Ceylon Journal of Science 55 (4) 2026: 1675-1696

RESEARCH ARTICLE

DNA Barcoding of Freshwater Fish Species from Upper Lake: A Ramsar Site, Bhopal, India: Implications for Species Authentication and Reference Barcode Database Development

N. Uikey, S. Yadav and R.K. Garg

Highlights

- DNA barcoding based on the mitochondrial *cox1* gene was used to identify fish species from Upper Lake, a Ramsar wetland in central India
- 560–640 bp barcode sequences were generated for 24 specimens representing 20 fish species.
- Low intraspecific divergence and high interspecific divergence provided a clear barcode gap for accurate species discrimination.
- Nearest-neighbour analysis demonstrated the effectiveness of *cox1* barcodes for reliable species discrimination.

Ceylon Journal of Science 55 (4) 2026: 1675-1696
DOI:https://doi.org/10.4038/cjs.v55i4.9568

RESEARCH ARTICLE

DNA Barcoding of Freshwater Fish Species from Upper Lake: A Ramsar Site, Bhopal, India: Implications for Species Authentication and Reference Barcode Database Development

N. Uikey[1*], S. Yadav[2] and R.K. Garg[1]

[1]*Department of Bioscience, Barkatullah University, Bhopal 462026, Madhya Pradesh, India*
[2]*Sri Sathya Sai College for Women, Bhopal 462024, Madhya Pradesh, India*

Received: 03.02.2026 Accepted: 17.09.2026

Abstract

Upper Lake, Bhopal, a Ramsar designated wetland and one of the largest natural freshwater lakes in central India, supports rich fish diversity and provides important ecological, economic, and fisheries resources. Accurate identification is expected to make proper conservation of any animal species, yet traditional morphometric methods are challenging and time consuming. To address these limitations, this study employed DNA barcoding using the mitochondrial cytochrome-c oxidase subunit I (*mtcox1*) gene to generate DNA barcodes for 24 fish specimens representing 20 species from 7 orders collected from Upper Lake, Bhopal. We generated sequences ranging from 560 - 640 bp and deposited them in GenBank and the Barcode of Life Data System (BOLD Systems). The mean intraspecific divergence was extremely low (0.58%), exhibiting a significant separation as a clear barcode gap from the minimum interspecific divergence at the genus level (10.24%). Nearest-neighbor method supported species delimitation with a mean interspecific distance (14.7%) 122.5-fold greater than the mean intraspecific distance (0.12%). However, the Barcode Index Number (BIN) analysis revealed notable instances of taxonomic discordance with 24 sequences generating 18 BINs. Nucleotide composition analysis showed a mean GC content of 45.38% with a characteristic positional bias, highest at the first codon position (mean 56.25%) and lowest at the third (37.32%). This study established a novel genetic reference dataset for the fish fauna of this Ramsar designated wetland and demonstrated the utility of the *mtcox1* gene for species identification, biodiversity monitoring, and conservation management of freshwater ecosystems.

Keywords: DNA barcoding; Fish biodiversity; *mtcox1;* Ramsar site; Upper Lake

INTRODUCTION

Fish represent a significant component of animal and an important biological component for maintaining healthy aquatic ecosystems biodiversity (Sheraliev & Peng, 2021). Due to broad distribution and high species richness, fishes possess substantial nutritional value and contribute critically to ecological conservation (Oishi, 2016). India, acknowledged as one of the world's mega- biodiversity nations, ranks ninth in terms of freshwater biodiversity (Mittermeier et al., 1997) with an estimated 2,500 fish species, India possesses a highly diverse ichthyofauna, including 930 freshwater species and 1,570 marine species (Eschmeyer et al., 2010; Kar et al., 2003).

Accurate taxonomic delineation and proper identification of fish species from the foundation of successful biodiversity conservation efforts. Identification of fish based solely on morphology can be problematic, especially at early developmental stages like larvae, or when dealing with incomplete, fragmented, or processed specimens. As none of these species have been barcoded previously, the present study examines the DNA sequence patterns of fishes from Upper Lake, Bhopal, using the cytochrome-c oxidase subunit I (*mtcox1*) gene for DNA barcoding. DNA barcoding is now recognized as a rapid and effective tool for to species identification (DeSalle & Goldstein, 2019).

Hebert et al. (2003a) proposed the mitochondrial cytochrome-c oxidase subunit I (*cox1*) gene as a universal marker for distinguishing most animal species. DNA barcoding improves the potential for species level identification worldwide through a standardized and validated DNA based method (Bhattacharya et al., 2016). Like a UPC identifies products, DNA barcoding uses a short mitochondrial DNA sequence to efficiently

**Corresponding Author's Email: neelima.nu.uikey@gmail.com*
https://orcid.org/0009-0005-1686-2419

This article is published under the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and precisely identify species (Haldar & Nath, 2020). It offers a lot, particularly in the form of instruments that allow for clear specimen identification and evaluation of stock structure (Ward, 2000). According to Ward et al. (2005), DNA barcoding facilitates accurate identification of multiple biological forms, including larvae, corpse, protected species, and damaged specimens, with a high degree of certainty.

DNA barcoding has diverse application across sector, including conservation, pest control, species identification, and product authentication. For fish, DNA barcoding offers clear and reliable identification of most species, including eggs, larvae, and even processed samples (April et al., 2011). The molecular technique has demonstrated effectiveness in detecting numerous non-native species (Kundu et al., 2016). The DNA barcoding approach is not intended to replace traditional taxonomy but rather to serve as a complementary tool, particularly valuable in cases of ambiguous or disputed species identifications, including closely related fish species, thereby enhancing biodiversity assessment (Tang et al., 2023).

Integrating morphological, ecological, and physiological traits with DNA barcoding data strengthens species identification and discovery (Taylor et al., 2012). The BOLD system, established in 2005 as a free bioinformatics platform designed to facilitate the acquisition, storage, analysis, and publication of DNA barcode data. Notably, the system is publicly accessible, allowing unrestricted access for user. Although DNA barcoding has become a widely accepted tool for accurate species identification and biodiversity assessment, a comprehensive *cox1* based barcode reference library for the freshwater fishes of Upper Lake, a Ramsar designated wetland of central India, has not been available. Previous investigations of the lake have relied predominantly on morphological identification, which is often constrained by phenotypic plasticity, ontogenetic variation, and morphological similarity among closely related taxa, potentially leading to species misidentification. Consequently, the absence of an authenticated molecular reference database has limited reliable species verification, long term biodiversity monitoring, and molecular based conservation initiatives in this ecologically significant wetland. To address this knowledge gap, the present study generated and validated *mtcox1* DNA barcodes for fish species collected from Upper Lake and evaluated their discriminatory power using genetic distance, barcode gap, Barcode Index Number and Neighbor-Joining phylogenetic analysis. The scope of this study extends beyond species identification by establishing the first publicly accessible regional DNA barcode reference library for the ichthyofauna of Upper Lake, thereby providing a robust molecular framework for taxonomic verification and biodiversity assessment. The authenticated barcode dataset generated in this study will serve as baseline genetic information for future ecological monitoring, fisheries management, invasive species surveillance, and conservation planning. Furthermore, the findings provide scientific evidence that can support biodiversity policy decisions for Ramsar wetlands and other freshwater ecosystems facing similar conservation challenges.

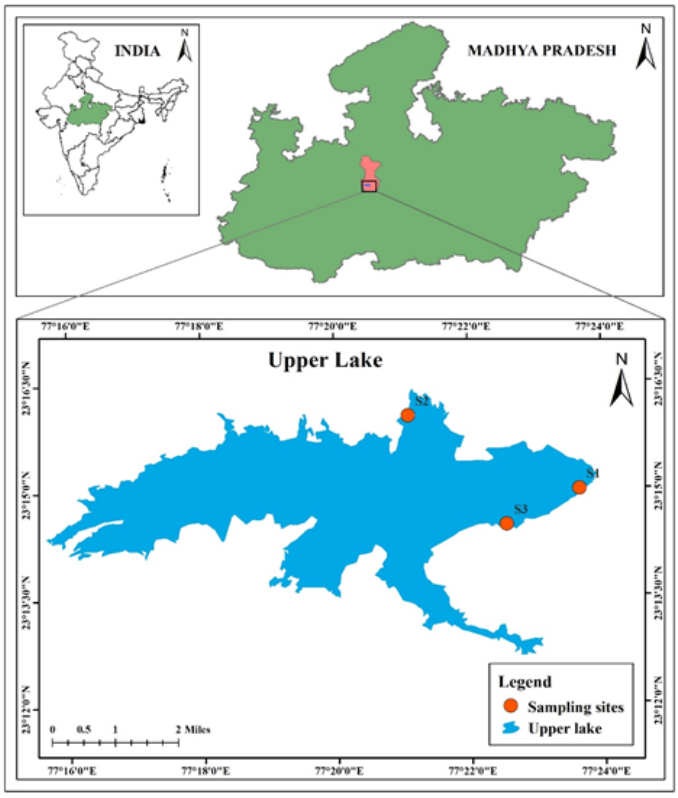


Figure 1: Location of Upper Lake, a Ramsar wetland in Bhopal District, Madhya Pradesh, India, showing the three fish sampling sites: S1 (Vardhman Sunset Park), S2 (Bairagarh), and S3 (Shyamla Hills)

MATERIAL AND METHODS

Study area

The present study was conducted at Upper Lake, located in Bhopal District of Madhya Pradesh, India, (23.2531ºN, 77.3382ºE). Locally known as Bada Talab, the lake was formed in the 11th century by constructing an earthen dam across the Kolans River. It has a catchment area of approximately 361 square kilometres and was designated as a Ramsar Site due to its ecological significance. The lake serves as a crucial habitat for a diverse array of aquatic organisms, particularly fish species. The region is characterized by rich biodiversity and requires effective conservation measures to safeguard the state's natural heritage. The wetland is home to a diverse range of plant and animals. Since wetlands contribute to a healthy and enjoyable environment in a variety of ways, their importance is gradually receiving the attention it deserves on a worldwide scale (Cohen et al., 2016; Richardson et al., 2016).

Sampling sites and sample collection

A total of 33 fish specimens were initially collected from the peripheral zones of various sites (covering marginal areas) of Upper Lake with the assistance of local fishermen (Figure 1). Preliminary species identification was performed based on morphological characteristics using taxonomic keys described by Gopalji (1998) and Jhingran (1991). The specimens were subsequently dissected, and tissue samples were preserved in 70% ethanol and stored at the Laboratory of Molecular Biology and Genomics, Department of Bioscience, Barkatullah University, Bhopal (M.P.), India, for further molecular analysis.

Total genomic DNA extraction and quantification

Total genomic DNA was extracted from the muscle and liver tissues using the phenol:chloroform:isoamyl alcohol (25:24:1) method provided by Sambrook and Russell (2001), with slight modifications. Approximately 100 mg of tissue sample was homogenized in digestion buffer, incubated at 55ºC, and subsequently centrifuged. The resulting supernatant was subjected to PCI treatment, and DNA in the aqueous phase was precipitated with ammonium acetate and chilled ethanol. Following incubation at -20ºC, the DNA pellet was recovered, washed with 70% ethanol, air-dried, and finally resuspended in TE buffer for storage at -20ºC. DNA quality and integrity were assessed by electrophoresis on a 1% agarose gel. To maintain an optimal absorbance ratio of 1.8 at 260/280 nm, samples were treated with Proteinase K or Ribonuclease (RNase). The purified genomic DNA was then diluted to a working concentration of 50 ng/µl for downstream polymerase chain reaction (PCR) amplification of the *mtcox1* gene.

PCR amplification *mtcox1* gene

The *mtcox1* gene was selectively targeted and amplified using a universal primer set comprising the forward primer FishF1 (TCAACCAACCACAAAGACATTGGCAC) and reverse primer FishR1 (TAGACTTCTGGGTGGGTGGCCAAAGAATCA). PCR amplification was performed in a 25µl reaction volume containing 12.5µl of 2X PCR master mix (3B089, 3B BlackBio Biotech, India), 1.0 µl each of forward and reverse primers (10 µM conc.), 1.0µl of DNA template, and 9.5 µl of molecular-grade distilled water. The PCR cycling conditions included an initial denaturation at 95ºC for 5min, followed by 35 cycles of denaturation at 95ºC for 30s, annealing at 53ºC for 1 min, and extension at 72ºC for 1 min. A final extension was performed at 72ºC for 10 min, after which the reaction was held at 4ºC for 10 min.

Agarose gel electrophoresis

PCR products were analysed and separated using a 1% agarose gel prepared in 1X TAE buffer and stained with ethidium bromide (EtBr) for visualization. Electrophoresis was performed, and DNA bands were examined under a UV transilluminator (LA 1069, Hi-UV™ Duo, HIMEDIA), with images captured using a camera mounted on a hooded system. A negative control was included in each PCR run to ensure the reliability of amplification. A 1 kb DNA ladder (MBT051, Himedia, India) was employed to estimate the size of the *mtcox1* gene amplicons.

Tools for bioinformatics and DNA sequencing

High-quality PCR products were purified using the HiPurA Multi-sample DNA purification kit (MB554, HiMedia). Sequencing was outsourced to M/s. BioKart Pvt. Ltd., Bangalore, India and Eurofins Genomic India Pvt., Ltd. Forward sequences were processed and analysed using Sequencing Analysis Software version 5.2 (Applied Biosystems Inc., CA, USA), the online nucleotide BLAST tool (https://blast.ncbi.nlm.nih.gov), and ORF Finder (https://www.ncbi.nlm.nih.gov) to eliminate low-quality reads and close sequence gaps. Final sequence alignments, incorporating both newly generated and publicly available genomic data, were aligned using MEGA version 11 software. The sequenced were submitted to GenBank and BOLD Systems under ULBNU project on BOLD (https://boldsystems.org). Within BOLD species identification was performed using the Identification Engine, while Barcode Index Number (BINs) were assigned using the BIN system. BOLD analytical tools were further used to generate BIN concordance analysis, BOLD similarity analysis, nucleotide frequency distribution, distance summary analysis, species accumulation curve, and barcode gap analysis. The Distance Summary Analysis was used to estimate overall pairwise genetic distances at different taxonomic levels (within species, genera, and families), whereas the Barcode Gap Analysis compared maximum intraspecific divergence with the nearest-neighbour interspecific divergence to evaluate species discrimination based on COI barcodes. The accumulation curve analysis, run across 20 iterations with the Kimura 2 Parameter model and BOLD's amino acid based HMM aligner. Pairwise genetic distances,

including nearest-neighbour distances and comparisons of intra-and interspecific genetic divergence, were calculated using the Kimura 2-Parameter (K2P) model implemented in BOLD. A Neighbor-Joining (NJ) phylogenetic tree was constructed in MEGA version 11 (Tamura et al., 2021) with 1000 bootstraps replicates to assess clustering Patterns.

RESULTS

Fish species diversity

In this study, a total of 24 fish specimens were sequenced, representing 20 fish species across 7 orders, 11 families, and 19 genera (Table 1, Figure 2). The analysis reveals that fish diversity in the Upper Lake spans several orders, with *cypriniformes* being the most dominant group, comprising numerous entries across multiple families like *cyprinidae* and *nemacheiliidae*. Other represented orders include *cichliformes, gobiiformes, siluriformes, osteoglossiformes, anabantiformes,* and *synbranchiformes*, each included only one or two species which suggested that *cypriniformes* is the most diverse taxonomic group in the sampled dataset, while the other orders show limited species. The bar plot analysis indicates a clear difference in fish species distribution among the three sites in Upper Lake, Bhopal. Vardhman Sunset Park (VSP) shows the highest species distribution (mean = 0.65) has greatest variability. Bairagarh (BR) has an intermediate distribution (mean = 0.40). Shyamla Hills (SH) has the lowest species distribution (mean = 0.20) and the least variability (Figure 4). The amplification of the *mtcox1* gene produced clear amplicons of approximately 600 bp from the examined fish specimens (Figure 3).

Molecular identification and conservation status

The majority of fish species identified based on morphological characteristics exhibited 99% to 100% sequence similarity with corresponding reference sequences in established databases. These fish species were classified as species of least concern by the IUCN conservation status (Table 1). After editing, all sequences ranged from 560 to 640 bp in length. The sequences have been successfully submitted to GenBank, and the corresponding accession numbers are provided in (Table 2). The sequences have also

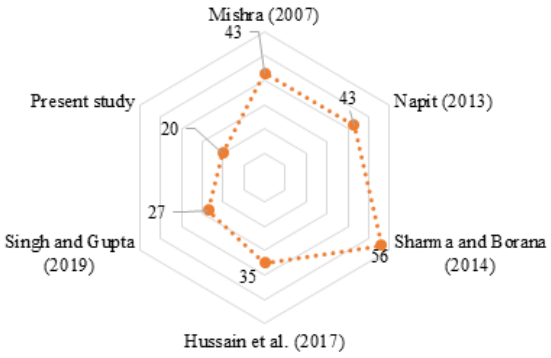


Figure 2: Comparison of fish species recorded in the present study with those reported in previous studies of Upper Lake, a Ramsar wetland in Bhopal District, Madhya Pradesh, India

been submitted on BOLD Systems, along with specimen and collection information, sequence file, trace files, and primer details. The corresponding barcode records are accessible through the ULBNU project on BOLD (https://boldsystems.org).

Barcode index number analysis

The Barcode Index Number (BIN) analysis revealed notable inconsistencies between molecular clusters and traditional taxonomy. The 24 barcode records were assigned to 18 BINs, of which 14 BINs comprising 19 records were taxonomically discordant, whereas 4 BINs comprising 5 records were taxonomically concordant. No singleton BINs were detected in the dataset. These mismatches suggest possible cryptic species, misidentifications, or inaccuracies within reference databases. Discordance occurred across multiple taxonomic ranks. For instance, ULBNU015-25 (*Notopterus notopterus*) clustered in a BIN BOLD: AAF2803 mixed between *Osteoglossiformes* and *Anabantiformes*, while ULBNU004-24 (*Labeo boggut*) BIN (BOLD: AAZ3554) grouped with *Bangana* and *Sinilabeo*. Species-level conflicts were also observed, such as ULBNU013-24 (*Garra mullya*) BIN (BOLD: AAY8310) clustering with *Garra annandalei*. Overall, the results highlight the effectiveness of BINs in assessing taxonomic validity (Table 3).

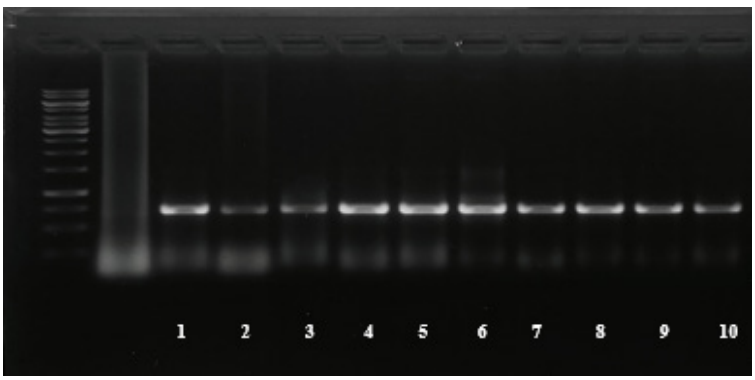


Figure 3: Representative agarose gel electrophoresis image showing PCR amplicons of the mitochondrial *cox1* (*mtcox1*) gene. The PCR amplicons were approximately 600 bp in size and were visualized on a 1% agarose gel alongside a 1 Kb Ladder. Lane 1: 1 kb DNA molecular weight ladder; lanes 2-11: PCR amplified products

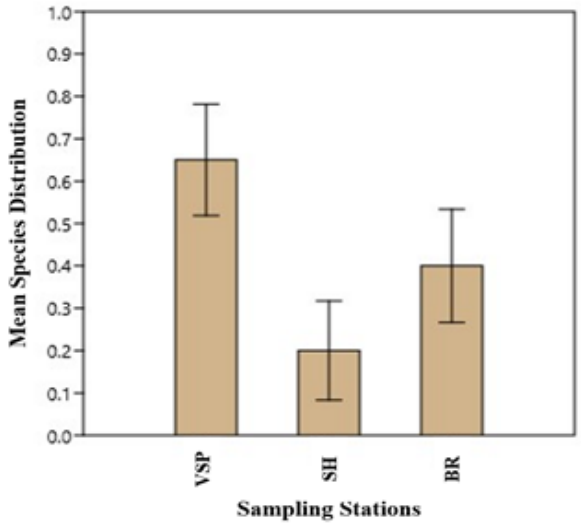


Figure 4: Mean species distribution across sampling stations in Upper Lake, Bhopal. The bar plot shows the mean species distribution at Vardhman Sunset Park (VSP), Shyamla Hills (SH), and Bairagarh (BR)

Distance summary analysis

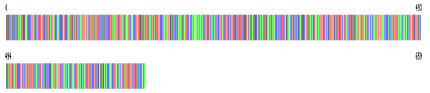
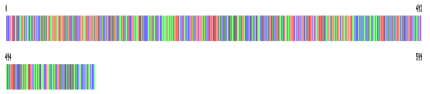
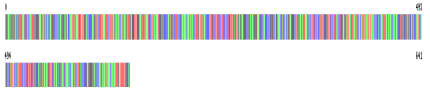
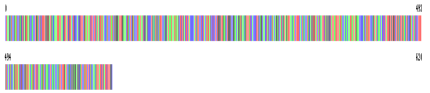
The distance summary using the COI-5P marker demonstrates a clear and pronounced barcode gap, affirming the marker's reliability for species discrimination. At the species level, the mean intraspecific divergence is extremely low 0.58% (range: 0.00% - 1.82%, SE= 0.18). Genetic distances varied across taxonomic levels. Within species, genetic distance ranged from 0.00% to 1.82%, with a mean distance of 0.58% (SE = 0.18%). Within the genus level, the genetic distance was 10.24%, with no variation among the comparisons (SE = 0.00%). At the family level, genetic distances ranged from 0.16% to 22.07%, with a mean distance of 16.98% (SE = 0.05%).

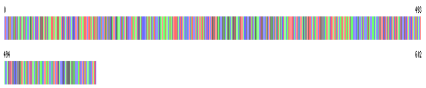
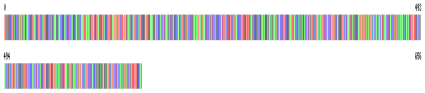
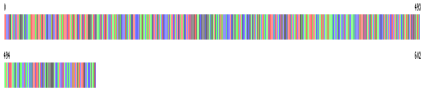
In sharp contrast, the interspecific divergence at the genus level is substantially higher, with a minimum divergence of 10.24% observed (mean = 10.24%, SE = 0.00). This wide separation where the maximum with-species distance (1.82%) is significantly lower than the minimum with-genus distance (10.24%) is a robust indicator of effective species discrimination (Figure 5). Furthermore, the frequency distribution histogram visually confirms this pattern, showing within-species divergences clustered tightly at low values (peak frequency at 75% near 0%), while within-genus divergence forms a discrete peak at 10.24%. This distinct separation is further supported by the normalized divergence histogram, which contrasts the low within-species values (blue bars, typically <1%) with the high between-species divergence (red bar, > 10%). The consistent divergence pattern, despite the limited number of comparisons at the genus level (n= 1), strongly supports the utility of the *mtcox1* gene for reliably differentiating the sampled fish species.

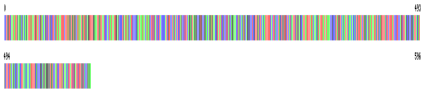
Table 1: Taxonomic classification of fish species recorded from Upper Lake, Bhopal, and their IUCN Red List status

Order	Family	Genus	Species	No. of samples	IUCN status
Cypriniformes	Cyprinidae	*Salmostoma*	*Salmostoma bacaila*	2	LC
	Cyprinidae	*Systomus*	*Systomus sarana*	1	LC
	Cyprinidae	*Garra*	*Garra mullya*	2	LC
	Cyprinidae	*Puntius*	*Puntius terio*	1	LC
	Cyprinidae	*Labeo*	*Labeo boggut*	2	LC
	Cyprinidae	*Osteobrama*	*Osteobrama cotio*	2	LC
	Cyprinidae	*Bangana*	*Bangana dero*	1	LC
	Cyprinidae	*Pethia*	*Pethia ticto*	1	LC
	Cyprinidae	*Puntius*	*Puntius chola*	1	LC
	Nemacheilidae	*Paracanthocobitis*	*Paracanthocobitis botia*	1	LC
Cichliformes	Cichlidae	*Oreochromis*	*Oreochromis niloticus*	1	LC
Gobiiformes	Gobiidae	*Glossogobius*	*Glossogobius giuris*	1	LC
	Ambassidae	*Chanda*	*Chanda nama*	1	LC
Siluriformes	Heteropneustidae	*Heteropneustes*	*Heteropneustes fossilis*	1	LC
	Bagridae	*Mystus*	*Mystus bleekeri*	1	LC
Osteoglossiformes	Notopteridae	*Notopterus*	*Notopterus notopterus*	1	LC
Anabantiformes	Channidae	*Channa*	*Channa punctata*	1	LC
	Nandidae	*Nandus*	*Nandus nandus*	1	LC
Synbranchiformes	Mastacembelidae	*Macrognathus*	*Macrognathus pancalus*	1	LC
	Mastacembelidae	*Mastacembelus*	*Mastacembelus armatus*	1	LC

Table 2: Voucher specimen images and corresponding DNA barcode information of the analysed fish species from Upper Lake, Bhopal. The photographs represent the voucher specimens used for DNA barcoding, whereas the illustrative barcode images represent DNA barcode visualizations generated by the BOLD Systems

S. No.	Specimen name	Voucher No.	GenBank accession no.	Photographs	Illustrative barcode
1.	*Salmostoma bacaila*	ULBR01A	PP812675		
2.	*Salmostoma bacaila*	ULVSP18	PP812676		
3.	*Systomus sarana*	ULBR03A	PP754214		
4.	*Orechromis niloticus*	ULBR10A	PP754242		
5.	*Pethia ticto*	ULVSP15	PP754350		
6.	*Glossogobius giuris*	ULVSP7	PP754341		
7.	*Gara mullya*	ULVSP17	PP754351		

8.	*Puntius terio*	ULSH16	PP809691		
9.	*Heteropneustes fossilis*	ULBR06A	PP754237		
10.	*Labeo boggut*	ULBR07A	PP812677		
11.	*Labeo boggut*	ULVSP16	PP812686		
12.	*Osteobrama cotio*	ULSH11	PP812679		
13.	*Osteobrama cotio*	ULSH13	PP812680		

14.	*Bangana dero*	ULSH21	PP812685		
15.	*Notopterus notopterus*	ULBR48	PV533813		
16.	*Chanda nama*	ULBR64	PV549597		
17.	*Mystus bleekeri*	ULVSP22	PX116877		
18.	*Channa punctata*	ULVSP24	PX116897		
19.	*Macrognathus pancalus*	ULVSP25	PX129264		

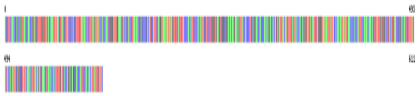
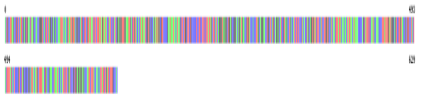
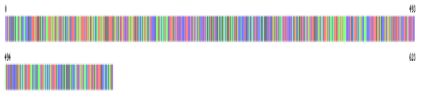
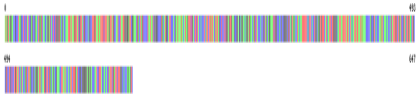
20.	*Garra mullya*	ULVSP26	PX130757		
21.	*Paracanthocobitis botia*	ULVSP27	PX129529		
22.	*Nandus nandus*	ULVSP31	PX130404		
23.	*Mastacembelus armatus*	ULBR71	PX129523		
24.	*Puntius chola*	ULBR74	PX129300		

Table 3: Detailed BIN discordance by taxonomic ranked and conflicting taxon

Process ID	Identification	Conflicting Taxon in BIN	Rank of Conflict	BIN	BIN Total Members	BIN Tax Variation
ULBNU015-25	*Notopterus notopterus*	*Osteoglossiformes*	Order	BOLD: AAF2803	304	*Osteoglossiformes* (294), *Anabantiformes* (1)
ULBNU019-25	*Macrognathus pancalus*	*Synbranchiformes*	Order	BOLD: AAF5455	157	*Synbranchiformes* (144), *Cypriniformes* (3)
ULBNU003-24	*Heteropneustes fossilis*	*Siluriformes*	Order	BOLD: ACR4875	162	*Siluriformes* (144), *Clupeiformes* (2)
ULBNU005-24	*Oreochromis niloticus*	*Cichliformes*	Order	BOLD: AAA8511	559	*Cichliformes* (557), *Cypriniformes* (2), *Clupeiformes* (1)
ULBNU022-25	*Nandus nandus*	*Anabantiformes*	Order	BOLD: AAZ8464	115	*Anabantiformes* (120), *Siluriformes* (2)
ULBNU010-24	*Glossogobius giuris*	*Gobiidae*	Family	BOLD: ABA9422	292	*Gobiida* (280), *Oxudercidae* (5)
ULBNU009-24	*Bangana dero*	*Bangana*	Genus	BOLD: AAZ3554	297	*Labeo* (232), *Bangana* (28), *Sinilabeo* (10)
ULBNU004-24	*Labeo boggut*	*Labeo*				
ULBNU012-24	*Labeo boggut*	*Labeo*				
ULBNU001-24	*Salmostoma bacaila*	*Salmostoma*	Genus	BOLD: ACA0489	70	*Salmostoma* (41), *Cabdio* (10), *Salmophasia* (1)
ULBNU014-24	*Salmostoma bacaila*	*Salmostoma*				
ULBNU021-25	*Paracanthocobitis botia*	*Paracanthocobitis*	Genus	BOLD: AAY8322	121	*Paracanthocobitis* (64), *Acanthocobitis* (48)
ULBNU024-25	*Puntius chola*	*Puntius*	Genus	BOLD: AAX7390	223	*Puntius* (193), *Pethia* (7)
ULBNU011-24	*Pethia ticto*	*Pethia*	Genus	BOLD: AAZ3076	101	*Pethia* (62), *Puntius* (38), *Amblypharyngondon* (1)
ULBNU008-24	*Puntius terio*	*Puntius*				
ULBNU013-24	*Garra mullya*	*Garra mullya*	Species	BOLD: AAY8310	120	*Garra mullya* (79), *Garra Annandale* (21), *Garr asp.* 1 KKL2013 (20)
ULBNU020-25	*Garra mullya*	*Garra mullya*				
ULBNU002-24	*Systomus sarana*	*Systomussarana*	Species	BOLD: AAY5233	194	*Systomus sarana* (156), *Systomus spilurus* (5), *Systomus cf. sarana* (1)
ULBNU017-25	*Mystus bleekeri*	*Mystus bleekeri*	Species	BOLD: AAZ9696	139	*Mystus bleekeri* (124), *Mystus armatus* (6)

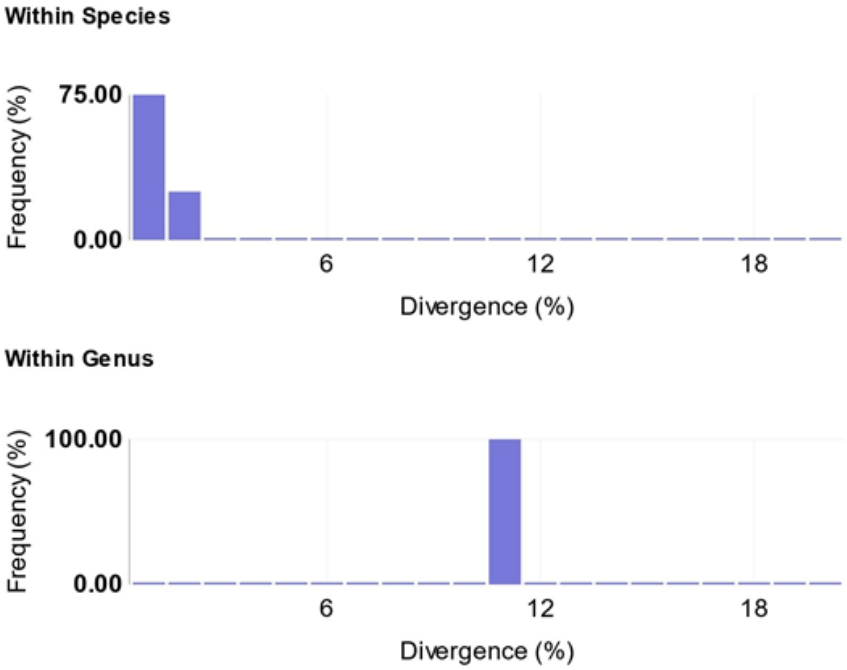


Figure 5: Comparison of pairwise COI-5P sequence divergence at the species and genus levels. Frequency distribution of Kimura 2-Parameter (K2P) pairwise genetic distances comparing intraspecific divergence with the minimum interspecific divergence at the genus level, demonstrating a clear barcode gap among the 24 analysed sequences

Nucleotide frequency distribution analysis

The statistics for nucleotide frequency distribution are presented in the (Table 4). The nucleotide composition analysis of 24 sequences revealed a variable distribution of base frequencies. The mean values for individual nucleotides were as follows: Guanine (G%) at 17.56%, Cytosine (C%) at 27.82%, Adenine (A%) at 25.90%, and Thymine (T%) at 28.72%. The overall GC content across sequences averaged 45.38%, with values ranging from 42.21% - 47.69%. Codon position specific GC distribution demonstrated clear positional bias. GC content was highest at the 1st codon position (mean = 56.25%), moderate at the 2nd position (GC = 42.58%), and lowest at the 3rd position (GC = 37.32%). This pattern aligns with the expected functional constraints of protein coding mitochondrial genes, where codon position 1 is most conserved, while codon 3 is more variable (SE = 0.8427).

***mtcox1* barcode length**

The distribution of sequence length of fish sample (Figure 6) where the sequences predominantly range from 560 to 640 bp (base pairs), which aligns with the expected length for the standard *mtcox1* (mitochondrial cytochrome-c oxidase subunit I) barcode region commonly used in fish species identification. The distribution around this range indicates high-quality, full length barcode sequences with minimal truncation or degradation. This reliable and uniform dataset provides a robust foundation for accurate taxonomic classification, biodiversity assessments, and phylogenetic analyses.

Accumulation curve using all data in valid groups

The accumulation curve based on 24 analysed records using the COI-5P marker illustrates the progressive build-up of taxonomic and sequence diversity across samples (Figure 7). The curve reveals that haplotype diversity increased more sharply than genus, species, and taxon levels, pointing to significant genetic variation within species. Meanwhile, the genus, species, and taxon curves followed a steady and parallel pattern with slightly lower values, suggesting that most taxonomic groups were consistently represented.

Barcode gap analysis

Barcode gap analysis of the COI-5P dataset (K2P distance, pairwise deletion; BOLD amino-acid HMM aligner; sequences ≥200 bp) shows (Figure 8) clear separation between intra- and interspecific variation. In both the Max intra-specific vs. Nearest Neighbour and Mean intra-specific vs. Nearest Neighbour plots, all data points lie well above the 1:1 line, indicating that nearest-neighbour (between-species) distances consistently exceed within-species divergence. Intraspecific distance is very low for the sampled taxa (generally near 0 and not exceeding 2%), whereas nearest neighbour distances are much larger (typically >10%), evidencing a strong barcode gap and supporting accurate species discrimination with COI-5P. The Individuals-per-species vs Max intra-specific plot indicates that most taxa are represented by one or two individuals, resulting in very low within-species genetic divergence, only a few species exhibit slightly higher values, which nonetheless remain well below their nearest-neighbour distances.

Table 4: Nucleotide composition of the analysed sequences, showing GC and AT percentages with minimum, mean, maximum, and standard error (SE) values

Sl. No.	GC (%), AT (%)	Minimum	Mean	Maximum	SE
1	G%	15.72	17.56	19.03	0.1657
2	C%	24.96	27.82	30.53	0.2687
3	A%	22.76	25.90	28.18	0.2532
4	T%	27.00	28.72	30.49	0.2175
5	GC%	42.21	45.38	47.69	0.2916
6	GC% Codon Pos 1	52.80	56.25	58.56	0.3249
7	GC% Codon Pos 2	41.21	42.58	44.09	0.1406
8	GC% Codon Pos 3	30.00	37.32	47.52	0.8427

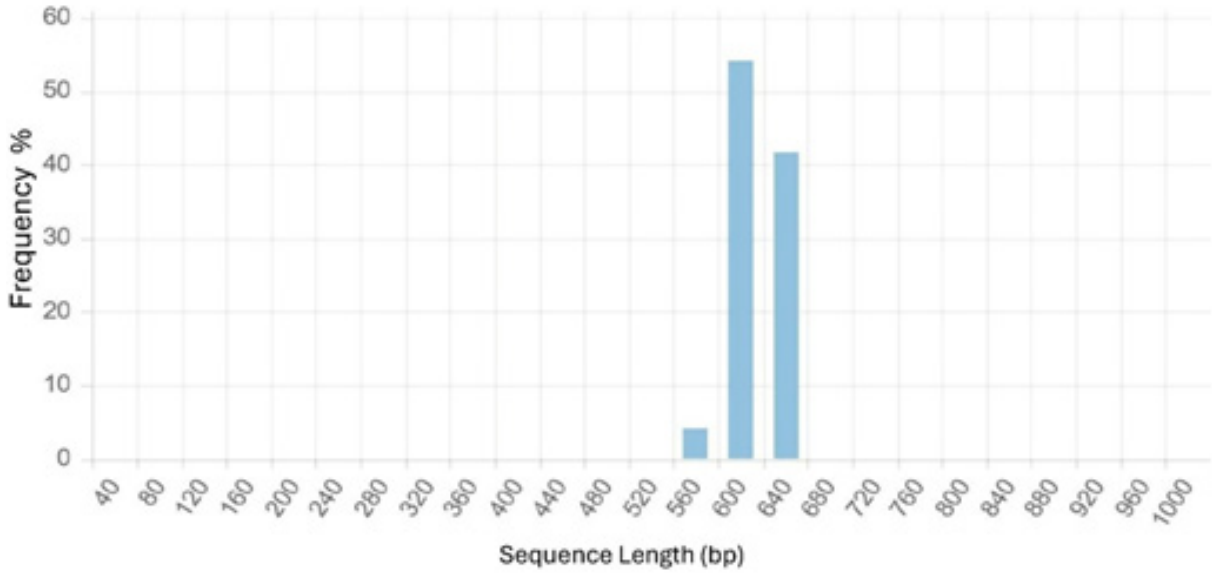


Figure 6: Distribution of *mtcox1* gene sequence lengths among the analysed fish specimens. Frequency distribution of *mtcox1* sequence read length (bp) obtained from the 24 analysed fish specimens collected from Upper Lake, Bhopal

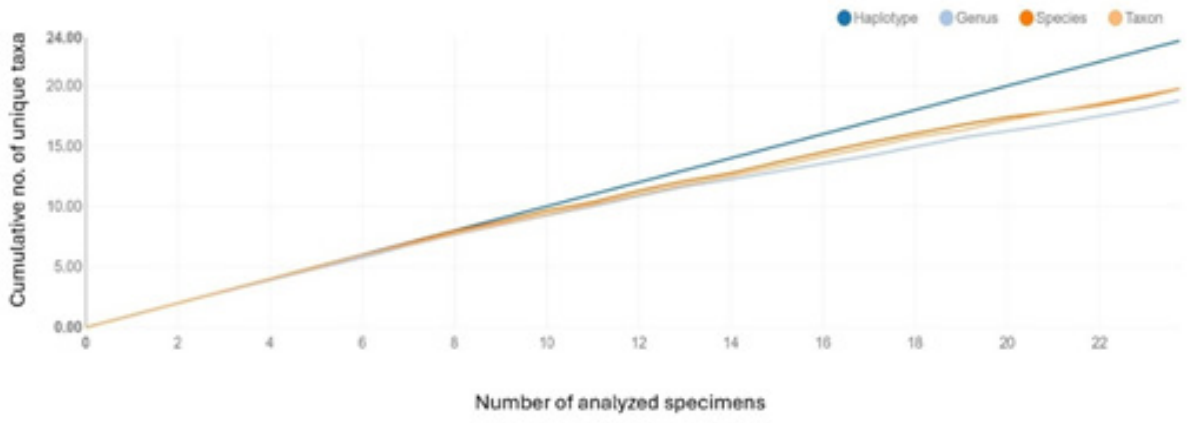


Figure 7: Cumulative accumulation of unique haplotypes, species, genera, and taxa among the analysed fish samples from Upper Lake, Bhopal

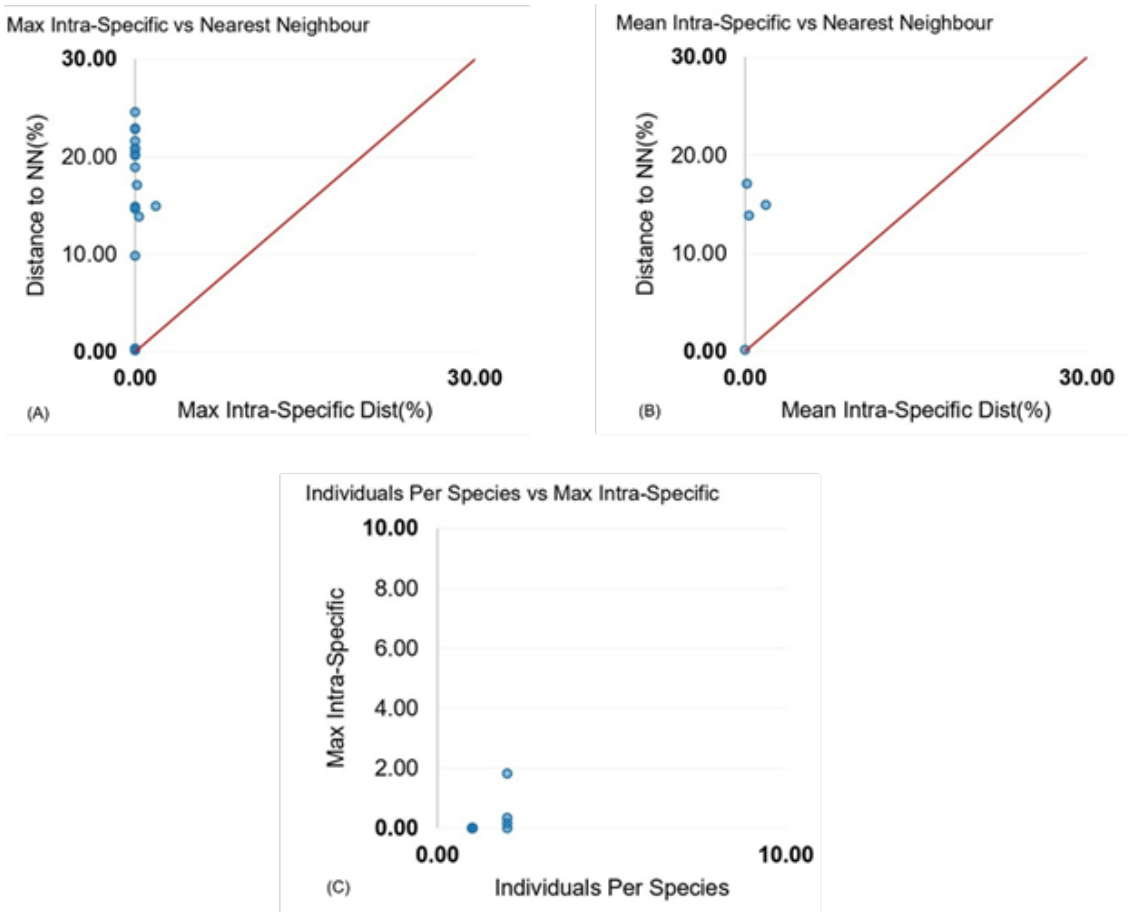


Figure 8: Scatterplots assessing the existence and magnitude of the DNA barcode gap among studied fish species. (A) Maximum intraspecific divergence versus distance to the nearest neighbour (NN); (B) Mean intraspecific divergence versus distance to the Nearest neighbour and (C) Number of individuals sampled per species versus maximum intraspecific divergence (%), showing the relationship between sampling effort and within-species genetic variation

Genetic divergence of barcode gap

The histogram analysis (Figure 9) of genetic divergences revealed a clear barcode gap in the examined fish samples. The mean intra-specific divergence was extremely low (0.12%), indicating a high genetic similarity among individuals of the same species. In contrast, the mean inter-specific divergence, measured as the distance to the nearest neighbour, was markedly higher (14.7%), with values ranging from 0.16% to 24.6%. This sharp contrast between intra- and inter-specific distances supports the efficiency of the COI gene as a reliable molecular marker for species discrimination in the studied taxa. The between-species mean is therefore 122.5-fold greater than the within-species mean (14.7 ÷ 0.12 = 122.5), which is strong numerical evidence of a pronounced barcode gap in this dataset. The presence of a distinct barcode gap confirms that most species can be accurately identified and differentiated using DNA barcoding.

The species-by-species comparison (Table S1) shows that, for most taxa, the nearest-neighbour (inter-specific) K2P distances (Kimura two-parameter distance) far exceed intra-specific divergence, supporting a clear barcode gap across the dataset. Many species are singletons in our sample (displayed as N/A for mean intra-specific values are not available, they are generally low (e.g. *Osteobrama cotio* max = 0.34%, *Salmostoma bacaila* max intra = 0.16%). At the other extreme, several nearest-neighbour distances are exceptionally high (e.g. *Glossogobius giuris* NN= 24.6%, *Notopterus notopterus* NN= 22.84%, *Chanda nama* NN = 22.94%), demonstrating that these taxa are clearly separated from their closest congeners in COI space and are reliably discriminated by the *mtcox1* marker. A small number of cases require closer scrutiny.

Two taxa *Bangana dero* and *Labeo boggut* show an exceptionally low interspecific distance to each other (NN= 0.16%), and *Pethia ticto* and *Puntius terio* also have low nearest-neighbour distances (0.33%). Such NN values below the 2% threshold (and approaching observed intra-specific maxima). *Garra mullya* displays the highest observed intra-specific divergence in table 5 (mean and max = 1.82%), which while still lower than its NN (=14.95%) may reflect population structure or cryptic diversity within that taxon. Overall, the table confirms a pronounced barcode gap for the majority of sampled species and therefore supports the utility of COI-5P for species discrimination in Upper Lake.

Table 5: Comparison of intra- and interspecific genetic divergence among the analysed fish species, showing mean and maximum intraspecific divergence and corresponding nearest-neighbour interspecific genetic distances based on the COI-5P dataset

Order	Family	Species	Mean Intra-Sp	Max Intra-Sp	Nearest Neighbour	Nearest Species	Distance to NN
Anabantiformes	Channidae	*Channa punctata*	N/A	0	ULBNU005-24	*Oreochromis niloticus*	20.89
Anabantiformes	Nandidae	*Nandus nandus*	N/A	0	ULBNU019-25	*Macrognathus pancalus*	21.63
Cichliformes	Cichlidae	*Oreochromis niloticus*	N/A	0	ULBNU023-25	*Mastacembelus armatus*	20.74
Cypriniformes	Cyprinidae	*Bangana dero*	N/A	0	ULBNU012-24	*Labeo boggut*	0.16
Cypriniformes	Cyprinidae	*Garra mullya*	1.82	1.82	ULBNU002-24	*Systomus sarana*	14.95
Cypriniformes	Cyprinidae	*Labeo boggut*	0	0	ULBNU009-24	*Bangana dero*	0.16
Cypriniformes	Cyprinidae	*Osteobrama cotio*	0.34	0.34	ULBNU024-25	*Puntius chola*	13.87
Cypriniformes	Cyprinidae	*Pethia ticto*	N/A	0	ULBNU008-24	*Puntius terio*	0.33
Cypriniformes	Cyprinidae	*Puntius chola*	N/A	0	ULBNU011-24	*Pethia ticto*	9.85
Cypriniformes	Cyprinidae	*Puntius terio*	N/A	0	ULBNU011-24	*Pethia ticto*	0.33
Cypriniformes	Cyprinidae	*Salmostoma bacaila*	0.16	0.16	ULBNU002-24	*Systomus sarana*	17.12
Cypriniformes	Cyprinidae	*Systomus sarana*	N/A	0	ULBNU004-24	*Labeo boggut*	14.9
Cypriniformes	Nemacheilidae	*Paracanthocobitis botia*	N/A	0	ULBNU002-24	*Systomus sarana*	20.3
Gobiiformes	Ambassidae	*Chanda nama*	N/A	0	ULBNU002-24	*Systomus sarana*	22.94
Gobiiformes	Gobiidae	*Glossogobius giuris*	N/A	0	ULBNU002-24	*Systomus sarana*	24.6
Osteoglossiformes	Notopteridae	*Notopterus notopterus*	N/A	0	ULBNU024-25	*Puntius chola*	22.84
Siluriformes	Bagridae	*Mystus bleekeri*	N/A	0	ULBNU003-24	*Heteropneustes fossilis*	20.14
Siluriformes	Heteropneustidae	*Heteropneustes fossilis*	N/A	0	ULBNU002-24	*Systomus sarana*	18.94
Synbranchiformes	Mastacembelidae	*Mastacembelus pancalus*	N/A	0	ULBNU023-25	*Mastacembelus armatus*	14.68
Synbranchiformes	Mastacembelidae	*Mastacembelus armatus*	N/A	0	ULBNU019-25	*Mastacembelus pancalus*	14.68

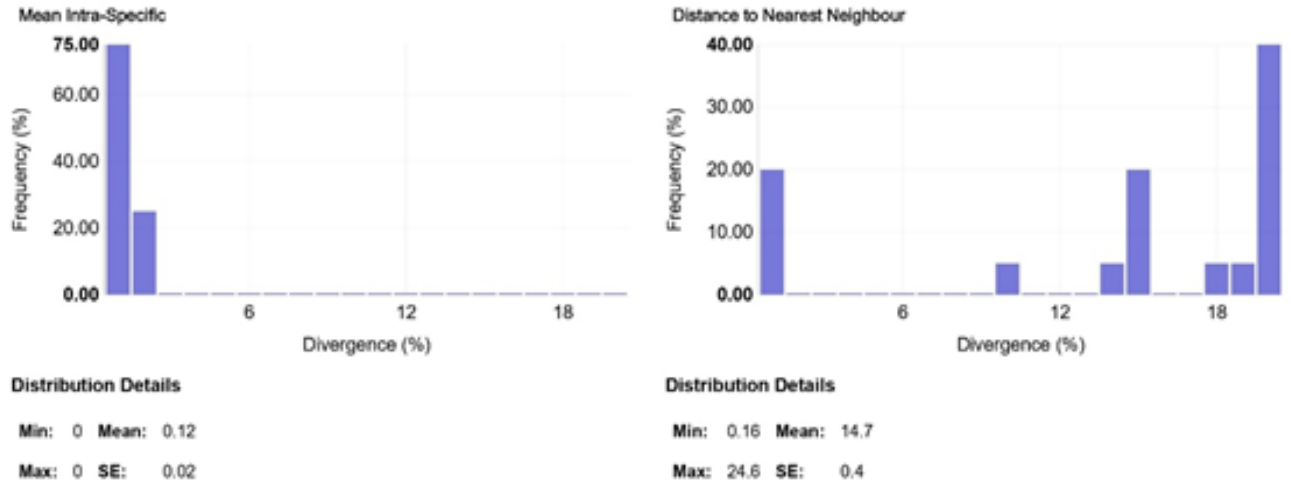


Figure 9: Histograms showing the distribution of genetic distances among the fish species analysed in the study

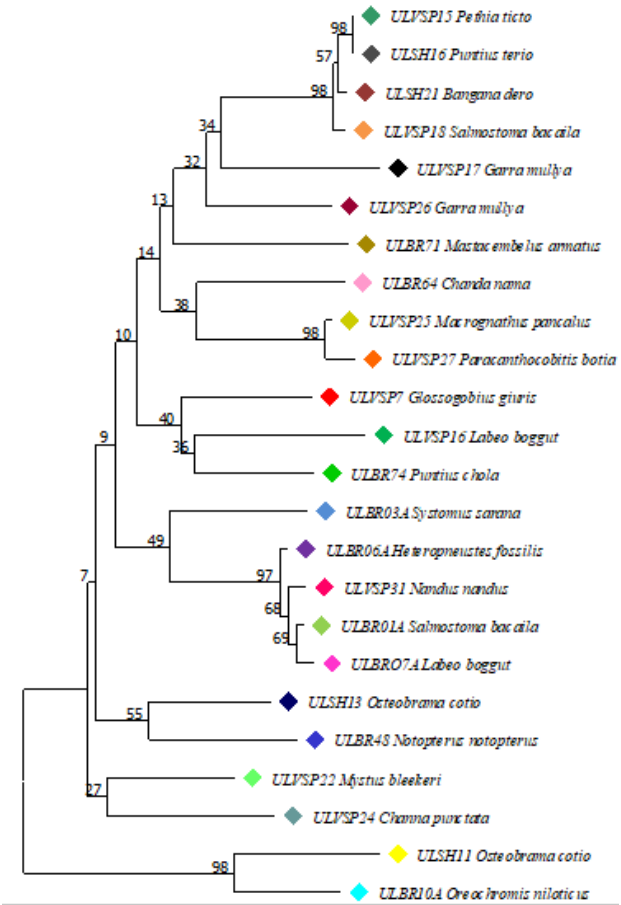


Figure 10: Neighbor- Joining (NJ) phylogenetic tree based on *mtcox1* sequences of the analysed fish species. The evolutionary distance was computed using the Maximum Composite Likelihood (MCL) method and the robustness of the inferred tree topology was assessed using 1000 bootstrap replicates

BOLD similarity analysis

The similarity scores of the top-hit matches showed 99.00% identity with entries in the Barcode of Life Data System (Figure S1). This clearly indicates that the present investigation successfully barcoded the DNA sequences of 24 fish samples and effectively identify the species using the developed dataset.

Neighbor-joining phylogenetic analysis

The evolutionary history was reconstructed using the Neighbor-Joining method, and the optima tree is presented in Figure 10. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown below the branches. The tree is drawn to scale, with branch lengths corresponding to the evolutionary distances used to infer the phylogeny. Evolutionary distances were calculated using the Maximum Composite Likelihood method and are expressed as the number of base substitutions per site. This analysis included 24 nucleotide sequences, with codon positions 1st+2nd+3rd, and Noncoding regions considered. All ambiguous positions were removed for each sequence pair using the pairwise deletion option. The final dataset comprised 670 positions.

DISCUSSION

DNA barcoding based on the COI gene has emerged as an important approach for species identification, drawing increasing interest, particularly in the study of fish species (McCusker et al., 2013; Knebelsberger et al., 2014). The ability to identify species quickly and correctly is central to ecological conservation initiatives and their overall success (Rasmussen & Morrissey, 2009). Lakra et al. (2010) emphasized that cytochrome c oxidase I based DNA barcoding serves a reliable approach for accurate identification of freshwater fishes in India. However, since the COI gene is maternally inherited, interspecific hybridization or endosymbiont infections may affect barcoding accuracy, highlighting limitations of relying solely on COI (Wu et al., 2018). Khedkar et al. (2014) examined ichthyofaunal diversity in the Narmada River employing the COI gene for species characterization. The Upper Lake (Bhojtal) in Bhopal, designated as a Ramsar site, holds considerable ecological importance (Vyas &

Veerwal, 2014).

Early studies by Mishra (2007) recorded 43 species, a trend also supported by later morphological surveys (Tamot et al., 2010). Notably Napit (2013) also reported 43 species, dominated by Cypriniformes. Hussain et al. (2017) observed 35 species but noted a strong dominance of Cypriniformes. The latest survey by Singh & Gupta (2019) recorded just 27 species (Figure 3). This sharp decline from earlier records highlights potential biodiversity loss. These historical findings stem from reliance on traditional morphological methods. In this study we generated 24 *mtcox1* sequences from 20 fish species, provides a high-quality genetic data. Similarly, Modeel et al. (2024) worked on Beas River fishes and Suryawanshi et al. (2025) analysed fish species from the Hirakud Dam, while Habib et al. (2021) utilised both COI and 16S rRNA markers in the Sundarbans Ramsar site to develop an integrated molecular morphometric reference database. Additionally, the COI gene shows a low rate of amino acid change (Lynch & Jarrell, 1993) and is highly conserved at the DNA level, particularly among individuals within the same species (Lambert et al., 2005).

Using BIN analysis, Chen et al. (2015) recognised 43 provisional species. Suryawanshi et al. (2024) found 83 BINs among 226 specimens, of which 73.49% matched their taxonomic assignments, while 26.50% were singletons, and no discordant BINs were detected. In this study the Barcode Index Number (BIN) revealing a significant discordance rate where 14 out of 18 generated BINs were inconsistent with the assigned species names. The conflicts span various taxonomic levels, evidenced by the unexpected clustering of *Notopterus notopterus* (*osteoglossiformes*) with sequences from a distant order *(anabantiformes*), and the grouping of *Labeo boggut* with different genera (*Bangana* and *Sinilabeo*). Furthermore, species-level conflicts, such as *Garra mullya* sharing a BIN with *Garra annandalei*, suggest that these taxa may be conspecific or require closer scrutiny.

Hebert et al. (2004) introduced the "10x rule," suggesting that genetic divergence between species is typically over ten times higher than the variation observed within a species. Intraspecific divergence is extremely low, averaging only 0.58% and peaking at 1.82%. In sharp contrast, the interspecific divergence at the genus level starts at a minimum of 10.24%. This non-overlapping range, where the maximum within-species distance (1.82%) is far exceeded by the minimum within-genus distance (10.24%), is visually confirmed by the divergence histograms. The within- species values are tightly clustered near 0% (pea frequency 75%), while between-species divergence forms a discrete, distant peak.

The sum of A and T bases (25.90% + 28.72% = 54.62%) exceeds the sum of G and C bases (17.56% + 27.82% = 45.38%). The average GC content of 45.38% confirms the A-T richness and a pronounced codon usage bias driven by the functional requirements of protein synthesis. This pattern reflects an AT-rich composition relative to GC, indicating a distinct base bias influenced by selective pressures (Bingpeng et al., 2018).

The most striking observation is the steeper accumulation of haplotype diversity compared to the curves for genus, species, and overall taxon levels. This pattern is a key indicator of significant intraspecific genetic variation within the sampled populations. The near-linear growth across all curves at the end of the analysis further confirms that the sampling efforts was adequate for capturing the fish diversity of Upper Lake.

In most species, K2P-estimated intraspecific divergence trends to be under 2% (Hebert et al., 2003b). The data show an extremely low mean intraspecific divergence (0.12%), reflecting minimal genetic variation among individuals of the species, in stark contrast to the markedly higher mean interspecific divergence (14.7%) to the nearest neighbor. The difference is highly significant, with the between-species mean being 122.5-fold greater than the within-species mean (14.7% ÷ 0.12% = 122.5), which clearly delineates species boundaries. This sharp separation between the two divergence classes, ranging from 0.16% to 24.6% for interspecific distance, confirms the reliability of the *mtcox1* gene.

Barcode gap analysis indicated that the nearest-neighbor (inter-specific) distances were markedly greater than the within-species divergences. The interspecific divergence is observed to be very low (mostly near 0% and not exceeding 2%), indicating a recent common ancestry for individuals within the same species. Conversely, nearest-neighbor distances are typically much larger (often greater than 10%), which established a strong barcode gap.

DNA barcoding offers a precise and universally comparable molecular record of fish species (Imtiaz et al., 2017). This approach is commonly used to assess the reliability of DNA barcode-based species identification (VanVelzen et al., 2012). Garg et al. (2017) reported a 99% similarity match for *Puntius ticto* using BOLD data. In this study, the *mtcox1* sequence showed a Top Hit of 100% similarity in BOLD, indicating strong conservation and a perfect match with validated reference barcodes. Most top matches showed ≥99.84% similarity, all corresponding to the same species and taxonomic hierarchy. This tight clustering strongly confirms the accurate identification of the specimen and its placement within a well-defined BOLD cluster.

The NJ-MCL method uses the Maximum Composite Likelihood (MCL) method to determine all evolutionary distances, and then the Neighbor-Joining (NJ) algorithm uses those distances to infer the phylogeny (Tamura et al. 2004). The inference of the evolutionary history, as depicted in the Neighbor-Joining (NJ) tree (Figure 10), offers a robust, distance-based estimation of the phylogenetic relationships among the 24 nucleotide sequences. The numbers below the branches represent bootstrap support values based on 1000 replicates. The evolutionary distances were meticulously calculated using

the Maximum Composite Likelihood (MCL) method. The inclusion of all codon positions (1st+2nd+3rd) and Noncoding regions maximize the informational content of the alignment, using a total of 670 positions in the dataset. Ward et al. (2005), who demonstrated that COI based NJ analysis effectively separates fish species into well-defined genetic groups. Comparable results were also observed by Lakra et al. (2010) in Indian fishes, where conspecific individuals formed closely related clusters with strong bootstrap support, confirming the utility of COI barcodes for reliable species identification.

CONCLUSIONS

Numerous studies have shown that DNA barcoding serves as an effective complementary approach for precise species identification and biodiversity assessment. The results validated the effectiveness of DNA barcoding for accurate identification of fish species in Upper Lake, Bhopal. In addition, this study updates and strengthens the ichthyofaunal record of Upper Lake, Bhopal. The DNA sequences submitted to BOLD Systems and GenBank (NCBI) are expected to help other researchers and policy makers correctly identify the fish species in Upper Lake, Bhopal. In conclusion, the integration of molecular with morphological approaches facilitates precise identification of fish species.

ACKNOWLEDGEMENT

Special thanks to Head of the Department of Bioscience from Barkatullah University for logistic support, laboratory availability at the Department. We are thankful to the local fishermen for their support during field sampling. We acknowledge the financial support for this work from the National Fellowship under the National Fellowship & Scholarship for Higher Education of ST Students (NFST) scheme.

DECLARATION OF COMPETING INTEREST

The authors declare that there are no conflicts of interest

FUNDING INFORMATION

This research was financially supported by the National Fellowship under the National Fellowship & Scholarship for Higher Education of ST Students (NFST) scheme, Ministry of Trible Affairs, Government of India (grant number 202425-NFST-MAD-02611).

AUTHOR CONTRIBUTION

Conceptualization, Data curation, Formal analysis, Investigation, methodology, Visualization, Writing the original draft: **N.U.**; Supervision, Manuscript editing and review: **S.Y.**; Conceptualization, Supervision, Manuscript editing and review: **R.K.G.**

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article.

REFERENCES

April, J., Mayden, R. L., Hanner, R. H., & Bernatchez, L. (2011). Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences*, **108**(26), 10602-10607. https://doi.org/10.1073/pnas.1016437108

Bhattacharya, M., Sharma, A. R., Patra, B. C., Sharma, G., Seo, E. M., Nam, J. S., Chakraborty, C., & Lee, S. S. (2016). DNA barcoding to fishes: current status and future directions. *Mitochondrial DNA Part A*, **27**(4), 2744-2752. https://doi.org/10.3109/19401736.2015.1046175

Bingpeng, X., Heshan, L., Zhilan, Z., Chunguang, W., Yanguo, W., & Jianjun, W. (2018). DNA barcoding for identification of fish species in the Taiwan Strait. *Public Library of Science one*, **13**(6), e0198109. https://doi.org/10.1371/journal.pone.0198109

Chen, W., Ma, X., Shen, Y., Mao, Y., & He, S. (2015). The fish diversity in the upper reaches of the Salween River, Nujiang River, revealed by DNA barcoding. *Scientific Reports*, **5**(1), 17437. https://doi.org/10.1038/srep17437

Cohen, M. J., Creed, I. F., Alexander, L., Basu, N. B., Calhoun, A. J., Craft, C., & Walls, S. C. (2016). Do geographically isolated wetlands influence landscape functions?. *Proceedings of the National Academy of Sciences*, **113**(8), 1978-1986. https://doi.org/10.1073/pnas.1512650113

DeSalle, R., & Goldstein, P. (2019). Review and interpretation of trends in DNA barcoding. *Frontiers in Ecology and Evolution*, **7,** 302. https://doi.org/10.3389/fevo.2019.00302

Eschmeyer, W. N., Fricke, R., Fong, J. D., & Polack, D. A. (2010). Marine fish diversity: history of knowledge and discovery (Pisces). *Zootaxa*, **2525**(1), 19-50.

Garg, R. K., Dubey, N., Batav, N., Pandey, P., & Singh, R. K. (2017). Mitochondrial COI gene sequence analyses of puntius ticto and compared with seven species of genus puntius of family cyprinidae: A finding for phylogenetic positioning and DNA barcoding as model study for cryptic species. *Journal of Proteomics Bioinformatics*, **10**, 214-221. https://doi.org/10.4172/jpb.1000445

Gopalji, S. (1998). Fishes of UP and Bihar. *Vishwavidalaya Prakashan Chowk, Varanasi, India*.

Habib, K. A., Neogi, A. K., Rahman, M., Oh, J., Lee, Y. H., & Kim, C. G. (2021). DNA barcoding of brackish and marine water fishes and shellfishes of Sundarbans, the world's largest mangrove ecosystem. *Plos one*, **16**(8), e0255110. https://doi.org/10.1371/journal.pone.0255110

Haldar, C., & Nath, S. (2020). DNA barcoding for fish species identification: current status and future prospective. *International Journal of Fauna and*

Biological Studies, 7, 72-75.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**(1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

Hebert PD, Ratnasingham S, DeWaard JR. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270** (1), 96-99.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *Public Library of Science Biology* **2**(10), e312. https://doi.org/10.1371/journal.pbio.0020312

Hussain, S., Ali, H., & Manohar, S. (2017). Ichthyofaunal diversity of freshwater lake: a case study of Upper Lake Bhopal (MP). *International Journal of Academic and Research for Multidisciplinary*, **5**(2), 70-75.

Imtiaz, A., nor, S. A. M., & Naim, D. M. (2017). Progress and potential of DNA barcoding for species identification of fish species. *Biodiversitas Journal of Biological Diversity*, **18**(4), 1394-1405. https://doi.org/10.13057/biodiv/d180415

Jhingran V.G. (1991). *Fish and Fisheries of India*. Hindustan Publishing Corporation, Delhi, India, 727 pp.

Kar, D., Kumar, C. B., Bohra, C., & Singh, L. K. (2003). Fishes of Barak drainage, mizoram and Tripura. *Environment, pollution and management, APH publishing corporation, New Delhi*, 604, 203-211.

Khedkar, G. D., Jamdade, R., Naik, S., David, L., & Haymer, D. (2014). DNA barcodes for the fishes of the Narmada, one of India's longest rivers. *Public Library of Science One*, ***9***(7), e101460. https://doi.org/10.1371/journal.pone.0101460

Knebelsberger, T., Landi, M., Neumann, H., Kloppmann, M., Sell, A. F., Campbell, P. D., Laakmann, S., Raupach, M.J. Carvalho, G.R. & Costa, F. O. (2014). A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. *Molecular Ecology Resources*, **14**(5), 1060-1071. https://doi.org/10.1111/1755-0998.12238

Kundu, S., Kumar, V., Laskar, B. A., Chandra, K., & Tyagi, K. (2016). Mitochondrial DNA effectively detects non-native Testudines: Invisible wildlife trade in northeast India. *Gene Reports*, **4**, 10-15. https://doi.org/10.1016/j.genrep.2016.02.002

Lakra, W. S., Sarkar, U. K., Kumar, R. S., Pandey, A., Dubey, V. K., & Gusain, O. P. (2010). Fish diversity, habitat ecology and their conservation and management issues of a tropical River in Ganga basin, India. *The Environmentalist*, **30**(4), 306-319. https://doi.org/10.1007/s10669-010-9277-6

Lambert, D. M., Baker, A., Huynen, L., Haddrath, O., Hebert, P. D. N., & Millar, C. D. (2005). Is a large-scale DNA-based inventory of ancient life possible?. *Journal of Heredity*, **96**(3), 279-284. https://doi.org/10.1093/jhered/esi035

Lynch, M., & Jarrell, P. (1993). A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics*, **135**(4), 1197-1208. https://doi.org/10.1093/genetics/135.4.1197

McCusker, M. R., Denti, D., Van Guelpen, L., Kenchington, E., & Bentzen, P. (2013). Barcoding Atlantic Canada's commonly encountered marine fishes. *Molecular Ecology Resources*, **13**(2), 177-188. https://doi.org/10.1111/1755-0998.12043

Mittermeier, R. A., and Goettsch Mittermeier, C. (1997). Megadiversity: Earth's biologically wealthiest nations (1st English ed.). CEMEX.

Mishra, S. M. (2007). Conservation and Management of Bhoj Wetland of Bhopal, India. *World Lake Vision Action Report (ILEC)*, 168-180.

Modeel, S., Negi, R. K., Sharma, M., Dolkar, P., Yadav, S., Siwach, S., Yadav P. & Negi, T. (2024). A comprehensive DNA barcoding of Indian freshwater fishes of the Indus River system, Beas. *Scientific Reports*, **14**(1), 2763. https://doi.org/10.1038/s41598-024-52519-0

Napit, M. K. (2013). Threatended Ichthyofauna of the upper lake Bhopal (MP). *International Journal for Life Sciences and Educational Research*, **1** (1), 43-46

Oishi, T. (2016). Ethnoecology and ethnomedicinal use of fish among the Bakwele of southeastern Cameroon. *Journal of Ethnoecology,* (10). https://doi.org/10.4000/ethnoecologie.2893

Rasmussen, R. S., & Morrissey, M. T. (2009). Application of DNA-based methods to identify fish and seafood substitution on the commercial market. *Comprehensive Reviews in Food Science and Food Safety*, **8**(2), 118-154. https://doi.org/10.1111/j.1541-4337.2009.00073.x

Richardson, C. J., Bruland, G. L., Hanchey, M. F., & Sutton-Grier, A. E. (2016). Soil restoration: the foundation of successful wetland reestablishment. *Wetland Soils Genesis Hydrology, Landscapes and Classification*, 469.

Sambrook, J., and Russell, D. W. (2001). Detection of DNA in agarose gels. *Molecular cloning, a laboratory manual*, 5-14.

Sheraliev, B., & Peng, Z. (2021). Molecular diversity of Uzbekistan's fishes assessed with DNA barcoding. *Scientific Reports*, **11**(1), 16894. https://doi.org/10.1038/s41598-021-96487-1

Singh, K., & Gupta, P. (2019). Conservation Management and Rehabilitation Aspect on the Fish Diversity of Upper Lake, Bhopal, India, *Journal of Critical Reviews*, **6**(4), 2394-5125.

Suryawanshi, R., Khedkar, G. D., Kalyankar, A., Khangale, D., & Shitale, K. (2024). DNA barcoding for the assessment of the taxonomy of fish from Chilika Lagoon, India. *Genetics of Aquatic Organisms*, **8**(3). http://doi.org/10.4194/GA812Suryawanshi, R. M., Khandagale, D., and Shitale, K. (2025). DNA Barcoding for the Identification of Fish Diversity from Hirakud Dam, India. *Russian Journal of Genetics*, **61**(6), 684-695.https://doi.org/10.1134/S1022795425700218

Tamot, P., & Awasthi, A. (2010). Biodiversity and conservation of indigenous fish species of Upper Lake, Bhopal (India). In *Proceedings of the national seminar on integrated management of water resources with*

reference to biodiversity and livelihood. Academy of Science and Technology, Bhopal, India pp, 378-388.

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences*, **101**(30), 11030-11035. https://doi.org/10.1073/pnas.0404206101

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, **38**(7), 3022-3027.https://doi.org/10.1093/molbev/msab120

Tang, Q., Deng, L., Luo, Q., Duan, Q., Wang, X., & Zhang, R. (2023). DNA barcoding of fish species diversity in Guizhou, China. *Diversity*, **15**(2), 203. https://doi.org/10.3390/d15020203

Taylor, H. R., & Harris, W. E. (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**(3), 377-388. https://doi.org/10.1111/j.1755-0998.2012.03119.x

VanVelzen, R., Weitschek, E., Felici, G., & Bakker, F. T. (2012). DNA barcoding of recently diverged species: relative performance of matching methods. *Public Library of Science One,* **7**(1), e30490. https://doi.org/10.1371/journal.pone.0030490

Vyas, V., & Veerwal, H. (2014). Spatial variations of waterbird diversity at Upper Lake, a Ramsar Site in Bhopal, MP, India. *Journal of Natural Sciences Research*, **4**(23), 123-126.

Ward, B.R. (2000). Declivity in steelhead (Oncorhynchus mykiss) recruitment at the Keogh River over the past decade. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**(2), 298-306. https://doi.org/10.1139/f99-243

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical transactions of the royal society B: biological sciences*, **360**(1462), 1847-1857. https://doi.org/10.1098/rstb.2005.1716

Wu, R., Zhang, H., Liu, J., Niu, S., Xiao, Y., & Chen, Y. (2018). DNA barcoding of the family Sparidae along the coast of China and revelation of potential cryptic diversity in the Indo-west Pacific oceans based on COI and 16S rRNA genes. *Journal of Oceanology and Limnology*, **36**(5), 1753-1770. https://doi.org/10.1007/s00343-018-7214-6

Appendix

Table S1: Comparison of intra- and inter-specific genetic divergence among the analysed fish species. The mean and maximum intraspecific genetic divergence and the corresponding nearest-neighbour (interspecific) genetic distances obtained from the COI-5P dataset to assess species-level genetic differentiation

Order	Family	Species	Mean Intra-Sp	Max Intra-Sp	Nearest Neighbour	Nearest Species	Distance to NN
Anabantiformes	Channidae	*Channa punctata*	N/A	0	ULBNU005-24	*Oreochromis niloticus*	20.89
Anabantiformes	Nandidae	*Nandus nandus*	N/A	0	ULBNU019-25	*Macrognathus pancalus*	21.63
Cichliformes	Cichlidae	*Oreochromis niloticus*	N/A	0	ULBNU023-25	*Mastacembelus armatus*	20.74
Cypriniformes	Cyprinidae	*Bangana dero*	N/A	0	ULBNU012-24	*Labeo boggut*	0.16
Cypriniformes	Cyprinidae	*Garra mullya*	1.82	1.82	ULBNU002-24	*Systomus sarana*	14.95
Cypriniformes	Cyprinidae	*Labeo boggut*	0	0	ULBNU009-24	*Bangana dero*	0.16
Cypriniformes	Cyprinidae	*Osteobrama cotio*	0.34	0.34	ULBNU024-25	*Puntius chola*	13.87
Cypriniformes	Cyprinidae	*Pethia ticto*	N/A	0	ULBNU008-24	*Puntius terio*	0.33
Cypriniformes	Cyprinidae	*Puntius chola*	N/A	0	ULBNU011-24	*Pethia ticto*	9.85
Cypriniformes	Cyprinidae	*Puntius terio*	N/A	0	ULBNU011-24	*Pethia ticto*	0.33
Cypriniformes	Cyprinidae	*Salmostoma bacaila*	0.16	0.16	ULBNU002-24	*Systomus sarana*	17.12
Cypriniformes	Cyprinidae	*Systomus sarana*	N/A	0	ULBNU004-24	*Labeo boggut*	14.9
Cypriniformes	Nemacheilidae	*Paracanthocobitis botia*	N/A	0	ULBNU002-24	*Systomus sarana*	20.3
Gobiiformes	Ambassidae	*Chanda nama*	N/A	0	ULBNU002-24	*Systomus sarana*	22.94
Gobiiformes	Gobiidae	*Glossogobius giuris*	N/A	0	ULBNU002-24	*Systomus sarana*	24.6
Osteoglossiformes	Notopteridae	*Notopterus notopterus*	N/A	0	ULBNU024-25	*Puntius chola*	22.84
Siluriformes	Bagridae	*Mystus bleekeri*	N/A	0	ULBNU003-24	*Heteropneustes fossilis*	20.14
Siluriformes	Heteropneustidae	*Heteropneustes fossilis*	N/A	0	ULBNU002-24	*Systomus sarana*	18.94
Synbranchiformes	Mastacembelidae	*Mastacembelus pancalus*	N/A	0	ULBNU023-25	*Mastacembelus armatus*	14.68
Synbranchiformes	Mastacembelidae	*Mastacembelus armatus*	N/A	0	ULBNU019-25	*Mastacembelus pancalus*	14.68

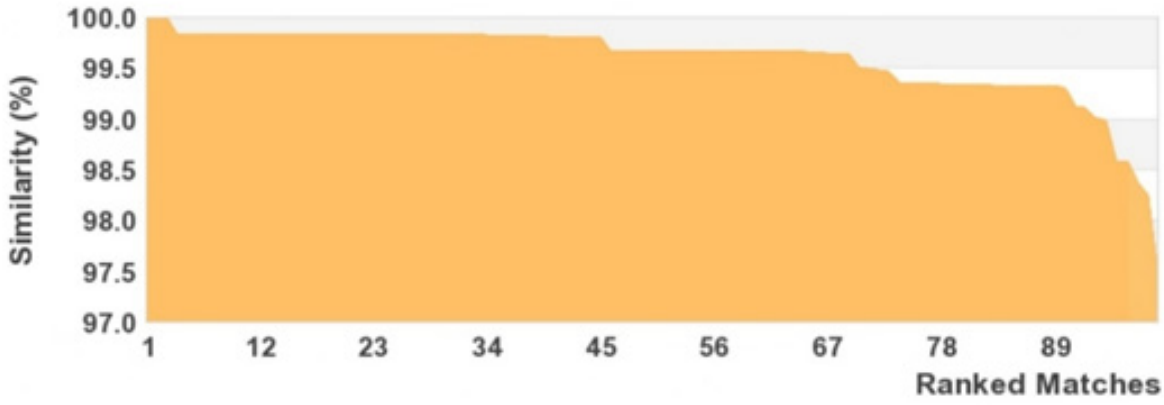


Figure S1: Similarity scores of the top 99 matches with database available on https:boldsystems.org