The ITS2 DNA sequence analysis in six species of barbin fishes with phylogenetic insights

Sandip Choudhury1*, Briyanka Kashyap2, Karabi Dutta2
1Department of Zoology, Nowgong Girls’ College, Nagaon, India.
2Department of Zoology, Gauhati University, Guwahati, India.

ABSTRACT

The internal transcribed spacer 2 (ITS2) nuclear DNA was sequenced for six species of barbin fishes from Assam in the North Eastern region of India. The variations in the sequences were investigated to estimate nucleotide composition, nucleotide diversity, transition-transversion bias, genetic distance, and phylogenetic relationship. From the sequence analysis, it was found that the average G + C content (64.8%) was more than A + T content (35.2%). The nucleotide diversity (Pi) was found to be 0.04737. The number of transition substitutions was more than transversion substitutions and the transition–transversion bias was 1.16. Overall mean genetic distance was found to be 0.050 with a range from 0.005716 (between Puntius sophore and Puntius chola) to 0.084536 (between Pethia gelius and Systomus sarana). The phylograms constructed by neighbor-joining and maximum likelihood methods resulted in a similar topology where the monophyly of the Pethia group was recovered which consists of P. gelius, Pethia ticto, and Pethia conchonius. The two Puntius species (P. sophore and P. chola) were not clustered together and S. sarana remained a distinct taxon. The results of the present study partially validated the utility of the ITS2 DNA sequence in genetic diversity and phylogenetic studies in the barbin fishes.

1. INTRODUCTION

The eukaryotic ribosomal DNA genes are arranged in tandemly repeated clusters with each cluster containing the genes for the 18S, 5.8S, and 28S ribosomal RNA (rRNA). The genes are separated by several spacers, the intergenic spacer or non-transcribed spacer, the external transcribed spacer, and the internal transcribed spacer (ITS). Availability of universal Polymerase Chain Reaction (PCR) primers designed from highly conserved regions flanking the ITS, its relatively small size and high copy number enable easy amplification of ITS regions and this makes the ITS regions an interesting subject for evolutionary/phylogenetic investigations [1]. The ITS2 which is flanked by 5.8S rRNA gene and 28S rRNA gene is a phylogenetic marker that has been of broad use in generic and infrageneric level classifications as its sequence evolves comparably fast [2]. Indeed after its first application in 1991 [3], it has been on a rapidly growing number of publications and most extensively it is used in fungal phylogenetics. In spite of its high use in phylogenetic analyses, the number of publications using this marker in phylogenetic studies in fishes is very less and thus provides great opportunity to test the potential of this genetic marker in fish phylogenetics.

The fishes of genus Puntius (Cypriniformes: Cyprinidae) which are commonly known as barbs are oriental and widely distributed in South Asia [4]. Because of their wide distribution in the world, these species are of great interest to fish biologists, ecologists, and evolutionary biologists. The status of Puntius is unclear, the demarcation and nomenclatural validity of the genus have remained unsettled [5–7]. Scantiness in the knowledge of interf generic and infrageneric relationships has made the status of this group obscure [8]. Kottelat described Puntius as a “catch-all” genus in which a large number of unrelated small barbs have been placed [9]. Pethiyagoda et al. [10] recognized five well-supported clades as distinct genera until then it was included in Puntius genus and their work was re-evaluated and supported [11] with
2. MATERIALS AND METHODS

2.1. Taxon Sampling

A total of 60 samples, 10 for each of the 6 barbin fishes under family Cyprinidae viz. *Puntius chola*, *Puntius sophore*, *Pethia gelius*, *Pethia conchonius*, *Pethia ticto*, and *Systomus sarana* were collected from Kamrup Metropolitan and Golaghat districts of Assam by random sampling. The GenBank accession numbers of the genetic markers used for the specified samples are given (Table 1). The species were identified by the use of taxonomic keys prepared by [4,12–14].

2.2. Extraction of Genomic DNA

For genomic DNA extraction, the phenol/chloroform extraction procedure [15] was used with some modifications. Quantity and quality assessment of extracted genomic DNA was done by spectrophotometry by reading the absorbance at 260 nm of wavelength [16,17] and ethidium—bromide-stained agarose gel electrophoresis technique [16].

2.3. Amplification and Sequencing

For amplification of the nuclear ITS2 DNA marker, the following universal primers were used—Forward primer ITS2 F1 (5′-gcaggacactattgatcagc-3′) and Reverse primer ITS2 F2 (5′-gtcttcatcttccttctctcgctg-3′). Amplification was done in 30 µl volume in each 0.2 ml labeled PCR tubes containing 30 pmol of each forward and reverse primers, 2.5 µl of 2.5 mM deoxyribonucleotide triphosphate (dNTPs), 3 µl of taq buffer (10×), 1.5 µl of Mgcl₂ (50 mM), 0.5 µl of taq polymerase (5 µ/µl), 11.5 µl nuclease-free double distilled water, and 5 µl template DNA (10 ng/µl). The thermal profile consisted of an initial denaturation step of 5 minutes at 95°C followed by 35 cycles at 94°C for 30 seconds, 65°C for 30 seconds, 72°C for 1 minute and a final elongated extension at 72°C for 10 minutes. The PCR products were kept in hold at 4°C. After this 1/10 volume (3 µl) of 6× gel loading dye was added to each labeled PCR tube and given brief spin. The PCR products were visualized on 1.2% Low melting (LM) agarose gel containing ethidium bromide and the elution was done using HiPura Agarose gel DNA purification Spin kit (Himedia) following the manufacturer’s protocol. The purified products of PCR were used for sequencing. The same PCR primers were used for sequencing. The PCR products were labeled using the BigDye Terminator v.3.1 cycle sequencing kit (Applied Biosystems). Sequencing was run bidirectionally and visualized on an ABI 3730 XL automated DNA Sequencer following the manufacturer’s instructions.

2.4. Sequence Analysis

The raw sequences were edited using BioEdit software version 5.0.9 [18]. The bad reads (noises) were removed from the sequences (as examined from the electropherogram) and then searched for homology alignment/similarity using Basic Local Alignment Search Tool (BLAST) program of National Center for Biotechnology Information (NCBI) [19]. The edited sequences were submitted to GenBank (NCBI) through Bankit submission tool (NCBI) following the instructors’ protocol. The sequences were aligned using the Clustal X 2.1 sequence alignment software [20]. The extent of sequence differences between species was calculated by averaging pair-wise Kimura-2-parameter [21] comparisons of the sequence difference across all the individuals using the Dnadist program of PHYLIP 3.69 package [22]. The number of polymorphic sites, total number of sites, conserved sites, parsimony informative sites, total number of mutations, and nucleotide diversity were estimated using the DnaSp 5.10.01 software [23]. The nucleotide compositions/frequencies, G + C and A + C content, rate of transitions and transversions, transition: transversion bias were calculated by the software MEGA6 [24]. In all analyses, gaps were considered as missing data.

2.5. Phylogenetic Analysis

Phylogenetic relationships were estimated using (1) Distance-based Neighbour Joining (NJ) method and (2) Character-based Maximum Likelihood (ML) method using the PHYLIP package version 3.69 [22]. The distance matrix for the sequences was constructed by the Kimura 2 parameter method [21]. The transition: transversion ratio was taken to be 2.0 and the sites assigned unweighted. Empirical base frequencies were also used. The distance matrix was then used to reconstruct the phylogenetic tree using NJ method in the Neighbour program of PHYLIP 3.69. For the ML method, the Dnaml program of PHYLIP 3.69 was used. The transition: transversion ratio was taken to be 2.0 and the sites assigned unweighted. Empirical base frequencies and slow but accurate analysis were employed. The statistical significance of the branching order was assessed by bootstrap resampling.

Table 1: GenBank accession numbers of the ITS2 DNA sequence of the species.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Taxon</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Puntius chola</em></td>
<td>KJ509175</td>
</tr>
<tr>
<td>2</td>
<td><em>Pethia conchonius</em></td>
<td>KJ509176</td>
</tr>
<tr>
<td>3</td>
<td><em>Pethia gelius</em></td>
<td>KJ509177</td>
</tr>
<tr>
<td>4</td>
<td><em>Systomus sarana</em></td>
<td>KJ509178</td>
</tr>
<tr>
<td>5</td>
<td><em>Puntius sophore</em></td>
<td>KJ509179</td>
</tr>
<tr>
<td>6</td>
<td><em>Pethia ticto</em></td>
<td>KJ509180</td>
</tr>
</tbody>
</table>
technique [25] considering 1,000 pseudo replicates using the Seqboot program of PHYLIP 3.69. The random dataset was analyzed in exactly the same way the original dataset was analyzed and the results from the random datasets were summarized by constructing an extended majority rule consensus tree using the Consense program of PHYLIP 3.69. The consense outtree files were then transferred to the TreeView 1.6.6 software [26] to draw the phylogenetic trees/phylograms with internal edge labels which represent the bootstrap values.

3. RESULTS

The ITS2 DNA sequences which also contain small partial sequences of 5.8S and 28S rRNA genes obtained from the six fish samples ranged in size from 412 bp for P. sophore to 375 bp for P. gelius. The BLAST search results were found to be good as revealed by the maximum scores and percentage similarity but for a limited number of species. It was found that there were very few available ITS2 sequences at the GenBank database for species under Cyprinidae family which possess significant similarity to the six new sequences. The frequency/composition of nucleotides, $G + C$ and $A + T$ content has been listed (Table 2). The multiple sequence alignment used for analyses of the six sequences resulted in 371 nucleotide sites. After excluding the indels/gaps/missing data, the net number of sites equals 349 nucleotides. Sites with alignment gaps/indels/missing data = 22. Invariable (monomorphic)/conserved sites = 308, variable (polymorphic) sites (S) = 41, total number of mutations (Eta) = 42, singleton variable sites = 29, parsimony informative sites = 12, singleton variable sites (two variants) = 29. Parsimony informative sites for two variants = 11 and for three variants = 1. Average number of nucleotide differences ($K$) = 16.533 and nucleotide diversity (Pi) = 0.04737.

ML estimate of substitution matrix (Table 3) revealed that transition substitutions were more than transversion substitutions. Overall transition and transversion were found to be 55.92 and 44.08, respectively. The estimated transition/transversion bias ($r$) was found to be 1.16. Pairwise nucleotide differences and genetic distance values (K2P) based on ITS2 DNA sequences are listed (Table 4). The overall mean distance was found to be 0.050 with a range from lowest 0.005716 (between P. sophore and P. chola) to highest 0.084536 (between P. gelius and S. sarana).

An extended majority rule consensus phylogram/tree was obtained (Fig. 1). No outgroup was used for the tree construction because of unavailability of ITS2 sequences to be appropriate as to be considered as outgroup species. The bootstrap values were indicated at branch nodes. The bootstrap supports were found to be strong for maximum branch nodes. The monophyly of genus *Pethia* was recovered with bootstrap support of 91.8%. The genus *Puntius* represented by *P. chola* and *P. sophore* was found to be non-monophyletic. *P. sophore* was found to be a close relative to the *Pethia* group. *Puntius chola* and *S. sarana* (the only representative of genus *Systomus*) did not show any clustering. The extended majority rule consensus NJ phylogram/

<table>
<thead>
<tr>
<th>Species</th>
<th>T%</th>
<th>C%</th>
<th>A%</th>
<th>G%</th>
<th>Total</th>
<th>(G + C)%</th>
<th>(A + T)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systomus sarana</td>
<td>18.7</td>
<td>35.6</td>
<td>16.0</td>
<td>29.7</td>
<td>357</td>
<td>65.3</td>
<td>34.7</td>
</tr>
<tr>
<td>Puntius chola</td>
<td>19.0</td>
<td>34.7</td>
<td>16.0</td>
<td>30.3</td>
<td>357</td>
<td>65.0</td>
<td>35.0</td>
</tr>
<tr>
<td>Puntius sophore</td>
<td>19.0</td>
<td>34.7</td>
<td>15.7</td>
<td>30.5</td>
<td>357</td>
<td>65.3</td>
<td>34.7</td>
</tr>
<tr>
<td>Pethia conchonius</td>
<td>18.0</td>
<td>34.4</td>
<td>17.5</td>
<td>30.1</td>
<td>355</td>
<td>64.5</td>
<td>35.5</td>
</tr>
<tr>
<td>Pethia ticto</td>
<td>17.5</td>
<td>35.6</td>
<td>17.5</td>
<td>29.4</td>
<td>354</td>
<td>65.0</td>
<td>35.0</td>
</tr>
<tr>
<td>Pethia gelius</td>
<td>19.5</td>
<td>34.3</td>
<td>16.8</td>
<td>29.4</td>
<td>364</td>
<td>63.7</td>
<td>36.3</td>
</tr>
<tr>
<td>Average</td>
<td>18.6</td>
<td>34.9</td>
<td>16.6</td>
<td>29.9</td>
<td>357</td>
<td>64.8</td>
<td>35.2</td>
</tr>
</tbody>
</table>

### Table 3: ML estimates of substitution matrix for ITS2 DNA marker sequences. Each entry is the probability of substitution ($r$) from one base (row) to another base (column).

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>4.00</td>
<td>7.71</td>
<td>15.61</td>
</tr>
<tr>
<td>T</td>
<td>3.74</td>
<td>-</td>
<td>20.72</td>
<td>6.59</td>
</tr>
<tr>
<td>C</td>
<td>3.74</td>
<td>10.74</td>
<td>-</td>
<td>6.59</td>
</tr>
<tr>
<td>G</td>
<td>8.85</td>
<td>4.00</td>
<td>7.71</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 4: Pairwise genetic distance matrix (K2P) based on ITS2 DNA sequences of the analysed taxa.

<table>
<thead>
<tr>
<th>Species</th>
<th>Systomus sarana</th>
<th>Puntius chola</th>
<th>Puntius sophore</th>
<th>Pethia conchonius</th>
<th>Pethia ticto</th>
<th>Pethia gelius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systomus sarana</td>
<td>0.00000</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Puntius chola</td>
<td>0.014460</td>
<td>0.00000</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Puntius sophore</td>
<td>0.014460</td>
<td>0.005716</td>
<td>0.00000</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Pethia conchonius</td>
<td>0.051291</td>
<td>0.041715</td>
<td>0.035639</td>
<td>0.00000</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Pethia ticto</td>
<td>0.080035</td>
<td>0.069898</td>
<td>0.063608</td>
<td>0.051348</td>
<td>0.00000</td>
<td>*</td>
</tr>
<tr>
<td>Pethia gelius</td>
<td>0.084536</td>
<td>0.080956</td>
<td>0.079389</td>
<td>0.078316</td>
<td>0.071176</td>
<td>0.00000</td>
</tr>
</tbody>
</table>
The nucleotide diversity value of 0.04737 also indicates the closeness of the species analyzed.

4. DISCUSSION

From the sequence analyses of the nuclear ITS2 DNA sequences, it was found that the mean \((G+C)\) content (64.8%) of all the species analyzed was more than the mean \((A+T)\) content (35.2%) and much higher \((G+C)\) content for ITS2 DNA sequences than \((A+T)\) content was reported in Brown Trout [27] and Pearl Oyster [28] also. The overall observed transition/transversion bias that was found to be 1.16 indicated the occurrence of transitions more than transversions. Marinho et al. [29] also found that for ITS2 DNA sequences transitions were more than transversions. Overall transition (55.92%) which was more than overall transversions (44.08%) is characteristic of the species that have diverged recently from a common ancestor [30]. Transition events more than transversion events were also reported in Salmo trutta [27] and five species of fishes of the subfamily Schizothoracinae [31]. The average value of genetic distance (0.050) was found to be low inferring that the species which were analyzed are closely related and are descendants of a near common ancestor. The genetic distance values of the ITS2 DNA sequences of the species analyzed indicate the ability of this marker in describing the interrelationships of the species of fishes studied up to some extent. The nucleotide diversity value of 0.04737 also indicates the closeness of the species analyzed.

In both the ML and NJ methods for the ITS2 DNA sequence analyses, the bootstrap values were satisfactory. This genetic marker was able to distinguish the Pethia group as a monophyletic one, which was also found in the previous studies [32,10,33–35,11]. Among the three Pethia species in the present study, P. ticto and P. gelius were found to be more closely related than they are to P. conchonius, and P. sophore was found to be more closely related to the Pethia group than P. chol. Based on the Recombination activating gene 2 (RAG2) nuclear gene [36], 16S rRNA gene and cytb gene [10,11], genus Puntius was recovered as non-monophyletic and in the present study also, the genus Puntius was recovered as non-monophyletic as P. chol. and P. sophore were not clustered together. So, the status of Puntius genus from the above and the present study is found to be obscure. The results of the nuclear ITS2 DNA sequence analyses somewhat resembled previous studies but to validate its potential and to test the robustness of this genetic marker for phylogenetic studies in fish, further studies using ITS2 secondary structure, larger sampling size are required. There are evidences that usually support increasing taxon sampling even at the expense of great quantities of character data for improved accuracy of topologies [37].

**Figure 1:** Extended majority—rule consensus ML phylogram (ITS2 DNA sequence datasets). Numbers at branch nodes = bootstrap values (out of 1,000 pseudo replicates).

**Figure 2:** Extended majority—rule consensus NJ phylogram (ITS2 DNA sequence datasets). Numbers at branch nodes = bootstrap values (out of 1,000 pseudo replicates).
5. CONCLUSION
In conclusion, ITS2 DNA sequence variation among the taxa present in the study was partially able to establish their phylogenetic relationship as the monophyly of genus Pethia was established but the status of genus Puntius was found to be obscure. Our understanding of the study of phylogenetic relationships of fish taxa using ITS2 DNA marker is growing but much more studies are required to validate its potential. The present study paves a way for further studies in evaluating the potential of this DNA marker in fish phylogenetics.

7. ACKNOWLEDGEMENTS
The authors are thankful to Dr. Pranjan Barman, Dept. of Biotechnology, Gauhati University and Dr. Lingaraj Sahoo, Dept. of Biotechnology, Indian Institute of Technology - Guwahati for providing laboratory facilities and guidance.

6. AUTHOR CONTRIBUTIONS
All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be an author as per the international committee of medical journal editors (ICMJE) requirements/guidelines.

7. FUNDING
There is no funding to report.

8. CONFLICTS OF INTEREST
The authors report no financial or any other conflicts of interest in this work.

9. ETHICAL APPROVALS
This study does not involve experiments on animals or human subjects.

10. PUBLISHER’S NOTE
This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES
11. Ren Q, Mayden RL. Comparative phylogenetics of the Asian genus “Puntius” and relatives (Cypriniformes; Cyprinidae): explorations into the potential impacts of taxon and character sampling. I3 Biodiversity 2019;3:1–19.
29. Marinho MAT, Junqueira ACM, Azeredo-Espin AML. Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure


How to cite this article: