

Weight Matrix Based Identification of Terpene Synthases Conserved Motifs in *Arabidopsis thaliana* Proteome

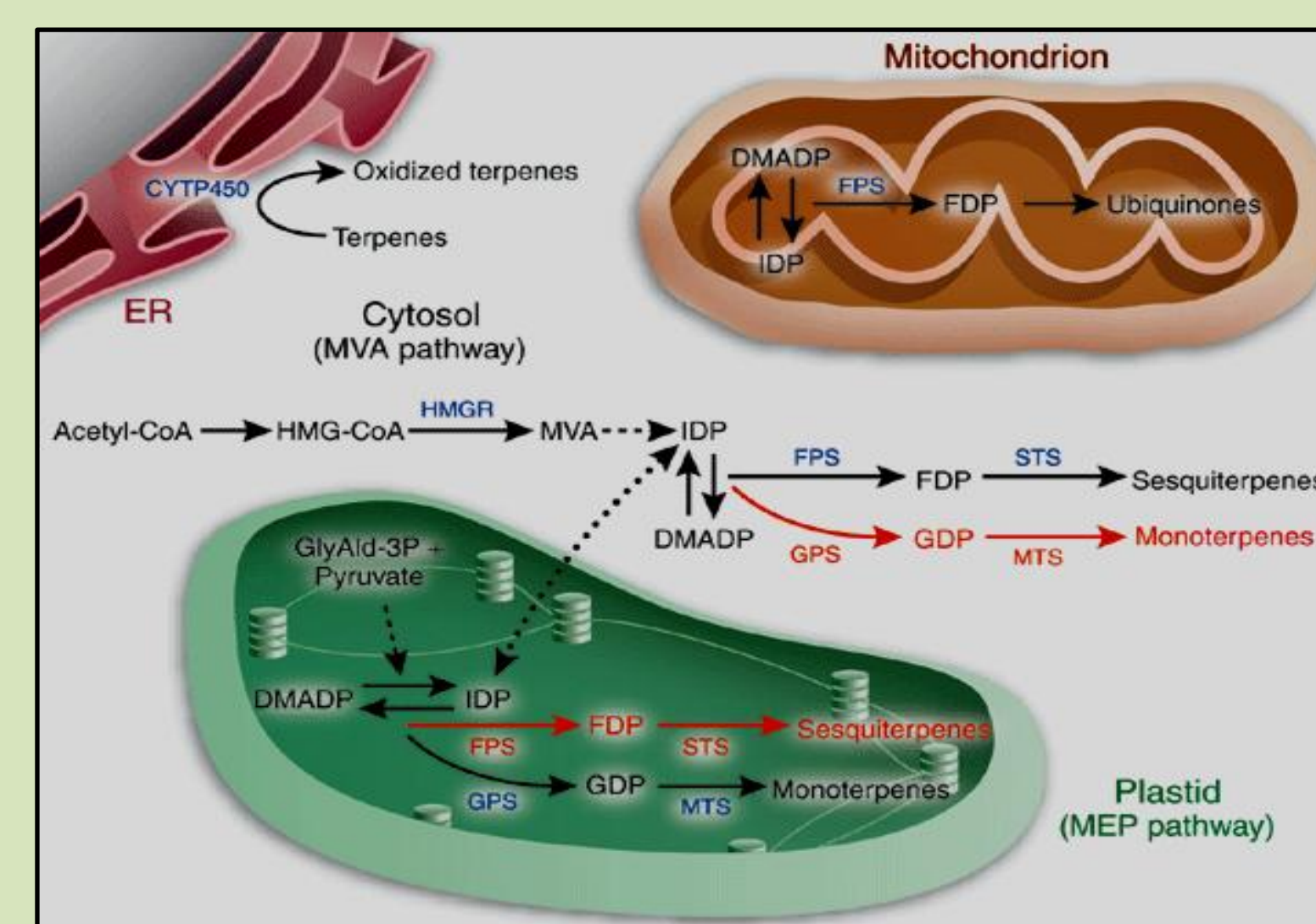
Suchita Srivastava, Suaib Luqman*, Feroz Khan

Central Institute of Medicinal & Aromatic Plants (CSIR), Lucknow



INTRODUCTION

- Terpenes comprise the most diverse collection of secondary metabolites and are the compounds that give wonderful aromas, tastes and pharmaceuticals.
- Terpene hydrocarbon scaffolds are generated by the action of mechanistically intriguing family of mono-, sesqui-, and diterpene synthases collectively termed as 'Terpene synthases'.
- Terpene synthases (TPS) promotes the Mg^{++} dependent expulsion of the pyrophosphate group from the substrate leading to formation of carbocations which further go through cyclization to form various terpenes.



METHODOLOGY

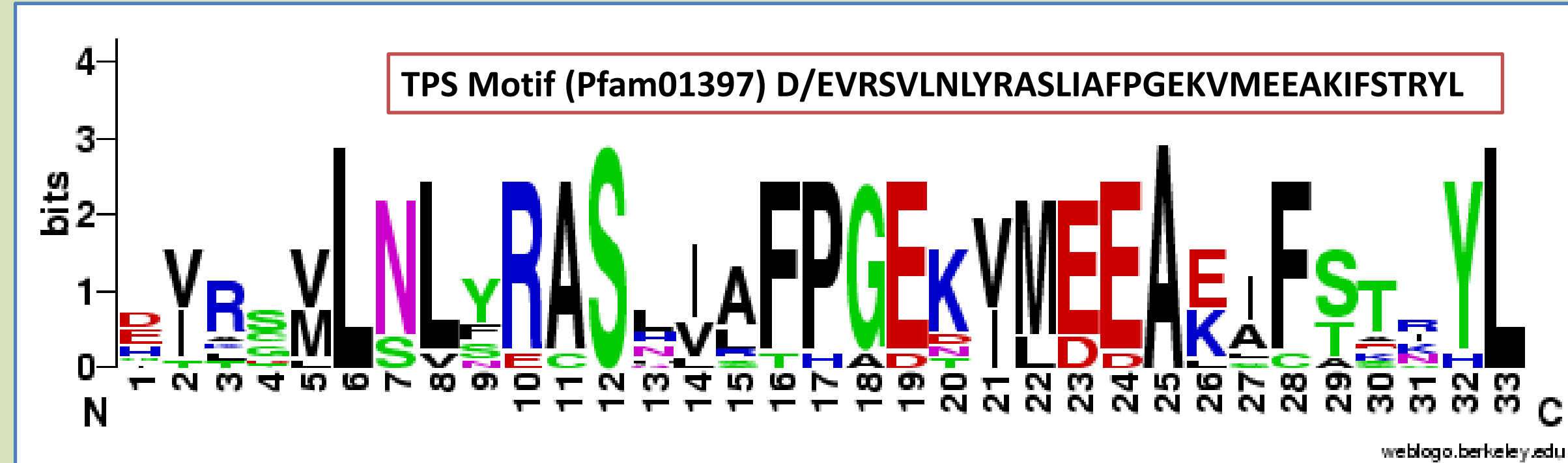
- Alignment of known orthologous terpene synthases from *A. grandis* (*agr*) & *N. tobaccum* (*nto*).
- Conserved domain and motif search in Prosite, Pfam, Interpro, Prodom, COG, TIGR databases.
- Retrieval of Arabidopsis proteome from SwissProtKB and NCBI Genbank databases.
- Position specific scoring matrix was constructed through MEME, MotifSampler, PossuMsearch tools.
- Weight matrix based pattern search of conserved TPS motifs in the proteome of *A. thaliana* through ESA, Lahead and Simple algorithm based search tools of PossuMsearch biosuite (Beckstette et al., 2006) in Linux openSuSe operating system.



RESULTS

PossuMsearch Results for Control Dataset of Terpene synthase (length 33, threshold 144, score range 0....236)

Gene ID	Motif Sequence	Position	E-value	Percent Similarity	Score
5EAS (nta)	DVLGLLNLYEASHVRTHADDILEDALAFSTIHL	142	1.05E-09	58.97	138
gi 62511188 (agr)	GVTDMNLVNRCSHVSPGETIMEEAKLCTERYL	444	3.58E-15	71.36	167
gi 62511183 (agr)	DVASMLNLYRASQLAFPGENILDEAKSFATKYL	389	6.62E-21	82.05	192
gi 62511235 (agr)	DIRTVLNLYRASFIAPFGEKVMEEAEIFSSRYL	212	2.44E-30	95.29	223
gi 17367921 (agr)	EIRSVLNLYRASLIAFPGEKVMEEAEIFSTRYL	214	2.50E-35	100	234
gi 17367924 (agr)	EIRGVLNLFRRASLIAFPGEKIMDEAEIFSTKYL	215	3.15E-29	94.01	220
gi 62511234 (agr)	EIKSVLNLFRRASLIAFPGEKVMEEAEIFSKIYL	217	3.15E-29	94.01	220
gi 17367918 (agr)	ITRSVLNLYRASLVAFPGEKVMEEAEIFSASYL	221	7.15E-29	93.58	219
gi 62511182 (agr)	HVRCMLSLSRASNILFPGEKVMEEAKAFTTNYL	176	3.17E-26	90.17	211
gi 62511181 (agr)	QVRSMLSLLRASEISFPGEKVMEEAKAFTREYL	163	1.79E-22	84.61	198



Chromosome wide distribution of TPS motif

<i>Arabidopsis thaliana</i> proteins	Threshold (E-value)				
	1e-05	1e-01	1e+00	1e+01	1e+02
Complete Proteome (NCBI)	2	31	35	62	146
Complete proteome (Uniprot)	27	57	60	74	190
Chromosome 1 NC003070	1	5	8	14	102
Chromosome 2 NC003071	-	1	1	5	88
Chromosome 3 NC003074	-	6	12	20	199
Chromosome NC003075	-	5	13	24	122
Chromosome 4 NC003076	-	1	4	12	84

Top ten PossuMsearch Results for *Arabidopsis thaliana* proteome dataset (length 33, threshold 86, score range 0....236)

Protein	Description	TPS-Motif	E-value	Score	Position (start)	Percent Similarity
KSA_ARATH	Q38802 Ent-kaurene synthase A, chloroplast precursor	AVTGMFNLYRASQLAFPREEILKNAKEFSYNYL	2.3418E-09	154	418	65.81
Q9LQ27_ARATH	Q9LQ27 F14M2.13 protein	DVRGMLQLYEAAHLATPFETILDEALSFTRYHL	6.0226E-06	135	192	57.69
Q9LRR2_ARATH	Q9LRR2 Terpene synthase-like protein	DIRGMLQLYEAAHLGTPSEDIMDEALSFTRYRL	1.3009E-05	133	122	56.83
Q9LIA1_ARATH	Q9LIA1 Terpene cyclase/synthase	DVRGLLQLYEAAHLGAPSEDIMDEALSFAHYHL	5.8863E-05	129	196	55.12
O04537_ARATH	O04537 F20P5.19 protein	DVRGMLSFYEASHFGTTTDEILEEAMSFQKHL	5.8863E-05	129	221	55.12
Q84UU9_ARATH	Q84UU9 Terpene synthase	DVRGMLSFYEASHFGTTTDEILEEAMSFQKHL	5.8863E-05	129	200	55.12
Q9ZVM4_ARATH	Q9ZVM4 T22H22.10 protein	DARGLLCLYEAAHWSHGEDIIDEALAFSRSHL	8.5309E-05	128	148	54.7
Q9FLW5_ARATH	Q9FLW5 (+)-delta-cadinene synthase	DAYGMLSLYEAAQWGTGEDIIDEALAFSRSHL	5.2549E-04	123	141	52.56
Q84YI3_ARATH	Q84YI3 Putative delta-cadinene synthase, PUP8	DAYGMLSLYEAAQWGTGEDIIDEALAFSRSHL	5.2549E-04	123	141	52.56
Q84UU4_ARATH	Q84UU4 Beta-caryophyllene/alpha-humulene synthase	DAYGMLSLYEAAQWGTGEDIIDEALAFSRSHL	5.2549E-04	123	141	52.56

CONCLUSION

This proteome wide search model paves the path to identify more terpene synthases genes in *A. thaliana*, as well as in other plant systems leading to better understanding of the unrevealed terpene biosynthetic pathways.

REFERENCE Beckstette M, Homann R, Giegerich R, Kurtz S. PoSSuMsearch: Fast index based algorithms and software for matching position specific scoring matrices. BMC Bioinformatics, 7 (1): 389, 2006.

ACKNOWLEDGEMENT Council of Scientific and Industrial Research, New Delhi

E mail: s.luqman@cimap.res.in

