

Molecular Cloning and Characterization of Triterpenoid Biosynthetic Pathway Gene HMGS in *Centella Asiatica* (Linn.)

Laiq Ur Rahman (✉ l.rahman@cimap.res.in)

Central Institute of Medicinal and Aromatic Plants CSIR

Shama Afroz

Central Institute of Medicinal and Aromatic Plants CSIR

Zafar Iqbal Warsi

Central Institute of Medicinal and Aromatic Plants CSIR

Kahkashan Khatoon

Central Institute of Medicinal and Aromatic Plants CSIR

Neelam S Sangwan

Central Institute of Medicinal and Aromatic Plants CSIR

Feroz Khan

Central Institute of Medicinal and Aromatic Plants CSIR

Research Article

Keywords: Isoprenoids, Centellosides, Transcriptome, Phylogenetic tree, Multiple sequence alignment, MEGA 10.1.6, Complementation test, *Saccharomyces cerevisiae*

Posted Date: January 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1198345/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The flux of isoprenoids and the total accumulation of triterpenoid saponins known as centellosides in *C. asiatica* are controlled by the key genes of the Mevalonate pathway (MVA). These genes were reported to have positive regulation of the pathway in providing isoprenoid moieties. For the study, full-length pathway gene Hydroxymethyl glutaryl-CoA-synthase (CaHMGS; GenBank accession number: MZ997833), was isolated from previously annotated transcriptome data of *Centella asiatica* leaves. HMGS has been successfully cloned and heterologously expressed in bacteria *E. coli* strain DH5 α . The cloned gene has been sequenced and further characterized through in silico studies by different bioinformatics tools. Also, the gene sequences have been submitted in NCBI. *In silico* studies of isolated gene sequence revealed the nature, characteristics of genes. The ORF of HMGS is 1449 bp encoding 482 amino acids. Predicted molecular weight (MW) of HMGS was 48.09 kDa and theoretical pI was 5.97. Blast results and Multiple sequence alignments of the gene showing the similarity with HMGS of other plants of their respective families. The Molecular Evolutionary Genetic Analysis (MEGA) version 10.1.6 was used to construct a phylogenetic tree. Differential tissue-specific expression of different plant parts was also checked. Tissue expression patterns unveiled that the highest expression level of the CaHMGS had been seen in the roots and lowest in the node of the plant. Functional complementation experiment of the CaHMGS in *Saccharomyces cerevisiae* wild strain YSC1021 and haploid strain YSC1021 which lack HMGS protein confirmed that the CaHMGS gene encodes functional CaHMGS that catalyzed the biosynthesis of mevalonate in yeast. Though, some information is available on the pathway and secondary metabolites. However, most of the pathway steps are not characterized functionally. As the genes were not reported in *Centella asiatica* yet. Hence, understanding this biosynthetic pathway gene will further help in the improvement of plants for enhanced secondary metabolites production.

Introduction

Centella asiatica, belongs to the family Apiaceae, is perennial herb. It is native to wetlands of European and different South Asian countries and commonly known as Indian pennywort, Jal Brahmi, Mandukaparni [1]. The plant is a great reservoir of various secondary metabolites including triterpenoid glycosides, saponin glycosides, flavonoids, phytosterols, volatile oils, tannins, sesquiterpenes and monoterpenes etc. [2]. The major glycone and aglycone pentacyclic triterpenes of the plant are asiatic acid, madecassic acid, ursolic acid, asiaticosides, madecassosides etc. [3]. These triterpenoid molecules of the plant are synthesized by MEP and MVA pathways. MVA pathway are localized in the cytosol [4] and contribute towards the biosynthesis of 5C moieties IPP and DMAPP [5, 6] for participation in the terpenoid pathway (Fig. 1). The enzyme hydroxymethylglutaryl-CoA synthase (HMGS) involved in production of 3-hydroxymethylglutaryl-CoA by the condensation of acetyl-CoA to acetoacetyl-CoA [7] which further leads to conversion into MVA by hydroxymethylglutaryl-CoA reductase (HMGR) [8]. Key genes of the pathways like AACT, HMGR, SQS, etc. among others which are reported to have positive regulation of the pathway in providing isoprenoid moieties [9, 10, 11]. Besides, the major four triterpenoid saponins are important and governed by glucosyltransferases and glucosidases. Genes of the

Mevalonate pathway are responsible for controlling the flux of isoprenoids and overall accumulation of triterpenoid saponins in *C. asiatica*. These genes were reported to have positive regulation of the pathway in providing isoprenoid moieties. The understanding of biosynthetic pathway genes is important for further improvement in the plants for secondary metabolites. Although earlier efforts in this direction have generated transcriptomic resource with annotated and mapped sequences of the pathway. HMGS had been reported in different medicinal plants viz. *Arabidopsis thaliana* [12], *Chamaemelum nobile* [13], *Matricaria chamomilla* [14], *Tripterygium wilfordii* [8], *Santalum album* [15], *Ginkgo biloba* [16, 17], *Panax notoginseng* [18], *Taxus media* [19], *Stereum hirsutum* [5], *Ganoderma lucidum* [20], *Solanum tuberosum* [21], *Gossypium hirsutum* [22], *Aconitum balfourii* [23]. HMGS have not been reported in *C. asiatica* till date. It is supposed that overexpression of HMGS may lead to the enhanced production of Centellosides yield in transgenic plants. In our study, *C. asiatica* HMGS (CaHMGS) gene was cloned and transformed into bacteria *E. coli* DH5 α . The plasmid of positive clones were isolated, sequenced [24] and then submitted to NCBI. Further, the gene was characterized *in silico* via different bioinformatics tools. Differential expression analysis by quantitative PCR (qRT-PCR) had been performed [25]. Complementation test using *Saccharomyces cerevisiae* have been performed [15, 17]. Though, some information is available on the pathway and secondary metabolites. However, most of the pathway steps are not characterized functionally. Functional elucidation of this pathway transcript in isoprenoids biosynthesis would reveal the finer control of the pathway that further will encourage genetic engineering of enzymes for increasing the medicinal value of the plant.

Material And Methods

Gene Sequence mining and primer designing

The full-length gene was mined from transcriptome previously sequenced from leaf samples of *C. asiatica* [26]. The Hydroxymethyl glutaryl-CoA-synthase (HMGS) gene sequence of *C. asiatica* was isolated followed by primer designing, RNA isolation, cDNA (complementary DNA) synthesis and further cloning processes. Genes specific primers (listed in Table 4) were synthesized by using OligoAnalyzer™ Tool (www.idtdna.com/pages/tools/oligoanalyzer) [27].

Total RNA isolation

Total RNA extraction was done from different tissues of *C. asiatica* by using TRI Reagent®(Sigma) using 100 mg of tissue samples. Isolated RNA was dissolved in Milli Q water and checked by Nanodrop spectrophotometer ND1000 and 0.8% agarose gel electrophoresis and further RNA was kept at -80° C [28] for applications in further applications.

cDNA synthesis

Revert Aid kit from Thermo Scientific (U.S.A.) was used to make cDNA from the isolated total RNA. Synthesized cDNA were confirmed by Semi quantitative PCR with Actin primers.

Amplification of the pathway gene CaHMGS by PCR

Amplification of the gene was done with gene specific primers (Table 4) taking synthesized cDNA as a template. For this the reagents like PCR master mix (Thermo Scientific), nuclease-free water, forward primers and reverse primers were added with 35X PCR cycles and temperatures as: Initial denaturation: 94°C:5 minutes; final denaturation: 94°C:35 seconds; primer annealing: 53°C:30 seconds, initial extension: 72°C:2:30 minutes; final extension: 72°C:7:00 minutes used for the successful amplification of HMGS gene. To visualize the bands of the gene, 0.8% agarose gel was casted (0.8g /100 ml 1x TAE) using 1x TAE (Tris-acetate-EDTA) as tank buffer. The amplified band of DNA fragments of interest from the gel was sliced/excised out with a scalpel or razor blade and extracted using the Gen Elute™ gene elution kit from Qiagen.

Cloning and selection of positive clones

The product eluted was further cloned in cloning vector pGEM®-T Easy to obtain a ligated product. The ligated product was then transformed into DH5α strain of *E. coli* and plated on Luria Agar medium with IPTG (Isopropyl β-d-1-thiogalactopyranoside) and X-gal (5-Bromo-4-Chloro-3-Indolyl-beta-D-Galactoside) as inducer and ampicillin as selection drug. The white colonies were taken for screening through PCR with the gene-specific primers to screen the putative transformants.

Colony PCR for screening the recombinant colonies and isolation of plasmid DNA from *E. coli* DH5α

Confirmation of the positive clones was done through colony PCR [29]. The conditions for the PCR used were 94°C:7 minutes; 94°C: 30 seconds; 53°C:40 seconds; 72°C: 2 minutes 30 seconds and final extension 72°C:7 minutes. The resulting product of colony PCR was checked on 0.8 % agarose gel. The plasmid was isolated through the alkaline lysis method and treated with RNaseA to remove RNA contamination. Positive clones were sequenced in an automated sequencer Genetic analyzer Model 3130xl using M13 forward and reverse universal primers [30].

Primary sequence analysis and phylogenetic study

Gene sequences were translated using EMBOSS Transeq tool (https://www.ebi.ac.uk/Tools/st/emboss_transeq). Obtained translated sequences were further undergoes *in silico* proteomics studies. First of all, HMGS (GeneBank ID:MZ997833) primary sequence of

the protein was subjected to BLASTp against the non-redundant (NR) database of NCBI for searching the closest homologs of CaHMGS. For multiple sequence alignment, all sequences with significant alignment were aligned by ClustalW tool [24] (<https://www.genome.jp/tools-bin/clustalw>) [25]. Identity cut off >90, E value -0.0, query coverage -100% with the CaHMGS. For the study of Evolutionary relationship of the CaHMGS gene with HMGS of other species, we constructed phylogenetic tree by Neighbour-Joining method having bootstrap value of 1000 replicates [25,28] by the software Molecular Evolutionary Genetic Analysis (MEGA) version 10.1.6. [31]. Various bioinformatics tools were used for protein sequence analysis like protein family and domain arrangement within the gene was explored by Pfam database (<http://pfam.sanger.ac.uk/>) [32], domain and motif was deduced through InterProScan and Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de>) Conserved Domain Database (CDD) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), was used to deduce the Conserved Domains of the protein. Besides, various parameters of physiochemical properties of the amino acid sequence of CaHMGS protein was also observed with the ExPaSy Protparam (<http://web.expasy.org/protparam>) tool of the proteomics server [32].

Theoretical modelling and its validation

The CaHMGS primary protein sequence was subjugated to BLAST against the Protein Data Bank database. The closest homolog found was *Brassica juncea* HMGS (7CQT.1) with >81% identity and >95% coverage was observed with CaHMGS. The sequences were deposited in the Swiss model (swissmodel.expasy.org) for homology modelling. Amongst different models, the best model having maximum favoured score was selected according to Ramachandran Plot. The model selected was then undergoes for refinement by Galaxy Refine server. Also, the obtained refined model was again verified through quantification of the amino acid residues shown in the Ramachandran plot. The stereochemical quality of the CaHMGS model was analyzed through PROCHECK [33,34] in the Structural Analysis and Verification Server (SAVES) (<http://nihserver.mbi.ucla.edu/SAVES/>) [25]. Also, the model has further visualized using UCSF Chimera-Molecular Modeling System Chimera [35]. Modelled protein was exposed to Protein Structure Analysis (ProSA) [36] and ERRAT (SAVES) tools for quality determination and validation of the modelled protein [37].

Tissue-Specific Expression Analysis

Differential tissue specific Expression of CaHMGS gene was studied in various tissues (node, petiole, leaves, and root) of *Centella asiatica* using quantitative real-time PCR. For the study, Total RNA from different tissues like leaves, node, petiole and root were isolated [16] using TRI Reagent® from Sigma. 5µg of total RNA was used to prepare cDNA using Gene Sure First Strand cDNA Synthesis Kit (Puregene) [38]. The PRIMER 3 was used to primers designing for real time PCR using having T_m > 50 °C, GC content of 50–60%, primer length of nucleotides approx.18-24 and amplicon size (expected) of 100–120 bp. RT-PCR specific primer sequences of CaHMGS (listed in Table 4). Concentration of cDNA synthesized was

optimized to 100-150 ng/ μ l. Relative expression of genes were also checked by semi-quantitative PCR. Actin gene was used as endogenous control for normalizing the expression of the target gene. Calculation of Relative expression was done by $2^{-\Delta\Delta Ct}$ method [25]. All the reactions have been done in triplicates to ensure the credibility of the data.

Functional complementation

Yeast (*Saccharomyces cerevisiae*) required MVA pathway for its survival [15] and interrupting MVA pathway genes in yeast could be lethal. pYES2 are autonomously replicating expression vectors that contain a yeast galactose-dependent promoter. Hence to characterized the function of CaHMGS, the vector pYES2-CaHMGS construct was prepared. For this, the cDNA prepared in previous experiment was amplified using primers (listed in Table 4) having restriction sites ScaI and XbaI, eluted and ligated in pYES vector underneath to the yeast galactose-dependent GAL1 promoter [39]. pYES2-CaHMGS construct was then transformed into diploid *Saccharomyces cerevisiae* strain YSC1021 and haploid strain YSC1021 (that lack one HMGS allele). Also, Wild type pYES2 plasmid construct was transformed as a positive control. Positive clones were then sequenced. Both yeast strains were streaked on YEPG (Himedia) and YEPD (Yeast Extract Peptone Dextrose) agar media plates either with or without G418 (Geneticin) and then incubated at 30°C for yeast growth.

Results And Discussion

Molecular cloning of HMGS gene

The cDNA of HMGS gene was successfully cloned and sequenced (Fig. 2; (a)-(d)). The sequence was subjected to Blastx. After confirmation sequence was submitted to NCBI databases [40] with GenBank accession number: MZ997833 (<https://submit.ncbi.nlm.nih.gov/about/bankit/>). The gene was further analyzed by various bioinformatics software.

Bioinformatics analysis of CaHMGS

CaHMGS nucleotide sequences was subjected to translate through Expasy translator (<https://web.expasy.org>) that revealed, it contained a 1449bp and 482 amino acids. Different domains present in the synthase superfamily family were identified by SMART against the Pfam database [41]. The results dictated that CaHMGS belongs to HMG_CoA_synt_C superfamily (Fig. 3). It contains two putative domains viz. Hydroxymethylglutaryl-coenzyme A synthase C terminal (Pfam: HMG_CoA_Synt_C) positioned between 211-465 amino acids and Hydroxymethylglutaryl-coenzyme A synthase N terminal (Pfam: HMG_CoA_Synt_N) comprised 5-165 amino acids (Table 1). The physiochemical studies revealed that CaHMGS is almost neutral in nature with a theoretical pI of 5.97. The instability index and aliphatic index of HMGS was 31.8 and 67.97 respectively. The molecular mass of CaHMGS protein was 48.09 kDa

(Table 2). Also, the derived amino acid sequence was searched for homology through PSI-BLAST and the result showed high sequence homology with other HMGS sequences from different plant species.

Multiple Sequence alignments

Multiple Sequence alignments of *Centella asiatica* Synthase gene HMGS with other plants was analyzed through ClustalW, with fourteen other protein sequence data: *Daucus carota* (XP_017236990), *Panax notoginseng* (AIK21781), *Catharanthus roseus* (AEC13715), *Actinidia chinensis* (PSS04126), *Arabidopsis thaliana* (NP_193291), *Panax ginseng* (ADI80347), *Gossypium hirsutum* (XP_016755596), *Ricinus communis* (XP_002509692), *Trachyspermum ammi* (AUZ98388), *Nicotiana tabacum* (XP_016481886), *Camellia sinensis* (XP_028098645), *Actinidia rufa* (GFZ06907), *Gossypium hirsutum* (XP_016716145), *Camellia sinensis* (AYQ58343) (Fig.4).

Phylogenetic Analysis

The CaHMGS protein was aligned with its other homologues in MEGA v10.1.6 for the study of evolutionary relationship [31,42] (Fig. 4). Taking the top hits having the lowest E-value, the highest query coverage and identity listed above into the consideration, phylogenetic tree was constructed by blastp tool. On the basis of differences in clade, phylogenetic tree of the CaHMGS was divided into cluster I and cluster II (Fig. 5). These findings help to conclude that the ancestral gene that was responsible for the evolution of the CaHMGS was ultimately shifted to other parentage of plant species.

Determination and description of 3D structure of *Centella asiatica* HMGS

Three-Dimensional structure of the CaHMGS (Fig. 6) was constructed on the basis of sequence identity with the *Brassica juncea* HMGS template (7CQT.1) by homology modelling in SWISS-MODEL [43]. The model has shown the similar upper and lower structural region as compared to other HMGS [44]. In addition, interface of the above-mentioned lower and upper region detects the binding pocket of acetoacetyl-CoA. The Upper structure region revealed active site and conserved motif i.e. NxD/NE/VEGI/VDx(2)NACF/YxG lies around the five-layered core structure. The conserved motif was responsible for regulating the catalytic activity of the CaHMGS on the substrate by being localized in the region of active site. Moreover, a catalytic cysteine residue is found at the C-terminus, which was responsible in ramping up HMG-CoA synthesis [45]. These detections will help us in better understanding of the functions of the CaHMGS and also create a pathway for further study towards terpenoids metabolic process.

The molecular modelling and model validation of CaHMGS

The secondary structure of CaHMGS protein was predicted by using PSIPRED (bioinf.cs.ucl.ac.uk/psipred) and UCSF Chimera-Molecular Modeling System Chimera [35] showed that the protein was actually a α -helical protein that contains 17 α -helices and 10 β -Sheets (CaHMGS). The obtained secondary structure of CaHMGS was further confirmed using 3D modeling in Swiss model server. Molecular modelling of CaHMGS showed that N terminal and C-terminal end of the protein hold α -helices in conjunction with β -Sheets and β -Sheets respectively.

The protein model was studied by residue and its stereochemical quality and analyzed in details through different bioinformatics model validation servers [46]. Ramachandran plot analysis of CaHMGS (Fig. 7) showed 90.4% (671) of amino acid residues were found to be in most favorable region. 8.8% (65) of residue were observed under generously favourable regions and 0.4 % (3) was in additionally favorable with only 0.4 % (3) of residues in unfavorable regions. ProSA revealed the overall quality of the modelled structure CaHMGS using Z-score (Fig. 8(a)) and knowledge-based energy graph (Fig. 8(b)). From this observation, it can be concluded that the CaHMGS model had a fine resolution and stereochemistry by comparing with template 7CQT (Table 3).

Differential tissue-specific expression

The results from Quantitative PCR revealed that the expression of the CaHMGS gene was ubiquitous in different tissues of *C. asiatica*, but the level of expression of the gene was found to be highest in the roots and lowest in nodes (Fig. 9; (a) and (b)).

Functional complementation

Wild type strain YSC1020 of yeast was able to grow on both non-induction medium YPD with G418 as well as on induction medium YPG with G418 but the mutant yeast strain harboring pYES2-CaHMGS could survive only on YPG medium but not survived on non-induction media [39] YPD medium either with or without G418 (Fig. 10). The results showed that the functional defect of strain YSC1020 could be complemented by the expression of the CaHMGS which possesses HMGS activity to catalyze the synthesis of mevalonate and can convert the acetoacetyl-CoA into the HMG-CoA.

Conclusions

In our study, HMGS gene had been first time reported in the *Centella asiatica*. The full-length HMGS gene was cloned from *C. asiatica* and heterologously expressed in bacteria DH5 α . Positive clones were sequenced. Confirmed HMGS gene sequences have been submitted to the NCBI. Further, bioinformatics analysis of the CaHMGS have been done. Blast results and Multiple sequence alignments showing the similarity with HMGS of other plants of their respective families. On the basis of Blast results and Multiple sequence alignment, phylogenetic tree has been constructed. The phylogenetic tree showed that the CaHMGS is the rate-limiting enzyme MVA pathway involved in the production of triterpenoids. Gene

was characterized by different bioinformatics tools. Differential tissue-specific expression in *C.asiatica* have been checked. The expression of the gene was highest in roots of the plant and lowest in the nodes. Complementation test showed that the CaHMGS gene encodes functional protein that catalyses the synthesis of Mevalonate in the MVA pathway. Our present study will lead to further understanding the role of gene in the pathway for production of triterpenoid saponins. Also, it can open up different channels for probable genetic engineering avenues that will surely improve the medicinal properties of the *Centella asiatica*.

Declarations

Acknowledgments:

Author is highly acknowledging the financial support by CSIR, New Delhi, India for awarding CSIR-Junior Research Fellowship (P-81). The authors are also thankful to the Director, CSIR-CIMAP for providing necessary facilities.

Funding:

The authors declare that the funds were received from CSIR, New Delhi, India during the experiments and preparation of this manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

Competing Interests:

The authors have no relevant financial or non-financial competing interests to disclose.

Author contributions:

All authors contributed to the study. Material preparation, data collection and analysis original manuscript writing were performed by Shama Afroz. *In silico* data analysis by Kahkashan Khatoon and Shama Afroz. Review and revision of manuscript by Zafar Iqbal Warsi. Manuscript editing done by Feroz Khan. Resources provided by Laiq Ur Rahman and Neelam S Sangwan. Supervision provided by Laiq Ur Rahman. All authors reviewed and approved the final manuscript.

Ethics approval:

Not applicable.

Consent to participate:

Not applicable.

Consent to publish:

Not applicable.

My manuscript have no associated data.

References

1. Joshi K, Chaturvedi P (2013) Therapeutic efficiency of *Centella asiatica* (L.) Urb. An underutilized green leafy vegetable: an overview. *International Journal of Pharma and Bio Sciences* 4(1):135–149
2. Gohil KJ, Patel JA, Gajjar AK (2010) Pharmacological review on *Centella asiatica*: a potential herbal cure-all. *Indian journal of pharmaceutical sciences* 72(5):546
3. Singh J, Sabir F, Sangwan RS, Narnoliya LK, Saxena S, Sangwan NS (2015) Enhanced secondary metabolite production and pathway gene expression by leaf explants-induced direct root morphotypes are regulated by combination of growth regulators and culture conditions in *Centella asiatica* (L.) urban. *Plant Growth Regul* 75(1):55–66
4. Van Deenen N, Bachmann AL, Schmidt T, Schaller H, Sand J, Prüfer D, Gronover CS (2012) Molecular cloning of mevalonate pathway genes from *Taraxacum brevicorniculatum* and functional characterisation of the key enzyme 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Mol Biol Rep* 39(4):4337–4349
5. Flynn CM, Schmidt-Dannert C (2018) Sesquiterpene synthase–3-hydroxy-3-methylglutaryl coenzyme A synthase fusion protein responsible for hirsutene biosynthesis in *Stereum hirsutum*. *Appl Environ Microbiol* 84(11):e00036–e00018
6. Carretero-Paulet L, Ahumada I, Cunillera N, Rodríguez-Concepción M, Ferrer A, Boronat A, Campos N (2002) Expression and molecular analysis of the *Arabidopsis* DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Plant Physiol* 129(4):1581–1591
7. Yoshida R, Yoshimura T, Hemmi H (2020) Reconstruction of the “archaeal” mevalonate pathway from the methanogenic archaeon *Methanosarcina mazei* in *Escherichia coli* cells. *Applied and environmental microbiology*, 86(6), e02889-19
8. Liu YJ, Zhao YJ, Zhang M, Su P, Wang XJ, Zhang XN, Huang LQ (2014) Cloning and characterisation of the gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase in *Tripterygium wilfordii*. *Molecules* 19(12):19696–19707

9. Bansal S, Narnoliya LK, Mishra B, Chandra M, Yadav RK, Sangwan NS (2018) HMG-CoA reductase from Camphor Tulsi (*Ocimum kilimandscharicum*) regulated MVA dependent biosynthesis of diverse terpenoids in homologous and heterologous plant systems. *Sci Rep* 8(1):1–15
10. Chaurasiya ND, Sangwan NS, Sabir F, Misra L, Sangwan RS (2012) Withanolide biosynthesis recruits both mevalonate and DOXP pathways of isoprenogenesis in Ashwagandha *Withania somnifera* L. *Plant Cell Rep* 31(10):1889–1897Dunal
11. Jadaun JS, Sangwan NS, Narnoliya LK, Singh N, Bansal S, Mishra B, Sangwan RS (2017) Over-expression of DXS gene enhances terpenoidal secondary metabolite accumulation in rose-scented *geranium* and *Withania somnifera*: active involvement of plastid isoprenogenic pathway in their biosynthesis. *Physiol Plant* 159(4):381–400
12. Montamat F, Guilloton M, Karst F, Delrot S (1995) Isolation and characterization of a cDNA encoding *Arabidopsis thaliana* 3-hydroxy-3-methylglutaryl-coenzyme A synthase. *Gene* 167(1–2):197–201
13. Cheng S, Wang X, Xu F, Chen Q, Tao T, Lei J, Li X (2016) Cloning, expression profiling and functional analysis of CnHMGS, a gene encoding 3-hydroxy-3-methylglutaryl coenzyme A synthase from *Chamaemelum nobile*. *Molecules* 21(3):316
14. Tao T, Chen Q, Meng X, Yan J, Xu F, Chang J (2016) Molecular cloning, characterization, and functional analysis of a gene encoding 3-hydroxy-3-methylglutaryl-coenzyme A synthase from *Matricaria chamomilla*. *Genes & Genomics* 38(12):1179–1187
15. Niu M, Yan H, Xiong Y, Zhang Y, Zhang X, Li Y, Ma G (2021) Cloning, characterization, and functional analysis of acetyl-CoA C-acetyltransferase and 3-hydroxy-3-methylglutaryl-CoA synthase genes in *Santalum album*. *Sci Rep* 11(1):1–13
16. Meng X, Song Q, Ye J, Wang L, Xu F (2017) Characterization, function, and transcriptional profiling analysis of 3-hydroxy-3-methylglutaryl-CoA synthase gene (GbHMGS1) towards stresses and exogenous hormone treatments in *Ginkgo biloba*. *Molecules* 22(10):1706
17. Meng X, Xu F, Song Q, Ye J, Liao Y, Zhang W (2018) Isolation, characterization and functional analysis of a novel 3-hydroxy-3-methylglutaryl-coenzyme A synthase gene (GbHMGS2) from *Ginkgo biloba*. *Acta Physiol Plant* 40(4):1–14
18. Liu WJ, Lv HZ, He L, Song JY, Sun C, Luo HM, Chen SL (2016) Cloning and Bioinformatic Analysis of HMGS and HMGR Genes from *Panax notoginseng*. *Chinese Herbal Medicines* 8(4):344–351
19. Kai G, Miao Z, Zhang L, Zhao D, Liao Z, Sun X, Tang K (2006) Molecular cloning and expression analyses of a new gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase from *Taxus media*. *Biol Plant* 50(3):359–366
20. Ren A, Ouyang X, Shi L, Jiang AL, Mu DS, Li MJ, Zhao MW (2013) Molecular characterization and expression analysis of GIHMGS, a gene encoding hydroxymethylglutaryl-CoA synthase from *Ganoderma lucidum* (Ling-zhi) in ganoderic acid biosynthesis pathway. *World Journal of Microbiology and Biotechnology* 29(3):523–531
21. Henriquez MA, Soliman A, Li G, Hannoufa A, Ayele BT, Daayf F (2016) Molecular cloning, functional characterization and expression of potato (*Solanum tuberosum*) 1-deoxy-d-xylulose 5-phosphate

- synthase 1 (StDXS1) in response to *Phytophthora infestans*. Plant Sci 243:71–83
22. Liu W, Zhang Z, Zhu W, Ren Z, Jia L, Li W, Ma Z (2019) Evolutionary conservation and divergence of genes encoding 3-hydroxy-3-methylglutaryl coenzyme A synthase in the allotetraploid cotton species *Gossypium hirsutum*. Cells 8(5):412
 23. Sharma E, Pandey S, Gaur AK (2016) Identification and expression analysis of DXS1 gene isolated from *Aconitum balfourii* Stapf. Acta Physiol Plant 38(10):1–9
 24. Wu SJ, Shi M, Wu JY (2009) Cloning and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene for diterpenoid tanshinone biosynthesis in *Salvia miltiorrhiza* (Chinese sage) hairy roots. Biotechnol Appl Chem 52(1):89–95
 25. Sharma R, Devi K, Modi MK, Sen P (2021) *In silico* characterization and differential expression analysis of 1-deoxy-d-xylulose-5-phosphate reductoisomerase (DXR) of *Centella asiatica*. 3 Biotech, 11, 1-9
 26. Shukla YM (2018) *In silico* primer designing of WRKY transcription factors in cumin (*Cuminum cyminum* L.) for wilt resistance from available transcriptome database. Journal of Pharmacognosy and Phytochemistry, 7(6), 404-410
 27. Sangwan RS, Tripathi S, Singh J, Narnoliya LK, Sangwan NS (2013) De novo sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. Gene 525(1):58–76
 28. Xu Y, Liu J, Liang L, Yang X, Zhang Z, Gao Z, Wei J (2014) Molecular cloning and characterization of three cDNAs encoding 1-deoxy-D-xylulose-5-phosphate synthase in *Aquilaria sinensis* (Lour.) Gilg. Plant physiology and biochemistry, 82, 133-141
 29. Chakraborty N, Besra A, Basak J (2020) Molecular cloning of an amino acid permease gene and structural characterization of the protein in common bean (*Phaseolus vulgaris* L.). Mol Biotechnol 62(3):210–217
 30. Kumar PN, Prasad YG, Prabhakar M, Shanker AK, Bhanu D (2017) Molecular and *in Silico* Characterization of *Achaea janata* Granulovirus Granulin Gene. Interdisciplinary Sciences: Computational Life Sciences, 9(4), 528-539
 31. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) Mol Biol Evol 30(12):2725–2729MEGA6: molecular evolutionary genetics analysis version 6.0
 32. Devi K, Patar L, Modi MK, Sen P (2017) An insight into structure, function, and expression analysis of 3-hydroxy-3-methylglutaryl-CoA reductase of *Cymbopogon winterianus*. Bioinform Biol insights 11:1177932217701735
 33. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283–291
 34. Laskowski RA, MacArthur MW, Thornton JM (2006) PROCHECK:validation of protein structure coordinates.. In: Rossmann MG, Arnold E (eds) International tables of crystallography, volume F, crystallography of biological macro-molecules, vol 25. Kluwer, Dordrecht, pp 722–725

35. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612
36. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(suppl2):407–410
37. Bhattacharya M, Hota A, Kar A, Chini DS, Malick RC, Patra BC, Das BK (2018) In silico structural and functional modelling of Antifreeze protein (AFP) sequences of Ocean pout (*Zoarces americanus*, Bloch & Schneider 1801). *Journal of Genetic Engineering and Biotechnology* 16(2):721–730
38. Khan S, Pandey SS, Shanker K, Khan F, Rahman LU (2017) Cloning and functional characterization of quinolinic acid phosphoribosyl transferase (QPT) gene of *Nicotiana tabacum*. *Physiol Plant* 160(3):253–265
39. Tong Y, Su P, Zhao Y, Zhang M, Wang X, Liu Y, Huang L (2015) Molecular cloning and characterization of DXS and DXR genes in the terpenoid biosynthetic pathway of *Tripterygium wilfordii*. *Int J Mol Sci* 16(10):25516–25535
40. Wang Q, Pi Y, Hou R, Jiang K, Huang Z, Hsieh MS, Tang K (2008) Molecular cloning and characterization of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (CaHDR) from *Camptotheca acuminata* and its functional identification in *Escherichia coli*. *BMB reports*, 41(2), 112-118
41. Rosli R, Chan PL, Chan KL, Amiruddin N, Low ETL, Singh R, Murphy DJ (2018) In silico characterization and expression profiling of the diacylglycerol acyltransferase gene family (DGAT1, DGAT2, DGAT3 and WS/DGAT) from oil palm, *Elaeis guineensis*. *Plant Sci* 275:84–96
42. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425
43. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31(13):3381–3385
44. Campobasso N, Patel M, Wilding IE, Kallender H, Rosenberg M, Gwynn MN (2004) *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase: crystal structure and mechanism. *Journal of Biological Chemistry*, 279(43), 44883-44888
45. Theisen MJ, Misra I, Saadat D, Campobasso N, Miziorko HM, Harrison DH (2004) 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in “real-time”. *Proceedings of the National Academy of Sciences*, 101(47), 16442-16447
46. Messaoudi A, Belguith H, Hamida JB (2013) Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β -lactamase. *Theoretical Biology and Medical Modelling* 10(1):1–10

Tables

Table 1

Conserved domain of CaHMGS in *C. asiatica* identified by the Conserved Domains Database.

| Gene | Name of Domain | Interval | E value |
|--------|----------------|----------|----------------------|
| CaHMGS | HMG CoA_Synt_C | 211-465 | 1.1e ⁻¹⁰⁰ |
| | HMG CoA_Synt_N | 5-165 | 3.1e ⁻⁷⁶ |

Table 2

Physicochemical properties of amino acid sequences of CaHMGS

| | |
|------------------------|------------|
| Total Amino acids | 482 |
| Molecular weight (kDa) | 48.09 kDa. |
| Theoretical pI | 5.97 |
| Acidic amino acids | 52 |
| Basic amino acids | 47 |
| Index of Instability | 31.88 |
| Aliphatic Index | 67.97 |
| Average hydropathicity | -0.271 |

Table 3

comparison of Ramachandran Plot Statistics of CaHMGS modelled structure with closely related homolog 7CQT.1

| Statistics of Ramachandran Plot | Template(7CQT.1) | | Modelled Structure of CaHMGS | |
|--|------------------|----------------|------------------------------|----------------|
| | Residues | Percentage (%) | Residues | Percentage (%) |
| Residues in most favoured regions | 1421 | 88.6 | 671 | 90.4 |
| Residues in additionally allowed regions | 166 | 10.4 | 65 | 8.8 |
| Residues in generously allowed regions | 12 | 0.7 | 3 | 0.4 |
| Residues in disallowed regions | 4 | 0.2 | 3 | 0.4 |
| Number of non-glycine and non-proline | 1603 | 100 | 742 | 100 |
| Number of end residues (excluding glycine and proline) | 12 | - | 4 | - |
| Number of glycine residues | 119 | - | 56 | - |
| Number of glycine residues | 63 | - | 32 | - |

Table 4

Primers list used during experiments

| Primer Code | Primer purpose | Primer sequences (5'→3') |
|-------------|----------------------------|---------------------------------|
| SAHMF | Semi Quantitative | ATGGATTCCCAAAAAAATGTTGG |
| SAHMR | Semi Quantitative | TTGACCATTTATCAAGGATCC |
| RTHGF1 | Quantitative Real Time PCR | CTGGCACCGTTTTCAAGTCT |
| RTHGR1 | Quantitative Real Time PCR | GTTGGGTCACCTTCTCAAGATC |
| RTHGF2 | Quantitative Real Time PCR | GCGAGCTCTGTTGATAAGGGTACT |
| RTHGF2 | Quantitative Real Time PCR | CGGCTAGCATAGCTTTCATCTC |
| YS2-HMGF | Functional complementation | CTCGACTCGATGGTCCGTGACCGCAACGA |
| YS2-HMGR | Functional complementation | CGGCTACGATGCCGAATACCGGTTTCAGTAA |

Figures

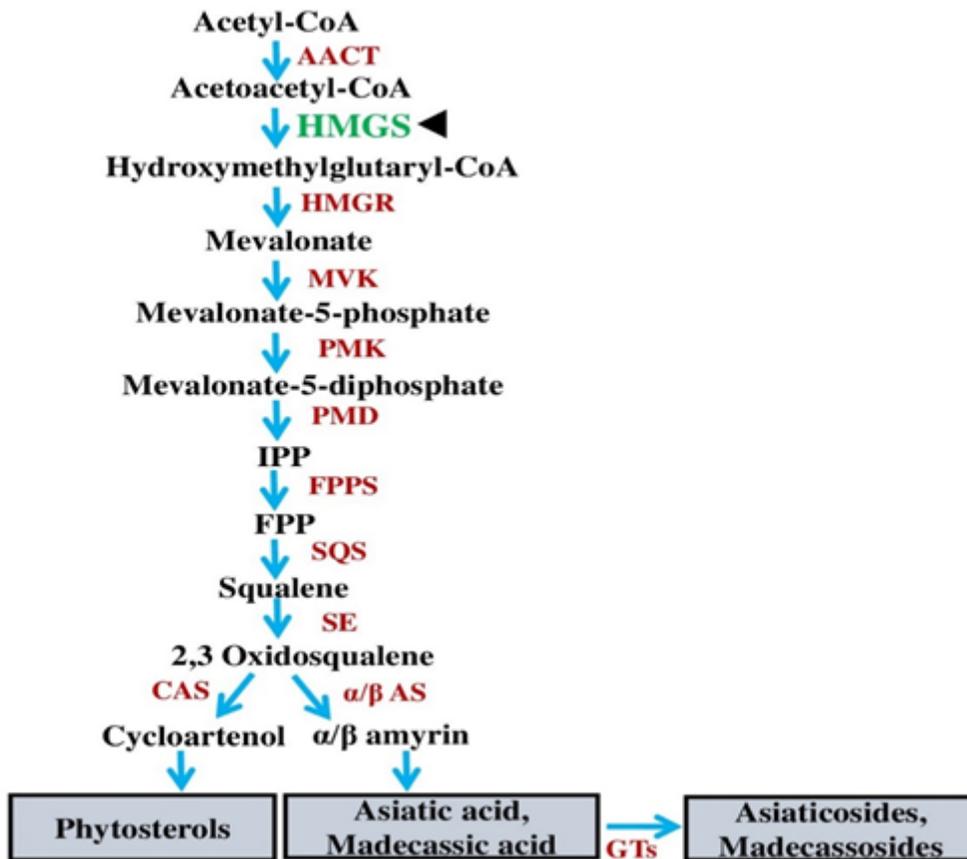


Figure 1

Mevalonate Pathway of isoprenoids biosynthesis in *Centella asiatica*; AACT: Acetyl coenzyme A (CoA) Acetyltransferase; HMGS: Hydroxymethyl glutaryl-CoA-Synthase; HMG

R: 3-Hydroxy-3-Methylglutaryl-CoA Reductase; MVK: Mevalonate Kinase; PMK: 5-Phopspho-Mevalonate Kinase; PMD: Mevalonate-5-Diphosphate Decarboxylase; IDI: Isopentenyl-Diphosphate Isomerase; FPP: Farnesyl Pyrophosphate Synthase; SQS: Squalene Synthase; SE: Squalene Epoxidase; CAS: Cycloartenol Synthase; α/β AS: α/β Amyrin Synthase GTs: Glycosyltransferase

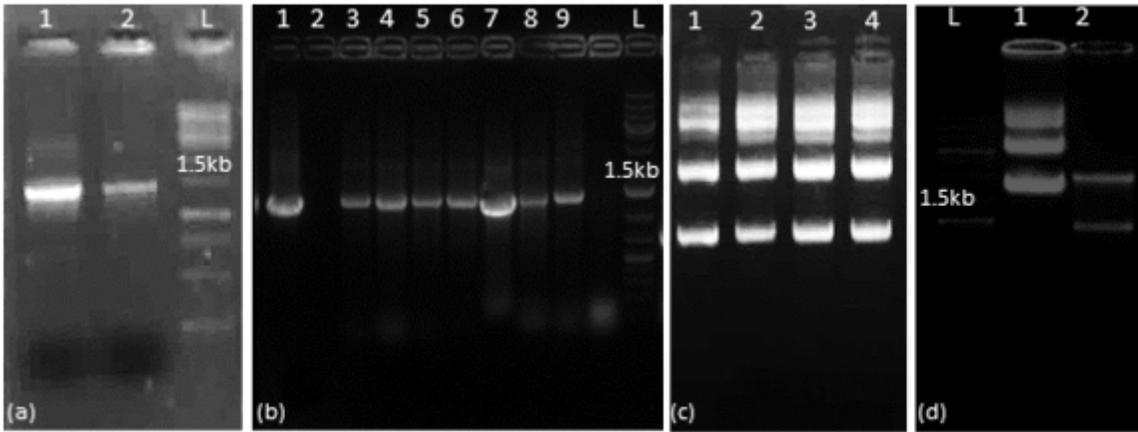


Figure 2

(a) PCR amplification of HMGS gene (lane: 1,2 1449bp gene; L-1kb ladder); **(b)** Colony PCR of HMGS gene (lane: 1- positive control; 2-negative control; 3-9-positive clones of gene of 1449bp; L-1kb⁺ ladder); **(c)** Plasmid isolation (lane: 1-4); **(d)** Confirmation of gene by restriction digestion of HMGS gene using restriction enzymes XbaI and SmaI (L-1kb⁺ ladder lane: 1-native plasmid; 2-digested plasmid of gene 1449bp gene)

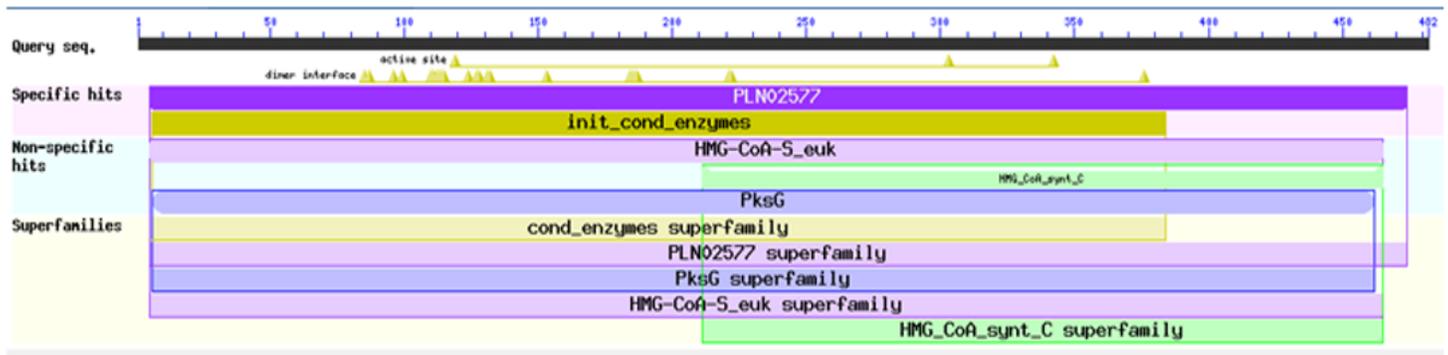


Figure 3

Different domain hits of CaHMGS proteins observed by Conserved Domain Database in NCBI



Figure 4

Multiple Sequence alignments of CaHMGS along with other homologous HMGS of different plants. Left side indicate the origin of the HMGS sequence and the right side indicate the number of amino acid residues. Red: identity = 100%; Dark green: $75\% \leq \text{identity} < 100\%$; light yellow: $50\% \leq \text{identity} \leq 75\%$. Dashes shows gaps introduced for the optimization of the alignment. Maroon box shows conserved motif and star sign indicates active site

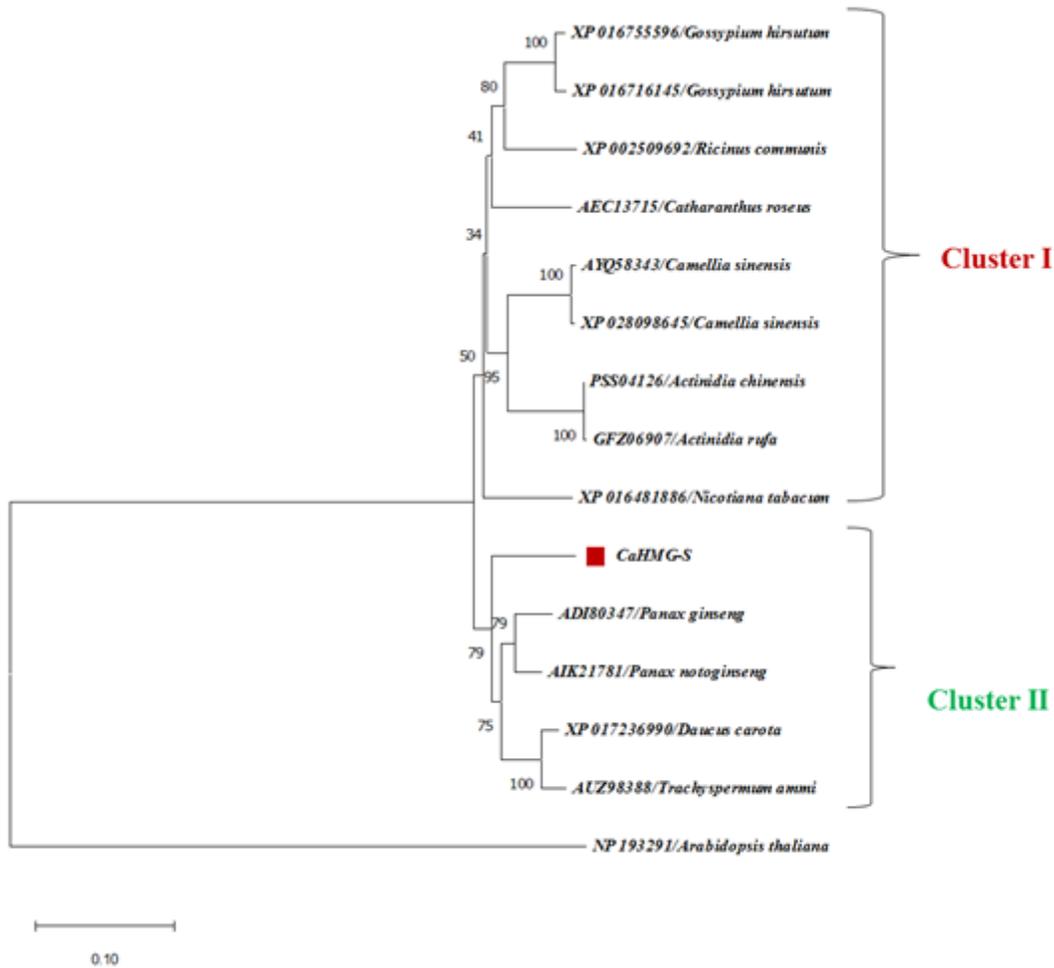


Figure 5

Phylogenetic tree of CaHMGS with its closest homologues was prepared by the Neighbor-Joining method in MEGA v10.1.6. The percentage of replicate trees in which the associated taxa were clustered together in the bootstrap test with value of 1000 replicates was displayed adjacent to the branches of the tree. The tree was constructed in scale with same units of the branch lengths in which the phylogenetic tree was concluded by evolutionary distances. p-distance method was used to estimate the evolutionary distances on the basis of number of amino acid differences per site

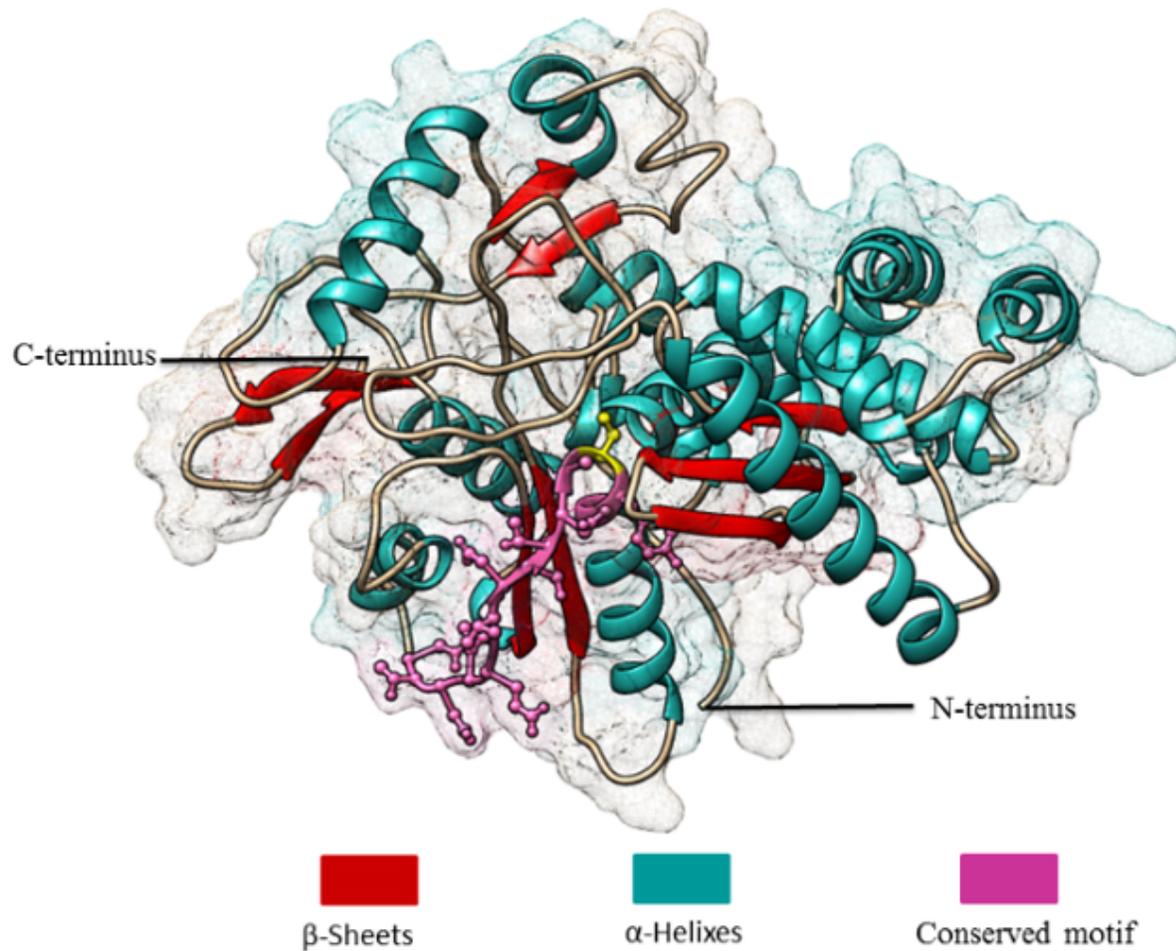


Figure 6

Three-dimensional model structure of CaHMGS. Helices shown in light sea green are α -helices and red arrows indicate β -sheets. Lines observed in brown color indicates turns and loops. Pink and yellow ball and stick revealed conserved motifs whereas yellow ball and stick showed, cysteine residue of active sites. Surface structure has shown N and C-terminal domain

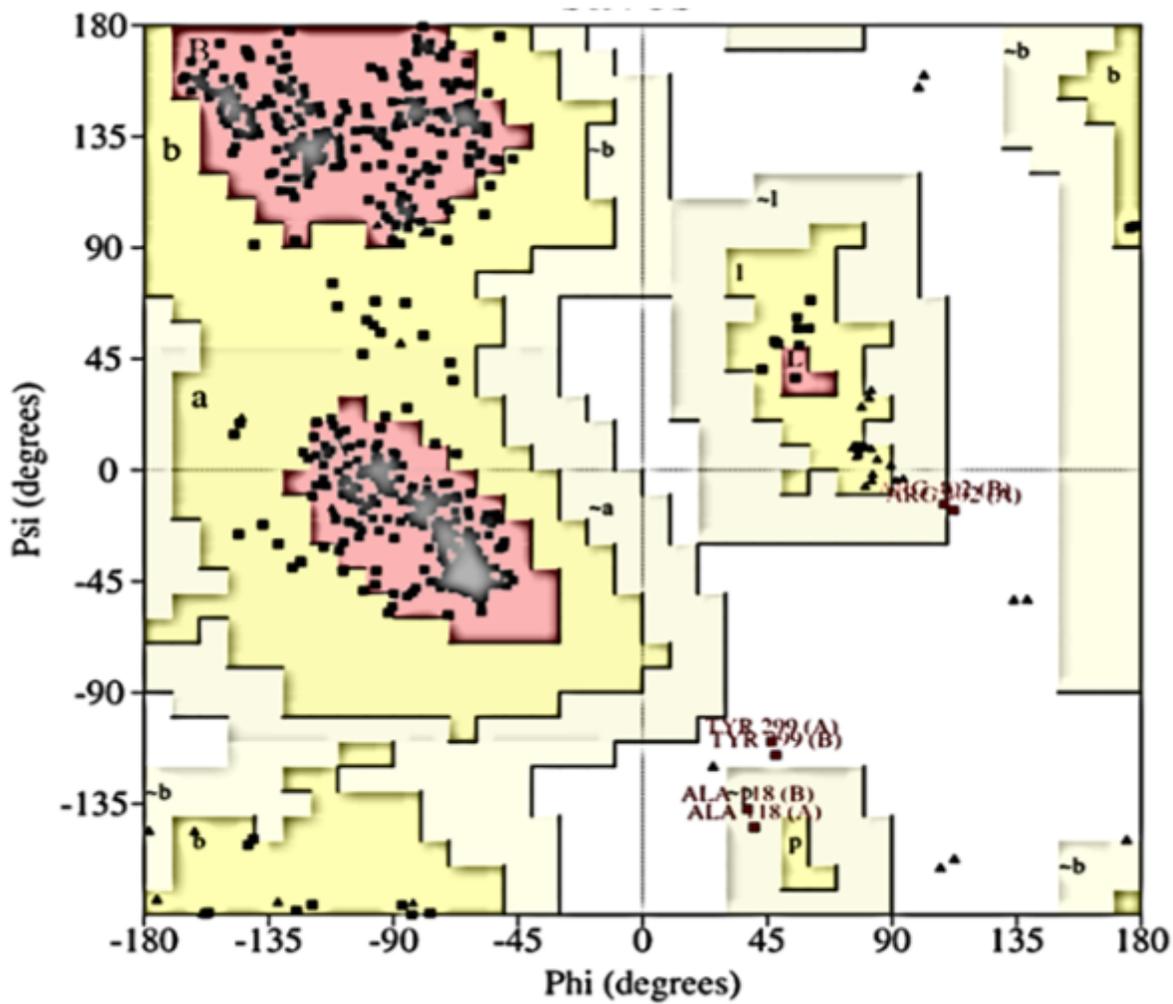


Figure 7

The energy-optimized model of CaHMGS was validated by PROCHECK. Ramachandran plot showed that only 0.4% under unfavourable region and rest 99.6% residues were under the most favoured region

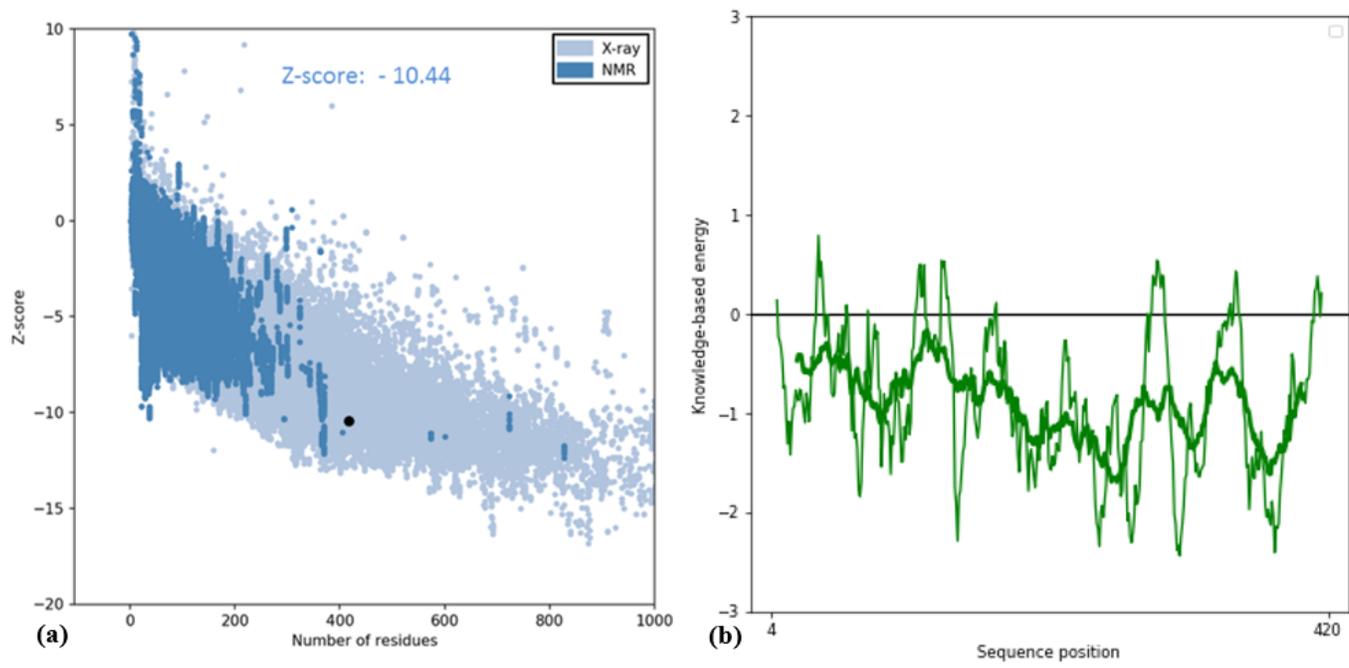


Figure 8

ProSA revealed the overall quality of the modelled structure CaHMGS using Z-score **(a)** and knowledge-based energy graph **(b)**

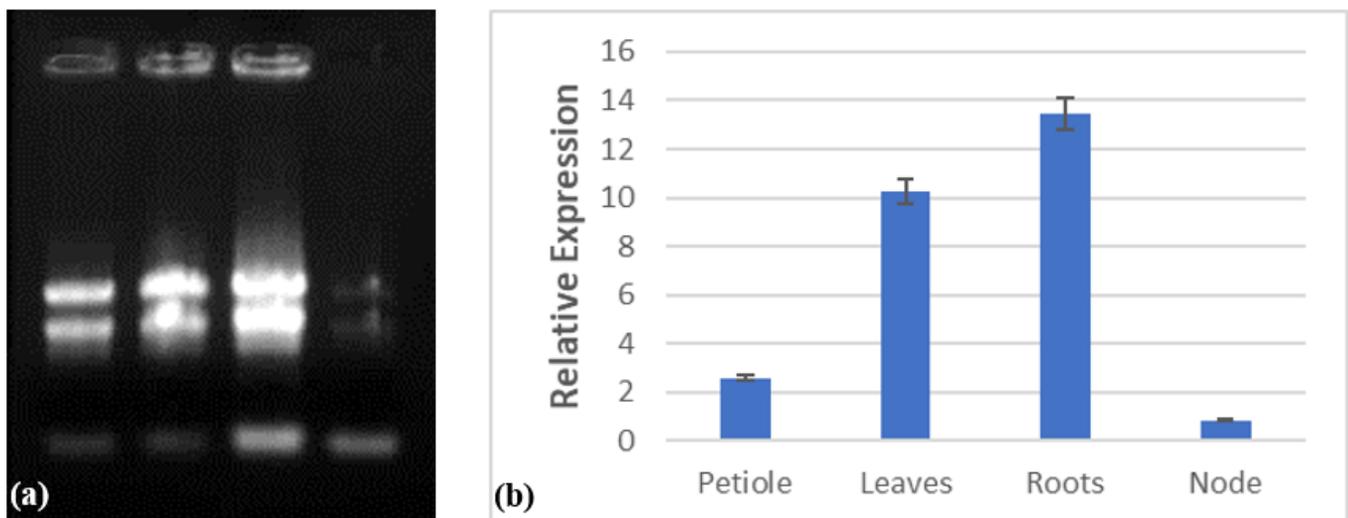


Figure 9

a) RNA bands showing concentration of gene in different tissues; **(b)** Relative expression of the gene confirmed by Real-Time PCR



Figure 10

Wild-type diploid yeast YSC1020 having pYES2-CaHMGS1 could easily grow on the non-induction medium YPD and induction YPG medium both (A& C), on the other hand, mutant haploid yeast YSC1020 containing pYES2-CaHMGS2 could only grow on induction media YPG + G418 (B) but not on non-induction media YPD (D)