International Journal of Advanced Research in Biological Sciences
ISSN: 2348-8069
www.ijarbs.com
(A Peer Reviewed, Referred, Indexed and Open Access Journal)
DOI: 10.22192/ijarbs Coden: IJARQG (USA) Volume 9, Issue 9 -2022

Research Article

DOI: http://dx.doi.org/10.22192/ijarbs.2022.09.09.001

Delineating species boundary through DNA barcodes of some fishes inhabiting river Murti in the foothills of Darjeeling Himalayas

Adwitiya Chaudhuri
Assistant Professor, Department of Zoology, Pingla Thana Mahavidyalaya
Maligram, Paschim Medinipur, Pin-721140, India
Corresponding author**:** *adwitiya.hiya@gmail.com*

Abstract

The aim of the present study was to establish a global bioidentification system in the form of a reference sequence library where fish species inhabiting river Murti were delineated by a particular sequence denoted as genetic 'barcode'.40 species representing 27 genera, 16 families and 5 orders were collected along different altitudinal gradients if river Murti. Mitochondrial DNA was isolated from the samples followed by amplification of COI gene, alignment by CLUSTAL W and submission to GenBank. The average nucleotide frequencies were calculated where all the fishes showed relatively higher GC content at second codon position. Haplotypes showed gradually decreasing genetic distance within orders, families, genera with very low genetic distance within species indicating marked genetic divergence beyond species boundaries. The average transitional pairs were found more frequent than average transversional pairs. The Neighbour Joining tree revealed distinct clusters formed by members of order Cypriniformes, Siluriformes, Perciformes, Synbranchiformes and Beloniformes, where the conspecific individuals were always found to cluster under the same node supported by high bootstrap value, while dissimilar species were clustered under separate nodes, ensuring unambiguous identification of species. The research work suggests that COI barcoding can be taken up as a sensible approach to eliminate obscurity in the identification of the fish fauna with applications in its management and conservation of the species inhabiting in specific microhabitats.

Keywords: COI gene, DNA Barcode, sequence library, unambiguous identification, conservation.

Introduction

Taxonomic ambiguity exists for several fish genera/species, and a proper identification is imperative for management and trade (Lakra et al., 2011). The task of routine species identification based on morphological features has four significant limitations: i) incorrect identification of species due to both phenotypic plasticity and genetic variability in the characters employed for species recognition; ii) inability to identify morphologically cryptic taxa, which are common in many groups (Knowlton 1993; Jarman & Elliott 2000); iii) morphological keys effective only for a particular life stage or gender, may not be applicable to others and iv) misdiagnoses due to lack of high level of expertise (Hebert et al., 2003). These limitations signal the need for a new approach to taxon recognition, which comes in the form of DNA sequence analysis that has been used for quite a long time to assist species identifications.

Hebert et al. (2003) proposed that a single gene sequence, instead of different sequences for different taxonomic groups, would be sufficient to differentiate all, or at least the vast majority of, animal species, and that the use of the mitochondrial DNA gene cytochrome oxidase subunit I (COI) as a global bioidentification system for animals. The sequence was like a barcode, with species being delineated by a particular sequence or by a tight cluster of very similar sequences. With the development of a reference sequence library, new specimens and products can be identified by comparing their DNA barcode sequences against this barcode reference library (Lakra et al., 2011). Tautz et al. (2002, 2003) made the case for a DNA-based taxonomic system. Empirical support for the barcoding concept ranges from studies on invertebrates to birds. Currently, DNA barcoding is being employed to a large variety of organisms ranging from yeasts to humans (Hebert et al. 2004a,b; Hogg & Hebert 2004; Moritz & Cicero 2004) including instances of marine and freshwater fishes (Hajibabaei et al. 2005; Steinke et al. 2005; Ward et al. 2005; Hubert et al. 2008; Lakra et al. 2009).

India has a rich natural heritage and nurtures a unique bio-diversity, placing it among the 12 most biodiverse countries and ranks 9th in term of freshwater mega biodiversity (Mittermeier and Mittermeier 1997). The ichthyodiversity of the eastern Himalaya and its foothills is unique in terms of the lotic water ecosystems (Acharjee and Barat 2011). River Murti is one such hill stream flowing through Jalpaiguri district, West Bengal, India, harbouring a rich fish biodiversity throughout the year. The freshwater resources of India are currently experiencing an alarming decline in fish biodiversity due to several factors and as a result, a sizeable portion of fresh water fishes have been categorized as threatened. River conservation and management activities suffer from inadequate knowledge of the constituent biota. This emphasizes an immediate need for initiating research globally to develop alternative conservation planning to protect freshwater biodiversity of these aquatic systems (Pusey et al. 2010; Margules and Pressey 2000; Lipsey and Child 2007). In this context, the present study was directed to formulate precise identification manual for the constituent fish species of River Murti through introspecting a small segment of the genome, referred to as genetic ‘barcodes’ and provides information about the COI sequence data of the same. Accurate and unambiguous identification of fish at different stages of the life cycle is important in enabling retail substitutions of species to be detected, assisting fisheries management for long-term sustainability, and improving ecosystem research and conservation (Ward et al., 2005).

Materials and Methods

Study Area and Fish Sampling:

Fish sampling was conducted at River Murti which takes its rise from the Mo forest (near the Neora Valley National Park) in Darjeeling Himalayas (2211m above sea level or asl) flowing its way along the foothills before it meets the Jaldhaka River (102m asl). The study sites were selected along altitudinal gradient at Rocky Island (MR-I) (27 °00.483′ N 88°48.107′ E), Samsing

(MR-II) (26 °59.014′ N 88°49.291′ E) and North Dhupjhora (MR-III) (26 °50.631′ N 88°49.704′ E) that is named after the River Murti (Table 1, Figure 1). Monthly sampling was carried out for 3 years (from July 2013 to June 2016) in the river at the sampling sites (for a stretch of 2km) using cast net (mesh size of 1cm and covering an area of about 4.5m^2) and gill net (20m in length with 3cm spacing between adjacent knots). The specimens were retrieved from the net and identified morphologically to the lowest taxonomic level (Shaw and Shebbeare, 1937; Day, 1958 and Talwar and Jhingran, 1991). All species names adhere to Fishbase (Froese and Pauly, 2018).

Table 1: Location and Physico-chemical characteristics in the sampling sites of the river Murti

Sampling sites	Latitudes (North)	Longitudes (East)	Altitudes, at river bed (m asl)	River Width (m)	River Depth (m)	Velocity of water (ms^{-1})	Dissolved Oxygen (mgl^{-1})	pH	Substratum
MR-I	27 °00.483′	88°48.107′	516	16	1.1	1.1	9.2	7.1	Sand, gravel, boulders and bedrocks
MR-II	26 °59.014′	88°49.291′	402	18	0.9	0.9	8.9	7.4	Sand, gravel, boulders and bedrocks
MR-III	26 °50.631′	88°49.704′	137	25	1.4	0.5	8.5	7.5	Sand, gravel, and bedrocks

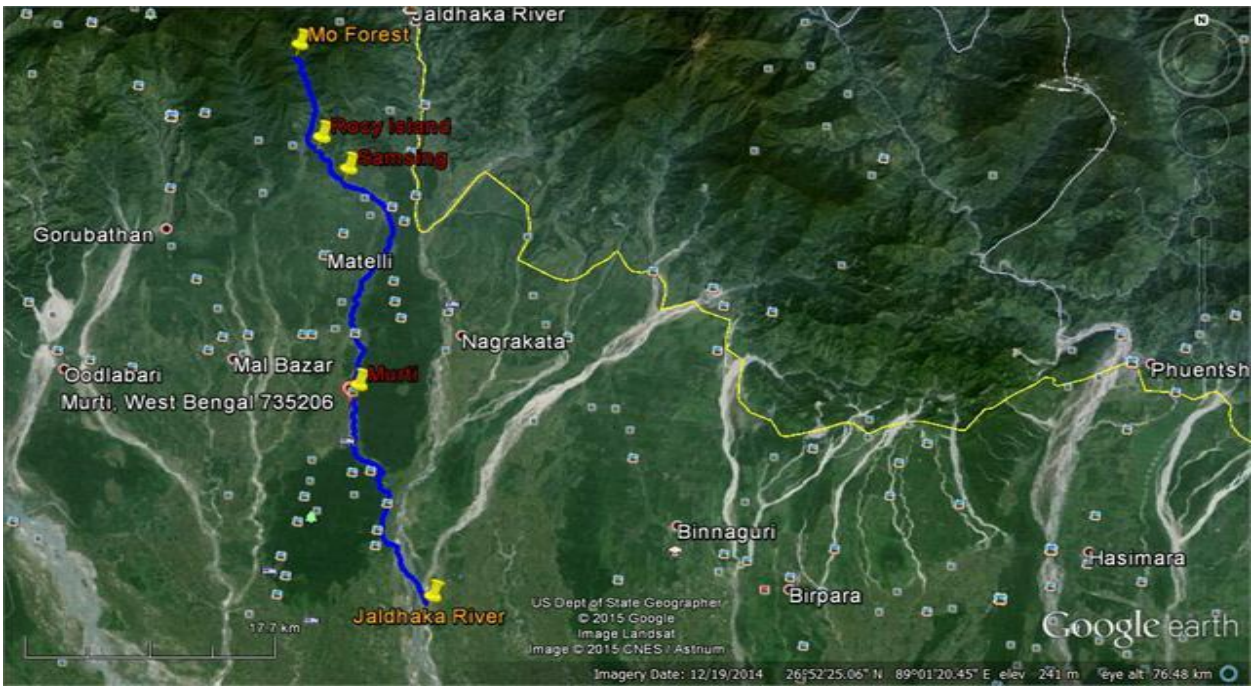

Fig 1: Study area and sampling sites in the River Murti (Courtesy- Google Earth)

Isolation of mitochondrial DNA:

Approximately 100 mg of white muscle tissue from 2 to 5 individuals of each species was preserved in 95% ethanol and brought to the laboratory under chilled condition. The samples were stored at –80° C until further use. Specimen details and GenBank accession numbersare shown in Table 2.DNA was isolated (Ruzzante et al., 1996) and its quality was checked in agarose gel. The concentration was estimated using a UV spectrophotometer. The DNA was diluted to a final concentration of 100 ng/µl. Approximately 655bp were amplified from the 5′ region of the COI gene in 50µl volume with 5 µl of 10XTaq polymerase buffer (Sibenzyme), 2µlof MgCl2 (50 mM), 2µl of dNTP (10 mM) mixture (Sibenzyme), 0.5 µl of each primer (0.01 mM), 0.5 U of Taq polymerase and 100 ng of genomic DNA(Sengupta and Homechaudhuri, 2013). The primers used for the amplification of the COI gene were (Ward et al., 2005 and Lakra et al., 2011):

F1 – 5′TCAACCAACCACAAAGACATTGGC AC3′ and
R1 – 5′ TAGACTTCTGGGTGGCCAAAGAAT CA3′

The thermal regime consisted of an initial denaturation for 2 min at 95° C followed by 35 cycles of 40 sec at 94° C, 40 sec at 54° C, 1 min 10 sec at 72° C, a final chain extension of 10 min at 72° C and held at 4° C. The PCR products were visualized on 1.2% agarose gels, and the most intense products were selected for sequencing after quantification through spectrophotometry. The products were purified using Qiagen gel extraction kit. Sequencing reactions were performed bidirectionally using 96 capillary high throughput sequencer ABI 3730 XL to generate sequences with accurate DNA base calling following manufacturer's instructions.

Data Analysis:

Sequences were aligned using CLUSTAL W (Thompson et al., 1994) and submitted to GenBank (Table 2). The extent of sequence difference between species was calculated by averaging pairwise comparisons across all individuals. The COI sequences of the conspecific individuals were aligned to yield a final sequence of 655 bp. The average nucleotide frequencies along with standard error were calculated. Pairwise evolutionary distance among haplotypes was determined by the Kimura two-parameter method (Kimura, 1980) using the software program MEGA (version 7) (Molecular Evolutionary Genetics Analysis) (Kumar et al., 2016). The mean nucleotide diversity (Tajima, 1989) was calculated. The neighbour-joining (NJ) tree was constructed using MEGA 7 and to verify the robustness of the internal nodes of NJ tree, bootstrap analysis was carried out using 1000 pseudo replications.

Results

a) All species

The results were presented for 40 species representing 27 genera, 16 families and 5 orders (Table 2).

These are *Acanthocobitis botia* (Hamilton, 1822),
Barilius barila (Hamilton, 1822),
Barilius bendelisis (Hamilton, 1807),
Barilius vagra (Hamilton, 1822),
Cabdio morar (Hamilton, 1822),
Crossocheilus latiuslatius (Hamilton, 1822),
Danio dangila (Hamilton, 1822),
Danio rerio (Hamilton, 1822),
Devario aequipinnatus (McClelland, 1839),
Devario devario (Hamilton, 1822),
Garra annandalei (Hora, 1921),
Garra gotylagotyla (Gray, 1830),
Garra kempi (Hora, 1921),
Labeo bata (Hamilton, 1822),
Neolissochilus hexagonolepis (McClelland, 1839,
Neolissochilus hexastichus (McClelland, 1839),
Neolissochilus stracheyi (Day, 1871),
Opsarius barna (Hamilton, 1822),
Opsarius tileo (Hamilton, 1822),
Pethia phutunio (Hamilton, 1822),
Puntius terio (Hamilton, 1822),
Puntius vittatus (Day, 1865),
Lepidocephalichthys guntea (Hamilton, 1822),

Psilorhynchus balitora (Hamilton, 1822),
Amblyceps mangois (Hamilton, 1822),
Chaca chaca (Hamilton, 1822),
Clarias batrachus (Linnaeus, 1758),
Mystus bleekeri (Day, 1877),
Olyra longicaudata (McClelland, 1842),
Pseudolaguvia foveolata (Ng, 2005),
Badis badis (Hamilton, 1822),
Chanda nama (Hamilton, 1822),
Channa marulius (Hamilton, 1822),
Channa orientalis (Bloch & Schneider, 1801),
Channa punctata (Bloch, 1793),
Channa stewartii (Playfair, 1867),
Trichogaster fasciata (Bloch & Schneider, 1801),
Macrognathus pancalus (Hamilton, 1822),
Mastacembelus armatus (Lacepède, 1800) and *Xenentodon cancila* (Hamilton, 1822). A total of 53 sequences were generated from these forty species using multiple specimens for all species. The universal primers amplified the target region in all the fishes generating barcodes of 655 bp. Sequence analysis revealed average nucleotide frequencies as T = 29.85%, C = 27.28%, A = 24.74%, G = 18.13% (Table 3). All the fish species studied showed relatively higher GC (45.30%) content (Figure 2). The average transitional pairs were more frequent than average transversional pairs, with an average ratio of 2.09. The mean nucleotide diversity (Tamura et al., 2007) was found to be 0.233. The average K2P distance (Thompson et al., 1994) in percentage within different taxonomic levels was calculated (Table 4). The average genetic distance within species, genera, families and orders were found to be 0.12%, 7.5%, 9.7% and 19.4% respectively. The NJ tree is shown in Figure 3. The details are discussed below.

b) Order: Cypriniformes

A single individual fish of *Acanthocobitis botia* belonging to family Nemacheilidae under the order Cypriniformes were analysed. The sequence analysis revealed nucleotide frequencies as T = 28.24%, C = 28.70%, A = 24.73%, G = 18.32% (Table 3).The highest GC content was found at the second codon (56.42%) position with progressive declining trend towards third (43.12%) and first (41.55%) codon positions (Figure 2). The average genetic distance of family Nemacheilidae with other families was 24.8%.The NJ tree revealed distinct clusters shared by *Acanthocobitis botia* with the members of order Cypriniformes having high bootstrap values.

Table 2: DNA barcoded fish species along with GenBank accession numbers

Order	Family	Genus	Species	Threat Status (According to BCPP-CAMP,	No. of individuals	GenBank accession No.
Cypriniformes	Nemacheilidae	*Acanthoco-bitis*	*botia*	Lower Risk- near threatened	1	MK174265
Cypriniformes	Cyprinidae	*Barilius*	*barila*	Vulnerable	1	MK174283
Cypriniformes	Cyprinidae	*Barilius*	*bendelisis*	Lower Risk- near threatened	2	MK125076 MK125077
Cypriniformes	Cyprinidae	*Barilius*	*vagra*	Vulnerable	1	MN062735
Cypriniformes	Cyprinidae	*Cabdio*	*morar*	Lower Risk- near threatened	1	MN062736
Cypriniformes	Cyprinidae	*Crossocheilus*	*latiuslatius*	Data Deficient	1	MN075127
Cypriniformes	Cyprinidae	*Danio*	*dangila*	Not evaluated	1	MN080525

Cypriniformes	Cyprinidae	*Danio*	*rerio*	Lower Risk- near threatened	1	MN080526
Cypriniformes	Cyprinidae	*Devario*	*aequipinnatus*	Lower Risk- near threatened	2	MK174280 MK174281
Cypriniformes	Cyprinidae	*Devario*	*devario*	Lower Risk- near threatened	1	MN080527
Cypriniformes	Cyprinidae	*Garra*	*annandalei*	Not evaluated	3	MK174277 MK174278 MK174279
Cypriniformes	Cyprinidae	*Garra*	*gotyla gotyla*	Vulnerable	1	MN080528
Cypriniformes	Cyprinidae	*Garra*	*kempi*	Vulnerable	1	MN080529
Cypriniformes	Cyprinidae	*Labeo*	*bata*	Lower Risk- near threatened	1	MN080530
Cypriniformes	Cyprinidae	*Neolissochilus*	*hexagonolepis*	Not evaluated	4	MK174266 MK174267 MK174268 MK174269
Cypriniformes	Cyprinidae	*Neolissochilus*	*hexastichus*	Not evaluated	1	MN096213
Cypriniformes	Cyprinidae	*Neolissochilus*	*stracheyi*	Not evaluated	1	MN096214
Cypriniformes	Cyprinidae	*Opsarius*	*barna*	Lower Risk- near threatened	2	MK125074 MK125075
Cypriniformes	Cyprinidae	*Opsarius*	*tileo*	Lower Risk- near threatened	1	MN062734
Cypriniformes	Cyprinidae	*Pethia*	*phutunio*	Lower Risk- least concern	1	MN096217
Cypriniformes	Cyprinidae	*Puntius*	*terio*	Lower Risk- near threatened	1	MN096218
Cypriniformes	Cyprinidae	*Puntius*	*vittatus*	Vulnerable	1	MN096219
Cypriniformes	Cobitidae	*Lepidocephali-chthys*	*guntea*	Not evaluated	3	MK174274 MK174275 MK174276
Cypriniformes	Psilorhyn-chidae	*Psilorhynchus*	*balitora*	Not evaluated	3	MK118746 MK174270 MK174271
Siluriformes	Amblycipiti-dae	*Amblyceps*	*mangois*	Lower Risk- near threatened	1	MK174264
Siluriformes	Chacidae	*Chaca*	*chaca*	Not evaluated	1	MN067436
Siluriformes	Clariidae	*Clarias*	*batrachus*	Vulnerable	1	MN075126
Siluriformes	Bagridae	*Mystus*	*bleekeri*	Vulnerable	1	MN096212
Siluriformes	Olyridae	*Olyra*	*longicaudata*	Not evaluated	1	MN096215
Siluriformes	Erethistidae	*Pseudolaguvia*	*foveolata*	Not evaluated	1	MN096216

Perciformes	Badidae	*Badis*	*badis*	Not evaluated	1	MK174282
Perciformes	Ambassidae	*Chanda*	*nama*	Not evaluated	1	MN067437
Perciformes	Channidae	*Channa*	*marulius*	Lower Risk-near threatened	1	MN067438
Perciformes	Channidae	*Channa*	*orientalis*	Vulnerable	1	MN075123
Perciformes	Channidae	*Channa*	*punctata*	Lower Risk-near threatened	1	MN075124
Perciformes	Channidae	*Channa*	*stewartii*	Not evaluated	1	MN075125
Perciformes	Osphronemidae	*Trichogaster*	*fasciata*	Lower Risk-near threatened	1	MN096220
Synbranchiformes	Mastacembelidae	*Macrognathus*	*pancalus*	Lower Risk-near threatened	1	MN080531
Synbranchiformes	Mastacembelidae	*Mastacembelus*	*armatus*	Not evaluated	2	MK174272 MK174273
Beloniformes	Belonidae	*Xenentodon*	*cancila*	Lower Risk-near threatened	1	MN096221

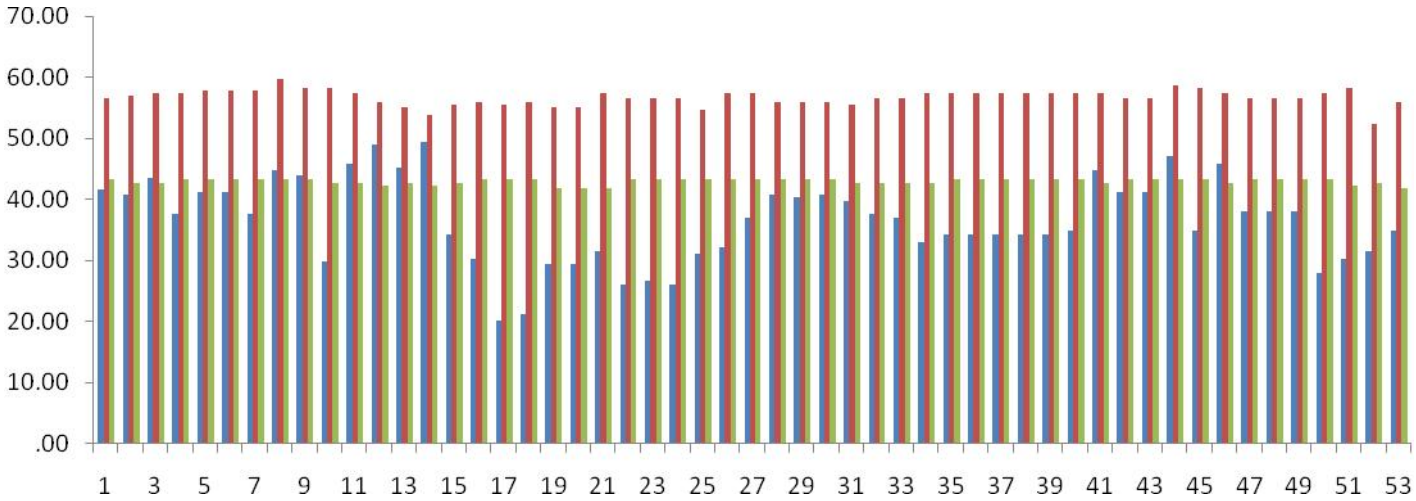

Fig 2: Variation in GC content for cytochrome c oxidase 1 (COI) gene sequence at first, second and third codon positions respectively, among the fishes under study

Key: 1, *Acanthocobitis botia*; 2, *Amblyceps mangois*; 3. *Badis badis;*4*; Barilius barila;* 5, *Barilius bendelisis* 1; 6, *Barilius bendelisis* 2; 7, *Barilius vagra*; 8, *Cabdio morar*; 9, *Chaca chaca*; 10, *Chanda nama*; 11, *Channa marulius*; 12, *Channa orientalis*; 13, *Channa punctata*; 14, *Channa stewartii;* 15, *Clarias batrachus*; 16, *Crossocheilus latiuslatius*; 17, *Danio dangila*; 18, *Danio rerio*; 19, *Devarioa equipinnatus* 1; 20, *Devarioa equipinnatus* 2; 21, *Devario devario*; 22, *Garra annandalei* 1; 23, *Garra annandalei* 2; 24, *Garra annandalei* 3; 25, *Garra gotyla gotyla*; 26, *Garra kempi*; 27, *Labeobata*; 28, *Lepidocephalichthys guntea* 1; 29, *Lepidocephalichthys guntea* 2; 30, *Lepidocephalichthys guntea* 3; 31, *Macrognathus pancalus*; 32, *Mastacembelus armatus* 1; 33, *Mastacembelus armatus* 2; 34, *Mystus bleekeri*; 35, *Neolissochilus hexagonolepis* 1; 36, *Neolissochilus hexagonolepis* 2; 37, *Neolissochilus hexagonolepis* 3; 38, *Neoloissochilushexagonolepis* 4; 39, *Neoloissochilushexasticus*; 40, *Neolissochilusstracheyi;* 41, *Olyra longicaudata*; 42, *Opsarius barna* 1; 43, *Opsarius barna* 2; 44, *Opsarius tileo*; 45, *Pethia phutunio*; 46, *Pseudolaguviafoveolata*; 47, *Psilorhynchusbalitora* 1; 48, *Psilorhynchusbalitora* 2; 49, *Psilorhynchus balitora* 3; 50, *Puntius terio*; 51, *Puntius vittatus*; 52, *Trichogaster fasciata*; 53, *Xenentodon cancila*

Twenty nine individual fishes of twenty one species belonging to eleven genera and family Cyprinidae under the order Cypriniformes were analysed. The sequence analysis revealed average nucleotide frequencies as T ranging from 26.56% in *Cabdio morar* to 34.50% in *Danio rerio* and *Danio dangila*, C ranging from 22.14% in *Danio rerio* and *Danio dangila* to 29.92% in *Cabdio morar*, A ranging from 22.21% in *Barilius bendelisis* to 27.18% in *Garra gaotyla gotyla*, G ranging from 16.4% in *Crossocheilus latiuslatius* to 20% in *Opsarius tileo* (Table 3). The highest GC content was found at the second codon (ranging from 54.59% in *Garra gotyla gotyla* to 58.26% in *Pethia phutunio*) position with progressive declining trend towards the third (ranging from 41.74% in *Devario aequipinnatus* and *Devario devario* to 43.12% in *Barilius barila, Barilus bendelisis, Barilius vagra, Crossocheilus latiuslatius, Danio dangila, Danio rerio, Garra annandalei, Garra gotyla gotyla, Garra kempi, Labeobata, Neolissocheilus hexagonolepis, Neolissocheilus hexastichus, Neolissocheilus stracheyi, Opsarius barna, Pethia phutunio* and *Puntius terio*) and first (ranging from 20.9% in *Danio dangila* to 41.10% in *Barilus bendelisis* and *Opsarius barna*) codon positions (Figure 2) for all fish species under the family Cyprinidae, except *Cabdio morar* and *Opsarius tileo*. The GC content was highest at the second codon (59.63% and 58.72%) position with a progressive declining trend towards the first (44.75% and 47.03%) and third codon (43.12% in both) positions for *Cabdio morar* and *Opsarius tileo* respectively. The average ratio of transition and transversion in the family is 2.95 showing that average transitional pairs were more frequent than average transversional pairs. The average genetic distance within family was 19.47%, whereas the average genetic distance of family Cyprinidae with other families was 25.5%.The NJ tree revealed distinct clusters shared by the species of the same family (Figure 3). All assemblages of conspecific individuals had high bootstrap values and the confamilial species formed the same clade.

Three individual fishes of *Lepidocephalichthys guntea* belonging to the family Cobitidae under the order Cypriniformes were analysed. The sequence analysis revealed average nucleotide frequencies as T = 29.77%, C = 29.16%, A = 23.72%, G = 17.35% (Table 3). The highest GC content was found at the second codon (55.96%) position with progressive declining trend towards the third (43.12%) and first (40.64%) codon positions (Figure 2). The average ratio of transition and transversion in the family is 1111989.88 showing that average transitional pairs were much more frequent than average transversional pairs. The average genetic distance within family was 0.10%. The average genetic distance of the family Cobitidae with other families under study (i.e. between families) was 24.9%, with value ranging from 21% to 27.1%. The NJ tree revealed distinct clusters shared by the conspecific individuals having high bootstrap values (Figure 3).

Three individual fishes of *Psilorhynchus balitora* belonging to the family Psilorhynchidae under the order Cypriniformes were analysed. The sequence analysis revealed average nucleotide frequencies as T = 31.30%, C = 26.56%, A = 22.90%, G = 19.24% (Table 3). The highest GC content was found at the second codon (56.42%) position with progressive declining trend towards the third (43.12%) and first (37.90%) codon positions (Figure 2). The average ratio of transition and transversion in the family is 0.50indicating equal number of transitional and transversional pairs. The average genetic distance within family was 0%. The average genetic distance of family Psilorhynchidae with other families under study (i.e. between families) was 25.3%, with value ranging from 22.9% to 28%. The NJ tree revealed distinct clusters shared by the conspecific individuals having high bootstrap values (Figure 3).

Table 3: Mean percentage base composition (with SE) comparing cytochrome c oxidase 1 (COI) sequence of the fishes under study

Fish Species	Average Nucleotide Composition (%)			
	T(U)	C	A	G
Acanthocobitis botia	28.24±0	28.70±0	24.73±0	18.32±0
Amblyceps mangois	30.38±0	28.09±0	22.90±0	18.63±0
Badis badis	29.16±0	29.47±0	23.05±0	18.32±0
Barilius barila	30.23±0	26.87±0	23.82±0	19.08±0
Barilius bendelisis	30.46±0.08	27.40±0.08	22.21±0.08	19.92±0.08
Barilius vagra	30.08±0	27.48±0	23.82±0	18.63±0
Cabdio morar	26.56±0	29.92±0	24.27±0	19.24±0
Chaca chaca	26.41±0	30.69±0	25.19±0	17.71±0
Chanda nama	33.13±0	25.19±0	23.36±0	18.32±0
Channa marulius	27.02±0	30.99±0	24.43±0	17.56±0
Channa orientalis	28.70±0	29.62±0	22.29±0	19.39±0
Channa punctata	29.16±0	29.92±0	23.21±0	17.71±0
Channa stewartii	28.55±0	30.38±0	23.05±0	18.02±0
Clarias batrachus	29.16±0	26.41±0	26.72±0	17.71±0
Crossocheilus latiuslatius	30.23±0	26.72±0	26.72±0	16.40±0
Danio dangila	34.05±0	22.14±0	26.41±0	17.40±0
Danio rerio	34.05±0	22.14±0	25.95±0	17.86±0
Devario aequipinnatus	33.44±0	24.05±0.08	24.58±0	17.94±0.08
Devario devario	31.91±0	25.34±0	24.58±0	18.17±0
Garra annandalei	31.09±0.05	24.78±0.05	27.02±0	17.10±0
Garra gotyla gotyla	29.92±0	25.80±0	27.18±0	17.10±0
Garra kempi	29.16±0	26.72±0	26.72±0	17.40±0
Labeo bata	28.85±0	27.63±0	25.34±0	18.17±0
Lepidocephalichthys guntea	29.77±0	29.16±0	23.72±0.05	17.35±0.05
Macrognathus pancalus	27.94±0	29.31±0	26.11±0	16.64±0
Mastacembelus armatus	28.78±0.08	28.32±0.08	25.80±0	17.10±0
Mystus bleekeri	29.77±0	27.63±0	25.95±0	16.64±0
Neolissochilus hexagonolepis	28.85±0	27.33±0	26.26±0	17.56±0
Neoloissochilus hexasticus	29.01±0	27.18±0	26.11±0	17.71±0
Neolissochilus stracheyi	28.85±0	27.33±0	26.11±0	17.71±0
Olyra longicaudata	29.77±0	28.24±0	21.98±0	20.00±0
Opsarius barna	29.77±0	27.02±0	23.36±0	19.85±0
Opsarius tileo	27.18±0	29.62±0	23.21±0	20.00±0
Pethia phutunio	29.16±0	26.26±0	25.50±0	19.08±0
Pseudolaguvia foveolata	27.94±0	28.70±0	23.51±0	19.85±0
Psilorhynchus balitora	31.30±0	26.56±0	22.90±0	19.24±0
Puntiusterio	30.38±0	24.89±0	26.87±0	17.86±0
Puntius vittatus	29.77±0	26.56±0	26.72±0	16.95±0
Trichogaster fasciata	34.35±0	24.58±0	23.51±0	17.56±0
Xenentodon cancila	31.30±0	25.95±0	24.58±0	18.17±0

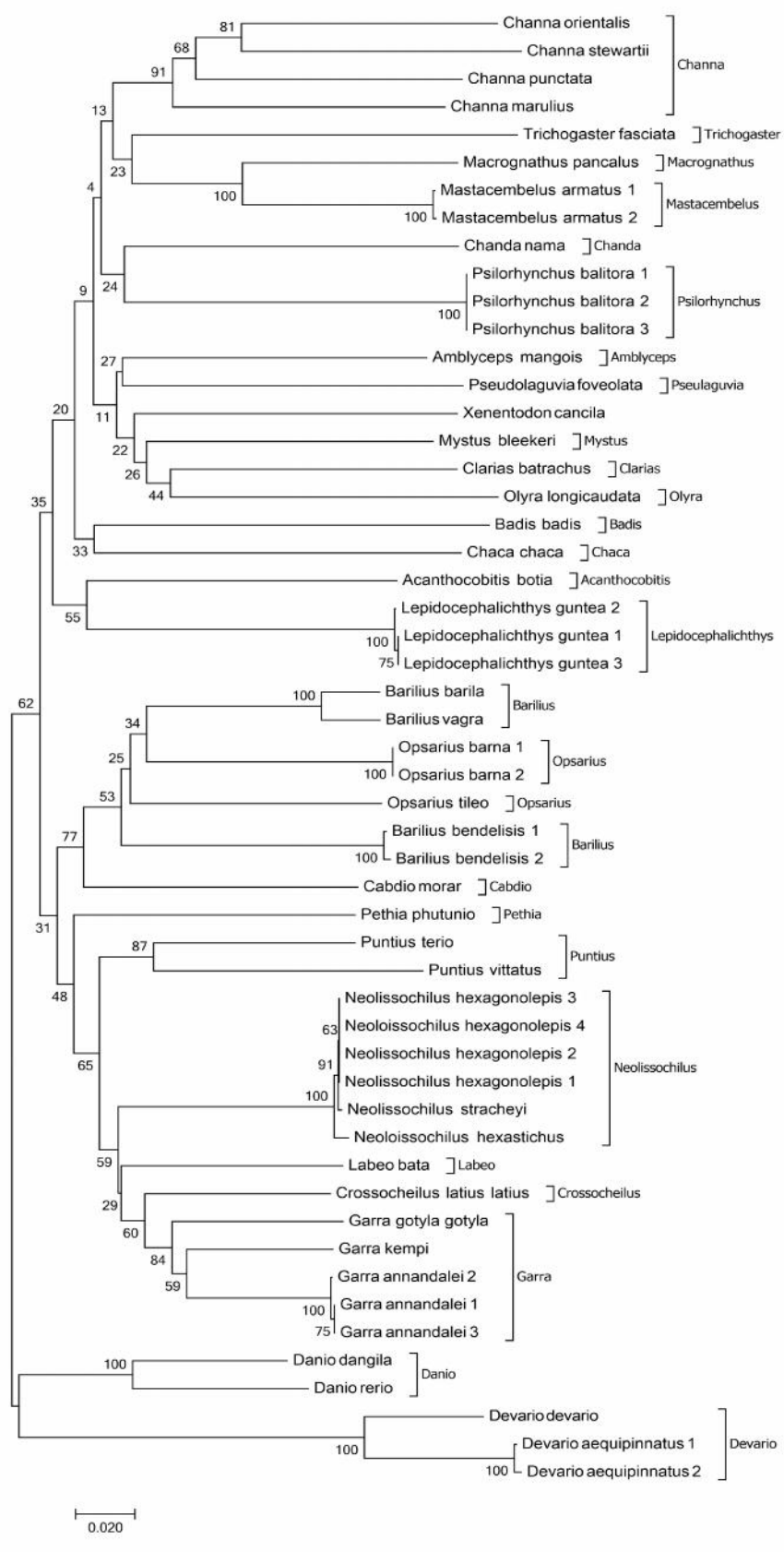

Fig 3: Neighbour-joining tree constructed with the sampled fishes on the basis of COI gene sequence with bootstrap values

c) Order: Siluriformes

Six individual fishes namely *Amblyceps mangois*, *Chaca chaca*, *Clarias batrachus*, *Mystus bleekeri*, *Olyra longicaudata* and *Pseudolaguvia foveolata* belonging to family Amblycipitidae, Chacidae, Clariidae, Bagridae, Olyridae and Erethistidae respectively under the order Siluriformes were analysed. The sequence analysis revealed average nucleotide frequencies as T = 30.38%, C = 28.09%, A = 22.90%, G = 18.63% for *Amblyceps mangois*; T = 26.41%, C = 30.69%, A = 25.19%, G = 17.71% for *Chaca chaca*; T = 29.16%, C = 26.41%, A = 26.72%, G = 17.71% for *Clarias batrachus*; T = 29.77%, C = 27.63%, A = 25.95%, G = 16.64% for *Mystus bleekeri*; T = 29.77%, C = 28.24%, A = 21.98%, G = 20% for *Olyra longicaudata* and T = 27.94%, C = 28.70%, A = 23.51%, G = 19.85% for *Pseudolaguvia foveolata* (Table 3). The highest GC content was found at the second codon (56.88%, 55.50% and 57.34%) position with progressive declining trend towards the third (42.66% for all the three species) and first (40.64%, 34.25% and 32.88%) codon positions (Figure 2) for *Amblyceps mangois*, *Clarias batrachus* and *Mystus bleekeri* respectively. Whereas for *Chaca chaca*, *Olyra longicaudata* and *Pseudolaguvia foveolata* respectively the GC content was found to be highest at the second codon (58.26%, 57.34% and 57.34%) position with progressive declining trend towards the first (43.84%, 44.75% and 45.66%) and third (43.12%, 42.66% and 42.66%) codon positions (Figure 2). The average genetic distance between families were 23.9%, 26%, 24.1%, 23.5%, 25.3% and 25% for Amblycipitidae, Chacidae, Clariidae, Bagridae, Olyridae and Erethistidae respectively. The NJ tree revealed distinct clusters shared by the species of the same family (Figure 3).

d) Order: Perciformes

Four fish species, each represented by a single individual, belonging to family Channidae under the order Perciformes were analysed. The sequence analysis revealed average nucleotide frequencies as T = 27.02%, C = 30.99%, A = 24.43%, G = 17.56% for *Channa marulius*; T = 28.70%, C = 29.62%, A = 22.29%, G = 19.39% for *Channa orientalis*; T = 29.16%, C = 29.92%, A = 23.21%, G = 17.71% for *Channa punctata* and T = 28.55%, C = 30.38%, A = 23.05%, G = 18.02% for *Channa stewartii* (Table 3). The highest GC content was found at the second codon (57.34%, 55.96%, 55.05% and 53.67%) position with progressive declining trend towards first (45.66%, 48.86%, 45.21% and 49.32%) and third (42.66%, 42.20%, 42.66% and 42.20%) codon positions (Figure 2) for *Channa marulius*, *Channa orientalis*, *Channa punctata* and *Channa stewartii* respectively. The average ratio of transition and transversion in the family is 2.70 indicating that average transitional pairs were more frequent than average transversional pairs. The average genetic distance within family was 19.67% whereas the average genetic distance of family Channidae with other families under study (i.e. between families) was 25.6%.The NJ tree revealed distinct clusters shared by species of the same genera (Figure 3).

Three individual fishes namely *Badis badis*, *Chanda nama* and *Trichogaster fasciata* belonging to family Badidae, Ambassidae and Osphronemidae respectively under the order Perciformes were analysed.The sequence analysis revealed average nucleotide frequencies as T = 29.16%, C = 29.47%, A = 23.05%, G = 18.32% for *Badis badis*; T = 33.13%, C = 25.19%, A = 23.36%, G = 18.32% for *Chanda nama* and T = 34.35%, C = 24.58%, A = 23.51%, G = 17.56% for *Trichogaster fasciata* (Table 3). The highest GC content was found at the second codon (58.26% and 52.29%) position with progressive declining trend towards the third (42.66% for both) and first (29.68% and 31.51%) codon positions (Figure 2) for *Chanda nama* and *Trichogaster fasciata* respectively. Whereas for *Badis badis* the GC content was found to be highest at the second codon (57.34%) position with progressive declining trend towards the first (43.38%) and third (42.66%) codon positions (Figure 2). The average genetic distance between families was 27.2%, 25% and 26.7% for Badidae, Ambassidae and Osphronemidae respectively. The NJ tree revealed distinct clusters shared by the species of the same family (Figure 3).

All assemblages of the confamilial species had relatively low bootstrap value.

e) Order: Synbranchiformes

Three individual fishes belonging to two genera and family Mastacembelidae under the order Synbranchiformes were analysed. The sequence analysis revealed average nucleotide frequencies as T = 27.94%, C = 29.31%, A = 26.11%, G = 16.64% for *Macrognathus pancalus* and T = 28.78%, C = 28.32%, A = 25.80%, G = 17.10% for *Mastacembelus armatus* (Table 3). The highest GC content was found at the second codon (55.50% and 56.42%) position with progressive declining trend towards the third (42.66% for both the species) and first (39.73% and 37.21%) codon positions (Figure 2) for *Macrognathus pancalus* and *Mastacembelus armatus* respectively (Figure 2). The average ratio of transition and transversion in the family is 3.68 showing that average transitional pairs were more frequent than average transversional pairs. The average genetic distance within family was 9.26%, whereas the average genetic distance of family Mastacembelidae with other families was 24.4%.The NJ tree revealed distinct clusters shared by the species of the same family (Figure 3). All assemblages of conspecific individuals had high bootstrap values and the confamilial species formed the same clade.

f) Order: Beloniformes

A single individual fish of *Xenentodon cancila* belonging to family Belonidae under the order Beloniformes were analysed. The sequence analysis revealed nucleotide frequencies as T = 31.30%, C = 25.95%, A = 24.58%, G = 18.17% (Table 3). The highest GC content was found at the second codon (55.96%) position with progressive declining trend towards third (41.74%) and first (34.70%) codon positions (Figure 2). The average genetic distance of family Belonidae with other families was 24.4%.The NJ tree revealed distinct clusters shared by *Xenentodon cancila* with the members of order Siluriformes having low bootstrap values.

Table 4: Summary of genetic divergence (K2P percentage) within various taxonomic levels calculated on the basis of COI sequence of the fishes under study

Comparison within	Minimum	Distance mean	Maximum	Standard error
Species	0	0.0012	0.003063	0.0005
Genera	0	0.075	0.196659	0.022
Families	0	0.097	0.195976	0.044
Orders	0.092574	0.194	0.243569	0.035

Discussion

In this study, the COI gene of forty species belonging to 27 genera, 16 families and 5 orders, collected from river Murti was sequenced (Table 2). Unambiguous identification of these fishes at any life-history stage should be the first step for efficient management and long term sustainability. A global DNA-based barcode identification system is believed to provide a simple yet universal tool for identification of the fishes (Ruzzante et al., 1996).

The mitochondrial genome of animals is preferred for analysis over the nuclear genome because of three main reasons: i) lack of introns, ii) limited exposure to recombination and iii) haploid mode of inheritance (Saccone *et al.* 1999). The cytochrome *c* oxidase I gene (COI) is targeted to focus the analysis on as it has three important advantages. First, insertions and deletions are rare (as they mostly lead to a shift in the reading frame) (Hebert et al., 2003). Second, the universal primers for this gene are very robust, enabling recovery of its 5′ end from representatives of most, if not all, animal phyla (Folmer *et al.* 1994; Zhang & Hewitt 1997). Third, COI appears to possess a greater range of phylogenetic signal than any other mitochondrial gene (Knowlton & Weigt 1998).

The universal primers amplified the target region in 53 individual fishes under study generating barcodes of 655 bp per taxon suggesting that NUMTs (nuclear DNA sequences originating from mitochondrial DNA sequences) were not sequenced because vertebrate NUMTS are typically smaller than 600 bp (Zhang & Hewitt 1996). The sequences were found to be diverse enough to permit unambiguous identification of fish species. No insertions, deletions or stop codons were observed in any sequence. The lack of stop codons is consistent with the fact that all amplified sequences being functional mitochondrial COI sequences (Ward et al., 2005).

Profound shifts in nucleotide usage, which may have serious impacts on phylogenetic analyses, (Philippe and Laurent, 1998; Foster and Hickey, 1999 and Mooers and Holmes, 2000) were shown by mitochondrial genomes among major taxonomic groups (Jermin et al., 1994). The average GC content was found to be 45.30% (Fig 2) in this study. This result was in agreement with those on complete mitochondrial genomes of Osteichthyes species, deriving GC contents of 43.2% (Zhang & Hewitt 1996) and 47.1% (Ward et al., 2005) for the COI gene sequence. The highest GC content (49.62%) was found in *Opsarius tileo*, whereas the lowest (39.54%) was observed in *Danio dangila*. In another study involving 16 animal phyla, the GC content was found to vary within a range of 22–53%, with bony fish species *Chirocentrus dorab* (52.8%) and *Salanxaria kensis* (53.4%) showing higher GC content in general among all other species studied (Clare et al., 2008). These findings may be indicative of the fact that fishes in general display higher GC content.

The average transition and transversion ratio was 2.09, showing bias for transitional events. The mean nucleotide diversity (*Pi*) among all the species was estimated as 0.233. The COI locus harbours a high mutational rate for mtDNA (Saccone et al., 1999). The average K2P distance of individuals within species was estimated as 0.12%, whereas it was 7.5% at the genus level, 9.7% for families and 19.4% for orders (Table 4). Hence, there was more than 62-fold sequence difference among species of the same genus than conspecific individuals. The variation was even more (80.83 fold) among the confamilial species and very high (161.67 fold) among individuals belonging to different orders. Hence, mean genetic distance within individuals of a species was much smaller than the average distance between individuals of different species belonging to different families and orders indicating marked genetic divergence beyond species boundaries that has been referred to as the barcoding gap (Meyer and Paulay, 2005). All these observations correspond reasonably well with the studies on Australian fish species (Ward et al., 2005), Alaskan skates (Spies et al., 2006) and Indian marine fishes (Lakra et al., 2011).

Although barcode analysis was performed to delineate species boundaries only, there was some phylogenetic signal in the COI sequence data (Ward et al., 2005). The NJ tree (Figure 3) revealed distinct clusters formed by members of order Cypriniformes, Siluriformes, Perciformes, Synbranchiformes and Beloniformes. The conspecific individuals were always found to cluster under the same node supported by high bootstrap value, whereas dissimilar species were clustered under separate nodes, enabling unambiguous identification of species. However, it is inappropriate to assume that the true phylogeny of fishes from a 655 bp fragment of mitochondrial DNA through K2P distance and NJ method can provide the final answer to species identification; rather they are an additional system (Okumus and Ciftci, 2003). Hence more gene regions should be used (including nuclear genes) and additional analytic methods including maximum parsimony and maximum likelihood should be deployed (Ward et al., 2005).

The task of barcode generation is of utmost importance since it permits the accurate and unambiguous identification of the vast majority of fish species through species-specific signals present in the COI sequence. Our results suggest that COI barcoding can be taken up as a sensible approach to eliminate obscurity in the identification of the fish fauna with applications in its management and conservation. Since river

Murti harbours fish fauna that are likely to survive only in specific microhabitats of hill streams, urbanization along with increased anthropogenic activity may pose serious threats to such ichthyofauna. Collection of such diverse fish fauna, followed by proper identification through DNA barcodes and their introduction in some freshwater protected area (FPA) may help in sustainable maintenance of the habitat specific fish stock which can later be utilized for rehabilitation work in degraded aquatic systems. Hence DNA barcoding can be a precious tool for fisheries ecologists to demonstrate the benefits of healthier ecosystems to the community and stakeholders.

Acknowledgments

The author is thankful to the University Grants Commission for financial assistance, the Head of the Department of Zoology, University of Calcutta for providing infrastructural facilities and Dr. Subhasree Sengupta and Prof. Sumit Homechaudhuri for technical guidance.

References

1. Acharjee, M.L. and Barat, S., 2011. Spatio-temporal Dynamics of Physico-Chemical Factors of River Relli in Darjeeling Himalaya, West Bengal, India. NBU *J. Anim. Sci.*, 5, 24-33.
2. Clare, E. L., Kerr, K. C. R., von Konigslow, T. E., Wilson, J. J. and Hebert, P. D. N.,2008. Diagnosing mitochondrial DNA diversity: applications of a sentinel gene approach. J. Mol. Evol., 66, 362–367.
3. Day, F., 1958. *The Fishes of India*, *Vols 1 and 2*, William Dawson, London.
4. 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. Mol. Mar. Biol. Biotechnol. 3, 294–299.
5. Foster, P. G. and Hickey, D. A.,1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol., 48, 284–290.
6. Froese, R. and D. Pauly. Editors. 2018. FishBase. World Wide Web electronic publication. www.fishbase.org, Version (02/2018)
7. Hajibabaei, M., de Waard, J.R., Ivanova, N., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M. and Hebert, P.D.N., 2005. Critical factors for assembling a high volume of DNA barcodes. Philosophical Transactions of the Royal Society B, 360, 1959–1967.
8. Hebert, P.D.N., Penton E.H., Burns J.M, Janzen D.H., Hallwachs, W., 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptesfulgerator. Proceedings of the National Academy of Sciences USA, 101, 14812–14817.
9. Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004b. Identification of birds through DNA barcodes. Public Library of Science, 2, 1657–1663.
10. Hebert, P.D.N., Cywinska, A., Ball, S.L. and Ward, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London, Ser. B, 270, 313–322.
11. Hogg, I.D., Hebert, P.D.N., 2004. Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. Canadian Journal of Zoology, 82, 749–754.
12. Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., 2008. Identifying Canadian freshwater fishes through DNA barcodes. Public Library of Science, 3, e2490. doi:10.1371/journal.pone.0002490.
13. Jarman, S.N. & Elliott, N.G., 2000. DNA evidence for morphological and cryptic Cenozoicspeciations in the Anaspididae, 'living fossils' from the Triassic. J. Evol. Biol. 13, 624–633.
14. Jermin, L.S., Graur, D., Lowe, R.M. and Crozier, R.H., Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. J. Mol. Evol., 1994, 39, 160–173.

15. Kimura, M.,1980. A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**, 111–120.
16. Knowlton, N. &Weigt, L. A., 1998. New dates and new rates for divergence across the Isthmus of Panama. Proc. R. Soc. Lond. B 265, 2257-2263. (DOI 10.1098/rspb.1998.0568.)
17. Knowlton, N., 1993. Sibling species in the sea. A. Rev. Ecol. Syst. 24, 189–216.
18. Kumar, S., Stecher, G., and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.Molecular Biology and Evolution 33:1870-1874.
19. Lakra, W.S., Goswami, M., Gopalakrishnan, A., 2009. Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: perciformes, Sciaenidae) based on 16S rRNA and cytochrome oxidase subunit I mitochondrial genes. Molecular Biology Reports, 36, 831–839.
20. Lakra, W.S., Sarkar, U.K., Dubey, V.K., Sani, R. and Pandey, A., 2011.River inter linking in India: status, issues, prospects and impli-cations on aquatic ecosystems and freshwater fish diversity. *Rev. Fish Biol. Fisheries*, **21**, 463–479.
21. Lipsey, M.K. and Child, M.F., 2007. Combining the fields of reintroduction biology and restoration ecology. *Conserv. Biol.*, 21, 1387-1388. doi: 10.1111/j.1523-1739.2007.00806.x
22. Margules, R.C. and Pressey, L.R., 2000. Systematic conservation planning. *Nature* (London), 405, 243-253. doi:10.1038/35012251.
23. Meyer, C. P. and Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol., 3, 2229–2238.
24. Mittermeier, R.A. and Mittemeier, C.G., 1997. Megadiversity: Earth's Biologically Wealthiest Nation. In: McAllister, D. E., Hamilton, A. L. and Harvery B. (Eds.). *Global Freshwater Biodiv.* Sea Wind, Cemex, Mexico City, 11, 1-140.
25. Mooers, A.O. and Holmes, E.C., 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol., 15, 365–369.
26. Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. Public Library of Science, 2, e354.
27. Okumus, I. and Ciftci, Y., 2003. Fish population genetics and molecular markers: II – molecular markers and their applications in fisheries and aquaculture. Turk. J. Fish. Aqcult., 3, 51–79.
28. Philippe, H. and Laurent, J., 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev., 8, 616–623.
29. Pusey, B.J., Arthington, A.H., Stewart-Koster, B., Kennard, M.J. and Read, M.G., 2010. Wide spread omnivory and low temporal and spatial variation in the diet of fishes in a hydrologically variable northern Australian river. *J Fish Biol.*, 77, 731-753. doi: 10.1111/j.1095- 8649.2010.02712.x.
30. Ruzzante, D.E., Taggart, C.T., Cook, C. and Goddard, S., 1996. Genetic differentiation between inshore and offshore Atlantic cod *Gadus morhua* off New found land: microsatellite DNA variation and antifreeze level. *Can. J. Fish. Aquat. Sci.*, 53, 634–645.
31. Saccone, C., DeCarla, G., Gissi, C., Pesole, G. &Reynes, A., 1999. Evolutionary genomics in the Metazoa: the mitochondrial DNA as a model system. Gene 238, 195–210.
32. Shaw, G.E. and Shebbeare, E.O., 1937. The fishes of North Bengal. *J. R. Asiat. Soc. Bengal*, III, 1–137.
33. Spies, I.B., Gaichas, S., Stevenson, D.E., Orr, J.W. and Canino, M.F., 2006. DNA based identification of Alaska skates (Amblyraja, Bathyraja and Raja: Rajidae) using cytochrome c oxidase submit (COI) variation. J. Fish Biol., 69, 283–292.
34. Steinke, D., Vences, M., Salzburger, W., Meyer, A., 2005.TaxI: a software tool for DNA barcoding using distance methods. Philosophical Transactions of the Royal Society, London, Series B, 360, 1847–1857.

35. Sengupta, S. and Homechaudhuri, S., 2013. DNA barcodes of some threatened freshwater indigenous fishes in India, Current Science, VOL. 105, NO. 1, 84.
36. Tajima, F., 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.
37. Talwar, P.K. and Jhingran, A., 1991.*Inland Fishes of India and Adja-cent Countries*, Oxford and IBH Publishing Co Pvt Ltd, New Delhi, two vols, pp. xix + 1158.
38. Tamura, K., Nei, M., and Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101:11030-11035.
39. Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) Software version 4.0. Mol. Biol. Evol., 24, 1596–1599.
40. Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. &Vogler, A. P., 2002. DNA points the way ahead in taxonomy. Nature 418, 479. (doi:10.1038/418479a.)
41. Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P., 2003. A plea for DNA taxonomy. Trends Ecol.
42. Thompson, J.D., Higgins, D.G. and Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence align-ment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
43. Ward, R.D., Zemlac, T.C., Innes, B.H., Last, P.R. and Hebert, P.D.N., 2005. DNA barcoding of Australia's fish species. *Philos. Trans. R. Soc. London, Ser. B*, **360**, 1847–1857.
44. Zhang, D.X. & Hewitt, G.M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol.Evol. 11, 247–251. (doi:10.1016/0169-5347(96)10031-8.)
45. Zhang, D.X. & Hewitt, G.M., 1997. Assessment of the universality and utility of a set of conserved mitochondrial primers in insects. Insect Mol. Biol. 6, 143–150.

Access this Article in Online	
	Website: www.ijarbs.com
Quick Response Code	Subject: Phylogenetics
DOI:10.22192/ijarbs.2022.09.09.001	

How to cite this article:
Adwitiya Chaudhuri. (2022). Delineating species boundary through DNA barcodes of some fishes inhabiting river Murti in the foothills of Darjeeling Himalayas. Int. J. Adv. Res. Biol. Sci. 9(9): 1-16.
DOI: http://dx.doi.org/10.22192/ijarbs.2022.09.09.001