

Phylogenetic relationship and molecular identification of five Indian Mahseer species using *COI* sequence

Jyoti Sati¹, Seema Sah¹, Himani Pandey², Shahnawaz Ali², Prabhati Kumari Sahoo²,
Veena Pande¹ and Ashoktaru Barat^{2*}

¹Department of Biotechnology, Kumaun University, Nainital-263 136, India

²Molecular Genetics Laboratory, Directorate of Coldwater Fisheries Research, Bhimtal-263 136, India.

*Corresponding Author email : abarot58@hotmail.com

Publication Info

Paper received:
26 September 2012

Revised received:
10 December 2012

Accepted:
20 February 2013

Abstract

This study examined the phylogenetic relationship and identification of five Mahseer species (*Tor putitora*, *Tor tor*, *Tor khudree*, *Tor chelynoidea* and *Neolissochilus hexagonolopis*) using partial sequencing of a Cytochrome Oxidase I (*COI*) DNA barcodes. The sequence analysis data showed that 134 (21.61%) sites out of 628 sites were variable without insertion or deletion. Rate of transition (70.5%) were higher than transversion (29.41%). There was a high inter-species divergence (range 4.1% to 12.2%) in Mahseer species as compared to intra-specific sequence divergence (1.7% for *T. putitora*, 1.2% for *T. tor*, 1.4% for *T. khudree*, 3.0% for *T. chelynoidea*, 0.26 % for *N. hexagonolopis*). The phylogenetic tree, constructed by maximum parsimony, maximum likelihood and unweighted pair group average methods revealed similar results suggesting that *T. putitora*, *T. tor* and *T. khudree* had a close relationship to each other while maximum divergence was observed in *T. chelynoidea*, which was also confirmed by the genetic distance data. The results indicate that *COI* sequencing or bar-coding is useful in unravelling phylogenetic relationship and identification of Mahseer species.

Key words

Bar-coding, Cytochrome Oxidase I, Genetic distance, Mahseer, Phylogenetic relationship

Introduction

Mahseer is the common name used for genera *Tor*, *Neolissochilus* and *Naziritor* belongs to family Cyprinidae. At present 46 Mahseer species are known of which 23 species belong of genus *Tor*, 22 species to genus *Neolissochilus* and one specie to genus *Naziritor* respectively (Eschmeyer *et al.*, 2004). Mahseers are important food as well as game fishes of the Asian region. They are endemic to Asia and are distributed across a number of countries such as Nepal, Pakistan, India, Sri Lanka, Malaysia and some South-east countries of Asia (Shrestha, 1990; Nguyen *et al.*, 2008). However, species characterization using morphology and anatomical characters sometimes cause errors in proper identification of closely related species. Because of these issues, molecular markers have been used as a complementary tool for taxonomic identification (Hebert *et al.*, 2003). Very little information is available regarding

species characterization using molecular tools in Mahseer species. A comprehensive literature survey revealed that Nguyen *et al.* (2008) studied Mahseer species using molecular tools and examined the phylogenetic relationships using mitochondrial gene. Mitochondrial DNA (mtDNA) is one of the widely used molecular markers for studying intra-specific and interspecies variation in animals (Avice, 1986; Billington and Hebert, 1988). The mtDNA data has provided new perspectives on taxonomically debatable taxa and confusing questions of phylogeny in which choice of gene is also of great significance (Simon *et al.*, 1994; Groves and Shields, 1996; Lunt *et al.*, 1996). Hebert *et al.* (2003) proposed that the analysis of sequence diversity in Cytochrome Oxidase I (*COI*) gene can provide an effective tool for species diagnosis. They argued that sequence diversity in this gene could be use to create a “barcoding” system that would enable the identification of all animal life. *COI* gene has a faster evolutionary rate (Simon *et al.*, 1994) and is

thus capable of providing a better resolution at the interspecific level. Hence, the present study aims to infer the species identification and phylogenetic relationships of different Mahseer species using *COI* sequences.

Materials and Methods

Sample collection: A total of 22 individuals of 5 species belonging to two genera *Tor* (*Tor putitora*, *Tor tor*, *Tor khudree* and *Tor chelynooides*) and *Neolissochilus* (*Neolissochilus hexagonolopis*) were collected from different geographically isolated location of India (Fig. 1, Table 1). Caudal fin samples of each fish were cut and placed in 2 ml vials containing 75% ethanol and voucher fish specimens immediately fixed in 4% formalin. The fin samples were kept at -20°C until DNA extraction.

Genomic DNA isolation: Genomic DNA was isolated from alcohol preserved 50mg fin tissue samples using proteinase k and phenol chloroform method (Sambrook et al., 1989). Isolated genomic DNA was precipitated with 2-2.5 volume of chilled ethanol. The DNA pellet was washed twice with

70% ethanol, air dried and resuspended in 1X TE (10mM Tris-HCl, pH 8.0 and 1mM ethylene diaminetetraacetic acid disodium salt) buffer and kept at -20°C till further use. The quality of DNA was checked by 0.8% agarose gel electrophoresis and the concentration of DNA was estimated in UV-VIS spectrophotometer (Thermo Scientific, England) at 260nm and 280nm absorbance.

PCR amplification and sequencing of PCR product: Partial sequence of Cytochrome Oxidase I (*COI*) gene was amplified by PCR (Eppendorf, Mastercycler gradient) using universal primers; CIF: 5'-AGTATAAGCGTCTGGGTAGTC-3' and COIA(L): 5'-CCTGCAGGAGGAGGAGAYCC-3' (Palumbi et al., 1991). Amplification was conducted in 50µl reaction volume containing 5µl of 10x PCR buffer (100mM Tris, pH 9.0, 500mM KCl, 15 mM MgCl₂, 0.1% Gelatin) (B-Genei, India) and 1 unit of Taq DNA polymerase (B-Genei, India), 200 µM of each dNTPs (dATPs, dCTPs, dGTP, dTTPs) (B-Genei, India), 25 pmol of each primers and 100ng of genomic DNA. The thermal profile used to amplify *COI* gene consisted of an initial denaturation of 95°C for 2 min; followed by 35 cycle of 94°C for 1 min, 54°C for 50 sec, 72°C for 50 sec and

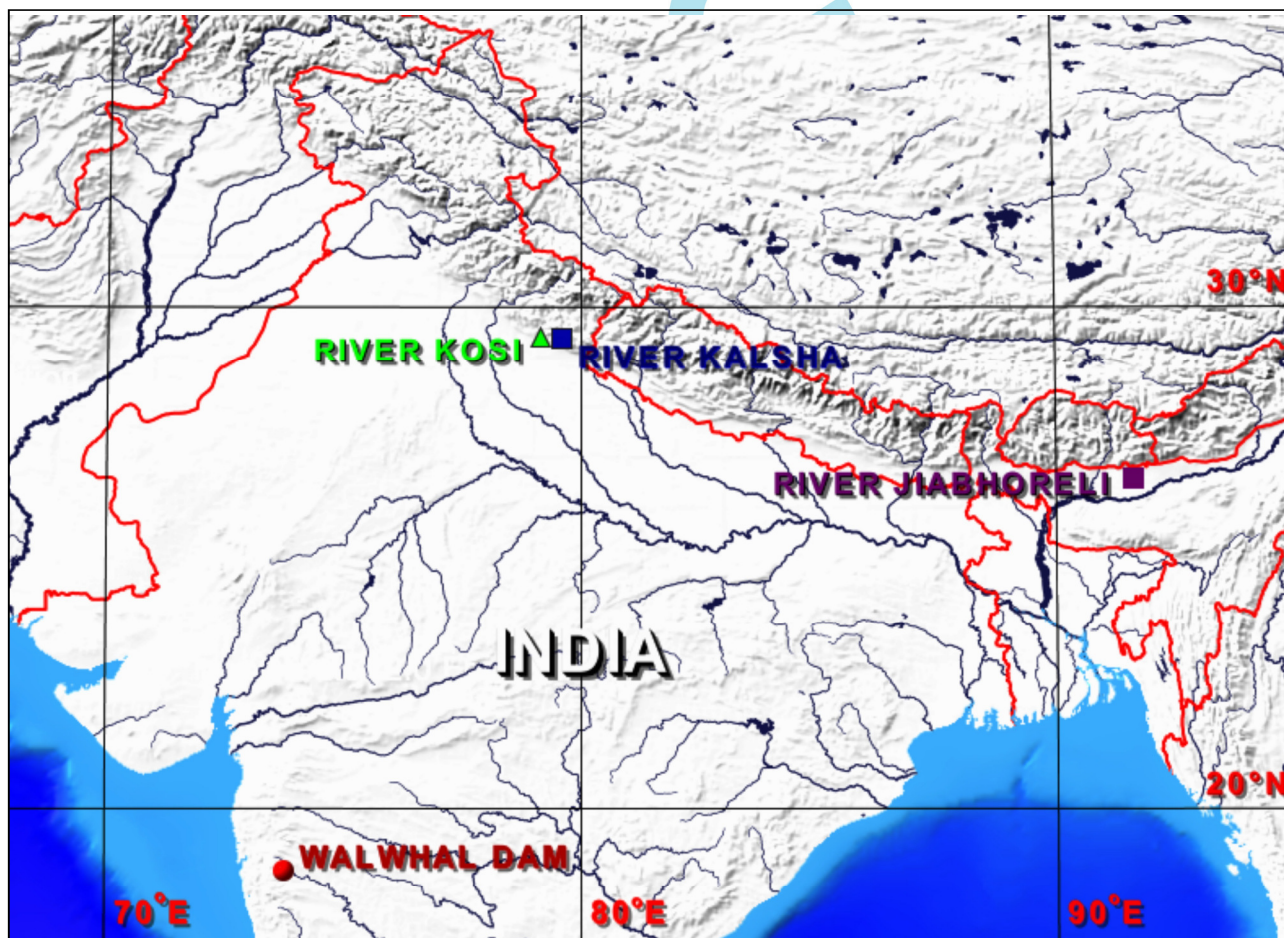


Fig. 1 : Sites of Mahseer species collection

Table 1 : Species, drainages, collection sites, number of haplotype and GenBank Accession numbers of specimens for this study

Species	No. of specimen	Drainages and collection sites	Coordinate location	Genbank accession numbers
<i>Tor putitora</i>	5	River Jia bhoreli, Bhalukpong (Assam-Arunachal Pradesh)	27°02'N; 92°35'E	JX204443-JX204447
<i>Tor putitora</i>	1	River Kosi, Ramnagar (Uttarakhand)	29°24'N; 79°07'E	JX204448
<i>Tor tor</i>	3	River Kosi, Ramnagar (Uttarakhand)	29°24'N; 79°07'E	JX204431-JX204433
<i>Tor khudree</i>	4	Walwhan dam, Lonavla (Maharashtra)	18°45'N; 73°45'E	JX204439-JX204442
<i>Neolissochilus hexagonolopis</i>	5	River Jia bhoreli, Bhalukpong (Assam-Arunachal Pradesh)	27°02'N; 92°35'E	JX204434-JX204438
<i>Tor chelynoides</i>	4	River Kalsa, Chanfi (Uttarakhand)	29°22'N; 79°34'E	JX204427-JX204430

a final extension at 72°C for 5 min. PCR products were checked in 1.2% agarose gel in 1X TBE (Tris-HCl, boric acid, EDTA, pH 8.0) buffer and visualized with ethidium bromide (Sambrook *et al.*, 2001) under UV-Gel-Documentation system (Alpha Imager 3400, Alpha Innotech Corporation, USA). Molecular weights were determined using 100 bp DNA marker (Fermentas, Canada).

The amplicons were purified before sequencing with Qiaquick columns (Qiagen, USA) as per manufacturer's instructions. Sequencing was performed in ABI Prism 3100 automated sequencer (Applied Biosystems, USA) using Bigdye terminator with same primers used for amplification of the target gene.

Sequence data analysis: The CHROMAS (Version 1.45) program was used to display the fluorescence based DNA sequencing analysis. The multiple sequence alignments were done using the CLUSTAL X program version 1.81 (Thompson *et al.*, 1997). Numbers of invariable, variable, singleton variable, parsimoniously informative sites and number of haplotypes were calculated using software DNAsp version 4.5 (Rozas *et al.*, 2003). Sequence divergence within and between the species were calculated using DNA STAR. The MEGA version 4.0 software (Tamura *et al.*, 2007) was used to construct the phylogenetic relationship among five species of Mahseer based on UPMGA, maximum-parsimony (MP) method. CLC genomics work Bench (version 5.1) was used to construct the maximum likelihood (ML) tree. Bootstraps support was calculated using 1000 replication.

Results and Discussion

The mtDNA *COI* gene of 628 bp length was successfully amplified and sequenced for 22 individuals in this study. The sequences obtained were aligned and compared with other GeneBank *COI* sequences. All the sequences representing *COI* gene were submitted to the GenBank with accession numbers given in Table 1. Empirical base frequency were A=27.2%, C= 23.9%, G= 18.0% and T=30.9%. Nucleotide sequences of *COI* gene in Mahseer species were A+T rich (58.1%) with anti-G bias of 18.0%, a

characteristics of the mitochondrial genome (Cantatore *et al.*, 1994), which were also reported in many fishes (Johns and Avise, 1998; Luhariya *et al.*, 2012).

The nucleotide sequences of five species of Mahseer were aligned to determine the variable sites (Table 2). The alignment data showed that 134 sites (21.61%) out of 628 were variable without any insertion or deletions. Among these 134 variable sites, 108 sites (80.59%) were parsimony information polymorphic while 26 sites (19.40%) were singleton variable sites. Most of the variable sites (98 sites, 73.13%) were in the first codon. It indicated several year of evolution involved in the genetic evolution of different cyprinid species (Springer and Douzery, 1996; Wang *et al.*, 2002; Barat *et al.*, 2012). A total of 153 mutations were observed in 134 sites. Rate of transition (108, 70.5%) were higher than transversion (45, 29.41%), which include all three codon position. A high transition bias is well known in vertebrate mtDNA (Meyer, 1993).

So far, 20 haplotypes were observed within these 5 species; 4, 3, 4, 6 and 3 haplotype in *T. khudree*, *T. chelynoides*, *T. putitora*, *T. tor* and *Neolissochilus hexagonolopis* respectively and no haplotype were shared by the five species. These haplotypes were used for further phylogenetic analysis.

The sequence divergence between the species (Table 3) ranged from 4.1% to 12.2%, showed close relationship between *T. tor* and *T. khudree* while maximum divergence were observed for *T. chelynoides*. Furthermore, average percentage divergence of individual species of *T. putitora* was 1.7%, *T. tor* was 1.2%, *T. khudree* was 1.4%, *T. chelynoides* was 3.0% and *N. hexagonolopis* was 0.26%. There was high inter-species sequence divergence for Mahseer species as compared to intra-specific sequence divergence. According to Hebert *et al.* (2003) intra-species divergence value were typically <3%.

The nucleotide sequences of *COI* gene were aligned in order to determine the phylogenetic relationships among five species of Mahseer. The topology of the ML, UPMGA and MP tree estimated were identical (Fig. 2). The

Nucleotide position →	4	5	6	7	8	9	1	1	1	2	2	2	2	2	2	3	4	4	4	5	5	5	5	6	6	7	7	7	7	7	8	8	8	8	9	9	9	0	0	1	2	2	2	3	3	1	1	1	1	1
Species↓	0	3	6	0	4	6	7	8	9	1	0	1	3	0	1	4	6	8	5	7	0	3	4	6	8	0	2	5	8	9	1	4	7	3	9	8	1	4	6	3	7									
<i>Tor khudree</i> 1	T	G	C	A	G	G	A	A	G	T	C	T	A	A	T	C	T	T	G	A	C	G	A	T	A	C	C	T	C	C	G	T	T	A	T	T	C	C	C	G	G	A	C							
<i>Tor khudree</i> 2	C	T	T	.	.	C	G	.	T			
<i>Tor khudree</i> 3	.	.	.	C	C		
<i>Tor khudree</i> 4	.	.	.	C	C		
<i>Neolissochilus hexagonolopis</i> 1	C	T	T	.	.	C	.	C	.	.	A		
<i>Neolissochilus hexagonolopis</i> 2	C	T	T	T	.	.	.	C	.	.	A		
<i>Neolissochilus hexagonolopis</i> 3	C	T	T	.	.	.	C	.	.	A	
<i>Neolissochilus hexagonolopis</i> 4	C	T	T	.	.	.	C	.	.	A	
<i>Neolissochilus hexagonolopis</i> 5	C	T	T	.	.	.	C	.	.	A	
<i>Tor chelynoidea</i> 1	C	.	.	G	.	G	C	C	T	.	A	.	T	A	.	.	G	.	C	T	T	T	A		
<i>Tor chelynoidea</i> 2	G	.	C	.	.	G	.	G	C	.	A	.	C	T	T	.	.	A	.	A	.	G	.	C	T	T	T	A	A	T			
<i>Tor chelynoidea</i> 3	C	C	.	.	G	.	G	C	C	T	.	A	.	G	T	A	.	.	G	.	C	T	T	T	A	
<i>Tor chelynoidea</i> 4	C	.	.	G	.	G	C	C	T	.	A	.	.	A	.	.	G	.	T	T	T	A		
<i>Tor putitora</i> 1	.	T	C	.	.	G	.	C	G	T	.	T	.	T	A	.	.	.	A
<i>Tor putitora</i> 2	G	C	A	G	C	.	.	G	.	C	T	G	C	.	T	.	T	.	T	A	.	.	.	A
<i>Tor putitora</i> 3	.	C	G	.	.	A	G	.	.	.	C	.	.	G	.	C	.</																																	

Journal of Environmental Biology, September 2013

[illegible]

	<i>Tor khudree</i>	<i>Tor tor</i>	<i>Tor putitora</i>	<i>Neolissochilus hexagonolopis</i>	<i>Tor chylenoides</i>
<i>Tor khudree</i>	-	4.1	5.2	16.2	11.4
<i>Tor tor</i>	0.04	-	4.7	6.2	11.6
<i>Tor putitora</i>	0.05	0.04	-	5.9	11.2
<i>Neolissochilus hexagonolopis</i>	0.06	0.06	0.05	-	12.2
<i>Tor chylenoides</i>	0.09	0.10	0.10	0.11	-

they also observed similar results. They observed close relationship between *T. tor*, *T. khudree* while *N. hexagonolopis* make a different cluster.

The mean genetic distance among five species of Mahseer ranged from 0.04 to 0.11 (Table 3). The lowest pair wise distance was observed between *T. tor* and *T. khudree* and between *T. tor* and *T. putitora* while the maximum divergence was observed for *T. chelynooides*. This revealed a closer relationship between the *T. tor*, *T. khudree* and *T. putitora*.

Finding of the present study showed high *COI* sequence divergence that occurred between the species and this study not only confirmed the capacity of DNA barcodes for species identification but also revealed deep divergence of *T. chelynoides* with other *Tor* species. The genetic data strongly indicated that *T. tor*, *T. khudree* and *T. putitora* belong to same genus (*Tor*) but unable to resolved

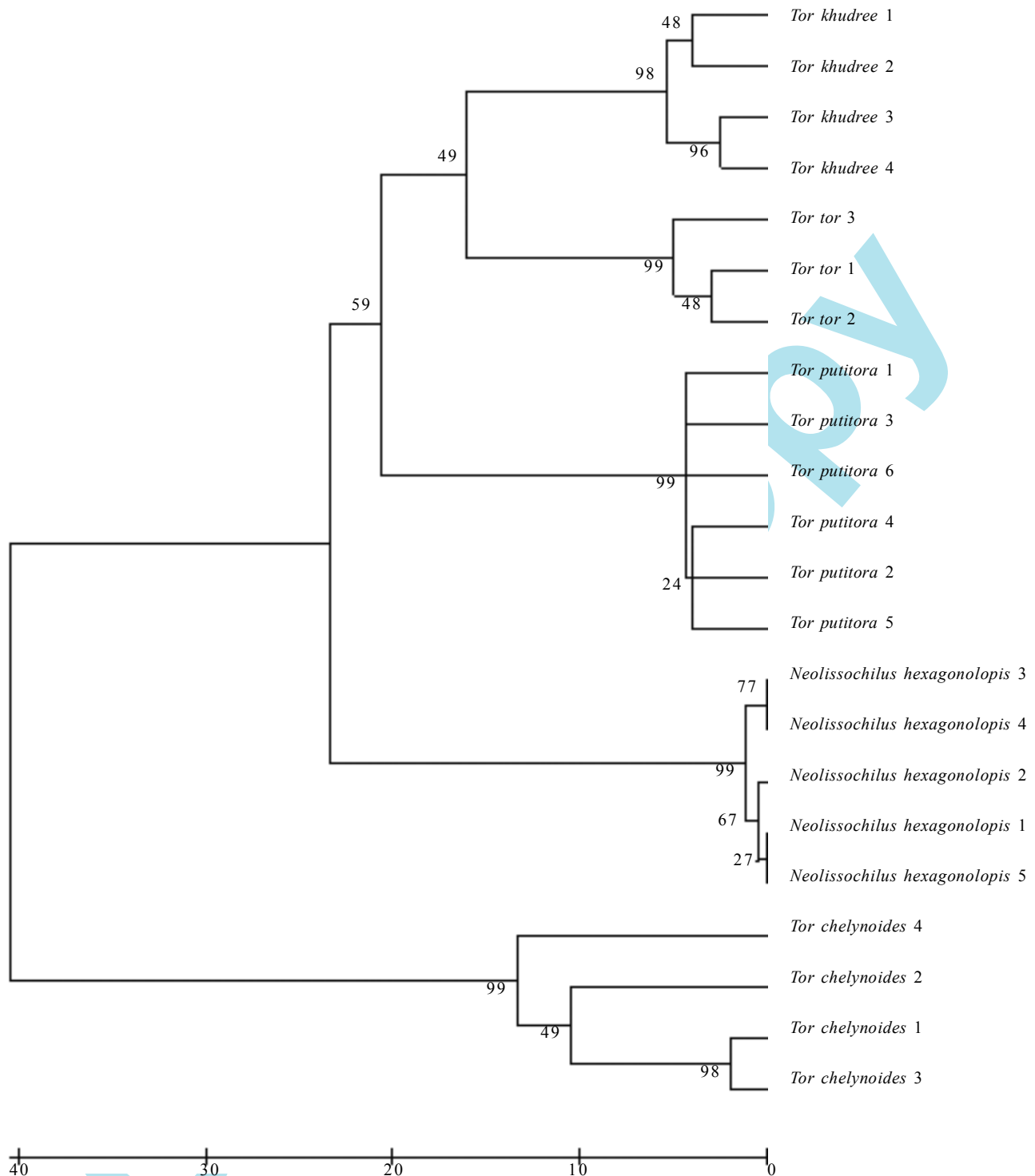


Fig. 2 : Phylogenetic tree (MP tree) based on 628 bp mitochondrial Cytochrome Oxidase I DNA sequences for five species of Mahseer

the relationship of *T. chelynoides* with other *Tor* species. Analysis based on Cytochrome Oxidase I gene are capable of discriminating Mahseer species with high accuracy, each species has a characteristics *COI* sequences that offers the

prospect to identify typical Mahseer species. Further studies are required to investigate the phylogenetic relationship of *Tor chelynoides* with other *Tor* species based on more mtDNA genes. This study should provide a

benchmark data for other studies.

Acknowledgments

The authors are thankful to Dr. P.C. Mahanta, Director, DCFR, Bhimtal for providing facilities and guidance. Thanks accorded to Indian Council of Agriculture Research, New Delhi, India for financial support to carry out this research work.

References

- Awise, J.C.: Mitochondrial DNA and evolutionary genetics of higher animals. *Phil. Trans. R. Soc. Lond.*, **312**, 325-342 (1986).
- Barat, A., S. Ali, J. Sati and G.K. Sivaraman: Phylogenetic analysis of fishes of the subfamily Schizothoracinae (Teleostei: Cyprinidae) from Indian Himalayas using Cytochrome b gene. *Indian J. Fish.*, **59**, 43-47 (2012).
- Billington, N. and P.D.N. Hebert: Mitochondrial DNA variation in Great Lakes walleye (*Stizostedion vitreum*) populations. *Can. J. Fish. Aquat. Sci.*, **45**, 643-654 (1988).
- Cantatore, P., M. Roberti, G. Pesole, A. Ludovico, F. Milella, M.N. Gadaletta and C. Saccone: Evolutionary analysis of Cytochrome b sequence in some perciformes: evidence for a slower rate of evolution than in mammals. *J. Mol. Evol.*, **39**, 589-597 (1994).
- Eschmeyer, W.N., Jr. C.J. Ferraries, M.D. Hoang and D.J. Long: The catalog of fishes, on-line, species of fishes, <http://www.calacademy.org/research/ichthyology/catalog/intro.html>. (2004).
- Groves, P. and G.F. Shields: Phylogenetics of the Caprinae on cytochrome b sequence. *Mol. Phyl. Evol.*, **5**, 467-476 (1996).
- Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. deWaard: Biological identification through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.*, **270**, 313-322 (2003).
- Jones, G.C. and J.C. Awise: A comparative summary of genetic distance in the vertebrates from the mitochondrial cytochrome b gene. *Mol. Biol. Evol.*, **15**, 1481-1490 (1998).
- Luhariya, R.K., K.K. Lal, R.K. Singh, V. Mohindra, P. Punia, U.K. Chauhan, A. Gupta and W.S. Lakra: Genetic divergence in wild population of *Labeo rohita* (Hamilton, 1822) from nine Indian rivers analyzed through mtDNA Cytochrome b region. *Mol. Biol. Rep.*, **39**, 3659-3665 (2012).
- Lunt, D.H., D.X. Zang, J.M. Szymura and G.M. Hewitt: The insect Cytochrome Oxidase I gene: Evolutionary patterns and conserved primers for phylogenetic studies. *Insect. Mol. Biol.*, **5**, 153-165 (1996).
- Meyer, A.: Evolution of mitochondrial DNA in fishes. *Biochem. Mol. Biol. Fishes*, **2**, 1-38 (1993).
- Nguyen, T.T.T., U. Na-Nakorn, S. Sukmanom and C. ZiMing: A study on phylogeny and biogeography of Mahseer species (Pisces: Cyprinidae) using sequence of three mitochondrial DNA gene region. *Mol. Phyl. Evol.*, **48**, 1223-1231 (2008).
- Palumbi, S.R., A. Martin, S. Romano, W.O. McMillan, C. Stice and G. Grabowski: The simple fool's guide to PCR, version 2.0. *Univ. Hawaii, Honolulu, HI*. pp.47. (1991).
- Rozas, J., J.C. Sanchez-DelBarrio, X. Messeguier and R. Rozas: DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497 (2003).
- Sambrook, J., E.F. Fritsch and T. Maniatis, Molecular cloning: A Laboratory Manual. 2ndEdn., Cold Spring Harbour Laboratory Press, New York. USA. (1989).
- Sambrook, J. and D.W. Russell: Molecular cloning: A Laboratory Manual. 3rd Edn., Cold Spring Harbour Laboratory Press, New York. USA. (2001).
- Shrestha, T.K.: Rare fishes of Himalayan waters of Nepal. *J. Fish Biol.*, **37**, 213-216 (1990).
- Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook: Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction "primers". *Ann. Entomol. Soc. Am.*, **87**, 651-701 (1994).
- Springer, M.S. and E. Douzery: Secondary structure and patterns of evolution among mammalian mitochondrial 12SrRNA molecules. *Mol. Biol. Evol.*, **43**, 905-911 (1996).
- Tamura, K., J. Dudley, M. Nei and S. Kumar: MEGA 4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **24**, 1596-1599 (2007).
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins: The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acids Res.*, **24**, 4876-4882 (1997).
- Wang, H.Y., M.P. Tsai, M.C. Tu and S.C. Lee: Universal primers of the complete mitochondrial 12SrRNA gene in vertebrates. *Zool. Stud.*, **39**, 61-66 (2002).