Taylor & Francis
Taylor & Francis Group

MITO COMMUNICATION

🔓 OPEN ACCESS 　 Check for updates

DNA barcoding of freshwater fishes from Brahmaputra River in Eastern Himalaya biodiversity hotspot

Shantanu Kundu 　(iD), Kailash Chandra 　(iD), Kaomud Tyagi, Avas Pakrashi and Vikas Kumar

Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Kolkata, India

ABSTRACT

The genetic diversity of freshwater fishes is still anonymous in several drainage systems in northeast India. Moreover, the comparative genetic analysis is largely sporadic to judge their actual diversity and true status. We generated 89 DNA barcodes of 40 morphologically identified fishes collected from two major tributaries of Brahmaputra River. The comparative study revealed that most of the species were clearly discriminated by their estimated genetic distances and monophyletic clustering in Bayesian (BA) tree. Considering the genetic divergence (2%) for species discrimination boundary, the high genetic diversity (2.36–10.73%) was detected in 11 species (*Macrognathus pancalus*, *Channa punctata*, *Puntius terio*, *Bangana ariza*, *Garra arupi*, *Badis badis*, *Mystus vittatus*, *Rita rita*, *Gagata cenia*, *Mastacembelus armatus*, and *Danio dangila*), which signified the occurrence of concealed genetic diversity in this eco-zone. However, the insignificant genetic distances were also noticed in few reportedly valid species: *Channa stiktos* and *C. ornatipinnis* (1.43%); *Mystus ngasep*, *M. rufescens*, and *M. carcio* (0.4%); *Glyptothorax trilineatus*, *G. churamanii*, and *G. verrucosus* (0.4%); *Botia almorhae*, *B. histrionica*, *B. loha-chata*, and *B. rostrata* (0–0.4%); *Barilius barilia* and *B. vagra* (0.4%); *Batasio merianiensis* and *B. tengana* (1.2%); *Puntius chola* and *P. fraseri* (0%), *Schistura beavani* and *S. paucireticulata* (0%); hence to validate this species, generation of more barcode data was required from their types or topotypes. The present study would help to develop conservation schemes for the native species and collegiate ecosystem, which associated with the livelihoods of millions of ethnic communities in this region.

1. Introduction

The northeastern region of India shares two biodiversity hotspots, Eastern Himalaya and Indo-Burma. These two hotspots are rich in biodiversity and considered as a prolific land of various floral and faunal diversity (Myers et al. 2000). The unique biogeography, climatic condition, vegetation, and inland water systems furnished a safe adobe for various indigenous living systems including ichthyofauna in this ecozone. The four river basins, the Bramhaputra, Barak, Chindwin, and Koladyne harbour a tremendous diversity of freshwater fishes. In general, these species not only play a substantial role in the ecosystem but also directly link with humans (as food resources, remedies, hobbies, and recreation purposes) and aid to flourish the country's economy. During the last two decades, ichthyologists exposed deep attention in the survey of piscatorial resources and consequently many new taxa have been discovered from this region (Darshan et al. 2018). However, the comprehensive checklist has emerged in 2012 which listed the potential occurrence of about 422 species from this region (Goswami et al. 2012). Further, the knowledge on fish diversity is exploring rapidly from this region. Almost 89 new species were added to the freshwater fish

fauna in northeast India in the last decade (Animal Discoveries 2009–2017; Eschmeyer 2012). Addition to this, due to the Miocene tectonic activity, this region is also reported by a considerable number of endemic fishes (He et al. 2001; Ruber et al. 2004). The status of freshwater fishes in eastern Himalaya was recorded as Critically Endangered (1%), Endangered (3%), Vulnerable (10%), Near Threatened (9%), Least Concern (50%), and Data Deficient (27%) as per IUCN Red List category (Allen et al. 2010). Nevertheless, these organisms face major threats in terms of habitat loss and degradation, untenable fishing practices, development of dams, illegitimate trade, and other anthropogenic activities (Allen et al. 2005; Carrizo et al. 2017). Thus, the rigorous assessment of freshwater fish diversity in each river basin is urgently required for ichthyology and implication of conservation policies (Molur and Walker 1998).

Owing to the advancement of taxonomic and biodiversity research, the utility of molecular tools, especially DNA barcoding has successfully evidenced to determine the fish diversity worldwide including India (Hubert et al. 2008; April et al. 2011; Chakraborty and Ghosh 2014; Khedkar et al. 2014; Chen et al. 2015). The notable studies on freshwater fishes

---

through DNA barcoding have been elucidated by estimating the diversity of catfish species from Barak drainage (Bhattacharjee et al. 2012) and ornamental fishes from both Brahmaputra and Barak drainage (Dhar and Ghosh 2015). This supplementary technique was manifested to discriminate the closely related species as well as resolve taxonomic ambiguities (Laskar et al. 2013, 2018). This additional tool was also corroborated to update the checklist of ichthyofauna from Ranganadi River in Assam state, and intensification of freshwater fish database from India (Nagpure et al. 2015; Kaushik and Bordoloi 2016). A significant study was also endeavouring to evaluate the genetic diversity of fishes from Diphlu River inside Kaziranga National Park (Laskar et al. 2019). Further, the large-scale attempt has been accomplished in recent past to generate the barcode data of 109 freshwater fishes from Indo-Myanmar biodiversity hotspot; considering Barak, Brahmaputra, Chindwin, and Karnafuli Rivers (Barman et al. 2018). However, the publicly available barcode data of freshwater fishes are still inadequate from each drainage system. Due to the inaccessibility in hilly terrain, lack of colloquial skill with native ethnic communities, and other legislative concerns; the survey and assessment of fishes are largely sporadic in the inland waters of Arunachal Pradesh, northeast India. These northeastern-most state is mainly encompassed by three major hills, Miri, Abor, and Mishmi as well as contiguous with three neighbouring nations, Bhutan, China, and Myanmar. Although political boundaries visibly demarcated this state from other countries, the streams of the underlying riverine system naturally share bio-resources including freshwater fishes. Nevertheless, the invasion of non-native species and their impacts on the native ichthyofauna has obtained less attention in the globe including northeast India (Collins et al. 2012). Moreover, the indigenous inhabitants of this region have frequently used ichthyofauna as folklore therapeutics and ethnozoological practices (Chinlampianga et al. 2013). Hence, the present study is aimed to rapidly evaluate the fish species diversity, hunted by the local tribes, from two focal tributaries (Siang and Dibang) of the Brahmaputra River and generate the DNA barcode data of taxonomically identified species. Besides, the comparative analysis of the genetic data within the native freshwater fishes has meagerly examined from northeast India. These open approaches will help to (i) recognize the fish genetic diversity in two drainage systems of the Brahmaputra River, (ii) enhance the ichthyofauna checklist of the state as well as from the region, and (iii) enrich the global DNA barcode library. The present study will be helpful to formulate the strategies for sustainable management and conservation of ichthyofauna bio-resources.

2. Materials and methods

2.1. Taxon sampling and morphological identification

A total of 89 freshwater fish samples were collected from the local tribes during the fishing activities in five different localities (28.15N 95.66E, 28.19N 95.55E, 28.08N 95.33E, 28.27N 95.19E, and 28.14N 95.10E) in two main streams of the Brahmaputra River, Siang and Dibang at two districts East

Siang and Lower Dibang Valley in Arunachal Pradesh (Figure 1a). Desire size fish samples were preserved in 70% alcohol for further taxonomic and molecular study. The morphological identification and nomenclature were confirmed by the available taxonomic keys in recently published literature, books, and existing information at Catalog of Fishes (Animal Discoveries 2009–2017; Eschmeyer 2012; Vishwanath et al. 2014). A meagre amount of muscle tissue was collected from each sample using sterile forceps, scissors, and surgical blades for DNA analysis. The voucher specimens were properly tagged and preserved in the National Zoological Collections (NZCs) of the Freshwater Fish Section, Zoological Survey of India (ZSI), Kolkata.

2.2. Genomic DNA extraction, amplification and sequencing

Genomic DNA extraction, PCR, and purification were carried out from each tissue sample following standard protocol with published primer pairs (Ward et al. 2005). PCR was performed using Veriti® Thermal Cycler (Applied Bio systems, Foster City, CA). The purified PCR products were cycle sequenced cleaned following by published protocol and sequenced bi-directionally using ABI-3730 DNA analyzer at the Zoological Survey of India, Kolkata (Tyagi et al. 2017).

2.3. Sequence quality control measures and data acquisition

The forward and reverse chromatograms (>600 bp) of the partial mitochondrial cytochrome c oxidase subunit 1 (mtCOI) gene were obtained for each studied sample. The consensus sequence of each sample was annotated carefully by the alignment of both forward and reverse complement of reverse sequence. The online tool, nucleotide BLAST (https://blast.ncbi.nlm.nih.gov) was used to check the insertion and deletion within the annotated sequences, and amino acid array of vertebrate mitochondrial genome was assured through online NCBI-ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). The availability of the DNA barcode data of the known Indian freshwater fish species was screened and total 140 sequences were retrieved from GenBank database, which were published in peer reviewed journals (Table S2). The database sequence of *Apristurus ampliceps* (order Carcharhiniformes) was also acquired from GenBank to use as out-group. Further, the list of sampled freshwater fishes (40 species) and their congeners (105 species) known from northeast India were finalized for comparative genetic analysis following the published literatures (Goswami et al. 2012; Barman et al. 2018).

2.4. Genetic divergence and cluster analysis

To form a final dataset, all the generated and publicly available database sequences were aligned through ClustalX software (Thompson et al. 1997). Further, to evade the incongruous outcomes in genetic distance and tree analysis, the dataset was prepared to equal length of 609 bp. The
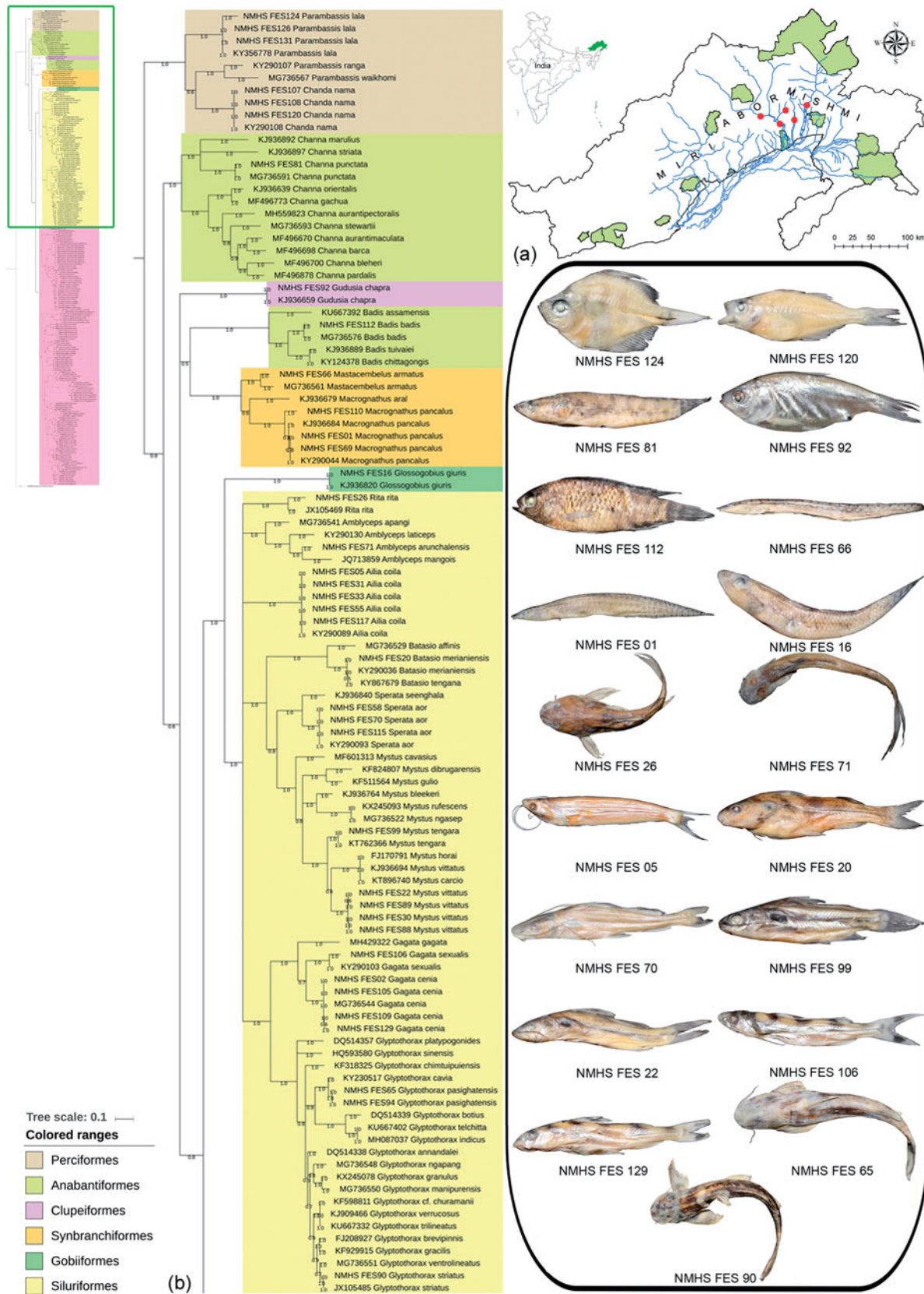
Figure 1. (a) Map with red dots showing the collection localities in Arunachal Pradesh, northeast India. (b) Pruned Bayesian (BA) tree by partial mtCOI gene inferred the delimitation of the studied freshwater fishes in northeast India. Lateral and dorsal images of the representative fish species were merged beside the BA tree. (c) Pruned Bayesian (BA) tree by partial mtCOI gene inferred the delimitation of the studied freshwater fishes in northeast India. Lateral and dorsal images of the representative fish species were merged beside the BA tree.
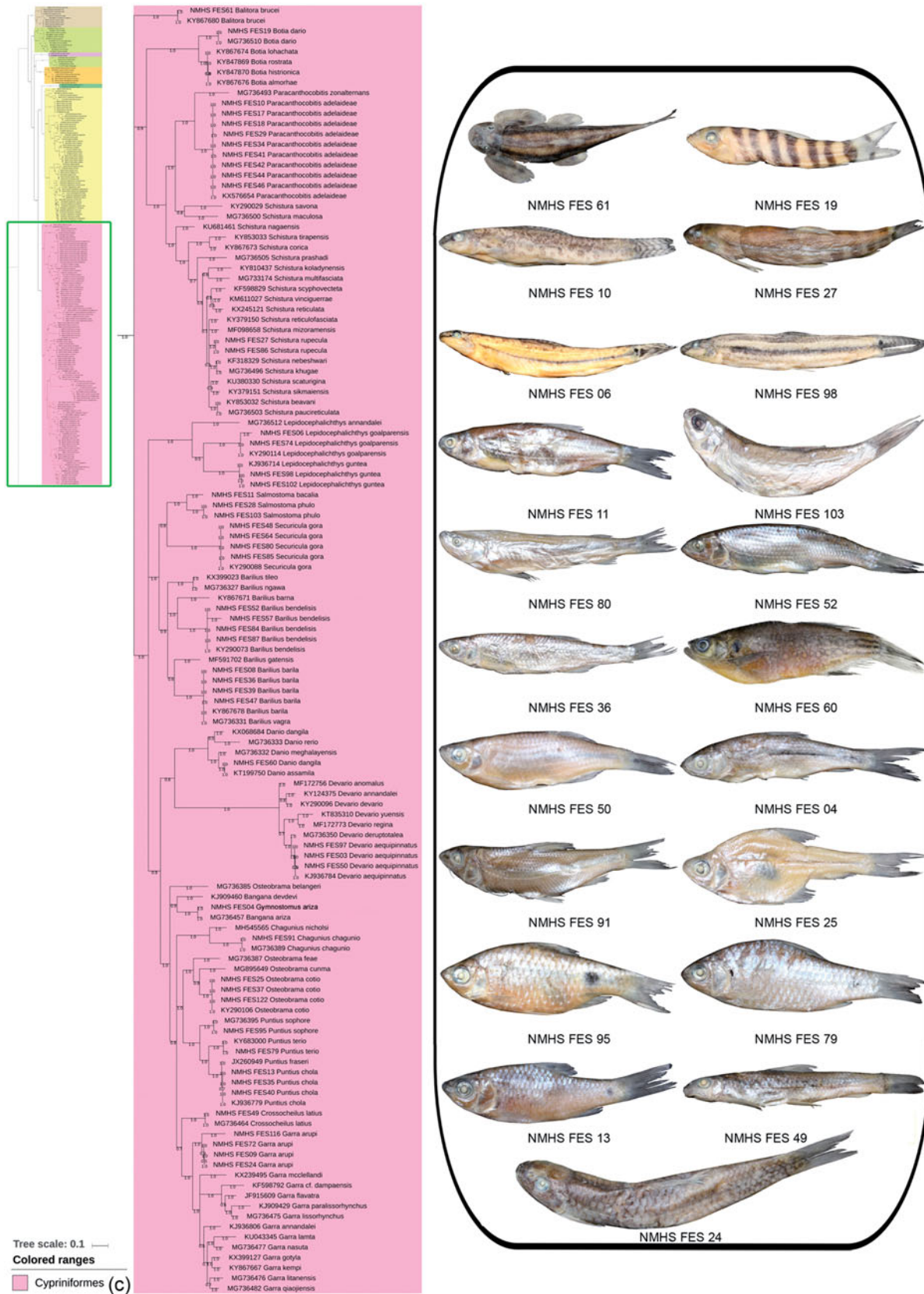
Figure 1. Continued.

genetic divergence within and between each taxonomic group was calculated in MEGAX (Kumar et al. 2018). The best fit model for this dataset was estimated using Mr. MODELTEST v2with lowest BIC (Bayesian Information Criterion) score (Nylander 2004). The Bayesian (BA) tree was constructed in Mr. Bayes 3.1.2by selecting nst = 6 for GTR + G + I model and four (one cold and three hot) metropolis-coupled Markov Chain Monte Carlo (MCMC), was run for

10,000,000 generations with 25% burn in with trees saving at every 100 generations (Ronquist and Huelsenbeck 2003). The MCMC analysis was used to generate the convergence metrics, till the standard deviation (SD) of split frequencies reached under 0.01 and the potential scale reduction factor (PSRF) for all parameters approached 1.0. To represent the generated BA tree, the web based iTOL tool (https://itol.embl.de/) was used (Letunic and Bork 2007).

3. Results and discussion

3.1. DNA barcoding and delimitation of taxa

The present study generated 89 DNA barcode sequences of 40 morphologically identified species from two major streams of Brahmaputra River in Arunachal Pradesh (Figure 1b,c; Table S1). All the studied morphospecies were closely related to the publicly available published database sequences of the same species and well discriminated within seven orders by partial mtCOI gene and reciprocal monophyly (Figure 1b, c). The genetic distances were 11.7%, 12.9%, 20.2%, 22.1%, and 22.5% in Synbranchiformes, Perciformes, Siluriformes, Anabantiformes, and Cypriniformes, respectively. The genetic distances between the above said orders were ranging from 24.4% to 30.0%. The present dataset deals with 229 DNA barcode data on 145 species within 32 genera reported from northeast India. Excluding the singleton species, the genetic distances within the genus were ranging from 3.21% (*Paracanthocobitis*) to 18.98% (*Channa*). The previous molecular study suggested that the variety of snakehead species is higher in Eastern Himalaya due to the concealed diversity and endemism or the fusion of more well-defined lineages (Conte-Grand et al. 2017). The resulted high genetic distances within the genus *Channa* reflected hidden species diversity in the studied region. Further, the genetic distances between the genera were ranged from 12.4% to 44.9%. Considering the previous thoughts on K2P distances is greater than 2% for accurate species differentiation in vertebrates (Johns and Avise 1998; Hebert et al. 2004), the present dataset observed high genetic distances within 11 species; *Macrognathus pancalus* (2.36%), *Channa punctata* (2.80%), *Puntius terio* (2.80%), *Bangana ariza* (3.20%), *Garra arupi* (3.45%), *Badis badis* (4.87%), *Mystus vittatus* (6.20%), *Rita rita* (6.93%), *Gagata cenia* (7.03%), *Mastacembelus armatus* (7.49%), and *Danio dangila* (10.73%). Although numerous studies has been intended to establish the actual barcode gap for species delimitation (Meyer and Paulay 2005; Steinke et al. 2009), the genetic distances for evaluating the barcode gap were inconsistent relating to the sampling strategies of different taxa (Meier et al. 2008). Hence, by observing the high genetic distances within the species, we assumed that the studied species may have possible cryptic diversity or more hidden species diversity from this region.

3.2. Comparative DNA barcodes analysis

In recent years, the generation of barcode data of freshwater fishes has massively contributed in the global GenBank database from different riverine systems, lakes, and reservoirs (Díaz et al. 2016; Lakra et al. 2016). Indeed, this rapid enrichment of database helps to identify the species promptly without prolonged appraisal through morpho-taxonomy (Hebert et al. 2003; Tautz et al. 2003; Ward et al. 2009). However, several technical issues such as erroneous annotation, and biased towards the publicly available sequences, are the major impediments for accurate species identification (Moritz and Cicero 2004; Shen et al. 2013). Further, the DNA barcodes of erroneously identified species or biological contamination often yield wrong similarity search result in open access Nucleotide Basic Local Alignment Search Tool (BLASTn) search (https://blast.ncbi.nlm.nih.gov) (Ebach and Holdrege 2005). Several DNA barcodes neither contained any collateral supplementary information like vouchers IDs, coordinates for collection localities nor published electronically. Hence, dealing with those open access sequences is often misleading during the dataset preparation and comparative analysis. Keeping all those obstacles in mind, the database sequences were retrieved and incorporated in the present comparative analysis. The northeast India is reported by seven genera each with single species, *Bangana devdevi Chanda nama*, *Mastacembelus armatus*, *Glossogobius giuris*, *Gymnostomus ariza*, *Rita rita*, and *Securicula gora*, which showed distinct genetic discrepancies as compared to other species. The species *Bangana ariza* is recently synonymized under *Gymnostomus* (Yang et al. 2012) and however showed sister cladding with *Bangana devdevi* in the BA tree.

The genus *Channa* is known by 15 species from northeast India (Eschmeyer 2012; Lalramliana Knight et al. 2018). Eleven species of *Channa* (*C. aurantimaculata*, *C. aurantipectoralis*, *C. barca*, *C. bleheri*, *C. gachua*, *C. marulius*, *C. orientalis*, *C. pardalis*, *C. punctata*, *C. stewartii*, and *C. striata*) showed 4.1% to 28.4% genetic distances with monophyletic clustering in BA tree. Further, *C. stiktos* from Kaladan River drainage, Mizoram is morphologically closer and showed low genetic divergence (1.43%) with the Burmese species, *C. ornatipinnis* and sufficient genetic distinctiveness (18.4–26.7%) with northeast Indian species. The recently described species, *C. bipuli* showed clear genetic distances (4.8–23%) and distinct clade with the northeast Indian species. The genus *Badis* is known by nine species from northeast India (Darshan et al. 2018; Eschmeyer 2012). Among them, four species (*B. assamensis*, *B. badis*, *B. chittagongis*, and *B. tuivaiei*) showed 3.2–18.9% genetic distance and close relationship in BA tree. The genus *Macrognathus* is known by four species from northeast India. Among them, two species, *M. aral* and *M. pancalus* maintained 17.7% genetic distance and closely assemblage in BA tree. The genus *Amblyceps* is known by eight species from northeast India (Darshan et al. 2016). Among them, four species (*A. apangi*, *A. arunchalensis*, *A. laticeps*, and *A. mangois*) showed 9.7–17.2% genetic distances and monophyletic clustering in BA tree.

The genus *Batasio* is known by three species from northeast India (Darshan et al. 2018; Eschmeyer 2012). *Batasio affinis* showed 13% and 13.5% genetic distances with *B. merianiensis* and *B. tengana,* respectively. However, significantly low genetic divergence (1.2%) was revealed between two species, *B. merianiensis* (KY290036) and *B. tengana* (KY867679). The genus *Sperata* is known by three species

from northeast India. Among them, two species *S. aor* and *S. seenghala* were well discriminated from each other with 9.8% genetic distance. The genus *Mystus* is known by 10 species from northeast India. All the congeners of *Mystus* were showed 0.4–24.7% genetic distances and monophyletic clustering in BA tree. However, *M. ngasep* (MG736522) and *M. rufescens* (KX245093) as well as *M. carcio* (KT896740) and *M. horai* (FJ170791) revealed insignificant genetic distance (0.4%) with each other, which need further reevaluation. The genus *Gagata* is known by four species from northeast India. Later on, *G. gasawyuh* was synonymized with *G. dolichonema* which is reported to be distributed in Chindwin and Tenesserim basins. The other three species, *G. cenia*, *G. gagata*, and *G. sexualis* revealed sufficient genetic distances ranging from 13.2% to 19.6% and showed monophyletic clustering in BA tree.

The genus *Glyptothorax* is known by 21 species from northeast India (Darshan et al. 2018). Later on, five species, *G. churamanii*, *G. dikrongensis*, *G. jayarami*, *G. pasighatensis*, and *G. verrucosus* were described from northeast India (Arun Kumar 2016). Two *Glyptothorax* specimens (NMHS FES 65 and NMHS FES 94) were collected from the same type locality in Siang River of Arunachal Pradesh and showed similarity with *G. pasighatensis*. The subsequent taxonomic revision further synonymized *G. coheni* with *G. saisii*, distributed in the Subarnarekha River of eastern India; and *G. chindwinica* with *G. burmanicus*, distributed in Irrawaddy and Salween River drainages in China and Myanmar. The examined genetic distance was negligible between *G. indicus* (MH087037) and *G. telchitta* (KU667402) as well as *G. churamanii* (KF598811) and *G. verrucosus* (KJ909466). Further, *G. trilineatus* (KU667332) showed 0.4% genetic distance with *G. churamanii* and *G. verrucosus* in the present dataset. Excluding the low genetic distances of above-mentioned *Glyptothorax* congeners, the genetic distances were ranging from 2.4% to 19.3%. However, all the *Glyptothorax* species showed monophyletic clustering in BA tree as suggested earlier (Jiang et al. 2011).

The genus *Botia* is known by five species from northeast India (Goswami et al. 2012). Among them *B. dario* (MG736510) showed 7.9–8.3% genetic distances and monophyletic clustering in BA tree with other four species (*B. almorhae*, *B. histrionica*, *B. lohachata*, and *B. rostrata*). However, the publicly available database sequences of four species; *B. almorhae* (KY867676), *B. histrionica* (KY847870), *B. lohachata* (KY867674), and *B. rostrata* (KY847869) showed low genetic distance (0–0.4%) in the present dataset, which suggests further investigation. The genus *Paracanthocobitis* is known by three species from northeast India. *Paracanthocobitis adelaideae* and *P. zonalternans* were distributed in the Salween River drainage, Thailand and Myanmar. However, these two species were also reported in northeast India. The generated DNA barcode sequences of these two species (KX576654 and MG736493) showed 15.9% genetic distance and monophyletic clustering in BA tree. The genus *Lepidocephalichthys* is known by eight species from northeast India. Later on, *L. manipurensis* was synonymized with *L. micropogon* and *L. menoni* was synonymized with *L. annandalei* (Eschmeyer 2012). Three species, *L. annandalei*,

L. goalparensis, and *L. guntea* showed 20.1–21% genetic distances and monophyletic clustering in BA tree.

The genus *Salmostoma* is known by four species from northeast India (Goswami et al. 2012). The present study collected two *Salmostoma* species, *S. bacalia* (NMHS FES 11) and *S. phulo* (NMHS FES 28 and NMHS FES 103) and showed monophyletic clustering with 13% genetic distance. The genus *Barilius* is known by 15 species from northeast India. Among them, nine species, *B. bakeri*, *B. barna*, *B. barnoides*, *B. chatricensis*, *B. dogarsinghi*, *B. lairokensis*, *B. ngawa*, *B. shacra*, and *B. tileo*, were transferred to genus *Opsarius* (Eschmeyer 2012). Later on, two species, *B. arunachalensis* and *B. profundus* were discovered from the D'Ering Memorial Wild Life Sanctuary, Arunachal Pradesh, and Koladyne River, Mizoram, respectively. In the present dataset, we compared the genetic distinctiveness of four *Barilius* species, *B. barila*, *B. bendelisis*, *B. gatensis*, and *B. vagra*, which showed 2–21.9% genetic distances with monophyletic cladding in BA tree. Further, the comparative analysis also noted insignificant genetic distance (0.4%) between two species, *B. barilia* and *B. vagra*.

The genus *Danio* is known by three species (*D. dangila*, *D. jaintianensis*, and *D. rerio*) from northeast India (Goswami et al. 2012). Later on, two new species, *D. assamila* and *D. meghalayensis* were discovered from the Brahmaputra River drainage in northeastern India (Eschmeyer 2012). The comparative DNA barcode analysis of remaining four species revealed 5.3–14.4% genetic distances between each other and showed monophyletic clustering in BA tree. The genus *Devario* is known by 10 species from northeast India including *D. assamensis* (Goswami et al. 2012). However, the current status of *D. assamensis* is uncertain because subsequently it was mentioned as a junior synonym of *D. aequipinnatus* (Kullander et al. 2017). Later on, one species, *D. deruptotalea* was discovered from Chindwin River drainage, Manipur (Eschmeyer 2012). The comparative analysis of seven species (*D. aequipinnatus*, *D. annandalei*, *D. anomalus*, *D. deruptotalea*, *D. devario*, *D. regina*, and *D. yuensis*) revealed 3.5–15% genetic distances between each other and showed monophyletic cladding in BA tree. Further, the DNA barcode data of three species (*D. acuticephala*, *D. horai*, and *D. naganensis*) is lacking in the GenBank database. The genus *Chagunius* is known by two species (*C. chagunio* and *C. nicholsi*) from northeast India. The generated and database sequences of both the species showed 12.7% genetic distance and monophyletic clustering in BA tree.

The genus *Puntius* is known by 24 species from northeast India (Goswami et al. 2012). Later on, most of the species are synonymized under four genera (*Pethia*, *Naziritor*, *Dawkinsia*, and *Systomus*) (Eschmeyer 2012). The present comparative study deals with four species, *P. chola*, *P. fraseri*, *P. sophore*, and *P. terio*. Among them, the generated and database sequences of *P. chola* (NMHS FES 13, NMHS FES 35, NMHS FES 40, KJ936779) and *P. fraseri* (JX260949) showed negligible genetic distances; however, *P. sophore* and *P. terio* showed 9.1–13.5% genetic distance between all the four species. All the studied species showed monophyletic clustering in BA tree. The genus *Garra* is known by 20 species from northeast India. Later on, 30 new species were also discovered from

Bramhaputra, Chindwin, Karnaphuli, and Koladyne River basin in northeast India (Moyon and Arunkumar 2018). Further, the recent morphology and DNA barcode based assessment reported the range expansion of *G. qiaojiensis* from the Upper Irrawaddy River basin, Yunnan, China in northeast India (Barman et al. 2018). Based on the generated sequences and available database sequences, the comparative analysis was performed with 13 species (*G. annandalei*, *G. arupi*, *G. dampaensis*, *G. flavatra*, *G. gotyla*, *G. kempi*, *G. lamta*, *G. lissorhynchus*, *G. litanensis*, *G. mcclellandi*, *G. nasuta*, *G. paralissorhynchus*, and *G. qiaojiensis*). The genetic distances (4–23.2%) clearly discriminated all the studied *Garra* species with monophyletic clustering in BA tree.

3.3. DNA barcoding data hinted paraphyly of some taxa

The genus *Parambasis* is known by three species (*P. lala*, *P. ranga*, and *P. waikhomi*) from northeast India. These three species were distinctly separated by 16.2–19.9% genetic distances. However, the BA tree showed ambiguous clustering; with generated and database sequences of *P. lala* with distinct clade while *P. ranga* (KJ936704) and *P. waikhomi* (MG736567) showed close assemblage with *C. nama* clade. Hence, the congeners of *Parambasis* represented paraphyletic clades in the present dataset. The genus *Schistura* is known by 26 species from northeast India (Goswami et al. 2012). Among them eight species *S. arunachalensis*, *S. corica*, *S. inglisi*, *S. minutus*, *S. montanus*, *S. multifasciatus*, *S. prashadi*, and *S. tigrinum* are considered as *S. tirapensis*, *Nemacheilus corica*, *S. rupecula*, *S. minuta*, *S. montana*, *S. multifasciata*, *Mustura prashadi*, and *S. tigrina* respectively. In recent past, six new species (*S. koladynensis*, *S. maculosa*, *S. mizoramensis*, *S. nebeshwari*, *S. paucireticulata*, and *S. scyphovecteta*) were discovered from Pharsih, Tuivai, Tuirivang, and Koladyne River in Mizoram state of northeast India and one species, *S. larketensis* was discovered from a limestone cave, east Jaintia hills, Meghalaya (Eschmeyer 2012). The collected specimen (NMHS FES 27 and NMHS FES 86) from Arunachal Pradesh was identified as *S. rupecula*, which is distributed in India and Nepal. The present dataset represents the comparative analysis of 18 *Schistura* species. Most of the species showed cohesive clustering in BA tree; except *S. savona* and *S. maculosa*, which exhibited close assemblage with *Paracanthocobitis* clade. Hence, the congeners of *Schistura* represented paraphyletic clade, which needs further confirmation with more sampling and other genetic markers. The genetic distance was negligible for two species, *S. beavani* (KY853032) and *S. paucireticulata* (MG736503). Excluding these two species, the remaining 16 species (*S. corica*, *S. khugae*, *S. koladynensis*, *S. mizoramensis*, *S. multifasciata*, *S. nagaensis*, *S. nebeshwari*, *S. prashadi*, *S. reticulata*, *S. reticulofasciata*, *S. rupecula*, *S. scaturigina*, *S. scyphovecteta*, *S. sikmaiensis*, *S. tirapensis*, and *S. vinciguerrae*) showed 2.4–22.7% genetic distance with monophyletic cladding in BA tree. The genus *Osteobrama* is known by four species (*O. belangeri*, *O. cotio*, *O. cunma*, and *O. feae*) from northeast India. The barcode sequences of *Osteobrama* congeners are clearly

discriminated by genetic distances (12.5–23.9%). Further, *O. belangeri* (MG736385) showed ambiguous cladding with other three species and depicted paraphyletic clade in BA tree.

4. Conclusion

Traditional taxonomy is largely assessed to discover and naming of the organisms since Linnaeus time. Over time, scientists proposed various alternative biological species concept, especially phylogenetic theory to recognize the diversity and evolutionary thoughts (Avise 1994; Bickford et al. 2007). However, the high degrees of phenotypic similarity in cryptic species, sibling species, and sexual dimorphism have posed challenges in morphotaxonomy (Hebert et al. 2003; Tyagi et al. 2017). Nonetheless, to measure the Environmental Impact Assessments (EIA) under the Convention on Biological Diversity (CBD), promotion of social benefit sharing, and amendment of sustainable schemes for animal welfare, the assessment of biodiversity is rapidly encountering in the last few decades (Gopal 2015). As a result, researchers have frequently unwrapped the faunal diversity with overlooked or overemphasized morphological characters (Hebert et al. 2004; Collins and Cruickshank 2013). At this juncture, DNA barcoding tool with different species delimitation methods play a pivotal role to resolve several taxonomic dilemmas and provide truthful conceptions in systematics research (Laskar et al. 2013; Yang et al. 2015). This supplementary technique has also equally assisted in taxonomic revision by synonymizing many taxa as well as clarifies the anecdotal thoughts of species distribution (Bergsten et al. 2012; Stuckas et al. 2013). Hence, to validate the taxonomic status of the reported freshwater fish species from the northeastern region, the comparative genetic analysis could have fundamentals for evolutionary studies, implications for biogeography and conservation planning. Most of the studied congeners from northeast India were distinctly discriminated by the estimated genetic distances and tree analysis. However, for in-depth evidence of paraphyletic clade resulted in the congeners of *Osteobrama Parambasis*, and *Schistura*; we suggested more sampling and generation of DNA data of multiple molecular markers of all extant congeners from northeast India and other known range distributions. The present comparative analysis further revealed very low genetic distance between few species of *Channa*, *Barilius*, *Batasio*, *Botia*, *Glyptothorax*, *Mystus*, *Puntius*, and *Schistura*; which need further confirmation by assessing their topotypes through integrated approaches. The DNA barcode data of taxonomically identified freshwater fishes strengthen the global database for prompt species identification. In addition, the present study also helps to validate the checklist of ichthyofauna from Arunachal Pradesh as well as from northeast India. Collectively, the present genetic analysis of freshwater fishes would be helpful to recognize the diversity, detection of invasive species, as well as formulation of strategies for sustainable management and conservation.

Acknowledgements

Author contributions

S.K. and K.C. conceived and designed the experiment. S.K. collected the specimens, performed taxonomic identification, and captured photographs. K.C. and V.K. contributed to chemicals, laboratory facilities, and provided logistics support. S.K., K.T., and A.P. generated molecular data and data acquisition. S.K., K.T., and V.K. analyzed the data and prepared the table, figures. S.K., K.C., K.T., and V.K. wrote and reviewed the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Shantanu Kundu http://orcid.org/0000-0002-5488-4433
Kailash Chandra http://orcid.org/0000-0001-9076-5442

References

Allen JD, Abell R, Hogan Z, Revenga C, Taylor BW, Welcomme RL, Winemiller K. 2005. Overfishing of inland waters. Bioscience. 55:1041–1051.

Allen DJ, Molur S, Daniel BA. 2010. The status and distribution of freshwater biodiversity in the Eastern Himalaya. Cambridge, UK and Gland, Switzerland: IUCN, and Coimbatore, India: Zoo Outreach Organization.

Animal Discoveries: New Species and New Records. 2009–2017. Published by the Director, Zoological Survey of India. Ministry of Environment, Forest and Climate Change. Govt. of India. http://zsi.gov.in.

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.

Arun Kumar L. 2016. *Glyptothorax pasighatensis*, a new species of catfish (Teleostei: Sisoridae) from Arunachal Pradesh, Northeastern India. Int J Pure App Zool. 4:179–185.

Avise JC. 1994. Molecular markers, natural history and evolution. New York: Chapman & Hall.

Barman AS, Singh M, Singh SK, Saha H, Singh YJ, Laishram M, Pandey PK. 2018. DNA barcoding of freshwater fishes of Indo-Myanmar biodiversity hotspot. Sci Rep. 8:8579

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN, et al. 2012. The effect of geographical scale of sampling on DNA barcoding. Syst Biol. 61:851–869.

Bhattacharjee MJ, Laskar BA, Dhar B, Ghosh SK. 2012. Identification and re-evaluation of freshwater catfishes through DNA barcoding. PLoS One. 7:e49950

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol (Amst.). 22:148–155.

Carrizo SF, Jähnig SC, Bremerich V, Freyhof J, Harrison I, He F, Langhans SD, Tockner K, Zarfl C, Darwall W, et al. 2017. Freshwater megafauna: flagships for freshwater biodiversity under threat. BioScience. 67:919–927.

Chakraborty M, Ghosh SK. 2014. An assessment of the DNA barcodes of Indian freshwater fishes. Gene. 537:20–28.

Chen W, Ma X, Shen Y, Mao Y, He S. 2015. The fish diversity in the upper reaches of the Salween River, Nujiang River, revealed by DNA barcoding. Sci Rep. 5:17437.

Chinlampianga M, Singha RK, Shukla AC. 2013. Ethnozoological diversity of Northeast India: Empirical learning with traditional knowledge holders of Mizoram and Arunachal Pradesh. Ind J Tradit Know. 12:18–30.

Collins RA, Armstrong KF, Meier R, Yi Y, Brown SDJ, Cruickshank RH, Keeling S, Johnston C. 2012. Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. PLoS One. 7:e28381.

Collins RA, Cruickshank RH. 2013. The seven deadly sins of DNA barcoding. Mol Ecol Resour. 13:969–975.

Conte-Grand C, Britz R, Dahanukar N, Raghavan R, Pethiyagoda R, Tan HH, Hadiaty RK, Yaakob NS, Rüber L. 2017. Barcoding snakeheads (Teleostei, Channidae) revisited: discovering greater species diversity and resolving perpetuated taxonomic confusions. PLoS One. 12:e0184017.

Darshan A, Abujam S, Das DN. 2018. Biodiversity of fishes in Arunachal Himalaya: systematics, classification, and taxonomic identification. 1st ed. London: Academic Press. 284 pp.

Darshan A, Kachari A, Dutta R, Ganguly A, Das DN. 2016. *Amblyceps waikhomi*, a new species of catfish (Siluriformes: Amblycipitidae) from the Brahmaputra drainage of Arunachal Pradesh, India. PLoS One. 11:e0147283.

Dhar B, Ghosh SK. 2015. Genetic assessment of ornamental fish species from North East India. Gene. 555:382–392.

Díaz J, Villanova GV, Brancolini F, Del Pazo F, Posner VM, Grimberg A, Arranz SE. 2016. First DNA barcode reference library for the identification of South American Freshwater Fish from the Lower Paraná River. PLoS One. 11:e0157419.

Ebach MC, Holdrege C. 2005. DNA barcoding is no substitute for taxonomy. Nature. 434:697.

Eschmeyer WN. 2012. Catalog of fishes. San Francisco (CA): California Academy of Sciences; [accessed 2019 Jan 31] Available: http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

Gopal B. 2015. Guidelines for rapid assessment of biodiversity and ecosystem services of wetlands. Version 1.0. Jahangirpuri, Delhi: National Institute of Ecology, 134 pp.

Goswami UC, Basistha SK, Bora D, Shyamkumar K, Saikia B, Changsan K. 2012. Fish diversity of North East India, inclusive of the Himalayan and Indo Burma biodiversity hotspots zones: a checklist on their taxonomic status, economic importance, geographical distribution, present status and prevailing threats. Int J Biodivers Conserv. 15:592–613.

He SP, Cao WX, Chen YY. 2001. The uplift of Qinghai-Xizang (Tibet) Plateau and the vicariance speciation of glyptosternoid fishes (Siluriformes: Sisoridae). Sci China C Life Sci. 44:644–651.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci. 270:313–321.

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*. Proc Natl Acad Sci USA. 101:14812–14817.

Hebert PD, Ratnasingham S, deWard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci. 270:S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. Plos Biology. 2:e312–1663.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, Watkinson D, Dumont P, Curry A, Bentzen P, et al. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One. 3:e2490.

Jiang W, Ng HH, Yang J, Chen X. 2011. Monophyly and phylogenetic relationships of the catfish genus *Glyptothorax* (Teleostei: Sisoridae) inferred from nuclear and mitochondrial gene sequences. Mol Phylogenet Evol. 61:278–289.

Johns GC, Avise JC. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. Mol Biol Evol. 15:1481–1490.

Kaushik G, Bordoloi S. 2016. Ichthyofauna of Ranganadi River in Lakhimpur, Assam, India. Check List. 12:1872.

Khedkar GD, Jamdade R, Naik S, David L, Haymer D. 2014. DNA barcodes for the fishes of the Narmada, one of India's longest rivers. PLoS One. 9:e101460.

Kullander SO, Rahman MM, Norén M, Mollah AR. 2017. *Devario* in Bangladesh: species diversity, sibling species, and introgression within danionin cyprinids (Teleostei: Cyprinidae: Danioninae). PLoS One. 12: e0186895.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGAX: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 35:1547–1549.

Lakra WS, Singh M, Goswami M, Gopalakrishnan A, Lal KK, Mohindra V, Sarkar UK, Punia PP, Singh KV, Bhatt JP, et al. 2016. DNA barcoding Indian freshwater fishes. Mitochondrial DNA A DNA Mapp Seq Anal. 27:4510–4517.

Lalramliana Knight JDM, Lalhlimpuia DV, Singh M. 2018. Integrative taxonomy reveals a new species of snakehead fish, *Channa stiktos* (Teleostei: Channidae), from Mizoram, North Eastern India. Vertebr Zool. 68:165–175.

Laskar BA, Bhattacharjee MJ, Dhar B, Mahadani P, Kundu S, Ghosh SK. 2013. The species dilemma of Northeast Indian Mahseer (Actinopterygii: Cyprinidae): DNA barcoding in clarifying the riddle. PLoS One. 8:e53704.

Laskar BA, Kumar V, Kundu S, Darshan A, Tyagi K, Chandra K. 2019. DNA barcoding of fishes from River Diphlu within Kaziranga National Park in northeast India. Mitochondrial DNA A DNA Mapp Seq Anal. 30: 126–134.

Laskar BA, Kumar V, Kundu S, Tyagi K, Chandra K. 2018. Taxonomic quest: validating two mahseer fishes (Actinopterygii: Cyprinidae) through molecular and morphological data from biodiversity hotspots in India. Hydrobiologia. 815:113–124.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 23:127–128.

Meier R, Zhang G, Ali F. 2008. The use of mean instead of smallest inter-specific distances exaggerates the size of the 'barcoding gap' and leads to misidentification. Syst Biol. 57:809–813.

Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3:e422.

Molur S, Walker S. 1998. Report of the workshop 'Conservation, assessment and management plan (CAMP) for freshwater fishes of India'. Coimbatore: Zoo Outreach Organization, Conservation Breeding Specialist Group. 156 p.

Moritz C, Cicero C. 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2:e354.

Moyon WA, Arunkumar L. 2018. *Garra moyonkhulleni*, a new labeonine species (Cyprinidae: Labeoninae) from Manipur, Northeastern India. Int J Fish Aquat. 6:107–115.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403:853–858.

Nagpure NS, Rashid I, Pathak AK, Singh M, Pati R, Singh SP, Sarkar UK. 2015. FMiR: a curated resource of mitochondrial DNA information for fish. PLoS One. 10:e0136711.

Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Uppsala: Evolutionary Biology Centre, Uppsala University.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Ruber L, Britz R, Kullander SO, Zardoya R. 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. Mol Phylogenet Evol. 32:1010–1022.

Shen YY, Chen X, Murphy RW. 2013. Assessing DNA barcoding as a tool for species identification and data quality control. PLoS One. 8: e57125.

Steinke D, Zemlak TS, Hebert P. 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLoS One. 4:e6300.

Stuckas H, Gemel R, Fritz U. 2013. One extinct turtle species less: *Pelusios seychellensis* is not extinct, it never existed. PLoS One. 8:e57116.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2003. A plea for DNA taxonomy. Trends Ecol Evol. 18:70–74.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Tyagi K, Kumar V, Singha D, Chandra K, Laskar BA, Kundu S, Chakraborty R, Chatterjee S. 2017. DNA barcoding studies on Thrips in India: cryptic species, species complexes. Sci Rep. 7:1–14.

Vishwanath W, Nebeshwar K, Lokeshwar Y, Shangningam BD, Rameshori Y. 2014. Freshwater fish taxonomy and a manual for identification of fishes of North-east India. Imphal, Manipur, India: Manipur University & National Bureau of Fish Genetic Resources.

Ward RD, Hanner R, Hebert PD. 2009. The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol. 74:329–356.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert P. 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360: 1847–1857.

Yang L, Arunachalam M, Sado T, Levin BA, Golubtsov AS, Freyhof J, Friel JP, Chen W-J, Hirt MV, Manickam R, et al. 2012. Molecular phylogeny of the cyprinid tribe Labeonini (Teleostei: Cypriniformes). Mol Phylogenet Evol. 65:362–379.

Yang L, Sado T, Vincent Hirt M, Pasco-Viel E, Arunachalam M, Li J, Wang X, Freyhof J, Saitoh K, Simons AM, et al. 2015. Phylogeny and polyploidy: resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). Mol Phylogenet Evol. 85:97–116.