Research Article

Molecular Divergence and Phylogenetic Positioning in Siluriformes Species Using mtDNA COI Gene

Mohd Imran*

Department of Zoology, Human Genetics and Toxicology Laboratory, Section of Genetics, Faculty of Life Sciences, Aligarh Muslim University, Aligarh, Uttar Pradesh, INDIA.

Submission Date: 14-08-2023; Revision Date: 15-09-2023; Accepted Date: 22-09-2023.

ABSTRACT

Aim: Molecular divergence between species or a group of species is efficient tool to trace the relation between them and assist in their better phylogenetic positioning with good bootstrap supports. The study has emphasised on a chosen group of siluriformes species available in the nearest water resource which includes species from family Clariidae, Heteropneustidae, Bagridae, and Siluridae. **Materials and Methods:** Standard barcoding protocol was followed where the intraspecific and interspecific pair wise sequence divergences were calculated based on Kimura-2-Parameter model, followed by phylogenetic tree clustering. **Results:** A considerable gap was found in the divergence values between species which is sufficient to differentiate and delimit them. For intraspecific divergence, the average was found less than 1% whilst the interspecific divergence varied from 16 to 24%. When the comparison was made on an average basis, the divergence between species (20.40%) was found approx. 25 times higher than intraspecific divergence (0.811%). When the divergence was calculated separately at each codon position, the maximum contribution was found from 3rd codon position (78.4%) to the combined divergence which was followed by 1st codon position (6.87%) and the least contribution from 2nd codon position (1.82%). The clustering analysis with Neighbor-Joining and Maximum likelihood methods delimit the species with parallel phylogenetic clustering supporting the divergence trends. **Conclusion:** The work provides a thorough picture of the relationship between divergence values and how this gap in divergence values plays a critical role in differentiating species while at the same time assisting in deciphering their taxonomic and phylogenetic positioning.

Keywords: Molecular divergence, Phylogenetic clustering, Siluriformes.

Correspondence:
Mohd Imran
Department of Zoology, Human Genetics and Toxicology Laboratory, Section of Genetics, Faculty of Life Sciences, Aligarh Muslim University, Aligarh-202002, Uttar Pradesh, INDIA.

Email: mimran@myamu.ac.in

INTRODUCTION

DNA barcoding aims to ease and speed up the process of species identification and to develop a comprehensive database with barcoding sequences from as many faunas as possible.[1,2] This is to make the identification of an organism accessible to the various ventures dealing with animals in different forms at different life stages, whether it is an organisation involved in their conservation or a food industry dealing with processed animal products, the two of the most important fields with an utmost requirement of identification. This approach could even be applicable in cases of the complete absence of morphology-based identification keys, like identifying illegally traded animal parts.[3] To generate these barcodes, the 5' 650 bp segment of the 1545 bp long mitochondrial Cytochrome c Oxidase I gene (COI) has been considered an efficient molecular marker. It is claimed that it provides deeper phylogenetic insight in comparison to other mitochondrial markers[1] and also has the advantage of sets of primers which has provided amplification across the enormous datasets from different classes of vertebrates- Pisces,[4,5]

SCAN QR CODE TO VIEW ONLINE

www.ajbls.com

DOI: 10.5530/ajbls.2023.12.63

Amphibia,[6] Reptilia,[7] Aves[8,9] and mammalia.[10] Hence, COI found quite relevant to conduct a huge experiment encompassing hundreds of species. This method of barcoding is based on the fact that no two species have identical COI sequences and there is a permissible divergence range which keeps the individuals of a species together despite having minor divergence variations among them, with the factual consideration that the intraspecific divergence is always lower than the interspecific divergence and there is, generally, a huge gap between the two divergence ranges which is called as the barcoding gap and thus separates one species from other. [11,12]

The catfishes are among the commercially important group of freshwater species in the Indian subcontinent with a contribution of 143 species from the Indian freshwaters.[13] They are widely consumed, especially in the North Indian states which are penetrated with a vast tributary network of the Ganges riverine system. Six catfish species from different families as *Clarias batrachus* and *C. gariepinus* from the family Clariidae; *Heteropneustes fossilis* from the family Heteropneustidae; *Sperata seenghala* and *Rita rita* from the family Bagridae, and *Wallago attu* from the family Siluridae were included in this study to determine the intraspecific and interspecific divergence ranges. While choosing these species, emphasis has been given to cover the divergence at different taxonomic levels. The rationale of choosing these species is that no such study has been done before on these species to decipher the trend of divergence range in catfishes. Besides this, they are widely distributed, commonly available and commercially important as well. So, in quest to find the divergence trend from the barcoding perspective alongside covering the higher taxonomic level viz. family, such a combination of species was chosen for this study. The sampling was done from distant locations to consider population effect on the divergence value, as the distant populations sometimes show unexpected large divergences overlapping with interspecific divergence range.[1,5]

Although standard morphometric keys are available for most of the catfish species,[14] but these keys are life stages dependent viz. the individuals of *Clarias batrachus* and *C. gariepinus* are quite difficult to distinguish at their young stages. So, to overcome such drawback of life stage dependency and for aforesaid application of barcoding, the COI based identification system is taken into consideration for barcoding these species, since one of the key advantages which comes with barcoding is its life stage independency. The process of barcoding any species is based on a simple methodology of Neighbour-Joining (NJ) clustering of the species using Kimura 2-Parameter (K2P), which differentiates them into separate cluster without giving any emphasis on their phylogenetic relationships. But as some previous studies claimed of getting the phylogenetic signals in NJ tree,[2,10,15,16] the clustering in the present study is further analysed for the same using Maximum Likelihood (ML) analysis. Hence, the objective of the present study emphasised on deciphering the percent sequence divergence at different taxonomic level in catfish species, and to look for the phylogenetic information content in Neighbor-Joining (NJ) clustering.

MATERIALS AND METHODS

Collection sites

The sample collection sites were two distant locations 560 km away on northern riverine system of Ganges i.e. Aligarh (27.88°N 78.08°E) and Varanasi (25°19'08"N 83°00'46"E) with the help of local fishermen using hook and line, fishing net and from the nearby local fish markets (Figure 1). Whenever possible specimens were brought to laboratory for identification or otherwise identified at the sampling spot using standard identification features.[14]

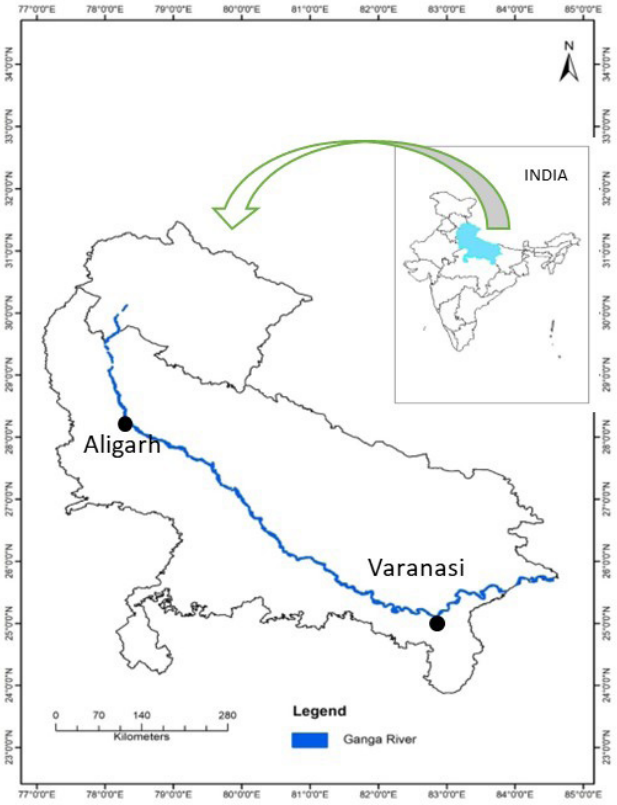

Figure 1: Specimen sampling was done from the districts of Aligarh and Varanasi from the river Ganga in the northern part of India.

DNA extraction and sequencing

The specimens were anaesthetized using appropriate dose of Tricaine Methanesulphonate (Sigma-Aldrich). Approximately 0.3-0.5 mL blood were taken via heart/caudal vein in Ethylenediaminetetraacetic acid (EDTA) coated vacutainer vial. The blood samples were placed in -20ºC for a short duration until before they subjected to DNA extraction. The high Salt Method[17] with minor changes was used for total DNA extraction including both nuclear and organellar DNA. The quantification of DNA samples was done using Nanodrop (Implan, Germany). All the samples with absorbance ratio ($A^{260/280}$) near to 1.8 were selected for PCR to avoid any kind of impurity and discrepancy in the polymerase chain reactions. Partial 'Cytochrome c Oxidase I (COI) gene' was amplified using the universal primers: Forward- fishF1, fishF2 and Reverse- fishR1, fishR2.[2] It has amplified a segment of approx. 650-700 bps from all the catfish species in PCR reaction mixture of 25 µL with 50-100 ng DNA template, 10X PCR buffer, 2 mM $MgCl_2$, 0.5 µL dNTPs mix, 0.25 µL each forward and reverse primer and 1 unit TaqDNA polymerase. PCR amplifications were done in Peqlab thermal cycler (Model: PEQSTAR 2X), with initial denaturation (95ºC/5 min) trailed by 35 thermal cycles of denaturation at 95ºC, 60 sec; annealing at 55-62ºC, 45 sec; extension at 72ºC, 60 sec., with a single final extension at 72ºC, 5 min. Multiple replicates were amplified for the individuals of each species to check the reproducibility of results, and to avoid any experimental error. This was followed by purification of all the PCR amplicons which then finally subjected for the further process of DNA sequencing with Sanger method (sequencer: ABI 3730XL) with BigDye Terminator v3.1 Sequencing Kit (applied biosystems). The sequences were deposited to GenBank sequence database with accession no. MH047225-MH047238.

DNA Sequence analysis

All the DNA sequences were observed individually for low resolution 5' and 3' terminals and their subsequent trimming using software BioEdit (version 7.2.5).[18] The multiple COI sequence alignment was done in software MEGA 7[19] using default parameters. The aligned set of 644 bp long COI was then examined for the basic characteristic of a mitochondrial protein coding gene like A+T richness, nucleotide frequency, anti-guanine bias to further confirm the sequences' natural identity and then proceeded for the intraspecific and interspecific pairwise divergence using K2P. The software DnaSP (Ver. 6.12.01)[20] was used to trace the site polymorphism including monomorphic, polymorphic, singleton variable and parsimony informative sites. The percent contribution of each nucleotide from all three nucleotide codon positions was evaluated. Standard barcoding protocol was followed where the intraspecific and interspecific pair wise sequence divergence based on K2P model was calculated first, followed by the NJ tree based on the same model using Mega 7. The NJ method is used because of its speed and strong performance with low sequence divergence[11] and its dependence on the coalescence of conspecific populations and the monophyly of species.[21] The NJ tree was constructed with K2P. Since, NJ gives outcome similar to that of likelihood or parsimony- based phylogenetic methods,[22] ML analysis was also done to check if there is any phylogenetic signal in COI sequence data and if it is there, then does NJ tree provide any phylogenetic insight or just distinguishes the species, irrespectively, into separate clade. For ML approach, the best fit substitution model was assessed in "Find Best DNA/Protein Models (ML)" option in MEGA (Version7.0). According to Akaike information criterion (AIC), K2P+G model was selected for the given data. The K2P method was used since it takes transitional and transversional bias into consideration. Since it is a protein coding gene where the three codon positions evolve at a different rate,[23] the gamma distribution was used to characterize the nucleotide substitution rate variation among sites.[24]

The percent divergence values were also calculated separately for all three nucleotide codon positions (1st to 3rd) to trace the contribution of individual codon position to the overall divergence.

RESULTS

The COI gene was found to be adenine and thymine (A+T) rich with average percent nucleotide contribution as A: 27.1%; T: 29.5%; G: 16.8%; and C: 26.6% with lowest contribution from guanine (G) because of "anti-guanine bias" at 3rd and 2nd codon positions (Table 1). The transition (ts) substitutions were higher than transversion (tv) substitutions, and their ratio (ts/tv) was found to be 2.30, the highest transitional substitution were found at position 3rd followed by position 1st and 2nd with ts/tv ratio on three codon positions were found in the order of 3rd (3.58) >1st (7.63) >2nd (1.24) using the method of Kimura.[25] Besides this, 436 nucleotide sites were found to be monomorphic; parsimony informative: 205; singleton variable: 03; 2-fold: 90 and 4 fold degenerate: 105. Sequences were found devoid of any insertion/deletion or nonsense mutation.

Sequence divergence

Table 1: Percent composition of Thymine(T), Cytosine(C), Adenine(A), Guanine(G) at three nucleotide codons, calculated by MEGA(Version7.0).

Fish Species	1st Codon				2nd Codon				3rd Codon			
	T	C	A	G	T	C	A	G	T	C	A	G
W. attu	18	26.5	25.1	30.2	41	29.4	16.4	13.1	27	28.4	35.3	9.3
H. fossilis	19	26.5	25.6	29.3	41	29.0	16.4	13.6	32	21.4	40.0	6.5
R. rita	19	25.1	26.5	29.3	41	29.0	15.9	14.0	30	24.7	39.5	5.6
C. batrachus	20	25.1	26.0	28.8	42	29.0	15.9	13.6	25	26.0	41.4	7.9
C. gariepinus	20	25.6	25.6	29.3	41	29.4	16.4	13.1	27	24.2	40.5	8.8
S. seenghala	20	25.1	25.6	29.8	41	29.0	15.9	14.0	28	25.6	40.5	6.0
Average	19	25.7	25.7	29.5	41	29.1	16.1	13.6	28	25.0	39.5	7.4

Table 2: Mean percent intraspecific divergence.

Species	Divergence (%)	Family
C. batrachus	0.16	Clariidae
C. gariepinus	2.22	Clariidae
S. seenghala	0.62	Bagridae
R. rita	0.62	Bagridae
H. fossilis	0.31	Heteropneustidae
W. attu	0.94	Siluridae

A trend of increasing K2P divergence was found from conspecific individuals to species belonging to different family (Table 2). The divergence in the conspecific individuals found < 1% ranging from 0.16% to 0.9% except in *C. gariepinus* where it reaches to 2.2%. The divergence between different families (Tables 3 and 4) reaches maximum to 24.28% between Clariidae (*C. gariepinus*) and Bagridae (*S. seenghala*). A K2P divergence between confamilial genera found equal to 21.14% between *R. rita* and *S. seenghala* from the family Bagridae but the divergence between the two congeneric species of genus *Clarias* found equal to 15.82% between *C. gariepinus* and *C. batrachus*. The divergence value between the non-congeneric

Table 3: Percent evolutionary divergence calculated over sequence pairs between species groups (Model used: K2P+gamma distribution).

Fish species	*Wa*	*Hf*	*Rr*	*Cb*	*Cg*	*Ss*
W. attu (Wa)	-	-	-	-	-	-
H. fossilis (Hf)	21.63	-	-	-	-	-
R. rita (Rr)	19.03	19.49	-	-	-	-
C. batrachus (Cb)	20.75	18.85	21.12	-	-	-
C. gariepinus (Cg)	22.60	16.12	19.42	15.82	-	-
S. seenghala (Ss)	22.00	21.23	21.14	22.60	24.28	-

Table 4: Mean percent divergence within and between families (range given in parenthesis, where applicable). Within family divergence values are marked with asterisk (*) and paragraph (¶) marks to indicate reason behind the difference between corresponding values. Asterisk (*) is indicating the 'within family divergence' from the same genus (*Clarias*) of the family Clariidae. Paragraph (¶) mark is indicating the 'within family divergence' from the two genera (*Sperata* and *Rita*) of same family Bagridae. No marks are placed on between families divergence values.

Divergence -Within and between families				
family	**Clariidae**	**Bagridae**	**Heteropneustidae**	**Siluridae**
Clariidae	15.82*	-	-	-
Bagridae	21.85 (19.42-24.28)	21.14¶	-	-
Heteropneustidae	17.48 (16.12-18.85)	20.36 (19.49-21.23)	-	-
Siluridae	21.67 (20.75-22.60)	20.51 (19.03-22.00)	21.63	-

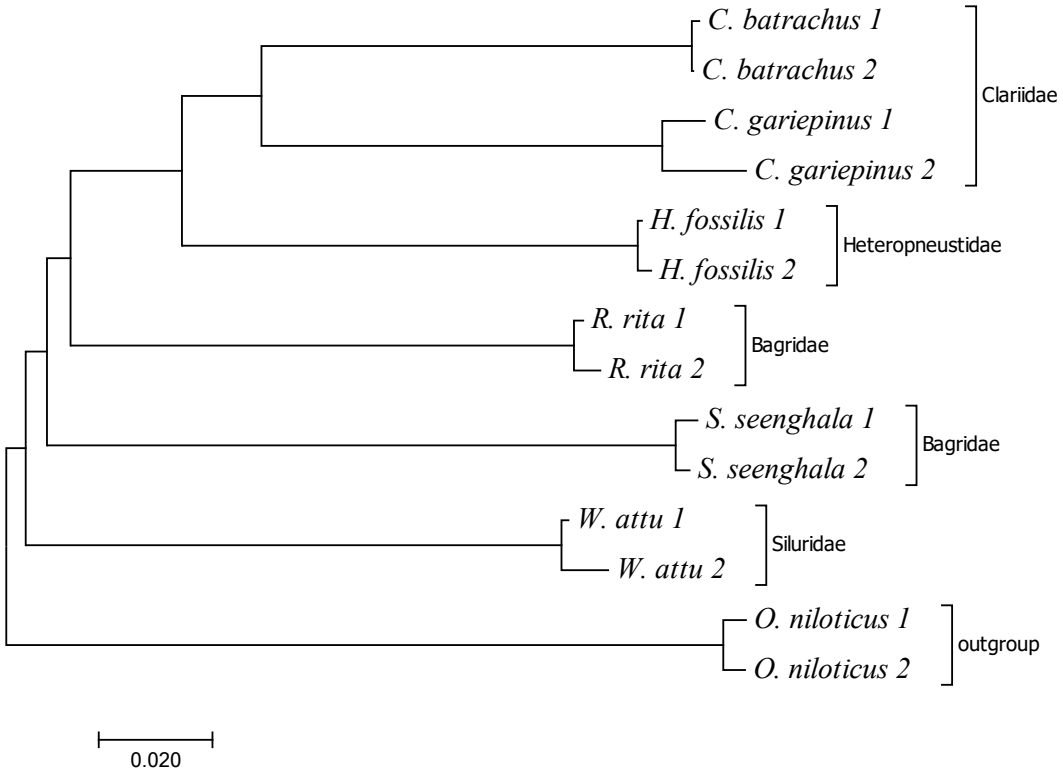

Figure 2: Evolutionary relationships of species using the Neighbour-joining method.[27] Gamma distribution was used to characterize the rate variation among sites.

H. fossilis and *C. gariepinus* (16.12%) found closer to the congeneric divergence found within genus *Clarias* and the possible reason is the relationship between Clariidae and Heteropneustidae as they are revealed as member of the same clade in the mitogenomic study by Kappas *et al.*[26]

Finally, the NJ-K2P clustering analysis has successfully separated all six species into separate clade (Figure 2). When the comparison was made on the average basis, the K2P divergence between species (20.40%) was found approx. 25 times higher to intraspecific divergence (0.811%). When the divergence was calculated separately at each codon position (Table 5), the maximum contribution was found from 3rd codon position (78.4%) to the combined divergence followed by first codon position (6.87%) and the least contribution form 2nd codon position (1.82%).

Table 5: Percent evolutionary divergence between species groups at three codon positions using K2P+gamma distribution. Table 5(a) representing divergence at 1st codon position; Table 5(b) representing divergence at 2nd codon position; Table 5(c) representing divergence at 3rd codon position.
Table 5(a): Divergence at 1st codon position.

Codon Position 1st						
Fish species	***Wa***	***Hf***	***Rr***	***Cb***	***Cg***	***Ss***
W. attu (Wa)						
H. fossilis (Hf)	0.0539					
R. rita (Rr)	0.0467	0.0673				
C. batrachus (Cb)	0.0592	0.0541	0.0697			
C. gariepinus (Cg)	0.0536	0.0487	0.0748	0.0515		
S. seenghala (Ss)	0.0794	0.0781	0.1083	0.0893	0.0971	
Average divergence=0.0687=6.87%						

Continued...

Table 5: Cont'd.
Table 5(b): Divergence at 2nd codon position.

Codon Position 2nd						
	Wa	***Hf***	***Rr***	***Cb***	***Cg***	***Ss***
W. attu (Wa)						
H. fossilis (Hf)	0.0214[a]					
R. rita (Rr)	0.0214[a]	0.0166[c]				
C. batrachus (Cb)	0.0239[b]	0.0094[d]	0.0166[c]			
C. gariepinus (Cg)	0.0239[b]	0.0142	0.0094[d]	0.0118		
S. seenghala (Ss)	0.0166[c]	0.0154	0.0166[c]	0.0227	0.0238[b]	
Average divergence=0.0182=1.82%						

(a, b, c and d are indicating identical divergence values).

Table 5(c): Divergence at 3rd codon position.

Codon Position 3rd						
	Wa	***Hf***	***Rr***	***Cb***	***Cg***	***Ss***
W. attu (Wa)						
H. fossilis (Hf)	0.9164					
R. rita (Rr)	0.7730	0.6928				
C. batrachus (Cb)	0.8210	0.7257	0.8194			
C. gariepinus (Cg)	1.0556	0.5553	0.7135	0.5504		
S. seenghala (Ss)	0.8699	0.7915	0.7070	0.8387	0.9446	
Average divergence=0.7849=78.49%.						

Phylogenetic analysis

Some phylogenetic signal seems appeared in the NJ tree with K2P. To confirm this, phylogenetic analysis was done with the character state approach based ML method (Figure 3). Depending upon the data, K2P was used as a statistical test since it takes into consideration

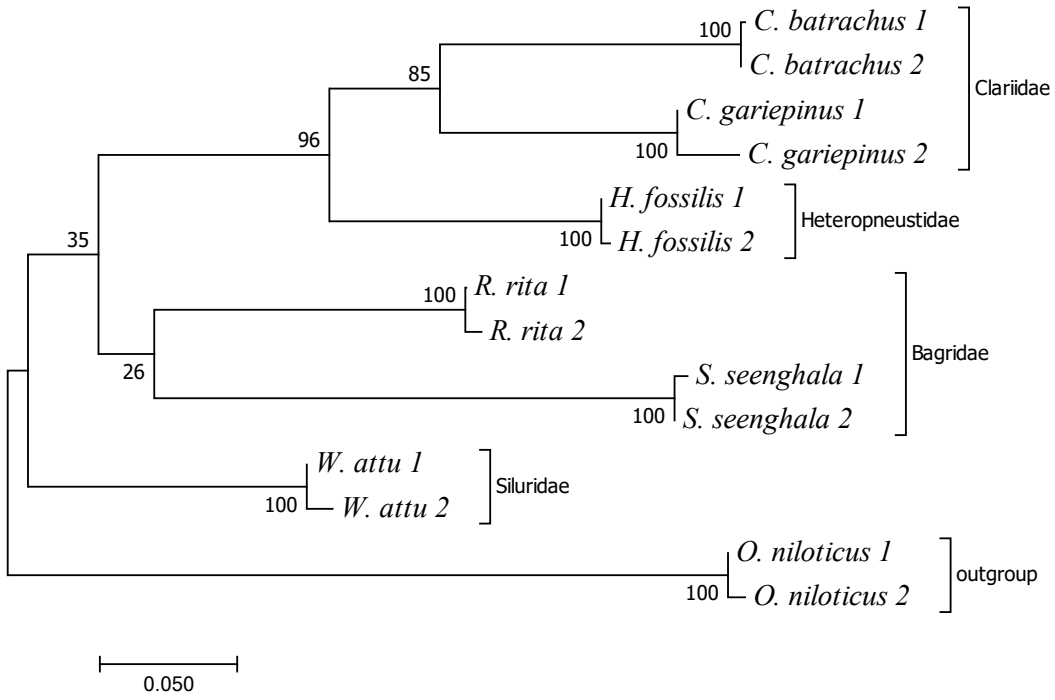

Figure 3: Maximum likelihood evolutionary tree based on the Kimura 2-parameter model.[25] Gamma distribution was used to characterize the rate variation among sites.

the ts/tv bias at different codon positions which varies considerably. *Oreochromis niloticus* was taken as an outgroup and bootstrapping was performed with 100 replicates. The ML analysis clearly confirmed the phylogenetic signals which appeared in the NJ tree with bootstrap support of 81 between the two species of *Clarias* and a very high bootstrap support of 95 between the family Clariidae and Heteropneustidae.The clade consisting of *R. rita* and *S. seenghala* although showing low bootstrap support but the clustering is true since the species belong to the same family Bagridae. This shows that phylogenetic signals appeared in the NJ tree were true.

DISCUSSION

This study has characterized a distinct trend of molecular divergence at different taxonomic level from conspecific individuals to species belonging to different families (Tables 2 and 3). We found a very low permissible limit of <1% of intraspecific divergence similar to that reported in numerous other fish species viz. 207 Chondrichthyes,[28] 391 ornamental fish species,[29] 09 Bagrid species.[30] Not just in fish species, a diverse number of species from other vertebrate group have shown the similar trends of <1% in intraspecific divergence viz, in amphibians-several species of frogs;[31,32] several species of reptiles including species of snakes, Geckos, Chameleon, Lizard;[33,34] in North American bird species, an average intraspecific divergence of 0.27% was reported,[11] similarly, 0.24% intraspecific distance reported in Scandinavian bird species.[8] Even among the invertebrate a huge number of insects species i.e. 28,619 species have shown a value in the range of <1% intraspecific divergence.[35] In the present study, this intraspecific value ranges from 0.16% to 0.9% except in *C. gariepinus* where it reaches to a value of 2.2%. The wide geographical distance between the sampling sites poses an increasing effect on the intraspecific divergence range.[36] Since the intraspecific divergence is usually not more than 2%,[37] a higher value indicates geographically isolated populations[1] or it could be the case of taxonomic uncertainty involving cryptic sibling species as discovered in Collembolans-*Sminthurides malmgreni* and *Folsomia quadrioculata* with a higher intraspecific divergence of 5% and 13%, respectively, against the well-noted <1% divergence.[38] So, the probable reason behind this exceptionally high divergence may be the different populations haven't been intermixed or hybridized for a long time or having a very low population mixing i.e. geographical effect. The similar kind of high divergence has also been previously reported in Indo-pacific fish species,[4] in amphibians like salamander species, toads and frogs.[39,40] The extensive analysis of COI sequences from the individuals of both populations will contribute some interesting results to the concept of DNA barcoding where the divergence value from different populations sometimes raises questions about the authenticity of its general consideration.[41]

As we moved towards the higher taxonomic level, the divergence value increases as expected, in terms of correlation with the morphological taxonomy where differences in morphological identification keys increases from lower to higher taxonomic level.[14] A similar increasing trend in COI divergence along with hierarchical ranks from species onwards is also reported in the reptilian order Testudines.[42] The divergence between different families (Table 4) reach maximum to 24.28% between Clariidae (*C. gariepinus*) and Bagridae (*S. seenghala*). A K2P divergence between confamilial genera found equal to 21.14% between *R. rita* and *S. seenghala* from the family Bagridae which is close to the 20.8% reported in bagrid catfish species and 22.6% reported in other catfish species[43] but the divergence between the two congeneric species of genus *Clarias* from the same family Clariidae found equal to 15.82% between *C. gariepinus* and *C. batrachus*. The divergence value between the non-congeneric *H. fossilis* and *C. gariepinus* (16.12%) was found to be closer to the congeneric divergence found within the genus *Clarias* and the possible reason could be the relationship between Clariidae and Heteropneustidae as they are revealed as member of the same clade in the mitogenomic study by Kappas *et al.*[26] The reptilian species have also shown the similar divergence trend between species within the same families where it ranges over 13.4% to 29.8% in different families of Madagascar reptilian lineages.[7] When the comparison was made on the average basis, the K2P divergence between species (20.40%) was found to be approx. 25 times higher to intraspecific divergence (0.811%) which is found similar to the values reported in marine fish species[29] and birds[11] where divergence between species was found 26 times and 24 times higher than intraspecific divergence, respectively. Since the data represents a combination of different families, we found an increasing divergence as we move up from intraspecific divergence to between family divergence as: within species<between species (within congeneric species<within familial genera<between species from different families). In this comparison of divergence, we found that the sequence divergence between different species which belongs to different families varies over

a range of 16-24% as discussed earlier. This divergence could be an important indication of the family level divergence range. But setting a divergence limit for the resolution of higher taxonomic level such as family will only be possible when an extensive sampling is done encompassing the maximum possible families from an order and comparing their divergence values. While doing this, the sampling uniformity should be kept in mind; the pattern of sampling should be similar from all the families, means, the count of congeneric, non-congeneric species should be similar in order to have a proper divergence range because the random sampling is not sufficient/ ideal for having a better understanding of divergence range.[44] Likewise for the order level, in a similar manner, the best way would be to take different possible families, keeping the specimens from confamilial genera similar in proportion in terms of congeneric and non-congeneric specimens, since the uniformity in sampling is the key to have a better understanding of the permissible divergence range at different taxonomic level. The same way, permissible divergence range could possibly be deduced for the generic level with exploiting the maximum possible confamilial genera and comparing its divergence with the divergence range of the similar data/other confamilial genera. Such basic comparison could possibly give the inference of "reliability of COI" at higher taxonomic levels. In our analysis, although with deviations, we have also found an increasing order of divergence values, where within species divergence is just below 1%, at the same time the congeneric divergence and divergence between families ranged as 15.8% and 16.12-24.28%, respectively. Such a huge difference between the divergence values at different taxonomic level clearly shows that there exists a divergence gap which was considered as the barcoding gap differentiating the intraspecific divergence from interspecific divergence. Such mitochondrial COI based barcoding gap are also well reported in other vertebrate species other than fish species viz. avians, anuran and caudata species.[45] The concept of barcoding gap is not just limited to vertebrates only, numerous invertebrate species especially arthropod species complies with the barcoding gap in their difference between intraspecific and interspecific divergences, for example, numerous genera of true bugs,[46] dragonflies and damselflies.[47] But one thing is confirmed, extensive sampling encompassing different geographical location is not just necessary for genetic variability studies or determining the population dynamic of a species, but also is essential for performing barcoding studies.[48]

The phylogenetic signals in NJ tree were found true when confirmed with the ML analysis where the clustering was found supported with high bootstrap values and as per the morphologically established relationships. The clustering of *Clarias* spp. in one clade, family clariidae and heteropneustidae as sister clade, and family bagridae as a separate clade with its both species. The low bootstrap support to the family bagridae consisting of *R. rita* and *S. seenghala* probably because of limited number of taxa as this family consist of a total 221 species and limiting the no. of species could affect the bootstrap support but this clustering of confamilial *R. rita* and *S. seenghala* is true since they belong to the same family. This shows that all the phylogenetic signals appeared in the NJ tree were true irrespective of the low bootstrap support in one clade, hence suggesting that NJ method not just resolves the taxa but at the same time also places them according to their phylogeny.

The divergence contribution from each codon position showed that the codon position 3rd (Table 5c) contributing the maximum with an average contribution of 78.49% to the combined divergence followed by 1st position (6.87%) (Table 5a), and then the 2nd position (Table 5b) which is contributing the least (1.82%). The reason behind the least contribution by 2nd codon position is its functional relevance which makes it most conserved among the three positions while the position 3rd is a wobble position which is liable to change. Besides, contributing the least, the 2nd codon position has also shown a peculiar feature of identical divergence between different species which are represented by the superscripts a, b, c and d (Table 5b). These identical divergence values between different species have again emphasized on the least contribution of 2nd codon position in the overall divergence, at the same time, when positions 1st and 3rd showing high variability, thus depicting its functional importance. The maximum contribution of codon position 3rd in the divergence values again depicts its significance that most of the information for identification is coming from this wobble position, and the same is reported by Ward and Holmes[49] where they did an extensive examination of COI nucleotide sequences and corresponding amino acid variabilities in fishes.

The COI is claimed as barcoding marker not just for the fish species or other vertebrate groups but it has been explored as barcoding marker in the invertebrate species as well. Since COI provided barcoding application from diverse set of species from invertebrate to vertebrates, it is claimed as the universal barcoding marker, although other mitochondrial markers have also shown this capability of species identification viz. mitochondrial 16s rRNA gene[50] but unlike rRNA gene, insertion and

deletion are not prevalent in protein coding COI gene and that also add to its advantage and ease the comparative analysis between diverse species. The advancement of next generation sequencing has further made the COI barcoding more applicable where the concept of metabarcoding involves sequencing of short fragment of less than 150bp called as mini-barcode, instead of 650 bp in sanger sequencing, which is sufficient to characterize the wide no. of species from environmental DNA (eDNA) samples,[51] thus overcoming the limitation associated with the sanger sequencing based COI barcoding and making the barcoding method more economical and widely applicable.

The findings in the study which explore the role of intraspecific and interspecific divergence distribution in differentiating species gives a clear cut picture of how the molecular divergence is relevant in barcoding species with the applicability of the concept of barcoding gap at different hierarchical level from species onward, thus further strengthening the concept of barcoding gap and its utilization in delimiting the catfish species. The findings of increase in molecular divergence with the rise in taxonomic rank from species onward shows the reliability of the COI marker not just to barcode a species alone but also classify interspecific relationship as well which are obvious with the NJ clustering. The findings could be of huge relevance in further exploring the molecular divergence over a broad range of catfish species to get a further in-depth picturization of their divergence dynamics in light of establishing the standard divergence limit for the characterization of catfishes, in general.

CONCLUSION

The sole purpose of barcoding is to ease and speed up the process of species identification. This identification is not just important from the taxonomic point of view but also is essential in various sectors which utilize any organism in any form and need a symbol for its identity whether it is trade, food, medicine, recreation and above everything its conservation due to the imposing risk of extinction. The study has successfully revealed the importance of deciphering the intraspecific and interspecific divergence in species differentiation among the catfish species, and at the same time have also brought to the notice the appearance of phylogenetic relationships in NJ tree with support from ML clustering. Additionally, emphasising on the sampling strategy, a proper and organized sampling should be taken into account while performing such studies in order to get accurate results which comes with the simplicity of the method. In this manner a more comprehensive picture could be obtained about the utility of COI divergence at different taxonomic levels.

ACKNOWLEDGEMENT

The Author sincerely acknowledges Prof. Mohammad Afzal, Human Genetics and Toxicology Laboratory, Section of Genetics, Department of Zoology, Aligarh Muslim University (AMU) for his noble insights into this work. The author also acknowledges the Department of Zoology, AMU, for providing necessary facility. The Council of Scientific and Industrial Research (CSIR), New Delhi is also gratefully acknowledged for providing funding support.

FUNDING SOURCE

The research was funded by Council of Scientific and Industrial Research (CSIR), New Delhi, India.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

mtDNA: Mitochondrial DNA; **COI:** Cytochrome c oxidase subunit I; **K2P:** Kimura-2-Parameter; **ML:** Maximum Likelihood; **NJ:** Neighbor Joining; **MEGA7:** Molecular Evolutionary Genetic Analysis (version 7.0).

SUMMARY

The molecular identification has been seen as an alternative, fast and more economical system in comparison to the morphological identification, where mitochondrial COI gene is hailed as the universal barcoding marker based on a simple approach of divergence calculation accompanied with clustering analysis. Using this approach, the study has shown the divergence pattern between species and within species where a trend of increasing K2P divergence is found from conspecific individuals to species belong to different families, which is also reflected in the NJ and ML phylogenetic clustering analysis.

REFERENCES

1. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270(1512):313-21. doi: 10.1098/rspb.2002.2218, PMID 12614582.
2. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci. 2005;360(1462):1847-57. doi: 10.1098/rstb.2005.1716, PMID 16214743.

3. Guglich EA, Wilson PJ, White BN. Forensic application of repetitive DNA markers to the species identification of animal tissues. J Forensic Sci. 1994;39(2):353-61. doi: 10.1520/JFS13606J, PMID 8195749.
4. Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN. DNA barcoding reveals overlooked marine fishes. Mol Ecol Resour. 2009;9;Suppl s1:237-42. doi: 10.1111/j.1755-0998.2009.02649.x, PMID 21564983.
5. Quyen VD, Oanh TT, Phuong TT, Thuoc TL, Binh DT. Molecular phylogeny of catfishes (*Teleostei*: *Siluriformes*) inferred from mitochondrial markers-Implication for lower Mekong River basin. Eur J Adv Res Biol Life Sci. 2018;6:3.
6. che J, Chen HM, Yang JX, Jin JQ, Jiang K, Yuan ZY, *et al*. Universal COI primers for DNA barcoding amphibians. Mol Ecol Resour. 2012;12(2):247-58. doi: 10.1111/j.1755-0998.2011.03090.x, PMID 22145866.
7. Nagy ZT, Sonet G, Glaw F, Vences M. First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLOS ONE. 2012;7(3):e34506. doi: 10.1371/journal.pone.0034506, PMID 22479636.
8. Johnsen A, Rindal E, Ericson PGP, Zuccon D, Kerr KCR, Stoeckle MY, *et al*. DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. J Ornithol. 2010;151(3):565-78. doi: 10.1007/s10336-009-0490-3.
9. Patel S, Waugh J, Millar CD, Lambert DM. Conserved primers for DNA barcoding historical and modern samples from New Zealand and Antarctic birds. Mol Ecol Resour. 2010;10(3):431-8. doi: 10.1111/j.1755-0998.2009.02793.x, PMID 21565042.
10. Schäffer S, Zachos FE, Koblmüller S. Opening the treasure chest: a DNA-barcoding primer set for most higher taxa of Central European birds and mammals from museum collections. PLOS ONE. 2017;12(3):e0174449. doi: 10.1371/journal.pone.0174449, PMID 28358863.
11. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLOS Biol. 2004;2(10):e312. doi: 10.1371/journal.pbio.0020312, PMID 15455034.
12. Luo A, Zhang A, Ho SY, Xu W, Zhang Y, Shi W, *et al*. Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. BMC Genomics. 2011;12:84. doi: 10.1186/1471-2164-12-84, PMID 21276253.
13. Sarkar UK, Pathak AK, Sinha RK, Sivakumar K, Pandian AK, Pandey A, *et al*. Freshwater fish biodiversity in the River Ganga (India): changing pattern, threats and conservation perspectives. Rev Fish Biol Fisheries. 2012;22(1):251-72. doi: 10.1007/s11160-011-9218-6.
14. Talwar PK, Jhingran AG. Inland fishes of India and adjacent countries. Vol. 2. Rotterdam: Balkema Publishers; 1992.
15. Moftah M, Abdel Aziz SHA, Elramah S, Favereaux A. Classification of sharks in the Egyptian Mediterranean waters using morphological and DNA barcoding approaches. PLOS ONE. 2011;6(11):e27001. doi: 10.1371/journal.pone.0027001, PMID 22087242.
16. Tran HTT, Tran TN, Tran HNA, Nguyen HT. DNA barcoding and phylogenetic relationships of nine catfish species from Mekong Basin, Vietnam. J Mol Biomark Diagn. 2017;8:6.
17. Montgomery GW, Sise JA. Extraction of DNA from sheep white blood cells. N Z J Agric Res. 1990;33(3):437-41. doi: 10.1080/00288233.1990.10428440.
18. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95-8.
19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870-4. doi: 10.1093/molbev/msw054, PMID 27004904.
20. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, *et al*. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol Biol Evol. 2017;34(12):3299-302. doi: 10.1093/molbev/msx248, PMID 29029172.
21. Wiemers M, Fiedler K. Does the DNA barcoding gap exist?-a case study in blue butterflies (Lepidoptera: *Lycaenidae*). Front Zool. 2007;4:8. doi: 10.1186/1742-9994-4-8, PMID 17343734.
22. Roch S. Toward extracting all phylogenetic information from matrices of evolutionary distances. Science. 2010;327(5971):1376-9. doi: 10.1126/science.1182300, PMID 20223986.
23. Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.
24. Uzzell T, Corbin KW. Fitting discrete probability distributions to evolutionary events. Science. 1971;172(3988):1089-96. doi: 10.1126/science.172.3988.1089, PMID 5574514.
25. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111-20. doi: 10.1007/BF01731581, PMID 7463489.
26. Kappas I, Vittas S, Pantzartzi CN, Drosopoulou E, Scouras ZG. A time-calibrated mitogenome phylogeny of catfish (Teleostei: Siluriformes). PLOS ONE. 2016;11(12):e0166988. doi: 10.1371/journal.pone.0166988, PMID 27907107.
27. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454, PMID 3447015.
28. Ward RD, Holmes BH, White WT, Last PR. DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. Mar Freshw Res. 2008;59(1):57-71. doi: 10.1071/MF07148.
29. Steinke D, Zemlak TS, Hebert PDN. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLOS ONE. 2009;4(7):e6300. doi: 10.1371/journal.pone.0006300, PMID 19621079.
30. Zou R, Liang C, Dai M, Wang X, Zhang X, Song Z. DNA barcoding and phylogenetic analysis of bagrid catfish in China based on mitochondrial COI gene. Mitochondrial DNA A DNA Mapp Seq Anal. 2020;31(2):73-80. doi: 10.1080/24701394.2020.1735379, PMID 32126856.
31. Biju SD, Garg S, Mahony S, Wijayathilaka N, Senevirathne G, Meegaskumbura M. DNA barcoding, phylogeny and systematics of Golden-backed frogs (Hylarana, Ranidae) of the Western Ghats-Sri Lanka biodiversity hotspot, with the description of seven new species. Contrib Zool. 2014;83(4):269-S4. doi: 10.1163/18759866-08304004.
32. Dang NX, Sun FH, Lv YY, Zhao BH, Wang JC, Murphy RW, *et al*. DNA barcoding and the identification of tree frogs (Amphibia: Anura: Rhacophoridae). Mitochondrial DNA A DNA Mapp Seq Anal. 2016;27(4):2574-84. doi: 10.3109/19401736.2015.1041113, PMID 26004249.
33. Liu Q, Zhu F, Zhong G, Wang Y, Fang M, Xiao R, *et al*. COI-based barcoding of Chinese vipers (Reptilia: Squamata: Viperidae). Amphibia Reptilia. 2015;36(4):361-72. doi: 10.1163/15685381-00003012.
34. Vasconcelos R, Montero-Mendieta S, Simó-Riudalbas M, Sindaco R, Santos X, fasola M, *et al*. Unexpectedly high levels of cryptic diversity uncovered by a complete DNA barcoding of reptiles of the Socotra archipelago. PLoS One. 2016;11(3):e0149985. doi: 10.1371/journal.pone.0149985, PMID 26930572.
35. Zhang H, Bu W. Exploring large-scale patterns of genetic variation in the COI gene among Insecta: implications for DNA barcoding and threshold-based species delimitation studies. Insects. 2022;13(5):425. doi: 10.3390/insects13050425, PMID 35621761.
36. Lukhtanov VA, Sourakov A, Zakharov EV, Hebert PDN. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Mol Ecol Resour. 2009;9(5):1302-10. doi: 10.1111/j.1755-0998.2009.02577.x, PMID 21564901.
37. Avise JC. Phylogeography: the history and formation of species. Cambridge: Harvard University Press; 2000.
38. Hogg ID, Hebert PDN. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zool. 2004;82(5):749-54. doi: 10.1139/z04-041.
39. Che J, Chen HM, Yang JX, Jin JQ, Jiang K, Yuan ZY, *et al*. Universal COI primers for DNA barcoding amphibians. Mol Ecol Resour. 2012;12(2):247-58. doi: 10.1111/j.1755-0998.2011.03090.x, PMID 22145866.
40. Rivera-Correa M, Correa-Medina H, Venegas-Valencia K, Daza JM. Genetic diversity, acoustic signal and geographic distribution of a colourful rain frog of the genus Pristimantis (Anura: Strabomantidae). Herpetol Notes. 2022;15:215-17.
41. Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. PLOS Biol. 2005;3(12):e422. doi: 10.1371/journal.pbio.0030422, PMID 16336051.
42. Mohd Salleh MH, Esa Y, Mohamed R. Global terrapin character-based DNA barcodes: assessment of the mitochondrial COI gene and conservation status revealed a putative cryptic species. Animals (Basel). 2023;13(11):1720. doi: 10.3390/ani13111720, PMID 37889683.
43. Wong LL, Peatman E, Lu J, Kucuktas H, He S, Zhou C, *et al*., Na-na. DNA barcoding of catfish: species authentication and phylogenetic assessment.

PLOS ONE. 2011;6(3):e17812. doi: 10.1371/journal.pone.0017812, PMID 21423623.

44. Johnson NK, Cicero C. New mitochondrial DNA data affirm the importance of *pliestocene* speciation in North American Birds. Evolution. 2004;58(5):1122-30. doi: 10.1111/j.0014-3820.2004.tb00445.x, PMID 15212392.
45. Velo-Antón G, Chambers EA, Poyarkov NA Jr, Canestrelli D, Bisconti R, Naumov B, *et al*. COI barcoding provides reliable species identification and pinpoints cryptic diversity in western Palearctic amphibians. Amphibia Reptilia [published online ahead of print:1-15]:1-15. doi: 10.1163/15685381-bja10148.
46. Bianchi FM, Gonçalves LT. Borrowing the Pentatomomorpha tome from the DNA barcode library: scanning the overall performance of cox1 as a tool. J Zool Syst Evol Res. 2021;59(5):992-1012. doi: 10.1111/jzs.12476.
47. Koroiva R, Kvist S. Estimating the barcoding gap in a global dataset of cox1 sequences for Odonata: close, but no cigar. Mitochondrial DNA A DNA Mapp Seq Anal. 2018;29(5):765-71. doi: 10.1080/24701394.2017.1357709, PMID 28752773.
48. Dasmahapatra KK, Mallet J. Taxonomy: DNA barcodes: recent successes and future prospects. Heredity. 2006;97(4):254-55. doi: 10.1038/sj.hdy.6800858, PMID 16788705.
49. Ward RD, Holmes BH. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (Cox1) in fishes. Mol Ecol Notes. 2007;7(6):899-907. doi: 10.1111/j.1471-8286.2007.01886.x.
50. Vences M, Thomas M, van der Meijden A, Chiari Y, Vieites DR. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Front Zool. 2005;2(1):5. doi: 10.1186/1742-9994-2-5, PMID 15771783.
51. Subrata SA, Siregar SRT, André A, Michaux JR. Identifying prey of the Javan mongoose (Urva javanica) in Java from fecal samples using next-generation sequencing. Mamm Biol. 2021;101(1):63-70. doi: 10.1007/s42991-020-00086-y.

Cite this article: Imran M. Molecular Divergence and Phylogenetic Positioning in Siluriformes Species Using mtDNA COI Gene. Asian J Biol Life Sci. 2023;12(3):473-83.