PLOS ONE

The Species Dilemma of Northeast Indian Mahseer (Actinopterygii: Cyprinidae): DNA Barcoding in Clarifying the Riddle

Boni A. Laskar, Maloyjo J. Bhattacharjee, Bishal Dhar, Pradosh Mahadani, Shantanu Kundu, Sankar K. Ghosh*

Department of Biotechnology, Assam University, Silchar, Assam, India

Abstract

Background: The taxonomic validity of Northeast Indian endemic Mahseer species, *Tor progeneius* and *Neolissochilus hexastichus*, has been argued repeatedly. This is mainly due to disagreements in recognizing the species based on morphological characters. Consequently, both the species have been concealed for many decades. DNA barcoding has become a promising and an independent technique for accurate species level identification. Therefore, utilization of such technique in association with the traditional morphotaxonomic description can resolve the species dilemma of this important group of sport fishes.

Methodology/Principal Findings: Altogether, 28 mahseer specimens including paratypes were studied from different locations in Northeast India, and 24 morphometric characters were measured invariably. The Principal Component Analysis with morphometric data revealed five distinct groups of sample that were taxonomically categorized into 4 species, viz., *Tor putitora*, *T. progeneius*, *Neolissochilus hexagonolepis* and *N. hexastichus*. Analysis with a dataset of 76 DNA barcode sequences of different mahseer species exhibited that the queries of *T. putitora* and *N. hexagonolepis* clustered cohesively with the respective conspecific database sequences maintaining 0.8% maximum K2P divergence. The closest congeneric divergence was 3 times higher than the mean conspecific divergence and was considered as barcode gap. The maximum divergence among the samples of *T. progeneius* and *T. putitora* was 0.8% that was much below the barcode gap, indicating them being synonymous. The query sequences of *N. hexastichus* invariably formed a discrete and a congeneric clade with the database sequences and maintained the interspecific divergence that supported its distinct species status. Notably, *N. hexastichus* was encountered in a single site and seemed to be under threat.

Conclusion: This study substantiated the identification of *N. hexastichus* to be a true species, and tentatively regarded *T. progeneius* to be a synonym of *T. putitora*. It would guide the conservationists to initiate priority conservation of *N. hexastichus* and *T. putitora*.

Introduction

The term 'mahseer' refers to a group of freshwater cyprinid fishes easily distinguishable by relatively larger size of scales on their body compared to the other cyprinid fishes [1,2]. The members of mahseer belong to two genera, viz., *Tor* and *Neolissochilus*. These two genera are distinguished by the presence of a continuous labial groove in *Tor* but interrupted in *Neolissochilus*, and 10–14 gill rakers on the lower arm of first gill arch in the former and 6–9 in the latter [3,4]. They inhabit in the mountain streams and distributed in the range from Pakistan throughout Southern Asia to Southeast Asia up to the Malay Peninsula and the larger Indonesian islands across Sumatra, Borneo and Java [5,6]. However, species composition within each genus varies in different locations, like Southeast Asian species are different from

Southern Asian species. Furthermore, within India, many species of mahseer are discontinuously distributed and mostly endemic in the South, Central and Northeast India. Among the mahseer of the Indian subcontinent, *Tor putitora* is widely distributed in Pakistan, India, Nepal and Bhutan; while *Neolissochilus hexagonolepis* is distributed in Nepal, Bhutan, North India and Northeast (NE) India [7,8]. A few studies suggest that the angling of mahseer provides superlative thrills than any other sport fishes except European Salmon [9,10]. They are highly sought-after because of great attraction to recreational anglers and are important components of the Angling-tourism pursuit [11]. In developing countries, there are many instances where the tourism industry has added recreational fishing to their attractions [12]. Owing to the growing value, the mahseer has become popular and considered as a cultural icon of diverse economic, recreation, and conservation

standpoint in rivers of eleven Asian nations [13]. Above all, the mahseer is an integral component of the aquatic ecosystem, serves as an important indicator of its health and supports the livelihood of many rural and indigenous ethnic groups in Asia [14]. However, the important mahseer fishes are threatened in the NE India as well as other distribution areas due to the growing harvest pressure as well as anthropogenic effects [15,16]. The two most threatened species, viz., *Tor putitora* and *Neolissochilus hexagonolepis* are regarded as the flagship species in NE India (http://www.nbfgr.res.in/). The conservation of mahseer has been hampered because the taxonomy of mahseer is most confusing due to the morphological variations they exhibit [17] that poised the understanding of actual species composition, distribution, autecology and biology at large.

Historically, with the pioneering work of Hamilton-Buchanan (1822) [1], many new descriptions of different species of mahseer have been proposed from Indian waters by distinguished naturalists. McClelland (1839) [18] recorded 4 new species from NE India, viz., *Tor progeneius*, *T. macrocephalus*, *Neolissochilus hexagonolepis* and *N. hexastichus*. McClelland, however, admitted difficulty in identifying Hamilton's *Cyprinus* (now *Tor*) *putitora* and particularly emphasized on a large cellular appendage to the apex of the lower jaw for *T. progeneius*, and the color gray on the back and reddish yellow on rest of the body for *N. hexastichus* [18]. The taxonomy of *T. progeneius* had long been in doubtful status, and it has been considered as a junior synonym of *T. putitora* [19]. Sen and Jayaram (1982) [20] characterized *T. progeneius* and elucidated with some new characteristics. Later, Rainboth (1985) [3] noted that *T. progeneius* is confusing to be classified whether within the genus *Neolissochilus* or *Tor*. It was further noted that most of the McClelland's type specimens were misplaced and some constituted curatorial nightmare [3]. Yet, McClelland's descriptions of two distinct species, viz., *Neolissochilus hexagonolepis* and *Tor progeneius* are recognized to be valid; while *T. macrocephalus* and *N. hexastichus* have been considered to be not valid rather the former was synonymized with *T. putitora* and the latter with *T. tor* [5,21].

Thus, the traditional taxonomy of mahseer in NE India has been facing several problems due to (1) lack of morphometric details in original description, (2) presence of very few holotypes of mahseer species, (3) indiscernible morphological nuances in them, and (4) disagreements in recognizing specific morphological characters. Consequently, the taxonomy of a few mahseer species has been extremely chaotic and described severally [2,4,5,20,21,22,23]. The mahseer species composition in the region is poorly understood and the identification of two species, viz., *T. progeneius* and *N. hexastichus*, has been difficult due to inconsistent taxonomic descriptions. Therefore, species level identification of mahseer is needed to be strengthened to facilitate the autecological study of mahseer and to develop conservation strategy for sustainable utilization in recreational fishing based tourism. Genomic approaches of taxon diagnosis have been found to be resourceful to aid traditional taxonomy [24,25]. In this context, the mitochondrial genome is a better target than nuclear genome because it evolved faster and can thus give more information to discriminate close species. Lately, a partial fragment of mitochondrial cytochrome oxidase c subunit I (COI) gene has been proposed to be sufficient singly to differentiate all, or at least the vast majority of animal species [26]. As such, this partial locus (COI) has been extensively tested for its efficacy in fish species identification and recognized as a unique marker of species identification with high confidence and called as "DNA barcode" [27,28,29]. The concept of DNA barcode based species identification is easy, rapid and accurate for being sequencing and web based; as such it has gained great attention worldwide [30,31,32]. Recently, the catfish diversity in NE India has been re-evaluated through DNA barcoding [33]. Therefore, morphological and DNA barcode data in combination can help to resolve the species dilemma of Northeast Indian mahseer, particularly *T. progeneius* and *N. hexastichus*, for effective conservation and management of the species.
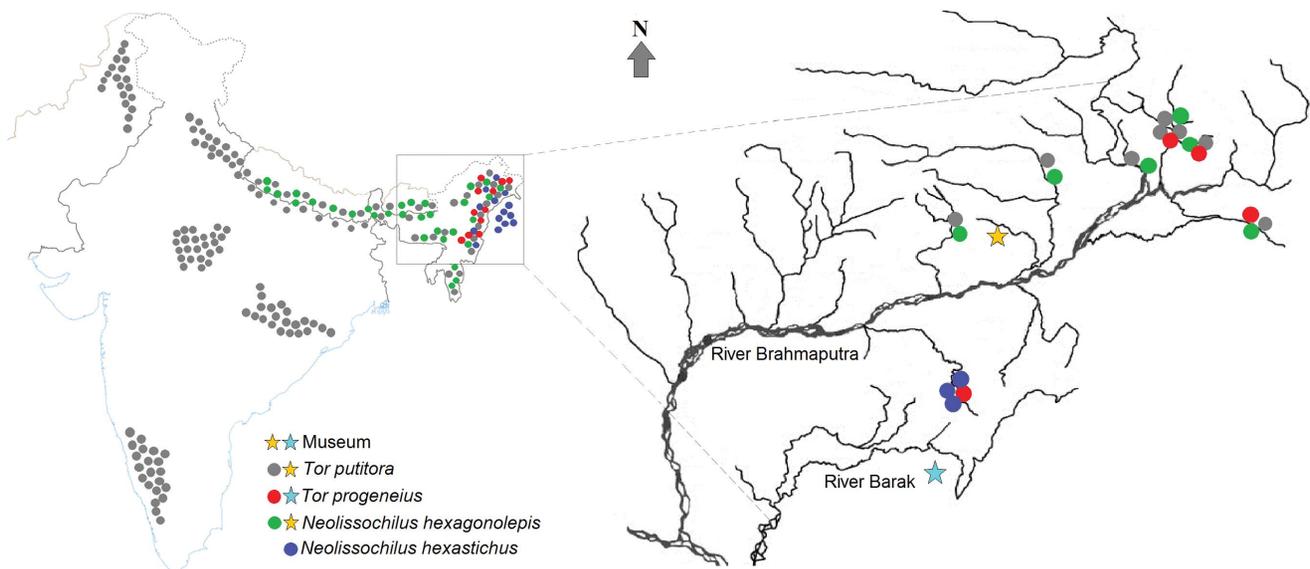
Figure 1. Map of the study site showing the known distribution of the studied species and the collection sites in different river drainages. The figure shows that the Northeastern region of India is drained mostly by River Brahmaputra and partly by River Barak. The studied specimens were collected from the drainages of River Brahmaputra. The topography of the region restricts the convergence of Southeast Asian fish composition with this region.
doi:10.1371/journal.pone.0053704.g001

Table 1. Morphological grouping of the studied organisms along with the corresponding codes.

Group	Nomenclature in practice	Sequence accession number used in molecular analysis/ catalogue number of paratypes in museum	Sample code used in morphological analysis (PCA)
N₂	*Neolissochilus hexastichus*	SGBL-BMF35	A
		JX127237	B
		JX127239	C
		JX127235	D
		JX127236	E
		JX127238	F
		SGBL-BMF36	*
N₁	*N. hexagonolepis*	JX127232	G
		JX127234	H
		JX127231	I
		JX127233	*
		** RGUMF-0036	V
		** RGUMF-0037	W
		** RGUMF-0038	X
T₂	*Tor progeneius*	JX127229	J
		***	K
		JX127228	L
		***	M
		JX127230	N
T₃	*T. putitora*	** RGUMF-0034	Aa
		JX127240	O
		JX127224	P
		JX127241	Q
		JX127242	U
T₁	*T. putitora*	JX127227	R
		JX127226	T
		JX127225	S
		** RGUMF-0035	Y
		***	Z
		***	Ab

The grouping was done based on scatter plot from Principal Component Analysis (PCA) as well as following the authoritative taxonomic keys. Sequence accession numbers in GenBank are used in the presentation of molecular analysis and the sample codes in PCA. Alphabetic sample codes replacing the full name of organisms are ascribed for ease of presentation those however clearly mentioned in Table S2.
*big specimen from market whose morphometric not done.
**paratypes from museum preserved in formaline whose sequencing not done.
***previously identified specimens preserved in formaline whose sequencing not done.
doi:10.1371/journal.pone.0053704.t001

Materials and Methods

Sample Collection

Fish specimens belonging to the group mahseer in the range of sub-adult to adult size were collected through participatory sampling with the marginal fishers engaged in commercial fishing. The specimens were from various locations in the hills and foothills across the Northeast India, particularly in the drainages of River Brahmaputra (Figure 1). The method of sample collection was approved by the Ministry of Science and Technology, Department of Biotechnology, Government of India (vide No. BT/HRD/01/002/2007). Some known voucher specimens within the genera *Tor* and *Neolissochilus* were examined from the Museum of Biodiversity in Rajiv Gandhi University, Arunachal Pradesh (voucher numbers are given in Table 1). The morphometrics of previously identified specimens from collection of *T. putitora* and *T. progeneius*, as well as the type specimens of *T. putitora* and *N. hexagonolepis* were included in the analysis. The type specimens of *T. progeneius* and *N. hexastichus* are not available in the museum. In lieu of examining type specimens of *T. progeneius* a small review on the existing contradictions among the taxonomists regarding the taxonomic descriptions and opinions on the status of the species is given in Supporting Information S1. Concerning the identification of *T. progeneius* and *N. hexastichus*, the original descriptions were emphasized. A total of 19 fresh specimens belonging to 4 species were studied in association with 5 paratypes and 4 previously collected specimens. Muscle tissue samples were invariably collected aseptically from behind of dorsal fin of the fresh specimens and taken in 500 µL of TES buffer (50 mM Tris HCl, 25 mM EDTA and 150 mM NaCl). The whole body
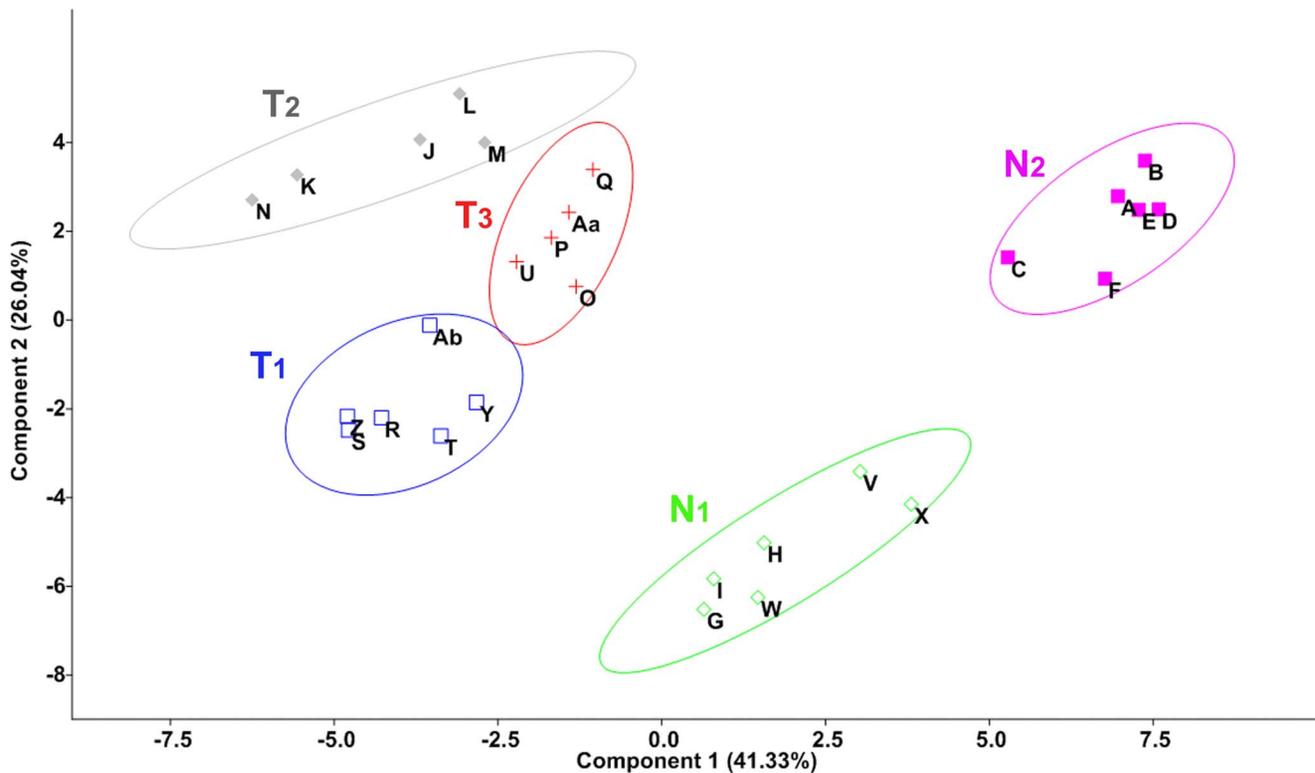
Figure 2. Principal Component Analysis (PCA) on 24 morphometric variables of the study samples including paratypes. The clusters of samples obtained from PCA were assigned to respective taxa based on meristic counts as well as non-quantitative characters of samples following authoritative taxonomic keys. The groups are like T_1, T_2 and T_3 comprising *Tor* congener, and N_1 and N_2 comprising *Neolissochilus* congener.
doi:10.1371/journal.pone.0053704.g002

specimens are preserved and stored at the Department of Biotechnology of Assam (Central) University, Silchar, Assam, India, for frequent examination and record of vouchers (vouchers' details are provided in Table S1).

Taxonomic Identification and Nomenclature

Specimens were categorized systematically based on the taxonomic characters available from the original description as well as subsequent re-descriptions and taxonomic reviews. Altogether 24 morphometric variables along with 6 important meristic counts were measured following standard literatures [23,34] (Figure S1) and the measurements were recorded using digital slide caliper (0.01 mm). The morphological characters those are non-quantitative yet taxonomically relevant, e.g. color pattern on the body and fins, presence or absence of tubercles, appearance and diagonal shape of mouth, etc. were also recorded from all the specimens. The measurements were taken at least three times independently and mode of each parameter was finally considered to minimize the error. The samples were designated into the respective species as per the authoritative taxonomic keys [4,20,23] and the species nomenclature was adopted as per the updated catalogue [8].

PCR Assay and Purification

DNA was extracted with standardized Phenol-Chloroform-Isoamyl alcohol method [35]. COI gene fragment (~655 bp) was amplified using the set of published primers: FishF1-5′TCAA-CCAACCACAAAGACATTGGCAC 3′ and FishR1-5′TAGAC-TTCTGGGTGGCCAAAGAATCA 3′[27]. The amplification was performed in 25 μl reaction mixture of 1X PCR buffer, 2 mM MgCl$_2$, 10 pmol of each primer, 0.25 mM of each dNTPs, 0.25 U high-fidelity polymerase and 100 ng of DNA template. PCR conditions were: initial denaturation at 94°C (2 minutes) followed by 30 cycles at 94°C (45 seconds), 50°C (45 seconds) and 72°C (1 minute), and a final elongation at 72°C (8 minutes). The PCR-amplified products were checked in 1% agarose gels containing ethidium bromide (10 mg/ml) and the single uniform band was then purified using QIAquick^R Gel extraction kit (QIAGEN, USA). The amplicons were bi-directionally sequenced in an automated DNA sequencer (ABI 3500, Applied Biosystems Inc., CA, USA).

Sequence Quality Control Measures

Two chromatograms that represent sequences of both the DNA strands were obtained for each sample. The PCR amplified products as well as their corresponding DNA sequences were larger than 600 bp that assured the sequences being not Numts as the limit of Numt hardly reaches 600 bp [36]. The noisy sequences were trimmed at both end and greater than 2% ambiguous bases were discarded, using quality value of >40 for bidirectional reads. BLASTN program was used to compare the sequences retrieved from the two chromatograms [37], and the fragment showing 100% alignment with no gap or indel (insertion/deletions) was selected. In some cases of discrepancy, both the sequences were reviewed and quality value of the sequences were considered to determine the most likely nucleotide using the software SeqScanner Version 1.0 (Applied Biosystems Inc., CA, USA). The selected fragments of the sequence were aligned using ClustalX software [38]. Finally, each of the sequences was compared in NCBI through BLASTN to examine

Table 2. Summary of PCA on 24 morphometric measurements of 28 samples within 4 species.

	PC 1	PC 2
% variance	41.336	26.04
Eigen value	19.4715	12.2662
Variable	Loadings	
SL	−0.3807	−0.1508
PrDL	−0.1599	0.2042
PoDL	0.1929	−0.1689
HtCF	0.298	0.1092
HL	−0.2624	0.3204
HtPF	0.1026	0.09338
HtDF	0.2399	0.3027
HtAF	0.1577	0.1663
HtDS	0.1129	0.1021
DP&V	−0.1125	−0.4824
LnCP	−0.1836	−0.1738
BDdf	0.4392	−0.1007
HDop	0.2464	−0.1186
HDe	0.1665	−0.03199
BWdf	0.1187	−0.1101
HWe	0.1814	0.00083
SnL	−0.1136	0.1563
ED	0.0623	0.07016
LnLF	−0.182	0.363
LHtCP	0.1185	−0.00832
HtVF	0.1842	0.07278
DVF&AF	−0.08024	−0.4201
LnBDF	0.2209	−0.09064
LnBAF	0.03401	−0.05084

Proportion of variance, Eigen values and coefficients (loadings) of the first two principal components (PC1 and PC2) for the % total length of the morphometrics of studied mahseer species.
doi:10.1371/journal.pone.0053704.t002

the complete alignment with the partial coding sequence of fish mitochondrial COI gene. The sequences were translated using the online software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and aligned through BLASTP to examine whether the partial amino acid codes were coherent with the fish mitochondrial COI gene frame and without any stop codon. In this way, the generated sequences were confirmed to be the fragments of mitochondrial COI gene. All the analyzed sequences were then deposited in GenBank (details of accession numbers are given in Table S1). The sequences were also submitted in a FISH-BOL project entitled "DNA barcoding of Mahseer fishes from Northeast India" and the code name 'MFISH'.

Data Analysis

Morphometry. Principal Component Analysis (PCA), a multivariate statistical procedure commonly used to reveal patterns in measured correlated variables, was used to differentiate the samples into possible groups and any variation among the samples of same species and the paratypes. The morphometric measurements were transformed into percentage of the total body length to develop the relative data of each variable for the samples

of different size and species. The analysis was performed using PAST version 2.17 b (http://folk.uio.no/ohammer/past). The PCA output is presented as scatter plot showing the groups of the samples with designated codes.

COI sequence data analysis. The sampled specimens were invariably sequenced and their congeneric sequences were acquired from the databases (GenBank and BOLD) to examine the level of intraspecific variation. Most of the database sequences lack geographical information yet they were assumed to be at least from distant locations. The analysis was based on a total data set of 76 COI barcode sequences of mahseer containing 21 *denovo* sequences and 55 database sequences. Additionally, 2 sequences of *Hypsibarbus wetmorei* and 3 sequences of *Puntius sarana* were acquired from GenBank to represent the out-group in the study. Geographical information and GenBank accession numbers of the developed as well as acquired sequences are given in Table S1. The calculation of Kimura 2-parameter (K2P) congeneric and conspecific distance [39] as well as phylogenetic analysis through Neighbor Joining (NJ) method were performed using MEGA Version 5.1 [40]. The tree topology obtained through NJ method was double-checked by Maximum Likelihood (using MEGA Version 5.1) and Bayesian approach (using MrBayes 3.2.0) [41].

Results

Morphological Characteristics

The PCA yielded 24 components which correspond to the 24 morphometric measurements. Projection of the morphometric data of studied mahseer species on first 2 principal axes showed the separation of the samples into 5 groups at 75% concentration ellipse level (Figure 2). The first 2 principal components contributed to 67.37% of total variance (PC1 = 41.33% and PC2 = 26.04%) (Table 2). The third, fourth and fifth components contributed to 8.57%, 4.93% and 3.55%, respectively, but did not improve the separation of the samples. These 5 groups were categorized into 2 broad groups and each corresponds to a genus, as per the authoritative taxonomic keys. The meristic count of the samples is presented in Table 3 which depicts a prominent difference in number of gill rakers on the lower arm of first arch between the two genera. The rakers were 8–9 in *Neolissochilus* and 13–14 in *Tor*. The other meristics were almost similar in all the samples. In the PCA scatter plot, the samples within the genus *Neolissochilus* further formed two distinct groups, one of which grouped with the paratypes of *N. hexagonolepis* but the other group stood distant indicating both the groups belonging to different species. The samples within the genus *Tor* appeared to be in a single but very stretched out group indicating a wide range of variation. In this group, some samples formed two slightly distant groups, yet each of the groups assembled with at least one of the paratypes of *T. putitora* while the rest few samples formed a slightly separate group and remained away from the paratypes. The non-quantitative characters of samples within *Tor* and the prevailing taxonomic descriptions suggested two possible species name. The groups of samples appeared in PCA were designated as T_1, T_2 and T_3 comprising *Tor* congener, and N_1 and N_2 comprising *Neolissochilus* congener. The constituent samples within each group were given the alphabetic sample code (Table 1), like S, R, T, Y, Z and Ab fall within T_1; J, K, L, M, and N fall within T_2; Aa, O, P, Q and U fall within T_3; G, H, I, V, W and X fall within N_1; and A, B, C, D, E and F fall within N_2. The meristic counts and morphometric data are given in Table 3 and supplementary Table S2 respectively.

Table 3. Important meristic counts of three specimens in each species.

Organism name (Species)	Replicates	Parameters Gill rakers on first arch (upper arm+lower arm)	Scales on lateral line	Dorsal fin rays	Ventral fin rays	Pectoral fin rays	Anal fin rays
Tor putitora	a	2+14	25	9+ii	8+i	15+i	6+i
	b	2+14	25	9+ii	8+i	15+i	6+i
	c	2+14	25	9+ii	8+i	15+i	6+i
Tor progeneius	a	2+14	26	9+ii	8+i	15+i	6+i
	b	2+13	26	9+ii	8+i	15+i	6+ii
	c	2+14	26	9+ii	8+i	15+i	6+i
Neolissochilus hexagonolepis	a	2+8	27	9+ii	8+i	14+i	6+i
	b	2+8	27	9+ii	8+i	14+i	6+i
	c	2+8	27	9+ii	8+i	14+i	6+i
Neolissochilus hexastichus	a	2+9	24	10+ii	7+i	14+i	7+i
	b	2+9	24	10+ii	7+ii	14+i	7+i
	c	2+9	24	10+ii	8+i	14+i	7+i

The table shows that the number of gill rakers on the lower arm of first arch is a very important distinguishing character between the two genera. This character is very easily identifiable and based on this character the first hand classification of mahseer in to respective genera can be easily done.
• Lowercase roman numerals are used to denote the simple rays in fin ray count.
doi:10.1371/journal.pone.0053704.t003

● GQ469780|T.m.mahanadicus] T. mmahanadicus
GQ469802|T.putitora] T. putitora
● HQ609722|T.m.mahanadicus] T. mmahanadicus
GQ469808|T.putitora] T. putitora
90 ● GQ469827|T.macrolepis] T. macrolepis
GQ469806|T.putitora] T. putitora
GQ469804|T.putitora] T. putitora
94 ● GQ469828|T.macrolepis] T. macrolepis
● GQ469829|T.macrolepis] T. macrolepis
GQ469803|T.putitora] T. putitora

● JX127226|T.putitora
64 ● JX127224|Tputitora
77 ● JX127225|T.putitora
80 ● JX127227|T.putitora
39 71 ▲ JX127228|T.putitora
62 ▲ JX127229|T.putitora
55 ● JX127240|T.putitora
● JX127242|T.putitora
44 ▲ JX127230|T.putitora
● JX127241|T.putitora
86

T. putitora

18 EU714114|T.tor
EU714115|T.tor
65 EU714112|T.tor
EU714111|T.tor
96 EU714116|T.tor
66 EU714117|T.tor
73 EU714120|T.tor
76 EU714118|T.tor
72 EU714119|T.tor
33

T. tor

JQ665787|T.tambroides
JQ665788|T.tambroides
99 JQ665789|T.tambroides
JQ665790|T.tambroides
JQ665791|T.tambroides
59

T. tambroides

GQ469790|T.khudree
GQ469791|T.khudree
99 GQ469789|T.khudree
GQ469788|T.khudree
GQ469787|T.khudree

T. khudree

EU714109|T.mosal
EU714110|T.mosal
99 EU714108|T.mosal
15 EU714107|T.mosal
39 EU714106|T.mosal
17

T. mosal

JN646100|T.douronensis] T. douronensis
HM536900|T.sinensis] T. sinensis
99
26 99 HM585023|T.malabaricus] T. malabaricus
HM585024|T.malabaricus
41 GQ469800|T.mussullah
GQ469801|T.mussullah
99 GQ469799|T.mussullah] T. mussullah
GQ469797|T.mussullah
GQ469798|T.mussullah

RCYY293-11|N.stracheyi] N. stracheyi

77 ● JX127235|N.hexastichus
42 ● JX127236|N.hexastichus
88 ● SGBL-BMF36|N.hexastichus
● JX127237|N.hexastichus
26 ● JX127238|N.hexastichus
92 24 ● SGBL-BMF35|N.hexastichus
● JX127239|N.hexastichus
76

N. hexastichus

17 EU714098|N.hexagonolepis
14 EU714100|N.hexagonolepis
32 EU714096|N.hexagonolepis
59 EU714099|N.hexagonolepis
67 ● HM536922|N.stracheyi] N. stracheyi
EU714097|N.hexagonolepis] N. hexagonolepis
99

N. hexagonolepis

98 99 ● JX127232|N.hexagonolepis
95 ● JX127233|N.hexagonolepis
● JX127234|N.hexagonolepis
44 FJ459521|N.hexagonolepis
34 FJ459520|N.hexagonolepis
21 FJ459524|N.hexagonolepis
● JX127231|N.hexagonolepis
33 FJ459523|N.hexagonolepis
43 FJ459522|N.hexagonolepis

N. hexagonolepis

99 RCYY145-11|H.wetmorei] H. wetmorei
RCYY146-11|H.wetmorei

JQ667570|P.sarana
99 FJ459432|P.sarana] P. sarana
92 FJ459431|P.sarana

0.01

Figure 3. Neighbor Joining (NJ) tree developed using K2P distance among 76 COI sequences of mahseer. In most cases the studied samples (marked as red dots) showed cohesive clustering with their conspecific database sequences. The samples morphologically identified as *Tor progeneius* (accession numbers JX127229, JX127230 and JX127228) clustered conspecific with developed sequences as well as database sequences of *T. putitora*. All samples of *N. hexastichus* clustered cohesive as the same species and distinct from *N. hexagonolepis* sequences. Few database sequences of 3 species revealed aberrant clustering {*Neolissochilus stracheyi* (accession number HM536922), *Tor mosal mahanadicus* (accession numbers HQ609722, GQ469780), *Tor macrolepis* (accession numbers GQ469827-29)}. ● The numbers at the nodes are bootstrap values based on 1000 replications. ● The specimens' GenBank accession number and species name are shown for each taxon. ● Red dots and black triangles correspond to the sequences developed in this study. Black triangles also correspond to the sequences of samples, although morphologically identified as *Tor progeneius* but were found conspecific with *Tor putitora* based on COI sequence data analysis. Black dots correspond to the cases of abnormal clustering.
doi:10.1371/journal.pone.0053704.g003

Tor Congener

The first taxonomic key to differentiate species within *Tor* is based on the relative head length to the body depth. In this study, all *Tor* congener possessed slightly longer head than body depth. There was no stringent variation in meristic counts among the *Tor* congener (Table 3); and both T_1 and T_2 samples were similar in most of the other taxonomic features (Table S2). But, T_3 differed from T_1 and T_2; firstly on having long (up to the margin of maxilla) mental lobe (also called lower labial flap) vs. short/absent, and secondly on having longer upper jaw and with skin like flap extending behind upper lip vs. both the jaws equal and upper lip without a flap. The mental lobe length was 4.29% to 5.91% of total length in T_3 samples vs. 1.47% to 1.99% in T_1 and T_2 samples. The longer upper jaw in T_3 samples correspondingly shared to greater head length and snout length than in T_1 and T_2 samples. It appeared that the presence of both upper and lower lips as being relatively more fleshy and the mental lobe being prominent and long in all the samples of T_3 differentiate them from T_1 and T_2 samples. The observed morphological features of all the samples within T_1 and T_2 bear close affinity with the described features of *Tor putitora*. Therefore, despite minor differences, T_1 and T_2 samples were considered to be belonging to the same species and named accordingly. The particular lip character in T_3 samples resemble with the original descriptions of *Tor progeneius*. It was observed that the three groups though bear minor variation in morphometrics but they are not discernible except the particular differentiating features of *Tor progeneius* that appeared to be unique and very much noticeable. Thus, T_3, T_1 and T_2 samples are tentative considered as morphs and the former is designated as long mental lobed while the latter 2 as short mental lobed.

Neolissochilus Congener

The N_1 samples were distinguished from N_2 due to absence of mental lobe vs. prominent, and interrupted groove behind the lower lip vs. continuous groove. Both these features of N_2 samples resembled with the *Tor* congener. But, the gill rakers in them were 9 vs. 13–14 in *Tor* congener. Tubercles were mostly present on the cheeks in N_1 but entirely absent in N_2. Mouth smoothly rounded in N_2 vs. truncate in N_1, edge of lower jaw blunts in N_2 vs. sharp in N_1. The color of the back, bases of caudal and dorsal as well as the upper part of the head in N_2 samples was greenish gray, reddish yellow on rest of the body, and the tips of the fins red. These observed features in N_2 samples have been originally emphasized
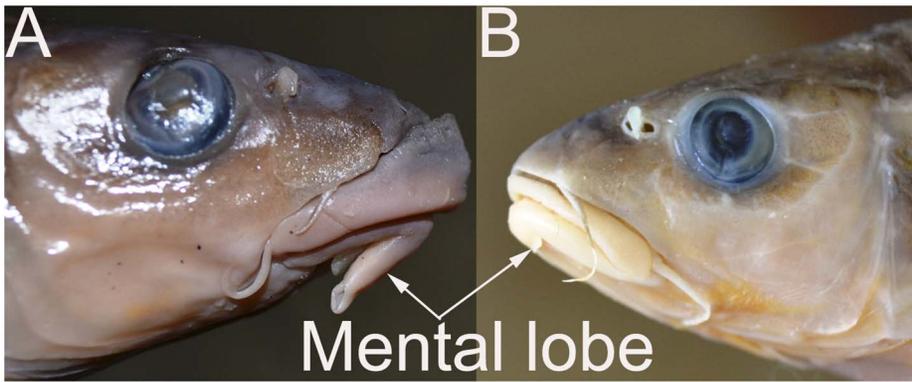
Figure 4. Illustrating the counter appearances of mouth in (A) *T. progeneius* **and (B)** *T. putitora*. Showing fleshy lips, a semicircular flap extending behind the upper lip, and a fleshy appendix extending from the lower lip up to the margin of maxilla (very long mental lobe) in (A), but absent in (B).
doi:10.1371/journal.pone.0053704.g004

to describe *N. hexastichus* as a distinct species. Thus, the N_2 samples were named as *N. hexastichus* according to the authoritative descriptions. The morphometrics of N_1 samples were mostly similar to both *N. hexagonolepis* and *N. stracheyi*. The N_1 samples were uniquely identified to be *N. hexagonolepis*, based on color pattern having scales coppery colored with a tinge of red above lateral line and fins deep slate paling towards their margins. As per the prevailing taxonomic description, *N. hexagonolepis* is different from *N. stracheyi* due to the absence of a lateral black stripe as in N_1 samples. Thus, following the taxonomic keys, the N_1 samples were named as *N. hexagonolepis*.

DNA Barcoding Analyses

The K2P divergence matrix of the dataset (as shown in Table S3) revealed that the congener of *Tor* maintained divergences in the range of 3.5% to 7.4% with the congener of *Neolissochilus*. The maximum K2P divergence among T_1 and T_2 samples was 0.2% and the comparison of both T_1 and T_2 samples with T_3 samples also revealed a maximum K2P divergence of 0.2%. The maximum divergence of all the samples belonging to T_1, T_2 and T_3 with the closest database sequences of *Tor putitora* was 0.8%. The divergence matrix suggested that all samples of T_1, T_2 and T_3 are conspecific of *T. putitora* in the absence of any database sequence of *T. progeneius*. Therefore, COI gene sequences of these 3 groups were submitted to both GenBank and BOLD under the putative species *T. putitora*. The maximum divergence within N_1 samples was 0.6% while their divergences with the conspecific database sequences were in the range of 0.4% to 0.8%. The divergence matrix suggested that N_1 samples are conspecific of *Neolissochilus hexagonolepis*. The within group divergences of N_2 samples were in the high range up to 0.9% possibly due to a particular sequence. Excluding the particular sequence (accession number JX127239), the within group divergence of N_2 samples remained nil in the absence of any conspecific sequence in the database.

The averages of conspecific and congeneric divergences were determined from the matrix to be 0.5% ±0.2% and 2.8% ±0.7% respectively. In the dataset, the minimum distance between the closest species (closest congener) was 1.5%. Therefore, the closest congeneric divergence among mahseer species was calculated to be 3 times higher than the mean conspecific divergence, which is called as the 'barcode gap'. Based on the barcode gap, T_1, T_2 and T_3 samples were found to be conspecific with *T. putitora*; N_1 samples were conspecific with *N. hexagonolepis*; and N_2 samples

including JX127239 were discrete in the absence of any database sequence of *N. hexastichus*.

The NJ as well as Maximum Likelihood (ML) and Bayesian tree based cluster revealed that the congener of *Tor* and *Neolissochilus* formed two related clades while *Puntius sarana* and *Hypsibarbus wetmorei* remained as out-group (Figure 3, Figure S2, and Figure S3). This also revealed that *T. putitora*, *T. tor*, *T. khudree*, *T. sinensis*, *T. mussullah*, *T. mosal*, *T. malabaricus*, *T. douronensis*, *T. tambroides*, *N. hexagonolepis* and *N. stracheyi* clustered separately and are distinct species. All the samples of T_1, T_2 and T_3 clustered in the same clade and nearest to *T. putitora*; N_1 samples clustered with *N. hexagonolepis*; and the 6 N_2 samples clustered in the same clade while the other N_2 sample remained a bit distant. In addition, some database sequences reflected aberrant clustering like, 1) all sequences of *T. mosal mahanadicus* and *T. macrolepis* clustered with *T. putitora* and 2) a single sequence of *N. stracheyi* (accession number HM536922) clustered with *N. Hexagonolepis*.

Discussion

In this study, all the possible mahseer habitats across the Northeast India were surveyed. Altogether three morphologically distinct groups of mahseer within the genus *Tor* and two within *Neolissochilus* were identified from the study site. DNA barcoding analyses however recognized all the three groups belonging to a single species within the genus *Tor* and conspecific of *T. putitora*. The *T. putitora* is a widely distributed species and it has been reported to be exhibiting polymorphism in geographically isolated populations [42]. Among the study samples, T_3 samples possessed long fleshy appendage to the lower lip (mental lobe) while the others lack this feature. This feature corresponds to the original description of *T. progeneius* where this particular feature was specially emphasized for nomenclature [18,21]. This species had been also considered closely allied to *T. tor* in view of its lower lip character; consequently these two species have been synonymized very often [5,19]. *T. progeneius* was however differently described after its original description probably due to lack of original holotype [3] and non-availability of fresh specimens [21]. It was identified to be distinct from *T. tor* due to length of head almost equal to depth of the body in the former vs. length of head considerably shorter than depth of the body in the latter [20]. Subsequently, based on archival specimens (Zoological survey of India, Kolkata; specimens' catalogue details not mentioned), it was characterized to be having 8–10 rakers on the lower arm of first gill arch, tubercles on the cheek and lacking completely a mental

lobe. Based on such characters this species was remarked to be doubtful to place in either in *Tor* or *Neolissochilus* [3]. According to one proposition, there are two types among the yellow finned mahseer: i) the lips are fleshy and the lower one is produced backwards into a long fleshy appendage, and ii) the lips are of normal type and the lower lip does not form an appendage [21]. Based on such descriptions, Hamilton's *Cyprinus* (present *Tor*) *putitora* and *C. mosal* have been stated to be the same species and the nature of their lips was stated to be adaptive characters [5]. Besides, the description of a fan-shaped structure behind upper jaw in *T. progeneius* [21] was stated to be an abnormal formation based on archival specimens (Zoological Survey of India, Kolkata; specimens' catalogue details not mentioned) [5]. In contrary, Menon (1992) [5] described *T. progeneius* to be possessing of 27–31 numbers of lateral line scales on the body. It seems that Menon (1992) was so influenced by this feature of *T. progeneius* that he used it as a taxonomic key to species. Secondly, in contrary to all previous descriptions except Rainboth (1985) [3], Menon (1992) noted the presence of cheek tubercles in *T. progeneius*. On the other hand, according to original description as well as the prevailing adoption of taxonomic character for this species indicate that the number of scales on the lateral line was never more than 26, and the extension of singular appendage from lower lip has been largely emphasized. This species has long been remained unreported, that might be due to the above mentioned morpho-taxonomic perplexity arising from vague and varied presentation of its specific characters incongruent to the original description [18]. All the T_3 samples were observed to be possessing of maximum of 26 lateral line scales, 13–14 gill rakers, the slightly longer head than body depth and particularly the fleshy lips with long angular appendage to the lower jaw (long mental lobe) that is in contrast to the short mental lobe in both T_1 and T_2 samples (Figure 4). The different lower lip structure in T_3 samples could be an adaptive [5,21] or a sexually dimorphic feature [43]. Moreover, different geographical populations of *T. putitora* have been reported for significant Nuclear Organiser Region polymorphism [42] that indicates the possibility of the presence of a polymorphic form of this species in northeast India. Because, the collection site of T_3 samples in the drainages of river Brahmaputra is phylogeographically poorly connected with the other Himalayan streams such as Ganga. Therefore, notwithstanding such noticeable differences in mouth structure, following DNA barcoding results, we conclude that *T. progeneius* is a synonymous species of *T. putitora*. This study contributed 10 replica barcode sequences in GenBank of *T. putitora*. In elsewhere, DNA barcoding approach has been successful in describing different nominal species in one [44]. This study would guide the conservationists to turn away the focus of conservation endeavor from *T. progeneius* to *T. putitora*.

The present study recognized two morphologically distinct groups of mahseer within the genus *Neolissochilus*. Among them, the N_1 and N_2 samples were identified to be *Neolissochilus hexagonolepis* and *N. hexastichus* respectively. DNA barcoding also differentiated both the species with considerable barcode gap and hence their identifications were confirmed. This study added in GenBank 3 replica barcode sequences of *N. hexagonolepis* and 7 new barcode sequences of *N. hexastichus*. The latter species has long been concealed since its first description in around 175 year back [18] due to lack of its morpho-taxonomic details and mis-identification with *T. tor*. The species *N. hexastichus* is though reported from other locations in the Salween basin [45] and Myanmar [46] but there are almost no biological data available on this species. Yet, it was first categorized into 'Vulnerable' [15] and subsequently to 'Near Threatened' status [16]. In this study, two species of mahseer, viz., *T. putitora* and *N. hexagonolepis* were found frequently in all the

mahseer habitats in the study area. On the other hand, the species *N. hexastichus* was absent in all the surveyed habitats except a particular river (25.420 N 92.993 E) in the entire study area that raises a serious concern about the future sustainability of this species. Although this river also harbors the other two most common species of mahseer but we observed illegal harvest of fishes through destructive fishing in the river. Thus, the mahseer species in this river are assumed to have been threatened from anthropogenic activities that demands mass awareness. This study would provide benefit to generate life history parameters of *N. hexastichus* for its conservation standpoint and development of aquaculture package of practice for sustainable utilization. Therefore, this study suggests to initiate priority conservation of *N. hexastichus*.

One of the differentiating characters of two genera *Neolissochilus* and *Tor* is based on the presence of labial groove interrupted in the former and continuous in the latter. This generic character was found to be confusing because this difference was not evident in *N. hexastichus*. Therefore, we consider that interrupted labial groove would be confusing to treat as the generic character of *Neolissochilus*. On the other hand, the characteristic difference of the number of gill rakers on the first arm of gill arch was found to be a very pronounced generic character of the two genera that may be emphasized in genus categorization.

The NJ, ML and Bayesian cluster showed that the genera *Puntius* and *Hypsibarbus* remained as out-group with respect to the two genera *Tor* and *Neolissochilus* of mahseer. In another study the two genera of mahseer have been proposed to be in a distinct clade compared to other six different clades within the subfamily Cyprininae [47]. So, the grouping of mahseer in a separate tribe [34,48] appeared justified, but, the particular tribe name is contentious. In the NJ phylogenetic analysis, some sequences, e.g., *T. macrolepis* (2 sequences) and *T. mosal mahanadicus* (3 sequences), though carried distinct names in the database but clustered cohesively with a popular species *T. putitora*. Such a wrong clustering of sequences may arise either due to misidentification or due to the occurrence of synonymous species, such as *T. macrolepis* has been stated to be a synonym of *T. putitora* [11,17]. Besides, the two samples of *Neolissochilus stracheyi* did not cluster with each other and have been possibly misidentified in the database.

In the history of taxonomy, the dawn of DNA barcoding technique has sufficiently helped in troubleshooting of many species identification where morphological characters were over-looked or overemphasized [29]. Yet, the reference database is found to be lacking of information on many extant species of mahseer. Hence, development of both new barcodes and replica barcodes from wide spatial scale would be important to enrich the DNA barcode reference library. New barcodes are particularly essential to achieve the objective of DNA barcoding to complete the digital taxonomic guide of earth's biota, while the replica barcodes from wide geographical ranges would substantiate the range distribution of the extant species.

Supporting Information

Figure S1 Scheme of measurement of morphometric variables on Fish. (adopted from Jayaram (1999) [23].
(TIF)

Figure S2 ML phylogeny. The tagging of the sequences with red and black dots as well as black triangles follow the same description as given for NJ phylogenetic tree in Figure 3.
(TIF)

Figure S3 Bayesian phylogeny. The specimens' GenBank accession number and species name are shown for each taxon. The

sequences highlighted with red and blue colour correspond to the sequences developed in this study while blue coloured sequences alone correspond to the sequences of samples morphologically identified as *Tor progeneius*, but are found conspecific with *Tor putitora* in this study hence, marked as *Tor putitora*. The green coloured sequences correspond to the cases of abnormal clustering.
(TIF)

Table S1   List of the studied species, GenBank Accession of the analyzed sequences and the geographical positions of the sample.
(DOC)

Table S2   Morphometric details of the studied species.
(DOC)

Table S3   Pairwise K2P divergence matrix between the sequences.
(XLS)

Supporting Information S1   Comparison of taxonomic descriptions based on morphology of *T. Progeneius* from time to time.
(DOC)

Author Contributions

Conceived and designed the experiments: SKG BAL MJB. Performed the experiments: BAL MJB BD PM SK. Analyzed the data: BAL MJB PM SK. Contributed reagents/materials/analysis tools: BD PM SK. Wrote the paper: BAL MJB SKG.

References

1. Hamilton F (1822) An Account of the Fishes found in the River Ganges and its Branches. Edinburg. 405 p.
2. Desai VR (2003) Synopsis of biological data on the *Tor* mahseer *Tor tor* (Hamilton, 1822). FAO Fisheries synopsis, Rome, FAO. 158: 36 p.
3. Rainboth WJ (1985) *Neolissochilus*, a new genus of south Asian Cyprinid fishes. Beaufortia. 35: 25–35.
4. Talwar PK, Jhingran A (1991) Inland fishes of India and adjacent countries. Oxford and IBH publishing, New Delhi. 1 & 2: 1158 p.
5. Menon AGK (1992) Taxonomy of mahseer fishes on the genus *Tor* of Gray with description of new species from Deccan. J. Bombay Nat. Hist. Soc. India. 89: 244–258.
6. Roberts TR (1999) Fishes of the cyprinid genus Tor in the Nam Theun watershed, Mekong Basin of Laos, with description of a new species. Raffles Bulletin of Zoology. 47: 225–236.
7. Bagra K, Kadu K, Sharma KN, Laskar BA, Sarkar UK, et al. (2009) Ichthyological survey and review of the checklist of fish fauna of Arunachal Pradesh, India. Check List. 5(2): 330–350.
8. Eschmeyer WN (ed) (2012) Catalog of Fishes. California Academy of Sciences. Available: http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp. Accessed 2012 Aug 04.
9. Thomas HS (1897) The Rod in India. Thacker and Company, London. 317 p.
10. McDonald JA (1948) Circumventing the Mahseer and other sporting fish in India and Burma. Natraj publication, Dehradun. 306 p.
11. Nguyen TTT (2008) Population structure in the highly fragmented range of *Tor douronensis* (Cyprinidae) in Sarawak, Malaysia, revealed by microsatellite DNA markers. Freshwater Biology. 53: 924–934.
12. FAO (2005) Fisheries and Aquaculture topics. Recreational fisheries. Topics Fact Sheets. Text by Andrew Smith. In: FAO Fisheries and Aquaculture Department [online]. Rome. Updated 27 May 2005. Accessed on 19 July 2012. http://www.fao.org/fishery/topic/14831/en.
13. Nautiyal P (2006) Rising awareness and efforts to conserve the Indian mahseers. Current Science. 91: 1604.
14. FAO (2006) FAO newsletter No.35. http://www.fao.org/docrep/013/a0595e/a0595e00.htm.
15. Molur S, Walker S (1998) Report of the workshop "Conservation, Assessment and Management plan (CAMP) for freshwater fishes of India. Zoo Outreach Organization, Conservation Breeding specialist group, Coimbatore. 156 p.
16. IUCN (2012) IUCN Red List of Threatened Species. Version 2012.1. <www.iucnredlist.org>. accessed on 04 August 2012.
17. Mohindra V, Khare, Praveen Lal KK, Punia P, Singh RK, et al. (2007) Molecular discrimination of five Mahseer species from Indian peninsula using RAPD analysis. Acta Zoologica Sinica 53: 725–732.
18. McClelland J (1839) Indian Cyprinidae. Asiatic Researches Calcutta, Bishop College Press. 19(2): 217–468.
19. Day F (1878) The fishes of India: being a natural history of the fishes known to inhabit the seas and fresh waters of India, Burma and Ceylon. Vols. I & II. Williams and Norgate, London. 778 p.
20. Sen TK, Jayaram KC (1982) The Mahseer Fishes of India: A review. Rec. Zool. Surv. India, Occ. Paper 39: 1–38.
21. Hora SL (1941) The game fishes of India: VIII. Mahseers of the large scaled barbels of India, 6. The Jungha of Assamese, *Barbus* (*Tor*) *progeneius* McClelland. Journal of Bombay Natural History Society 12: 526–532.
22. Hora SL, Mukerji DD (1936) Fish of the Naga Hills, Assam. Records of Indian Museum 38: 328–330.
23. Jayaram KC (1999) The freshwater fishes of Indian region. Narendra publishing house, Delhi. 551p.
24. Avise JC (1994) Molecular markers, natural history and evolution. New York: Chapman & Hall.
25. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol. Evol. 18: 70–74.
26. Hebert PDN, Stoekle M, Zemlak T, Francis CM (2004) Identification of birds through DNA barcodes. Plos Biology 2: 1657–1668.
27. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding of Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360: 1847–1857.
28. Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329–356.
29. April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci USA 108: 10602–10607.
30. Clare EB, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2006) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Molecular Ecology Notes 7: 184–190.
31. Smith MA (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc Natl Acad Sci USA 105: 12359–12364.
32. Frezal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. Infection, genetics and evolution. Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 8: 727–736.
33. Bhattacharjee MJ, Laskar BA, Dhar B, Ghosh SK (2012) Identification and re-evaluation of freshwater catfishes through DNA barcoding. PLoS ONE 7(11): e49950. Doi:10.1371/journal.pone.0049950.
34. Rainboth WJ (1996) FAO species identification field guide for fishery purposes, fishes of the Cambodia Mekong. FAO, Rome. 265 p.
35. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.
36. Zhang DX, Hewitt GM (1996) Nuclear integrations: challenge for mitochondrial DNA markers. Trends in Ecology and Evolution 11: 247–251.
37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
38. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Current protocols in Bioinformatics/editorial board, Andreas D Baxevanis [et al] Chapter 2: Unit 2.3.
39. 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.
40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 10: 2731–2739.
41. Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.
42. Singh M, Kumar R, Nagpure NS, Kushwaha B, Mani I, et al. (2009) Extensive NOR site polymorphism in geographically isolated populations of Golden mahseer, Tor putitora. Genome 52: 783–789.
43. Byrkjedal I, Rees DJ, Willassen E (2007) Lumping lumpsuckers: Molecular and morphological insights into the taxonomic status of Eumicrotremus spinosus (Fabricius, 1776) and E. eggvinii Koefoed, 1956 (Teleostei: Cyclopteridae). J. Fish Biol. 71: 111–13.
44. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences, USA 101: 14812–14817.
45. Doi A (1997) A review of taxonomic studies of cypriniform fishes in Southeast Asia. Jap. J. Ichthyol. 44: 1–33.

46. Oo W (2002) Inland fisheries of the Union of Myanmar. In T. Petr and DB Swar (eds.) Cold Water Fisheries in the Trans-Himalayan Countries. FAO Fish. Tech. Pap. 431: 70–76.

47. Yang L, Mayden RL, Sado T, He S, Saitoh K, et al. (2010) Molecular phylogeny of the fishes traditionally referred to Cyprinini sensu stricto (Teleostei: Cypriniformes) Zoologica Scripta 39: 527–550.

48. Devi RK, Indra TJ (1997) Check List of the Native Freshwater Fishes of India. Available: zsi.gov.in/checklist/Native%20freshwater%20Fishes%20of%20India. Accessed 2012 Aug 04.