Identification and Re-Evaluation of Freshwater Catfishes through DNA Barcoding

Maloyjo J. Bhattacharjee, Boni A. Laskar, Bishal Dhar, Sankar K. Ghosh*
Department of Biotechnology, Assam University, Assam, India

Abstract

Background: Catfishes are globally demanded as human food, angling sport and aquariums keeping thus are highly exploited all over the world. North-East India possess high abundance of catfishes and are equally exploited through decades. The strategies for conservation necessitate understanding the actual species composition, which is hampered due to sporadic descriptions of the species through traditional taxonomy. Therefore, actual catfish diversity in this region is important to be studied through the combined approach of morphological and molecular technique of DNA barcoding.

Methodology/Principal Findings: Altogether 75 native catfish specimens were collected from across the North-East India and their morphological features were compared with the taxonomic keys. The detailed taxonomic study identified 25 species belonging to 17 genera and 9 families. The cytochrome oxidase c subunit-I gene fragment were then sequenced from the samples in accordance with the standard DNA barcoding protocols. The sequences were compared with public databases, viz., GenBank and BOLD. Sequences developed in the current study and from databases of the same and related taxa were analyzed to calculate the congeneric and conspecific genetic divergences using Kimura 2-parameter distance model, and a Neighbor Joining tree was created using software MEGAS.1. The DNA barcoding approach delineated 21 distinct species showing 4.33 folds of difference between the nearest congeners. Four species, viz., Amblyceps apangi, Glyptothorax telchitta, G. trilineatus and Erethistes pusillus, showed high conspecific divergence; hence their identification through molecular approach remained inconclusive. On the other hand, the database sequences for three species, viz., Mystus horai, Bagarius yarrelli and Clarias batrachus, appeared mislabeled.

Conclusion: The efficiency of DNA barcoding was reaffirmed from its success by easily identifying the major share (84%) of the studied catfish into 21 distinct species. The study contributed 27 new barcodes for 7 species and confirmed the range expansion of 2 important species in NE India.

Introduction

Catfishes are members of the order Siluriformes (Actinopterygii) and inhabit inland and marine ecosystems. They are generally bottom dwellers and feed upon almost any kind of plant or animal matters, hence, play an important role in transferring energy throughout the food web [1]. Most of the species hold demand all over the world, including North-East (NE) India, as human food and aquarium keeping [2]. However, due to pressure from unregulated harvest for commercial sale along with other anthropogenic and environmental threats in NE India, some of the native catfish species have become threatened (www.iucnredlist.org). NE India is rich in biodiversity and shares two of the Biodiversity Hotspots in the world, viz., the Eastern Himalaya and the Indo-Burma [3]. The region is bestowed with numerous water bodies of diverse nature and is home to around 267 species of fishes, including many endemic catfishes such as Amblyceps apangi, Amblyceps annachalensis, Glyptothorax striatus [4]. The inventories of fishes from this region were plethoric and entirely based on conventional taxonomy. Due to adherent impediments with traditional taxonomy [5], a few species once claimed new are remarked to be not valid. Instead, many cases of synonym species have been uncovered and await taxonomic revision. For example, the occurrence of M. vittatus in NE has been debated repeatedly [6,7,8] and is a great concern for the systematic. Moreover, taxonomic confusions exist with some other species of the genera Sperata, Ompok, Eutropiichthys, Clupisoma, Gagata and Nangra. The congeners of Sperata are distinguishable by either round or spatula-shaped snout or length of maxillary barbells that either extends to base of caudal fin or no further than pelvic fins. The congeners of Eutropiichthys are differentiable based on length of maxillary barbells and number of fin rays. Since all such characters are prominent only in adults hence the specimens at early stage are difficult to identify [6]. Ompok bimaculatus has been considered “restricted in southern India” and its conspecificity throughout the Indian subcontinent was remarked to be doubtful. Rather, the populations of the species from different areas of the subcontinent were assumed to be representing different species [9,10]. Nevertheless, the congeners of Amblyceps from NE India were often synonymized with Amblyceps mangois [11]. Pterygoplichthys indicus
was often described endemic in NE India but mostly remarked doubtful about its validity [12]. Many such cases are due to unresolved issues related to proper documentation of catfish diversity in NE India. Indeed, the earlier inventories seem to be non-exclusive and sporadic and a correct checklist of catfish diversity in the region is unavailable. Therefore, evaluation of actual catfish diversity using molecular tools is important to resolve the perplexity in identification of many catfish species of the region can be resolved and made easier through the intervention of the advanced DNA barcode based species identification technique.

The DNA barcoding concept has been launched as a rapid, accurate, automatable, and globally accessible procedure for species delimitation and identification [13]. The effectiveness of this method relies on the relatively conserved stretch of approximately 655 nucleotides of the mitochondrial cytochrome oxidase c subunit-I (COI) gene. Based on the nucleotide sequences, accurate identification of organisms at the species level is reasonably straightforward and has been applied to numerous animal taxa [14,15,16,17]. The DNA barcode reference library is rapidly growing by the contributions of the global community in the Barcode of Life Data Systems (BOLD) [18]. With the glory as an attractive species identifier in fish biodiversity research [19], the application of DNA barcoding has recently reflected 28% increase of North American freshwater fish diversity [20]. The technique further bears application in monitoring fish products for health safety [21,22] and in regulating the exploitation of fish species under aquarium trade [23,24,25].

DNA barcoding technique was adopted in NE India to study the actual diversity of catfishes inhabiting in the region. This will also enrich the database by contributing both new barcodes and

### Table 1. Summary of identification based on each species consensus barcoded sequence using BLASTN search from GenBank and BOLD Identification System (BOLD-IDS).

| Sl. No. | Studied species | Species match by name | GenBank (BLASTN) | BOLD-IDS | % Similarity | GenBank (BLASTN) | BOLD-IDS |
|---------|-----------------|-----------------------|------------------|----------|--------------|------------------|----------|
| 1.      | Rita rita (1)   | Rita rita             | Rita rita        | 99       | 100          | Rita rita        | 99       |
| 2.      | Mystus bleekeri (8) | Mystus bocourtii     | No match         | 88       | No match     | Mystus bocourtii | 89       |
| 3.      | M. cavusi (4)   | Mystus ocoulatus      | No match         | 89       | No match     | Mystus ocoulatus | 89       |
| 4.      | M. vittatus (7) | Mystus vittatus       | Mystus vittatus  | 99       | 99.83        | Mystus horai     | 99       |
| 5.      | Sperata aor (3) | Sperata aor           | No match         | 86       | No match     | Sperata aor     | 86       |
| 6.      | Hembagurus menoda (2) | Hembagurus menoda | No match         | 100      | 100          | Hembagurus menoda | 100     |
| 7.      | Bagarius bagarius (3) | Bagarius bagarius   | Bagarius bagarius | 100 | 100          | Bagarius bagarius | 100 |
|         | Bagarius yarrelli | Bagarius yarrelli    | Bagarius yarrelli | 100 | 100          | Bagarius yarrelli | 100 |
| 8.      | Gagata cenia (6) | Gagata cenia          | Gagata cenia     | 99       | 99.5         | Gagata cenia    | 99       |
| 9.      | Gagata sexualis (2) | Gagata sexualis     | Gagata sexualis | 99       | 99           | Gagata sexualis | 99       |
| 10.     | Glyptothorax telchitta (3) | Glyptothorax telchitta | No match       | 93       | No match     | Glyptothorax telchitta | 93 |
| 11.     | G. striatus (1)  | Glyptothorax striatus | Glyptothorax striatus | 97 | 97.6         | G. striatus     | 97       |
| 12.     | G. trilineatus (1) | Glyptothorax trilineatus | No match       | 96       | No match     | G. trilineatus | 96       |
| 13.     | Sisor rhabdophorus (4) | Sisor rhabdophorus   | Sisor rhabdophorus | 100 | 100          | Sisor rhabdophorus | 100 |
| 14.     | Ailia coila (3)  | Ailia coila           | Ailia coila      | 99       | 99.67        | Ailia coila     | 99       |
| 15.     | Clupisoma garua (1) | Laiides hexanema     | Laiides hexanema | 91       | No match     | Clupisoma garua | 91 |
| 16.     | Eutropiichthys murius (3) | Pangasius karnauidi | Pangasius karnauidi | 86   | No match     | Eutropiichthys murius | 86 |
| 17.     | E. vacha (7)     | Laiides hexanema     | Laiides hexanema | 89       | No match     | E. vacha        | 89       |
| 18.     | Ompok bimaculatus (2) | Ompok bimaculatus   | Ompok bimaculatus | 99 | 99.84        | Ompok bimaculatus | 99       |
| 19.     | O. pabo (2)      | O. pabo               | O. pabo          | 99       | 99.33        | O. pabo         | 99       |
| 20.     | Wallago attu (2) | Wallago attu          | Wallago attu     | 100      | 100          | Wallago attu    | 100     |
| 21.     | Clarias batrachus (2) | Clarias batrachus   | Clarias batrachus | 98 | 96.64        | Clarias batrachus | 98    |
|         | Clarias batrachus | Clarias batrachus    | Clarias batrachus | 98 | 96.64        | Clarias batrachus | 98 |
| 22.     | Heteropeoneutes fossilis (3) | Heteropeoneutes fossilis | No match       | 100 | No match     | Heteropeoneutes fossilis | 100 |
| 23.     | Eretisthes pusilis (2) | Eretisthes pusilis | Eretisthes pusilis | 93 | No match     | Eretisthes pusilis | 93 |
| 24.     | Amblyceps apangi (1) | Amblyceps apangi    | Amblyceps apangi | 95       | No match     | Amblyceps apangi | 95 |
| 25.     | Olyra longicaudata (1) | Olyra longicaudata  | Olyra longicaudata | 90 | No match     | Olyra longicaudata | 90 |

- Similarity description used in the study: 97%–100%– significant, 92%–96%– moderate, ≤91%– insignificant.
- Bolded words correspond to problematic identification of species in the present study using either one or both the databases. Details are further discussed in the text.
- Numbers in brackets indicate the number of individual sequences of each species.

<doi:10.1371/journal.pone.0049950.t001>
replicating existing barcodes thereby enabling the evaluation of taxonomic status of the native catfish diversity in NE India. Here we studied the first DNA barcode based taxonomic resolution of freshwater catfishes from NE India to resolve key areas of doubt arising from morphological taxonomy. This investigation not only proves the potential use of DNA barcoding as a tool to aid traditional taxonomy of freshwater catfish but also will help further in easy identification of the studied species from any of their body parts and at any stage of life. Nevertheless the sequences generated from this study would be accessible to establish the conspecificity of NE Indian catfish with other geographical location and vice versa.

Materials and Methods

Sample collection

The native catfishes were collected from different natural water bodies during different seasons of the year from random geographical locations within the NE region of India (28°09′N 97°24′E on the East to 27°49′N 88°15′E on the West, and 29°18′N 96°04′E on the North to 29°59′E on the South). Fresh specimens were spot examined for specific morphological characters that define the catfish and sampled from different wild habitats as and when caught by the professional fishers. Each of the catch was investigated by an experienced fish taxonomist to ensure correct sampling and labeling. Upon every spate of collection, the tissue samples from each of the specimens were collected aseptically and preserved in 90% ethanol. Major taxonomic keys of each of the fish specimens were noted and measurements were taken using a digital caliper. Species level identification of the specimens was confirmed by comparing with the described characters and the taxonomic keys available in the leading taxonomic guides of the fishes in India [6,26]. However, the nomenclature of species follows the Catalogue of Fishes [27]. The comparisons of the observed characters in respect of each species with their described characters along with the particular dispute are presented in Table S1. Altogether 75 fish specimens belonging to 25 species within 17 genera and 9 families were collected and included in this study. All the voucher specimens have been deposited in the Department of Biotechnology, Assam University, Silchar. The specimen information, IUCN Red list status and distribution of the studied species are given in Table S2.

Since the studied fishes were routine caught by the professional fishers for sale hence no permission was required for their sampling.

DNA extraction

20 mg of anal fin tissue was taken aseptically and dissolved in 500 μL of TES buffer (50 mM Tris HCl, 25 mM EDTA and 150 mM NaCl) in a microcentrifuge tube. The extraction of DNA was performed with Phenol-Chloroform-Isoamylalcohol method [28].

Figure 1. Neighbor joining (NJ) tree developed using K2P distance among 101 CO1 sequences. Notable anomalies in clustering are shown by 4 species (Mystus horai (accession number FJ170791), Bagarius yarrelli (accession number DQ508069), Clarius batrachus (accession number HQ654701) and Amblyceps apangi (accession number EU490873)). Deep conspecific divergences are shown by 3 species (Glyptothorax trilineatus, Glyptothorax telchitta and Erethistes pusillus). * The numbers at the nodes are bootstrap values based on 1000 replications. Specimen GenBank accession number and species name are shown for each taxon. Red and black dots correspond to the sequences acquired from database. Red dot alone corresponds to the cases of abnormal clustering and deep conspecific divergence.

doi:10.1371/journal.pone.0049950.g001
collected specimens. The COI amplicons were recovered from all the sequencer (ABI 3500, Applied Biosystems Inc., CA, USA), amplicons were bidirectionally sequenced in an automated DNA kit (QIAGEN, USA), following manufacturer’s instructions. The uniform band was then purified using QIAquick® Gel extraction kit (Qiagen, USA). The PCR-amplified products were analyzed in 1% agarose gels with two elongation step at 72°C for 1 minute, and concluded with a final elongation step at 72°C for 8 minutes followed by a hold at 4°C. The PCR-amplified products were analyzed in 1% agarose gels containing ethidium bromide staining (10 mg/ml) and the single uniform band was then purified using QIAquick® Gel extraction kit (Qiagen, USA), following manufacturer’s instructions. The amplicons were bidirectionally sequenced in an automated DNA sequencer (ABI 3500, Applied Biosystems Inc., CA, USA), through the best known service of GCC Biotech India Pvt. Ltd. (Kolkata, India). The COI amplicons were recovered from all the collected specimens.

**PCR amplification and purification**

The COI gene (655 bp) was amplified using the set of published primers [29] as follows. FishF1-5'/ TCAACCAACGCAAGAA-GATTGGCAC 3' and FishR1-5'/ TAGACTTCTGGGTGGC-CAAAGATC 3'/ in a Veriti Mastercycler (Applied Biosystems Inc., CA, USA). The amplification reactions were performed in a total volume of 25 μl comprising 1X PCR buffer, 2 mM MgCl₂, 10 pmol of each primer, 0.25 mM of each dNTPs, 0.25 U of high-fidelity Taq polymerase (Applied Biosystems Inc., CA, USA) and 100 ng of DNA template. The thermal profile of the PCR reaction was as follows: An initial denaturation at 94°C for 2 minutes, 30 cycles at denaturation temperature of 94°C for 45 seconds, annealing temperature of 50°C for 45 seconds and elongation temperature of 72°C for 1 minute, and concluded with a final elongation step at 72°C for 8 minutes followed by a hold at 4°C. The PCR-amplified products were analyzed in 1% agarose gels containing ethidium bromide staining (10 mg/ml) and the single uniform band was then purified using QIAquick® Gel extraction kit (Qiagen, USA), following manufacturer’s instructions. The amplicons were bidirectionally sequenced in an automated DNA sequencer (ABI 3500, Applied Biosystems Inc., CA, USA), through the best known service of GCC Biotech India Pvt. Ltd. (Kolkata, India). The COI amplicons were recovered from all the collected specimens.

**Sequence quality control measures**

Both the PCR amplified products and their corresponding DNA sequences were larger than 600 bp that assured no NUMTs being amplified as the limit of NUMIT hardly reaches 600 bp [30]. Ends of the noisy sequences were trimmed and more than 600 bp sequences were used for the final analysis (except in four cases, accession numbers JN697602, JN628915, JN628929 and JN628930). For each sample two chromatograms that represent sequences of both the strands of DNA were obtained. BLASTN [31] program was used to compare the sequences from the two chromatograms, and the fragment of the two sequences showing 100% alignment with no gap or indels (insertion/deletions) was selected. In case of any discrepancy, both the sequences were reviewed and quality value of the sequence was considered to determine the most likely nucleotide using the software SeqCleaner Version 1.0 (Applied Biosystems Inc., CA, USA). In most cases, the sequence quality values were above 50. The selected fragments of the sequences for all the specimens were aligned using ClustalX software [32] and found no indels in any of the sequences. Finally, each of the sequences were subjected to BLASTN searches at the National Centre for Biotechnology Information [31], that showed alignment with the partial coding sequence of fish mitochondrial COI gene without any indels. The sequences were translated using the online software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and aligned through BLASTP [31] that revealed coherent partial amino acid codes with fish mitochondrial COI gene frame without any stop codon. Therefore, it was confirmed that the generated sequences were fragments of mitochondrial COI gene. All the analyzed sequences were then deposited in GenBank and received valid accession numbers (Table S1). The sequences were also submitted by creating a FISH-BOL project in BOLD in the code name of ‘CFISH’ entitled “DNA barcoding of freshwater catfishes of Northeast India”.

**Data analysis**

The BOLD provided 778 COI sequences for catfish (accessed on 17 November, 2011) among them the database sequences of the same and/or related taxa are used in association with developed sequences for evaluating the taxonomic status of our target species. The total dataset included 101 COI barcode sequences for 27 catfish species among which 75 sequences belonging to 25 species were developed de-novo, and 26 sequences representing same and related taxa were acquired from GenBank only because there were no additional sequences available in BOLD other than those mined from GenBank source. Geograph-
The Neighbor Joining (NJ) cluster analysis (Figure 1) revealed straight forward identification showing either a single or distinct cluster of individual(s) for 10 of our studied species. These include 11 accurately identified species which showed significant species specific similarity and 7 species which showed insignificant non-specific match (in parity with ambiguous match category-5). However, ambiguities persisted for 7 other species and showed three distinct patterns: 1) same and different named-species clustered together (in parity with ambiguous match category-1, e.g., all sequences of M. vitatus clustered with M. horai (accession number EJ170791) and all sequences of B. bagarius clustered with B. yarrelli (accession number DQ508069), 2) same named-species clustered both jointly and distinctly (in parity with ambiguous match category-2, e.g., C. batrachus of accession number GQ466402 and HQ534701), and same named-species clustered only distinctly (in parity with ambiguous match category-4, e.g., A. apangi of accession number EU490873), and 3) high range of clustering differences with conspecific query sequences (in parity with ambiguous match category-3, e.g., G. telchitta of accession number DQ514362, G. trilineatus of accession number DQ508077, E. pusillus of accession number DQ508074 and DQ508079, and A. apangi of accession number DQ508066).

The minimum congeneric and maximum conspecific K2P divergences were determined to be 0.104 and 0.024 respectively and presented in Figure 2. Based on these divergence values, a 4.33 folds barcoding gap was calculated. 72% (18) of the studied species identified through the NJ clustering have also been easily delineated following the barcoding gap. Few database sequences of the same and/or related species have not obeyed this gap and hence designated as ambiguous. For example, 1) a few congeneric sequences merged within the range of conspecific divergence (e.g., M. horai of accession number EJ170791 and B. yarrelli of accession number DQ506089) and vice versa (e.g., C. batrachus of accession number HQ534701 and A. apangi of accession number EU490873), 2) a few individual sequences were widely dispersed from their conspecific sequences and did not reach the congeneric threshold divergence (e.g., G. trilineatus of accession number DQ508077, G. telchitta of accession number DQ514362, E. pusillus of accession numbers DQ508074 and DQ508079 and A. apangi of accession number DQ508066), and finally 3) congeneric distances of G. striatus with G. trilineatus have remained much below the congeneric threshold divergence.

**Discussion**

This study of identification of catfishes from NE India was based on the morphological investigation followed by DNA barcoding approach. The morphological study of the specimens has raised a few questions on the observed features versus the described features. In a few cases, morphological species keys were difficult to discern. Moreover, disparities relating to the species keys were observed in a few cases between the two leading taxonomic guide books of fishes in India (Table S1). The DNA barcoding approach resolved some identification issues and explained the actual species composition in the region.

Among the 25 studied catfish species, the similarity search approach revealed two straightforward cases for 18 species. Firstly, 11 species, viz., R. rita, S. aor, G. cenia, G. sexualis, G. straitus, S. rabdophorus, A. cola, O. bimaculatus, O. paho, W. attu and H. fossili, H. fossili latter was identified by GenBank alone. The rest of the studied species (14) showed ambiguous match categories, like, 1) significant but equally species specific and non-specific (e.g. M. vitatus and B. Bagarius), 2) species specific but equally significant and insignificant (e.g. C. batrachus), 3) species specific but moderate (e.g. G. telchitta, G. trilineatus, E. pusillus and A. apangi), 4) species specific but insignificant (e.g. A. apangi accession number EU490873), and 5) non-specific and insignificant (e.g. M. bleekeri, M. casavis, H. menoda, C. gurra, E. murius, E. vacha and O. longicaudata).

The Neuristic information and GenBank accession numbers of the developed as well as acquired database sequences are given in Table S2. The sequence similarity search for species identification was done in two public databases, viz., BOLD and GenBank. The highest percent pairwise identity for each sequence blasted (BLASTN) at NCBI were compared with the percent similarity scores of the same sequence within the BOLD-IDS (BOLD Identification System) [10]. The query species that matched either with the same or different species in the databases has been termed as ‘specific’ or ‘non-specific’ respectively. The similarity range of 97%–100%, 92%–96% and ≤91% between the query and the database sequence have been expressed as significant, moderate and insignificant respectively, Kimura 2-parameter (K2P) congeneric and conspecific variation [33] and Neighbor Joining (NJ) tree construction were done using the computer program MEGA Version 5 [34]. Maximum conspecific and minimum congeneric divergences have been determined considering the sequences showing cohesive NJ clustering within a species and remained distinct from other species. The number of times the minimum congeneric divergence differs from the maximum conspecific divergence is the lowest divergence between congener and has been assumed to be the threshold level of species delineation and thereby considered as a barcoding gap in this study.

**Results**

Comprehensive species identification of the studied catfishes based on BOLD and GenBank databases is depicted in Table 1. The study helped in straightforward identification of 10 species that showed significant species specific similarities in both the databases. The species are Sperata aor, Sisor rabdophorus, Wallago attu, Gagata sexualis, Rita rita, Gagata cenia, Glyptothorax straitus, Aila cola, Ompok bimaculatus and Ompok paho. GenBank sequences showed moderate species specific similarity for both Glyptothorax telchitta and Erethistes pusillus at 93%, G. trilineatus at 96% and significant species specific similarity for Heteropneustes fossilis at 100%. It also showed insignificant non-specific similarity (≤91%) for seven species, viz., Mystus bleekeri, M. casavis, Hemibagrus menoda, Chrysoma gurra, Eutropichthys narias, E. vacha and Olra longicaudata. Both the databases showed significant species specific similarity (≥98%) for M. vitatus, B. Bagarius and Clarias batrachus, as well as significant non-specific similarity (≥99%) for M. vitatus (query) with M. horai (database accession number EJ170791), and B. yarrelli (query) with B. yarrelli (database accession number DQ508069). GenBank alone showed insignificant species specific similarity (90%) for C. batrachus with a database accession number HQ534701. GenBank concurrently showed moderate species specific similarity (95%) and insignificant species specific similarity (88%) for Amblyphyes apangi with database accession numbers DQ508066 and EU490873 respectively.

The similarity search result thereby confirmed definitive identity showing significant species specific match in GenBank and BOLD for 11 species, viz., R. rita, S. aor, G. cenia, G. sexualis, G. straitus, S. rabdophorus, A. cola, O. bimaculatus, O. paho, W. attu and H. fossili. H. fossili latter was identified by GenBank alone. The rest of the studied species (14) showed ambiguous match categories, like, 1) significant but equally species specific and non-specific (e.g. M. vitatus and B. Bagarius), 2) species specific but equally significant and insignificant (e.g. C. batrachus), 3) species specific but moderate (e.g. G. telchitta, G. trilineatus, E. pusillus and A. apangi), 4) species specific but insignificant (e.g. A. apangi accession number EU490873), and 5) non-specific and insignificant (e.g. M. bleekeri, M. casavis, H. menoda, C. gurra, E. murius, E. vacha and O. longicaudata).
reflected the lack of barcode reference data for these species in both the databases. However, all the sequences of the above mentioned 18 species showed conspecific NJ clustering by the specimens within each species having well supported bootstrap proportion (>95%) [35]. Further, all the 18 species were definitely delineated considering barcoding gap principle (Figure 2) and identified as true species based on the combined approaches, including 7 species whose barcode data were not available in the databases previously. So, the study contributed new barcode data in the global database for those seven species. Range expansion of O. hinnacularus in NE India was also evident from this study through observed high conspecificity among the queries and database sequences from the Ganga basin (FH230073–4) in India. We concentrated to delineate the species based on the threshold level of species divergence taking the maximum conspecific versus minimum congeneric divergence into account rather than considering the conventional mean value of congeneric and conspecific divergence. This has led to the reflection of the lowest barcoding gap of 4.33 folds compared to the previous DNA barcoding studies of fishes [23,29,36,37,38].

While morphological identification were convincing up to species-level for all the studied species, DNA barcoding even remained inconclusive for 7 species (viz., M. vittatus, B. bagarius, C. batrachus, A. apangi, G. trilineatus, G. telchita, and E. pusillus), at its first hand approach. Combined approach has confirmed that database sequence of M. horai and M. vittatus are conspecific. Moreover, the query sequences of M. vittatus showed conspecificity with a same named database sequence of accession number DQ508093. Previous morphological studies have already raised doubtfulness on the taxonomic validity of M. horai and the species was remarked to be not recorded from any other location than its type locality (in Indus drainage) [27,39]. Since the molecular evidences also reckon the previous morphological debate, hence the study tentatively considered M. horai as a synonym of M. vittatus. As such the study identified M. vittatus being a true species and recognized its range expansion in NE India. Therefore, the study resolved the debate surrounding the existence of M. vittatus in NE India [7,8]. Again, one of the two database sequences of B. yarrelli (accession number DQ508069) revealed conspecific divergence with B. bagarius while the other sequence (accession number EU490855) maintained congeneric divergence with the same. This reflected that the former sequence of B. yarrelli is mislabeled in the database. This study thus confirmed both B. bagarius and B. yarrelli being true species and met with the previous argument [40]. In case of C. batrachus, a single database sequence (accession number HQ654701) showed congeneric divergence with other conspecific sequences. On the other hand, the queries as well as the other database sequences of C. batrachus showed conspecificity. This confirmed C. batrachus to be a true species and indicated a clear mislabeling of the said sequence in the database. Similar cases of mislabeling have also been reported earlier [22]. Ignoring the few mislabeled database sequences, the identification of above three species was confirmed.

In another case, one of the two database sequences of Amblyceps apangi (accession number EU490873) was identified to be a distinct congener of the query sequence. This may be again a case of mislabeling because the two species within the genus Amblyceps are endemic to NE India [41] and inadequately described or frequently synonymised [11,42]. The other database sequence of A. apangi (accession number DQ508066) and all the database sequences of G. trilineatus, G. telchita, and E. pusillus were not conspecific with the respective query sequences, and remained below the congeneric threshold thereby revealed deep conspecific divergences. This indicated possible cases of independently evolving lineages of a species from different geographical location [20] or cryptic species with low divergence, or even recently-diverged overlooked species [23,29]. As such, the identification of species under the genera, viz., Glyptothorax, Erethistes, and Amblyceps remained inconclusive due to inadequate and perplexing descriptions from conventional taxonomy. Hence, in order to develop a correct barcode reference library, there is a paramount need of extensive revision, combining morphological and DNA barcoding of the extant species under these genera.

Thus, it can be concluded that 21 species representing 84% of the studied catfish species were identified straightforward through DNA barcoding that reaffirmed the efficacy of the technique. The study resolved some cases of synonymy, clarified the range distribution and revealed the catfish diversity in NE India. Occurrence of Percocypris indicus in NE India was not evident in the study and hence holds to agree upon arguments on the doubtful status of this species [12]. However, remaining 16% of the studied species representing 3 genera remained inconclusive and warrant further evaluation. Few database sequences were observed to be bearing misidentified species caption among the species those possess either confusing morphological description or share crypticism. Given that the database is enriched with the multiple sequences for a target species and for the extant species within a target genus from the range of distribution, the species taxonomy would be rectified and assessment of biodiversity would be correct and easier.

Supporting Information

Table S1 Morphological taxonomic keys observed versus described. (DOC)

Table S2 List of the studied species, GenBank accession numbers of the analyzed sequences, the geographical position, and IUCN status. (DOC)

Acknowledgments

Our humble acknowledgement goes to the Department of Biotechnology, Government of India for providing infrastructure facilities (BT/HRD/01/002/2007) and post-doc fellowship to Dr. Boni A Laskar in the Department of Biotechnology, Assam University Silchar. Comments on the manuscript from reviewers have helped to enrich the revised version to present in a better way and are highly acknowledged.

Author Contributions

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

References

1. Layher WG, Maughan OE (1985) Relations between habitat variables and channel catfish populations in prairie streams. Transactions of the American Fisheries Society 114: 771–781.

2. FAO (2008–2012) Small-scale fisheries-website. Nutrition and food security. In: www.fao.org/fishery/topic/16603/en. Accessed 2012 Mar 19.
14. Hebert PDN, Stoekle M, Zemlak T, Francis CM (2004) Identification of birds through DNA barcodes. Plos Biology 2: 1657–1668.
15. Clare EB, Lim KKP (1993) Freshwater fishes of Sarawak and Brunei Darussalam: a preliminary annotated checklist. Sarawak Museum Journal 48: 227–258.
16. Ng HH, Hadiaty RK (2009) Ompok brevirictus, new catfish (Teleostei: Siluridae) from Sumatra. Zootaxa 2232: 50–60.
17. Vishwanath W, Linthoingambi I (2007) Redescription of catfishes Amblyceps spp. from the Indus basin. Zootaxa 1418: 1–62.
18. Ferraris CJ Jr (2007) Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes) and catalogue of siluriform primary types. Zootaxa 1418: 1–62.
19. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270: 313–321.
20. Hebert PDN, Stoeckle M, Zemlak T, Francis CM (2004) Identification of birds through DNA barcodes. Plos Biology 2: 1637–1660.
21. 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.
22. Smith MA, Rodriguez JJ, Whittington JB, Deans AR, Janzen DH, et al. (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.
23. 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.
24. Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System. Molecular Ecology Notes 7: 355–364. Available: http://www.barcodinglife.org.
25. Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329–356.
26. Jayaram KC (1999) The freshwater fishes of the Indian region: Narendra Pub. House.
27. Jayaram KC (1999) Catfish Diversity – A DNA Barcoding Approach
28. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.
29. 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.
30. Zhang DX, Hewitt GM (1996) Nuclear integrations: challenge for mitochondrial DNA markers. Trends in Ecology and Evolution 11: 247–251.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
32. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics/editorial board, Andrea D Baxevanis [et al] Chapter 2: Unit 2.3.
33. 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.
34. 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.
35. Felsenstein J (1983) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.
36. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, et al. (2008) Identifying Canadian Freshwater fishes through DNA Barcodes. PLoS ONE 3: e2490.
37. Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, et al. (2011) DNA barcoding of Indian marine fishes. Molecular Ecology Resources 11: 60–71.
38. Mabragaña E, Díaz de Astداء JM, Hanner R, Zhang J, González Castro M (2011) DNA Barcoding Identifies Argentine Fishes from Marine and Brackish Waters. PLoS ONE 6: e20653. doi:10.1371/journal.pone.0020653.
39. Talwar PK, Blyth AG (1991) Inland fishes of India and adjacent countries: Oxford & IBH Pub. Co. Vol 2. pp554–1158.
40. Roberts TR (1983) Revision of south and southeast Asian sisorid catfish genus Bagarius, with description of a new species from the Mekong. Copeia 2: 435–445.
41. Nath P, Dey SC (1989) Two new species of the genus Amblycheptus Blyth from Arunachal Pradesh, India. Journal of Assam science 32: 1–6.
42. Ng HH (2005) Amblycheptus cannam, a new species of hillstream catfish from Myanmar (Teleostei: Amblycipitidae). Raffles bulletin of Zoology 53: 243–249.