



MIPS sequences: a promising molecular consideration in angiosperm phylogeny and systematics

ANJAN HAZRA^{1,3}, PARAMITA NANDY², CHANDAN SENGUPTA³, SAUREN DAS^{1*}

¹Agricultural and Ecological Research Unit, Indian Statistical Institute, Kolkata, India

²Department of Botany, Sister Nibedita Government General Degree College for Girls, Judges Court Road, Kolkata, India

³Department of Botany, University of Kalyani, Kalyani, Nadia, India

Abstract

The diversity among myo inositol 1-phosphate synthase (MIPS) complete protein sequences, the essential inositol synthesis regulatory enzyme, reported for almost 38 species representing 19 plant families (from Dicotyledons and Monocotyledons), undoubtedly endorses its applicability in molecular systematics. The sequences retrieved from GenBank and their potential application to angiosperm systematics has been emphasized here. Site-specific amino acid residue signatures and their variation among angiosperm genera have been studied. Multiple sequence alignment using Multiple Sequence Comparison by Log-Expectation (MUSCLE) and further manual analysis exposed a major part of ±510 aa residues that have been found to be highly conserved. Moreover, variations of MIPS sequences are very informative in systematics and the phylogeny of angiosperm. The MIPS sequences aligned based on the Maximum Likelihood method and the generated phylogenetic tree strongly support the modern taxonomic hierarchy, especially the Angiosperm Phylogeny Group-IV (APG-IV, 2016) scheme.

Key words: angiosperm phylogeny, APG IV, cladogram, MIPS, multiple sequence alignment

Introduction

Utilization of DNA or protein sequence data in angiosperm systematics and evolution has been widely practiced over the last two decades. Molecular biology applications helped decipher many unsolved issues in systematics and evolutionary biology over time. In this respect, sequence variation in the coding and spacer regions of a number of chloroplast, mitochondrial, and nuclear genes (such as *atpF* or ATP synthase subunit b; *rbcL* or Ribulose biphosphate carboxylase large chain; *matK* or Maturase K; and *ycf2*, a plastid gene) has been considered in elucidating the evolutionary history of organisms or plants (Huang et al., 2010). For this purpose both the coding and the intergenic spacer regions (ITS) of the ribosomal genes have been utilized (Alvarez and Wendel, 2003; Baldwin et al., 1995; Biswal et al., 2012; Feliner and Rosselló, 2007; Hilu and Liang, 1997). Based on a single gene or multigene datasets, in different angiosperms, several clades were identified (Jansen et al.,

2007; Moore et al., 2007; Soltis et al., 2000; Yin-Long et al., 1999). However, only a few nuclear genes could be used in inferring the phylogenetic history of flowering plants (Blattner, 2016; Li et al., 2017; Morton, 2011; Zhang et al., 2012). Still, several interrelationships in the angiosperm phylogenies remain debatable, and thus, for a better understanding it is necessary to use information from different nuclear genes in combination with genes encoded in organellar DNA, as well as from morphological or phenotypic data.

The phylogenetic analysis methods follow substitutions at site that exist due to a mutation, which might be related to natural selection. Mutations take place in the coding as well as noncoding regions with differential rates of substitution (Sang, 2002). Usually, mutations in noncoding regions become silent and are not directly associated to the gene product; however, mutations in the coding parts of genes can be reflected in the protein sequences that may impose changes in their functions.

* Corresponding author: Agricultural and Ecological Research Unit, Indian Statistical Institute, Kolkata, India; e-mail: sauren@isical.ac.in

Proteins need to retain their functions by maintaining a core conserved architecture. Logically, the sequences of functional domains of proteins, as well as the respective coding sequences, are substituted more neutrally and specifically than other nonconserved protein domains or noncoding portions of DNA, which may vary in differential selective compulsions (Rech et al., 2014). Hence, in a phylogenetic analysis, the usage of gene coding region or protein sequences has important applications. Thus, information about evolutionary relationships among distantly or closely related organisms might be established by examining reliable datasets.

MIPS (myo inositol 1-phosphate synthase, EC 5.5.1.4) is a key enzyme regulating the myo inositol biosynthesis pathway, reported in a wide range of organisms (Hazra and Nandy, 2016). The MIPS protein is well known for its ability to increase abiotic stress tolerance in several plant species; it is involved in the resistance to biotic stresses, as well as some developmental processes such as hormonal and growth regulation, cell membrane biogenesis, seed development, programmed cell death, and so on (Abid et al., 2009; Abreu and Aragão, 2006; Das-Chatterjee et al., 2006; Goswami et al., 2014; Kaur et al., 2013; Tan et al., 2013; Zhai et al., 2016). In eukaryotes, this multiple stress-related enzyme is a homotetramer. Each subunit is composed of 530 amino acids. By exploiting MIPS DNA and protein sequences of various organisms, phylogenetic analyses have been performed (Majumder et al., 2003). Currently, the availability of a larger number of sequences covering a wide range of taxa paves the way for molecular phylogenetic and systematics studies. The utilization of molecular data in angiosperm systematics along with the morphological parameters has recently attracted much attention. The present work focuses on the use of some MIPS protein sequences for correlating phylogenetic classification of seed plants and has been examined in addition to excavation of some novel aspects as well.

Materials and methods

Data mining and plant samples

The publicly available MIPS sequences from plant species belonging to angiosperm were extracted from GenBank (National Center for Biotechnology Information, USA) and filtered (the sequences designated as “clone, hypothetical, uncharacterized partial, putative, similar,

like and predicted” were eliminated). The primarily selected 40 sequences were saved in a FASTA format along with their important metadata. Plant taxa (Table 1) were arranged according to their usual taxonomic hierarchy.

Multiple sequence alignment

Multiple and pairwise sequence alignments of full-length protein sequences were performed with the MUSCLE (Edgar, 2004) program. All statistical analyses were performed utilizing the MEGA7 (Kumar et al., 2016) package (optimized parameters; gap open: -2.9, gap extend: 0, hydrophobicity multiplier: 1.2, maximum alignment iterations: 2,865, clustering method: UPGMB). The alignment was further refined by removing ambiguously aligned sequences from the dataset. Finally, 38 sequences were selected for further analysis. The variability of the sequences among individuals were manually examined to identify taxa-specific signature sites or domains.

Phylogenetic tree estimation

The molecular phylogeny of the MIPS proteins within the studied plant group (Angiosperms) was reconstructed using the Molecular Evolutionary Genetics Analysis (MEGA) software (version 7). For application, its evolutionary history was inferred using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-5285.4321) was shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the JTT model. The validity of phylogeny reconstruction was tested following a 100 bootstrap replications.

Comparison with Angiosperm Phylogeny Group-IV (APG) classification

The phylogenetic classification scheme of angiosperms chiefly following Angiosperm Phylogeny Website, APweb (2016) and APG IV (The Angiosperm Phylogeny, 2016) were retrieved from the website. Based on the molecular phylogenetic data, a reference tree was drawn by picking the families covered in the MIPS characteristics and from the APG hypothetical tree. Both trees were examined to check for similarities and dissimilarities in the branching and clustering patterns to evaluate the significance of MIPS sequences as systematic markers.

Table 1. List of curated MIPS amino acid sequences from angiosperm members taken for the study

Plant	Family	Accession id	Submitted by
<i>Avicennia marina</i>	<i>Acanthaceae</i>	AAK21969	Jithesh et al., 2001
<i>Actinidia deliciosa</i>	<i>Actinidiaceae</i>	AFV31635	Cui et al., 2012
<i>Actinidia eriantha</i>	<i>Actinidiaceae</i>	AHB62409	Cui et al., 2013
<i>Actinidia rufa</i>	<i>Actinidiaceae</i>	AHB62410	Cui et al., 2013
<i>Brassica napus</i>	<i>Brassicaceae</i>	ACJ65008	Ray and Georges, 2008
<i>Arabidopsis thaliana</i>	<i>Brassicaceae</i>	AAA85390	Johnson, 1995
<i>Cardamine hirsuta</i>	<i>Brassicaceae</i>	AII30313	Cartolano et al., 2015
<i>Suaeda salsa</i>	<i>Chenopodiaceae</i>	AAL28131	Wang et al., 2001
<i>Arthrocnemum glaucum</i>	<i>Chenopodiaceae</i>	ABP63338	Mahdy et al., 2001
<i>Citrullus lanatus subsp. vulgaris</i>	<i>Cucurbitaceae</i>	BAI52951	Lee et al., 2009
<i>Sporobolus alterniflorus</i>	<i>Cyperaceae</i>	ADC33414	Baisakh and Subudhi, 2010
<i>Hevea brasiliensis</i>	<i>Euphorbiaceae</i>	AFD61599	Saha et al., 2011
<i>Ricinus communis</i>	<i>Euphorbiaceae</i>	ACU30131	Dai and Wei, 2009
<i>Jatropha curcas</i>	<i>Euphorbiaceae</i>	NP_001292951	Ref-Seq
<i>Glycine max</i>	<i>Fabaceae</i>	AAK72098	Hitz et al., 2001
<i>Medicago truncatula</i>	<i>Fabaceae</i>	XP_003601987	Tang et al., 2014
<i>Phaseolus vulgaris</i>	<i>Fabaceae</i>	AAK69514	Johnson et al., 2001
<i>Medicago falcata</i>	<i>Fabaceae</i>	ABO77439	Wang et al., 2007
<i>Trifolium repens</i>	<i>Fabaceae</i>	ADD09590	Hand et al., 2010
<i>Gossypium hirsutum</i>	<i>Malvaceae</i>	ACJ11714	Pang et al., 2009
<i>Gossypium arboreum</i>	<i>Malvaceae</i>	KHG09555	Mudge et al., 2014
<i>Morus notabilis</i>	<i>Moraceae</i>	XP_010090558	He and Zhao, 2014
<i>Passiflora edulis</i>	<i>Passifloraceae</i>	ABF51620	Abreu and Aragao, 2007
<i>Sesamum indicum</i>	<i>Pedaliaceae</i>	NP_001291338	Chun et al., 2015
<i>Triticum aestivum</i>	<i>Poaceae</i>	AGK06903	Ma et al., 2013
<i>Oryza coarctata</i>	<i>Poaceae</i>	ACI94875	Majumder et al., 2010
<i>Zea mays</i>	<i>Poaceae</i>	AFW89388	Schnable et al., 2013
<i>Avena sativa</i>	<i>Poaceae</i>	BAB40956	Yoshida et al., 2002
<i>Oryza sativa Japonica</i>	<i>Poaceae</i>	BAA25729	Yoshida et al., 1999
<i>Lolium perenne</i>	<i>Poaceae</i>	AAN52772	Amiard et al., 2002
<i>Citrus paradise</i>	<i>Rutaceae</i>	P42802	Abu-Abied and Holland, 1994
<i>Populus trichocarpa</i>	<i>Salicaceae</i>	XP_002307107	Tuskan et al., 2013
<i>Nicotiana tabacum</i>	<i>Solanaceae</i>	Q9LW96	Hara et al., 2000
<i>Nicotiana paniculata</i>	<i>Solanaceae</i>	Q9SSV4	Hashimoto et al., 1999
<i>Solanum lycopersicum</i>	<i>Solanaceae</i>	NP_001296998	Aoki et al., 2010
<i>Theobroma cacao</i>	<i>Sterculiaceae</i>	XP_007048015	Mockaitis et al., 2014
<i>Camellia sinensis</i>	<i>Theaceae</i>	AJO70149	Yue et al., 2015
<i>Xerophyta viscosa</i>	<i>Velloziaceae</i>	AAP85531	Majee et al., 2003

Results and discussion

The MIPS sequences were reported from a wide range of seed plant families (Table 1). A total 1,206 sequences were retrieved from the National Center for Biotechnology Information (NCBI) by searching the “inositol phosphate synthase” of the flowering plants origin. Thereafter, the sequences were curated by removing partial, probable, putative, predicted, and redundant ones. After a precise elimination of unsuitable varieties, 38 complete sequences representing 19 families were used, among which 30 belonged to the dicots and the remaining 8 were monocots (Table 1). The total number of families covered in the analysis was 19, including *Acanthaceae* (1), *Actinidiaceae* (3), *Brassicaceae* (3), *Chenopodiaceae* (2), *Cucurbitaceae* (1), *Cyperaceae* (1), *Euphorbiaceae* (3), *Fabaceae* (5), *Malvaceae* (2), *Moraceae* (1), *Passifloraceae* (1), *Pedaliaceae* (1), *Poaceae* (6), *Rutaceae* (1), *Salicaceae* (1), *Solanaceae* (3), *Sterculiaceae* (1), *Theaceae* (1), and *Velloziaceae* (1). The alignment of these sequences revealed the presence of at least 30 phylogenetically informative amino acid residues. Most of the changes in nucleotides were related to the degeneration of codons (e.g., both GCA and GCC codes for alanine) and posttranslational modifications (e.g., mutations that remain silent) that do not change amino acids in protein sequences. Site-specific sequence changes in the proteins are the driving force of molecular divergence and evolution of the proteins among the individuals (Halabi et al., 2009). Therefore, a diversity analysis at a protein sequence level is more informative, accurate, and relevant in the evolutionary systematics of higher plants.

Although the MIPS gene was found in various plants with an almost similar sequence composition, their activity level has not been conserved (Khurana et al., 2012; Ray et al., 2010). These differences might also be attributed to the transcriptional and/or translational regulation or changes in the sequences. Correlating all these factors as well as molecular phenotypes involves different approaches of statistical validation and simulation modeling (Liberles et al., 2012). The phylogenetic study on steroid hormone receptors revealed that their affinities for various ligands have been changing during evolutionary processes (Bridgham et al., 2009). The observations and analyses revealed that a few changes in coding or amino acids sequences are sufficient to cause

functional alterations, while consequences of very few mutations have also posed new functions. Leucine-responsive regulatory protein (Lrp) and its paralog AsnC (an autogenously regulated activator of asparagine synthetase) in γ -Proteobacteria group have conserved sequence signatures that allow for their definite annotation; moreover, specific residues were correlated with global versus local regulatory roles, for example, among the members versus within the member itself, which were further reflected in protein functions (Unoarumhi et al., 2016). Halabi et al. (2009) demonstrated that in S1A family of serine proteases, evolutionary units of a three-dimensional protein structure have multiple sectors. The evolutionary divergence of S1A family amino acid sequences has been reflected in the changes of certain protein sectors; importantly, some of them have roles in functional specificity. Therefore, its members express a variation in substrate specificities and effectively classify sequences to the organism type.

A keen observation of the MIPS proteins by multiple sequence alignment (Fig. 1, Supplementary Fig. 1) allowed to determine taxa-specific variable sites. The per residue signature observations indicated that glycine in position 16 (Gly 16) was conserved among all monocots tested, whereas dicots were quite unstable at the site. Likewise, in position 23, valine (V) residue was present in all dicots, whereas monocots had aspartic acid (D). Whereas glutamate (E) and tyrosine (Y) were present in position 43 in most dicots, but serine (S) was present in all monocots. Arginine (R) or lysine (K) were present in position 44 in all the monocots analyzed; however, most dicots had glutamic acid (E) in this position. The alanine residue present in position 82 in all monocots was substituted by glycine (G) in dicots (excluding *Brassicaceae* and *Euphorbiaceae*, which have alanine). Tyrosine (Y) in position 120 was conserved among all the monocots and a few dicots tested; however, most dicots had phenylalanine (F) in that position. Threonine (T) from position 158 was shared by all the monocots as a conserved amino acid within the class. A polymorphic 243 site consisted of cysteine (C) or serine (S) in all the monocots, whereas valine (V) and isoleucine (I) were present in dicots. However, G in position 279 was common among all the monocots, but asparagine (N) among most dicots. Another class-specific site was present at the terminal part of the protein, in position 487, where

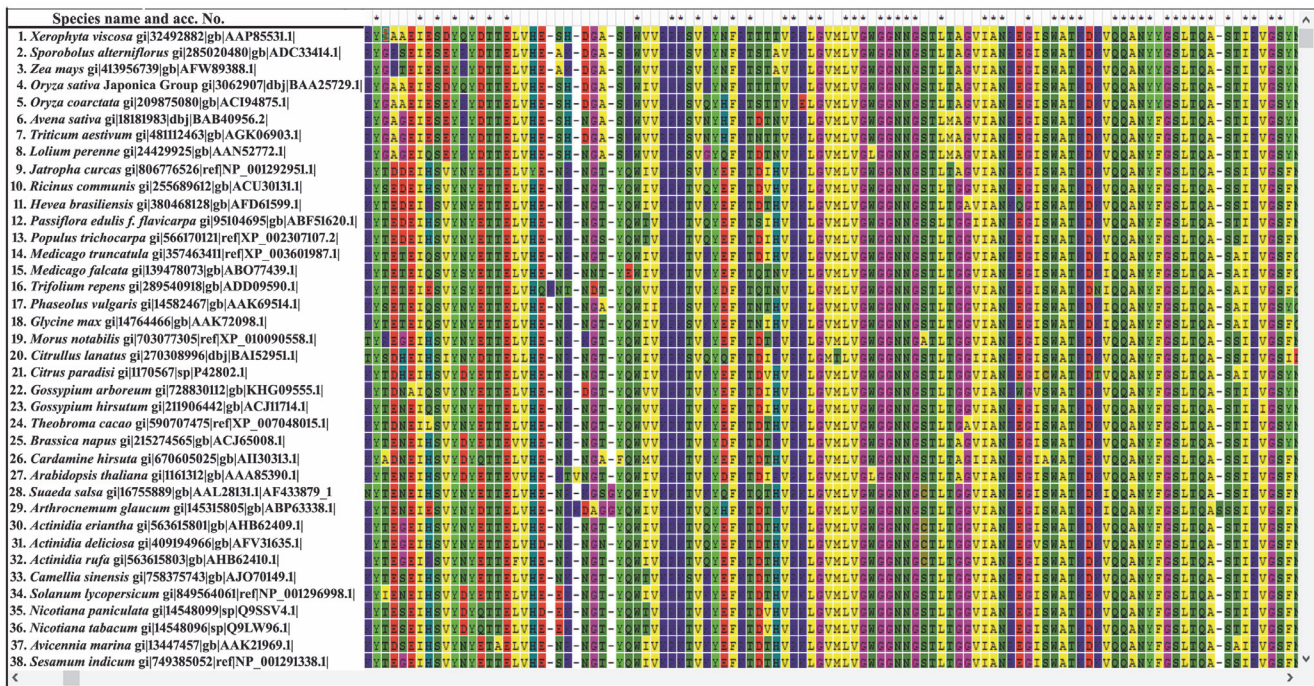


Fig. 1. Screenshot showing a part of the multiple sequence alignment using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) among myo inositol 1-phosphate synthase (MIPS) amino acid sequences

A was shared by the monocots and S by all the dicots. Some family-specific conservensness in the polymorphic positions was also found, for example, A, in position 114, was a conserved signature in *Fabaceae*; S, in position 504 in *Actinidiaceae*; and A in position 372 in *Brassicaceae* and *Chenopodiaceae*. The consequences of these site-specific residue divergences in the evolution of these protein/enzymes functions are still to be confirmed.

The number of domains constituting MIPS proteins is conserved throughout different groups of plants (Majumder et al., 2003). In some domains, a major part of the sequence is conserved (Hazra and Nandy, 2016) and highly conserved along the entire length among the seed plants (observed in the present study), but still some signature sequences are evidently informative in molecular systematics. Similar analyses performed on xanthine dehydrogenase revealed that some domains were well-conserved, although they possessed some variability that could be implicated in the angiosperm phylogenetic and systematic research (Morton, 2011).

Most of the phylogenetic tree constructed using MIPS protein sequences (Fig. 2) supported the concurrent angiosperm phylogeny classification. Notably,

monocots and dicots were distinctly separated with a strong bootstrap support (100). Moreover, clustering of all species into families according to their usual taxonomic lineages, close relatedness of *Euphorbiaceae*, *Salicaceae*, and *Passifloraceae*; monophyly of *Sterculiaceae* and *Malvaceae*; evolution of Brassicales (*Brassicaceae*) and Caryophyllales (*Chenopodiaceae*) from a common ancestor; and constellation of *Acanthaceae* and *Pedaliaceae* into a single group (Order: Lamiales) has been reflected in the study, in congruence with the APG-IV classification system (The Angiosperm Phylogeny, 2016). However, some other observations are not in line with the previously established phylogenetic relationships, and they include these pointing to *Solanales* and *Lamiales* of Lamids being clearly separated; *Moraceae* of Fabids and *Rutaceae* of Malvids being sister groups; *Fabaceae* and *Cucurbitaceae* being distantly related; or *Hevea brasiliensis* being a common ancestor of *Euphorbiaceae*, *Salicaceae*, and *Passifloraceae*. A statistical analysis inferred more relative strength of the branches below the order level. Hence, practically, it would help as an inter- or intra-family marker in angiosperm systematics, though more sequence tracking is necessary for enriching the novelty of the subject itself.

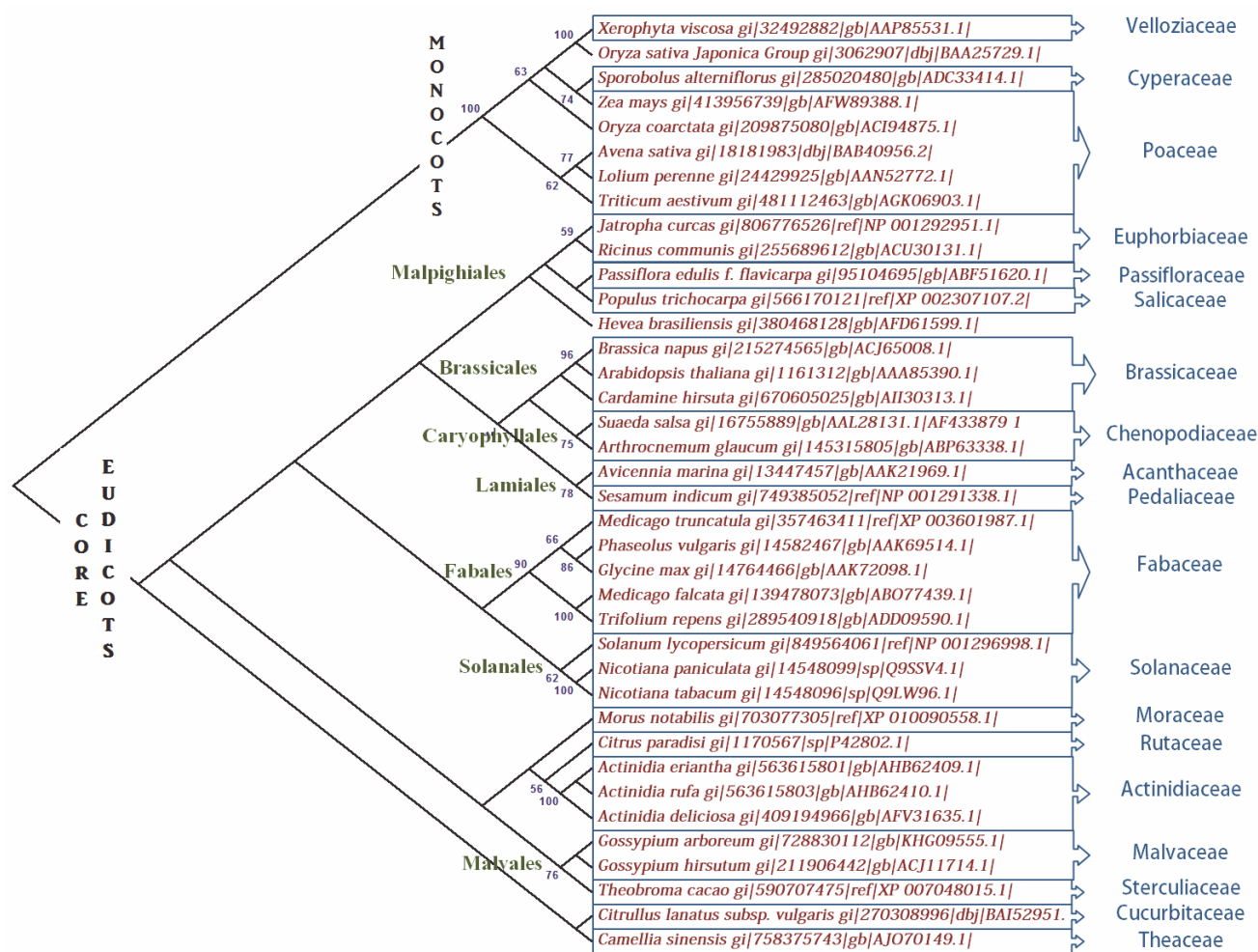


Fig. 2. Molecular phylogenetic analysis by the Maximum Likelihood method (Guindon and Gascuel, 2003); the evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model (Jones et al., 1992); initially, the tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model; evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016); bootstrap support values more than 50% are shown in branch nodes

Conclusions

Nucleic acid and amino acid sequences are still commonly used for inferring evolutionary taxonomy and classification of organisms. The validation and utilization of low-copy nuclear coding sequences and their respective protein sequences are difficult to accomplish. Sequences of extensively studied MIPS enzyme were examined to check their efficiency in the phylogenetic position. The potentiality of using these sequences in angiosperm phylogeny and systematics has been demonstrated. Moreover, due to very specific diversity among organisms and groups, we see its usefulness and recommend integration of MIPS sequences into the established convenient tools

in molecular phylogeny combining chloroplast, mitochondrial, nuclear, and morphological data.

Compliance with ethical standards

Funding: The author, A.H., thanks National Tea Research Foundation, Tea Board of India for providing a research fellowship. The infrastructure facility was provided by the Indian Statistical Institute, Kolkata.

Conflict of interest: none declared.

References

Abid G., Silue S., Muhovski Y., Jacquemin J.-M., Toussaint A., Baudoin J.-P. (2009) *Role of myo-inositol phosphate synthase and sucrose synthase genes in plant seed development*. *Gene* 439(1): 1–10.

- Abreu E.F., Aragão F.J. (2006) *Isolation and characterization of a myo-inositol-1-phosphate synthase gene from yellow passion fruit (Passiflora edulis f. flavicarpa) expressed during seed development and environmental stress*. Ann. Bot. 99(2): 285–292.
- Alvarez I., Wendel J. (2003) *Ribosomal ITS sequences and plant phylogenetic inference*. Mol. Phylogen. Evol. 29(3): 417–434.
- Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechowski M.F., Campbell C.S., Donoghue M.J. (1995) *The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny*. Ann. Missouri Bot. Garden 82: 247–277.
- Biswal D.K., Debnath M., Kumar S., Tandon P. (2012) *Phylogenetic reconstruction in the Order Nymphaeales: ITS2 secondary structure analysis and in silico testing of maturase k (matK) as a potential marker for DNA bar coding*. BMC Bioinformatics 13(17): S26.
- Blattner F.R. (2016) *TOPO6: a nuclear single-copy gene for plant phylogenetic inference*. Plant Systemat. Evol. 302(2): 239–244.
- Bridgham J.T., Ortlund E.A., Thornton J.W. (2009) *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution*. Nature 461(7263): 515–519.
- Das-Chatterjee A., Goswami L., Maitra S., Dastidar K.G., Ray S., Majumder A.L. (2006) *Introgression of a novel salt tolerant L myo inositol 1 phosphate synthase from Porteresia coarctata (Roxb.) Tateoka (PcINO1) confers salt tolerance to evolutionary diverse organisms*. FEBS Lett. 580(16): 3980–3988.
- Edgar R.C. (2004) *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucl. Acids Res. 32(5): 1792–1797.
- Feliner G.N., Rosselló J.A. (2007) *Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants*. Mol. Phylogen. Evol. 44(2): 911–919.
- Goswami L., Sengupta S., Mukherjee S., Ray S., Mukherjee R., Majumder A.L. (2014) *Targeted expression of L-myoinositol 1-phosphate synthase from Porteresia coarctata (Roxb.) Tateoka confers multiple stress tolerance in transgenic crop plants*. J. Plant Biochem. Biotechnol. 23(3): 316–330.
- Guindon S., Gascuel O. (2003) *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood*. Systematic Biol. 52(5): 696–704.
- Halabi N., Rivoire O., Leibler S., Ranganathan R. (2009) *Protein sectors: evolutionary units of three-dimensional structure*. Cell 138(4): 774–786.
- Hazra A., Nandy P. (2016) *Myo-inositol 1-phosphate synthase – the chosen path of evolution*. BioTechnologia 97(2): 95–108.
- Hilu K., Liang H. (1997) *The matK gene: sequence variation and application in plant systematics*. Amer. J. Bot. 84(6): 830–830.
- Huang J.L., Sun G.L., Zhang D.M. (2010) *Molecular evolution and phylogeny of the angiosperm ycf2 gene*. J. Systemat. Evol. 48(4): 240–248.
- Jansen R.K., Cai Z., Raubeson L.A., Daniell H., Leebens-Mack J., Müller K.F., Guisinger-Bellian M., Haberle R.C., Hansen A.K., Chumley T.W. (2007) *Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns*. Proc. Natl. Acad. Sci. USA 104(49): 19369–19374.
- Jones D.T., Taylor W.R., Thornton J.M. (1992) *The rapid generation of mutation data matrices from protein sequences*. Comput. Applicat. Biosci. CABIOS 8(3): 275–282.
- Kaur H., Verma P., Petla B.P., Rao V., Saxena S.C., Majee M. (2013) *Ectopic expression of the ABA-inducible dehydration-responsive chickpea L-myoinositol 1-phosphate synthase 2 (CaMIPS2) in Arabidopsis enhances tolerance to salinity and dehydration stress*. Planta 237(1): 321–335.
- Khurana N., Chauhan H., Khurana P. (2012) *Expression analysis of a heat-inducible, Myo-inositol-1-phosphate synthase (MIPS) gene from wheat and the alternatively spliced variants of rice and Arabidopsis*. Plant Cell Rep. 31(1): 237–251.
- Kumar S., Stecher G., Tamura K. (2016) *MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets*. Mol. Biol. Evol.: msw054.
- Li Z., De La Torre A.R., Sterck L., Cánovas F.M., Avila C., Merino I., Cabezas J.A., Cervera M.T., Ingvarsson P.K., Van de Peer Y. (2017) *Single-copy genes as molecular markers for phylogenomic studies in seed plants*. Genome Biol. Evol. 9(5): 1130–1147.
- Liberles D.A., Teichmann S.A., Bahar I., Bastolla U., Bloom J., Bornberg Bauer E., Colwell L.J., De Koning A., Dokholyan N.V., Echave J. (2012) *The interface of protein structure, protein biophysics, and molecular evolution*. Protein Sci. 21(6): 769–785.
- Majumder A.L., Chatterjee A., Ghosh Dastidar K., Majee M. (2003) *Diversification and evolution of L myo inositol 1 phosphate synthase 1*. FEBS Lett. 553(1–2): 3–10.
- Moore M.J., Bell C.D., Soltis P.S., Soltis D.E. (2007) *Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms*. Proc. Natl. Acad. Sci. USA 104(49): 19363–19368.
- Morton C.M. (2011) *Newly sequenced nuclear gene (Xdh) for inferring angiosperm phylogeny1*. Annal. Missouri Bot. Garden 98(1): 63–89.
- PF S. APWeb [Internet]. Available from: www.mobot.org/MOBOT/research/APweb
- Ray S., Patra B., Das-Chatterjee A., Ganguli A., Majumder A.L. (2010) *Identification and organization of chloroplastic and cytosolic L-myoinositol 1-phosphate synthase coding gene (s) in Oryza sativa: comparison with the wild halophytic rice, Porteresia coarctata*. Planta 231(5): 1211–1227.
- Rech G.E., Sanz-Martín J.M., Anisimova M., Sukno S.A., Thon M.R. (2014) *Natural selection on coding and noncoding DNA sequences is associated with virulence genes in*

- a plant pathogenic fungus*. *Genome Biol. Evol.* 6(9): 2368–2379.
- Sang T. (2002) *Utility of low-copy nuclear gene sequences in plant phylogenetics*. *Crit. Rev. Biochem. Mol. Biol.* 37(3): 121–147.
- Soltis D.E., Soltis P.S., Chase M.W., Mort M.E., Albach D.C., Zanis M., Savolainen V., Hahn W.H., Hoot S.B., Fay M.F. (2000) *Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences*. *Bot. J. Linnean Soc.* 133(4): 381–461.
- Tan J., Wang C., Xiang B., Han R., Guo Z. (2013) *Hydrogen peroxide and nitric oxide mediated cold and dehydration induced myo inositol phosphate synthase that confers multiple resistances to abiotic stresses*. *Plant, Cell Environ.* 36(2): 288–299.
- The Angiosperm Phylogeny G. (2016) *An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV*. *Bot. J. Linnean Soc.* 181(1): 1–20.
- Unoarumhi Y., Blumenthal R.M., Matson J.S. (2016) *Evolution of a global regulator: Lrp in four orders of γ -Proteobacteria*. *BMC Evolut. Biol.* 16(1): 111.
- Yin-Long Q., Lee J., Bernasconi-Quadroni F., Soltis D.E. (1999) *The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes*. *Nature* 402(6760): 404.
- Zhai H., Wang F., Si Z., Huo J., Xing L., An Y., He S., Liu Q. (2016) *A myo inositol 1 phosphate synthase gene, IbMIPS1, enhances salt and drought tolerance and stem nematode resistance in transgenic sweet potato*. *Plant Biotech. J.* 14(2): 592–602.
- Zhang N., Zeng L., Shan H., Ma H. (2012) *Highly conserved low copy nuclear genes as effective markers for phylogenetic analyses in angiosperms*. *New Phytologist* 195(4): 923–937.