International Journal of Advanced Biotechnology and Research(IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol5, Issue3, 2014, pp283-294
http://www.bipublication.com

Comparative Mitochondrial DNA Sequence and Amino Acid Analysis of the Cytochrome C Oxidase Subunit I (COI) from Two EEL Species, *MONOPTERUS CUCHIA* and *MONOPTERUS ALBUS*

Papari Devi[1], Chittaranjan Baruah[2] and Dhirendra K. Sharma[1,*]

[1]Bioinformatics Centre, Department of Zoology, Gauhati University, Guwahati – 781014, Assam, India
[2]Department of Zoology, Darrang College, Tezpur-784001, Assam, India
*Corresponding author: gauhatiuniv.btisnet@nic.in

[Received-19/04/2014, Accepted-01/07/2014]

ABSTRACT:

The present investigation includes *in silico* sequence analysis, tertiary structure prediction and phylogenetic analysis of cytochrome C oxidase subunit I (COI) from seven freshwater eels to identify the molecular differences of *Monopterus cuchia* (Hamilton) and *Monopterus albus* (Zuiew). The analyses were performed using the sequence data of COI gene and its encoded cytochrome C oxidase subunit I protein. The evolutionary analyses were performed using Maximum Likelihood and Maximum Parsimony methods. The structures of COI proteins were predicted using the template of Bovine Heart Cytochrome C Oxidase (PDB ID: 1V54). The computed instability index (23.92 to 27.49) classifies COI of eel species as stable protein. After verification, the structures of COI protein have been deposited to Protein Model Database (PMDB). The predicted structures of COI could be of use for further evaluation of molecular mechanism of function. The study has revealed several interesting differences between *M.albus* from *M. cuchia* at molecular level, which clarifies their genetic distinctness rather than species complex as suggested by conventional morphometric studies.

Key words: COI, eels, in-silico, enzyme, phylogeny

[I] INTRODUCTION

Freshwater habitats provide the occurrence of various species of freshwater eel – a fish with an elongated snake like structure. The freshwater air-breathing mud eel- *Monopterus cuchia* and swamp eel- *Monopterus albus*, are tentatively identified as belonging to the synbranchid genus *Monopterus* [1, 2]. They are regarded as species complex and require taxonomic revision [3]. Both *M. cuchia* and *M. albus* are economically important freshwater fishes, recorded from India, Nepal, Bangladesh, Myanmar and Pakistan [4 - 6]. Therefore, in the present investigation, an attempt has been made to study the molecular variation between two Synbranchid eel species- *Monopterus cuchia* and *Monopterus albus*. The nucleotide and protein sequence of COI of five other eel shaped fishes belonging to the families Anguillidae (*Anguilla bengalensis*) and Mastacembelidae (*Mastacembelus armatus, Macrognathus pancalus, Macrognathus aral, Macrognathus aculeatus*) also included in the present study to establish the evolutionary relationships of *Monopterus albus* and

Monopterus cuchia with other eel species. The present study provides a comparative account of the taxonomic differences of eel like species of freshwater habitats.

The enzyme cytochrome C oxidase (COI; EC 1.9.3.1) is a large transmembrane protein comlex encoded by mitochondrial DNA in eukaryotes. It is the last enzyme in the respiratory electron transport chain of mitochondrial DNA. The size and structure of cytochrome oxidase subunit 1 (COI) gene has been well conserved in the animal groups analyzed so far, a feature which makes it especially suitable for evolutionarystudies [7]. The COI genes are preferred over other mitochondrial genes as it has lower mutation rates [8] and the third position of the codons shows a high incidence of nucleotide substitution as compared to other protein genes [9]. There are three subunits of cytochrome C Oxidase (COI, II, III), which are widely conserved and have been used for phylogenetic inference [10-11]. In the present study, an attempt has been made for DNA and amino acid sequence analysis, tertiary structure prediction [12] and phylogenetic profile of COI from seven eel-like fish species to establish the genetic and evolutionary relationships of *Monopterus cuchia* and *Monopterus albus*.

[II] MATERIALS AND METHODS

2.1. Acquisition and alignment of sequences

The study was extended to data mining and sequence analyses of COI gene and COI protein from the sequence information extracted from GenBank (NCBI) and Protein Knowledgebase (UniProtKB), respectively [13-14] (**Table 1** and **2**). The sequences were simultaneously aligned using CLUSTAL-W [15] and Modeller version 9.12 [16] programs.

2.2. Comparative sequence analysis

The nucleotide and protein (COI) sequence analyses were performed in the CLC Genomics Workbench 7.0.3 (CLC Bio, Hyderabad). The physico-chemical parameters of COI were computed using CLC Genomics Workbench and ProtParam [17]. The important calculations for the amino acid composition, atomic composition, theoretical pI, molecular weight, Formula, extinction coefficients, half-life, instability index, aliphatic index, hydrophobicity, charge *vs.* pH were carried out under sequence analysis.

2.3. Molecular Phylogenetic analysis

The nucleotide and protein sequence of COI of other eel shaped fishes (from GenBank and UniProtKB) belonging to the families Anguillidae (*Anguilla bengalensis*), Mastacembelidae (*Mastacembelus armatus*, *Macrognathus pancalus*, *Macrognathus aral*, *Macrognathus aculeatus*) also included in the phylogenetic analysis to establish the evolutionary relationships of *Monopterus albus* and *Monopterus cuchia* with other eel species.The nucleotide and amino acid sequences of COI were separately aligned using ClustalW 1.6 [18]. Evolutionary analyses were conducted in MEGA6 [19]. The evolutionary history was inferred by using the Maximum Parsimony [20] and Maximum Likelihood methods [21]. Nucleotide substitution model that best fits each dataset and the model parameters were estimated using Akaike information criterion implemented in the program MODELTEST version 3.7 [22] (**Table 3**).

2.4. Comparative modelling of COI protein from *M. albus* and *M. cuchia*

BlastP [23] and FASTA [24] searches were performed independently with PDB [25] for obtaining a suitable template. The significance of the BLAST results was assessed by expect values (e-value) generated by BLAST family of search algorithm [26]. The target-template alignment was carried out using Modeller version 9.12 programmes [27]. Comparative modelling was conducted by the Modeller version 9.12 programme [28]. The loop regions were modeled using MODLOOP server [29]. The final structures were obtained by optimization of a molecular probability density function (pdf) of Modeller [30]. The molecular pdf for homology modelling was optimized with the variable target function procedure in

Cartesian space that employed the method of conjugate gradients and molecular dynamics with simulated annealing [31].

The structures for COI were evaluated [32] by ERRAT [33] and ProCheck [34] programmes. After fruitful verification, the coordinate files were successfully deposited to PMDB [35]**.** All the graphic presentations of the 3D structures were prepared using Chimera version 1.8.1 [36].

[III] RESULTS

3.1. Comparative sequence analysis

The COI genes of the present study ranged from 605(COI of *Anguilla bengalensis*) to 655 (COI of *Macrognathus pancalus* and *Macrognathus aral*) nucleotide long and with molecular weights of 185.602 kDa (in *A. bengalensis*) to 200.758 kDa (in *Macrognathus aral*) respectively. The melting temperature ranged from 83.40 (COI of *A. bengalensis*) to 84.49 (COI of *M. albus*) at 0.1M salt concentration (**Table 1**). The frequency of AT in COI mRNA (cDNA) sequence in different fishes of the present study ranged from 0.522 (in COI of *M. albus*) to 0.615 (in COI of *Monopterus cuchia*). On the other hand frequency of GC ranged from 0.385 (in COI gene of *Monopterus cuchia*) to 0.478 (in COI gene of *Monopterus albus*) (**Table 1**)**.** The COI gene sequences were found to be A:T rich (**Table 1**; **Figure 1**). The transition/ transversion frequency for the nucleotides of the COI gene are- A=>T = 0.05, A=>C = 0.04, A=>G= 0.1, T=>A = 0.04, T=>C= 0.19, T=>G= 0.2, C=>A= 0.04, C=>T= 0.2, C=>G=0.02, G=>A= 0.19, G=>T=0.05, G=>C=0.04.

The sizes of protein sequences of COI enzyme in the present study ranged from 216 (COI of *Anguilla bengalensis*) to 518 (COI of *M. albus*) amino acids. The amino acids Lucine (frequency=0.118 in *M. albus* to 0.157 in *M. cuchia*) followed by Alanine (frequency=0.089 in *M. albus* to 0.111 in *M. cuchia*) have been found predominantly rich in the COI of the eel species of the present study (**Figure 2**). Sequence analysis of COI protein revealed +ve hydropathy on average (0.722 in *M. albus* and 0.913 in *M. cuchia*) (**Table 2**; **Figure 3**). The molecular weight of COI in the present study ranged from 23.151 kDa (in COI of *Monopterus cuchia*) 57.519 kDa (in COI of *M. albus*). The isoelectric point of the COI ranged from 4.94 to 6.23 (**Table 2**; **Figure 4**). The Instability index of COI of the present study ranged from 23.92 to 27.49 (**Table 2**). The pair-wise sequence alignment of amino acid sequence revealed following differences in *M.albus* from *M. cuchia* - T=>M (at position 31), S=>C (at position 42), V=>I (at position 64) , S=>A and V=>I (at position 116 and 118 respectively), S=>G (at position 134), A=>S and I=>V (at position 187 and 188 respectively) and V=>I (at position 193) (**Figure 5**).

3.2. Molecular evolution of COI

The Maximum-likelihood model parameters for data sets as estimated in Modeltest are listed in Table 3. Pairwise distances of COI gene and COI protein have been depicted in the **Tables 4** & **5** respectively. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed [37].

3.2.1. COI *gene phylogeny*

The Pair-wise distance of COI gene sequences among the different eel species of the present study revealed shortest genetic distance (0.003) between *Macrognathus aral* and *Macrognathus aculeatus.* The longest genetic distance (0.375) exists between *Monopterus albus* and *Monopterus cuchia* followed by second longest genetic distance (0.356) between *Monopterus albus* and *Macrognathus aculeatus* (**Table 4**). The COI gene phylogenic analysis involved 21 nucleotide sequences for, where there were a total of 559 positions in the final dataset.

A. The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 9 most parsimonious trees (length = 419) is shown. The consistency index is (0.702128), the retention index is (0.891892), and the composite index is 0.653486 (0.626222) for all

sites and parsimony-informative sites. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei and Kumar, 2000) (**Figure 6A**).

B.The evolutionary history was inferred by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano model [38]. The tree with the highest log likelihood (-2588.6145) is 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 Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (*+G*, parameter = 0.6077)]. The rate variation model allowed for some sites to be evolutionarily invariable ([*+I*], 48.2788% sites) (**Figure 6B**).

The COI gene MP tree formed two distinct clades and revaled that *M. cuchia* is a sister taxa of *A. bengalenis* plus *M. albus* with 100% bootstrap support.*Macrognatus pancalus* is their successive sister taxa for ming a distinct clade. The second clade is formed by *Magrognathus aculeatus* and *Macrognathus aral* (boot strap support 100%) plus *Mastacembalus aramatus* as their successive sister taxa (100% bootstrap support) (**Figure 6A**).

The COI gene ML tree also formed two distinct clades with bootstrap support 99%. The first clade is formed by *Macrognathus aral* plus *M. aculeatus* followed by *Macrognathus pancalus* plus *Mastacembelus aramatus* (bootstrap support 100%) as their successive sister taxa. The second calde is formed by *Monopterus cuchia* (bootstrap support 100%) followed by *A.bengalensis* plus *Monopterus albus* as is successive sister taxa. (**Figure 6B**).

3.2.2. COI *protein phylogeny:*

The Pairwise distance of COI protein sequences among the different eel species of the present study revealed shortest genetic distance (0.001) between *Macrognathus aral* and *Macrognathus aculeatus.* The longest genetic distance (0.177) exists between *Monopterus albus* and *Anguilla bengalensis.* (**Table 5**).

The analysis involved 18 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 164 positions in the final dataset.

A. The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 10 most parsimonious trees (length = 41) is shown. The consistency index is (0.888889), the retention index is (0.965517), and the composite index is 0.918419 (0.858238) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [39] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates) (**Figure 7A**).

B. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Reversible Mitochondrial + Freq. model [40]. The tree with the highest log likelihood (-706.9901) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial 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 (**Figure 7B**).

The COI protein MP tree formed two distinct clades and revaled that *M. cuchia* is a successive sister taxa of *A. bengalenis* and *M. albus.* The second clade is formed by *Macrognathus aculeatus* and *Macrognathus aral* plus *Magrognathus aculeatus* and *Macrognathus aral* (boot strap support 893%) followed by *Mastacembalus aramatus* as their successive sister taxa (87% bootstrap support) (**Figure 7A**).

The COI protein ML tree also formed two major clades and revaled that *Macrognathus aculeatus* and *Macrognathus aral* are having close evolutionary relationship, *Macrognathus puncalus* is their sister taxa and these three taxa

formed a clade with their successive sister taxon *Mastacembelus armatus* (bootstap support 91%). The second clade is formed by *Monopterus cuchia* followed by its successive sister taxa of *M. albus* and *A. bengalensis* (**Figure 7B**).

3.3. Tertiary structures of COI from *M. albus* and *M. cuchia*

The tertiary structure of COI for *M. cuchia* has 9 helices, 15 helix-helix interacs, 10 beta turns, 1 gamma turns, whereas the tertiary structure of COI for *M. albus* has 22 helices, 60 helix-helix interacs, 28 beta turns, 2 gamma turns(Figure 8 A-B). The ERRAT verification revealed that the overall quality factor for the predicted tertiary structures of COI is around 95%. ProCheck verification revealed structural accuracy as more than 90% of the aminoacid residues in the predicted 3D structures of COI are in the range of most favoured region, which confirms the validity of the predicted structures (**Figure 9**).

[IV] DISCUSSION

Efficient identification of the two Synbranchid eel species of the present study is critical for aquaculture management as well as for eel conservation [41]. Thus, identification of *M. cuchia* and *M. albus* has been supported by molecular characterization in the present study instead of conventional methods [42]. The present study have revealed an interesting point of difference for identification of the two Synbranchid species that the cDNA sequence COI-gene of *M. cuchia* is more A:T rich than that of *M. albus* (**Figure 1**). Moreover the COI protein of *M. cuchia* has more Lucine (L) and Alananine (A) residues frequency than that of *M. albus* (**Figure 2**). The Instability index in COI is in the range of stable molecule. The hydropathicity plot (**Figure 3**) revealed that the COI protein is hydrophobic in nature. The COI gene of present study has showed higher transversion frequency than the transition frequency for the nucleotides of the eel species (A=>G= 0.1, G=>A= 0.19). The graph of electrical charge as a function of pH for COI (**Figure 4**) revealed significant differences in the isoelectric points of *M. cuchia* (*pI* =4.94) and *M. albus* (*pI* = 6.23). The pair-wise sequence alignment of amino acid sequence revealed 10 specific positions where amino acid sequence differences exist between *M.albus* from *M. cuchia* (**Figure 5**).

Both *Monopterus albus* and *Monopterus cuchia* are regarded as species complex and demands taxonomic revision [43]. However, in the present study, the genetic distance of COI gene (0.375) revealed that the two *Monopterus albus* and *Monopterus cuchia* are not within a species complex but they are two well established distantly related species. The COI gene MP tree revealed that *M. albus* has more close evolutionary relationship with *A. bengalensis* than that of *M.cuchia* and *M. albus* (**Figure 6A**).The COI gene ML tree also revealed more evolutionary close relatedness between *A.bengalensis* and *Monopterus alcus* than that of *M. albus* and *M. cuchia* (**Figure 6B**).

The MP tree of COI protein revealed distinct evolutionary profile of *M. cuchia* and *M. albus* and out of this two species *M. albus* has a very close evolutionary relationship with *Anguilla bengalensis*) (**Figure 7A**).The ML tree of COI protein revealed that *M. albus* has a very close evolutionary relationship with *Anguilla bengalensis* than that of *M.cuchia* of the same genus. All the phylogenetic trees revealed an interesting point that *Monopterus cuchia* (family- Synbranchidae) is intermediate taxa of the taxa belonging to family Mastacembelidae and 'Anguillidae plus Synbranchidae' (**Figure 7B**). The tertiary structures of COI is found to be structurally conserved and revealed that COI is a alpha-domain protein.

[V] CONCLUSION

Understanding of genetic relationships between the two species of eels under the genus *Monopterus* is listed as high priority in the IUCN record [43]. The present study will certainly be helpful in understanding genetic variation between *M. cuchia* and *M. albus* and will clarify taxonomic uncertainties mentioned by earlier workers. Further, a microsattelite

based studies on northeast Indian population of *Monopterus* species complex by the present authors (unpublished data) also showed genetic distinctness of *M. cuchia* and *M. albus* rather than a species complex.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Biotechnology, Govt. of India for financial support to establish the Bioinformatics Infrastructure Facility (BIF) at the Zoology Department, Gauhati University, which has been utilized in the present study.

REFERENCES

1. Collins TM, Trexler J, Nico L, Rawlings TA (2002). Genetic diversity in a morphologically conservative invasive taxon: multiple introductions of swamp eels to the southeastern United States. Conserv. Biol., 16: 1024-1035.
2. Li WT, Liao XL, Yu XM, Cheng L, Tong J (2007). Isolation and characterization of polymorphic microsatellites in a sex reversal fish, rice field eel (*Monopterus albus*). Mol. Ecol. Notes, 7: 705-707.
3. Dahanukar, N. 2010. *Monopterus cuchia*. In: IUCN 2013. IUCN Red List of Threatened Species. Version 2011.2. www.iucnredlist.org
4. Menon, A.G.K. (1999). *Check list - fresh water fishes of India.* Records of the Zoological Survey of India, Occasional Paper No. 175.
5. Mirza, M.R. and Alam, M.K. (2002). A checklist of the fishes of the Punjab, Pakistan. *Records of the Zoological Survey of Pakistan* 14: 31-35.
6. Zhou R, Cheng H, Zhang Q, Guo Y, Richard RC, Terrence RT (2002) SRY-related genes in the genome of the rice field eel (Monopterus albus). Genet Sel Evol 34:129–137.
7. Lunt D. H., Zhang D. X., Szymura J. M. and Hewitt G. M. (1996). The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol. Biol*. 5, 153–165.
8. Yi, S., Ellsworth, D. L., and Li, W. H. (2002) Slow molecular clocks in old world monkeys, apes and humans. *Mol. Biol and Evol*. 19, 2191-2198.
9. Mcclellan, D. A. (2000) The codon-degeneracy model of molecular evolution. *J. Mol. Evol.* 50, 131-140.
10. Hatefl Y. (1985).The mitochondrial electron transport and oxidative phosphorylation system. *Ann Rev Biochem* 54: 1015-1069.
11. Attardi G, Schatz G. (1988). Biogenesis of mitochondria. *Ann Rev Cell Biol* 4: 289-333.
12. Zemla A, Venclovas C, Moult J, Fidelis K (1999). Processing and analysis of CASP3 protein structure predictions Proteins, Suppl. 3: 22–29.
13. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nuc. Ac. Res. 31:365-370.
14. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2004). UniProt: the Universal Protein knowledgebase. Nuc. Ac. Res. 32:D115-119.
15. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc. Ac. Res. 22:4673-4680.
16. Fiser A, Do RK, Sali A (2000). Modelling of loops in protein structures, Prot Sci. 9: 1753-1773.
17. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005). Protein Identification and Analysis Tools on the ExPASy Server ;(In) *John M. Walker (ed): The Proteomics Protocols Handbook,* Humana Press pp. 571-607.
18. Thompson JD, Higgins DG, Gibson TJ, (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc. Ac. Res. 22: 4673–4680.
1. 19.Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*30: 2725-2729.
19. Eck, R. V., Dayhoff, M. O., (1966). Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Maryland, USA.
20. Jones DT, Taylor WR, Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Applic. Biosci. 8: 275-282.
21. Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics. 14: 817–818.
22. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc. Ac. Res. 25(17): 3389-3402.

23. Pearson WR (1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics. 11(3): 635-50.
24. Berman H, Henrick K, Nakamura H, Markley JL (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nuc. Ac. Res. 35(Database issue):D301-3.
25. Altschul SF (1991). Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219(3):555–565.
26. Lassmann T, Sonnhammer EL (2005). Automatic assessment of alignment quality. Nuc. Ac. Res. 33(22):7120-7128.
27. Martí-Renom MA, Yerkovich B, Sali A (2002). Comparative protein structure prediction. John Wiley & Sons, Inc. Curr. Protocol. Prot. Sci.1: 2.9.1 - 2.9.22.
28. Fiser A, Sali A (2003). ModLoop: automated modelling of loops in protein structures Bioinformatics 18: 2500-01.
29. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006). Comparative Protein Structure Modelling Using MODELLER. Ed: Coligan JE, Dunn BM, Speicher DW, Wingfield PT. Current Protocols in Protein Science Unit. 2.9 1-31.
30. Sali A, Blundell TL (1993).Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234: 779-815.
31. Giorgetti A, Raimondo D, Miele AE, Tramontano A (2005). Evaluating the usefulness of protein structure models for molecular replacement. Bioinformatics 21(2): 72–76.
32. Colovos C, Yeates TO (1993). Verification of protein structures: Patterns of nonbounded atomic interections. Prot. Sci. 2: 1511-1519.
33. Laskowski RA, Watson JD, Thornton JM (2003). From protein structure to biochemical function. J. Struct. Funct. Genomics. 4: 167–177.
34. Tiziana C, Paolo D'Onorio DM, Domenico C, Talamo IG, Tramontano A (2006). The PMDB Protein Model Database CASPUR, Nuc. Ac. Res. 2006 34(Database Issue): D306-D309.
35. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin, TE (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 25:1605-1612.
36. Felsenstein J (1985). Confidence limits on phytogenies: An approach using the bootstrap. Evol. 39: 783-791.
37. Hasegawa M., Kishino H., and Yano T. (1985). Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22:160-174.
38. Nei M, Kumar S (2000). Molecular evolution and phylogenetics. Oxford University Press, Oxford.
39. Adachi J. and Hasegawa M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Molecular Evolution* 42:459-468.
40. Dudu A., Georgescu S. E.,& Costache M. 2010. PCR-RFLP method to identify fish species of Importance. *Archiva Zootechnica*, 13(1):53-59.
41. Huang J. P., Han Y. S., Tzeng W. N. 2001. Species identification of Anguillid eels by polemerase chain reaction/restriction fragment length polymorphism analysis of the gonadotrophin II-β subunit gene. Acta Zoologica Taiwanica, 12(2): 41-49.
42. IUCN. 2013. IUCN Red List of Threatened Species (ver. 2013.2). Available at: www.iucnredlist.org.

Table 1. Nucleotide sequence statistics of the COI cDNA sequence.

Statistical parameter	***M. cuchia***	***M. albus***	***Anguilla bengalensis***	***Mastacembelus armatus***	***Macrognathus pancalus***	***Macrognathus aral***	***Macrognathus aculeatus***
GenBank Accession numbers	KF742427	AP002945	JX887590	JX983365	FJ459512	HQ219137	JX260905
Length (bp)	652bp	655bp	605bp	652bp	655bp	655bp	651bp
MW in single stranded condition (kDa)	200.473 kDa	200.571 kDa	185.602 kDa	199.973 kDa	200.715 kDa	200.758 kDa	199.515 kDa
Melting temperature (^{0}C) [salt] = 0.1M	80.68	84.49	83.40	83.83	83.80	82.86	82.79
Frequency of A + T	0.615	0.522	0.549	0.538	0.539	0.562	0.564
Frequency of C + G	0.385	0.478	0.451	0.462	0.461	0.438	0.436

Table 2. COI protein statistics

Statistical parameter	*M. cuchia*	*M. albus*	*Anguilla bengalensis*	*Mastacembelus armatus*	*Macrognathus pancalus*	*Macrognathus aral*	*Macrognathus aculeatus*
UniProtKB Accession number	W0I5X7	Q94SG6	S4SNW0	S4SMK7	B1A2X4	G3DQ16	I1TM91
No. of amino acids	217aa	518aa	216aa	217aa	218aa	218aa	217aa
MW (kDa)	23.151 kDa	57.519 kDa	23.157 kDa	23.163 kDa	23.34 kDa	23.354 kDa	23.207 kDa
pI	4.94	6.23	6.02	4.94	4.94	4.94	4.94
-ve charged residues	9	25	8	9	9	9	9
+ve charged residues	4	17	6	4	4	4	4
Formula	$C_{1084}H_{1680}N_{254}O_{281}S_{12}$	$C_{2728}H_{4056}N_{638}O_{673}S_{30}$	$C_{1083}H_{1654}N_{256}O_{280}S_{13}$	$C_{1086}H_{1684}N_{254}O_{282}S_{1}$	$C_{1096}H_{1695}N_{255}O_{284}S_{11}$	$C_{1097}H_{1697}N_{255}O_{284}S_{11}$	$C_{1088}H_{1688}N_{254}O_{283}S_{11}$
II	24.45	26.78	27.49	23.92	24.74	24.25	24.31
AI	123.23	106.02	105.23	125.02	124.45	124.45	125.02
GRAVY	0.913	0.722	0.702	0.917	0.915	0.916	0.907

MW: Molecular weight; pI: Isoelectric point; II: Instability index; AI: Aliphatic index; GRAVY: Grand average of hydropathicity.

Table 3. Maximum-likelihood model parameters for data sets as estimated in Modeltest (Posada and Crandall, 1998)

Parameter	**COI gene**	**COI protein**
Model	HKY+G+I	mtREV24
Bayesian Information Criterion (BIC) scores	5581.044	1698.75
Akaike Information Criterion, corrected (AICc) value	5249.71	1501.84
Maximum Likelihood value (*lnL*)	-2579.71	-717.53
Gamma distribution (*G*)	1.4366	n/a
invariable (*I*)	0.4832	n/a
Transition/Transversion bias (*R*)	2.132	0.019
Total positions in the final dataset	559	164

Table 4. Pairwise distance COI gene

		1	2	3	4	5	6	7
1	*Monopterus cuchia*	-						
2	*Monopterus albus*	**0.375**	-					
3	*Anguilla bengalensis*	0.277	0.320	-				
4	*Macrognathus aculeatus*	0.282	0.356	0.300	-			
5	*Macrognathus aral*	0.285	0.352	0.297	**0.003**	-		
6	*Macrognathus pancalus*	0.295	0.355	0.292	0.178	0.183	-	
7	*Mastacembelus armatus*	0.273	0.340	0.317	0.175	0.180	0.171	-

Table 5. Pairwise distance COI protein

		1	2	3	4	5	6	7
1	*Monopterus cuchia*	-						
2	*Monopterus albus*	0.047	-					
3	*Anguilla bengalensis*	0.150	**0.177**	-				
4	*Macrognathus aculeatus*	0.028	0.072	0.166	-			
5	*Macrognathus aral*	0.028	0.072	0.166	**0.001**	-		
6	*Macrognathus pancalus*	0.038	0.082	0.166	0.014	0.014	-	
7	*Mastacembelus armatus*	0.028	0.072	0.166	0.009	0.009	0.019	-

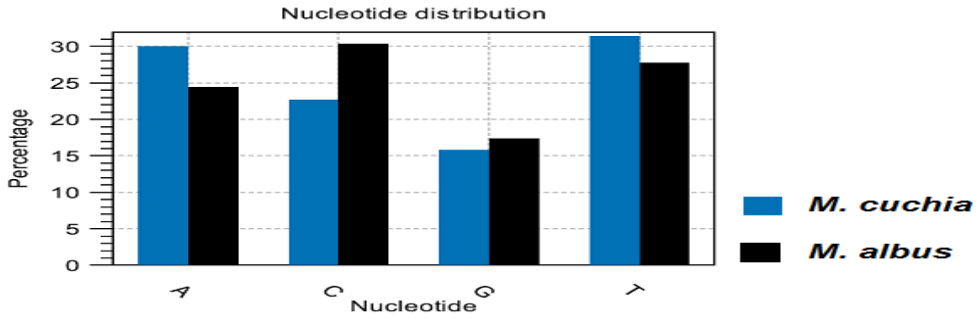


Figure 1. Comparative nucleotide composition (% in average) in the COI cDNA sequence of *M. cuchia* and *M. albus.*

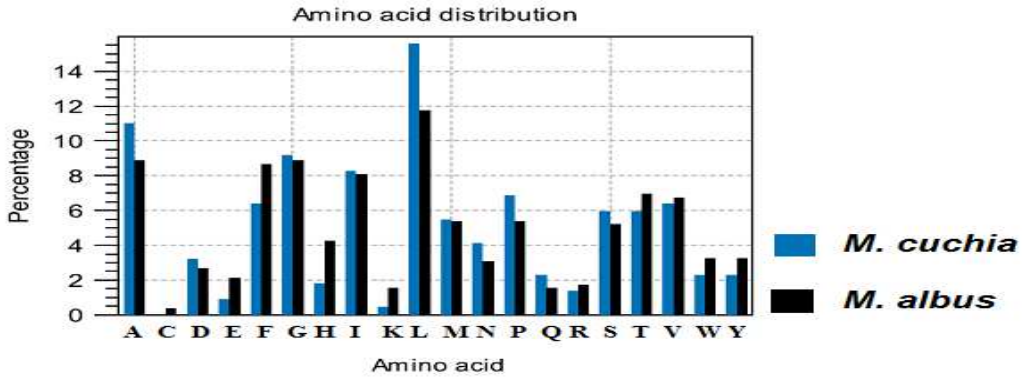
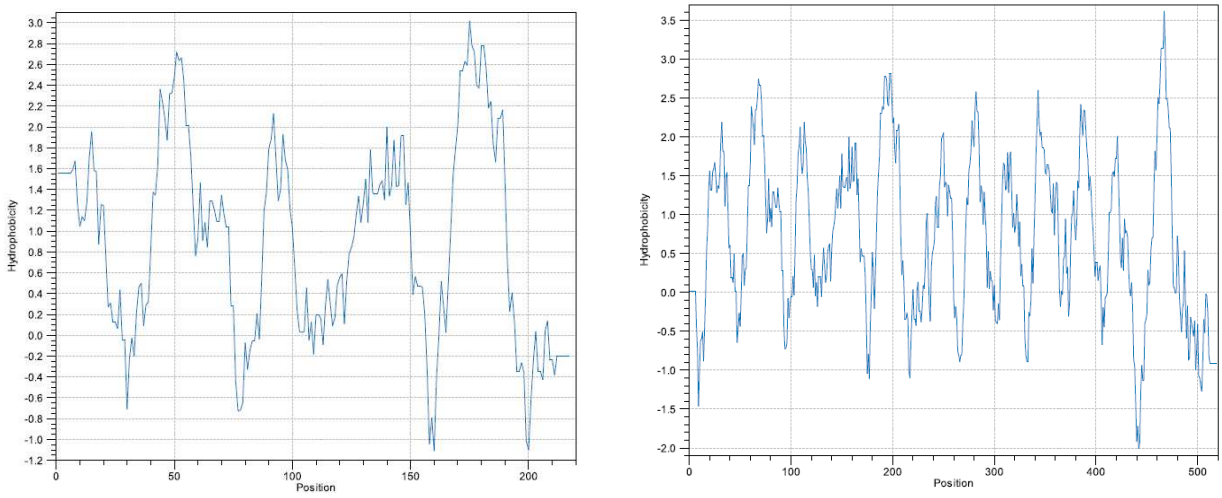


Figure 2. Distribution of amino acids for COI protein of *M. cuchia* and *M. albus.*



A. B.

Figure 3. Plot of local Hydropathy for COI (Kyte-Doolittle scale). A. COI of *M. cuchia*, B. COI of *M. albus*

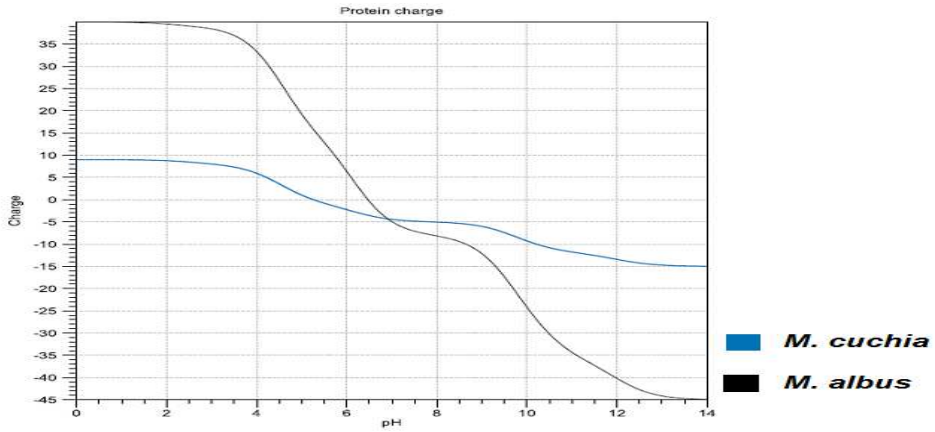


Figure 4. Electrical charge as a function of pH for COI in the two eel species.

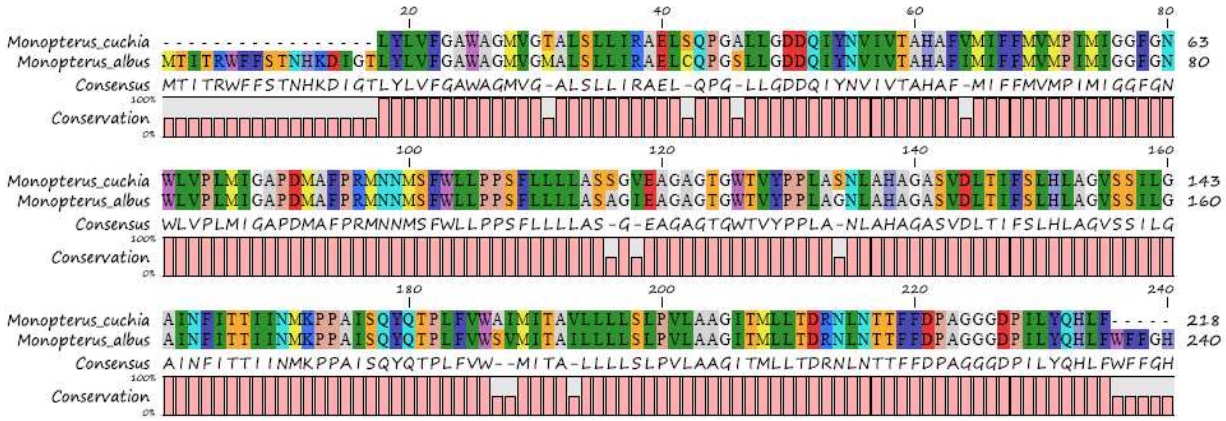
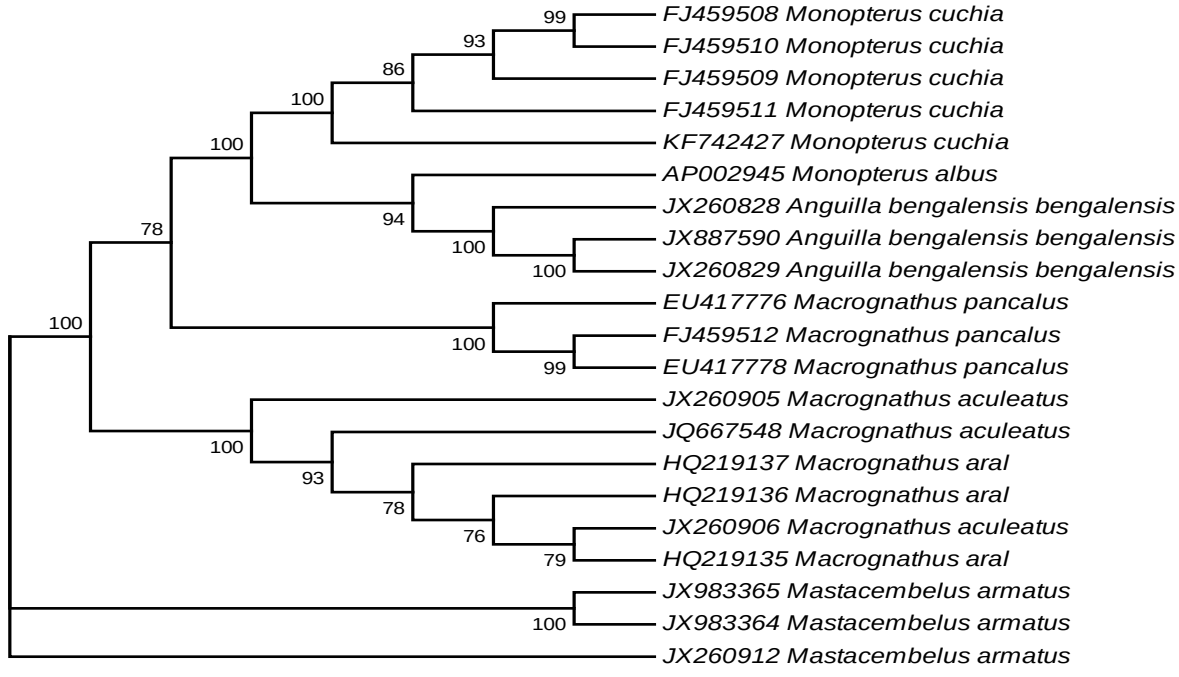
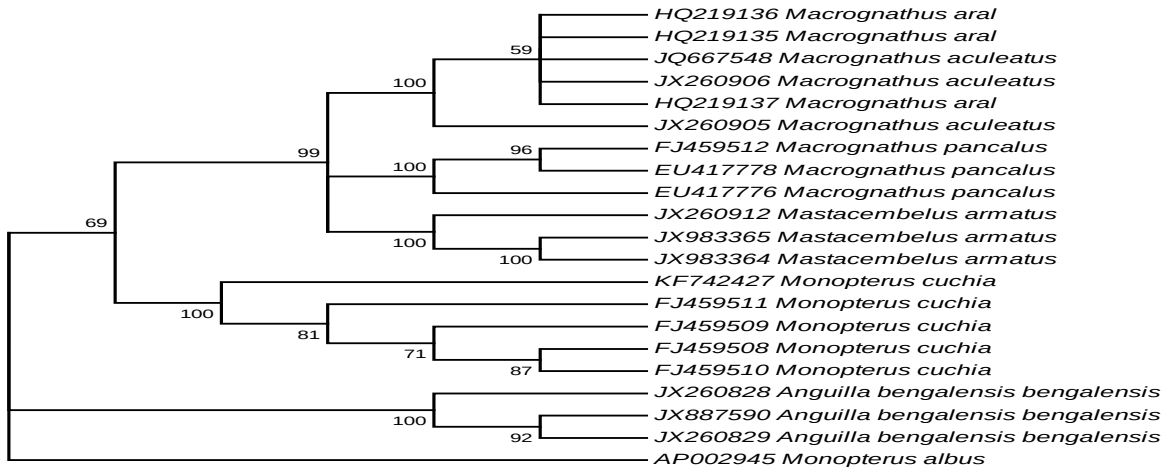


Figure 5. Multiple amino acid sequence alignment of COI protein in *M. cuchia* and *M. albus.* '-' represent sequence not conserved. The sizes of the bar diagram represent the degree of conservation of respective amino acid in each alignment position.

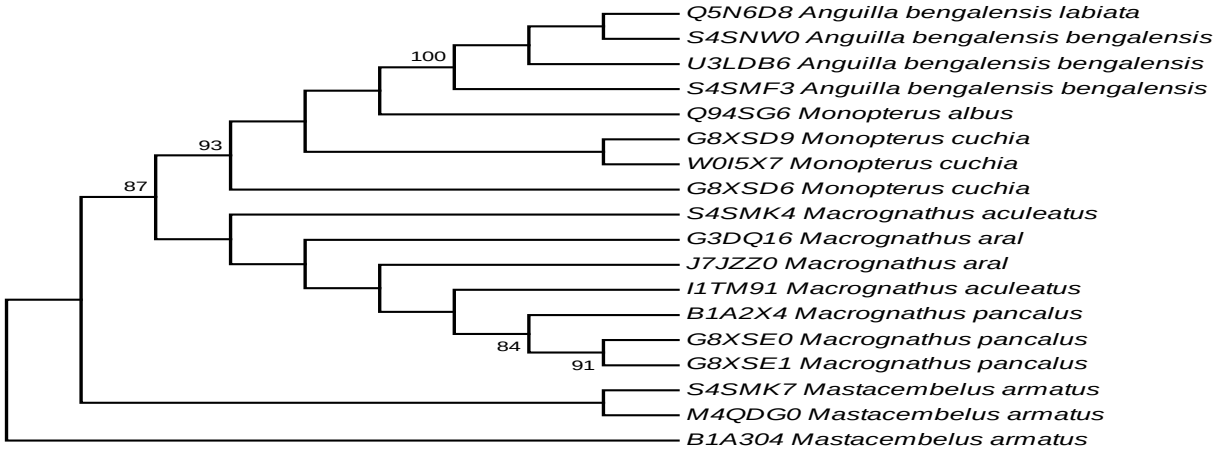


A.

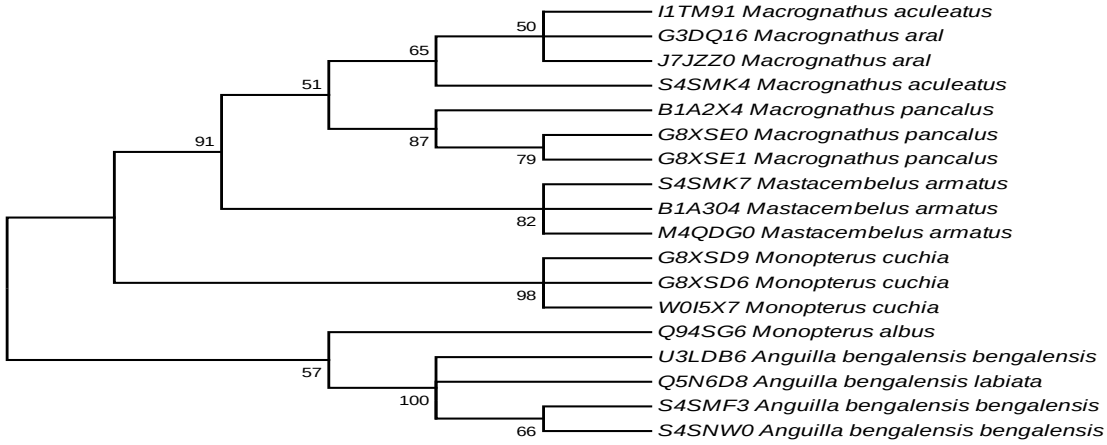


B.

Figure 6. Molecular phylogenetic anaylsis of COI gene. A. Maximum Parsimony tree, B. Maximum Likelihood tree based on Hasegawa-Kishino-Yano model (Hasegawa *et a*l., 1985). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985).The scale bars represent the branch lengths measured in the number of changes (substitutions per site) over the whole sequence.



A.



B.

Figure 7. Molecular phylogenetic anaylsis of bony fish COI protein. A. Maximum Parsimony tree, B. Maximum Likelihood tree based on the General Reversible Mitochondrial + Freq. model (Adachi and Hasegawa, 1996). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein,1985).The scale bars represent the branch lengths measured in the number of changes (substitutions per site) over the whole sequence.

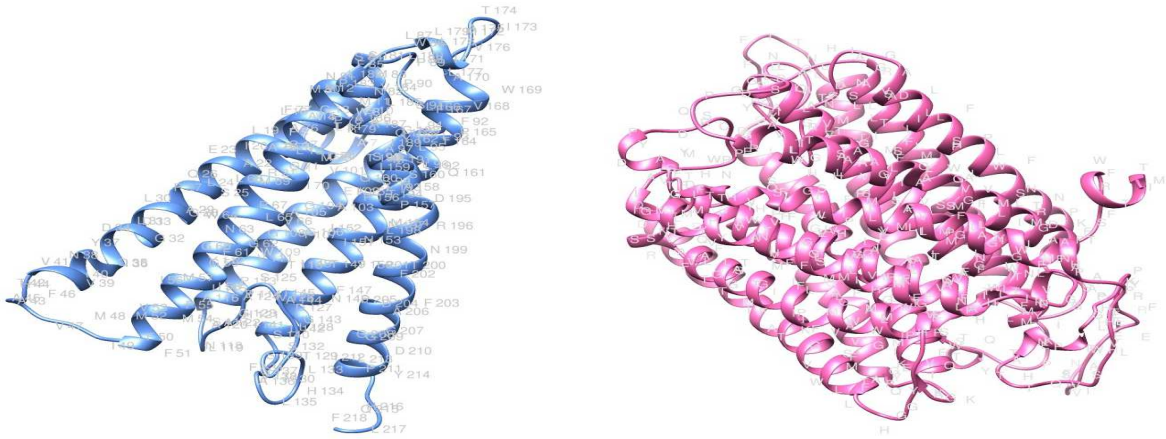


Figure 8. The predicted homology model of COI structure, as displayed by UCSF Chimera. A. *M. cuchia*, B. *M. albus*

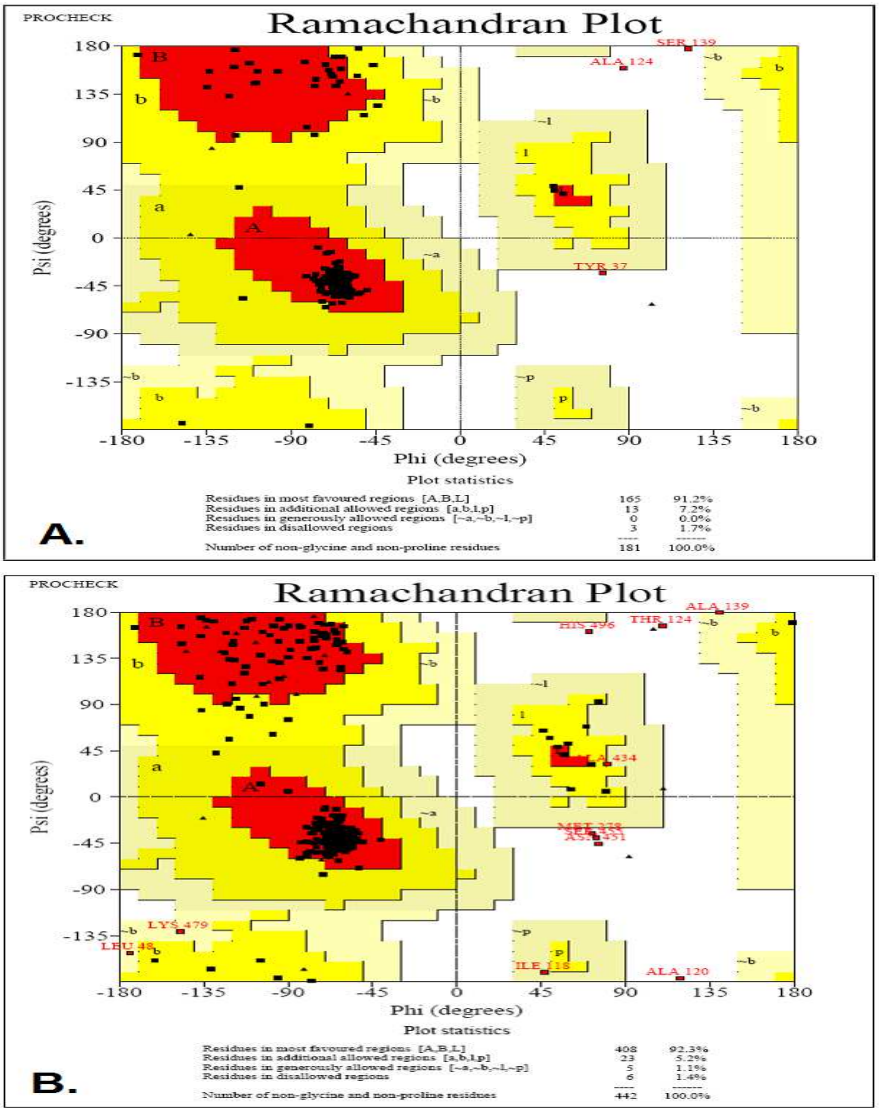


Figure 9. Ramachandran analysis of the backbone dihedral angles PSI (y) and PHI (f) for the final structure of COI from A. *Monopterus cuchia*, B. *Monopterus albus*. Red region represents the most favored region, yellow = allowed region, light yellow = generously allowed region, white = disallowed region [ProCheck].