Genetics of Aquatic Organisms 9(2), GA878
https://doi.org/10.4194/GA878

PROOF
RESEARCH PAPER

GENAQUA
GENETICS OF AQUATIC ORGANISMS

Molecular Evolution of Cytochrome c Oxidase I Marker Encoded Protein Across the Fish Families from The Bay of Bengal, Odisha Coast, India

Bijayalaxmi Sahu[1,3], Raghunath Satpathy[2], Tapan Kumar Barik[3], Amiya Kumar Patel[1,*]

[1]Sambalpur University, Department of Biotechnology and Bioinformatics, Jyoti Vihar-768019, Burla, Odisha, India.
[2]Gangadhar Meher University, School of Biotechnology, Amruta Bihar-768004, Sambalpur, Odisha, India.
[3]Berhampur University, P.G. Department of Zoology, Bhanja Bihar-760 007, Berhampur, Odisha, India.

How to Cite

Sahu, B., Satpathy, R., Barik, T.K., Patel, A.K. (2025). Molecular Evolution of Cytochrome c Oxidase I Marker Encoded Protein Across the Fish Families from The Bay of Bengal, Odisha Coast, India. *Genetics of Aquatic Organisms, 9(2), GA878*. https://doi.org/10.4194/GA878

Article History
Received 05 February 2025
Accepted 20 May 2025
First Online 02 June 2025

Corresponding Author
E-mail: amiya.gene@gmail.com

Keywords
Bay of bengal
Fish
DNA barcoding
Protein structure
Structural modeling

Abstract

The BOLD database contains approximately 20.5 million Cytochrome c Oxidase I (COI) barcode sequences belonging to 3.6 lakhs species, among which 2.6 lakhs are animal species, which have made the COI gene fragments of the animal kingdom's most thoroughly sequenced gene area. Additionally, COI barcodes present a distinctive resource for investigating functional limitations on DNA evolution, as the barcode region is situated at the center of cellular energy production, where COI is one of the Cytochrome c Oxidase (COX) protein's building parts. This study has compared the amino acid variations in COI barcode areas across different fish families, as they are the major contributors to the current vertebrate population. The COI sequences produced by the DNA barcoding program in the Bay of Bengal, Gopalpur-on-sea, India, were explored in the present study. Among the families under study, 28 amino acid variations were detected inside the COI barcode sequences. The Sciaenidae family displayed the greatest number of variants among all the families, with variations in the 19 amino acid positions. Overall, these results show DNA barcodes of how the fish species may be used for purposes far beyond taxonomy and identification, followed by a wide range of research.

Introduction

The existence of short standard DNA regions in all target lineages has enough sequence variations to distinguish between species through DNA barcoding, which is a rapid and accurate method for identifying the species (Hebert et al., 2003). Besides, the intra-species divergence is less compared to the inter-species divergence, which predicts the discriminating potential of the DNA barcodes (Kochzius et al., 2010). Thus, the DNA barcoding approach is the most reliable molecular method for identifying alien species (Cross et al., 2011), which makes it easier to identify the individuals whose identities are uncertain. Cytochrome c oxidase subunit I (COI) of mitochondria offers the standard approach for categorizing most forms of life out of various potential barcode sections that have already been developed (Hebert et al., 2003 a,b; Radulovici et al., 2010). Effective species identification is possible using COI sequences, which frequently exhibit higher inter-specific and lower intra-specific divergence (Hebert et al., 2003; Ward et al., 2009). The quantity of partial sequences of COI gene made accessible in public databases has increased as a result of the ground-breaking concept. The 'Barcode of Life Datasystems database' (BOLD) contains approximately 20.5 million COI barcode sequences belonging to 3.6 lakh species as of January 2025, and the use of COI barcodes to identify and find new life forms has been covered in a staggering number of research articles. Such attempts have made the COI gene

fragments of the animal kingdom's most thoroughly sequenced gene area. The majority of the DNA barcoding experiments that have been published interpret gene region only as an identification tag, which is exactly in line with the idea of the legibly readable barcode. Nevertheless, the barcode region is situated at the center of cellular energy production, where COI is one of the cytochrome c oxidase protein's building parts (COX).

The Cytochrome c oxidase protein is made up of two components, which are identical and exist as a dimer. These are made up of several amino acid chains and metallic ligands including two iron atoms attached with heme groups along with three coppers, one zinc and one magnesium (Tsukihara et al., 1996; Balsa et al., 2012). COX is used as an enzyme in the electron transport chain, which reduces oxygen and pumps protons across the inner membrane of mitochondria. Hence, the variations in the sequences of amino acids that alter the structure of the protein may have an impact on energy metabolism. Such alterations are likely to be mediated by modifications around the enzymatically active sites or contact areas of the amino acid chain (Mathews et al., 2013).

DNA variation patterns in the barcode region may indicate potential and limitations for protein development. To comprehend the relevance of the DNA barcoding technique for taxon-specific goals of species discrimination, as well as to derive functional insights from the vast amount of data accumulated to date, these two layers of information are prerequisites, which are interconnected. The COI sequences produced by the DNA barcoding program were examined in the present study to compare the variations in amino acids in DNA barcode areas across the different fish families. In order to transfer the conserved and variable amino acid sites onto a three-dimensional model of the COI protein structure, the variations at each amino acid site were quantified and subsequently, the observed variation is projected onto the evolutionary tree.

Materials and Methods

Sampling

Specimens of the fishes that had been caught by the local fishermen immediately after fishing from the Bay of Bengal's coastal waters were collected from the Gopalpur-on-Sea, Odisha, India. The specimens were brought to the laboratory, morpho-taxonomically identified using the taxonomic keys stated in renowned taxonomic publications such as Commercial Sea Fishes of India (Talwar & Kacker, 1984), Fishes of The World (Nelson et al., 2016), and Fauna of India and Adjacent Countries (Talwar, 1995) (S1 appendix). Nevertheless, the specimen's species-level classification was accomplished by utilizing the approved taxonomic keys specified in Eschmeyer's Catalogue of Fishes (Fricke et al., 2022). Later, fin and muscle tissues were extracted under aseptic circumstances and stored at -20°C for further investigations.

DNA Isolation, Amplification, and Sequencing

Morphologically species delineations were reconfirmed by molecular taxonomy using DNA barcoding. Total genomic DNA was extracted from muscle tissue utilizing a revamped version of the salting-out approach proposed by Sambrook and Russell (2001). The extracted DNA was then analyzed using 1.5 percent agarose gel electrophoresis and kept at -20°C for later use. mtCOI gene was amplified using the isolated DNA as a template and oligonucleotide primers reported in the earlier study by Ivanova et al. (2007). For the amplification, 25 microliters of reaction mixture containing 10 µM of each primer, 1.0 u Dream Taq DNA polymerase (Thermo Scientific, USA), 10 mM dNTPs mix, and 1X PCR Assay buffer containing 20 mM $MgCl_2$, was prepared. The following conditions were used for the PCR: 95°C for 2 mins; 35 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min; 72°C for 10 mins. Later, the amplified products were subjected to 1.5 percent agarose gel electrophoresis. Following the successful amplification of the targeted COI gene from the DNA of collected samples, the amplified samples were commercially sequenced by Bioserve Biotechnologies Pvt. Ltd., Hyderabad, India, using the di-deoxy chain termination technique. Electropherograms were base-called utilizing the program Codoncode aligner (v. 11.0.2, Codon Code Corporation), and consistent sequence repeats of superior-quality bases (Quality value (QV)>20) were regarded for further investigation. The BioEdit program (Hall 1999), was used to remove the noisy sequences from both ends. Accurate barcode sequences were generated, aligned, and further consensus sequences were generated using procedures documented by Hall (1999). To detect the existence of any insertions, stop codons, or deletions, the Expasy portal was utilized. The BLASTN feature on the NCBI (National Center for Biotechnology Information) web service was used to verify if the sequence corresponded to the desired locus; following that, the fragment featuring perfect alignment with no gaps, insertions, or deletions was picked. Sequences were eventually submitted to NCBI, and accession numbers were obtained.

Nucleotide sequences along with the amino acid sequences generated from the selected fish families were considered for the current study. The accession numbers and protein ID of the respective sequences are mentioned in Table 1. The amino acid sequences of respective nucleotide sequences as well as the sequences of the same species from different locations were downloaded from the NCBI database (S2 appendix). Thereafter, the amino acid sequences were then aligned with ClustalW using MEGA X software (Kumar et al., 2018).

Recognition and Revelation of Variable Amino Acid Sites

The distribution of amino acids within the sequences was determined using MEGA X software. Entropy [uncertainty; H(x)] values for all the positions were computed in BioEdit to assess the variation at each amino acid position in the studied sequence dataset. Entropy has a value of 0 when there is zero variation. The entropy value rises which corresponds to the rise in variability. The variable amino acids were subsequently separated into four (arbitrary) entropy-increasing groups based on the results: H(x) 0.5-0.7, 0.71-0.9, 0.91-1.1, and >1.1. Entropy values less than 0.5 were deemed to be non-variable amino acid locations, whereas residues with zero variation were deemed to be conserved. Amino acids were separated into standard groups with well-defined chemical properties at each site: non-polar aliphatic (G, A, V, L, M, I); polar uncharged (S, T, C, P, N, Q); aromatic (F, Y, W); positively charged (K, R, H); and negatively charged (D, E) as described by earlier study of Pentinsaari et al. (2016). When the amino acids at a given position exhibited variation exclusively among amino acids within the groups, the site was considered as a non-variable region and treated as equal to those sites with entropy <0.5.

Mapping of Amino Acid Variations on Phylogeny

To evaluate the phylogenetic relationship among the selected eleven fish families, the translated amino acid sequences of the generated DNA barcodes were aligned, and a Neighbor-joining phylogenetic tree with 1000 bootstrap repetitions was constructed using Mega X software.

Structural Modelling

Cattle (*Bos taurus*) was used as a suitable model to locate the COI barcode region within the COI protein as its Cytochrome c Oxidase protein structure is predominantly well-researched at a fine resolution (Tsukihara et al., 1995; 1996; 2003). The 3D homology models of the COI barcode region were developed by using the Swiss model server (https://swissmodel.expasy.org/) followed by their structure quality validation using the saves server (https://saves.mbi.ucla.edu/). Homology models were generated depending on the bovine protein X-ray structure with protein ID: 3abm.1.A for all the families except for Priacanthidae and Sciaenidae families, bovine protein X-ray structure with protein ID: 8d4t.1.A was used as a template. Ten models per protein were generated and from all the models, the lowest energy model was chosen for visualization in PyMOL Molecular Graphics System 1.7 (Schrödinger, LLC). The marked focus was given to the amino acid positions with the amino acid substitutions involving different chemical groups having entropy >0.5.

Results

Amino Acid Variation Within Species

For the study of amino acid variation in the COI sequences of *Acanthurus mata,* 234 amino acids were considered. In evidence of functional constraints influencing diversity in amino acid sequences, 233 out of the 234 amino acids were found to be completely conserved among the *A. mata*. At position 150, only one sequence of *A. mata* was from the Arabian Sea on the coast of India, where the non-polar amino acid leucine (L) was substituted by polar charged amino acid proline (P). Although the substitution of amino acid was from different chemical groups and therefore it was not considered as variable, because the entropy value was less than 0.5. For the study of amino acid variation in the COI sequences of both *Ephippus orbis,* 231 amino acids were considered. In *E. orbis*, all 231 amino acids were observed to be completely conserved. It is evident from the study that there were no variations in amino acids in the COI sequence of *E. orbis.* For the study of the amino acid variations in the COI sequences of *Myripristis botche,* 210 amino acids were considered. All the 210 amino acids were found to be completely conserved among the *M. botche*. The study revealed that there were no variations in amino acids in the case of all the sequences of *M. botche*.

For the study of the amino acid variations in the COI sequences of *Upeneus sulphureus* and *Upeneus pori,*

Table 1. COI sequences from the studied fish families for structural modelling

Sl. No.	Fish family	Species	Accession No.	Protein ID
1.	Acanthuridae	*Acanthurus mata*	ON222817	UOU35138
2.	Ephippidae	*Ephippus orbis*	ON197105	UOM32510
3.	Holocentridae	*Myripristis botche*	ON384426	UPQ55960
4.	Mullidae	*Upeneus pori*	ON182865	UOK00551
5.	Scorpaenidae	*Pterois russelii*	ON248008	UOX99372
6.	Synanceiidae	*Minous monodactylus*	ON246993	UOX76502
7.	Gerreidae	*Gerres filamentosus*	OK353776	UBU02971
8.	Monodactylidae	*Monodactylus argenteus*	ON384430	UPQ55961
9.	Nemipteridae	*Nemipterus japonicus*	OK346330	UBT68516
10.	Priacanthidae	*Priacanthus tayenus*	ON248034	UOX99380
11.	Sciaenidae	*Dendrophysa russelii*	ON248017	UOX99378

235 and 233 amino acids were considered respectively. In evidence of functional constraints influencing diversity in amino acid sequences, the variations were found at position 20 of the 235 amino acids among the *U. sulphureus*. However, the significant variables (entropy >0.5) were observed at the amino acid positions 101, 179, and 215. Other positions were considered as non-variable regions because the entropy values were less than 0.5 or substitutions of the amino acids had the same chemical group. At position 101, about 9 sequences of *U. sulphureus* showed the non-polar amino acid leucine (L) and 12 sequences showed the aromatic amino acid phenylalanine (F). At the amino acid positions 179 and 215 of the 9 sequences, the non-polar amino acid alanine (A) was substituted by polar uncharged amino acid serine (S). Besides, the study indicated that there were two variations found in one sequence of the *U. pori* amino acid sequences, which were considered as the non-variable regions due to the minimal value of entropy less than 0.5. Thus, the analysis suggested that there were no significant variations in the amino acid sequences found among *U. pori.* About 233 amino acids were considered for the study of the amino acid variations in the COI sequences of *Pterois russelii*. The study indicated that there were variations in 8 amino acid positions among *P. russelii.* Surprisingly, all 8 variations were in a single sequence, which was from Bangladesh having the same substitution *i.e.* from methionine (M) to isoleucine (I). As the amino acid substitution was from the same chemical group and the entropy value was less than 0.5, it was considered as non-variable region. For this purpose, 234 amino acids were considered to study the amino acid variations in the COI sequences of *Minous monodactylus*. The study suggested that the variations were found in 2 of 234 amino acid positions among *M. monodactylus*, which were considered as non-variable due to the minimal value of entropy less than 0.5.

For the study of the amino acid variation in the COI sequences of *Gerres filamentosus* and *Gerres erythrourus,* 236 and 235 amino acids were considered respectively. The study revealed that there were variations in the 6 amino acid positions in both *G. filamentosus* and *G. erythrourus*. Besides, all the positions were considered as non-variable, because the substitutions were from the same chemical group with entropy values less than 0.5 in both *G. filamentosus* and *G. erythrourus*. Accordingly, 240 amino acids were considered to determine amino acid variation in the COI sequences of *Monodactylus argenteus*. In evidence of functional constraints influencing diversity in amino acid sequences, 239 out of the 240 amino acids were found to be completely conserved among the *M. argenteus*. At position 80, about 9 sequences of *M. argenteus* showed the non-polar amino acid valine (V) and 7 sequences showed the non-polar amino acid isoleucine (I). Although the entropy value for such variation was more than 0.5, the substitution of amino acid was from the same chemical group and hence considered as non-variable. For the study of the amino acid variation in the COI sequences of *Nemipterus japonicus* and *Scolopsis vosmeri,* 236 and 230 amino acids were considered respectively. Based on the estimated entropy values, it was observed that, although the amino acid variations occurred at 6 different positions in the case of the sequence datasets belonging to species *N. japonicus,* they all were considered as non-variable. The study clearly suggested that 3 variations out of these 6 variations were considered to be within-group variations and the rest showed the entropy values less than 0.5. Besides, the study also revealed that there was only one variation in the case of the species *S. vosmeri*. However, the substitution was considered as non-variable, because the entropy value contributed by such variation was found to be less than 0.5 and the substitution of amino acid was from the same chemical group. For the study of the amino acid variation in the COI sequences of *Priacanthus tayenus,* 225 amino acids were considered. The study revealed that variations were found in the 2 of 225 amino acid positions among the *P. tayenus*. Besides, all the positions were considered as non-variable, because the substitutions were from the same chemical group and minimal entropy values less than 0.5. About 235 amino acids were considered for analysing the amino acid variation in the COI sequences of *Dendrophysa russelii*. The study revealed that variations were found in the 3 of 235 amino acid positions among the *D. russelii*. Besides, all the positions were considered as non-variable, because the substitutions were from the same chemical group and entropy values were less than 0.5.

Amino Acid Variations Among Various Fish Families

The amino acids distribution in the COI sequences of different fish families was presented in Figure 1. As anticipated for a membrane-bounded protein, the segment coded by the DNA barcode sequence consisted mostly of nonpolar amino acids with lowest (52.8%) in the Gerreidae family and highest (58%) in the Priacanthidae family. It was followed by the polar uncharged amino acids with an average of 25%. Family Sciaenidae had the highest polar uncharged amino acid whereas the Priacanthidae family showed the lowest aromatic amino acid content. However, both the positive and negatively charged amino acids were equally distributed in all the families (Figure 2). The cysteine amino acid was totally absent in all the families. Besides, the distribution of lysine was found to be negligible in all the families. Relatively highest distribution of leucine followed by alanine amino acid was observed in all the families. The study unequivocally showed that the distribution of amino acids among the fish families varied greatly (Figure 1).

Further, 235 amino acids were considered to elucidate the amino acid variations in the COI sequences of different families (Figure 3). In evidence of functional constraints influencing diversity in amino acid

sequences, there were 28 variations in amino acids, and 207 out of 235 amino acids were found to be completely conserved in all the families (Table 2). Among all the variations, 3 amino acid positions were considered as variable (0.5<entropy<0.7). At the amino acid position 34 of the Ephippidae family, the amino acid serine (S) is substituted by asparagine (N) of the same polar uncharged chemical group, whereas it is substituted with alanine (A) of other nonpolar chemical group in the Mullidae family. Most families had non-polar amino acid alanine (A) at position 38, which was substituted with serine (S) of the polar uncharged group in the families Gerreidae, Priacanthidae, and Sciaenidae. At position 168, the non-polar amino acid isoleucine (I) was substituted with polar uncharged threonine (T) and valine (V) in the Acanthuridae and Sciaenidae families respectively. The rest 25 amino acid positions were either substituted with the same chemical group with the entropy value less than 0.5 and hence were considered as non-variable.

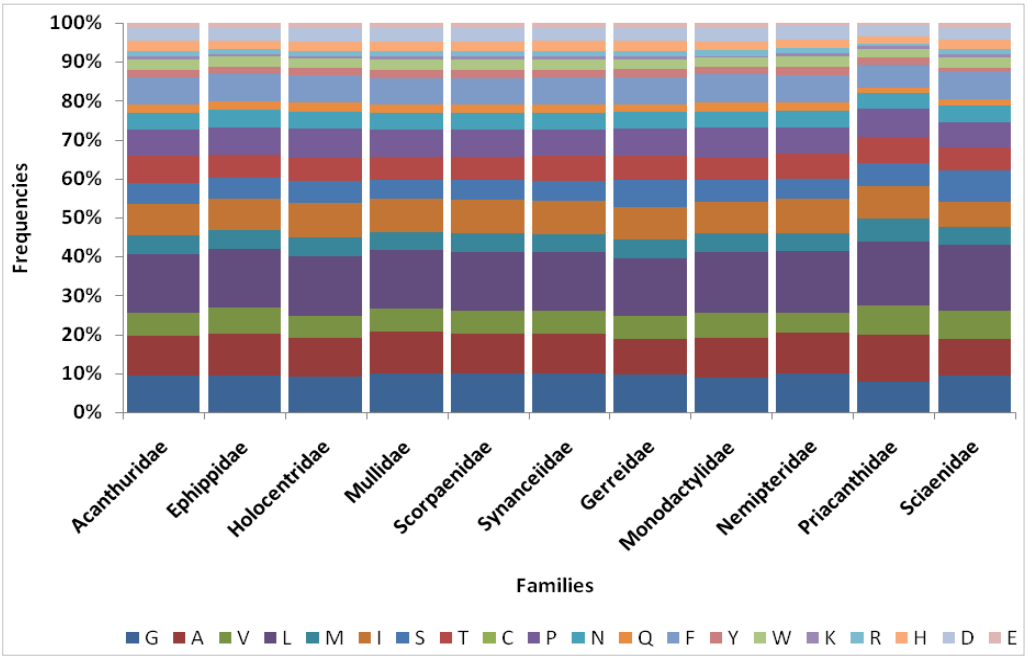

Figure 1. Average amino acid frequencies in barcode sequences of the selected fish families. Single-letter standard codes were provided for amino acids. Amino acids were grouped in accordance to their biochemical properties.

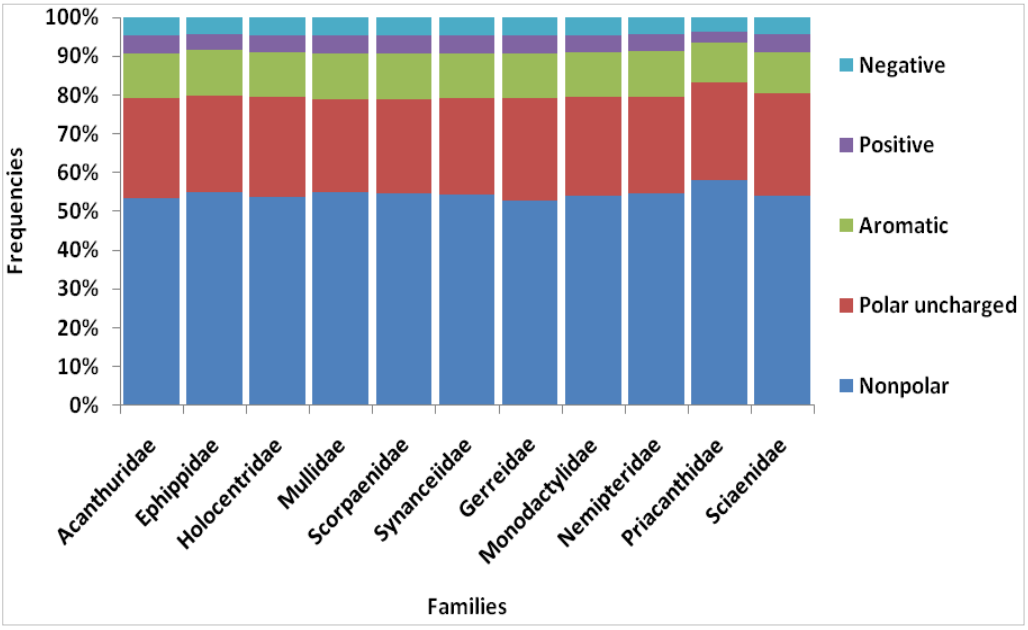

Figure 2. Average amino acid frequencies of the amino acid groups in barcode sequences of fish families. The amino acids are grouped by their biochemical properties as explained in Methods. The different groups of amino acids are nonpolar, polar uncharged, aromatic, positive, and negatively charged amino acids.

Mapping of Amino Acid Variations

Major amino acid substitutions from one biochemical group to another have been indicated in the NJ phylogenetic tree (Figure 4), which was also supported by the Maximum Likelihood (ML) phylogenetic tree (S3 appendix). Among the Root mean square deviation (RMSD) values (0-0.113) of the superimposed structures of the studied families, RMSD values of the Priacanthidae family (0.111) and Sciaenidae family (0.113) were found to be much higher as compared to the other families (0-0.010), which was also supported by the NJ phylogenetic tree. The study revealed that the Sciaenidae family was separated as an outgroup among all the selected fish families due to the occurrence of the highest number of variations in amino acids, which evidenced the effect of functional constraints on biodiversity.

Structural Variations

The animal DNA barcode fragments were covered with 230 amino acids showing the enzymatically active component Pfam.cox1 domain with the amino acid positions 4 to 233, where the transfer of electrons takes place from Cu to heme group. The protein segment coded by the DNA barcoding region of the COI gene was superimposed on the complete COI gene of *Bos taurus* to locate the DNA barcoding region (Figure 5). The secondary structure of the barcode portion comprises of six α -helices such as Helix 1, Helix 2, Helix 3, Helix 4, Helix 5 and Helix 6 starting from the N-terminus joined by five loops such as Loop 1–2, Loop 2–3, Loop 3-4, Loop 4-5 and Loop 5-6. The three-dimensional protein structures in the study suggested that most of the conserved amino acid residues were located in the helices that perforate the inner mitochondrial

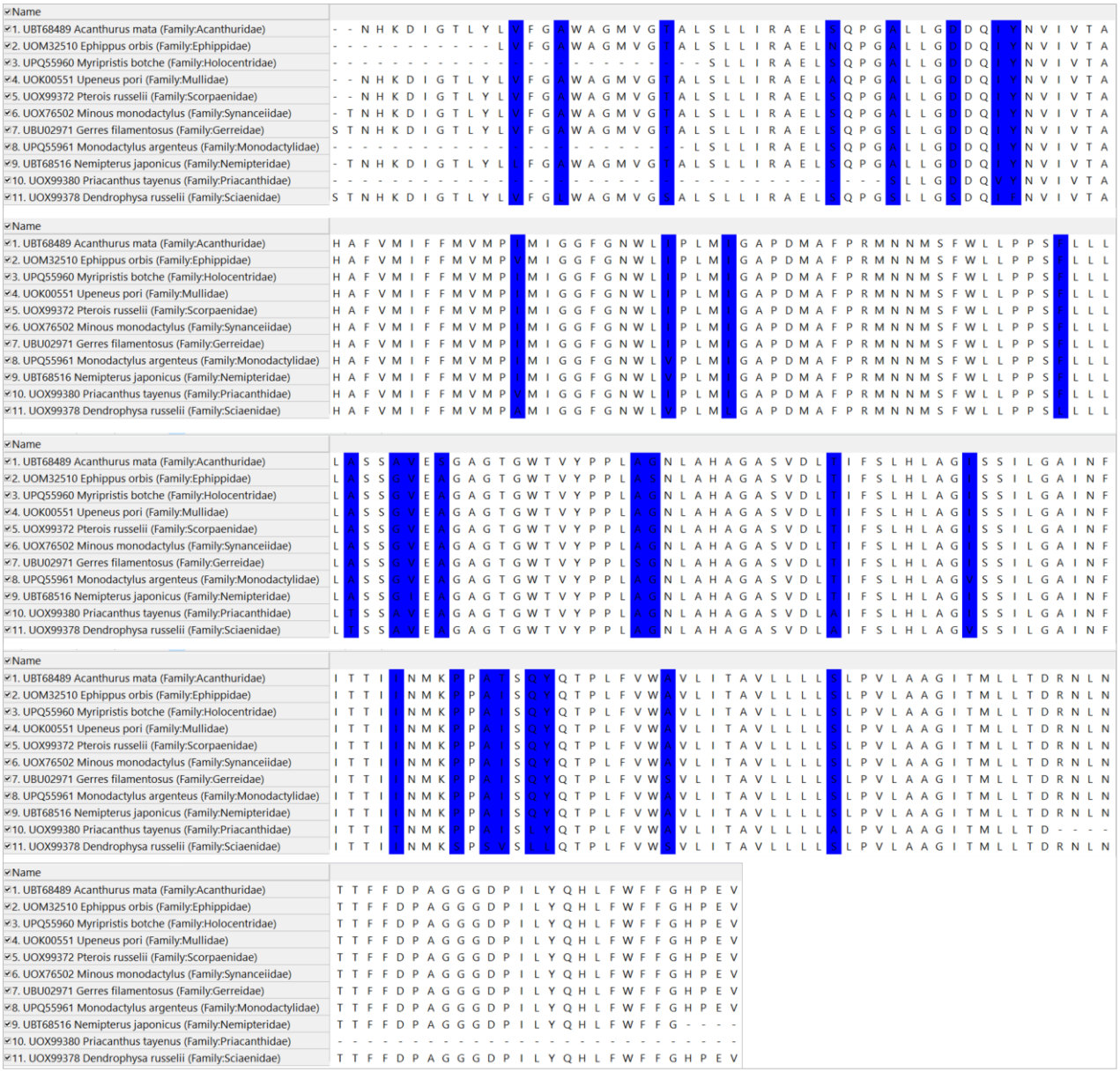

Figure 3. Amino acid sequence alignment of the COX1 sequences of the species representing different families. Variable amino acids are highlighted in blue colour, and the rest of the amino acids are conserved.

membrane (Figure 6). However, three of the loops (Loop 2-3, Loop 3-4, and Loop 5-6) of the protein were also characterized by conserved amino acids. Across the dataset, three amino acids showed variability. The detected variation occurred at sites that were far away from the active site of protein, and hence unlikely to have any impact on the protein function. Major amino acid transitions from one biochemical group to another at these sites have been shown in the phylogenies (Figure 6).

Discussion

The COI region not only demonstrates enough conservation within the species, but also large variations between the species, which allows the accurate identification of taxon (Hebert et al., 2003). Particularly in the cytochrome oxidase genes, the amino acid changes are uncommon (Pesole et al., 1999; Castoe et al., 2008). Based on the DNA-level differences, the Folmer region of the COI is mostly employed for species discrimination (Folmer et al., 1994). While being different between species, the area is sufficiently preserved within each species. Due to its location in the COI's active enzymatic domain, this region participates in the respiratory chain by transferring electrons from the Cu to heme. As a result, this area is functionally constrained, and mutations that occur near heme should be fatal. To map the conserved and variable amino acid sites onto the three-dimensional model of COI protein structure, the variations at each amino acid site were evaluated.

As anticipated for membrane-bounded protein, the segment coded by the DNA barcode sequence consisted largely of nonpolar amino acids, which were 52.8% (lowest) in the Gerreidae family to 58% (highest)

Table 2. Amino acid substitutions in the COI barcode sequences of studied fish families revealed huge variations with respect to their distribution pattern.

Sl.	Amino acid position	Fish family	Amino acid variation
1.	13	Nemipteridae	V to L
2.	16	Sciaenidae	A to L
3.	23	Sciaenidae	T to S
4.	34	Ephippidae	S to N
		Mullidae	S to A
5.	38	Gerreidae	A to S
		Priacanthidae	A to S
		Sciaenidae	A to S
6.	42	Sciaenidae	D to S
7.	45	Priacanthidae	I to V
8.	46	Sciaenidae	Y to F
9.	65	Ephippidae	I to V
		Priacanthidae	I to V
		Sciaenidae	I to A
10.	75	Monodactylidae	I to V
		Nemipteridae	I to V
		Sciaenidae	I to V
11.	79	Sciaenidae	I to L
12.	101	Sciaenidae	F to L
13.	106	Priacanthidae	A to T
		Sciaenidae	A to T
14.	109	Acanthuridae	G to A
		Priacanthidae	G to A
		Sciaenidae	G to A
15.	110	Nemipteridae	V to I
16.	112	Acanthuridae	A to S
17.	125	Gerreidae	A to S
18.	126	Ephippidae	G to S
19.	138	Priacanthidae	T to A
		Sciaenidae	T to A
20.	147	Monodactylidae	I to V
		Sciaenidae	I to V
21.	161	Priacanthidae	I to T
22.	165	Sciaenidae	P to S
23.	167	Sciaenidae	A to S
24.	168	Acanthuridae	I to T
		Sciaenidae	I to V
25.	170	Priacanthidae	Q to L
		Sciaenidae	Q to L
26.	171	Sciaenidae	Y to L
27.	179	Gerreidae	A to S
		Sciaenidae	A to S
28.	190	Priacanthidae	S to A

in the Priacanthidae family. It was followed by the polar uncharged amino acids with an average of 25%. Differences in the amino acid composition of the barcode region of the COI sequences among studied families were noticed. Alanine (A) and valine (V) were lower in the Gerreidae family and highest in the Priacanthidae family compared to other families. Cysteine was completely absent in the barcodes of all families. Besides, lysine (K) was present in small quantities in all families with the lowest in Priacanthidae. In the Priacanthidae family, two other amino acids such as arginine (R) and glutamic acid (E) were likely to be encoded in small amounts. Similar findings were earlier reported by Pentinsaari et al. (2016) and Sabir et al. (2019). Thus, COI barcodes can reveal information about how the encoded protein has evolved as well as their functional attributes at various taxonomic scales.

The variation of DNA in different portions of the COI barcoding sequence should be related to the functional role of these sections in the protein (Pentinsaari et al., 2016). From the amino acid variation study between the selected families, 28 variations were detected inside the COI barcode sequences among eleven families. Among them, 3 positions of the amino acids were considered as variable. The remaining positions were considered as non-variable as the amino acid substitution was from the same chemical group or the entropy value was less than 0.5. The substitution of amino acids with the same chemical attribute at any

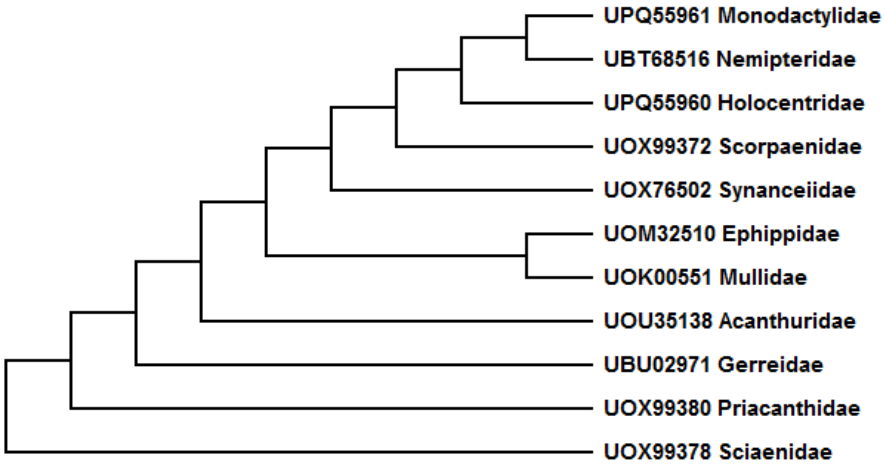

Figure 4. Neighbor-joining phylogenetic tree using selected families. The evolutionary history was inferred using the Neighbour-Joining method (Saitou & Nei, 1987) and analyzed using MEGA X (Kumar et al., 2018). The evolutionary history of the taxa under study was assumed to be represented by a bootstrap consensus tree obtained from 1000 replicates (Felsenstein, 1985). Evolutionary distances were computed using the Poisson correction method (Zuckerkandl & Pauling, 1965).

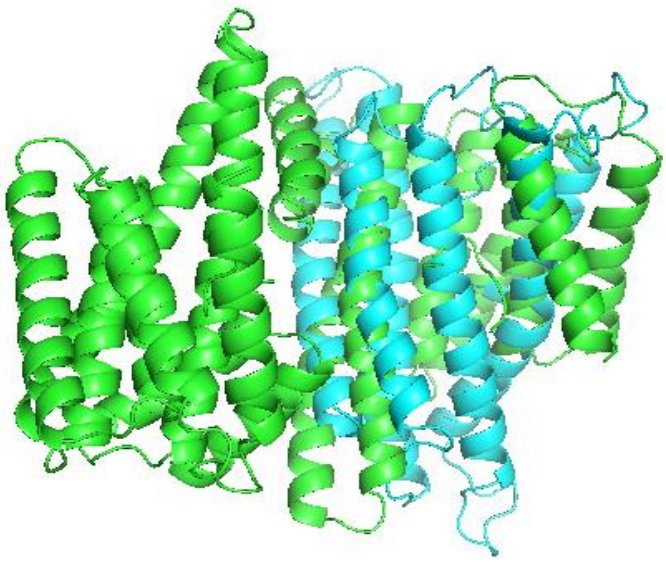

Figure 5. Location of protein fragment coded by DNA barcoding region of COI gene. Most of the fragments were located in the transmembrane region of cytochrome oxidase (COX) protein. The entire COX molecule reconstructed from cattle was displayed in green, with the barcode sequence shown in cyan.

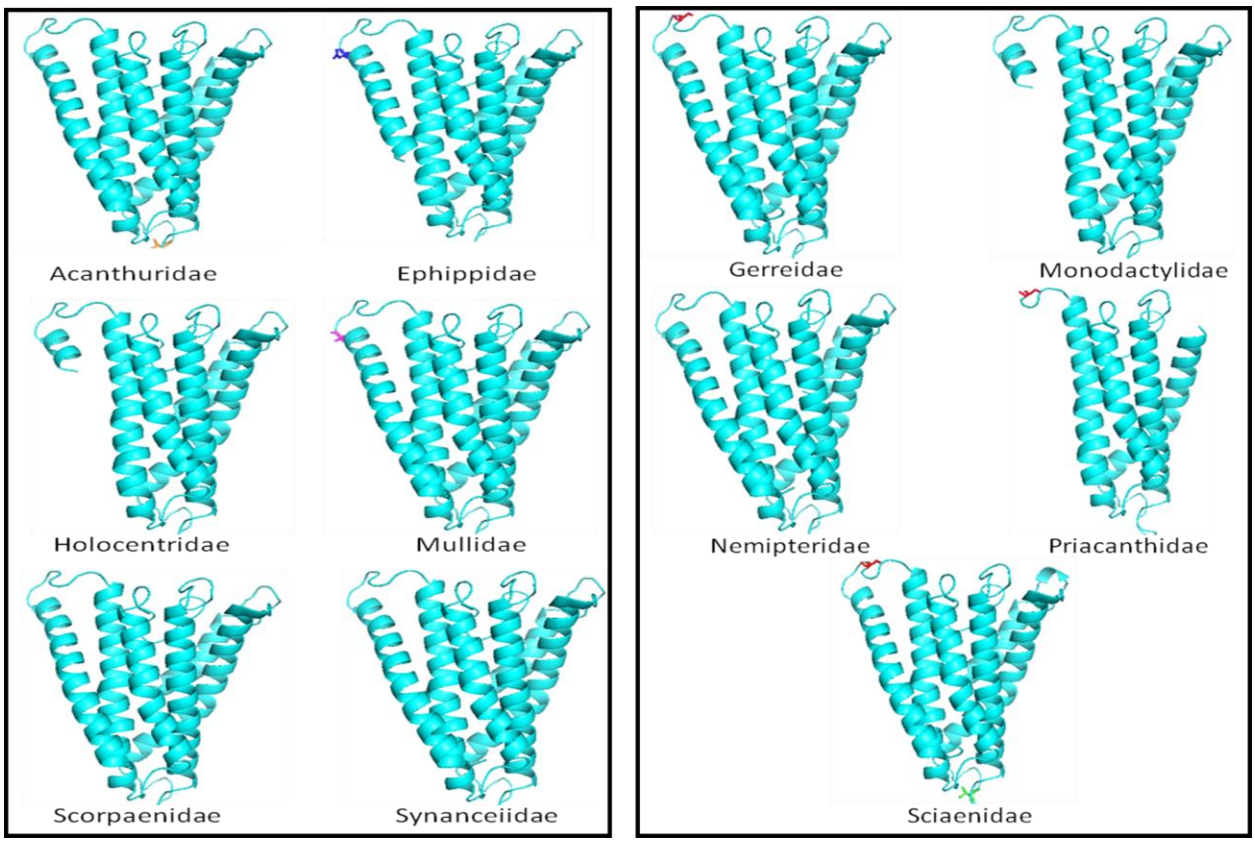

Figure 6. Homology models and amino acid variation of the COI barcode sequences in fish families. Variable amino acids were highlighted as sticks. Pink colour and blue colour reflected amino acid variation S→A and S→N respectively at position 34. The red colour reflected amino acid variation A→S at position 38. The orange and green colours reflected amino acid variation I→T and I→V respectively at position 168. The lower end of the models points to the mitochondrial matrix and the upper end to the intermembrane space.

given location is not thought to affect significantly on enzyme performance (Pentinsaari et al., 2016; Sabir et al., 2019). As the variable amino acid positions were in the loop region, it may not affect the function of the COI protein. Among all the families, the Sciaenidae family showed the highest number of variations, with variations in the 19 amino acid positions. Due to the highest number of amino acid variations, it was reflected in the protein structure of the Sciaenidae family. The root mean square deviation (RMSD) value of the fish family Sciaenidae was found to be relatively higher among all the studied families, which was also supported by the NJ phylogenetic tree. The study indicated that the Sciaenidae family was separated as an outgroup from the other families, as reflected in the NJ tree.

The variations were identified in regions considered to be functionally redundant (such as the loop structures of protein), and the highest degree of conservatism is closer to the active site, where the functional constraints are likely to limit the differences in the amino acid sequence. The results were consistent with the prior findings, which suggested that the loop structures exhibit amino acid change more frequently compared to rigid helices (Panchenko et al., 2005). The loops in COI are primarily extra-membranous. In contrast to membrane-bounded helices, which are constrained by their lipophilic and crowded surroundings, they may permit greater flexibility in amino acid charge and size. However, it has been suggested by Ulmschneider et al. (2005) that the extra-membranous loop shapes of some transmembrane proteins influence membrane dynamics and protein stability. Therefore, more research will be necessary to determine whether variations in the COI loop sequences have functional implications (Pentinsaari et al., 2016). Further, laboratory analysis is recommended to know the effects of changes in amino acids consequently on body metabolism in the target species.

Conclusions

These results of the current study show how the global DNA barcode library can be used for purposes far beyond identification and will be the subject of much future research. Combining phylogenetic and biochemical studies on the COX enzyme is encouraged as a particularly viable route for further exploitation of the sequence data produced by the worldwide DNA barcoding program. This research will provide an intriguing perspective on evolution. The results of this small-scale investigation also point to the need for additional research on the structural variations in the COX protein and amino acid composition among the entire fish biota.

Ethical Statement

Since this study did not disturb environmental populations, ethical approval was not required.

Funding Information

The authors received no specific funding for this work.

Author Contribution

Conceptualization: BS, TKB & AKP, Methodology: BS, RS & TKB, Validation: BS, RS, & TKB, Formal analysis: BS & RS, Investigation: BS, RS & TKB, Resources: TKB, Original Draft Preparation: BS Review & Editing: BS, TKB & AKP, Visualization: BS & RS, Supervision: TKB & AKP.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The names, degrees, and affiliations who have contributed substantially to a study but do not fulfill the criteria for authorship can be listed in the Acknowledgments section.

References

Balsa, E., Marco, R., Perales-Clemente, E., Szklarczyk, R., Calvo, E., Landázuri, M. O., & Enríquez, J. A. (2012). NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. *Cell Metab, 16* (3), 378-86. https://doi.org/10.1016/j.cmet.2012.07.015

Castoe, T. A., Jiang, Z. J., Gu, W., Wang, Z. O., & Pollock, D. D. (2008). Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One, 3*, e2201. https://doi.org/10.1371/journal.pone.0002201

Cross, H. B., Lowe, A. J., & Gurgel, C. F. D. (2011). DNA barcoding of invasive species. Fifty years of invasion ecology: the legacy of Charles Elton (ed. by D. Richardson), 289–299. Wiley, Hoboken.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution, 39*, 783-791. https://doi.org/10.2307/2408678

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology Biotechnology, 3*, 294–299. https://doi.org/10.1371/journal.pone.0013102

Fricke, R., Eschmeyer, W. N., & Fong, J. D. (2022). "Eschmeyer's Catalog of Fishes: Genera/Species by Family/Subfamily (http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp), Electronic version.

Hall, T. A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings Biological sciences, 270*(1512), 313-21. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings Biological sciences, 270*(1), S96-S99. https://doi.org/10.1098/rsbl.2003.0025 https://www.ncbi.nlm.nih.gov/ https://saves.mbi.ucla.edu/ https://swissmodel.expasy.org/

Ivanova, N. V., Zemlak, T., Hanner, R., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding, *Molecular Ecology Notes, 7*(4):544-548, https://doi.org/10.1111/j.1471-8286.2007.01748.x

Nelson, J. S., Grande, T. C., & Wilson, M. V. H. **(**2016**)**. Fishes of the World (Fifth ed.), John Wiley & Sons, Hoboken, New Jersey, USA.

Kochzius, M., Antoniou, A., Botla, S., Campo, D., Cariani, A., Garcia-Vazquez, E., Hauschild, J., Hervet, C., Hjörleifsdottir, S., Hreggvidsson, G., Kappel, K., Landi, M., Magoulas, A., Marteinsson, V., Nölte, M., Planes, S., Tinti, F., Turan, C., & Blohm, D. (2010). Identifying Fishes through DNA Barcodes and Microarrays. *Plos One, 5*. https://doi.org/10.1371/journal.pone.0012620

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution, 35*, 1547-1549. https://doi.org/10.1093/molbev/msy096

Mathews, C. K., Holde, K. E. van, Appling, D. R., & Anthony-Cahill, S. J. (2013). Biochemistry 4th edition (Pearson).

Panchenko, A.R., Wolf, Y. I., Panchenko, L. A., & Madej, T. (2005). Evolutionary plasticity of protein families: Coupling between sequence and structure variation. *Proteins, 61*, 535–544. https://doi.org/10.1002/prot.20644

Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports, 6*, 35275. https://doi.org/10.1038/srep35275

Pesole, G., Gissi, C., De Chirico, A., & Saccone, C. (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution, 48*, 427-434. https://doi.org/10.1007/PL00006487

Radulovici, A. E., Archambault, P., & Dufresne, F. (2010). DNA barcodes for marine biodiversity: moving fast forward? *Diversity, 2,* 450–472. https://doi.org/10.3390/d2040450

Sabir, J., Rabah, S., Yacoub, H., Hajrah, N., Atef, A., Almatry, M., Edris, S., Alharbi, M., Ganash, M., Mahyoub, J., al-hindi, R., Alghamdi, K., Hall, N., Bahieldin, A., Kamli, M., & Ratherid, I. (2019). Molecular evolution of cytochrome C oxidase-I protein of insects living in Saudi Arabia. *PLos One 14*, 1-18. https://doi.org/10.1371/journal.pone.0224336

Saitou, N., & Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution 4*:406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Talwar, P. K., & Kacker, R. K. (1984) Commercial Sea Fishes of India, Zoological Survey of India, Calcutta**.**

Talwar, P. K. (1995). Fauna of India and adjacent countries, Zoological Survey of India, Kolkata.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., & Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. *Science 24*, 272 (5265), 1136-44. https://doi.org/10.1126/science.272.5265.1136

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono R, & Yoshikawa S (1995). Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 A. Science, 25, 269 (5227):1069-74.

Tsukihara, T., Shimokata, K., Katayama, Y., Shimada, H., Muramoto, K., Aoyama, H., Mochizuki, M., Shinzawa-Itoh, K., Yamashita, E., Yao, M., Ishimura, Y, & Yoshikawa, S. (2003). The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proceedings of the National Academy of Sciences of USA, 23*, 100 (26), 15304-9. https://doi.org/10.1073/pnas.2635097100

Ulmschneider, M. B., Tieleman, D. P. & Sansom, M. S. P. (2005). The role of extra-membranous inter-helical loops in helix–helix interactions. Protein Engineering, *Design and Selection,* 18, 563–570. https://doi.org/10.1093/protein/gzi059

Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, Fish-Bol. *J Fish Biol, 74,* 329–356. https://doi.org/10.1111/j.1095-8649.2008.02080.x

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, 97-166. Academic Press, New York. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6