Molecular characterization of marine and coastal fishes of Bangladesh through DNA barcodes

Md. Sagir Ahmed[1], Sujan Datta[2], Tonmoy Saha [2], and Zarif Hossain[3]

[1]University of Dhaka
[2]Jagannath University
[3]University of Dhaka Faculty of Earth and Environmental Sciences

June 11, 2020

Abstract

Abstract: This study attempted to molecular characterization of marine and coastal fishes of Bangladesh based on mitochondrial cytochrome c oxidase subunit I (COI) gene as a marker. A total of 376 mitochondrial COI barcode sequences were obtained from 185 species belonging to 146 genera, 74 families, 20 orders and two classes of fishes. The mean length of the sequences was 652 base pairs. For all the samples, %G was significantly lower compared to the other nucleotides and %GC was significantly lower compared to %AT ($p < 0.005$). Also, a significantly lowered %GC content was observed in second and third codon position compared to the first codon position in all the samples (1st>2nd>3rd, p-value < 0.005). In Elasmobranchii (Sharks and rays) the average Kimura two parameter (K2P) distances within species, genera, families and orders were 1.2%, 6.07%, 11.08% and 14.68%, respectively and for Actinopterygii, the average K2P distances within species, genera, families and orders were 0.40%, 6.36%, 14.10% and 24.07%, respectively. The mean interspecies distance was 16-fold higher than the mean intraspecies distance. The K2P neighbor-joining (NJ) trees based on the sequences generally clustered species in accordance with their taxonomic position. A total of 21 species were newly recorded in Bangladesh. High efficiency and fidelity in species identification and discrimination were demonstrated in the present study by DNA barcoding, and we concluded that COI sequencing can be used as an authentic identification marker for Bangladesh marine fish species. Key words: COI, Barcoding, Elasmobranchii, Actinopterygii, Genetic diversity, Phylogeny

ABSTRACT

This study attempted to molecular characterization of marine and coastal fishes of Bangladesh based on mitochondrial cytochrome c oxidase subunit I (COI) gene as a marker. A total of 376 mitochondrial COI barcode sequences were obtained from 185 species belonging to 146 genera, 74 families, 20 orders and two classes of fishes. The mean length of the sequences was 652 base pairs. For all the samples, %G was significantly lower compared to the other nucleotides and %GC was significantly lower compared to %AT ($p < 0.005$). Also, a significantly lowered %GC content was observed in second and third codon position compared to the first codon position in all the samples ($1^{st}>2^{nd}>3^{rd}$,p- value < 0.005). In Elasmobranchii (Sharks and rays)the average Kimura two parameter (K2P) distances within species, genera, families and orders were 1.2%, 6.07%, 11.08% and 14.68%, respectively and for Actinopterygii, the average K2P distances within species, genera, families and orders were 0.40%, 6.36%, 14.10% and 24.07%, respectively. The mean interspecies distance was 16-fold higher than the mean intraspecies distance. The K2P neighbor-joining (NJ) trees based on the sequences generally clustered species in accordance with their taxonomic position. A total of 21 species were newly recorded in Bangladesh. High efficiency and fidelity in species identification and discrimination were demonstrated in the present study by DNA barcoding, and we concluded that COI sequencing can be used as an authentic identification marker for Bangladesh marine fish species.

Key words: COI, Barcoding, Elasmobranchii, Actinopterygii, Genetic diversity, Phylogeny

INTRODUCTION

Bangladesh has vast coastal and marine resources along its south edge as the Bay of Bengal is situated in the south of Bangladesh. There is a total of 166,000 square km water areas including the Exclusive Economic Zone (EEZ) which is larger than the country's total land area of 147,500 square km. The country is rich not only in terms of its vast water areas but also in terms of biological diversity. The marine fisheries sector plays an important role in the economy of Bangladesh in terms of nutrition, income, employment and foreign exchange earnings (DoF, 2018). Fish provides about 60% of animal protein in the daily dietary requirement of 160 million people of the country. Marine fisheries alone contributes 654,687 metric tons which is 15.31% of the country's total fish production (DoF, 2018).

Nonetheless, description and information of marine fishes of Bangladesh are scattered throughout a wide range of scientific publications (Shafi & Quddus, 1982; Rahman et al., 1995; Rahman et al., 2009; Hussain et al., 1971). Estimates of total fish species vary from 170 (Shafi & Quddus 1982) to 402 (Rahman et al., 2009), 442 (IUCN, 2000) or 475 (Hussain et al., 1971) including the migratory and estuarine species. If we compare this number with our neighboring countries, diversity estimates for Bangladeshi marine fishes seems to be underestimated. It is evident from the published articles, books and review papers on Bangladeshi marine fishes (Ahmed et al., 2019a; Rahman et al., 2009; Shafi & Quddus 1982; Hussain et al., 1971; IUCN, 2000) that the ichthyofaunal diversity statistics are incomplete. Moreover, no systematic survey or in-depth taxonomic study was undertaken on marine fish faunal diversity since 1969 (Hussain et al., 1971).

The accurate identification of fish species is a pivotal component to protect the extant ichthyofaunal biodiversity and to perform regular assessments of local fish faunas for conservation planning (Ahmed et al., 2019a). Currently, partial cytochrome c oxidase subunit I (COI) sequences (DNA barcodes) are applying as a standardized and routine species identification (Hebert et al. 2003; Shehata et al., 2019; Chin et al., 2016; Filonzi et al., 2010) instead of traditional mopho taxonomy. The marked divergence and lack of overlap between intraspecific and interspecific genetic distances is the primary reason for the selection of COI as the standard barcode gene (Hebert et al. 2003). More importantly, COI evolution is sufficiently rapid to allow the discrimination of very closely related species in most groups, as well as taxonomically significant intraspecific variation associated with geographic structure (Buckling et al., 2010).

DNA barcoding has been successfully identified marine ichthyofauna and provided the wealth of DNA barcode information in many places, such as Australia (Ward et al., 2005), Canada (Hubert et al., 2008; Steinke et al., 2009), India (Lakra et al, 2011), China (Zhang et al., 2011; Zhang & Hanner, 2012; Wang et al., 2018), Portugal (Costa, et al., 2012), Germany (Knebelsberger et al., 2014), Taiwan (Bingpeng et al., 2018; Chang et al., 2017) and Vietnam (Thu et al., 2019).

Considering the economic importance of marine fishery and the expected richness of the fish fauna and in the absence of an expert based taxonomy, DNA barcoding may be an important component in biological conservation and management of biodiversity and fishery of Bangladeshi marine fishes. As a highly overpopulated country, anthropogenic activities, overfishing, habitat destruction and natural disasters have generated significant impacts on the biodiversity and structure of the fish community in Bangladesh. Unfortunately, the marine ichthyofauna of Bangladesh remains unexplored due to the lack of taxonomists. Hence, adopting an authentic and quick identification method is essential to assist fishery managers, scientists and policy makers for sustainable management of this invaluable marine resources. This study aims to build a DNA based barcode library of the morphologically identified marine and coastal fish species of Bangladesh using partial COI gene sequence.

MATERIALS AND METHODS

2.1 Study area and specimen collection

Fish samples were collected from marine and coastal habitats, fish landing centers, fish markets or from the local fishermen from July 2015 to June 2019. Most of the specimens were collected from the Cox's Bazar and Patuakhali regions. At least three specimens were collected for each species. In case of rare ones single

specimen were analyzed. Personal fishing was also conducted to collect some rare and non-commercial fish species whenever necessary. Digital photographs of all the fishes were taken immediately and taxonomic identification of specimens was done following previous reports (Talwar and Jhingran, 1991; Rahman et al., 2009; Siddiqui et al., 2007; Nakabo, 2002; Carpenter et al., 2002a, 2002b; Last et al., 2016). Immediately after collecting the specimens, tissue samples were excised and stored in 90% ethanol. Voucher specimens were fixed with 10% formalin and then transferred to 70% ethanol solution for preservation. Voucher specimens were transported to Dhaka and deposited in the Dhaka University Zoology Museum (DUZM).

DNA barcoding

DNA was isolated from muscle sample using the QIAGEN DNeasy Blood & Tissue Kit following the manufacturer's protocol, under sterile condition. The concentration of the isolated DNA was measured in NanoDrop spectrophotometer to evaluate its quality and quantity. A 658 bp long fragment from the 5' region of the COI gene was PCR-amplified by the primer pair FishF2 5'TCGACTAATCATAAAGATATCGGCAC3' and FishR2 5'ACTTCAGGGTGACCGAAGAATCAGAA3'. The primer pair FishF1 5'TCAACCAACCACAAAGACATTGGCAC3' and FishR1 5'TAGACTTCTGGGTGGCCAAAGAATCA3' were used for the amplification of COI that failed to amplify using FishF2/FishR2. The PCR amplification of each sample was conducted in a 25 µl volume comprised of 23 µl of PCR Master Mix and 2 µl of template DNA that was subsequently spun for 30 seconds for homogenization. The components of the PCR Master Mix were as such: 12.5 µl Taq Polymerase, 8.5 µl Nano Pure water, 1 µl forward primer and 1 µl reverse primer. The PCR amplification was carried out in the ASTEC Thermal Cycler GeneAtlas (Astec Co. Ltd.) where the thermocycling profile was customized as such: an initial denaturation at 95 for 5 minutes followed by 41 cycles of denaturation at 95 for 30 s, primer annealing at 54 for 30 seconds, extension for 72 for 1 minute, and a final extension at 72 for 5 minutes. The PCR products were kept at room temperature for 15 minutes, and preserved afterward at -20 until further downstream application. The amplicons were further identified through electrophoresis in a 1% agarose gel. Then they were purified using PureLink PCR purification kit. The good quality amplicons were then parceled for sequencing to the First BASE laboratories, Malaysia where they conducted the Sanger's dideoxy sequencing using ABI PRISM 3730xl Genetic Analyzer exploiting the BigDye® Terminator v3.1 cycle sequencing kit chemistry. The assembled contigs were prepared by the CAP3 DNA assembly program. These newly obtained sequences were uploaded in the BLASTn suite to check whether they meet the threshold value of [?] 97% for both the percent identity and query coverage. The sequences of the high-fidelity amplicons were then deposited in the GenBank with the help of the Barcode Submission Tool with detailed source information and feature annotation.

2.3 Bioinformatic and statistical analyses

DNA sequences were translated in-silico by the Translation Tool provided in the ExPASy portal to check for any pseudogenes and avoid sequencing errors. The composition of nucleotides was calculated on the CLC Workbench v7.7.1 and Mega X. The alignment of the amplicons was carried out in MUSCLE with the default specifications. Kimura two parameter (K2P) model was employed to calculate pairwise genetic distance at the level of species, genus and family (Kimura, 1980). Using this distance matrix, a Neighbor-Joining (NJ) algorithm was implemented to generate the phylogenetic tree topology in Mega X with a bootstrapping value of 1000 replications (Saitou and Nei, 1987, Felsenstein, 1985). Basic statistical analyses were done in Excel 2013 and the hypotheses testing (e.g.: t -test, F -test, etc.) was done in RStudio-1.2.5001.

3. RESULTS

3.1 General inference

A total of 748 tissue samples were collected from Bangladesh coast, among which 376 COI sequences were obtained (Table 1). Based on morphological and molecular identifications, these samples represented 185 species of 146 genera, 74 families, 20 orders and two classes (Table 1). Among the 185 species, 21 fishes were new records in Bangladesh. The length of all barcode sequences ranged from 477 to 683 bp, with an average of 652 bp and 89% of sequences were longer than 600 bp. No stop codon, insertion, or deletion was observed in any of the obtained sequences. The lack of stop codons in these sequences indicates that they

are functional mitochondrial COI sequences, together with the fact that each of the amplified sequences was about 658 bp in length. Hence, it suggests that Nuclear DNA Sequences Originating from Mitochondrial DNA Sequences (NUMTs) were not sequenced, as vertebrate NUMTS are typically less than 600 bp (Zhang & Hewitt 1996).

3.2 Elasmobranchii (Sharks and rays)

A total of 52 samples were sequenced belonging to 5 order, 8 families, 13 genera and 21 species (Table 1). Among these 21 species, two species have been found as endangered (EN), five as vulnerable (VU) and seven as near threatened (NT). One species was least concern (LC), two were not evaluated (NE) and the remaining four species were data deficient (DD) (IUCN, 2020). *Chiloscylliumburmensis* and *C. hasseltii* were reported as two new records (Ahmed et al., 2019b) in this area. The sequence analysis indicated the average nucleotide frequencies to be- A: 25.90%, T:32.70%, G:16.20% and C:25.20%. The base composition analysis for the COI sequence showed that the average percent T content was the highest and the average percent G content was the lowest; the AT content (58.60%) was higher than the GC content (41.40%). The GC contents at the first, second and third codon positions for the Elasmobranchii were 53.30%, 43.60% and 27.70%, respectively. At the first codon position, the usage of T (20.00%) was the lowest, and the usages of the other bases were 23.50%, 27.40% and 29.50% for C, A and G respectively. At the second codon position, the content of T (42.00%) was highest, and the percentage of the other bases were 29.40%, 14.50% and 14.20% for C, A and G respectively. At the third codon position, the base usage was T: 37.00%, C: 22.80%, A: 35.70% and G: 4.90%; the G content being the lowest, exhibited a clear pattern of anti-G bias.

The K2P genetic distances within each taxonomic level are summarized in Table 2. The average genetic distance within species, genus, family and order were 1.2+-0.0007%, 6.07+-0.014%, 11.08+-0.018% and 14.68+-0.04%, respectively. The NJ tree clearly distinguished all the species and the species belonging to 8 families were represented by eight distinct clades (Fig. 1).

3.3 Actinopterygii (Ray-finned fishes)

A total of 324 sequences were generated belonging to 15 orders, 66 families, 133 genera and 164 species (Table 1). The sequence analysis revealed average nucleotide frequencies to be A: 23.9%, T: 29.6%, G: 18.3% and C: 28.2% (Fig. 3). The base composition analysis for the COI sequence showed that the average T content was the highest and the average G content was the lowest; the AT content (53.50%) was higher than the GC content (46.50%). The GC contents at the first, second and third codon positions for the fish species were 54.62%, 43.66% and 38.04% respectively (Fig. 4). The pattern of %GC content at different codons was $1^{st}>2^{nd}>3^{rd}$ (p -value <0.005) The K2P genetic distances within each taxonomic level are summarized in Table 3. The average genetic distance within species, genus, family and order were 0.40 +- 0.002%, 6.36 +- 0.008%, 14.10 +- 0.01% and 24.07 +- 0.02%, respectively (Table 3).

The NJ tree of all generated sequences included 164 species is provided in Figure 2. Most of the specimens of the same species were clustered together, which reflected the prior taxonomic assignment based on morphology. No taxonomic deviation was detected at the species level, indicating that the majority of the examined species could be authenticated by the barcode approach.

3.3.1 Order Clupeiformes

This order includes many of the most important forage and food fish. A total of 45 samples were sequenced belonging to three families, 12 genera and 19 species. Among the three families Clupeidae is the most valuable family where a single species *Tenualosa ilisha* contributes over 12% of total fish production of the country (DoF, 2018). The overall mean nucleotide base frequencies observed for these sequences were T: 28.60%, C: 27.80%, A: 24.20% and G 19.40%. The AT content (52.80%) was higher than the GC content (47.20%) (Fig. 3). The GC contents at the first, second and third codon positions were 49.10%, 48.20% and 44.40% respectively (Fig. 4). The K2P distances of the COI sequence within species, genus and family were 1.81, 6.55 and 13.41, respectively (Table 3). The NJ tree clearly distinguished all the species (Supplementary Fig.1). The species belonging to family Clupeidae, Engraulidae and and Pristigasteridae were represented

by three distinct clades.

3.3.2 Order Perciformes

Perciformes is the most dominant order among the marine fishes of Bangladesh and contributes to over 50% of total exploited species. A total of 195 samples were sequenced belonging to 35 families, 79 genera and 100 species. Among the 35 families, Gobiidae was the most dominant one followed by Carangidae, Sciaenidae and Polynemidae. The overall mean nucleotide base frequencies observed for these sequences were- T: 29.90%, C: 28.10%, A: 23.80% and G: 18.20%. The AT content (53.70%) was higher than the GC content (46.30%) (Fig. 3). The GC contents at the first, second and third codon positions were 56.70%, 42.70% and 39.50% respectively. The K2P distances of the COI sequence within species, genus and family were 0.58, 6.26, and 15.62, respectively. In the NJ tree, most of the specimens belonging to the same species were clustered together bolstering the prior taxonomic assignment based on morphology (Supplementary Fig. 2).

3.3.3 Order Siluriformes

A total of 14 samples were sequenced belonging to four families, six genera and six species. The overall mean nucleotide base frequencies observed for these sequences were—T: 29.30%, C: 28.10%, A: 24.80% and G: 17.70%. The AT content (54.10%) was higher than the GC content (45.90%). The GC contents at the first, second and third codon positions were 56.70%, 42.70% and 38.10% respectively. The K2P distances of the COI sequence within species and between species were 0.14 and 26.62 respectively.

3.3.4 Order Pleuronectiformes

Fourteen samples were sequenced belonging to three families, five genera and six species. The overall mean nucleotide base frequencies observed for these sequences were—T: 30.80%, C: 26.70%, A: 24.10% and G: 18.30%. The AT content (54.90%) was higher than the GC content (45.10%). The GC contents at the first, second and third codon positions were 55.50%, 42.10% and 37.60%, respectively. The K2P distances of the COI sequences within species and family were 0.36 and 17.65 respectively.

3.3.5 Order-Beloniformes

Fifteen samples were sequenced belonging to four families, six genera and seven species. The overall mean nucleotide base frequencies observed for these sequences were—T: 32.40%, C: 25.90%, A: 24.90% and G: 16.80%. The AT content (57.30%) was higher than the GC content (42.70%). The GC contents at the first, second and third codon positions were 55.10%, 42.60% and 30.60% respectively. The K2P distances of the COI sequence within species, genus and family were 0.24, 4.77, and 9.67 respectively.

DISCUSSION

DNA barcoding has been adopted as a global bio-scanner to provide an efficient molecular technique for species-specific identification using the partial sequence of the mitochondrial COI gene. It is evident from the decade long studies (Ward et al., 2005; Hubert et al., 2008; Zhang et al., 2011; Lakra et al., 2011; Chang et al., 2017; Thu et al., 2019) that the DNA barcoding can discriminate the marine fish species from the different geographic regions, including Australia, Canada, China, India, Taiwan and Vietnam. Here, we have profiled the barcode of marine fishes collected from the coast of Bangladesh and also have demonstrated the promise of barcoding to identify these, exploiting the partial sequence of mitochondrial COI genes. Barcodes were generated for 185 species of Elashmobranchii and Actinopterygii from Bangladesh belonging to 146 genera and 74 families and 20 orders (Table 1). We observed no insertions/ deletions or codon stops after translating the nucleotide sequences, supporting the view that all of the amplified sequences denote functional mitochondrial COI sequences. Moreover, the average length of the amplified sequences was larger than 650bp, the limit typically observed for nuclear DNA sequences originating from mtDNA (NUMTs) (Gunbin et al., 2017). All of these species were differentiable based on the individual COI barcodes. Hence, this study has strongly validated the efficiency of COI barcodes for identifying fish species.

Within the Elasmobranchii, a total of 12 rays and nine sharks species including two new records (*Chiloscyllium burmensis* and*Chiloscyllium hasseltii*) were confirmed through barcoding. The overall AT and GC

content was 58.60% and 41.40%, respectively. But the mean GC content of the 11 barcoded ray species was higher than the 8 shark species (43.67% versus 38.59%). This was largely due to the GC variation in the 3rd codon position (33.59% versus 20.38%).

In this study, the COI barcode sequences for 164 teleost fish species were successfully amplified (Table 1, Fig. 2). The base composition analysis of the COI sequences revealed AT content (53.50%) to be higher than GC content (46.50%), similar to the pattern observed in Australian (Ward et al., 2005), Canadian (Steinke et al., 2009) and Cuban fish species (Lara et al., 2010). The GC contents in the first, second and third codon positions were 54.62%, 43.66% and 38.04%, respectively. At the first codon position, the usage of G (20.00%) was the lowest, and the usages of the other bases were 28.60%, 27.90% and 27.00% for C, A and T, respectively. At the second codon position, the content of T (32.00%) was highest, and the contents of the other bases were—C: 29.50%, A: 20.20% and G: 18.30. At the third codon position, the base usage was—T: 30.00%, C: 26.40%, A: 23.50% and G:16.50% (Fig. 4). There was a significantly higher overall GC content in the 164 species of bony fish compared to the 21 species of sharks and rays (46.50% versus 41.40% with a p -value ¡0.005). This difference was attributable to the GC content at the 2nd (47.80% versus 43.60%) and, especially, the 3rd codon base (42.80% versus 27.70%). The pattern of %GC content at different codons for all these fishes was invariably 1st>2nd>3rd(p -value < 0.005) and for Pleuronectiforms 1st>2nd>3rd(p -value < 0.05, n=6).

Kimura 2-parameter distance values of 6.36 $\pm$ 0.008%, 14.10 $\pm$ 0.01% and 24.07 $\pm$ 0.02% were obtained for within genus, within family and within order respectively (Table 3, Fig 5). Consistent with previously published fish barcoding data, pairwise genetic distance values were increasing at higher taxonomic levels. This increase in the genetic distance through the higher taxonomic levels supports the significant change in genetic divergence at the species boundaries (Hubert et al., 2008; Lakra et al., 2011).

In this study, the average within species K2P distance was 0.40%, compared with 6.36% for within genera. The mean interspecific distance was found to be 16-fold higher than the mean intraspecific distance. More than 13.9-fold difference was observed in the marine fishes commonly encountered in the Canadian Atlantic (Steinke et al., 2009), Indian (Lakra et al., 2011) and Australian marine fishes (Ward et al., 2005). This result corresponds to the DNA barcoding principle that interspecific divergence sufficiently outscores intraspecific divergence.

The accuracy of species identification through DNA barcoding mostly depends on both interspecific and intraspecific divergence. In our study, the average genetic distance within species was found 0.40$\pm$ 0.002%. Mean intraspecific genetic distance was calculated as <1% in previous studies; Hubert et al. (2008) found 0.30% (0–7.42%) for 194 fish species from Canadian ichthyofauna; Ward et al. (2005) 0.39% (0–14.08%) for 207 marine fish species from Australia; Thu et al. (2019) 0.34% for 458 ray finned species in Vietnam and Bingpeng et al. (2018) found 0.21% for 85 genera in Taiwan strait (Table 4).

Phylogenetic relationship of barcoded species of Elasmobranchii and Actinopterygii were shown in separate NJ tree (Fig. 1 & 2). Each species was associated with a specific DNA barcode cluster and the relationship among these species was clearly revealed. Closer species in terms of genetic divergence, were clustered at the same nodes and the distance between the terminal branches of the NJ tree widened as they got more distinct.

Our study suggests that DNA barcoding has been successful in identifying and discriminating the vast majority of marine ichthyofauna. The DNA barcoding method has been proven to be an effective tool for species identification, particularly with specimens that are damaged, incomplete, or consisting of several morphologically distinct stages (Pečnikar & Buzan, 2014; Bingpeng et al., 2018). Nevertheless, DNA barcoding also has its limitations. In some cases, related species may present identical sequences making DNA barcodes useless for species discrimination. Therefore, DNA barcoding can serve as a complementary tool for species identification, though it cannot replace the traditional morpho-taxonomy. Through this study, a reliable DNA barcode reference library for the marine fish in the Bay of Bengal, Bangladesh has been established, which could be used to assign fish species by screening sequences against it in the future. We hope this

would appreciably contribute to achieving better monitoring, conservation, and management of fisheries in this over exploited region.

Acknowledgements

We acknowledge the financial support from the Ministry of Education, Government of the People's Republic of Bangladesh as a grant for advanced research in Education (Grants No. MRS2017448). We are also thankful to Dr. Mohammad Abdul Baki, Nusrat Jahan Sanzida, JBM Aysha Akter, Ayesha Akhter Zhilik, Nishat Jahan Chowdhury, Sumaiya Salam, Tamim Afrin, Noor Aida Arfin, Mysha Mahjabin, Sajid Ahmed and Afroza Kaonine Haque for their cooperation during this research.

REFERENCES

Ahmed, M.S., Chowdhury, M.M.K., & Nahar, L. (2019a). Molecular characterization of small indigenous fish species (SIS) of Bangladesh through DNA barcodes. *Gene*, **684**, 53-57.

Ahmed, M.S., Chowdhury, N.Z., Datta, S.K., & Zhilik, A.A. (2019b). New Geographical Record of the Burmese Bamboo Shark, *Chiloscyllium burmensis* (Orectolobiformes: Hemiscylliidae), from Bangladesh Waters. *Thalassas: An International Journal of Marine Sciences,* **35** (2), 347-350.

Bingpeng, X., Heshan, L., Zhilan, Z., Chunguang, W., Yanguo, W., & Jianjun, W. (2018). DNA barcoding for identification of fish species in the Taiwan Strait. *PloS one* , **13** (6): e0198109. https://doi.org/10.1371/journal.pone.0198109

Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual review of marine science* , *3* , 471-508.

Carpenter, K.E., & De Angelis, N. (Eds.). (2002a). The living marine resources of the Western Central Atlantic. Rome: Food and agriculture organization of the United Nations. **2** , 602-1373

Carpenter, K. E., & Niem, V. H. (2002). FAO species identification guide for fishery purposes. The living marine resources of the western central Atlantic. Volume 3. Bony fishes. Part 2.(Opistognathidae to Molidae), sea turtles and marine mammals. *FAO species identification guide for fishery purposes. The living marine resources of the Western Central Atlantic* , *3* .

Chang, C.H., Shao, K.T., Lin, H.Y., Chiu, Y.C., Lee, M.Y., Liu, S.H., & Lin, P.L. (2017). DNA barcodes of the native ray finned fishes in Taiwan. *Molecular ecology resources* , **17** (4), 796-805.

Chin, T.C., Adibah, A.B., Hariz, Z.D., & Azizah, M.S. (2016). Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving transparency in food market. *Food Control* , **64** , 247-256.

Costa, F.O., Landi, M., Martins, R., Costa, M.H., Costa, M.E., Carneiro, M., & Carvalho, G.R. (2012). A ranking system for reference libraries of DNA barcodes: application to marine fish species from Portugal. *PLoS One* , **7** (4): e35858. doi:10.1371/journal.pone.0035858

DoF. (2018). Yearbook of Fisheries Statistics of Bangladesh, 2017-18. Fisheries Resources Survey System (FRSS), Department of Fisheries. Bangladesh: Ministry of Fisheries and Livestock, Vol **35** , 129pp

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* , **39** (4), 783-791.

Filonzi, L., Chiesa, S., Vaghi, M., & Marzano, F.N. (2010). Molecular barcoding reveals mislabelling of commercial fish products in Italy. *Food Research International* , **43** (5), 1383-1388.

Gunbin, K., Peshkin, L., Popadin, K., Annis, S., Ackermann, R.R., & Khrapko, K. (2017). Data on the time of integration of the human mitochondrial pseudogenes (NUMTs) into the nuclear genome. *Data in brief* , **13** , 536-544.

Hebert, P.D., Cywinska, A., & Ball, S.L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* , **270** (1512), 313–21.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., & Zhang, J. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *PLoS one,* **3** (6): e2490.

doi:10.1371/journal.pone.0002490

Hussain, M.M. (1971). The commercial fishes of the Bay of Bengal. Survey for the development of fisheries in East Pakistan. Chittagong, UNDP Project PAK 22, Project Publication No. 1, 60 pp

IUCN Bangladesh. (2000). Red Book of Threatened Fishes of Bangladesh. IUCN- The World Conservation Union. Xii+116PP.

IUCN. (2020). The IUCN Red List of Threatened Species. Version 2020-1. <https://www.iucnredlist.org>

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.

Knebelsberger, T., Landi, M., Neumann, H., Kloppmann, M., Sell, A.F., Campbell, P.D., & Costa, F.O. (2014). A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. *Molecular ecology resources* , **14** (5), 1060-1071.

Lakra, W.S., Verma, M.S., Goswami, M., Lal, K.K., Mohindra, V., Punia, P., & Hebert, P. (2011). DNA barcoding Indian marine fishes. *Molecular Ecology Resources* , **11** (1), 60-71.

Lara, A., Ponce de León, J.L., Rodriguez, R., Casane, D., Cote, G., Bernatchez, L., & García Machado, E.R.I.K.L. (2010). DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. *Molecular ecology resources* , **10** (3), 421-430.

Last, P., Naylor, G., Séret, B., White, W., de Carvalho, M., & Stehmann, M. (Eds.). (2016). Rays of the World. CSIRO publishing.

Nakabo, T. (2002). Fishes of Japan with pictorial keys to the species, English edition I. Tokai University Press, Japan.

Pečnikar, Ž.F., & Buzan, E.V. (2014). 20 years since the introduction of DNA barcoding: from theory to application. *Journal of Applied Genetics* , **55** (1), 43-52.

Rahman, A.K.A., Khan, M.G., Chowdhury, Z.A., & Hussain, M.M. (1995). Economically important marine fishes and shell fishes of Bangladesh. Department of Fisheries, Dhaka 53pp.

Rahman, A.K.A., Kabir, S.M.H., Ahmad, M., Ahmed, A.T.A., Ahmed, Z.U., Begum, Z.N.T., Hassan, M.A., & Khondker, M.(eds). (2009). Encyclopedia of Flora and Fauna. Asiatic Society of Bangladesh, Dhaka **24** , 485pp

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution,* **4** (4), 406-425.

Shafi, M., & Quddus, M.M.A. (1982). Fisheries resources of Bay of Bengal (Bonggoposagarer Matshaya Sampad, (In Bengali). Bangla Academy, Dhaka.

Shehata, H.R., Bourque, D., Steinke, D., Chen, S., & Hanner, R. (2019). Survey of mislabelling across finfish supply chain reveals mislabelling both outside and within Canada. *Food research international* , **121** , 723-729.

Siddiqui, K.U., Islam, M.A., Kabir, S.M.H., Ahmad, M., Ahmed, A.T.A., Rahman, A.K.A., Haque, E.U., Ahmed, Z.U., Begum, Z.N.T., Hassan, M.A., Khondker, M., & Rahman, M.M. (eds.). (2007). Encyclopedia of Flora and Fauna of Bangladesh, Fresherwater Fishes. Asiatic Society of Bangladesh, Dhaka **23** , 300pp

Steinke, D., Zemlak, T.S., Boutillier, J.A., & Hebert, P.D. (2009). DNA barcoding of Pacific Canada's fishes. *Marine Biology* , **156** (12), 2641-2647.

Talwar, P.K., & Jhingran, A.G. (1991). Inland fisheries of India and adjacent countries. Vol**. I & II** , 1-1158.

Thu, P.T., Huang, W.C., Chou, T.K., Van Quan, N., Van Chien, P., Li, F., & Liao, T.Y. (2019). DNA barcoding of coastal ray-finned fishes in Vietnam. *PloSOne* , **14** (9). e0222631. https://doi.org/10.1371/journal.

pone.0222631

Wang, L., Wu, Z., Liu, M. et al. (2018). DNA barcoding of marine fish species from Rongcheng Bay, China. *Peer J* . 6 , e5013

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., & Hebert, P.D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* **360** , 1847-1857.

Zhang, J. (2011). Species identification of marine fishes in China with DNA barcoding. *Evidence-Based Complementary and Alternative Medicine* , **11:** 1-10. Article ID 978253, doi:10.1155/2011/978253

Zhang, D.X., & Hewitt, G.M. (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in ecology & evolution,* **11** (6), 247-251.

Zhang, J., & Hanner, R. (2012). Molecular approach to the identification of fish in the South China Sea. *PLoS One,* **7** (2). e30621. doi:10.1371/ journal.pone.0030621

Ethics statement

All fish species were caught in the offshore area (not national parks, other protected areas, or private areas, etc.), so no specific permissions were required for these locations/activities. However, specimen collection was conducted under a permit to the University of Dhaka, and approval of the Ethical Review Committee, Faculty of Biological Sciences, University of Dhaka.

Conflict of interest

The authors declare no conflict of interest.

Data availability

All the voucher specimens with their respective Voucher ID are keeping at the Museum of Department of Zoology, University of Dhaka, Bangladesh, and has public access with permision. The DNA barcoding data can be retrieving from the NCBI GenBank as have open public access.

Contribution

Ahmed MS initiated the project, acquired funding, managed project administration, specimen curation and internal reports, performed the taxonomic analyses, and prepared the whole manuscript. Datta SK and Saha T performed field collection, generated and analyzed molecular sequences, curated tissue samples, and participated in the writing of the final manuscript. Hossain Z contributed laboratory analyses, participated in generation of sequences and made contribution to the manuscript. All authors reviewed the manuscript.

Hosted file

`Tables 1-4.docx` available at https://authorea.com/users/332255/articles/458700-molecular-characterization-of-marine-and-coastal-fishes-of-bangladesh-through-dna-barcodes