



Received on 24 December 2018; received in revised form, 08 July 2019; accepted, 13 August 2019; published 01 September 2019

COMPARATIVE PHYLOGENETIC ANALYSIS OF METHYLERYTHRITOL-4-PHOSPHATE (MEP) PATHWAY GENES IN ASTERAECAE FAMILY

Tiahs Saha and Sheela Chandra *

Department of Bio-Engineering, Birla Institute of Technology, Mesra, Ranchi - 835215, Jharkhand, India.

Keywords:

NCBI, MEP,
Rebaudioside A,
MSA, Phylogenetics

Correspondence to Author:

Dr. Sheela Chandra

Assistant Professor,
Department of Bio-Engineering,
Birla Institute of Technology, Mesra,
Ranchi - 835215, Jharkhand, India.

E-mail: schandra@bitmesra.ac.in

ABSTRACT: Stevioside glycosides (SGs) are a class of diterpene glycosides being produced by leaves of *Stevia rebaudiana*. *Stevia* is getting agricultural importance due to the production of natural sweetener steviol glycosides, also called as zero-calorie sweeteners. Earlier it was thought that steviol was synthesized from Kaurene *via* MVA pathway but later on it was concluded that precursors of steviol are actually synthesized *via* the plastid localized methylerythritol 4-phosphate (MEP) pathway. In total, seventeen enzymes are used in this biosynthetic pathway. Each enzyme plays a significant role in the subsequent production of rebaudioside A and other SGs. In order to understand those enzymes functionally, understanding of those genes evolution basically in the Asteraceae family is highly warranted. In this study, the National Center for Biotechnology Information (NCBI) was searched to retrieve amino acid sequences that encode MEP pathway enzymes homologs. A number of MEP pathway enzymes' amino acid sequences from the Asteraceae family were phylogenetically analyzed to light the way for the evolution characteristics of sweetener-producing plants homologs. The study is about to interpret the evolutionary relationships of *Stevia rebaudiana* with other plants in the same or different genus in the family Asteraceae.

INTRODUCTION: Isoprenoids are a class of natural products having important biological activities. All isoprenoids are constructed from two precursors, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). In the recent years, two most important discoveries in isoprenoid biosynthetic studies are the elucidation of a second pathway (the methylerythritol phosphate (MEP) pathway) for isoprenoid biosynthetic and a modified mevalonate (MVA) pathway.

Isoprenoids are the largest and structurally diverse groups of natural products with more than 55,000 members that play key metabolic and regulatory roles in all life forms. Several monoterpenes, sesquiterpenes, and diterpenes produced as secondary metabolites are pheromones or defensive agents and are involved in many aspects of our life including flavor, fragrance, medicine and nutrition¹⁸. Among that, a natural sweetening agent is Stevioside Glycosides (SGs) which is being produced from *Stevia rebaudiana*.

Stevia is getting agricultural importance due to the production of natural sweetener steviol glycosides, also called as zero-calorie sweeteners. The genus *Stevia* is comprised of approximately 200 species, which are distributed in North and South America. The phylogenetic relationships have been constructed using sequences of ITS and cpDNA

QUICK RESPONSE CODE 	DOI: 10.13040/IJPSR.0975-8232.10(9).4293-03
	The article can be accessed online on www.ijpsr.com
DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.10(9).4293-03	

and estimated the divergence times of the major clade of this genus. Results suggested that *Stevia* originated in Mexico 7.0–7.3 million years ago¹⁹. Tissue-specific chemical analyses confirmed that SGs were accumulated in leaf cells but not in trichomes²⁰. Several steviol glycosides are produced by *Stevia*. Rebaudioside A is produced in the largest concentrations but possesses a bitter aftertaste. Whereas rebaudioside D is reported to have a superior flavor, but are produced in much smaller concentrations. The leaves of *Stevia rebaudiana* (Bertoni) is reported to accumulate at least eight steviol glycosides (SGs). Due to the differential glycosylation process, each SG

possesses distinctive organoleptic properties. Stevioside is reported to be 143 times sweeter than sucrose on a weight basis, but Rebaudioside-A is 242 times sweeter³. The aftertaste quality of Rebaudioside-A is better than stevioside because it is sweeter and less bitter⁴. World Health Organization (WHO) has now recognized that stevioside is not genotoxic and has been assigned an acceptable daily intake of 0-2 mg SGs/kg body weight⁵. Long back studies on Steviol concluded that it was synthesized from kaurene, via the mevalonate pathway (MVA)^{6, 7, 8, 9} similar to the synthesis of several diterpenes.

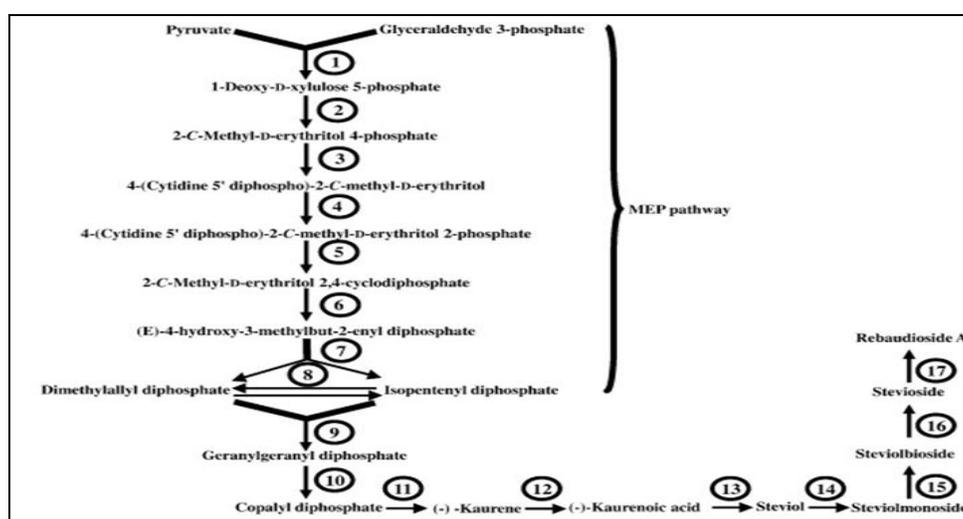


FIG. 1: MEP PATHWAY FOR PRODUCTION OF REBAUDIOSIDE A¹²

TABLE 1: THE NAME OF THE THOSE GENES OF THESE PATHWAYS AND THEIR ABBREVIATION IS GIVEN BELOW WITH THE COMPARISON AT THE PROTEIN LEVEL TO THEIR *ARABIDOPSIS* ORTHOLOGS

S. no.	Enzyes	GenBank Acc no.	% Identity/Similarity to Acc no.**
1	DXS	ACI43010.1	74/86 to Q38854
2	DXR	CAD22156.1	79/89 to AAF73140
3	CMS	ABB88837.2	82/91 to NP_565286.1
4	CMK	ABB88838.3	70/80 to NP_180261.1
5	MCS	ABG23395.1	77/87 to NP_850971
6	HDS	ABG75916.2	82/92 to AAM19840
7	HDR	ALJ30091.1	79/89 to AAN87171
8	GGDPS	ABD92926.2	63/80 to P34802
9	IDI*	--	-
10	CDPS*	--	-
11	KS	AAD34294.1	53/69 to NP_192187
12	KO	AAV42951.1	52/69 to AAC39443
13	KAH*	--	-
14	UGT85C2*	--	-
15	UGT*	--	-
16	UGT74G1*	--	-
17	UGT76G1	ACT33422.1	45/62 to NP_196207.1

(*for their absence in genebank databases those genes are excluded from our work.)

Deoxyxylulose-5-phosphate synthase (DXS), deoxyxylulose-5-phosphate reductoisomerase (DXR), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 4-diphosphocytidyl-2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS), 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate synthase (HDS) and 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate reductase (HDR), geranylgeranyldiphosphate synthase (GGDPS), copalyl diphosphate synthase (CDPS), kaurene synthase (KS), kaurene oxidase (KO), kaurenoic acid 13-hydroxylase (KAH), UDP-glycosyltransferase (UGT).

Later it was demonstrated that the precursors of steviol are actually synthesized via the plastid localized methylerythritol 4-phosphate (MEP) pathway **Fig. 1**.¹⁰ In total, seventeen enzymes are used in this biosynthetic pathway. Each enzyme plays a significant role in the subsequent production of rebaudioside A. In order to understand those enzymes functionally, understanding of those genes evolution basically in the Asteraceae family is highly warranted. The seventeen enzymes are enlisted in **Fig. 1**.¹¹

So in today's scenario, effort is required in order to find out the superior varieties which may produce a range of SGs or knowledge about the biosynthetic (MEP) pathway genes to undergo for flux change through metabolic engineering or even finding the relationship of *Stevia* with other plantlets in the same or different genus in the family Asteraceae. The elucidation of biochemical pathways to study plant secondary metabolism has been a long term interest in the scientific community. The emergence of genomic resources like whole genome sequencing and tools like high throughput expressed sequence tags (ESTs) have accelerated the understanding of these metabolic pathways¹.

In this respect, to understand exact relationships between *Stevia* genus, other similar plants members in the family Asteraceae, Steviol glycoside content and composition, more predictable approaches are necessary. To gain insights into this, using the computational tools could enable the researchers to understand the physical and biochemical properties of proteins and genes involved. So the present study has been undertaken to find out the similarities between members of the Asteraceae family and to compare the MEP pathway genes in the Asteraceae family.

MATERIALS AND METHODS:

A) Phylogenetic Analysis:

Retrieval of the Protein Sequences of Enzymes & Screening of the Databases: Each genes of *Stevia rebaudiana* are searched by doing BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) respect to the orthologs of *Arabidopsis sp.* Genbank accession number(as shown in **Table 1**¹¹ with the percentage of identity). The homology search of the genes was done by BLAST search tool of NCBI (<https://www.ncbi.nlm.nih.gov/BLAST/>) using blastp

algorithm with an amino acid sequence of Asteraceae (taxid4210) family. Different amino acid sequences were downloaded from GenBank Accession number of different species of Asteraceae family¹¹.

Multiple Sequence Alignment: All protein sequences of genes were aligned using MUSCLE by MEGA software (version 7.0.21) to find out the similarity present among the sequences of the same family¹².

Phylogenetic Tree Construction: Phylogenetic analysis of the sequences was done by Molecular Evolutionary Genetics Analysis (MEGA)¹³ software (version 7.0.21), using Neighbor-joining¹⁴ method. Each node was tested using the bootstrap approach by taking 100.¹⁵ The branch lengths were drawn to scale indicated.

B) Motifs' Search:

Retrieval of the Protein Sequences of Enzymes & Screening of the Databases: Each gene of *Stevia rebaudiana* are searched by doing BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) respect to the orthologs of *Arabidopsis sp.* Genbank accession number (as shown in **Table 1** with percentage of identity). The homology search of the genes was done by BLAST search tool of NCBI (<https://www.ncbi.nlm.nih.gov/BLAST/>) using blastp algorithm with amino acid sequence of Asteraceae (taxid4210) family. Different amino acid sequences were downloaded from genbank Accession number of different species of Asteraceae family¹¹.

Finding Motifs: Motif search is done by MEME-suite (<http://meme-suite.org/>)¹⁷ of those protein sequences consequently by keeping one occurrence per sequence in select site distribution option and thirty-five in the option of select the numbers of motifs and other parameters are taken as customized in that page.

RESULTS AND DISCUSSION:

A) Phylogenetic Analysis: In this study, amino acid sequences of a total of eleven enzymes belonging to the MEP pathway were obtained from NCBI Genbank and analyzed. The neighbor-joining method was utilized to analyze the evolutionary history¹⁴. MEGA7 was used to perform evolutionary analyses¹³. According to the neighbor-joining algorithm, groups and subgroups

were defined per protein in phylogenetic tree¹⁶. Convergence and divergence are two essential phylogenetic properties, which could be useful to identify the closely as well as distantly related groups of MEP pathway proteins. These findings suggest that all tested enzymes are conserved and all of the analyzed sequences evolved from a common ancestral enzyme. Each amino acid sequences of MEP pathway enzymes was first selected and screened, aligned and then subjected for phylogenetic analysis. So, a total number of eleven enzymes were taken and used for the final results processing. The results of each are showing and elaborating one after another.

No. (i) Deoxyxyulose-5-phosphate Synthase (DXS) Enzyme: The DXS enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to transketolase C super-family. From phylogenetic tree below based on the amino acid sequences of DXS enzyme, it is observed that *Stevia rebaudiana* and *Helianthus annuus* are too closely related and diversified from the same origin. Similarly, *crysanthemum x morifolium* and *Taraxacum kok-saghyz* are diversified from the same origin and are hence closely related to *Cynara cardunculus* var. *scolymus* related *Atractylodes ianica* are showing some similarity which means it is diversified from rest of species of Asteraceae family.

No. (ii)- Deoxyxyulose-5-phosphate Reducto-Isomerase (DXR) Enzyme: The DXR enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to deoxyxyulose-5-phosphate reductoisomerase C super-family.

Similar sequences are randomly selected for multiple sequence alignment and phylogenetic analysis. Total eleven different species of Asteraceae family were selected from blastp result. For phylogenetic tree based on the protein sequences of DXR enzyme, *Stevia rebaudiana* is closely related to *Taraxacum kok-saghyz* and *Helianthus annuus*; they are then more diversified to *Atractylodes ianica* and *Argentinan adenophora*. The both *Chrysanthemum* sp., *Artemisia* sp. and

Achillea sp. diversified from each other from the same origin; which means they are distantly related from *S. rebaudiana* along with *Tenacetum parthenium*.

No. (iii) 4- diphosphocytidyl-2-C -methyl-D-Erythritol Synthase (CMS) Enzyme: The CMS enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to the Asteraceae family. The blast results depict that the enzyme belongs to glycosyl-transferase super-family.

Sequences above and are 99% were selected for multiple sequence alignment and phylogenetic analysis. Total five different species of Asteraceae family were selected from blastp result. For phylogenetic tree based on the protein sequence of CMS enzyme, *Stevia rebaudiana* is closely related to *Taraxacum kok-saghyz* then diversified with *Cynara cardunculus* var. *scolymus*. Both *Helianthus annuus* and *Artemisia annua* are distantly related.

No. (iv) 4- diphosphocytidyl-2-C -methyl-D-Erythritol Kinase (CMK) Enzyme: The CMK enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to IspE superfamily.

Multiple sequence alignment and phylogenetic analysis are done with selected five sequences. For phylogenetic tree based on the protein sequence of CMK enzyme, *Stevia rebaudiana* is closely related to *Helianthus annuus* then to *Taraxacum kok-saghyz* then more diversified with *Cynara cardunculus* var. *scolymus*; *Artemisia annua* is distantly related.

No. (v) 4- diphosphocytidyl-2-C -methyl-D-Erythritol 2, 4-Cyclodiphosphate Synthase (MCS) enzyme: The MCS enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to MECDP synthase super-family.

Sequences which shows higher sequence similarity with six members of the family Asteraceae of MCS amino acid sequence of *Stevia rebaudiana*. Multiple sequence alignment and phylogenetic analysis are done with selected six sequences.

For phylogenetic tree based on the protein sequences of MCS enzyme, it is observed that *Stevia rebaudiana* is closely related to *Ageratina adenophora*; these two species are related to *Taraxacum kok-saghyz* and collectively these three species are related to *Helianthus annuus* and then these four species are related to *Cynara cardunculus var. scolymus*; *Artemisia annua* is distantly related *Stevia sp.*

No. (vi) 1-hydroxy-2-methyl-2(E)-butenyl 4-Diphosphate Synthase (HDS) Enzyme: The HDS enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with

respect to Asteraceae family. The blast results depict that the enzyme belongs to GcpE protein super-family.

Sequences above 80% were selected for multiple sequence alignment and phylogenetic analysis. Total five different species of Asteraceae family were selected from blastp result. For phylogenetic tree based on the protein sequences of HDS enzyme, it is observed that *Stevia rebaudiana* is closely related to *Artemisia annua* and these to are related with *Helianthus annuus*. *Taraxacum kok-saghyz* and *Cynara cardunculus var. scolymus* are distantly related to *Stevia sp.*

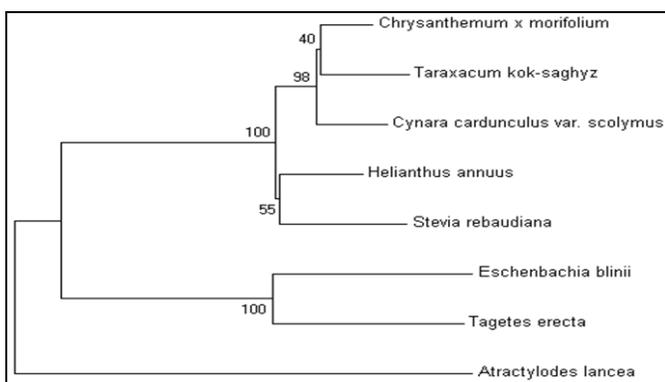


FIG. 2: PHYLOGENETIC TREE BASED ON DXS ENZYME PROTEIN SEQUENCES

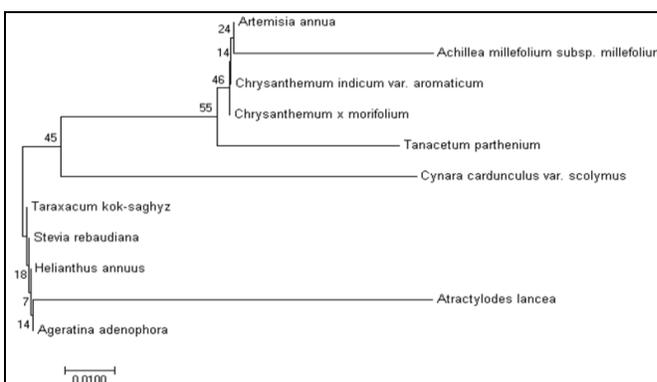


FIG. 3: PHYLOGENETIC TREE BASED ON DXR ENZYME PROTEIN SEQUENCES

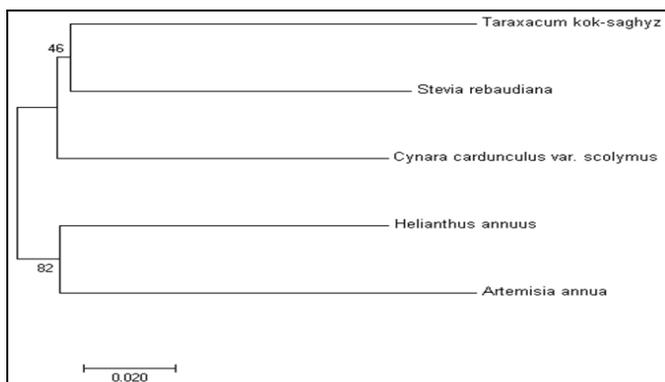


FIG. 4: PHYLOGENETIC TREE BASED ON CMS ENZYME PROTEIN SEQUENCES

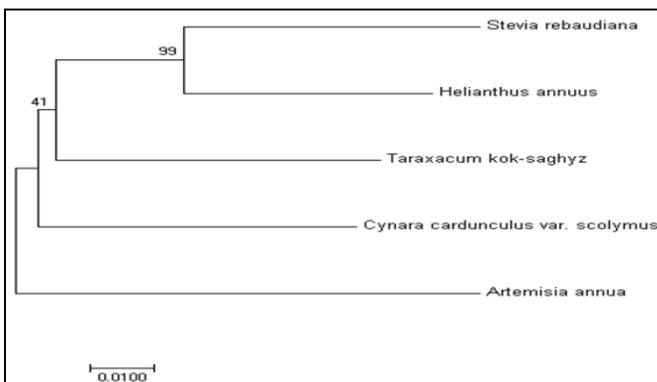


FIG. 5: PHYLOGENETIC TREE BASED ON CMK ENZYME PROTEIN SEQUENCES

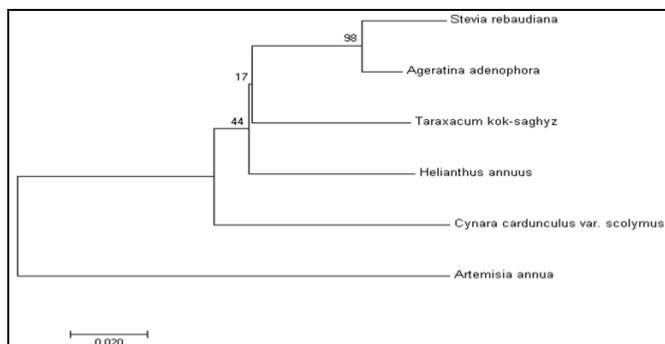


FIG. 6: PHYLOGENETIC TREE BASED ON MCS ENZYME PROTEIN SEQUENCES

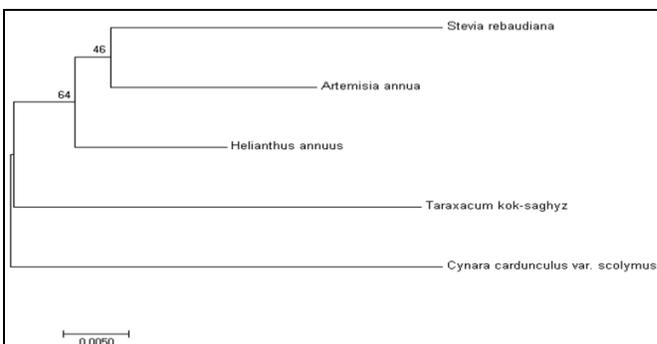


FIG. 7: PHYLOGENETIC TREE BASED ON HDS ENZYME PROTEIN SEQUENCES

No. (vii) 1-hydroxy-2-methyl-2(E)-butenyl 4-Diphosphate Reductase (HDR) Enzyme: The HDR enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to protein lytB and ispH super-family. Sequences above 99% were selected for multiple sequence alignment and phylogenetic analysis. Total six different species of asteraceae family were selected from blastp result. For phylogenetic tree based on the protein sequence of HDR enzyme, it is observed that *Stevia rebaudiana* is closely related to *Tanacetum parthenium* and *Artemisia annua* distantly related with *Cynara cardunculus* var. *scolymus*.

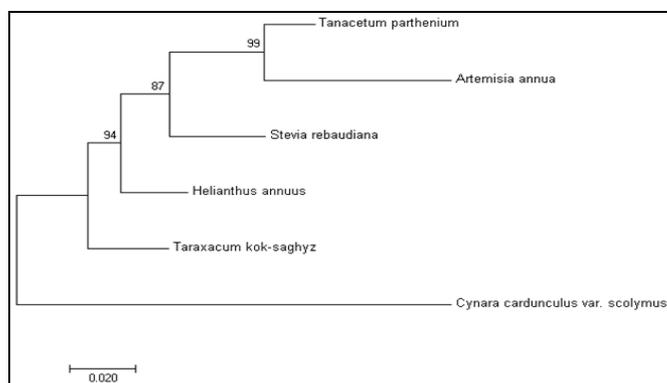


FIG. 8: PHYLOGENETIC TREE BASED ON HDR ENZYME PROTEIN SEQUENCES

No. (viii) Geranyl-geranyl-diphosphate Synthase (GGDPS) Enzyme -The GGDPS enzyme of *Stevia* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to protein isoprenoid biosynthesis class 1 super-family. Sequences above 100% identity were selected for multiple sequence alignment and phylogenetic analysis. Total of eight different species of Asteraceae family were selected from blastp result. For phylogenetic tree based on the amino acid sequence of GGDPS enzyme, it is observed that *Stevia rebaudiana* is distantly related to any other species of Asteraceae family.

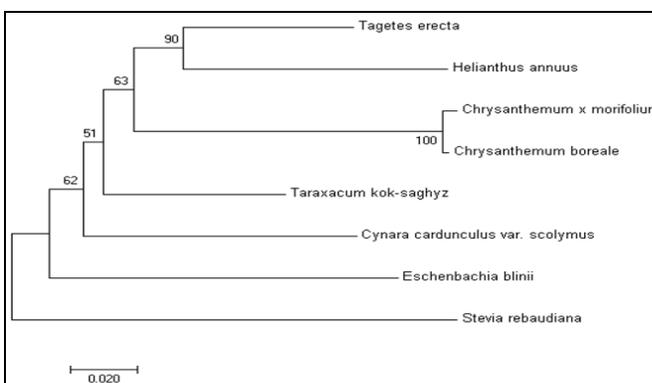


FIG. 9: PHYLOGENETIC TREE BASED ON GGDPS ENZYME PROTEIN SEQUENCES

No. (ix) Kaurene Synthase (KS) Enzyme: The KS enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to terpene synthase super-family.

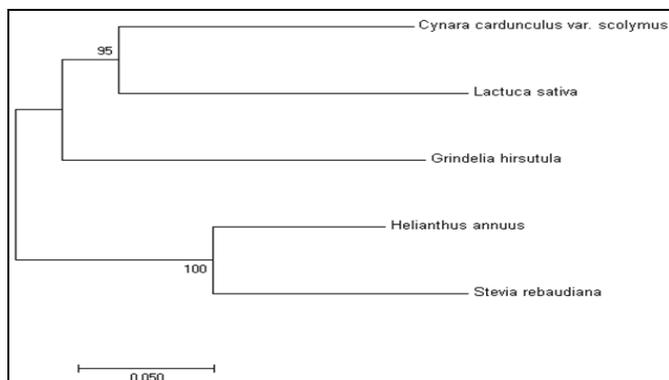


FIG. 10: PHYLOGENETIC TREE BASED ON KS ENZYME PROTEIN SEQUENCES

Sequences above 95% identity were selected for multiple sequence alignment and phylogenetic analysis. Total five different species of Asteraceae

family were selected from blastp result. For phylogenetic tree based on the amino acid sequence of KS enzyme, it is observed that *Stevia rebaudiana* is closely related to *Helianthus annuus* and distantly related to *Lactuca sativa* and *Cynara cardunculus* var. *scolymus*.

No. (x) Kaurene Oxidase (KO) Enzyme: The KO enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to terpene synthase super-family.

Sequences above 95% identity were selected for multiple sequence alignment and phylogenetic analysis. Total five different species of Asteraceae family were selected from blastp result. For phylogenetic tree based on the protein sequence of KO enzyme, it is observed here that *Stevia* is closely related with *Helianthus annuus* and these both are diversified from *Lactuca sativa*; distantly related with *Tanacetum cinerariifolium*.

No. (xi) UDP-glycosyltransferase 76G1 (UGT 76G1) Enzyme: The UGT76G1 enzyme of *Stevia rebaudiana* with a particular accession number was subjected for blastp with respect to Asteraceae family. The blast results depict that the enzyme belongs to glycosyltransferase GTB type super-family. Sequences above 98% identity were selected for multiple sequence alignment and phylogenetic analysis. Total five different species of Asteraceae family were selected from blastp

result. For phylogenetic tree based on the amino protein sequence of UGT76G1 enzyme, by different enzymes of species *Helianthus annuus*; selected from the alignment of the blastp result to allow the bootstrap process for tree construction. Here, we can see that UGT76B1 amino acid sequence is closely related to UGT76G1 of *Cynara cardunculus* var. *scolymus* but UGT76G1 of both *Stevia rebaudiana* and *Helianthus annuus* are diversified from the same origin.

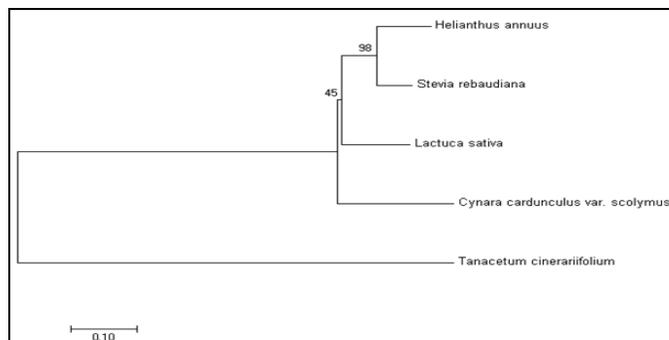


FIG. 11: PHYLOGENETIC TREE BASED ON KO ENZYME PROTEIN SEQUENCES

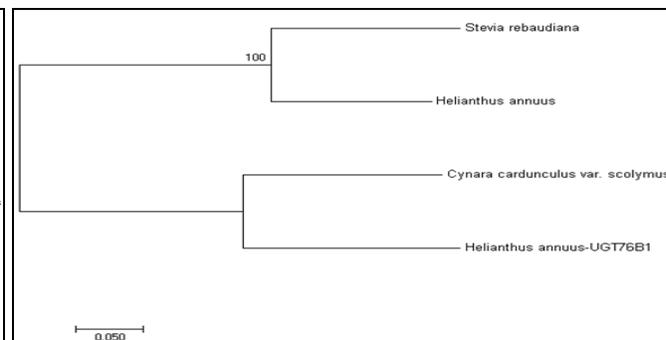


FIG. 12: PHYLOGENETIC TREE BASED ON UGT76G1 ENZYME PROTEIN SEQUENCES

B) Motifs' Search:

(i) For enzymes, **Deoxyxyulose-5-phosphate synthase (DXS) & Deoxyxyulose-5-phosphate reductoisomerase (DXR) & Geranyl-geranyl-diphosphate synthase (GGDPS):** Fig. 13A depicts the thirty-five motifs' locations in each protein sequences of three enzymes with their consequent Genbank accession number with a p-value ranging from $1.58e-74$ to $2.55e-119$. Some of the motifs of these three enzyme sequences are shown below.

The five different motifs as raw are-

No.1 TGIVGC	No.2 VHYDLWCRDY	No.3 FQVVDD
MCIAAC	VHDDL PAMDN	DQVVDD
No.4 YYIGPV	No.5 MMFKNEK	
NWDGPK	EMFIDDK	

In Fig. 13B some motifs which were found using MEME-suite; These five motifs are selected on the basis of E-value ranging from $1.0e+003$ to $1.1e+003$; varying with a length of six to ten; the different colors indicate the presence of different amino acids in the protein sequences.

(ii) For enzymes **4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS) & 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK):** Fig. 14A depicts the thirty-five motifs' locations on two protein sequences of CMS

& CMK enzymes with a p-value ranging between $3.95e-134$ to $5.25e-127$ also along with their accession numbers.

The five different motifs as raw are-

No.1 QLVCVND	No.2 YTTSRMRI VQEI	No.3 MSTMNA
EDVCIHD	YCFGRGEEVKDI	MAPLAA
No.4 QSSYTN	No.5 DSFVPQ	
QGNRYN	DSFANA	

The motifs in Fig. 14B are found using MEME-suite; These five motifs are selected on the basis of E-value ranging from $1.0e+002$ to $1.3e+004$; varying with a length of six to twelve; the different colors indicate the presence of different amino acids in the protein sequences.

(iii) For Enzymes **1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate reductase (HDR) & 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate synthase (HDS):** The thirty-five motifs located on HDR & HDS enzymes' protein sequences have a p-value ranging from $1.83e-159$ to $3.33e-148$.

The different motifs as raw are-

No.1 QRVNPGN	No.2 NNHFLS	No.3 HPIRNQT
IRIGPGN	DIHFAP	AGIAIQT
No.4 VQGWN Y	No.5 YYEYTW	
VGGWDS	NCESTH	

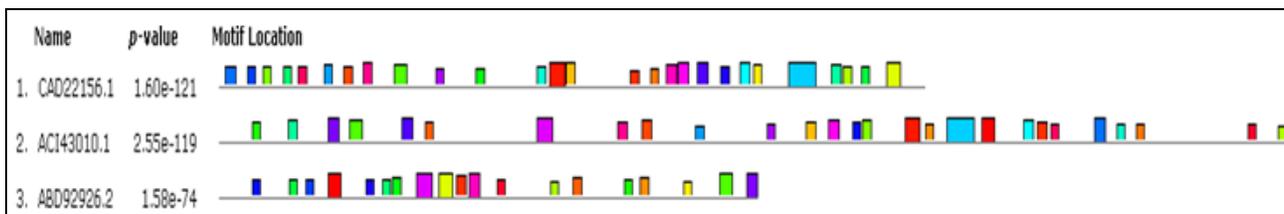


FIG. 13A: MOTIF LOCATIONS ON DXS, DXR & GGDPs ENZYMES' PROTEIN SEQUENCES WITH THEIR ACCESSION NUMBER

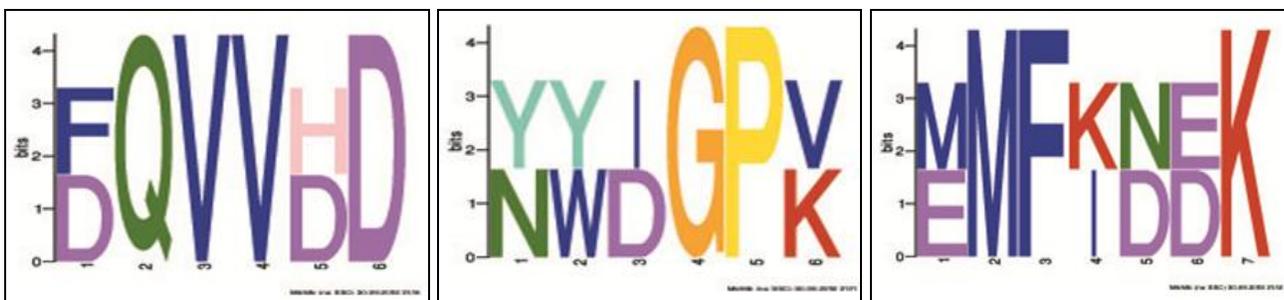
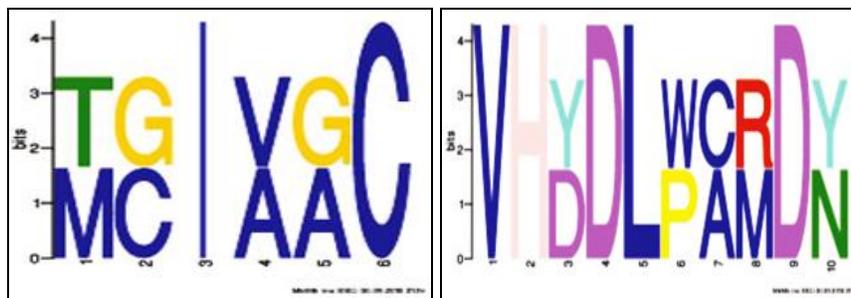


FIG. 13B: 5 DIFFERENT MOTIFS OF DXS, DXR & GGDPs ENZYMES' PROTEIN SEQUENCES

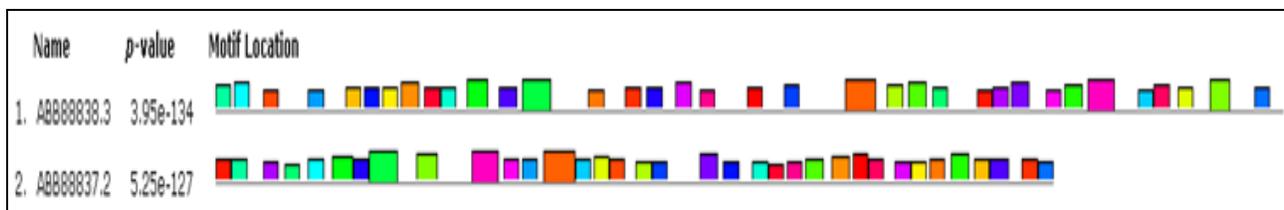


FIG. 14A: MOTIF LOCATIONS ON CMK, CMS ENZYMES' PROTEIN SEQUENCES WITH THEIR ACCESSION NUMBER

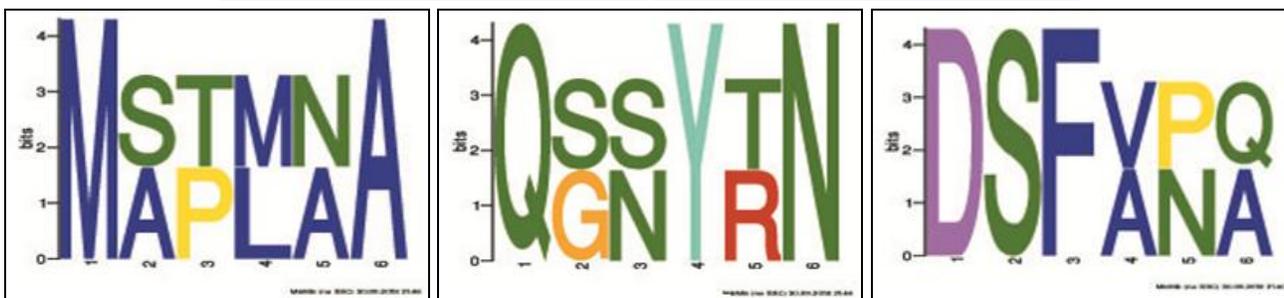
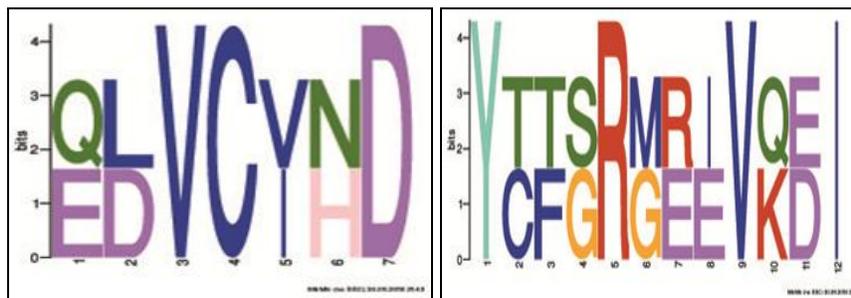


FIG. 14B: 5 DIFFERENT MOTIFS OF CMK, CMS ENZYMES' PROTEIN SEQUENCES

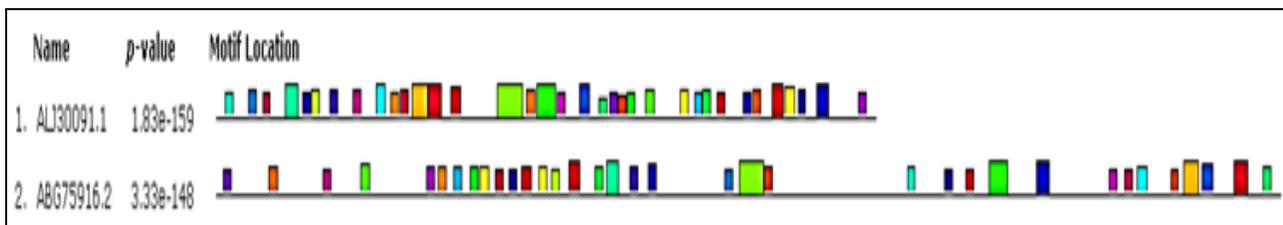


FIG. 15A: MOTIF LOCATIONS ON HDR, HDS ENZYMES' PROTEIN SEQUENCES WITH THEIR ACCESSION NUMBER

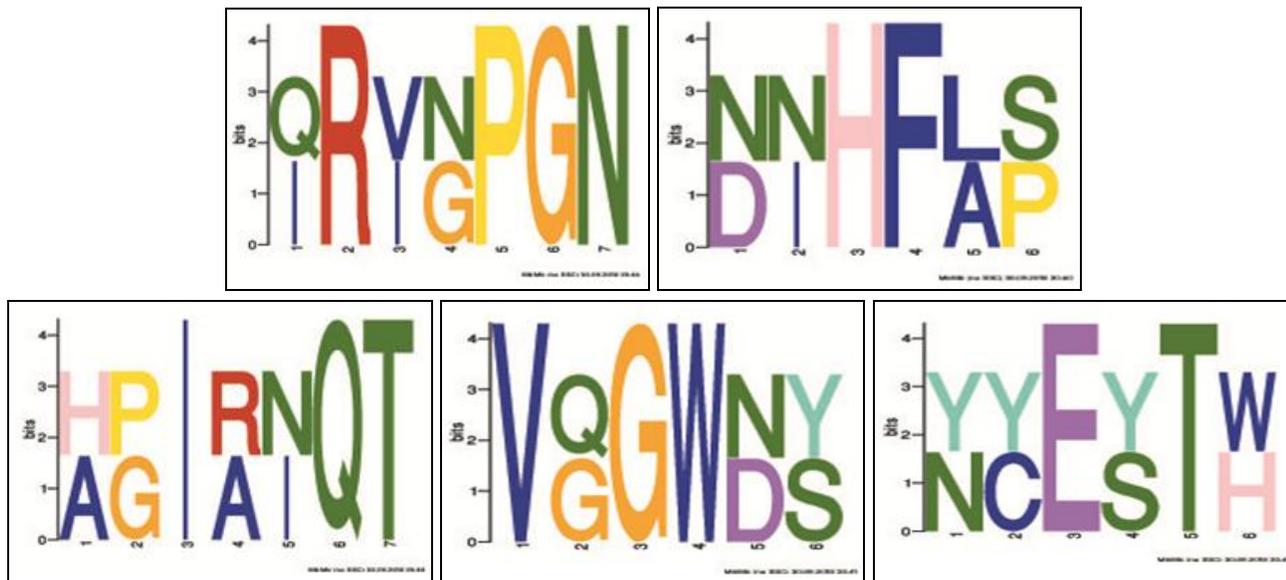


FIG. 15B: 5 DIFFERENT MOTIFS OF HDR, HDS ENZYMES' PROTEIN SEQUENCES

These five motifs are selected on the basis of E-value ranging from 1.1×10^3 to 1.8×10^1 ; varying with a length of six to seven Fig 15A and B.

(iv) For enzymes **Kaurene oxidase (KO) & Kaurene synthase (KS)**: The thirty-five motifs located on KO & KS enzymes' protein sequences have a p-value ranging from 1.04×10^{-134} to 2.30×10^{-144} .

The different motifs as raw are-

- | | | |
|-------------------------|--------------------------|-----------------------|
| No.1 DTTWVAM
DKAMVAM | No.2 YSMKTG
YQIKNG | No.3 GGYTVP
GGQHNP |
| No.4 FWNMYH
EWAMCE | No.5 DFPPYLNW
CFECLKW | |

These five motifs are selected on the basis of E-value ranging from 1.1×10^2 to 1.9×10^1 ; varying with a length of six to eight Fig 16A and B.

(v) For enzymes **UDP-glycosyltransferase 76G1 (UGT76G1) & 4-diphosphocytidyl-2-C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase (MCS)**: The thirty-five motifs located on UGT76G1 & MCS enzymes' protein sequences have a p-value ranging from 2.33×10^{-117} to 5.41×10^{-135} .

The different motifs as raw are-

- | | | |
|-------------------------|-----------------------|---------------------------|
| No.1 GELRRP
DALGLE | No.2 QTKASS
KGAASS | No.3 ATSSWYYT
ADQISCFT |
| No.4 ARYMSDV
AHS GDV | No.5 SIHAHV
NFAAHT | |

These five motifs are selected on the basis of E-value ranging from 1.0×10^5 to 1.7×10^3 ; varying with a length of six to eight Fig 17A and B.

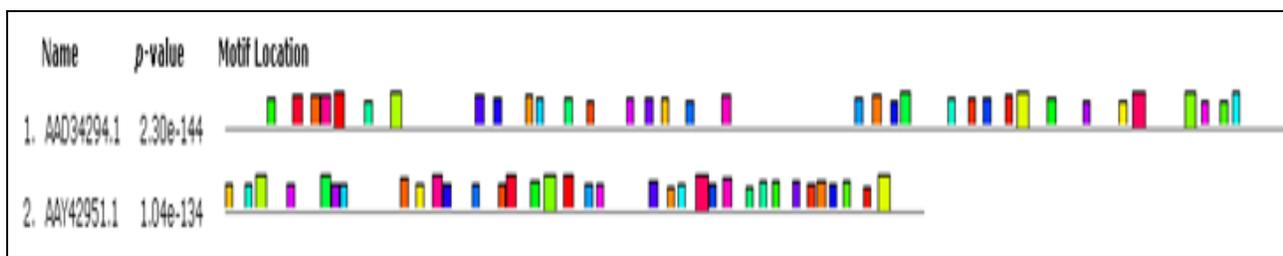


FIG. 16A: MOTIF LOCATIONS ON KS, KO ENZYMES' PROTEIN SEQUENCES WITH THEIR ACCESSION NUMBER

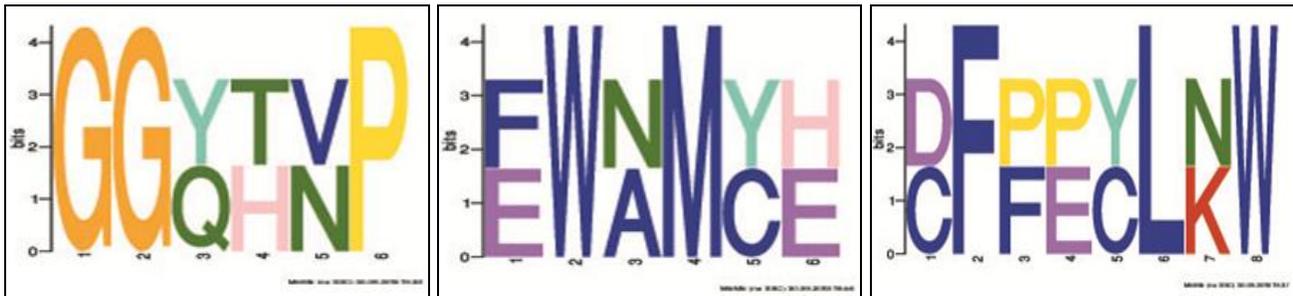
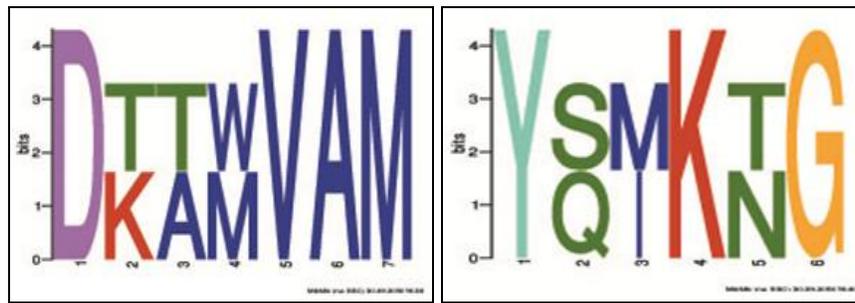


FIG. 16B: 5 DIFFERENT MOTIFS OF KS, KO ENZYMES' PROTEIN SEQUENCES

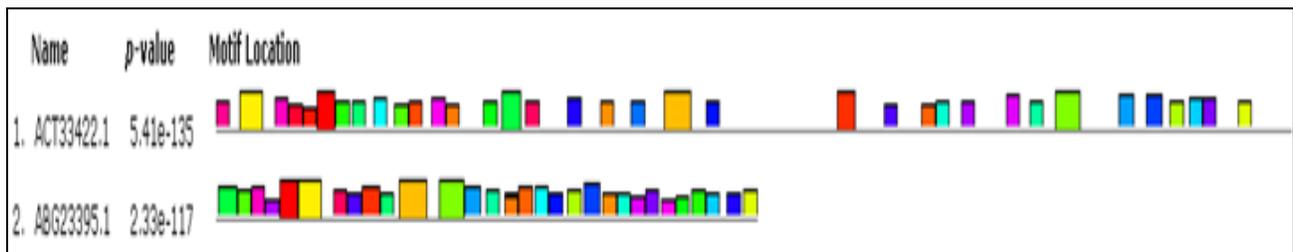


FIG. 17A: MOTIF LOCATIONS ON MCS, UGT76G1 ENZYMES' PROTEIN SEQUENCES WITH THEIR ACCESSION NUMBER

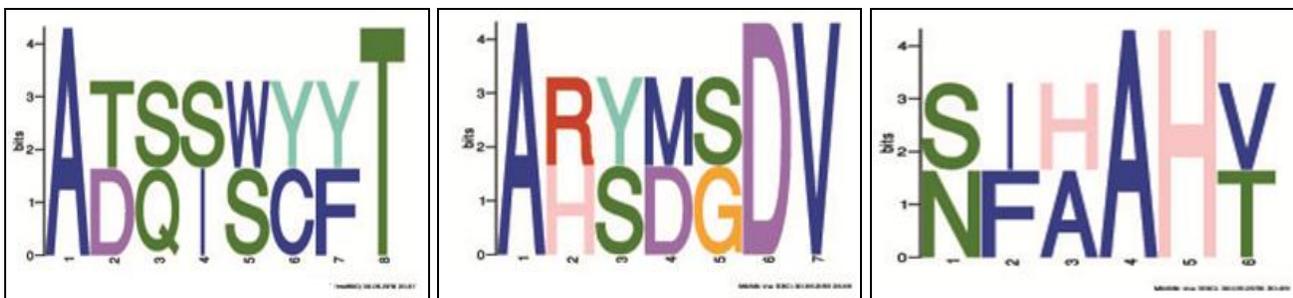
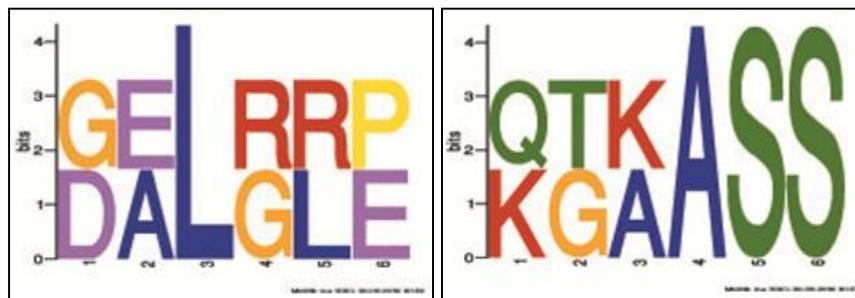


FIG. 17B: 5 DIFFERENT MOTIFS OF MCS, UGT76G1 ENZYMES' PROTEIN SEQUENCES

CONCLUSION: In conclusion, amino acid sequence analysis of Methylerythritol-4-phosphate pathway proteins showed that these plants have been related together as they evolved together with conserved regions in sweetener production; there are some common species of Asteraceae family

e.g., *Helianthus annuus* is common in all the phylogenetic trees; the other common species are *Cynara cardunculus* var. *scolymus* and *Taraxacum kok-saghyz*. There are a bunch of motifs found in motifs' search with two or three amino acids same and with a p-value ranging from 1.04e-134 to

5.41e-135. Motifs which are shown here mostly are with higher ranges from 1.0e+002 to 1.09e+001.

ACKNOWLEDGEMENT: The authors acknowledge BTISnet, Department of Bio-Engineering, Birla Institute of Technology, Mesra, Ranchi for providing Infrastructure and internet facility for Bioinformatics research.

CONFLICT OF INTEREST: The authors confirm that this article content has no conflicts of interest.

REFERENCES:

1. Soejima A, Tanabe AS, Takayama I, Kawahara T, Watanabe K, Nakazawa M, Mishima M and Yahara TL: Phylogeny and biogeography of the genus *Stevia* (Asteraceae: Eupatorieae): an example of diversification in the Asteraceae in the new world. *J Plant Res* 2017; 130: 953-72.
2. Kim MJ, Jin J, Zheng J, Wong L, Chua NH and Jang IC: Comparative transcriptomics unravel biochemical specialization of leaf tissues of *Stevia* for diterpenoid production. *Plant Physiology* 2015; 169: 2462-80.
3. Rodriguez-Concepcion M and Boronat A: Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol* 2004; 130: 1079-89.
4. Soejarto DD, Kinghorn AD and Farnsworth NR: Potential sweetening agents of plant origin. III. Organoleptic Evaluation of *Stevia* Leaf Herbarium Samples For Sweetness 1982; 45(5): 590-99.
5. Kasai R, Kaneda D, Tanaka O, Yamazaki K, Tsukamoto M, Morimoto K, Okada S, Kitahata S and Furukawa H: Sweet diterpene-glycosides of leaves of *Stevia rebaudiana* Bertoni. Synthesis and structure-sweetness relationship of rebaudiosides-A, D, E and their related glycosides of *Stevia* leaves and correlation between sweetness and chemical structure 1981; 5: 726-35.
6. DuBois GE and Stephenson RA: Diterpenoid sweeteners. Synthesis and sensory evaluation of stevioside analogues with improved organoleptic properties. *J Med Chem* 1985; 28: 93-98.
7. Beneford DJ, DiNovi M and Schlatter J: Steviol glycosides; safety evaluation of certain food additives, In: WHO food additive series, WHO, Geneva, 2006: 117-143.
8. Mosettig E, Beglinger U, Dolder F, Lichti H, Quitt P and Waters JA: The absolute configuration of steviol and isosteviol. *J Am Chem Soc* 1963; 85: 2305-09.
9. Ruddat M, Heftmann E and Lang A: Biosynthesis of steviol. *Archive of Biochemistry & Biophysics* 1965; 110(3): 496-99.
10. Bennett RD, Lieber ER and Heftmann E: Biosynthesis of steviol from (-)-kaurene. *Phytochem* 1967; 6(8): 1107-10.
11. Hanson JR and White AF: Studies in terpenoid biosynthesis-II: the biosynthesis of steviol. *Phytochemistry* 1968; 7(4): 595-97.
12. Kumar H, Kaul K, Bajpai-Gupta S, Kaul VK and Kumar SK: A comprehensive analysis of fifteen genes of steviol glycosides biosynthesis pathway in *Stevia rebaudiana* (Bertoni). *Gene* 2012; 492(1): 276-84.
13. Brandle JE and Telmer PG: Steviol glycoside biosynthesis. *Phytochemistry* 2007; 68(14): 1855-63.
14. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004; 32(5): 1792-97.
15. Kumar S, Stecher G and Tamura K: MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 2007; 33(4): 1870-74.
16. Saitou N and Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 1987; 4(4): 406-425.
17. Felsenstein J: Confidence limits on phylogenies: An approach using the bootstrap. *Evol* 1985; 39(4): 783-91.
18. Bailey TL and Elkan C: Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 1994; 2: 28-36.
19. Zhao L, Chang WC, Xiao Y, Liu HW and Liu P: Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu Rev Biochem* 2013; 82: 497-30.

How to cite this article:

Saha T and Chandra S: Comparative phylogenetic analysis of methylerythritol-4-phosphate (MEP) pathway genes in Asteraceae family. *Int J Pharm Sci & Res* 2019; 10(9): 4293-03. doi: 10.13040/IJPSR.0975-8232.10(9).4293-03.

All © 2013 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **ANDROID OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)