In-silico analysis of *7-dehydrocholestrol reductase* gene (*Sl7-DR2*) in tomato: Linking cholesterol metabolism to vitamin-D pathway

ABSTRACT:

Tomatoes are the most popular vegetable crop worldwide and contain many health-promoting compounds. Most foods contain very little amount of vitamin D, and plants are very poor sources in the production of vitamin D. The *7-dehydrocholesterol reductase* gene(*Sl7-DR2*) is responsible for the accumulation of 7-dehydrocholesterol, which is a precursor for vitamin D_3 synthesis. This gene is present in different plant species and plays a role in cholesterol metabolism as well as it links with the vitamin D pathway by inhibiting its actual function in the synthesis of cholesterol. In this study, sequencing analysis of the gene using multiple sequence alignment provided information about its similarity and their conserved sequence regions in different plant species. Using MEGA11 software, the phylogenetic analysis of this gene was done and it showed that, this gene present in chromosome 6 and chromosome 1 of *Solanum lycopersicum* has a close resemblance with *Solanum tuberosum* and *Datura stramonium* respectively. The physiochemical characteristics of *Sl7-DR2* were analyzed by using the Expasy protparam tool. Functional annotation and secondary structure analysis of *Sl7-DR2* were done using the Pfam, STRING, and SOPMA databases. The results obtained from this analysis showed that the protein was stable and has a low GRAVY Index indicating its higher affinity to water and the secondary structure analysis predicts the robust nature of this protein and the STRING analysis gave information related to proteins that are involved in the sterol pathway and their functions. These tools characterized the protein function of the *Sl7-DR2* gene and studied the role of 7-dehydrocholesterol reductase in converting 7-dehydrocholesterol to vitamin D_3 and its role in cholesterol metabolism. The pathway analysis using the Reactome database helps in characterizing the network between the cholesterol and vitamin D pathway. The 3D structure of the *Sl7-DR2* protein was predicted using an *ab initio* approach, Robetta, and the predicted structure was validated using the SAVESv6.0 (PROCHECK) server. The predicted model has 90.8% of total residues present in the most favoured regions and showed high readability and quality. These analyses demonstrated the significance of this gene in the vitamin D pathway and its potential for human health in preventing various diseases.

Keywords:Cholesterol;VitaminD;*7-Dehydrocholestrol reductase gene*;Pathway analysis; STRING;SOPMA;Molecular modelling;Robetta;PROCHECK

INTRODUCTION:

Tomato (*Solanum lycopersicum*) is an excellent vegetable that complements any diet because of its health advantages as it contains minerals, phytochemicals, lycopene, potassium, iron, folate, and vitamin C [1,2]. Tomato is suitable for understanding various plant metabolic pathways, including the cholesterol pathway that can be manipulated for accumulating vitamin D_3, a crucial human steroid hormone. The deficiency of vitamin D is widely prevalent among the human population. Vitamin D deficiency in humans causes inflammation, micronutrient deficiencies, cancer, depression, dementia, and rheumatoid arthritis. Vitamin D deficiency affects around one billion people globally(3). The major reason for the increase in vitamin D deficiency is largely because of inadequate dietary availability. The best dietary sources of vitamin D are fish and dairy products. Most foods contain little vitamin D, and plants are very poor sources of vitamin D_3 but its synthesis is recognized in tomatoes in the form of 7-Dehydrocholesterol(4).

The Vitamin D content in the blood serum is estimated by knowing the concentration of the 25-hydroxy vitamin D_3 (Calcidiol)[5]. The sufficient amount of 25(OH)D_3 required by humans is 50 nmol/L[8] and the optimal amount of 25(OH)D_3 required is at a concentration of 80 nmol/L [8,9]. A level of 20 ng/ml to 30 nmol/L of 25(OH)D_3 is considered deficient in people[6,7]. The National Institute of Health implemented the vitamin D standardization program (VDSP)[10,11], which establishes deficiency as anything below 50 nmol/L[12]. Moreover, vitamin D deficiency increases the risk of developing COVID-19 [13].

Vitamin D synthesis in humans is linked to the cholesterol pathway and its synthesis is regulated by the 7-dehydrocholesterol reductase enzyme. This enzyme has a common role in the cholesterol and vitamin D biosynthesis pathway. The main role of this enzyme is in the conversion of 7-dehydrocholesterol to cholesterol but due to the downregulation or inactivation of this enzyme, the amount of 7-dehydrocholesterol increases, and it is converted to vitamin D_3 through UV-B exposure at a wavelength of 290-315nm. The biosynthesis of the active form of vitamin D [1,25(OH)D_3], involves a series of metabolic processes that occur across various cellular compartments including extracellular regions, cytosol,

lysosomes, endoplasmic reticulum, and mitochondria[15].In humans for the synthesis of vitamin D two metabolic pathways are involved such as the mevalonate pathway and the Kandutsch-Russel branch of the cholesterol biosynthesis and these pathways help in the transition of 7-dehydrocholestrol to pre-vitamin D_3[14] by exposure to UV-B light by cleaving the bond between C9 and C10 carbons of 7-Dehydrocholestrol. Vitamin D produced in the skin enters the blood circulation and is transported to the liver by vitamin D binding protein (DPB) and then converted to 25-hydroxyvitamin D_3[25(OH)D_3] by CYP2R1(cytochrome P450 2R1)[16,17]. Finally, in the Kidneys, 25(OH)D_3 is transformed to its active form1,25-dihydroxyvitamin D_3[1,25(OH)2D_3], through CYP27B1 (cytochrome P450 family 27 subfamily B member 1) [18](Fig.1).

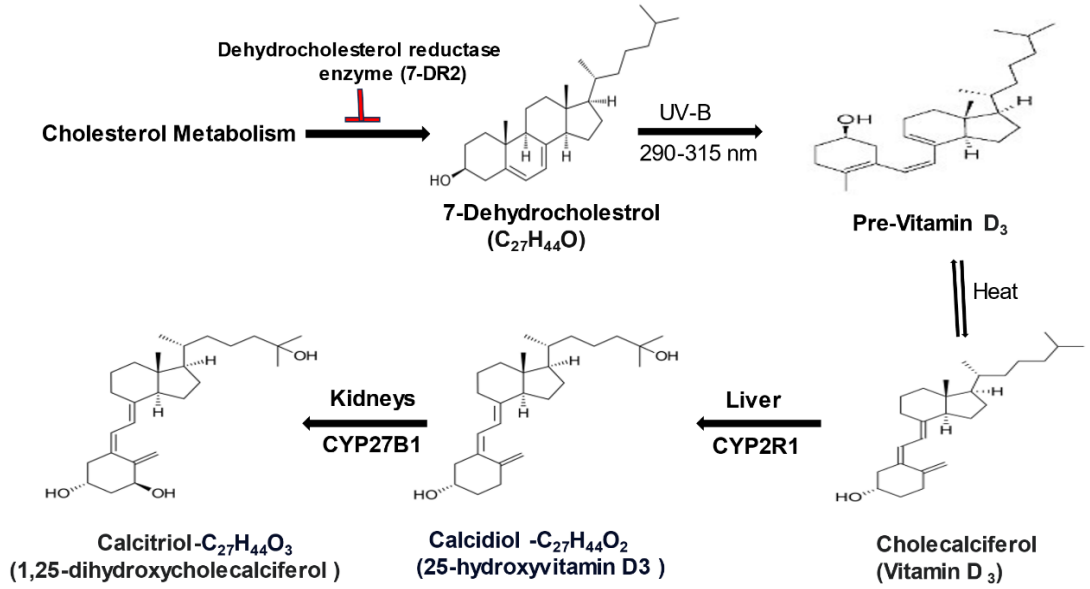


Fig.1.Vitamin D synthesis from 7-dehydrocholestrol in human

Plants producing steroidal glycoalkaloids (SGAs), such as α-solanine in potatoes and α-tomatine in tomatoes, are part of the Solanaceae family and accumulate cholesterol as a major precursor [19].These plants can serve as unique models for studying plant cholesterol metabolism[20] and manipulating the functional role of this gene through genome editing helps in the synthesis of vitamin D without conversion into cholesterol and this gene is present in different plant species and helps in studying its importance in the accumulation of 7-dehydrocholesterol.

In this paper, the *insilico* analysis of the *7-dehydrocholesterol reductase(Sl7-DR2) in Solanum lycopersicum* was studied that is involved in coding the dehydrocholesterol reductase protein, employing a series of interconnected analyses was studied and the gene annotation and sequencing analysis between the interspecies helps in characterization and its genetic constitution. The phylogenetic analysis and the functional characters of this gene were studied and helped in finding its genomic position and its nature in involving various biosynthetic pathways. The computational modeling of the pathway involved in vitamin D synthesis and cholesterol synthesis helps in studying the relation between the two pathways. The protein structure prediction and the modeling of the protein helped in studying the gene significance in tomatoes and also helped in finding its enzymatic potential.

2. MATERIALS AND METHODS:

2.1 Gene Annotation and Sequence Analysis:

For sequence analysis and gene annotation, 8 different plant species and a human being were selected to study protein sequence in terms of similarity and also to find the conserved protein domains, and motifs present in this gene. The *7-Dehydrocholestrol reductase gene* and protein sequence of different plant species *Solanum lycopersicum Chr-6* (NP_001353043), *Solanum lycopersicum Chr-1* (NP_001353044), *Solanum tuberosum (*NP_001305624), *Arabidopsis thaliana*(NP_001077693), *Capsicum annum* (XP_016577971), *Withiana somnifera*(AOX15301), *Datura stramonium*(MCD7446296), *Rosa chinensis* (XP_024181789), *Nicotiana tobaccum* (XP_016480318) and *Homo sapiens* (NP_001351.2) were retrieved in FASTA format from NCBI[21] and they were aligned to study the sequence similarity using crustal-W tool. The conserved protein domains and functional motifs within the protein sequence were studied by using the Inter ProScan tool [22].

2.1.1 Phylogenetic analysis :

The *7-Dehydrocholestrol reductase* homologous *gene* sequences from 8 different plant species (NP_001353043, NP_001353044, NP_001077693, NP_001305624, XP_016577971, AOX15301, MCD7446296, XP_024181789, XP_016480318) and homo sapiens(NP_001351.2) coding this gene were chosen and then they were aligned using the MEGA software and the phylogenetic tree was constructed among these species to infer the evolutionary relationship and branching pattern among the species. The bootstrap analysis was done to analyze the robustness of the phylogenetic tree.

2.2 Physio-Chemical Characterization:

The physio-chemical properties of 7-dehydrocholesterol reductase protein sequences and the amino acid composition of the gene from different species were determined using the ProtParam server[23] from the ExPASy tool. The isoelectric point (pI), the total number of negative (-R) and positive (+R) residues, extinction coefficients (EC), instability index (II), aliphatic index (AI), and grand average hydropathy were all calculated.

2.3 Functional Annotation and protein interaction analysis of the gene:

The retrieved protein sequence of the *7-dehydrocholesterol reductase gene* in chromosome 6 of *Solanum lycopersicum* was used for finding the conserved regions and similarities among sequences, the multiple sequence alignment among different organisms was characterized and the gene potential functions (biological, molecular, and cellular) were predicted by using the UniProt database[24] and STRING database. The protein family, domain, and functional site data were computed using the Prosite server[25] and for characterization of the chosen protein in relation to the particular protein family was done using Pfam[26]. The STRING database was used to find the proteins that are involved in the interaction with 7-dehydrocholestrol reductase and the pathways related to it.

2.4 Prediction of Secondary Structures:

The secondary structure of the *7-dehydrocholesterol reductase gene* was predicted by using the Self-Optimized Prediction Method with Alignment (SOPMA)[27] server. The default parameters of SOPMA, which encompass a similarity threshold of 8 and a window width of 17, were employed. SOPMA algorithmically assesses the content of α-helix, extended-sheets, beta turns and random coils, based on the amino acid sequence.

2.5 Computational Modeling of Vitamin D and Cholesterol Pathway in Higher Organisms:

The computational modeling helps in examining the convergence of cholesterol and vitamin D metabolism [28]. The interplay between these two pathways was studied using the system biology and helped in finding their biochemical interactions. This study helps in finding the role of other metabolisms involved in synthesizing vitamin D from the cholesterol pathway.

2.6 Molecular Modeling of 7-dehydrocholestrol reductase:

The three-dimensional structure of the full-length 7-dehydrocholesterol reductase protein from *Solanum lycopersicum* was not present in the PDB database. In light of this, the tertiary structure of the protein was generated through the utilization of the Robetta web server[29] using the protein query sequence. The quality of the predicted 3D model was subsequently assessed utilizing the SAVES v6.0 server [30]. This was done to evaluate the overall stereochemical properties of the protein, specifically, energetically allowed regions used for docking studies to know the relation of this gene to Vitamin D_3 conversion.

3. RESULTS AND DISCUSSION

3.1 Sequence and Phylogenetic Analysis:

The gene sequences of 8 different plant species and human being (NP_001353043, NP_001353044, NP_001351.2, NP_001077693, NP_001305624, XP_016577971, AOX15301, MCD7446296, XP_02 4181789, XP_016480318)were retrieved from the NCBI database (Annexure 1). The Phylogenetic tree predicts that this gene present in chromosome 6 and chromosome 1 of *Solanum lycopersicum* has a close resemblance with *Solanum tuberosum and Datura stramonium* having a bootstrap value of 43 and 40 respectively. This gene present in Homo sapiens and Arabidopsis have 100% resemblance.Fig.(2). The multiple sequence alignment analysis of this gene in the 8 plant species and human beings shows the similarity in the amino acid sequence and shows that they are conserved among the plant species(Annexure 1).

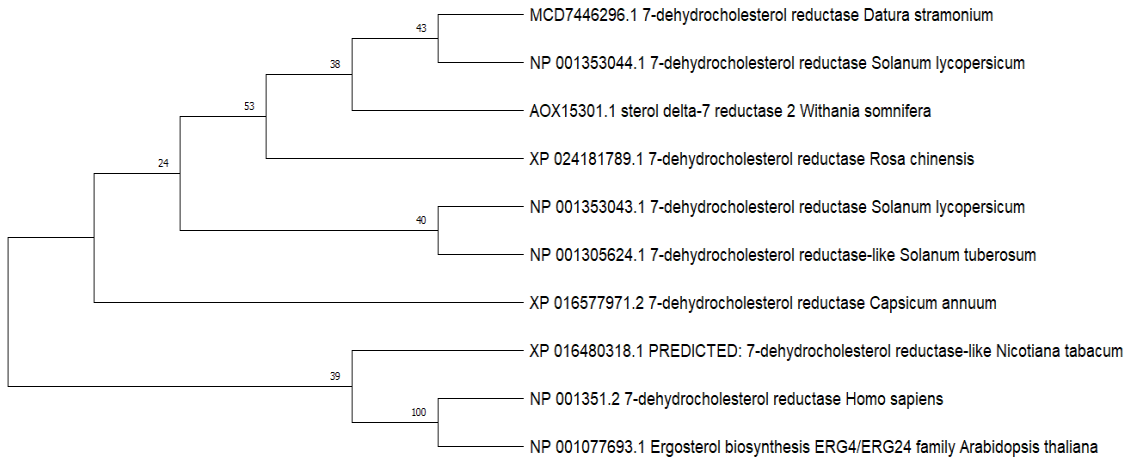


Fig.2.Phylogenetic analysis of 7-Dehydrocholestrol reductase in different species

3.2 Physio-Chemical characterization:

The utilization of Expasy 's Protparam tool allowed for the computation of the amino acid composition and physio-chemical properties of 7-dehydrocholesterol reductase protein from 8 different plant species and human beings (Annexure 2). The protein sequence of the *7-dehydrocholesterol reductase gene* exhibits a variance in length, ranging from 431 to 475 amino acids, and displays different molecular weights (Table 1). The isoelectric point (pI), which measures protein solubility and electrophoretic separation [31], was found to be greater than 7 for all the plant species and humans, indicating an elevated level of protein solubility and electrophoretic separation. The relative volume filled by aliphatic side chains (Ala, Val, Ile, and Leu) in a protein is characterized by its aliphatic index [32]. A positive index shows that the thermostability of globular proteins has increased. The results showed that the aliphatic index was higher in *Capsicum annum* and then in *Nicotiana tobaccum*, *Solanum lycopersicum* (tomato) than that of other species. Lastly, the instability index (II), which is a parameter for measuring protein stability, denotes a stable protein if the value is below 40[33]. The gene in all these species has a lower instability index below 40 and denotes stability in various species.

The 7-dehydrocholesterol reductase protein found in all the plant species and humans exhibits a low isoelectric point, suggesting that these proteins are stable. The solubility of proteins was measured by using the GRAVY index, whereby lower values indicate greater affinity for water molecules and higher values suggest greater hydrophobicity [34]. In comparison to the other species, *Solanum lycopersicum* displays a lower GRAVY index, indicating its potential for higher interactions with water molecules. Based on these physio-chemical properties, it is possible that the protein sequence of 7-dehydrocholesterol reductase in tomato has high stability due to its higher aliphatic index and lower GRAVY index denotes its solubility nature.

Table 1: Parameters computed using Expasy's Protparam tool

Protein Name	No.of Amino acids	MW	pI	-R	+R	EC	II	AI	GRAVY
S.lycopersicum-[Chromosome 6]	434	49786.32	8.77	28	36	132295	31.67	100.16	0.225
S. tuberosum	434	49715.25	8.85	27	36	130805	33.82	99.49	0.239
S.lycopersicum [Chromosome 1]	434	49671.37	8.92	28	38	131835	31.84	102.19	0.309
H. sapiens	475	54489.43	8.95	28	39	153960	29.94	96.08	0.232
A.thaliana	431	49404.87	8.65	28	35	137460	31.82	98.84	0.371
C. annum	434	49859.52	8.75	28	36	129440	28.06	103.73	0.294
W. somniferum	434	49755.39	9.02	28	39	131835	32.78	99.72	0.298
R. chiinensis	434	49722.13	8.97	26	36	139855	34.47	97.95	0.285
N..tobaccum	434	49768.33	8.66	27	34	136430	32.44	103.09	0.309
D. Somniferum	434	49699.35	8.95	27	38	133450	33.81	99.72	0.286

**MW:Molecular weight,*pl:Isoelectric Point,*-R:Number of negative residues,*+R:Number of positive residues,*EC:Extinction Coefficient at 280nm, *II:Instability Index,:*AI:Aliphatic Index,:*GRAVY:Grand Average Hydropathicity*

3.3 Functional Annotation and protein interaction analysis of the gene:

In *Solanum lycopersicum* the gene's potential functions biological, molecular, and cellular were characterized using UniProt and STRING database(Table.2). The domain identified by profile for the 7-dehydrocholestrol reductase enzyme was the cytoplasmic domain and the results of Pfam, Interpro was well characterized in table3. The *Sl7-DR2* gene is involved in the sterol biosynthetic process (GO:0016126), and lipid metabolic process and has oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor and it is a cellular component of transmembrane(GO:0016020) regions. The pattern and profile of the protein were assessed using Expasy's Prosite, which stores annotations for motif descriptors that can be used to identify protein families and domains[35]. The protein was also subjected to Pfam analysis, and it shows the importance of the gene that codes 7-dehydrocholestrol reductase protein belongs to the ergosterol biosynthesis ERG4/ERG24 family. This analysis helps in studying the efficiency of genome annotation and curating information on known protein families[36]. The STRING database predicted the interaction between 7-dehydrocholestrol reductase protein(AOA3Q7H0B6) to other proteins(A0A3Q7IVA9, A0A3Q7HGH7, A0A3Q7F016, A0A3Q7FC39, SSR2, SISSR1, CYP51, A0A3Q7H150, A0A3Q7F823, A0A3Q7G0R2), with similar domains and families help in correlating the characteristics of this protein(fig.3). The proteins A0A3Q7F823, SISSR1, SSR2, A0A3Q7F016, A0A3Q7FC39, A0A3Q7G0R2, A0A3Q7IVA9 are all characterized proteins and A0A3Q7H150, CYP51, A0A3Q7HGH7 are uncharacterized proteins. These characterized proteins are involved in sterol synthesis, metabolic pathways, and lipid synthesis. This analysis also predicted that there was no coexpression analysis exists between these proteins in *Solanum lycopersicum* but it exists in other organisms among the predicted characterized and uncharacterized proteins. The biological pathway related to the *7-dehydrocholestrol reductase* gene was analyzed using the Kyoto Encyclopedia of Genes and Genomes(KEGG) and Reactome database (Table 3). This information can be used to study a wide range of biological phenomena, from the functioning of individual cells to the evolution of entire ecosystems.

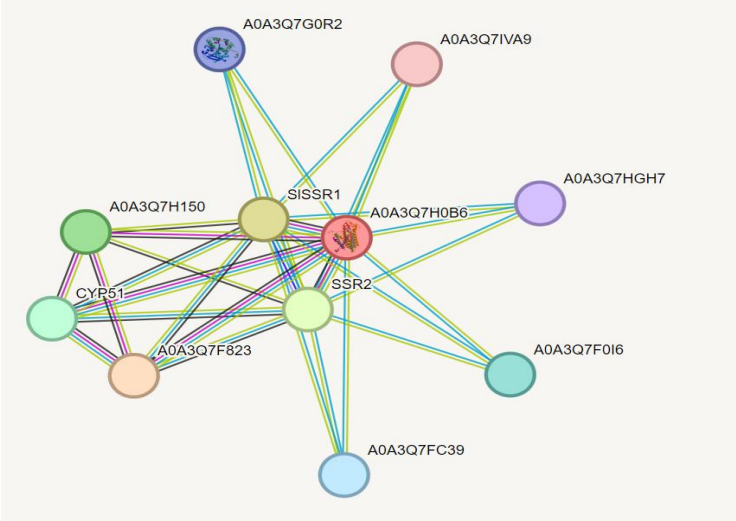


Fig.3 Interaction of 7-dehydrocholestrol reductase with other proteins using STRING database

Table.2 Biological ,Molecular and Cellular function of *Sl7-DR2* gene characterized through String database

Biological Function	Molecular Function	Cellular Function
Defense response to insect	Delta 24-sterol reductase activity	Endoplasmic reticulum
Brassinosteroid biosynthetic process	Triglyceride lipase activity	
Sterol biosynthesis process	Oxidoreductase activity ,acting on the CH-CH group of donors	
Steroid biosynthesis process	Hydrolase activity on ester bonds	
Plant-type secondary cell wall biogenesis	Catalytic activity	

Organic hydroxy compound biosynthesis process				
Unidimensional cell growth				
Lipid catabolic process				
Regulation of hormone levels				
Lipid metabolic process				
Metabolic process				

Table 3: Network analysis, protein domains, pathways related to *7-dehydrocholestrol reductase* gene

Local network cluster(STRING)	**Protein Domains(Pfam)**	**Protein Domains and Features (Interpro)**	**KEGG pathways**	**Reactome Pathways**
Delta(24)-sterol reductase,and Sterol delta 7 reductase activity	Putative esterase domain	Lipase, eukaryotic	Steroid biosynthesis	Formation of the cornified envelope
Cholesterol biosynthesis via lathosterol and Sterol reductase	FAD binding domain	Delta(24)-sterol reductase	Biosynthesis of secondary metabolites	Cholesterol biosynthesis via lathosterol
Steroid biosynthesis, and Squalene cyclase, N-terminal	Alpha/beta hydrolase fold	Partial AB-hydrolase lipase domain	Metabolic pathways	Cholesterol biosynthesis via desmosterol
Mixed, Phosphatase complex, and Epoxide hydrolase		Serine aminopeptidase		Keratinization
		Alpha/Beta hydrolase fold		Cholesterol biosynthesis

3.4 Prediction of secondary structure:

The secondary structure of the 7-dehydrocholestrol reductase protein of *Solanum lycopersicum and other plant species* and humans was predicted using SOPMA(Table 4). The results show that *Solanum lycopersicum* has a higher percentage of alpha helices and it dominates the other species followed by random coils, extended sheets, and beta strands.High alpha helix percentage demonstrating the robust nature of the 7-Dehydrocholestrol reductase protein.

Table4:Predicted secondary structures of different species that codes this gene

Different species (Locus id)	Amino acids	Alpha helix	Extended sheets	Beta turns	Random coils
Solanum lycopersicum (Chr-6)	434	47.74%	16.36%	5.53%	30.65%
Solanum tuberosum	434	44.24%	18.20%	5.53%	32.03%
Solanum lycopersicum (Chr-1)	434	44.01%	19.35%	6.91%	29.72%
Homo sapiens	475	43.16%	17.89	5.53%	33.89%
Arabidopsis thaliana	431	42.46%	18.79	5.80%	32.95%
Withiana somnifera	434	40.78%	20.28%	6.68%	32.26%
Datura stramonium	434	46.77%	17.97%	6.22%	29.03%
Capsicum annum	434	39.63%	21.66%	5.99%	32.72%
Rosa chinensis	434	45.39%	18.66%	5.76%	30.18%
Nicotiana tobaccum	434	42.63%	20.28%	6.68%	30.41%

3.5 Molecular pathway analysis of vitamin D and Cholesterol :

Using the Reactome database the cholesterol and vitamin D biosynthesis pathway was analyzed as well as the feedback mechanisms that involved in regulating their synthesis. The system biology graphical notation maps were helpful in studying the entire role of different pathways, enzymes, and proteins that are involved in vitamin D biosynthesis. Systems Biology Graphical Notation (SBGN) data describes the detailed vitamin D and cholesterol pathway[37].

The pathway starts in the cytosol with the conversion of acetyl coenzyme A (AC-CoA) by the enzyme acetyl-coenzyme A acetyltransferase 2(ACAT2) and then progresses through the endoplasmic reticulum, where key steps involve the conversion of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) to mevalonic acid (MVA) catalyzed by HMGCR[14]. This step is targeted by statin drugs for treating high cholesterol levels. The pathway branches into the Bloch pathway (forming cholesterol) and the Kandutsch–Russell pathway, leading to vitamin D biosynthesis(fig.4). The vitamin D-forming branch starts with the conversion of 7-dehydrocholesterol to cholecalciferol (Vitamin D_3) through UV radiation in the skin layers[38]. The feedback regulation involves (Sterol regulatory element binding protein) SREBP complexes and high cholesterol binds to them and retains them in the endoplasmic reticulum, while low cholesterol levels activate SREBP complexes, upregulating enzymes responsible for early pathway steps subsequently transformed into calcidiol and finally transforming to calcitriol[1,25α(OH)D_3][14].In this way, these pathways help in regulating the formation of cholesterol to vitamin D.

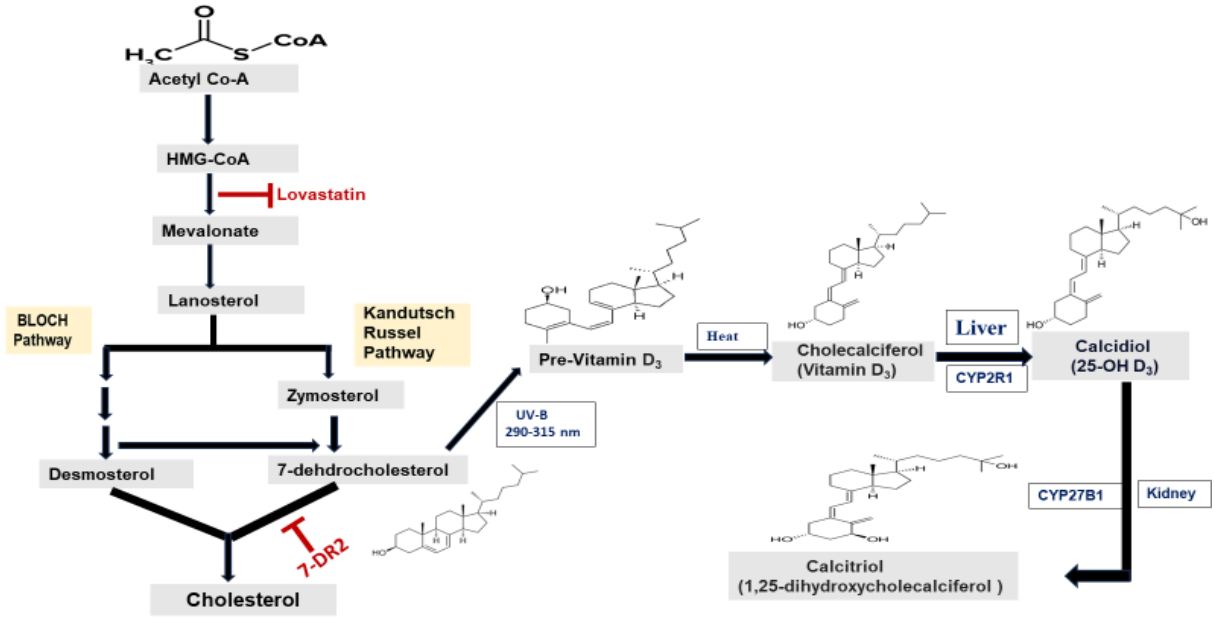


Fig.4 Network connecting the Cholesterol pathway to the Vitamin D biosynthesis pathway

3.6 3D-Modeling and validation of 7-dehydrocholestrol :

The tertiary structure is very important for studying the protein's function and active sites. The Robetta server was used to create the 3D model of the 7-dehydrocholestrol reductase protein by giving the query sequence and a 3D model for this protein was generated by using this server. The predicted model for this protein was visualized by using the PyMOL molecular visualization tool (Fig.5). For evaluating the predicted model SAVES server 6.0 was used and the model of this protein underwent PROCHECK analysis, which led to the generation of a Ramachandran plot (Fig.6). The Ramachandran plot analysis categorized residues into quadrangle regions. The graph's red sections denote the most permissible areas, while yellow areas indicate permissible areas. The use of Ramachandran map calculations, in conjunction with the PROCHECK tool, facilitated the assessment of the stereochemical quality of predicted models after refinement. The model structure's reliability and high quality are evidenced by the distribution of approximately 90.8% of the total residues in the most widely distributed region (Table 5)

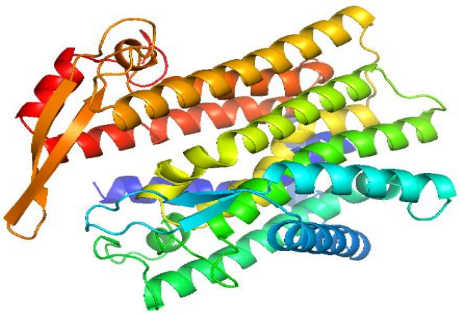


Fig.5. 3D structure of 7 -dehydrocholesterol reductase protein predicted by Robetta server and visualized through PyMOL

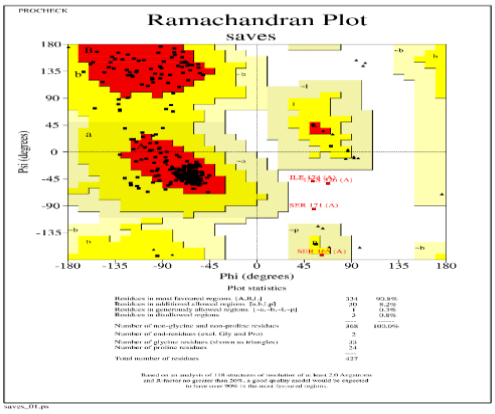


Fig. 6. Ramachandran plot showing the phi -psi torsion angles for all the residues of the 7-Dehydrocholestrol reductase protein: the distribution of amino acid residues to β-sheets (+psi, -phi), right-handed α-helix (-psi, -phi) and left-handed α-helix (+phi, +psi)

Table 5. Ramachandran plot calculation with the PROCHECK program

Ramachandran plot statistics	Distribution
Residues in most favoured regions	90.8%
Residues in additionally allowed regions	8.2%
Residues in generously allowed regions	0.3%
Residues in disallowed regions	0.8%

4. CONCLUSION:

The functional role of the *7-dehydrocholesterol reductase* gene in the process of vitamin D synthesis was studied. When the gene is upregulated, the 7-Dehydrocholestrol is converted into cholesterol, but due to the downregulation of the gene the 7-DHC content will increase and the pathway is shifted from cholesterol synthesis to vitamin D_3 synthesis and the Kandutsch Russel pathway clearly showed the transformation of 7-dehydrocholesterol to pre-vitamin D_3 due to the inactivation of the enzyme that restricts the conversion to cholesterol. From the sequence and phylogenetic analysis of the gene, we conclude that in *Solanum lycopersicum* this gene was present in two chromosomes 1 and 6, and has a resemblance to other species in terms of similarity. Furthermore, the physio-chemical analysis of the gene tells about its stability in tomatoes and its importance in the conversion of 7-DHC to vitamin D_3. Pathway analysis showed that the gene has a role in cholesterol metabolism and also provides information related to the compounds that are involved in changing the cholesterol pathway. The functional analysis of this protein represents its stability, hydrophilic nature, and solubility nature. In addition, molecular modeling has revealed a β-barrel fold structure of the protein that is indicative of enzymatic function in the conversion of 7-DHC to vitaminD_3 and the Ramachandran plot validated this model and the results represent 90.8% of amino acid residues in the favoured region. Overall, this study's results have clearly demonstrated this gene's stability and its significant role in the synthesis of vitamin D_3, which has important implications for human health, immunity, and crop enhancement. From this study, we can conclude that this gene in other species also has a role in vitamin D conversion and further helps in studying the role of this gene in different plant species in the conversion of dehydrocholesterol to vitamin D_3 by genome editing. Subsequent research could focus on exploring the gene's role in other

plant species, delving into the detailed synthesis mechanisms, and devising strategies to enhance the vitamin D content in crops.

5. ACKNOWLEDGEMENTS :

The authors express their gratitude to the Department of Plant Biotechnology at Tamil Nadu Agricultural University in Coimbatore for furnishing the requisite resources and facilities to carry out this research. We would also like to express our sincere appreciation to the Department of Bioinformatics for their generous support and guidance. Their expertise in bioinformatics was invaluable in the analysis of our data.

REFERENCES

1. Kumar KS, Paswan S, Srivastava S. Tomato-a natural medicine and its health benefits. Journal of Pharmacognosy and Phytochemistry. 2012;1(1):33-43.
2. Toor RK, Savage GP. Antioxidant activity in different fractions of tomatoes. Food research international. 2005 Jun 1;38(5):487-94.
3. van Schoor N, Lips P. Worldwide vitamin D status. Vitamin D. 2018 Jan 1:15-40.
4. Li J, Scarano A, Gonzalez NM, D'Orso F, Yue Y, Nemeth K, Saalbach G, Hill L, de Oliveira Martins C, Moran R, Santino A. Biofortified tomatoes provide a new route to vitamin D sufficiency. Nature Plants. 2022 Jun;8(6):611-6.
5. Cashman KD, van den Heuvel EG, Schoemaker RJ, Preveraud DP, Macdonald HM, Arcot J. 25-Hydroxyvitamin D as a biomarker of vitamin D status and its modeling to inform strategies for prevention of vitamin D deficiency within the population. Advances in nutrition. 2017 Nov;8(6):947-57.
6. Cashman KD, Dowling KG, Skrabakova Z, Gonzalez-Gross M, Valtuena J, De Henauw S, Moreno L, Damsgaard CT, Michaelsen KF, Mo lgaard C, Jorde R. Vitamin D deficiency in Europe: pandemic?. The American journal of clinical nutrition. 2016 Apr 1;103(4):1033-44
7. Bouillon R, Carmeliet G. Vitamin D insufficiency: Definition, diagnosis and management. Best practice & research Clinical endocrinology & metabolism. 2018 Oct 1;32(5):669-84
8. Dawson-Hughes B, Heaney RP, Holick MF, Lips P, Meunier PJ, Vieth R. Estimates of optimal vitamin D status. Osteoporosis international. 2005 Jul;16:713-6.
9. Henry HL, Bouillon R, Norman AW, Gallagher JC, Lips P, Heaney RP, Vieth R, Pettifor JM, Dawson-Hughes B, Lamberg-Allardt CJ, Ebeling PR. 14th Vitamin D Workshop consensus on vitamin D nutritional guidelines. The Journal of steroid biochemistry and molecular biology. 2010 Jul;121(1-2):4-6.
10. Sempos CT, Vesper HW, Phinney KW, Thienpont LM, Coates PM, Vitamin D Standardization Program (VDSP)*. Vitamin D status as an international issue: national surveys and the problem of standardization. Scandinavian Journal of Clinical and Laboratory Investigation. 2012 Apr 1;72(sup243):32-40.
11. Binkley N, Sempos CT, Vitamin D Standardization Program (VDSP). Standardizing vitamin D assays: the way forward. Journal of Bone and Mineral Research. 2014 Aug;29(8):1709-14.
12. Sempos CT, Heijboer AC, Bikle DD, Bollerslev J, Bouillon R, Brannon PM, DeLuca HF, Jones G, Munns CF, Bilezikian JP, Giustina A. Vitamin D assays and the definition of hypovitaminosis D: results from the First International Conference on Controversies in Vitamin D. British journal of clinical pharmacology. 2018 Oct;84(10):2194-207.
13. Lopez L, Hart LH, Katz MH. Racial and ethnic health disparities related to COVID-19. Jama. 2021 Feb 23;325(8):719-20.
14. Warren T, McAllister R, Morgan A, Rai TS, McGilligan V, Ennis M, Page C, Kelly C, Peace A, Corfe BM, Mc Auley M. The interdependency and co-regulation of the vitamin D and cholesterol metabolism. Cells. 2021 Aug 6;10(8):2007.
15. Christakos S, Ajibade DV, Dhawan P, Fechner AJ, Mady LJ. Vitamin D: metabolism. Rheumatic Disease Clinics. 2012 Feb 1;38(1):1-1.
16. Bhattacharyya MH, DeLuca HF. The regulation of rat liver calciferol-25-hydroxylase. Journal of Biological Chemistry. 1973 May 1;248(9):2969-73.
17. Christakos S, Ajibade DV, Dhawan P, Fechner AJ, Mady LJ. Vitamin D: metabolism. Endocrinol Metab Clin North Am. 2010 Jun;39(2):243-53.
18. Bikle DD, Patzek S, Wang Y. Physiologic and pathophysiologic roles of extra renal CYP27b1: Case report and review. Bone reports. 2018 Jun 1;8:255-67.
19. Milner SE, Brunton NP, Jones PW, O'Brien NM, Collins SG, Maguire AR. Bioactivities of glycoalkaloids and their aglycones from Solanum species. Journal of agricultural and food chemistry. 2011 Apr 27;59(8):3454-84.

20. Choi S, You MK, Jeon YA, Lee J, Kim J, Park YJ, Kim J, Park J, Kim JK, Choe S. Metabolic Engineering to Enhance Provitamin D_3 Accumulation in Edible Tomatoes. GEN Biotechnology. 2023 Jun 1;2(3):219-27.
21. Available:https://www.ncbi.nlm.nih.gov/
22. Available:https://www.ebi.ac.uk/interpro/search/sequence/
23. Available:https://web.expasy.org/protparam/
24. Available:https://www.uniprot.org/
25. Available: https://prosite.expasy.org/
26. Available:http://pfam-legacy.xfam.org/
27. Available:https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html
28. Mc Auley MT, Proctor CJ, Corfe BM, Cuskelly GJ, Mooney KM. Nutrition research and the impact of computational systems biology. Journal of Computer Science & Systems Biology. 2013 Oct 7;6(5):271-85.
29. Novere NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT. The systems biology graphical notation. Nature biotechnology. 2009 Aug;27(8):735-41.
30. Available:https://saves.mbi.ucla.edu/
31. Audain E, Ramos Y, Hermjakob H, Flower DR, Perez-Riverol Y. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. Bioinformatics. 2016 Mar 15;32(6):821-7.
32. Sivakumar K, Balaji S, Gangaradhakrishnan. In silico characterization of antifreeze proteins using computational tools and servers. Journal of Chemical Sciences. 2007 Sep;119:571-9.
33. Gamage DG, Gunaratne A, Periyannan GR, Russell TG. Applicability of instability index for in vitro protein stability prediction. Protein and peptide letters. 2019 May 1;26(5):339-47.
34. Magdeldin S, Yoshida Y, Li H, Maeda Y, Yokoyama M, Enany S, Zhang Y, Xu B, Fujinaka H, Yaoita E, Sasaki S. Murine colon proteome and characterization of the protein pathways. BioData mining. 2012 Dec;5(1):1-4.
35. Magdeldin S, Yoshida Y, Li H, Maeda Y, Yokoyama M, Enany S, Zhang Y, Xu B, Fujinaka H, Yaoita E, Sasaki S. Murine colon proteome and characterization of the protein pathways. BioData mining. 2012 Dec;5(1):1-4.
36. Liu Z, Ma H, Goryanin I. A semi-automated genome annotation comparison and integration scheme. BMC bioinformatics. 2013 Dec;14:1-2.
37. Novere NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT. The systems biology graphical notation. Nature biotechnology. 2009 Aug;27(8):735-41.
38. MacLaughlin J, Holick MF. Aging decreases the capacity of human skin to produce vitamin D_3. The Journal of Clinical Investigation. 1985 Oct 1;76(4):1536-8.

Annexure 1:

Protein Sequence of 7 -dehydrocholestrol reductase gene in different species

>XP_016577971.2 7-dehydrocholesterol reductase [Capsicum annuum]
MGDSQLVHPPLFTYISMLTLLTIVPPFVILMWYTNVHADGSVLQTFNFLKENGLQGLIDIWPRPTAVAGK
IIICYALFEAILQLLLPGKRVEGPISPTGHRPVYKANGVAAYIVTLITYLSLWWFGIFNPTIVYDHLGEI
LSTLNFGSLIFCLFLYIKGHVAPSSTDHGSSGNIIIDYYWGMELYPRIGKHFDIKVFTNCRFGMISWAVL
PITYCIKQYEEYGNLSDSMLVHTIITLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVFLPCIYTSP
GMYLVKHPVNLGPQLAISILVAGILCVYINYDCDRQRQEFRKTNGKALVWGRPPSKIVASYTTTTGETKT
SLLLTSGWWGLSRHFHYVPEILASFFWCVPALFNHFIPYFYVIYLIVLLLDRAKRDDERCKSKYGKYWKK
YCDKVPYRVIPGIY

>AOX15301.1 sterol delta-7 reductase 2 [Withania somnifera]
MGETKLVHSPLLTYASMLTLLSFTPSFVILMWYTNVHADGSVLKTFNHLREHGLQGLIDIWPKPTAIAGK
IIIFYALFEAALQLLLPGKRVEGPISPTGHRPVYKANGMAAYAVTLITYISLWWFGIFTPAIVYDHLGEI
LSTLVFGSLVFCVFLYIKGHVAPSSTDSGSSGNIIVDFYWGMELYPRIGKHFDIKVFTNCRFGMMSWAVL
AVTYCIKQHEEYGRVSDSMLVNTILMLVYITKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWVPSIYTSP
GMYLVKQPVNLGLQLSLYILVAGLLCIYINYDCDRQRQEFRRTNGKCTVWGKAPSKIVATYTTTSGEKKT
SLLLTSGWWGLARHFHYVPEILAAFFWSVTALFNHFIPYFYVIFLTILLFDRAKRDDDRCKAKYGKYWKL
YCEKVPYRIIPGVY

>MCD7446296.1 7-dehydrocholesterol reductase [Datura stramonium]
MAETKLVHSPLITYGSMLTLLSFTPPFVILMWYTNVHADGSILKTFNHLRENGLQGLINIWPKPTAIAGK
LIICYALFEAALQLLLPGKRVEGPISPTGHRPVYKANGMAAYAVTLITYISLWWFGIFNPAIVYDHLGEI
FSTLIFGSLVFCVFLYIKGHVAPSSTDSGSSGNIIVDFYWGMELYPRIGKHFDIKVFTNCRFGMMSWAVL
AVTYCIKQHEEYGSVSDSMLVNTILMLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWLPCIYTSP
GMYLVKQPVNLGLQLSLYILVAGLLCIYINYDCDRQRQEFRRTNGKCTVWGKAPSKIVASYTTTSGEKKT
SLLLTSGWWGLARHFHYVPEILASFFWTVPALFNHFIPYFYVIYLTILLLDRAKRDDDRCKAKYGKYWKL
YCEKVPYRVVPGIY

>NP_001351.2 7-dehydrocholesterol reductase [Homo sapiens]
MAAKSQPNIPKAKSLDGVTNDRTASQGQWGRAWEVDWFSLASVIFLLLFAPFIVYYFIMACDQYSCALTG
PVVDIVTGHARLSDIWAKTPPITRKAAQLYTLWVTFQVLLYTSLPDFCHKFLPGYVGGIQEGAVTPAGVV
NKYQINGLQAWLLTHLLWFANAHLLSWFSPTIIFDNWIPLLWCANILGYAVSTFAMVKGYFFPTSARDCK
FTGNFFYNYMMGIEFNPRIGKWFDFKLFFNGRPGIVAWTLINLSFAAKQRELHSHVTNAMVLVNVLQAIY
VIDFFWNETWYLKTIDICHDHFGWYLGWGDCVWLPYLYTLQGLYLVYHPVQLSTPHAVGVLLLGLVGYYI
FRVANHQKDLFRRTDGRCLIWGRKPKVIECSYTSADGQRHHSKLLVSGFWGVARHFNYVGDLMGSLAYCL
ACGGGHLLPYFYIIYMAILLTHRCLRDEHRCASKYGRDWERYTAAVPYRLLPGIF

>NP_001077693.1 Ergosterol biosynthesis ERG4/ERG24 family [Arabidopsis thaliana]
MAETVHSPIVTYASMLSLLAFCPPFVILLWYTMVHQDGSVTQTFGFFWENGVQGLINIWPRPTLIAWKII
FCYGAFEAILQLLLPGKRVEGPISPAGNRPVYKANGLAAYFVTLATYLGLWFGIFNPAIVYDHLGEIFSA
LIFGSFIFCVLLYIKGHVAPSSSDSGSCGNLIIDFYWGMELYPRIGKSFDIKVFTNCRFGMMSWAVLAVT
YCIKQYEINGKVSDSMLVNTILMLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWVPSVYTSPGMY
LVNHPVELGTQLAIYILVAGILCIYINYDCDRQRQEFRRTNGKCLVWGRAPSKIVASYTTTSGETKTSLL
LTSGWWGLARHFHYVPEILSAFFWTVPALFDNFLAYFYVIFLTLLLFDRAKRDDDRCRSKYGKYWKLYCE

KVKYRIIPGIY

>NP_001353043.1 7-dehydrocholesterol reductase [Solanum lycopersicum]
MAESQLVHPPLFTYISMITLLTLVPPFVILMWYTNVHADGSVLQTYNYLKENGLQGLIDIWPRPTAIAGK
IIICYALFEATLQLLLPGKRVEGPISPTGHRPVYKANGMAAYTVTLITYLSLWWFGIFNPTIVYDHLGEI
LSTLNFGSLIFCLFLYIKGHVAPSSTDHGSSGNIIVDYYWGMELYPRIGKHFDIKVFTNCRFGMISWGLL
PITYCIKQYEEYGSLSDSMLIHTIITLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVFLPCMYTSP
GMYLVKHPVNLGPQLAISILVAGILCVYINYDCDRQRQEFRRTNGKALVWGKAPSKIVASYTTTTGETKT
SILLTSGWWGLSRHFHYVPEILASFFWSVPALLNHFMPYIYVIYLTGLLLDRAKRDDERCKSKYGKYWKK
YCEKVPYRVVPGIY

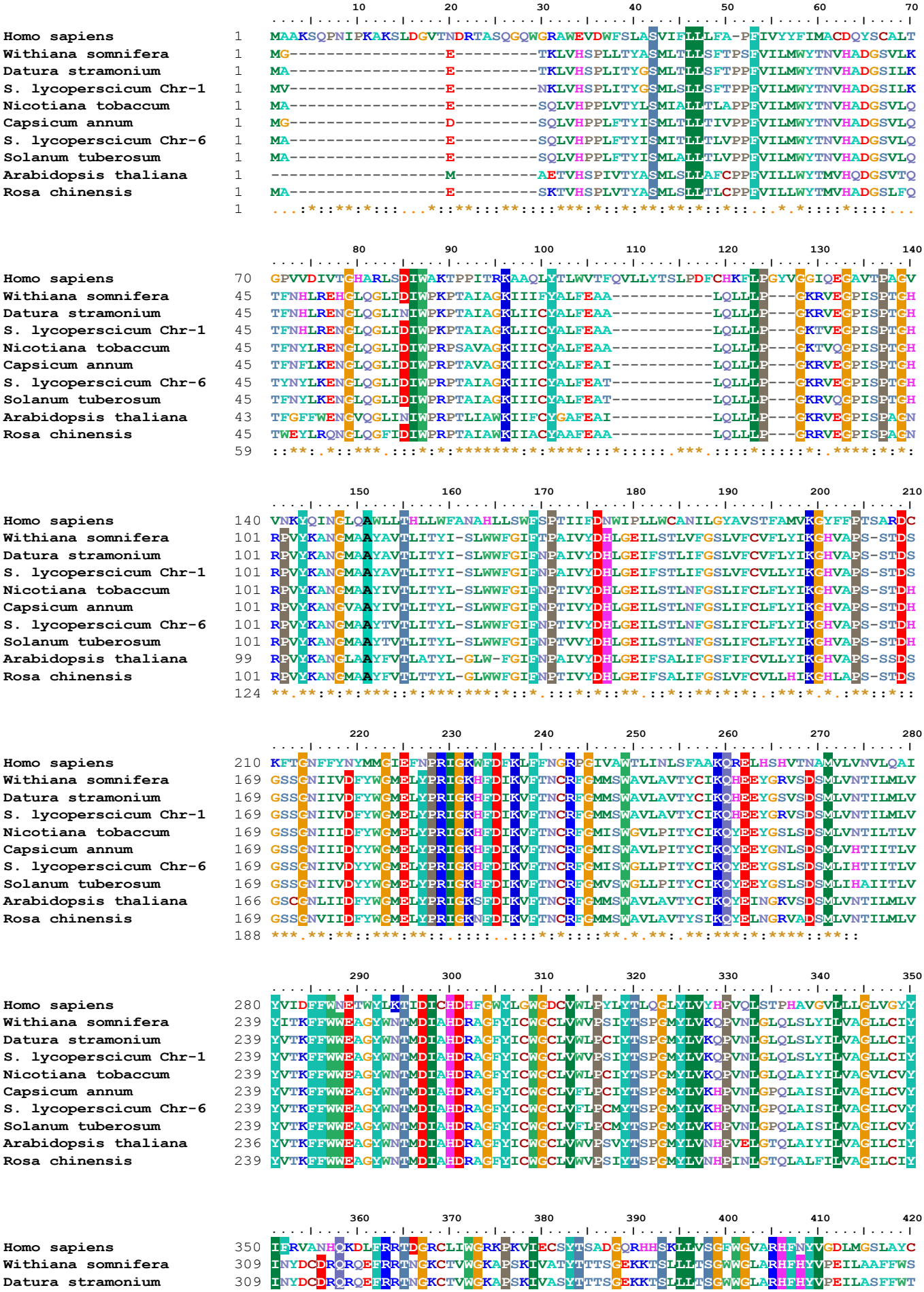
>NP_001305624.1 7-dehydrocholesterol reductase-like [Solanum tuberosum]
MAESQLVHPPLFTYISMLALLTLVPPFVILMWYTNVHADGSVLQTFNYLKENGLQGLIDIWPRPTAIAGK
IIICYALFEATLQLLLPGKRVQGPISPTGHRPVYKANGMAAYTVTLITYLSLWWFGIFNPTVVYDHLGEI
LSTLNFGSLIFCLFLYIKGHVAPSSTDHGSSGNIIVDYYWGMELYPRIGKHFDIKVFTNCRFGMVSWGLL
PITYCIKQYEEYGSLSDSMLIHAIITLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVFLPCMYTSP
GMYLVKHPVNLGPQLAISILVAGILCVYINYDCDRQRQEFRRTNGKALVWGKAPSKIVASYTTTTGETKS
SLLLTSGWWGLSRHFHYVPEILASFFWSVPALFNHIMPYFYVIYLTGLLLDRAKRDDERCKSKYGKYWKK
YCEKVPYRVIPGIY

>NP_001353044.1 7-dehydrocholesterol reductase [Solanum lycopersicum]
MVENKLVHSPLITYGSMLSLLSFTPPFVILMWYTNVHADGSILKTFNHLRENGLQGLIDIWPKPTAIAGK
LIICYALFEAALQLLLPGKTVEGPISPTGHRPVYKANGMAAYAVTLITYISLWWFGIFNPAIVYDHLGEI
FSTLIFGSLVFCVLLYIKGHVAPSSTDSGSSGNIIVDFYWGMELYPRIGKHFDIKVFTNCRFGMMSWAVL
AVTYCIKQHEEYGRVSDSMLVNTILMLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWVPSIYTSP
GMYLVKQPVNLGLQLSLYILVAGLLCIYINYDCDRQRQEFRRTNGKCTVWGKTPSKIVAAYTTTSGEKKT
SLLLTSGWWGLARHFHYVPEILAAFFWSVPALFNHFIPYFYVIFLIILLLDRAKRDDDRCKAKYGKYWKL
YCEKVPYRVIPGIY

>XP_024181789.1 7-dehydrocholesterol reductase [Rosa chinensis]
MAESKTVHSPLVTYASMLSLLTLCPPFVILLWYTMVHADGSLFQTWEYLRQNGLQGFIDIWPRPTAIAWK
IIACYAAFEAALQLLLPGRRVEGPISPAGNRPVYKANGMAAYFVTLTTYLGLWWFGIFNPTIVYDHLGEI
FSALIFGSLVFCVLLHIKGHLAPSSTDSGSSGNVIIDFYWGMELYPRIGKNFDIKVFTNCRFGMMSWAVL
AVTYSIKQYELNGRVADSMLVNTILMLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWVPSIYTSP
GMYLVNHPINLGTQLALFILVAGILCIYINYDCDRQRQEFRRTNGKCLVWGRAPSKIVASYTTTSGETKT
SLLLTSGWWGLSRHFHYVPEISAAFFWTVPALFNHFLPYFYVVFLTILLLDRAKRDDDRCRSKYGKHWKL
YCQKVPYRVIPGIY

>XP_016480318.1 PREDICTED: 7-dehydrocholesterol reductase-like [Nicotiana tabacum]
MAESQLVHPPLVTYLSMIALLTLAPPFVILMWYTNVHADGSVLQTFNYLRENGLQGLIDIWPRPSAVAGK
IIICYALFEAALQLLLPGKTVQGPISPTGHRPVYKANGMAAYIVTLITYLSLWWFGIFNPTIVYDHLGEI
LSTLNFGSLIFCLFLYIKGHVAPSSTDHGSSGNIIIDFYWGMELYPRIGKHFDIKVFTNCRFGMISWGVL
PITYCIKQYEEYGSLSDSMLVNTILTLVYVTKFFWWEAGYWNTMDIAHDRAGFYICWGCLVWLPCIYTSP
GMYLVKQPVNLGLQLAIYILVAGVLCVYINYDCDRQRQEFRRTNGKCLVWGKAPSKIVASYTTTTGETKT
SLLLTSGWWGLARHFHYVPEILASFFWSVPALFNHFIPYFYVIYLTILLFDRAKRDDDRCKSKYGKYWKL
YCEKVPYRVIPGIY

Multiple Sequence Alignment analysis of 7-dehydrocholesterol reductase



Annexure 2:

Amino acid composition of 7-dehydrocholestrol reductase in different plant species

Amino acids	Solanum lycopersicum (C-6)	Solanum tuberosum	Solanum lycopersicum Chr-1	Homo sapiens	Arabidopsis thaliana	Capsicum annum	Withiana somniferum	Datura stromatium	Nicotiana tobaccum	Rosa chinensis
Ala (A)	4.80%	5.30%	6.00%	7.80%	6.50%	4.6%	6.2%	6.0%	5.5	6.9
Arg (R)	3.50%	3.50%	3.50%	4.40%	3.90%	3.5%	3.7%	3.5%	3.5	4.6
Asn (N)	3.00%	3.00%	3.20%	3.60%	3.20%	3.2%	2.5%	3.2%	3.2	3.5
Asp (D)	3.20%	3.20%	3.50%	4.20%	3.50%	3.7%	3.5%	3.2%	3.5	3.5
Cys (C	2.50%	2.50%	2.50%	2.70%	3.00%	2.8%	2.3%	2.8%	2.8	2.5
Gln (Q)	1.80%	2.10%	1.60%	3.20%	1.90%	1.8%	1.6%	1.6%	2.3	2.1
Glu (E)	3.20%	3.00%	3.00%	1.70%	3.00%	2.8%	3.0%	3.0%	2.8	2.5
Gly (G)	7.60%	7.60%	7.40%	7.80%	7.70%	7.4%	7.4%	7.4%	7.4	7.1
His (H)	3.00%	3.00%	2.80%	3.60%	1.90%	3.0%	3.0%	2.8%	2.5	2.5
Ile (I)	8.80%	8.10%	8.30%	5.90%	8.10%	9.4%	7.8%	7.8%	8.5	7.1
Leu (L)	11.10%	11.30%	11.10%	11.60%	10.20%	10.8%	10.8%	11.1%	11.5	11.1
Lys (K)	4.80%	4.80%	5.30%	3.80%	4.20%	4.8%	5.3%	5.3%	4.4	3.7
Met(M)	2.50%	2.50%	2.50%	1.70%	2.30%	1.8%	2.5%	2.5%	2.1	2.5
Phe (F)	4.60%	5.10%	5.50%	6.70%	7.00%	5.5%	6.0%	5.5%	5.3	6.0
Pro (P)	5.50%	5.50%	4.80%	4.40%	4.40%	5.8%	4.4%	4.8%	5.3	4.8
Ser (S)	5.80%	6.00%	5.80%	4.40%	5.80%	5.3%	5.8%	5.8%	5.5	6.0
Thr (T)	6.90%	6.20%	5.80%	5.30%	5.60%	6.2%	6.7%	6.2%	6.2	6.2
Trp (W)	3.50%	3.50%	3.70%	4.20%	3.90%	3.5%	3.7%	3.7%	3.7	4.1
Tyr (Y)	7.60%	7.40%	6.70%	6.10%	6.70%	7.1%	6.7%	6.9%	7.4	6.2
Val (V)	6.20%	6.50%	7.10%	6.90%	7.20%	6.9%	7.1%	6.9%	6.7	6.9
Pyl (O)	0.00%	0.00%	0.00%	0.00%	0.00%	0.0%	0.0%	0.0%	0.0	0.0
Sec (U)	0.00%	0.00%	0.00%	0.00%	0.00%	0.0%	0.0%	0.0%	0.0	0.0