Original Research Paper

Zoology

Molluscan Protein Phylogeny: A Case Study

Bibhuti Prasad Barik

Post Graduate Department of Zoology, Khallikote University, Brahmapur-760001, Odisha, India

ABSTRACT

Large Molluscan protein datasets are existing and parallel rise of evolutionary studies relating to Molluscan species are missing. The phylum Mollusca comprises huge number of species. The inter specific diversity and phylogeny is poorly investigated. In the current study different protein sequences were retrieved and phylogenetic analysis were carried out resulting different trees. The trees showed similar species remain clustered together with few alterations. This may be attributed to adaptive radiation or mutations.

Introduction

Molluscs are soft-bodied shell bearing animals. The members of this phylum are highly diverse, successful and mostly aquatic occupying habitats ranging from the deep ocean to shallow waters to moist terrestrial niches. Phylogenetic studies within molluscs stays unanswered despite availability of computational and predictive database and the associated tools. In the current study an attempt has been made to study the phylogenetic relationships of Molluscan species found globally.

Materials & Methods

Retrieval of sequence

The protein sequences were retrieved from NCBI-GenBank (Benson, 2013). A key word 'Mollusca' in the search box with nucleotide option. The webserver generated a variety entries including INSDC (GenBank) nucleotide sequence, mRNA nucleotide sequence, and corresponding reference protein sequences. The server generated 42499 entries. The corresponding filter searched protein sequences are selected based on their sequence profile for further analysis.

Taxon sampling

The sequences were sorted using BioEedit software version 7.0.5.3(Hall, 1999). Altogether nine protein categories were retrieved and separated. The sequences were edited according to their accession number and lineage.

Phylogenetic analysis

The retrieved protein sequences were fasta formatted and subjected to multiple sequence alignment using clustalW (Thompson et al., 1994). For pair wise sequence alignment the gap opening penalty and gap extension penalties were 15 and 6.66 respectively. The aligned files were exported to the local machine for phylogenetic analysis using MEGA 7 (Kumar et al., 2016). Five different methods (maximum likelihood, neighbour joining, minimum evolution, UPGMA and maximum parsimony) were adopted to perform phylogenetic analysis. Trees are reconstructed with given topologies and branch lengths (table 1). All characters were equally weighted and unordered. Alignment gaps were treated as missing data.

Table 1: Branch length and indices of CI, RI and CI

Sl. No.	Protein	Sum of Branch length					Consistency index	Retention index	Composite index
		ML	NJ	ME	UPGMA	MP			
1.	Actin II	-247.85	0.44	0.44	0.37	19	0.84	0.93	0.78
2.	Cytochrome B	-5088.45	7.07	7.07	6.93	876	0.42	0.42	0.32
3.	Cytochrome C	-270.13	0.40	0.40	0.39	20	1.00	1.00	1.00
4.	Conotoxin	-251.59	1.03	1.03	1.01	30	0.96	0.94	0.91
5.	Elongation-factor 1 alpha	-1798.98	4.76	4.76	4.71	294	0.44	0.42	0.18
6.	Mitochondrial ATP synthase	-1305.69	0.42	0.42	0.42	99	0.63	0.64	0.40
7.	Myosin heavy chain type II	-742.10	1.80	1.80	1.80	100	0.61	0.63	0.38
8.	Rhodopsin	-896.17	2.06	2.06	2.02	125	0.54	0.52	0.37
9.	RNA Polymerase II	-1279.78	4.83	4.43	4.66	234	0.66	0.65	0.43

ML: Maximum Likelihood, NJ: Neighbour Joining, ME: Minimum Evolution, UPGMA: Unweighted Pair Group Method with Arithmetic Mean, MP: Maximum Parsimony, CI: Consistency Index, RI: Retention Index and CI: Composite Index

Results:

Morphological disparity among the key lineages of Mollusca has encouraged numerous conflicting phylogenetic hypotheses. Unfortunately, because of varying interpretations of features as derived or plesiomorphic, a lack of clear synapomorphies, and often unclear character homology, the ability of morphology to resolve such deep phylogenetic events is limited. Morphological and traditional molecular phylogenetic approaches have failed to robustly reconstruct mollusc phylogeny. Notably, several recent phylogenomic studies have significantly advanced our understanding of metazoan evolution by using sequences derived from genome and transcriptome

data. With this approach, numerous orthologous protein-coding genes can be identified and employed in phylogeny reconstruction. Here, in this investigation a phylogeneteic approach was attempted to know evolutionary relationships among the major lineages of Mollusca. Sequence data of proteins namely actin II, cytochrome B, cytochrome C, conotoxin, elongation-factor 1 alpha, mitochondrial ATP synthase, myosin heavy chain type II, rhodopsin and RNA polymerase II were collected from different operational taxonomic units. The phylogenetic trees showing different clades were depicted below (Fig. 1-9).

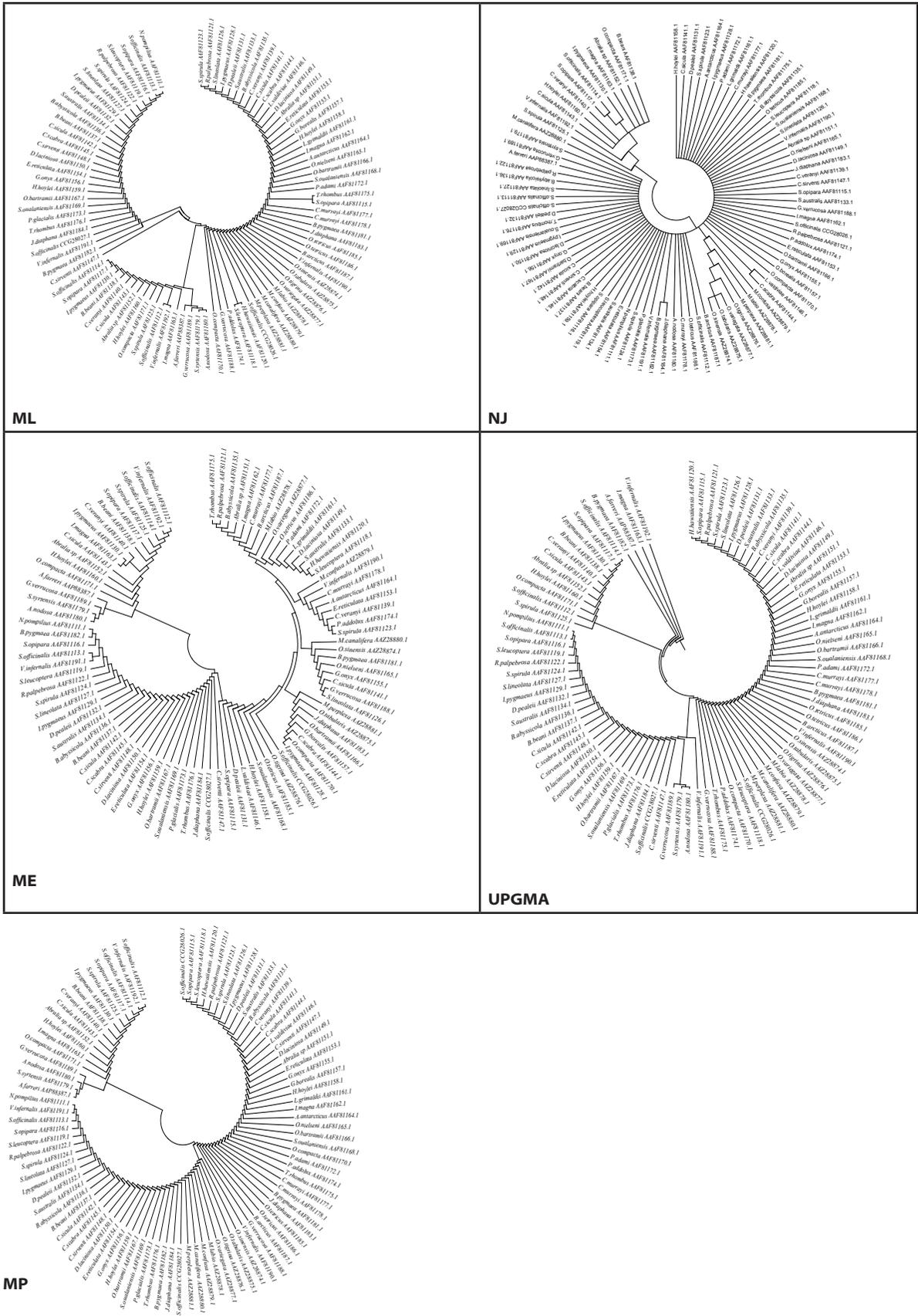**Fig.1 (a-e): Actin II protein based phylogenetic trees**

ML

NJ

ME

UPGMA

MP

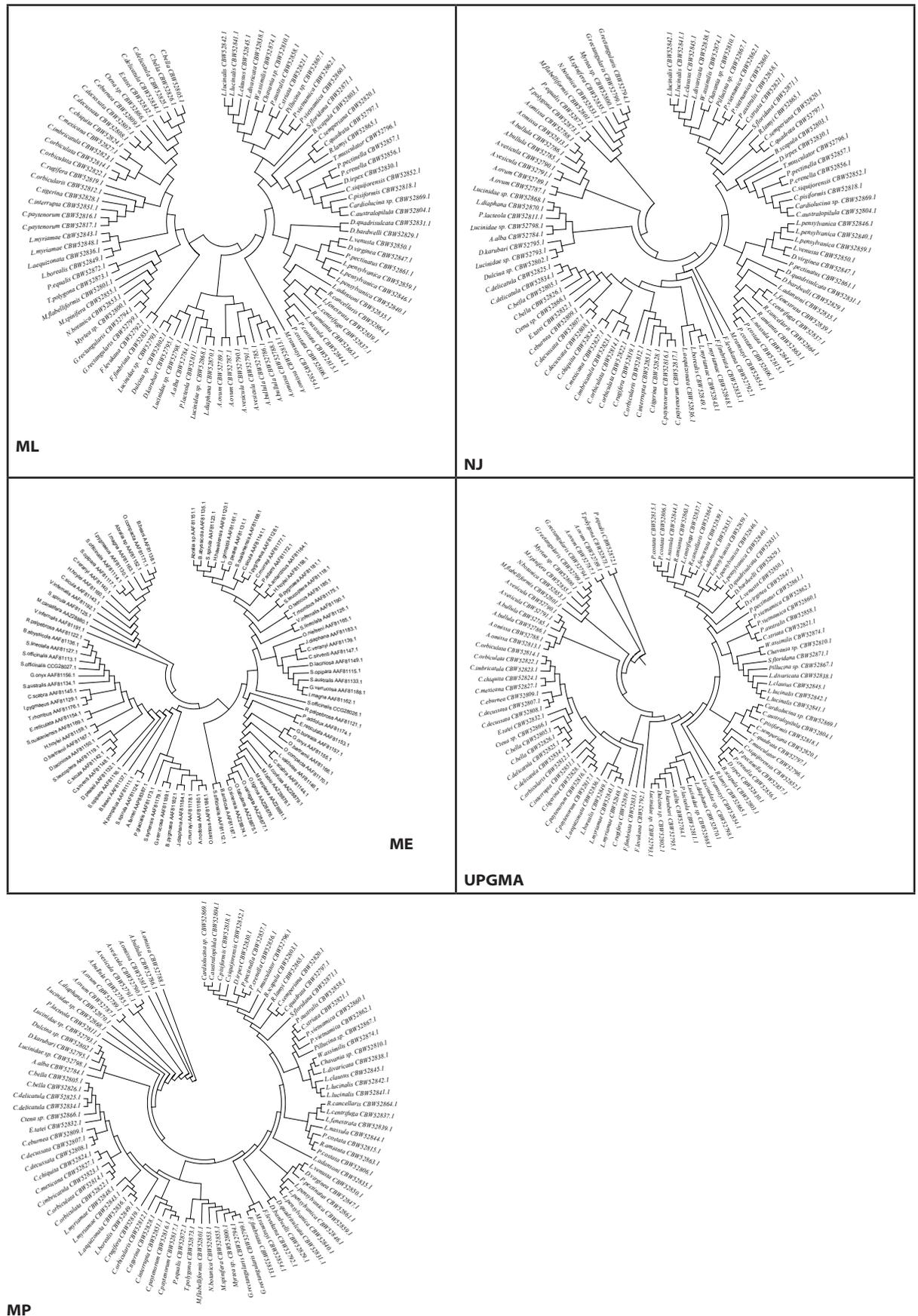**Fig.2 (a-e): Cytochrome B based phylogenetic trees**
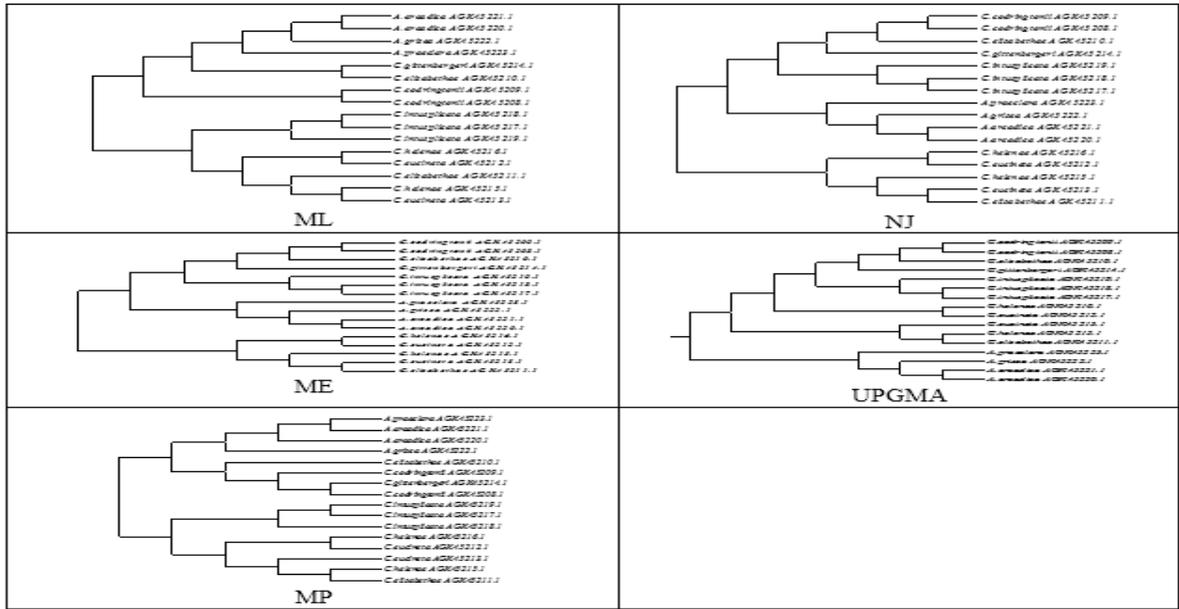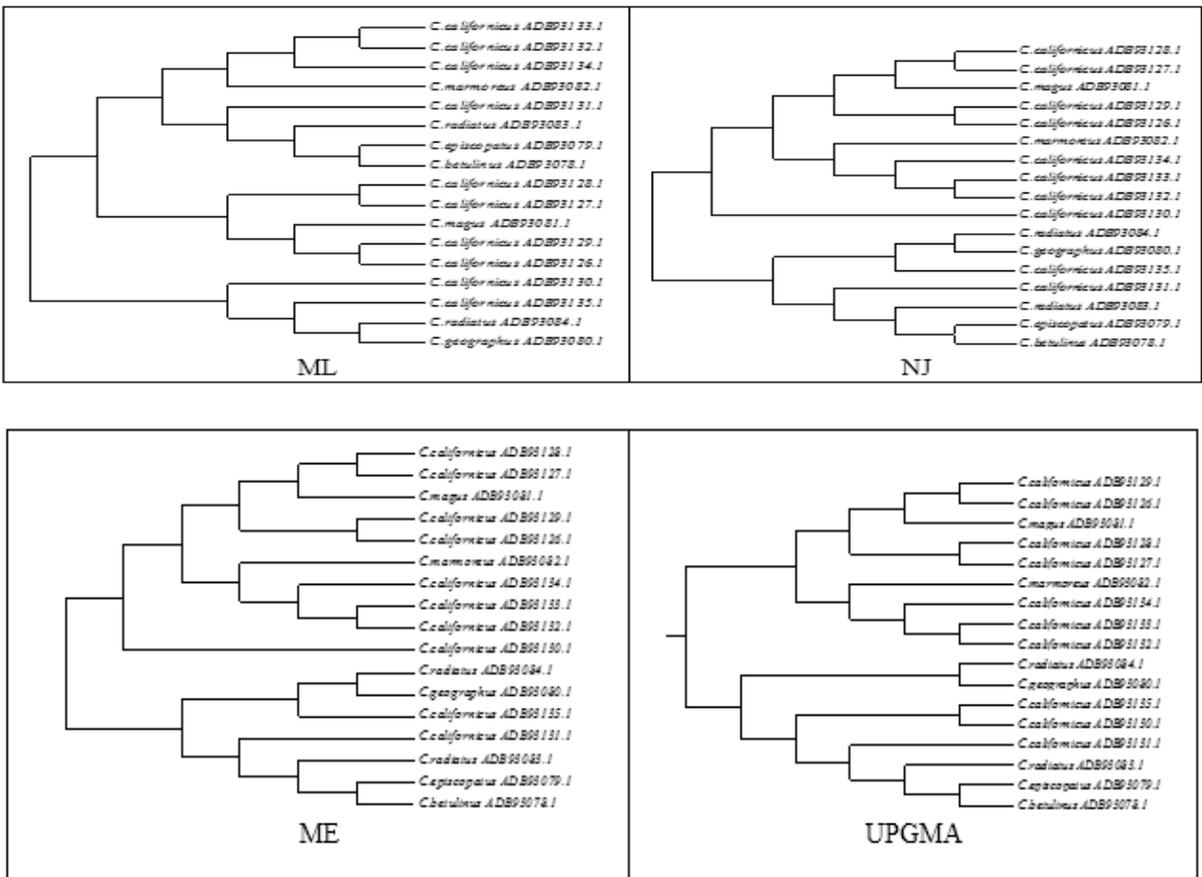
Fig. 3 (a-e): Cytochome C based phylogenetic trees
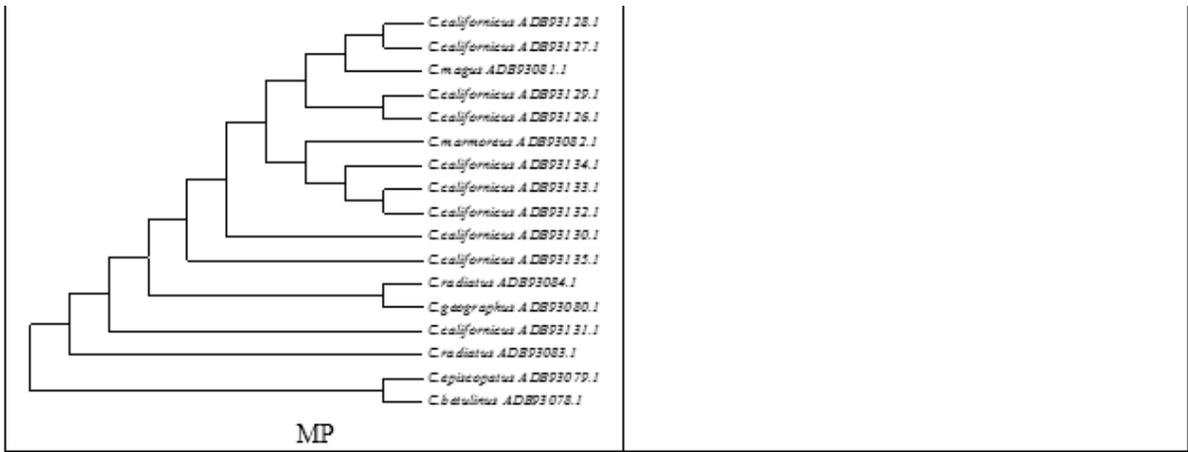

Fig. 4 (a-e): Conotoxin based phylogenetic trees

Fig. 5 (a-e): Elongation-factor 1 alpha based phylogenetic trees

Fig. 6 (a-e): Mitochondrial ATP synthase based phylogenetic trees

Fig.7 (a-e): Myosin heavy chain type II based phylogenetic trees
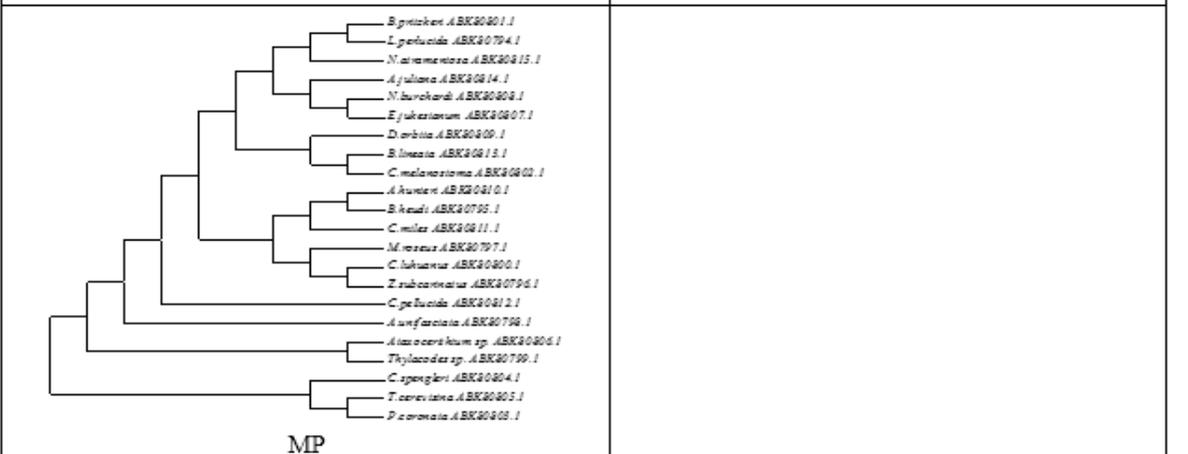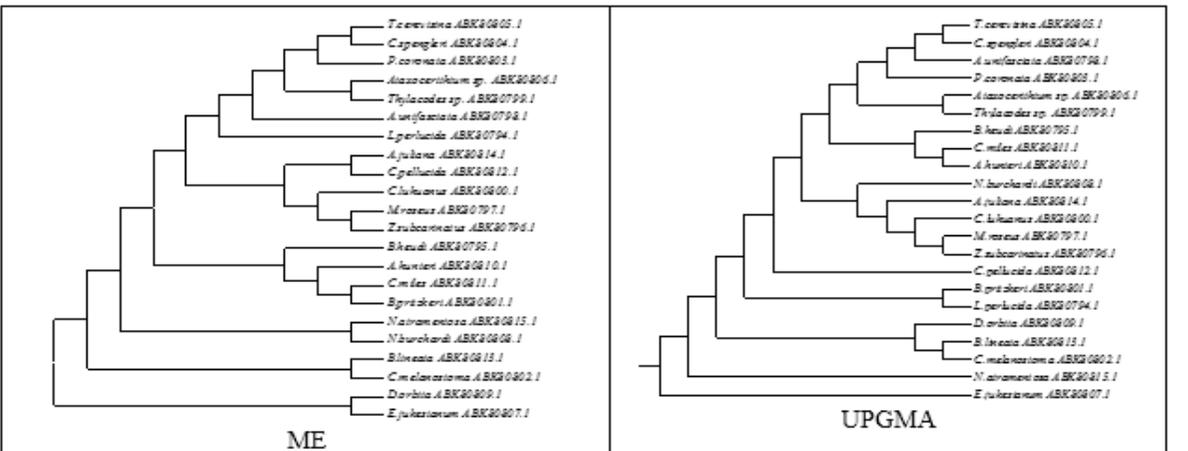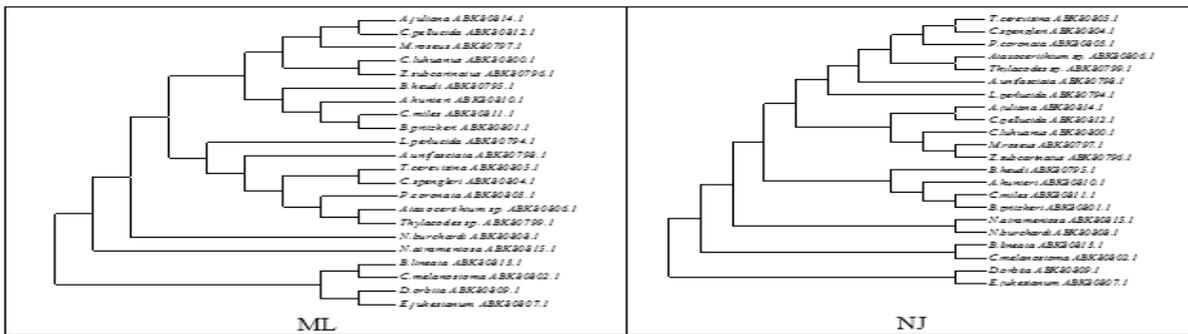

Fig.8 (a-e): Rhodopsin based phylogenetic trees

Fig.9 (a-e): RNA Polymerase II based phylogenetic trees

The evolutionary trees showed similar species remain cluster together with few alterations. This may be attributed to adaptive radiation or mutations. For especially for deep nodes, the outcomes of molecular phylogenetics should always be compared with, and eventually validated by, all the expertise in the field, merging to a widely accepted phylogenetic hypothesis, encompassing the whole evidence from protein to morphology. Further improvements of the present work along with proteome structural will increase the available dataset either by exploiting more molecular markers or by further enlarging the sample, with special reference to some underrepresented groups.

Conclusion:

The inter specific diversity and phylogeny is poorly investigated. In the current study different protein sequences were accessed, retrieved and phylogenetic analysis were carried out. The trees showed similar species remain cluster together with few alterations. This may be attributed to adaptive radiation or mutations. Further study of physio-chemical properties along with disorder study of protein structures may be used to explore the animal's response to different environmental conditions.

References

1.   Kumar S., Stecher G. and Tamura K., (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33(7): 1870–1874.

2.   Thompson J.D., Higgins D.G. and Gibson T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22 (22): 4673-4680.

3.   Hall T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp Ser. 41: 95–98.

4.   Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. and Sayers E.W. (2013). GenBank. Nucleic Acids Res. 41(Database issue): 36-42.