Environment Conservation Journal 16(3) 109-116, 2015
ISSN 0972-3099 (Print) 2278-5124 (Online)
Abstracted and Indexed

Molecular characterization of coldwater fishes of district Uttarkashi, Uttarakhand using DNA Barcoding

Madhu Thapliyal[1], Hemlata Pokhriyal[1*], Bipin K. Sati[3*], N.S. Nagpure[2], Mahender Singh[2] and Ashish Thapliyal[3]

Received: 12.09.2015 **Accepted: 21.11.2015**

Abstract

We explored fish fauna Ganga and Yamuna river in Uttarkashi district of Uttarakhand, to establish the molecular taxonomy database. A total of 133 samples were collected from various sampling sites along the entire stretch of rivers in district Uttarkashi. A region of cytochrome c-oxidase unit I (COI) gene (655 bp) was amplified using PCR and sequenced (DNA Barcoded). All the sequences have been uploaded into the NCBI GenBank (accession number assigned). Analysis of data generated showed that these 133 samples belonged to 22 species of 4 orders, 14 genera and 8 families. The genetic variability (K2P distance) distribution analysis was also carried out. The average mean distance of 22 species is 0.219 with 0.014 standard error. The mean genetic distance between 22 species ranged from 0.010%-0.362% while the mean genetic distance within the species ranged from 0.0006%- 0.0048%. The lowest pairwise genetic distance observed in *Schizothorax sinuatus* and *Schizothorax progastus* i.e. 0.010% with 0.003 standard error indicating a closer phylogenetic relationship between *Schizothorax sinuatus* and *Schizothorax progastus* than other species which was confirmed by the genetic distance data. Maximum divergence were observed between *Danio devario* and *Channa gachua* i.e. 0.362% with 0.028 standard error. The maximum sequence divergence within the species is observed in *Barilius bendelisis* with 0.0048% while minimum sequence divergence is observed in *Tor putitora* with 0.0006 of cyprinidae family. Our data suggests that there is high inter-specific sequence divergence as compared to intra-specific sequence divergence and also conclude that COI sequencing (barcoding) was found to be suitable for the identification of fresh water fish species.

Keywords: DNA Barcoding, Species identification, biodiversity, molecular taxonomy, Garhwal Himalaya, Uttarkashi

Introduction

Garhwal Himalayas are located between latitude 29˚26′ N and longitude 78˚ to 88˚E with an area of 30,090 km^2. Uttarakhand, the northern part of India is endowed with vast water resources in the form of rivers, streams, canals etc. that are famous for its rich aquatic fauna. Uttarkashi district extends from 28˚43′ to 31˚ 27′ N latitude and 77˚34′E to 81˚02′ E longitude. The two of India's largest rivers, the Ganges and the Yamuna, originate in the glaciers located in Uttarkashi district of Uttarakhand, where they are fed by myriad lakes, glacial melts and several tributaries. Some of the tributaries are Alaknanda, Bhagirathi, Bhilangana, Mandakini, and Kosi. These rivers are snow fed in origin whereas some of the springs fed tributaries are Henwal, HemGanga, Song, Suswa and AsiGanga. Besides these there are hundreds of rivulets as tributaries.These freshwater systems of Garhwal Himalaya are famous for harboring rich aquatic biodiversity and impart valuable and important contribution to cold water fisheries. Though a lot of literature is available on the fish species that are endemic to Uttarakhand, some of these species exhibit similar morphometrics and so there identification is difficult. DNA barcoding is considered to be one of the best molecular tool used to characterize and identify any fish species.

In the past 15 years mitochondrial DNA (mtDNA) has attracted a lot of attention and has been used for several studies, especially related to population and evolution [Avise, 2004]. Mitochondrial molecular markers has become popular marker and dominated genetic studies designed to answer questions of phylogeny and population structure in fish for more than a decade. Identification of stocks and analysis of mixed fishery can be done with the help of mtDNA studies

Author's Address
[1]Pt. LMS Govt. PG College, Rishikesh, Uttarakhand, India
[2]DNBFGR, Lucknow, U.P.
[3]Deptt. of Biotechnology, Graphic Era Univ.,Dehradun
E-mail: madhuthapliyal@gmail.com
***Contributed equally**

which provide information on hybridization and introgression between fishes, serve as a genetic marker in forensics analysis and provide critical information for use in the conservation and rehabilitation programs (Billington, 2003). DNA barcoding has gained global support as a rapid, accurate, cost-effective, and broadly applicable tool for species identification, particularly with respect to fishes as coordinated by the fish barcode of life (FISH-BOL; www.fishbol.org) campaign (Ward et al. 2009). In DNA Barcoding a short specific DNA sequences has been standardized for identification of organisms at species level (Hebert *et al*., 2003) and has analogy to Universal Product Code (UPC) system used to identify good in retail stores. A region of mitochondrial cytochrome *c* oxidase I (COI) serves as the core molecular bio-identification marker for animals (De salle, 2006). Standardization of this technique at international level has made it a valuable tool for base line databases about species and identification of cryptic species. (DNA Barcoding is a term given by Paul Hebert in the year 2003). DNA Barcoding typically targets a large number of species and it can be a tremendous tool to accelerate species discovery and initiate new species descriptions (De salle *et al*., 2005; De salle, 2006; Lakra *et al*., 2011). Apart from study on taxonomic collection and listing diversity of fishes, not much work has been done on DNA Barcoding of endemic fish species in the upper Ganges and Yamuna region in district Uttarkashi. Attempts have been made to generate DNA barcodes but most of the attempts have been limited to major rivers i.e Ganga and Yamuna and sampling has been done only on few sites near the major town. There has been no systematic sampling at fixed distance intervals in this region mainly due to the remote locations of sampling sites. Since, to date, there is no detailed database based on molecular taxonomy scanning the entire river length, we explored fish diversity with the aim to generate a molecular database (DNA Barcodes) for fish species of Ganga and Yamuna rivers in Uttarkashi district of Uttarakhand.

Materials and Methods

Sample collection: A total of 133 samples collected from twelve different sampling stations located in two river systems, the Ganga and the Yamuna in district Uttarkashi from Uttarakhand and few samples were taken from other locations also between April 2012 to May 2014 (Fig. 1, Table 1).

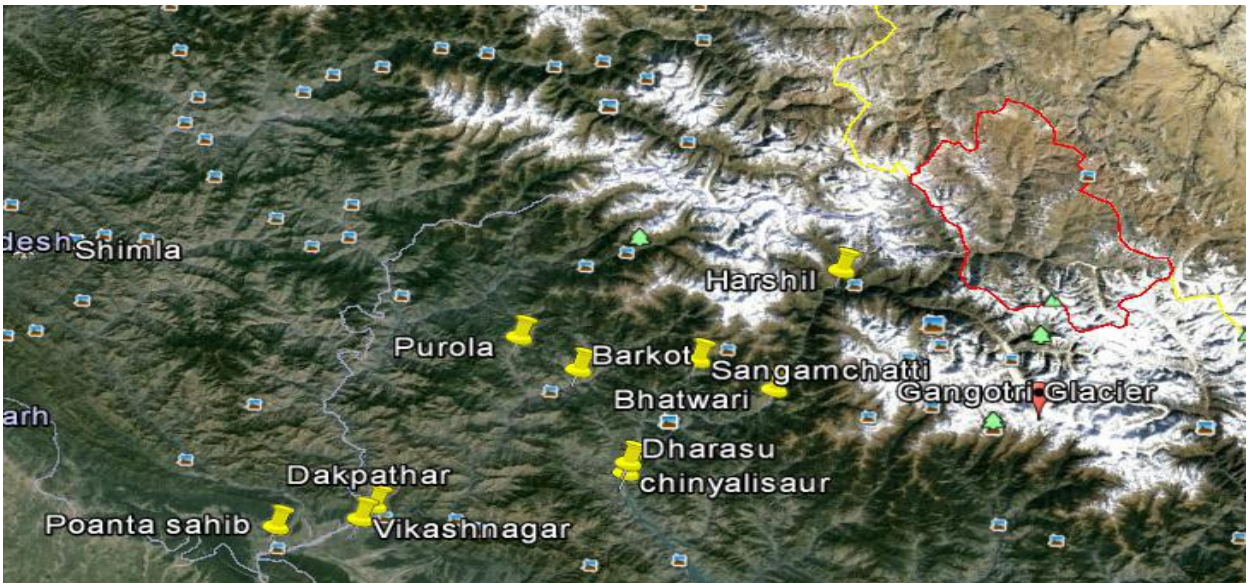

Fig. 1: Map of Ganga and Yamuna river showing sampling stations (Google earth). The sampling stations in river Ganges were Harsil, Bhatwari, Uttarkashi (Indravati), Maneri, Sangamchatti, Dharasu and Chinyalisaur. Sampling station in Yamuna were Mori, Barkot, Ponta Sahib, Naugaun and Dakpatthar.

Table 1: Species, Family, number of specimens, sampling sites and NCBI GenBank Accession number of specimens identified in the study.

Species	Family	No of specimens	Sampling sites	GenBank Accession No.
Acanthocobitis botia (AB)	Nemacheilidae	5	Poanta Sahib, Yamuna	KR809714, KU043312-KU043315
Badis badis (BB)	Badidae	8	Dakpathar, Yamuna	KR809715-KR809718, KU043316-KU043319
Barilius barna (BR)	Cyprinidae	5	Dharasu, Ganga	KR809719, KU043320-KU043323
Barilius bendelisis (BD)	Cyprinidae	5	Chinyalisaur, Ganga	KR809720-KR809721, KU043324-KU043326
Barilius vagra (BV)	Cyprinidae	7	Chinyalisaur, Ganga Dakpathar, Yamuna	KR809722-KR809728
Channa gachua (CG)	channidae	5	Poanta Sahib, Yamuna	KR809729, KU043327-KU043330
Channa punctata (CP)	channidae	5	Poanta Sahib, Yamuna	KR809730, KU043331-KU043334
Cyprinus carpio(CC)	Cyprinidae	6	Chinyalisaur, Ganga Barkot ,Yamuna	KR809731-KR809736
Danio devario (DD)	Cyprinidae	5	Poanta Sahib, Yamuna	KR809737-KR809738, KU043335-KU043337
Garra gotyla (GG)	Cyprinidae	5	Dharasu, Ganga	KR809739, KU043338-KU043341
Garra lamta (GL)	Cyprinidae	5	Chinyalisaur, Ganga	KR809740, KU043342-KU043345
Lepidocephalichthys guntea (LG)	Cobitidae	5	Poanta Sahib, Yamuna	KR809741-KR809742, KU043346-KU043348
Lepidocephalichthys sp. (LS)	Cobitidae	5	Poanta Sahib, Yamuna	KR809743, KU043349-KU043352
Mystus vittatus (MV)	Bagridae	5	Poanta Sahib, Yamuna	KR809744, KU043353-KU043356
Nemacheilus montana (NM)	Nemacheilidae	5	Dharasu, Ganga	KR809745, KU043358-KU043360

Pseudecheneis sulcata (PS)	Sisoridae	6	Indrawati, Ganga	KR809746-KR809748, KU043361-KU043363
Puntius chelynoides (PC)	Cyprinidae	15	Indrawati, Ganga Purola, Indrawati	KR809749-KR809763
Salmo trutta (ST)	Salmonidae	5	Harshil, Ganga	KR809764-KR809768
Schizothorax plagiostomus (SL)	Cyprinidae	5	Chinyalisaur, Ganga	KR809769, KU043364-KU043367
Schizothorax progastus (SP)	Cyprinidae	11	Sangamchatti, Ganga Naogaon, Yamuna	KR809770-KR809780
Schizothorax sinuatus (SS)	Cyprinidae	5	Dharasu, Ganga	KR809781-KR809782, KU043368-KU043370
Tor putitora (TP)	Cyprinidae	5	Chinyalisaur, Ganga Poanta Sahib, Yamuna	KR809783-KR809787

From the collected sample specimens, muscle tissue samples were cut from dorsal part and preserved in 95% ethanol and all the specimens were preserved in 10% formalin in the field as voucher (as specified by protocol of Barcode of life database (FishBOL).

Laboratory Procedures

DNA Extraction, PCR Amplification and Sequencing:

Total genomic DNA was isolated from muscle tissue by using the standard phenol-chloroform isolation protocol. Isolated DNA was checked on 0.7% Agarose Gel. Quantity of isolated DNA was checked in UV spectrophotometer (Nanodrop) by taking the optical density (OD) at absorbance values of 260 nm and 280 nm. Samples were amplified by Polymerase Chain Reaction (PCR) for COI gene using Primers: Fish F1 (5'– TCA ACC AAC CAC AAA GAC ATT GGC AC– 3') and Fish R1 (5'– TAG ACT TCT GGG TGG CCA AAG AAT CA – 3') (Ward et. al 2005). PCR amplifications were performed on a MJ research PTC-200 thermo-cycler in a 50µl reaction consisting of: 5µl of 10X buffer (100mM Tris, pH 9.0, 500mM KCl, 15mM MgCl , 0.1% Gelatin) (Genei, India), 200 µM each nucleotide (dNTP, Genei, India), 5pmole of each primer (Sigma Genosys, USA), 1.5U taq polymerase (Genei, India) and 1-2µl of total genomic DNA. The thermal regime consisted of an initial step of 3 min at 95°C followed by 34 cycles of 50s at 94°C, 1 min at 54°C and 45s at 72°C, followed in turn by 10 min at 72 °C. PCR products were stored at 4ºC. All PCR products were electrophoresed on 1.5% of Agarose gel followed by ethidium bromide staining **(Fig. 2)** and visualized under UV illumination in the Gel-Doc system (UVP) and purified using GeNeiTM Quick PCR purification kit (Genei, Bangalore, India) following the instructions given.

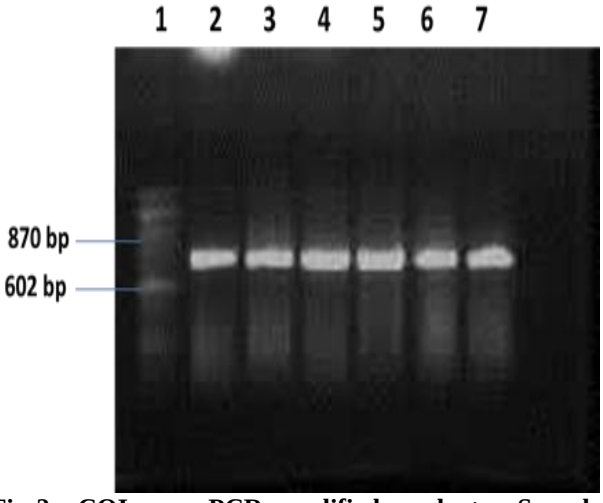

Fig 2 : COI genes PCR amplified products: Sample loaded in lanes are: 1 -DNA marker: -2-*Schizothorax progastus* : 3- *Tor putitora* : 4- *Salmo trutta* ; 5-*Barilius vagra*; 6-*Mystus vittatus* 7-*Badis badis*. These bands were used for sequencing after purification.

The most intense products were selected for sequencing. The cycle sequencing conditions was Initial Denaturation of 95° C for 30 s, annealing 50° C for 5 s and extension 60° C for 4 min repeated for 25 Cycles and then the samples were stored at storage temperature of 4° C.

Data Sequencing Analysis:

The DNA sequences were aligned using BioEdit sequence alignment editor version 7.5.2 (Hall, 1999) and Megalign 6.0 software. Phylogenetic and molecular evolutionary analysis was conducted using MEGA version 6 (Tamura et. al 2013). A neighbor-joining (NJ)and UPGMA dendrogram of K2P distances was created using MEGA v 6.0 software to provide a graphic representation of the patterning of divergence among species and bootstrapping was performed with 1000 replications. The genetic distances among and within species were calculated using the Kimura-2-Parameter (K2P) distance model (Kimura, 1980). The number and rate of transitions / transversions were also calculated using the program MEGA. All data (i.e electro-pherograms, primer details and collection localities) for each specimen and the assembled DNA sequences were also submitted to GenBank (List of NCBI accessions generated was given in Table 1).

Results and Discussion

In our study, we analyzed 133 samples collected from various sampling sites from Ganges and Yamuna in Uttarkashi district. These samples belonged to 22 species, 4 orders, 14 genera and 8 families. All amplified sequences were larger than 600bp and no insertions, deletions or stop codons were observed hence reducing the possibility of nuclear DNA. This also confirmed that the sequence originated from mitochondrial DNA. Nucleotide composition showed a CT bias A =24.7%, C= 28.0%, G= 18.2%, T= 29.0%. All sequences were deposited in GenBank (Accession no # KR809714-KR809787, KU043312-KU043370) and the barcodes, specimen and collection data, sequences, trace files and primers details are available in NCBI GenBank. A complete NJ tree of entire data is presented in fig.3a. The average mean distance of 22 species is 0.219% with 0.013 standard error. Pairwise genetic distance between the species is presented in table 2. The mean genetic distance among 22 species ranged from 0.010-0.362. The lowest pairwise genetic distance observed in *Schizothorax sinuatus* and *Schizothorax progastus* 0.010% with 0.003 standard error while the maximum divergence were observed between *Danio devario* and *Channa gachua* 0.362% with 0.028 standard error. They revealed a closer phylogenetic relationship between *Schizothorax sinuatus* and *Schizothorax progastus* than other species which was confirmed by the genetic distance data. Furthermore, the sequence divergence within these species is given in **table 3.** The maximum sequence divergence within the species is observed in *Barilius bendelisis* 0.0048% with 0.0018 standard error while minimum sequence divergence is observed in *Tor putitora* with 0.0006% of cyprinidae family. Our data suggests that there is high inter-specific sequence divergence as compared to intra-specific sequence divergence and also conclude that COI sequencing or 'barcoding' was found to be suitable for the identification of fresh water fish species. The COI based studies identification data shows that Ganga (Bhagirathi) and Yamuna River is having 22 species.

Neighbour-Joining Analysis of COI Barcode Sequences: The phylogenetic tree was constructed using Neighbour joining and Unweighted Pair Group Method with Arithmetic Mean (NJ and UPGMA) methods which were similar and the topologies of the both trees were identical, with 1000 bootstraps values. Both tree building methods (NJ and UPGMA) recovered each species as a monophyletic group (fig. 3a, fig 3b) which shows the major clades comprising the families *Nemacheilidae, Badidae, Cyprinidae, channidae, Cobitidae, Bagridae, Sisoridae, Salmonidae* placed each species separately indicating each 22 species. The N-J Analysis of COI gene showed that *Sshizothorax progastus* and *Schizothorax sinuatus* formed a monophyletic group (a result strongly supported by high bootstrap value of 98%) and then constitute one clade with *S. plagiostomus*; further they constitute another clade formed a different clusters of different species. The COI barcode is an effective tool for identification purposes. All species were resolved as monophyletic groups, despite low COI divergences between some individuals.

Table 2. Pairwise genetic distance (nucleotide Kimura 2 parameter) for cytochrome oxidase I (COI) gene sequences between species (Ganga and Yamuna) (Values above the diagonal line represent standard error and values below pairwise distance). The abbreviations in the table are the coded names of fish species as per table 1.

	AB	BB	BR	BD	BV	CG	CP	CC	DD	GG	GL	LG	LS	MV	NM	PS	PC	ST	SL	SP	SS	TP
AB		.024	.022	.023	.022	.021	.023	.021	.026	.022	.022	.021	.022	.023	.021	.023	.022	.025	.022	.022	.022	.020
BB	.275		.025	.025	.023	.024	.023	.024	.027	.024	.025	.024	.025	.028	.023	.025	.024	.024	.024	.024	.024	.025
BR	.238	.279		.005	.017	.023	.025	.018	.023	.021	.021	.023	.023	.024	.022	.022	.021	.023	.020	.020	.020	.020
BD	.248	.281	.019		.017	.023	.024	.018	.024	.021	.021	.023	.023	.024	.022	.022	.020	.023	.020	.020	.021	.021
BV	.234	.270	.175	.176		.024	.026	.017	.024	.019	.020	.022	.022	.026	.021	.021	.021	.025	.019	.018	.019	.020
CG	.255	.279	.285	.290	.291		.020	.023	.028	.025	.025	.023	.024	.024	.024	.023	.025	.023	.022	.023	.024	.023
CP	.285	.274	.273	.272	.286	.197		.023	.028	.025	.023	.023	.024	.024	.023	.023	.024	.024	.025	.025	.025	.023
CC	.199	.259	.209	.215	.190	.262	.253		.022	.017	.017	.022	.021	.023	.019	.021	.016	.022	.013	.012	.013	.015
DD	.315	.332	.288	.297	.291	.362	.343	.267		.024	.025	.025	.025	.027	.025	.027	.024	.028	.023	.022	.023	.023
GG	.218	.270	.223	.229	.197	.285	.285	.158	.277		.008	.022	.022	.023	.021	.022	.020	.025	.018	.018	.018	.020
GL	.218	.273	.221	.227	.214	.282	.253	.165	.301	.047		.021	.021	.020	.020	.021	.020	.025	.018	.018	.018	.018
LG	.225	.276	.254	.262	.238	.258	.264	.234	.299	.228	.231		.008	.024	.021	.023	.022	.024	.022	.022	.022	.020
LS	.231	.285	.255	.264	.236	.271	.280	.222	.301	.223	.226	.038		.024	.022	.024	.021	.024	.021	.021	.021	.020
MV	.253	.325	.257	.267	.297	.279	.273	.251	.325	.242	.202	.273	.275		.022	.018	.023	.022	.022	.022	.023	.021
NM	.215	.253	.240	.236	.215	.276	.264	.205	.305	.209	.202	.232	.235	.241		.022	.021	.024	.021	.021	.021	.020
PS	.265	.278	.260	.250	.239	.266	.266	.222	.323	.234	.217	.260	.267	.189	.235		.021	.023	.020	.020	.020	.022
PC	.224	.266	.220	.225	.231	.276	.269	.141	.282	.200	.198	.232	.231	.237	.225	.225		.024	.018	.017	.017	.012
ST	.279	.258	.271	.264	.293	.266	.261	.231	.350	.274	.270	.263	.269	.224	.268	.259	.261		.023	.023	.023	.023
SL	.225	.250	.223	.231	.192	.246	.272	.110	.280	.174	.179	.240	.221	.232	.221	.217	.160	.241		.004	.006	.017
SP	.222	.256	.217	.228	.193	.255	.275	.103	.275	.171	.174	.234	.213	.229	.219	.207	.155	.238	.013		.003	.016
SS	.226	.263	.224	.235	.198	.261	.278	.110	.288	.180	.183	.235	.214	.238	.225	.218	.164	.243	.022	.010		.017
TP	.202	.265	.217	.224	.216	.262	.261	.129	.280	.180	.174	.210	.209	.219	.214	.238	.088	.252	.155	.149	.159	

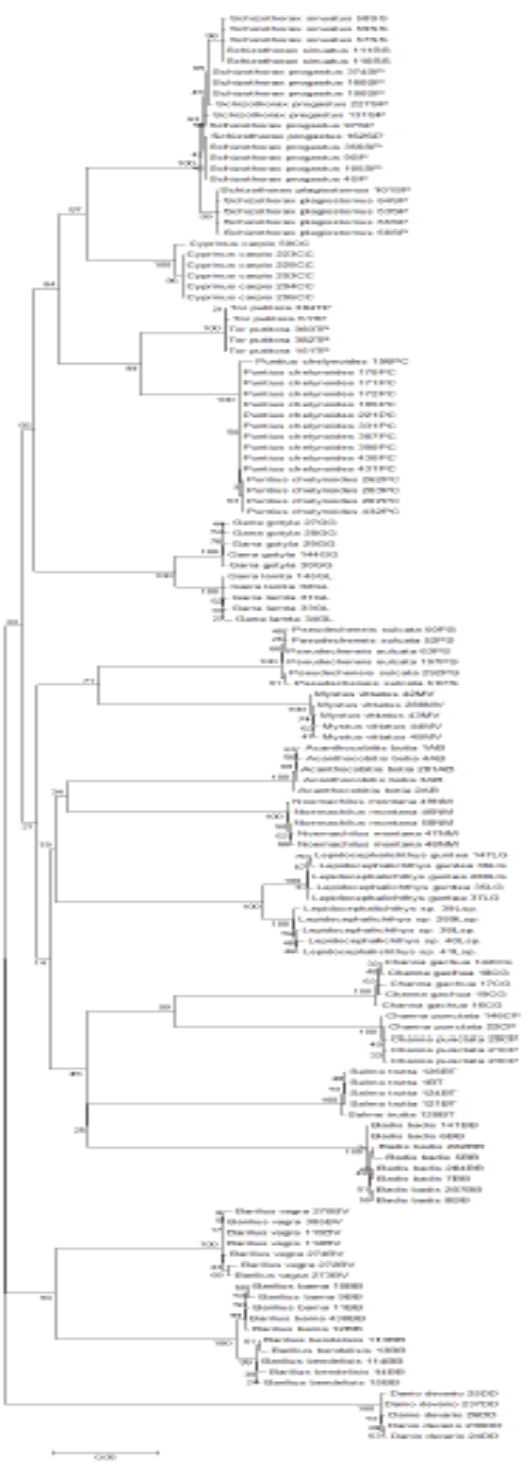

Fig. 3a: Neighbour-joining(NJ) tree of Kimura two-parameter (K2P) distances showing 22 analyzed species of 8 families from the river Ganga and Yamuna. Numbers above branches refer to bootstrap proportions among 1000 bootstrap replicates. Lowest pairwise distance observed in Schizothorax sinuatus and Schizothorax progastus indicating closer phylogenetic relationship between these two species and maximum pairwise distance is observed in *Danio devario* and *Channa gachua*

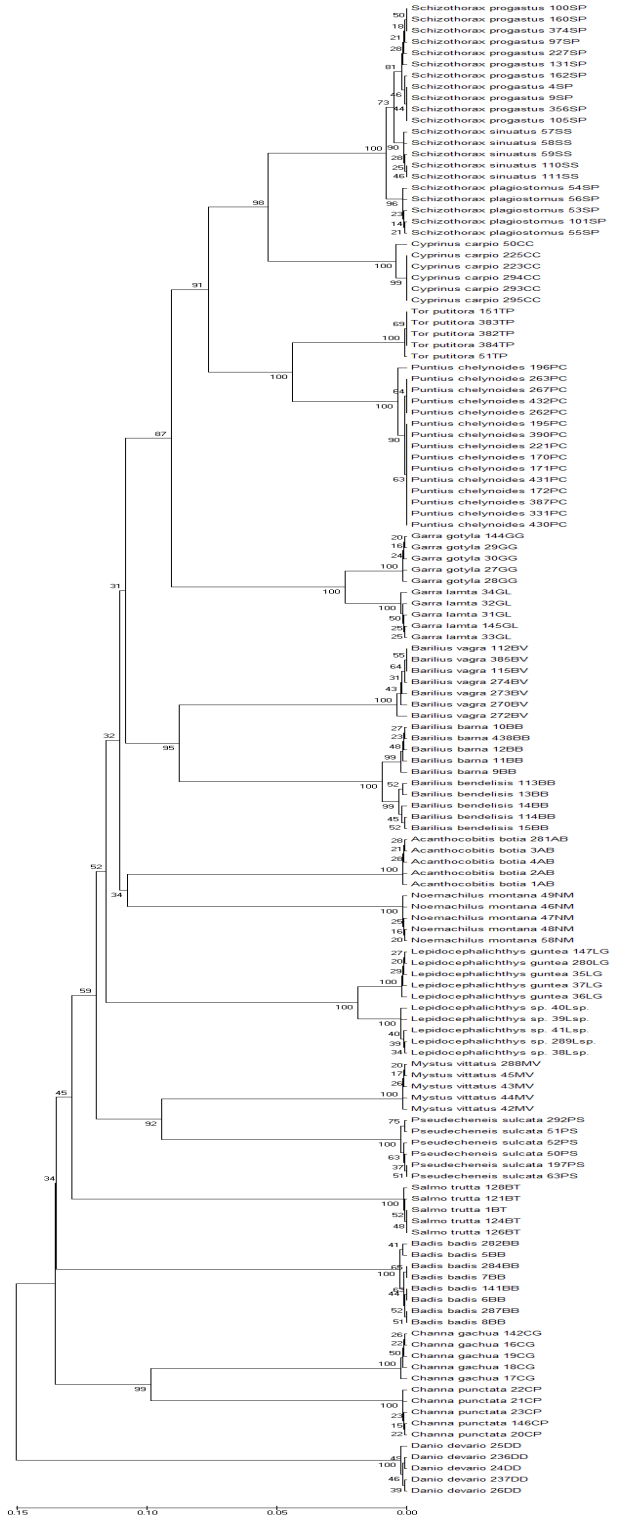

Fig. 3b: UPGMA (Unweighted Pair group Method with Arithmetic Mean) tree of Kimura two-parameter (K2P) distances showing 22 analyzed species of 8 families from the river Ganga and Yamuna. Numbers above branches refer to bootstrap proportions among 1000 bootstrap replicates

Conclusion

This study has ascertained the efficiency of COI barcodes for identifying species. COI based species identification study identified 22 species, with 0.219% mean distance among species. High COI sequence divergences existed between the species. Mean intra-specific and inter-specific COI sequence divergences differed by more than an order of magnitude. Extremely low sequence divergences between sister species and among species complexes are believed to be indicative of their recent origin. The study also suggests that probably, *S. progastus and S. plagiostamus and S. sinatus* are more abundant towards the upper reaches of Ganges any Yamuna. However this needs further investigations.

Acknowledgement

Sincere thanks to National Bureau of Fish Genetic Resources (NBFGR) Lucknow for lab resources in this study. Our sincere thanks to Uttarakhand council for Biotechnology, Govt. of Uttarakhand (Project No. 07/MT/RCUPG-Utk/NRIB/R&D-Project-1/2011) for financial assistant to Dr. Madhu Thapliyal.

REFERENCES

Avise, J.C. 2004. *Molecular Markers, Natural History and Evolution.* 2nd ed., Sinauer Associates, Sunderland.

Billington, N. 2003. Mitochondrial DNA. E. M. Hallerman (Ed.), *Population genetics: principles and applications for fisheries scientists*. American Fisheries Society, Bethesda, Maryland: 59-100.

DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: ***Response to Rubinoff. Cons. Biol***. 20, 1545-1547.

DeSalle, R., Egan, M.G., Siddall, M. 2005. *The unholy trinity: taxonomy, species delimitation and DNA barcoding*. Phil. Trans. R. Soc. B 360, 1905-1916.

Hall, T.A. 1999. *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series*, 41: 95-98.

Hebert, P.D.N.,Cywinska, A., Ball, S.L. & deWaard, J.R. 2003a. *Biological identifications through DNA barcodes*. Proceedings of the Royal Society of London, Series B 270: 313–321.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. ***Journal of Molecular Evolution***, 16: 111–120.

Lakra, W S, M S Verma, M Goswami, K K Lal, V Mohindra, P Punia, A Gopalakrishnan, K V Singh, R D Ward and P Hebert 2011. DNA barcoding Indian marine fishes. ***Molecular Ecology Resources*** 11(1): 60-71 (DOI: 1.1111/j.1755-0998.201.02894.x).

Park, L.K. and Moran, P. 1995. Developments in molecular genetic techniques in fisheries. G.R. Carvalho, and T.J. Pitcher (Eds.), *Molecular Genetics in Fisheries*. Chapman and Hall, London: 1-28

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6. ***Molecular Biology and Evolution*** 30: 2725-2729.

Ward R.D., R. Hanner and P.D.N. Hebert. 2009. The campaign to DNA barcode all fishes. FISH-BOL. ***Journal of Fish Biology*** 74:329-356.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. and Hebert, P.D.N. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B.* doi:1.1098/rstb.2005.1716 Published online.