Barathkumar TR and Thangaraj M (2020)
Notulae Scientia Biologicae 12(4):781-793
DOI:10.15835/12410822
Research Article

# Phylogenetic observation in Ariidae, Bagridae and Plotosidae catfishes by *COI* gene sequence analysis

Thathiredypalli R. BARATHKUMAR, Muthusamy THANGARAJ*

*Annamalai University, Faculty of Marine Sciences, Centre for Advanced Study in Marine Biology, Parangipettai, Tamil Nadu 608 502, India; biotricknologist@gmail.com; coralholder@yahoo.com (*corresponding author)*

## Abstract

To understand the phylogenetic status of Ariidae, Bagridae and Plotosidae catfishes, this study was planned using the barcode gene, cytochrome oxidase I (*COI*). Totally 71 species were used in phylogenetic reconstructions under maximum parsimony, maximum likelihood and Bayesian inference criteria. The one-way ANOVA showed that the three catfish families are significantly different (F = 19.79, d.f. = 3; 116, *P*< 0.0001 (Plotosidae); F = 44.21, d.f. = 3; 986, *P*< 0.0001 (Ariidae); F = 24.83, d.f. = 3; 1322, *P*< 0.0001 (Bagridae). In MP, ML and BI based phylogenetic tree of Ariidae, *Plicofollis* genus displayed as a monophyletic group with higher bootstrap and posterior probability values for all the species except two species of *Neoarius,* which intervened separating *P. polystaphylodon.* In the phylogenetic tree of Plotoside, *Plotosus* genus displayed as monophyletic group with higher bootstrap and posterior probability values for all the eight species. In the case of Bagridae phylogeentic tree, *Mystus* genus displayed as a monophyletic group with higher bootstrap and posterior probability values for all the species except *Mystus montanus* forming a distant and distinct clade whereas *Mystus tengara* collides into monophyletic clade when *Neotropius* genuswas removed. By this study we could establish a phylogenetic hypothesis for all the 36 catfish families and examine the monophyly status of the subfamilies and genera.

***Keywords:*** catfish; cytochrome oxidase I; monophyly; phylogeny; transition; transversion

## Introduction

In zoological and ecological literature, identification of unknown specimens based on cytochrome oxidase I (*COI*) has become known as DNA barcoding (Hebert *et al.*, 2003; Remigio and Hebert, 2003; Moritz and Cicero, 2004). DNA barcoding has found a wide range of applications, from identification of specimens in conservation biology and molecular ecology. DNA barcoding system for animal life could be based upon sequence diversity in 5′ section of *COI* gene. *COI* exhibits a greater range of phylogenetic signal than any other mitochondrial genes (Hebert *et al.*, 2003). As in other protein-coding genes, the third-position nucleotides of *COI* show a high incidence of base substitution, leading to a rate of molecular evolution that is about three times greater than that of 12S or 16S rRNA (Knowlton and Weigt, 1998). They also argued that 12S and 16S rRNA genes are having multiple insertions and deletions so they pose potential problems in their alignment. This problem would apply as well to the nuclear 28S rRNA and internal transcribed spacer regions (ITS).

*Received: 26 Sep 2020. Received in revised form: 24 Nov 2020. Accepted: 07 Dec 2020. Published online: 21 Dec 2020.*

The classical procedure for such molecular identification has been the use of Blast searches (Altschul *et al*., 1997). Blast offers no information to help researchers choose among multiple close matches. Whereas the local alignment problem can be circumvented using global alignments, the remaining two problems cannot be addressed without a statistical evaluation of the phylogenetic associations among species. This question is difficult to address as the evolutionary relationship among genetic markers may not truly reflect the evolutionary relationship among species. In cases, where reciprocal monophyly cannot safely be assumed, an analysis quantifying between-species and between-genera genetic variation forms a more correct basis of assignment. Such analyses, however, require comprehensive different methods of phylogenetic coverage that is generally not available to the biologist. This study addresses the species problem but instead attempt to devise and display different methods of phylogenetic analysis for the assignment of 16 species of catfishes to their family taxa. Different phylogenetic methods lead to improved accuracy and importantly, it provides a measure of statistical confidence associated with the barcoding assignment.

## Materials and Methods

*Sample collection*

The catfish samples were collected from Mudasalodai (11° 45′19″ N; 79° 47′45″ E), Cuddalore (11° 42′37″ N; 79° 46′28″ E), Perumal Lake (11° 34′30″ N; 79° 40′18″ E) and Vadavar River (11° 08′03″ N; 79° 27′05″ E) of Tamil Nadu state, India. They were identified by standard reference book (Jayaram, 1984). Dorsal fin tissues were taken out and preserved in 95% ethanol for DNA extraction.

*DNA isolation, PCR and sequencing*

Genomic DNA was isolated by standard phenol/ chloroform method (Sambrook *et al*., 1989) and the concentration of isolated DNA was estimated using a UV spectrophotometer. The DNA was diluted in TAE buffer to a final concentration of 100 ng/µL. Cytochrome c oxidase-1 (CO1) gene was amplified in a 50 µL volume PCR mix with 5 µL of 10X Taqpolymerase $MgCl_2$ (50 mM) buffer, 1µL of each dNTP (0.05 mM), 1 µL of each primer (0.01 mM), 0.6 U of Taqpolymerase, 2 µl of genomic DNA and 36 µl of double distilled water. The universal primer, *FishF1*-5' TCAACCAACCACAAAGACATTGG CAC3' and *FishR1*-'TAGACTTCTGGGTGGCCAAAGAATC 3' (Ward *et al*., 2005) was used for the amplification of the CO1 gene. The thermal regime consisted of an initial step of 2 min at 95 °C followed by 35 cycles of 40 s at 94 °C, 40 s at 54 °C and 1in 10 s at 72 °C followed in turn by final extension of 10 min at 72 °C. The PCR products were visualized on 1.5% agarose gels, and the most intense amplicons were elected for sequencing. The cleaned-up PCR product was sequenced by a commercial sequencing facility (Eurofins, Bangalore). The CO1 gene partial sequences of 16 individuals were edited using MEGA 5.0 (Kumar *et al*., 2011) and aligned with Clustal W 1.6, implemented in same software. Thehaplo type definitions have been submitted to the NCBI GenBank through BankIt.

*Phylogenetic analysis*

The probability of rejecting the null hypothesis that sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences (Disparity index test), (Kumar and Gadagkar, 2001). A Monte Carlo test (500 replicates) was used to estimate the *P*-values (Kumar and Gadagkar, 2001). *P*-values < 0.05 are considered as significant. The estimates of the disparity index per site are shown for each sequence pair. The analysis involved 45 (Ariidae), 16 (Plotosidae) and 52 (Bagridae) nucleotide sequences. Codon positions included were $1^{st}+2^{nd}+3^{rd}$+Noncoding. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA5.0 (Kumar *et al*., 2011).

Totally 71 species were used in phylogenetic reconstructions under maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) criteria. Among them,16 species were our own data, the remaining 55 species data were retrieved from the GenBank. Their accession number and source details are listed in Table 1. The final alignment of the mitochondrial 5′ section of *COI* partial gene included 639 bp (Ariidae), 589 bp (Bagridae) and 640 bp (Plotosidae).

**Table 1.** List of accession numbers and source of *COI* gene sequences used in this study

| S.No | NCBI Accession number | Species | Source | Family |
|---|---|---|---|---|
| 1. | KF824816 - KF824819 | *Neotropius atherinoides* | Present study | Bagridae |
| 2. | JX460965 | *Neotropius khavalchor* | Unpublished | Bagridae |
| 3. | KF824794 - KF824797 | *Mystus bleekeri* | Present study | Bagridae |
| 4. | KF824798 - KF824802 | *Mystus atrifasciatus* | Present study | Bagridae |
| 5. | KF824803 - KF824807 | *Mystus dibrugarensis* | Present study | Bagridae |
| 6. | KF824808 - KF824812 | *Mystus albolineatus* | Present study | Bagridae |
| 7. | KF824813 - KF824815 | *Mystus gulio* | Present study | Bagridae |
| 8. | JN228949 | *Mystus vittatus* | Unpublished | Bagridae |
| 9. | HQ219109 | *Mystus malabaricus* | Unpublished | Bagridae |
| 10. | JN228946 | *Mystus cavasius* | Unpublished | Bagridae |
| 11. | HQ219128 | *Mystus montanus* | Unpublished | Bagridae |
| 12. | FJ459517 | *Mystus tengara* | Unpublished | Bagridae |
| 13. | HQ009492 | *Mystus oculatus* | Unpublished | Bagridae |
| 14. | EU490863 | *Mystus bocourti* | Unpublished | Bagridae |
| 15. | JX177677 | *Mystus multiradiatus* | Unpublished | Bagridae |
| 16. | JX177678 | *Mystus rhegma* | Unpublished | Bagridae |
| 17. | JQ289146 | *Mystus singaringan* | Unpublished | Bagridae |
| 18. | FJ170791 | *Mystus horai* | Unpublished | Bagridae |
| 19. | JQ289145 | *Mystus mysticetus* | Unpublished | Bagridae |
| 20. | EU490857 | *Bagrus docmac* | Unpublished | Bagridae |
| 21. | JQ289151 | *Bagrichthys macracanthus* | Unpublished | Bagridae |
| 22. | EU490856 | *Bagrichthys macropterus* | Unpublished | Bagridae |
| 23. | JX260836 | *Batasio tengana* | Unpublished | Bagridae |
| 24. | HQ009500 | *Batasio travancoria* | Unpublished | Bagridae |
| 25. | EU490858 | *Batasio tigrinus* | Unpublished | |
| 26. | HQ009503 | *Horabagrus nigricollaris* | Unpublished | Bagridae |
| 27. | JQ289150 | *Pseudomystus siamensis* | Unpublished | Bagridae |
| 28. | JX460967 | *Horabagrus brachysoma* | Katwate *et al.*,2012 | Bagridae |
| 29. | HM882793 | *Bagrus filamentosus* | Nwani *et al.,* 2011 | Bagridae |
| 30. | HM882791 | *Bagrus bajad* | Nwani *et al.,* 2011 | Bagridae |
| 31. | KF824820 - KF824822 | *Arius arius* | Present study | Ariidae |
| 32. | KF824823 - KF824825 | *Arius jella* | Present study | Ariidae |
| 33. | KF824826 - KF824828 | *Arius maculatus* | Present study | Ariidae |
| 34. | KF824829 - KF824831 | *Arius gagora* | Present study | Ariidae |
| 35. | KF824832 - KF824834 | *Arius subrostratus* | Present study | Ariidae |
| 36. | HQ682626 | *Arius manillensis* | Santos and Quilang, 2011 | Ariidae |
| 37. | HQ682609 | *Arius dispar* | Santos and Quilang, 2011 | Ariidae |
| 38. | JX198217 | *Arius venosus* | Unpublished | Ariidae |
| 39. | KF824835 - KF824837 | *Plicofollis tenuispinis* | Present study | Ariidae |
| 40. | KF824838 - KF824840 | *Plicofollis platystomus* | Present study | Ariidae |

| | | | | |
|---|---|---|---|---|
| 41. | KF447876 | *Plicofollis polystaphylodon* | Unpublished | Ariidae |
| 42. | JX198180 | *Plicofollis argyropleuron* | Unpublished | Ariidae |
| 43. | FJ918912 | *Ariopsis felis* | Unpublished | Ariidae |
| 44. | GU702401 | *Bagre bagre* | Unpublished | Ariidae |
| 45. | GU225557 | *Bagre marinus* | Unpublished | Ariidae |
| 46. | JQ365264 | *Cathorops spixii* | Ribeiro*et al.,* 2012 | Ariidae |
| 47. | JX515603 | *Cephalocassis jatia* | Unpublished | Ariidae |
| 48. | JF493496 | *Galeichthys feliceps* | Unpublished | Ariidae |
| 49. | JF493494 | *Galeichthys ater* | Unpublished | Ariidae |
| 50. | JQ365364 | *Genidens genidens* | Ribeiro*et al.,* 2012 | Ariidae |
| 51. | JX124786 | *Genidens barbus* | Unpublished | Ariidae |
| 52. | EF609287 | *Neoarius midgleyi* | Ward and Holmes, 2007 | Ariidae |
| 53. | EF609286 | *Neoarius graeffei* | Ward and Holmes, 2007 | Ariidae |
| 54. | EF607328 | *Netuma thalassina* | Zhang, 2011 | Ariidae |
| 55. | JX124821 | *Notarius grandicassis* | Unpublished | Ariidae |
| 56. | JQ365228 | *Notarius luniscutis* | Ribeiro*et al.,* 2012 | Ariidae |
| 57. | EF609566 | *Osteogeneiosus militaris* | Lakra *et al.,* 2011 | Ariidae |
| 58. | EU751952 | *Potamarius nelsoni* | Valdez-Moreno *et al.,* 2009 | Ariidae |
| 59. | GU225657 | *Sciades assimilis* | Unpublished | Ariidae |
| 60. | FJ418755 | *Ageneiosus inermis* | Ardura *et al.,* 2010 | Ariidae |
| 61. | EU490849 | *Ageneiosus ucayalensis* | Unpublished | Ariidae |
| 62. | KF824841 - KF824843 | *Plotosus lineatus* | Present study | Plotosidae |
| 63. | KF824844 - KF824846 | *Plotosus limbetus* | Present study | Plotosidae |
| 64. | KF824847 - KF824849 | *Plotosus canius* | Present study | Plotosidae |
| 65. | JF494191 | *Plotosus nkunga* | Unpublished | Plotosidae |
| 66. | HM006974 | *Neosilurus hyrtlii* | Page and Hughes, 2010 | Plotosidae |
| 67. | HM006990 | *Tandanus tandanus* | Page and Hughes, 2010 | Plotosidae |
| 68. | EF609335 | *Cnidoglanis macrocephalus* | Unpublished | Plotosidae |
| 69. | HM006980 | *Porochilus rendahli* | Unpublished | Plotosidae |
| 70. | JN021312 | *Pangasius bocourti* | Unpublished | Pangasidae |
| 71. | JX997836 | *Pangasius bocourti* | Unpublished | Pangasidae |

The MP reconstructions were conducted in PAUP v. 4.0b10 (Swofford, 2002) via heuristic searches with random addition (RA) of sequences and tree-bisection-reconnection (TBR); clade support was evaluated using non-parametric bootstrapping with RA and TBR.

For ML and BI, the best-fit models of sequence evolution were estimated using the Akaike information criterion (AIC) in ModelTest v. 3.7 (Posada and Crandall, 1998). All analyses were run unpartitioned.

The ML analyses were performed in program RAxML v.7.04 (Stamatakis, 2006). RAxML searches were run in CIPRES portal v.1.13 under default configurations. ML nodal support was evaluated in RAxML using the rapid bootstrapping algorithm with automatic estimation of runs. For RAxML searches, several runs from random starting seeds were performed to check convergence of likelihood scores. Model parameters were estimated simultaneously (i.e., unfixed). Remaining settings were left at their default values.

The BI analyses were performed in MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003) via Markov chain Monte Carlo (MCMC) iterations. The MCMC analyses were conducted in triplicate using four chains and sampling trees every 100 generations. Conservatively, 25% of the first trees sampled in each MCMC run were discarded as burn-in. Marginal probabilities of summary parameters, consensus phylograms, and posterior probabilities of nodes were estimated from the postburn-in samples of the three independent runs.

To confirm that post-burn-in trees were sampled from the actual MCMC posterior distribution, marginal parameters (MrBayes log file) were analyzed using the Effective Sample Size (ESS) statistic in program Tracer (Drummond *et al*., 2007); ESS greater than 200 suggests that MCMC searches were run long enough to accurately represent the posterior distribution (Drummond *et al*., 2007).

## Results and Discussion

*Phylogenetic analysis at genus and family level*

Among the three reconstruction methods conducted (MP, ML-RAxML, BI) on the mitochondrial dataset, BI analysis resulted in least resolved tree. The consensus tree of three methods is shown in Figures 1 - 3. Bootstrap values and posterior probability values are labeled and indicated with gradient colour scheme for each of the respective node in format BI/ML/MP.

*COI* as a DNA marker has been able to discriminate between species phylogenetic relationships appropriately. While choosing as outgroup for a family level relationship, a phylogeneticist normally prefers distantly related family taxa. Hence in this case, Ageneiosidae (for Ariidae), Scheilbeidae (for Bagridae) and Pangasidae (for Plotosidae) have been chosen as outgroups as observed from the previous literature (Sullivan *et al*., 2006).

The length of the *COI* sequence of all the individuals of 16 catfish species ranged from 589 to 640 bp long. The sequence which consists more than 600 bp long suggested that PCR amplified products belongs to mitochondrial *COI* gene not to nuclear mitochondrial DNA (Numt). Nuclear insertions of mitochondrial origin are found throughout the human genome and are believed to have arisen from DNA transfer between the mitochondrial and the nuclear genomes during evolution. Numts suggested to be less than 600 bp long, show high sequence identity with the mitochondrial genome DNA sequence and can be large, the largest Numts identified was >14.6 kb, with 15 out of 296 Numts being greater than 5.8 kb (Mourier *et al*., 2001). The remaining sequences that consist less than 600 bp might be the effect of lower PCR amplification even though they belong to *COI* gene confirmed by sequence identity and usage of vertebrate mitochondrial translation pattern. All sequences were conceptually translated into protein sequences. It is important to assess the pseudogene status of amplified products. Moreover, there was no evidence for the presence of Numts in Actinopterygii (Bensasson *et al*., 2001).

*COI* nucleotide sequence data provide an opportunity to examine the rate of evolution and amount of phylogenetic information across various taxonomic levels. *COI* study was primarily interested in comparative series of taxa of different rank. The *COI* gene show a trend of increasing mean base composition distances with increasing rank of the groups compared, from within species to families. Heterogeneity of *COI* gene evolution rate, also significant in present data is widely known from earlier studies (Machordom and Macpherson, 2004). A one-way ANOVA (model with random effects for groups of the same size) showed that mean distances in four groups analysed were significantly different for three catfish families, $F= 19.79$, d.f.= 3; 116, $P< 0.0001$ (Plotosidae); $F= 44.21$, d.f.= 3; 986, $P< 0.0001$ (Ariidae) and $F= 24.83$, d.f.= 3; 1322, $P< 0.0001$ (Bagridae). However, to remember, this comparison is not quite correct for all of the DNA sequences compared, because it includes heterogeneous groups of catfish families of different size and sequence length.

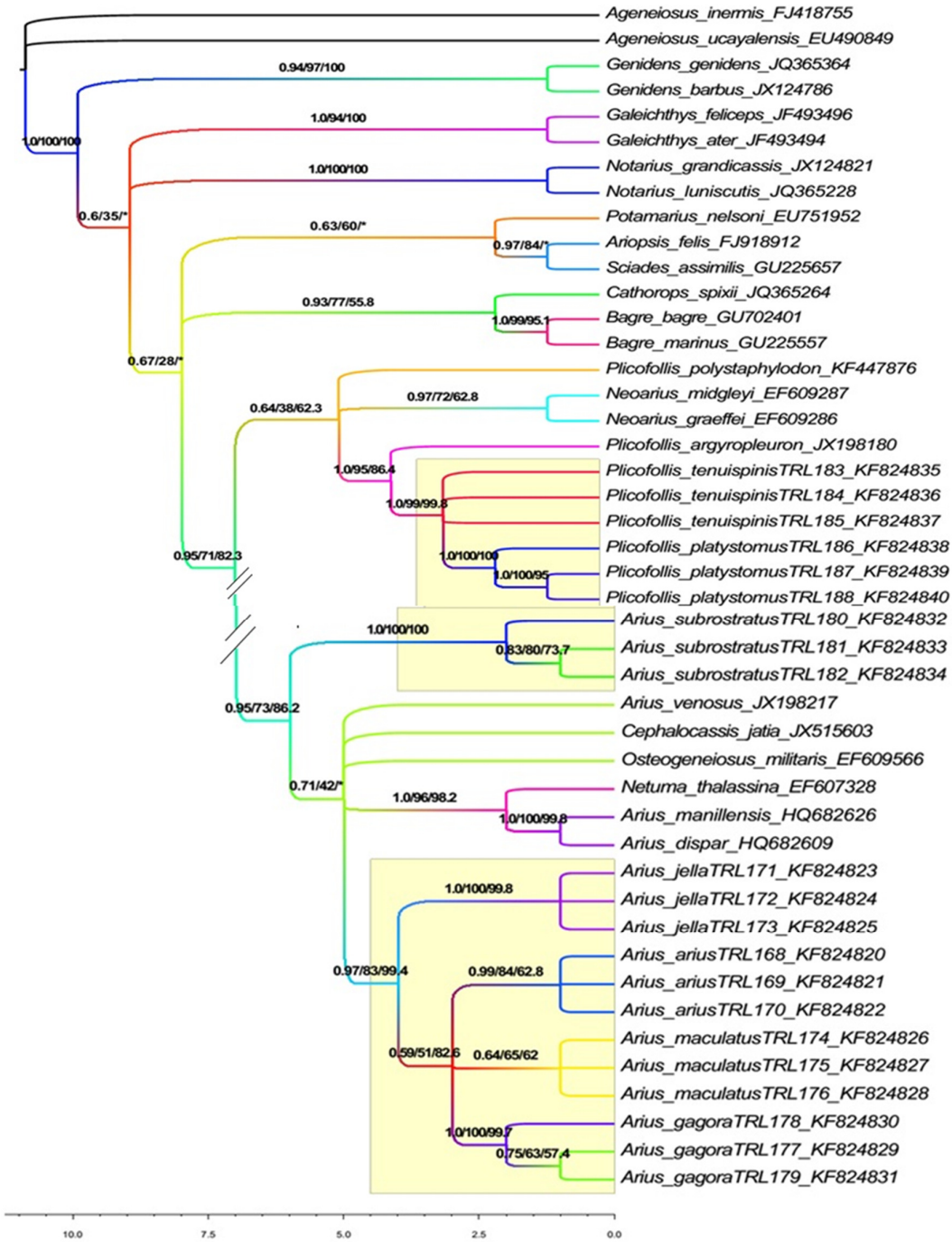


**Figure 1.** Phylogeny of Ariidae. Colored node labels indicate posterior probability and bootstrap values (BI/ML/MP) that are congruent with MP and ML
Asterisks (*) designate insufficient clade support.

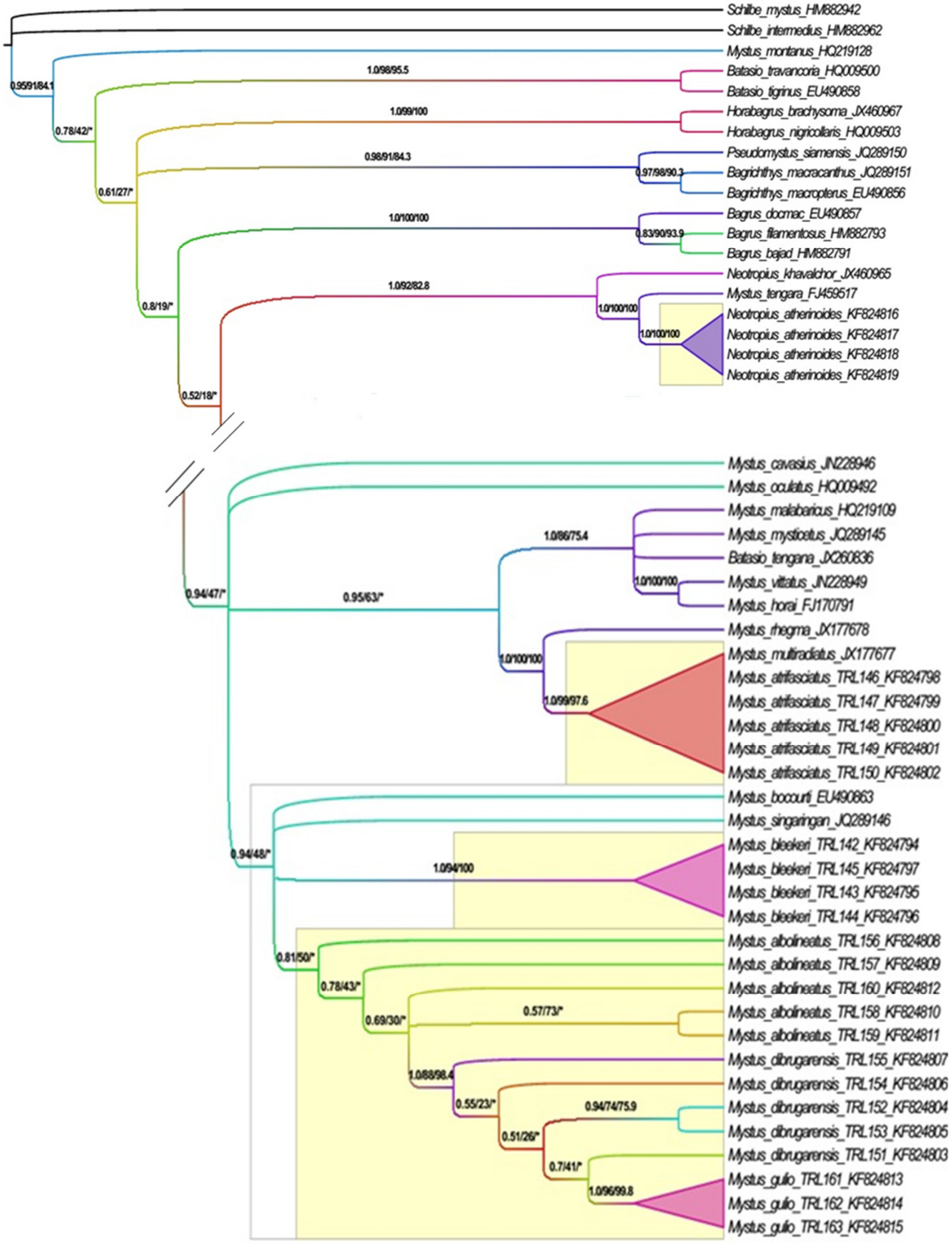


**Figure 2.** Phylogeny of Plotosidae based on *COI* gene sequence. Colored node labels indicate posterior probability and bootstrap values (BI/ML/MP) that are congruent with MP and ML
Asterisks (*) designate insufficient clade support.

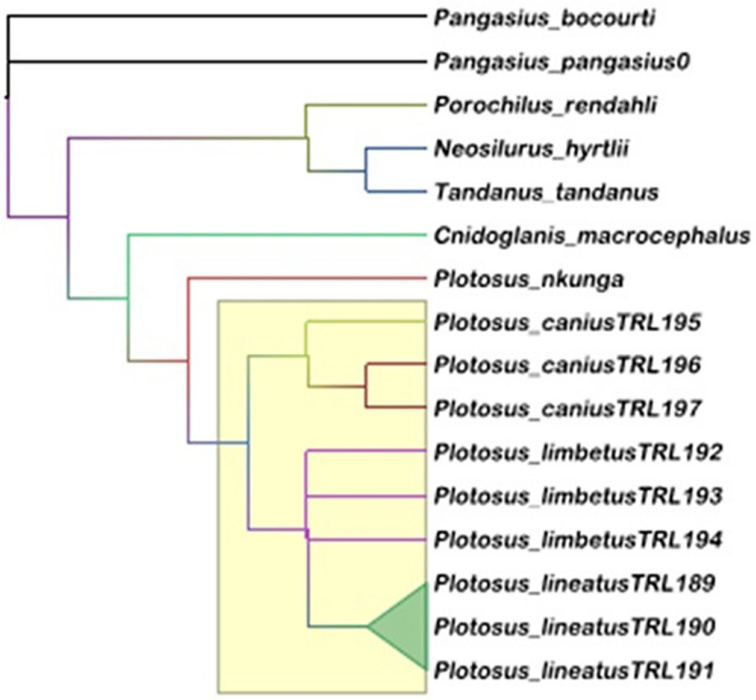


**Figure 3.** Phylogeny of Bagridae based on *COI* gene sequence. Colored node labels indicate posterior probability and bootstrap values (BI/ML/MP) that are congruent with MP and ML
Asterisks (*) designate insufficient clade support.

A common assumption in comparative sequence analysis is that the sequences have evolved with the same pattern of nucleotide substitution (homogeneity of the evolutionary process). Violation of this assumption is known to affect the accuracy of phylogenetic inference and tests of evolutionary hypotheses. In this research, disparity index, ID has been proposed to measure the observed difference in evolutionary patterns for a pair of sequences. Based on this index, all the *COI* sequences of three catfish families were evaluated to test the homogeneity of the observed patterns. This test does not require a priori knowledge of the pattern of substitutions, extent of rate heterogeneity among sites, or the evolutionary relationship among sequences. Homogeneity assumption was tested by calculating the probability of observing a composition distance greater than that expected under the null hypothesis of homogeneity (*i.e.,* ID more than zero).

Relative saturation rates of different sites and types of substitution were assessed to estimate substitution patterns. In order to use coding sequences to reconstruct phylogenies accurately, it is important to adjust for the relative rates at which different codon sites and different types of substitution (e.g., transitions vs. transversions (Ti/Tv) saturate. This study compared the relative rates at which transitions and transversions saturate across taxonomic categories by counting substitutions in all pairwise comparisons between sequences from the following categories: within species, between species within a genus, between genera within a family and away from outgroup.

Phylogenetic reconstruction from DNA or amino acid sequences relies heavily on suitable distance measures. *COI* gene region analyzed here have been frequently used to address phylogenetic relationships, but only at the generic level or below had its relationships successfully resolved. At the family level, this gene rarely gives satisfactory resolution (Brown *et al*., 1994; Miura *et al*., 1998) and often proves to be unreliable (Dowton and Austin, 1997; Mardulyn and Whitfield, 1999), while at still higher levels; cytochrome oxidase sequences are not suitable for resolving relationships (Howland and Hewitt, 1995; Frati *et al*., 1997). *COI* data set in this study in comparison to these higher-level phylogenetic relationships is a first step towards this objective.

*Ariidae*

The monophyly of the Ariidae has not been seriously questioned and is strongly supported on both molecular and morphological grounds (Sullivan *et al*., 2006; Acero and Betancur, 2007). The group is divided into two subfamilies, the monogeneric Galeichthyinae (four species) which predominantly belong to marine species and the Ariinae (remaining taxa) (Acero and Betancur, 2007), so that apart from Ariinae, *Galiechthys feliceps* and *Galiechthys ater* are the only utilized species in this study. These contribute 106 singletons, 89 two-fold and 123 four-fold degenerative sites from the whole Ariidae family. Ti/Tv bias was two-fold higher in 1st codon position (9.00) than the 3rd codon position (4.65). In *Arius* genus, this becomes inverse with Ti/Tv bias, which was six-fold higher in 3rd codon position (7.47) than the 1st codon position (1.46). But in whole Ariidae family, Ti/Tv bias was nearly equal in 1st (5.07) and 3rd (5.55) codon positions. From the phylogram of whole Ariidae, a galeichthyinae branch shows it had recently diverged with longer branch length in *G. feliceps* (Figure 1). Their base composition distance fall within thin limits of 5% level significance of Standard Error and does not affect the different levels of taxa. Different patterns of substitution have been attributed for *G. feliceps* against all the Ariid species (ID between 0.25 and 0.68) except *A. jella* and *A. subrostratus* at5% level of significance (Figure 1). This might be the reason for the inverse relationship of Ti/Tv bias at 1st and 3rd codon positions as explained in aforementioned statement. Even then, *G. feliceps* exhibit different substitution pattern (ID = 0.1522) against the same species *G. ater,* which was highly significant contributing the longer branch length in phylogram.

The whole *Arius* genus comprises eight species, out of which five are utilized in this study and the remaining was by from Santos and Quilang (2011). There are totally 531 conserved and 108 variant sites in *Arius* genus that consist 88 pi sites, out of which 81 are at 3rd position of a codon. Especially, in 40th nucleotide base, adenine was present in *Arius arius,* whereas in all other species it was cytosine. Initially Ti/Tv bias for the Ariidae family was found to be ten. But with addition of outgroups, according to the HKY+G+I model tested through BIC, Ti/Tv bias found to be 7.19. The Ti/Tv bias, A+T content, Pi sites at all the 1+2+3 coding sites observed for all the species of *Arius* genus maintain balancing selection at the nucleotide level without any drastic change. The base composition distance for all the species of *Arius* genus fell within limits of same species distance (<0.007) with the exception for *Arius gagora* (0.022). Against family comparison, *Arius* genus exhibited highest base composition distances when compared with *Bagre bagre*. Homogeneity pattern of substitution has been maintained for all the species of *Arius* genus accepting the null hypothesis with the exception between *A. jella* and *A. gagora* but with the lowest disparity index estimate (ID = 0.0819). In the phylogenetic tree (Figure 1), *Arius* genus displayed as monophyletic group with higher bootstrap and posterior probability values for all the species except *A. subrostratus* and the other species that are not sampled in this study. The phylogram shows that the branch lengths of the entire *Arius* genus showed negligible branch lengths depicting least substitution rates all throughout the Ariid family based on *COI* dataset. Although the basal arioid clades are well defined, much controversy has arisen regarding the phylogeny and classification of Ariid taxa, particularly within diverse Ariinae (Betancur, 2009).

The whole *Plicofollis* genus comprises four species, out of which two are utilized in this study and the remaining two (*P. tenuispinis* and *P. argyropleuron*) from unpublished data as reported in NCBI database. The A+T content at 3rd codon position were nearly 70% and the Ti/Tv bias was higher at 1st coding site (3.74) than 3rd coding site (2.22). Other than this, no peculiar phenomenon could be observed on singletons, Pi sites and 2-fold and 4-fold degeneracy sites at all the 1st+2nd+3rdcoding sites observed for all the species that may maintain balancing selection at the nucleotide level without any drastic change. The base composition distance for all the species of *Plicofollis* genus fell within limits of same species distance (<0.002), but with the exception for *Plicofollis platystomus* (0.015). Homogeneity pattern of substitution has been maintained for all the species of *Plicofollis* accepting the null hypothesis with the exception between *Plicofollis argyropleuron* with *Plicofollis tenuispinis* (ID = 0.084 and 0.096) and *Plicofollis platystomus* (ID = 0.176). In MP, ML and BIbased cladogram, *Plicofollis* genus displayed as a monophyletic group with higher bootstrap and posterior probability values for all the species except two species of *Neoarius,* which intervened separating *P. polystaphylodon*

(Figure 1). While this interruption was reported with less bootstrap value from ML construction. The phylogram shows that the branch lengths of the entire *Plicofollis* genus showed <0.1 depicting moderate substitution rates all throughout the Ariid family.

*Plotosidae*

Plotosid fishes have been called blunt-tail catfishes or eel tail catfishes which contributes very few reports for molecular marker studies. Previous results suggest that the family was originally marine, with an invasion into freshwaters (Page and Hughes, 2010). As currently defined the family has nine genera and about 40 species. Five of those genera are restricted to freshwaters in Australia and New Guinea, while the remaining four genera are marine. Of the marine genera, all but one is restricted to seas around Australia and New Guinea. The last marine genus, *Plotosus* is extremely widespread, occurring from South Africa and Japan to Australia.

The whole *Plotosus* genus comprises four species, out of which three are utilized in this study and the remaining one species (*P. nkunga*) from unpublished data as reported in NCBI database. The A+T content at 1st and 3rd codon position were around 60% and the Ti/Tv bias was higher at 1st coding site (5.39) than 3rd coding site (1.46). Other than this, abnormal extremities were not observed on singletons, Pi sites, two-fold and four-fold degeneracy sites at all the 1st+2nd+3rdcoding sites observed for all the species of *Plotosus* genus that may maintain balancing selection at the nucleotide level without any drastic change. There are 160 Pi sites in whole Plotosidae family and reduces to 98 sites when restricted to *Plotosus* genus. In that case, only one Pi site was observed in 2nd codon position at the 479 nucleotides site in this genus. The base composition distance for all the species of *Plotosus* fell within limits of same species distance (<0.006). Higher values of base composition distance (~0.25) across the entire genus within Plotosidae family beyond the values of outgroup comparisons (~0.08) were clearly noted. Two outliers and two far outlier values were observed for *Porochilus rendahli* and *Cnidoglanis macrocephalus* against the *Plotosus* genus. Homogeneity pattern of substitution has been maintained for all the species of *Plotosus* genus accepting the null hypothesis with the exception between *Porochilus rendahli* with all the species of *Plotosus tenuispinis* (ID = 0.35 and 0.43). In MP, ML and BI based cladogram, *Plotosus* genus displayed as monophyletic group with higher bootstrap and posterior probability values for all the species (Figure 2). The phylogram shows that the branch length of the entire *Plotosus* genus showed branch lengths <0.1 depicting moderate substitution rates throughout this genus as per *COI* dataset.

*Bagridae*

Bagrid catfishes constitute a very important group among siluriform having immense commercial importance from inland fisheries and aquaculture farming in south-east countries. Bagridae family, comprising of 27 genera (six in Indian region) is widely distributed in Asia and Africa (Talwar and Jhingran, 1991). With the available *COI* sequences, six genera have been utilized in this study with the importance given to *Mystus* genus. More pi characters (243) were observed on Bagridae for maximum parsimony tree construction. Minimized number of optimal trees (8) were retained as observed in Bagridae and highest (24 trees) for Ariidae. There are other examples of bursts of fish evolution, documented by molecular markers (Rutaisire *et al*., 2004; Duftner *et al*., 2005).

The whole *Mystus* genus comprises 17 species, out of which five are utilized in this study and the remaining species from unpublished data as reported in NCBI database. The 22 out of48 singletons were observed in 1st codon position which is higher than the 3rd codon position. Most singletons observed in *Mystus* genus were observed from the other species that were not utilized in this study. There are totally 189 variant sites in *Mystus* genus that consist 199 pi sites in total out of which 116 are at 3rd position of a codon. Initially Ti/Tv bias for the Ariidae family was found to be ten. But with addition of outgroups, according to the HKY+G model tested through BIC, Ti/Tv bias found to be 1.963. The Ti/Tv bias was higher at 1st and 2nd coding site than 3rd coding site seeking out alternative explanations for neutral theory of natural selection. A+T content and singletons reported higher values in 1st codon position. Pi sites (199) was highest when compared to the other catfish families at all the 1st+2nd+3rdcoding sites observed for all the species of *Mystus* that enable

enrichment for the phylogenetic tree construction. The base composition distance for all the species of *Mystus* changes dynamically (0 - 0.07) for highest value of same species distance (0.0720) between individuals of *M. albolineatus* and the null values arose from individuals of *Mystus bleekeri* and *Mystus atrifasciatus*. Between species base composition distance values lies between 0.058 - 1.523 where the highest value was noted between *M. atrifasciatus* and *M. monatanus* and the lowest between *M. gulio* and *M. cavasius.* Against family comparison, those values lie between 0.0136 - 1.0759, *M. atrifasciatus* exhibited highest base composition distances when compared with *Bagrus docmac* and the lowest between *M. albolineatus* and *Bagricthys macracanthus*. Homogeneity pattern of substitution has not been maintained for all the species of *Mystus* rejecting the null hypothesis with the highest disparity index estimate (ID = 1.5798) between *M. montanus*. In MP, ML and BI based cladogram, *Mystus* genus displayed as a monophyletic group with higher bootstrap and posterior probability values for all the species except *M. montanus* forming a distant and distinct clade whereas *M. tengara* collides into monophyletic clade when *Neotropius* genus was removed. The phylogram shows that the branch length of the entire *Mystus* genus showed longer branch lengths compared to other catfish families utilized in this study accumulating intense substitution rates all throughout the Bagrid family. This may alter balancing selection under neutral theory of natural selection. Longest branch length was denoted from *M. montanus*, which could affect the total phylogram through long branch attraction.

## Conclusions

This phylogenetic analysis is a first step toward this objective, although it still needs more comparative ecological data for a comprehensive analysis of the evolution of feeding habits. By this study we could establish a phylogenetic hypothesis for all the 36 catfish families and examine the monophyly status of the subfamilies and genera.

## Authors' Contributions

Both authors read and approved the final manuscript.

## Acknowledgements

The authors gratefully acknowledge the Dean and Director, Faculty of Marine Sciences and The authorities of Annamalai University for the facilities and encouragements.

## Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

## References

Acero PA, Betancur RR (2007). Monophyly, affinities, and subfamilial clades of sea catfishes (Siluriformes: Ariidae). Ichthyological Exploration of Freshwaters 18:133-143.

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. Nucleic Acids Research 25:3389-3402. *https://doi.org/10.1093/nar/25.17.3389*

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution 16:314-321. *https://doi.org/10.1016/s0169-5347(01)02151-6*

Betancur RR (2009). Systematics and evolutionary history of sea catfishes (Siluriformes: Ariidae). In: PhD Thesis, Auburn University, Auburn pp 200.

Booth AJ, Muwanika VB, Masembe C, Nyakaana S, Rutaisire J (2004). Evolution of *Labeo victorianus* predates the Pleistocene desiccation of Lake Victoria: Evidence from mitochondrial DNA sequence variation. South African Journal of Science 100:607-608.

Brown JM, Pellmyr O, Thompson JN, Harrison RG (1994). Mitochondrial DNA phylogeny of the Prodoxidae (Lepidoptera: Incurvarioidea) indicates rapid ecological diversification of Yucca moths. Annals of Entomological Society of America 87:795-802. *https://doi.org/10.1093/aesa/87.6.795*

Dowton M, Austin AD (1997). Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. Journal of Molecular Evolution 44:398-405. *https://doi.org/10.1007/PL00006159*

Drummond AJ, Ho SYW, Rawlence N, Rambaut A (2007). A rough guide to BEAST 1.4. University of Edinburgh, Edinburgh pp 1-41.

Duftner N, Koblmuller S, Sturmbauer C (2005). Evolutionary relationships of the Limnochromini, a tribe of benthic deep-water cichlid fish endemic to Lake Tanganyika, East Africa. Journal of Molecular Evolution 60:277-289. *https://doi.org/10.1007/s00239-004-0017-8*

Frati F, Simon C, Sullivan J, Swofford DL (1997). Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. Journal of Molecular Evolution 44:145-158. *https://doi.org/10.1007/PL00006131*

Hebert P, Cywinska A, Ball S, Dewaard J (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:313-321. *https://doi.org/10.1098/rspb.2002.2218*

Howland DE, Hewitt GM (1995). Phylogeny of the Coleoptera based on mitochondrial cytochrome oxidase I sequence data. Insect Molecular Biology 4:203-215. *https://doi.org/10.1111/j.1365-2583.1995.tb00026.x*

Jayaram KC (1984) Ariidae. In: Fischer W, Bianchi G (Eds). Western Indian Ocean Fishing area 51.1. FAO, Rome.

Knowlton N, Weigt LA (1998). New dates and new rates for the divergence across the Isthmus of Panama. Proceedings of the Royal Society of London. Series B: Biological Sciences 265:2257-2263. *https://doi.org/10.1098/rspb.1998.0568*

Kumar S, Gadagkar SR (2001). Disparity index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158:1321-1327.

Machordom A, Macpherson E (2004). Rapid radiation and cryptic speciation in squat lobsters of the genus *Munida* (Crustacea, Decapoda) and related genera in the South West Pacific: molecular and morphological evidence. Molecular Phylogenetics and Evolution 33:259-279. *https://doi.org/10.1016/j.ympev.2004.06.001*

Mardulyn P, Whitfield JB (1999). Phylogenetic signal in the COI, 16S and 28S genes for inferring relationships among genera of *Microgastrinae* (Hymeoptera; Braconidae): Evidence of a high diversification rate in this group of parasitoids. Molecular Phylogenetics and Evolution 12:282-294. *https://doi.org/10.1006/mpev.1999.0618*

Miura T, Maekawa K, Kitade O, Abe T, Matsumoto T (1998). Phylogenetic relationships among subfamilies in higher termites (Isoptera: Termitidae) based on mitochondrial COII gene sequences. Annals of Entomological Society of America 91:515-523.

Moritz C, Cicero C (2004). DNA barcoding: Promise and pitfalls. PLoS Biology 2:e354. *https://doi.org/10.1371/journal.pbio.0020354*

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001). The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Molecular Biology and Evolution 18:1833-1837. *https://doi.org/10.1093/oxfordjournals.molbev.a003971*

Page TJ, Hughes JM (2010). Comparing the performance of multiple mitochondrial genes in the analysis of Australian freshwater fishes. Journal of Fish Biology 77:2093-2122. *https://doi.org/10.1111/j.1095-8649.2010.02821.x*

Posada D, Crandall KA (1998). Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818. *https://doi.org/10.1093/bioinformatics/14.9.817*

Remigio E, Hebert P (2003). Testing the utility of partial CO1 sequences for phylogenetic estimates of gastropod relationships. Molecular Phylogenetics and Evolution 29:641-647. *https://doi.org/10.1016/s1055-7903(03)00140-4*

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574. *https://doi.org/10.1093/bioinformatics/btg180*

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: A laboratory manual. 2nd Edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, pp 1626.

Santos BS, Quilang JP (2011). DNA barcoding of *Arius* species (Siluriformes: Ariidae) in Laguna de Bay, Philippines using the cytochrome C oxidase subunit I gene. Philippine Agricultural Scientist 94:205-210.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690. *https://doi.org/10.1093/bioinformatics/btl446*

Sullivan JP, Lundberg JG, Hardman M (2006). A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. Molecular Phylogenetics and Evolution 41:636-662. *https://doi.org/10.1016/j.ympev.2006.05.044*

Swofford DL (2002). PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0 Beta. Sinauer Associates Inc., Sunderland, MA, USA.

Talwar PK, Jhingran AG (1991). Inland fisheries of India and adjacent countries. Oxford and IBH Publishing Co., New Delhi, India, pp 514.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731-2739. *https://doi.org/10.1093/molbev/msr121*

Ward RD, Zemlak TS, Innes BH, Last PA, Hebert PDN (2005). DNA barcoding Australia's fish species. Philosophical Transactions Royal Society B: Biological Sciences 360(1462):1847-1857. *https://doi.org/10.1098/rstb.2005.1716*