhanavade MJ, Fandilolu PM, Sonawane KD (2013) Homology modeling, molecular docking and DNA binding studies of nucleotide excision repair UvrC protein from *M. tuberculosis*. Protein J 32(6):467–476
- Dhanavade MJ, Parulekar RS, Kamble SA, Sonawane KD (2016) Molecular modeling approach to explore the role of cathepsin B from *Hordeum vulgare* in the degradation of Aβ peptides. Mol Biosyst 12(1):162–168
- CDD (2020) NCBI Domain search portal <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>
- Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7(3):211–221
- Mishra SK, Thakran P (2018) Intron specificity in pre-mRNA splicing. Curr Genet 64(4):777–784
- Harris KA, Breaker RR (2018) Large noncoding RNAs in bacteria. In: Regulating with RNA in bacteria and Archaea, pp 515–526
- Qu G, Piazza CL, Smith D, Belfort M (2018) Group II intron inhibits conjugative relaxase expression in bacteria by mRNA targeting. Elife 7:e34268
- Bulman S, Ridgway HJ, Eady C, Conner AJ (2007) Intron-rich gene structure in the intracellular plant parasite Plasmodiophorabraceae. Protist 158(4):423–433
- Shepelev MV, Tikhonov MV, Kalinichenko SV, Korobko IV (2018) Insertion of multiple artificial introns of universal design into cDNA during minigene construction assures correct transgene splicing. Mol Biol 52(3):430–435
- Frigola J, Sabarinathan R, Mularoni L, Muiños F, Gonzalez-Perez A, López-Bigas N (2017) Reduced mutation rate in exons due to differential mismatch repair. Nat Genet 49(12):1684–1692
- Guo L, Liu CM (2016) A single-nucleotide exon found in Arabidopsis. Sci Rep 5(1). <https://doi.org/10.1038/srep18087>
- Unland K, Pütter KM, Vorwerk K, van Deenen N, Twyman RM, Prüfer D, Schulze Gronover C (2018) Functional characterization of squalene synthase and squalene epoxidase in *Taraxacum koksaghyz*. Plant Direct 2(6):e00063
- Rong Q, Jiang D, Chen Y, Shen Y, Yuan Q, Lin H, Zha L, Zhang Y, Huang L (2016) Molecular cloning and functional analysis of squalene synthase 2 (SQS2) in *Salvia miltiorrhiza* Bunge. Front Plant Sci 7:1274

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)