Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Characterization of the complete mitochondrial genome of *Barilius malabaricus* and its phylogenetic implications

Vandana R. Prabhu[a], Hijam Surachandra Singha[a], Rahul G. Kumar[b], Achamveetil Gopalakrishnan[c], Muniyandi Nagarajan[a,*]

[a] Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Kasaragod 671316, Kerala, India
[b] Peninsular and Marine Fish Genetic Resources Centre, ICAR-NBFGR, CMFRI Campus, Kochi 682018, Kerala, India
[c] Central Marine Fisheries Research Institute, Ernakulam North PO, Kochi 682 018, Kerala, India

ARTICLE INFO

Keywords:
Barilius
Opsarius
Cyprinidae
Fish phylogeny
Mitochondrial genome
COI

ABSTRACT

Bariliine fishes are important components of the ornamental as well as subsistence fishery sectors in India. Many of the species in the genus *Barilius* are threatened by habitat loss and therefore need to be met with conservation initiatives. Effective conservation measures, however, require clarification of species identities and resolution of the validity of many species currently treated as synonymous or sub-species. The complete mitochondrial genome data provide better insight into phylogenetic information than the short fragment or single gene based analysis. Thus, we have sequenced the complete mitochondrial genome of *Barilius malabaricus,* one of the important fish species in the fresh water ornamental sector, for better understanding its phylogenetic status. The 16,519 bp mitochondrial genome consists of 37 genes which classifies as 22 tRNA, 13 protein coding and 2 ribosomal RNA genes and a control region. Overall, the mitochondrial genome bears the typical gene order and composition as in other fishes. Further, the COI, cytochrome *b* and 16S rRNA gene sequences revealed that, *B. malabaricus* is genetically closer to *B. canarensis* and *B. bakeri* compared to other *Barilius* species. Also, the *Barilius* species of west flowing rivers in Western Ghats were consistently recovered as a clade distinct from other species. We therefore suggest to retain the genus name *Barilius* for the species from the Western Ghats until a comprehensive analysis based on both morphological and molecular markers reveals the relationship between species now variously placed in the genera *Barilius* and *Opsarius* in greater detail.

1. Introduction

Barilius malabaricus (Jerdon, 1849) is a- commercially important fish belonging to the bariliine group of Cyprinidae family. Bariliine Cyprinids in general are distributed across Africa, and South and Southeast Asia, however *B. malabaricus* is restricted to the southern parts of Karnataka and the northern Kerala of India. It grows to about 9.1 cm in length and is characterised by a single or double row of bluish green lateral spots, with bright orange margined dorsal and anal fins [1]. Initially, Jerdon [2] identified *B. malabaricus* as a distinct species bearing significant morphological difference from *B. canarensis* and *B. gatensis*. However, Day [3,4], treated both *B. malabaricus* and *B. canarensis* as synonyms of *B. bakeri*. Recently, Knight et al. [1], on the basis of morphological features such as gills rakers, lateral line scales and dorsal fin colouration, reconsidered the earlier classification by Jerdon [2], and treated *B. malabaricus*, *B. canarensis* and *B. bakeri* as distinct species. Additionally, Knight et al. [1] introduced a new species *B. ardens* which differs morphologically from *B. malabaricus* and *B. canarensis*. The accurate identification of *Barilius* species is often a difficult task and a subject of debate over decades due to their morphological similarity and overlapping meristic features. Therefore, the usage of genetic markers for the identification of *Barilius* species would probably help to resolve the taxonomic ambiguity of the bariliine fishes.

Mitochondrial DNA in comparison to the nuclear DNA shows relatively higher mutation rate and therefore it is considered as one of the most efficient and reliable molecular marker for phylogenetic studies on fishes [5–7]. Previously, several studies have been conducted on Cyprinidae family using the mitochondrial gene fragments as well as whole mitochondrial genome which complied to the conventional classification system, while in certain species it defied the previous taxonomic classification [5,8,9]. Hitherto, on the genus *Barilius*, there have been only a few molecular phylogenetic studies carried out using 16S rRNA, cytochrome *b* and COI genes [6,10–13]. To date the availability of mitochondrial genome sequence is limited to *B. bendelisis*. In

* Corresponding author.
E-mail address: nagarajan@cukerala.ac.in (M. Nagarajan).

https://doi.org/10.1016/j.ygeno.2019.12.009
Received 27 September 2019; Received in revised form 11 November 2019; Accepted 12 December 2019

Table 1
Nucleotide composition and AT-GC skewness of the *B. malabaricus* mitogenome.

B. malabaricus	Size(bp)	A%	T%	G%	C%	AT.%	GC%	GC skew	AT skew
Whole mitogenome	16,519	30.5	26.4	16.6	26.5	56.9	43.1	−0.232	0.072
PCGs	11,280	28.2	28.9	15.9	27.0	57.1	42.9	−0.257	−0.011
tRNAs	1555	28.4	26.3	23.6	21.7	54.7	45.3	0.041	0.040
rRNAs	2623	34.8	20.1	21.4	23.7	54.9	45.1	−0.049	0.266
Control region	899	32.0	32.2	14.8	21.0	64.2	35.8	0.170	0.558

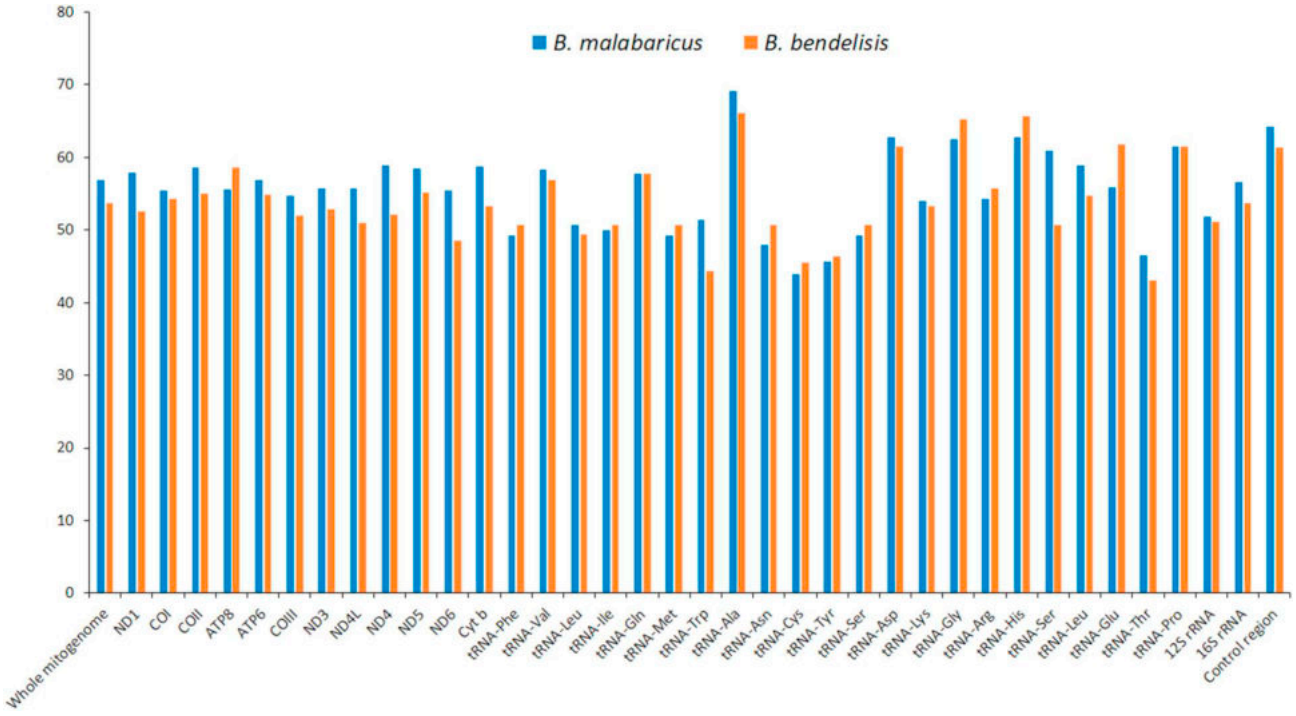

Fig. 1. The AT content of various constituent elements and complete mitochondrial genome of *B. malabaricus* and *B. bendelisis.*

the present study, for the first time, we have sequenced the whole mitochondrial genome of *B. malabaricus* using Illumina sequencing technology. The study provides the characteristic features of the *B. malabaricus* mitochondrial genome and its phylogenetic relationships with other *Barilius* species and subfamily Danioninae for further research.

2. Materials and methods

2.1. Sample collection

The *B. malabaricus* sample used in the study was collected from the Payaswini river, Sullia, Karnataka. This study does not include live animals and the samples used were dead and obtained from the commercial catches. The fin clips were collected from the dead fishes and stored in absolute ethanol at 4 °C for DNA isolation.

2.2. Mitochondrial genome sequencing

Genomic DNA was extracted from fin clips using DNeasy blood and tissue kit (Qiagen, Valencia, CA) as per the manufacturer's protocol. Subsequently, the entire mitochondrial genome was amplified by long-range PCR using two overlapping sets of primers BM1 (5′-GCAAACGACGCACTTATTGA-3′, 5′-GAAGGCTGGCTCTTCAAATG-′3) and BM2 (5′-TGATTCTTCTCCACCAACCA-3′, 5′-GATGAACGGGTGTTCTACCG-3′). The PCR was carried out in 50 µl reaction mixture containing 23 µl of nuclease free water, 10 µl of 5× PrimeSTAR GXL Buffer (Takara, Japan), 4 µl of dNTPs, 4 µl of each primer, 1 µl of PrimeSTAR GXL DNA Polymerase (Takara, Japan), and 4 µl of DNA template. PCR conditions applied for the amplification process were as follows: initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 68 °C for 1 min and extension at 68 °C for 10 min. The final extension was subjected to 68 °C for 10 min. The amplified PCR products were processed for library preparation using NEB Ultra DNA Library Prep Kit and the sequencing was performed on Illumina HiSeq 2500.

2.3. Data analysis

The adapter sequences were removed from the raw reads using CUTADAPT [14]. Further the low quality bases (Phred score, Q < 20) were trimmed using Sickle [15] and the duplicate reads were removed using FastUniq [16]. The pre-processed paired end reads were assembled using SeqMan Pro (DNASTAR, Inc., USA). The whole

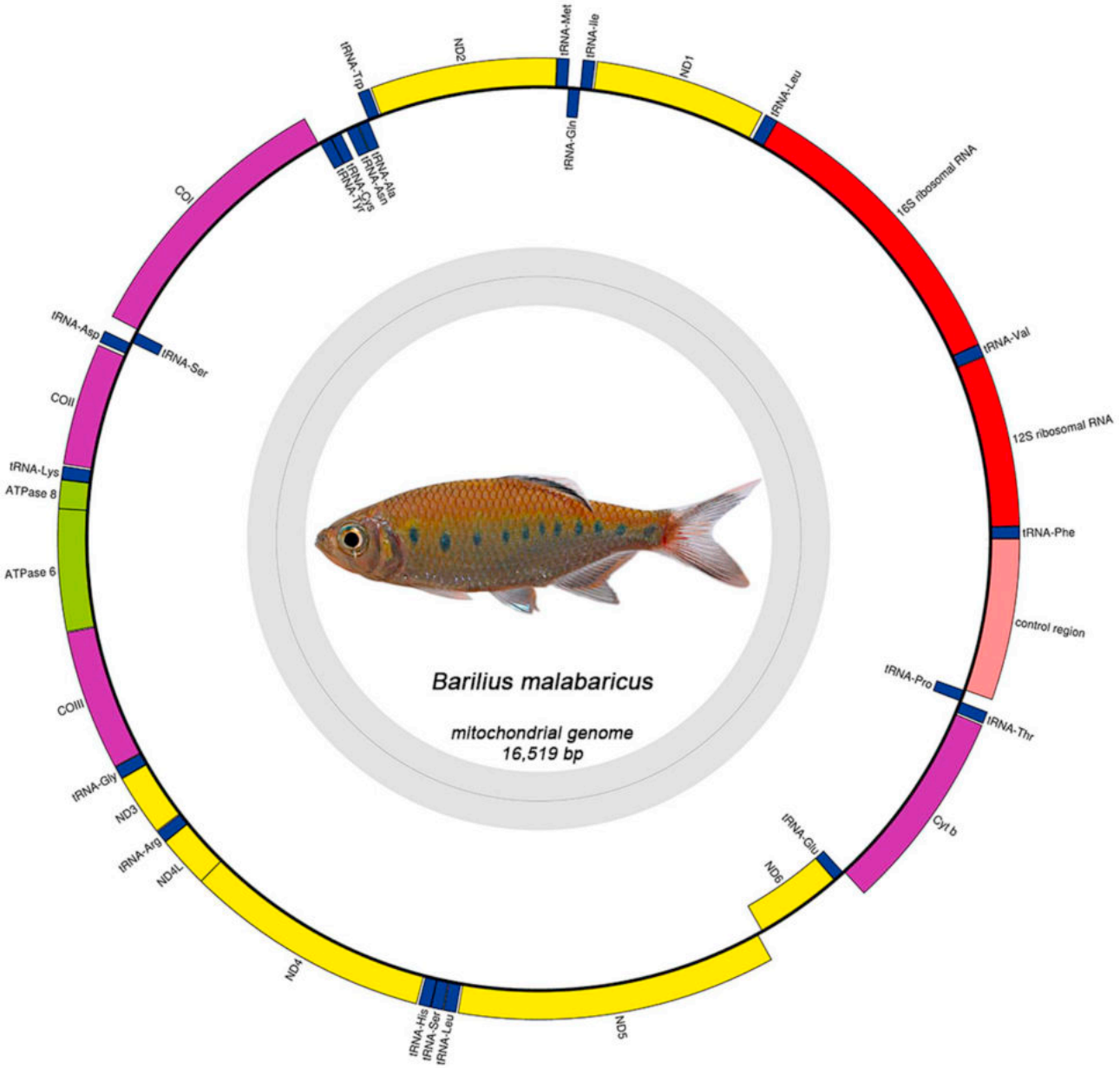

Fig. 2. Mitochondrial genome map of *B. malabaricus*.

mitochondrial genome sequence was edited and aligned using *MegAlign Pro* (DNASTAR, Inc., USA) and annotated using MITOS web server [17]. OGDRAW was used to construct the mitochondrial genome map [18]. Prediction of the secondary tRNA structure was done through MITOS webserver [17]. The software MEGA [19] was used to analyse the nucleotide composition, relative synonymous codon usage (RSCU) and Nei's genetic distance. To understand the base compositional bias of the mitochondrial genome AT skew and GC skew were estimated using the following formulae: AT skew = (A − T)/(A + T) and GC skew = (G − C)/(G + C) [20]. AMOVA was done using the software Arlequin [21]. The phylogenetic tree was constructed using the software MrBayes [22]. The general time reversible substitution model with the invariant site plus eight gamma categories was applied to the data. The MCMC chains were run for 5×10^5 cycles and the trees were sampled at every 100 generations. The sampled trees preceding stationarity were discarded with burnin = 2500 and the remaining tree samples were used for generating the tree. *Bangana tungting* was used as an outgroup for rooting the phylogenetic tree.

3. Results and discussion

3.1. Mitochondrial genome structure and organization

After quality filtering the raw reads, a total of 963,674 high quality clean reads were obtained which was used to assemble the *B. malabaricus* mitochondrial genome. The assembled mitochondrial genome consisted of 16,519 bp (The mitochondrial genome sequence has been submitted in the GenBank under the accession number MN650735),

Table 2
Characteristic features of mitochondrial genome of *B. malabaricus*.

Name	Start	End	Strand	Size (bp)	IGS (bp)	Anticodon	Start codon	Stop codon
tRNA-Phe	1	69	+	69	0	GAA		
12S rRNA	70	1021	+	952	1			
tRNA-Val	1023	1094	+	72	0	TAC		
16S rRNA	1095	2765	+	1671	−1			
tRNA-Leu	2765	2839	+	75	17	TAA		
ND1	2857	3810	+	954	7		ATA	CAA
tRNA-Ile	3818	3889	+	72	−2	GAT		
tRNA-Gln	3888	3958	−	71	2	TTG		
tRNA-Met	3961	4029	+	69	0	CAT		
ND2	4030	5067	+	1038	7		ATG	ATG
tRNA-Trp	5075	5144	+	70	1	TCA		
tRNA-Ala	5146	5213	−	68	1	TGC		
tRNA-Asn	5215	5287	−	73	66	GTT		
tRNA-Cys	5322	5387	−	66	0	GCA		
tRNA-Tyr	5388	5457	−	70	7	GTA		
COI	5465	6997	+	1533	12		ATC	GTA
tRNA-Ser	7010	7080	−	71	3	TGA		
tRNA-Asp	7084	7153	+	70	14	GTC		
COII	7168	7851	+	684	7		ATG	GAA
tRNA-Lys	7859	7934	+	76	0	TTT		
ATP8	7935	8096	+	162	4		ATG	TGG
ATP6	8093	8773	+	681	2		ATG	ATT
COIII	8776	9558	+	783	1		ATG	TCA
tRNA-Gly	9560	9631	+	72	0	TCC		
ND3	9632	9979	+	348	2		ATG	GAG
tRNA-Arg	9982	10,051	+	70	0	TCG		
ND4L	10,052	10,345	+	294	4		ATG	TGC
ND4	10,342	11,715	+	1374	8		ATG	TGA
tRNA-His	11,724	11,793	+	70	0	GTG		
tRNA-Ser	11,794	11,862	+	69	1	GCT		
tRNA-Leu	11,864	11,936	+	73	7	TAG		
ND5	11,944	13,719	+	1776	29		ATT	CTA
ND6	13,749	14,267	−	519	0		ATG	GTT
tRNA-Glu	14,268	14,335	−	68	4	TTC		
Cyt b	14,340	15,473	+	1134	7		ATG	GAA
tRNA-Thr	15,481	15,551	+	71	−1	TGT		
tRNA-Pro	15,551	15,620	−	70	0	TGG		
Control region	15,621	16,519		899				

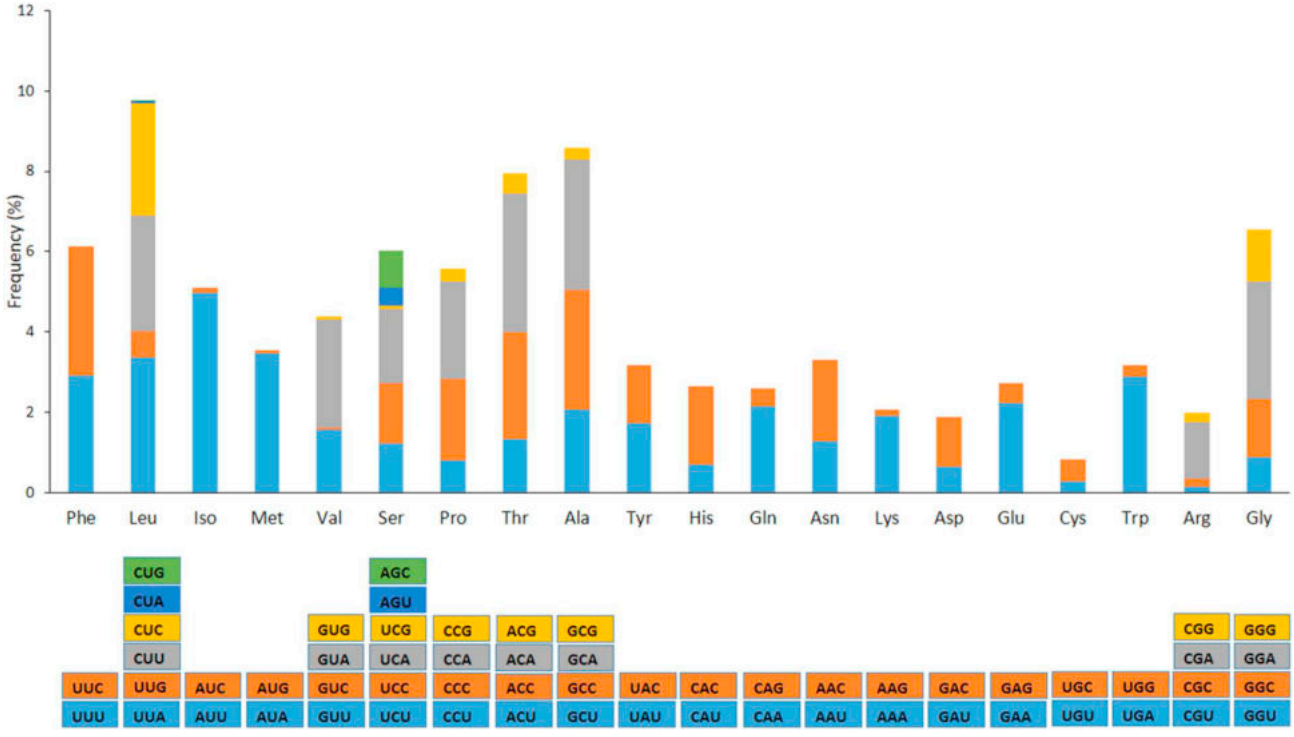

Fig. 3. The codon usage frequency of the mitochondrial protein-coding genes of *B. malabaricus*. Amino acid and codon usage frequency (%) are given on the x- axis and y-axis respectively.

Fig. 4. The relative synonymous codon usage (RSCU) of the mitochondrial protein-coding genes of *B. malabaricus*. Amino acid and RSCU values are given on the x-axis and y-axis respectively.

which was comparable to the previously reported *B. bendelisis* mitochondrial genome (AP011433). The overall base composition appeared to be 30.5% A, 16.6% G, 26.5% C and 26.4% T, which makes as AT and GC percentage of 56.9% and 43.1% respectively, thereby revealing that the mitochondrial genome is biased towards AT (Table 1). The AT.% was slightly higher in the *B. malabaricus* than that of the closely related *B. bendelisis* (Fig. 1). The base composition bias have been reported to play a vital role in the replication and transcription of mitochondrial genome [23]. Besides, it showed negative GC skew value (−0.232) indicating C base is more common than G base whereas AT skewness was positive (0.072) suggesting A base occurs more frequently than T base in the *B. malabaricus* mitochondrial genome (Table 1). The mitochondrial genome was essentially comprised of 22 tRNA, 13 protein coding and 2 ribosomal RNA genes and a control region (Fig. 2; Table 2), which is consistent with the mitochondrial genome of fishes [5,8,9] and other vertebrates [24,25]. The heavy (H) strand also known as the '+' strand codes for most of the genes except nine genes. The exceptional nine genes included eight tRNA genes coding for the amino acids Gln, Ala, Asn, Cys, Tyr, Ser, Glu and Pro, and a NADH dehydrogenase subunit 6 (*ND6*) gene which were located on the other complementary strand (the light or 'L' strand).

3.2. Protein coding genes (PCGs)

The overall length of the protein coding genes was 11,280 bp, which accounted for 68.29% of the whole mitochondrial genome. The base composition of the PCGs was found to be 28.2% A, 28.9% T, 15.9% G and 27% C, which reflects the prevalence of biasness towards AT content among the PCGs (Table 1). Moreover, the AT and GC skew value for the PCGs were − 0.25 and − 0.011 respectively which suggests that the nucleotides T and C have higher occurrence than their respective counterparts (Table 1). As observed in other fishes and vertebrates [5,24], the 13 protein coding genes were composed of seven subunits of NADH ubiquinone oxidoreductase complex (*ND1–6* and *ND4L*), three subunits of cytochrome *c* oxidase (*COI, COII* and *COIII*), a subunit of the ubiquinol cytochrome *b* oxidoreductase complex (*Cyt-b*) and two subunits of ATP synthases (*ATP-6* and *ATP-8*). Out of the 13 protein coding genes, 10 genes (*ND2*, *ND3*, *ND4L*, *ND4*, *ND6*, *COII*, *COIII*, *ATP6*, *ATP8* and *Cyt-b*) start with ATG start codon whereas *ND1*, *ND5* and *COI* use the start codon ATA, ATT and ATC respectively as reported in other fish species [26–29]. Among the protein coding genes, *ND5* (1776 bp) was the longest, whereas *ATP8* was found to be the shortest gene (162 bp). Similar to other vertebrates, there were four pairs of coding genes positioned adjacent to each other *ATP8-ATP6*, *ATP6-COIII*, *ND4L-ND4*, and *ND5-ND6* [22,23]. Among them, *ATP8-ATP6* and *ND4L-ND4* showed overlapping in the sequence spanning 4 bp, which is common characteristics of fishes and other vertebrates [5,8,9]. The comparative study of PCGs of *B. malabaricus* and *B. bendelisis* mitochondrial genome showed a similarity of 81.46% with each other.

Following the pattern of codon degeneracy, the amino acids serine and leucine were encoded by six synonymous codons while the rest of the amino acids were encoded by either four or two codons. The codon usage frequency analysis revealed that, the codons CUA (leucine) and AUU (isoleucine) were most frequently used, while the codons UCG (serine), AAG (lysine) and CGU (arginine) were the least utilised codons (Fig. 3). However, the RSCU values were highest for the codons CUA (leucine) and CGA (arginine). The RSCU values of the PCGs revealed that, the amino acids prefer the codons bearing adenine in the third codon position among their other synonymous alternates (Fig. 4).

3.3. Ribosomal RNAs (rRNA) and transfer RNAs (rRNA)

The subunits of rRNA in the mitochondrial genome were of two types-12S rRNA (952 bp) and 16S rRNA (1671 bp) which had overall

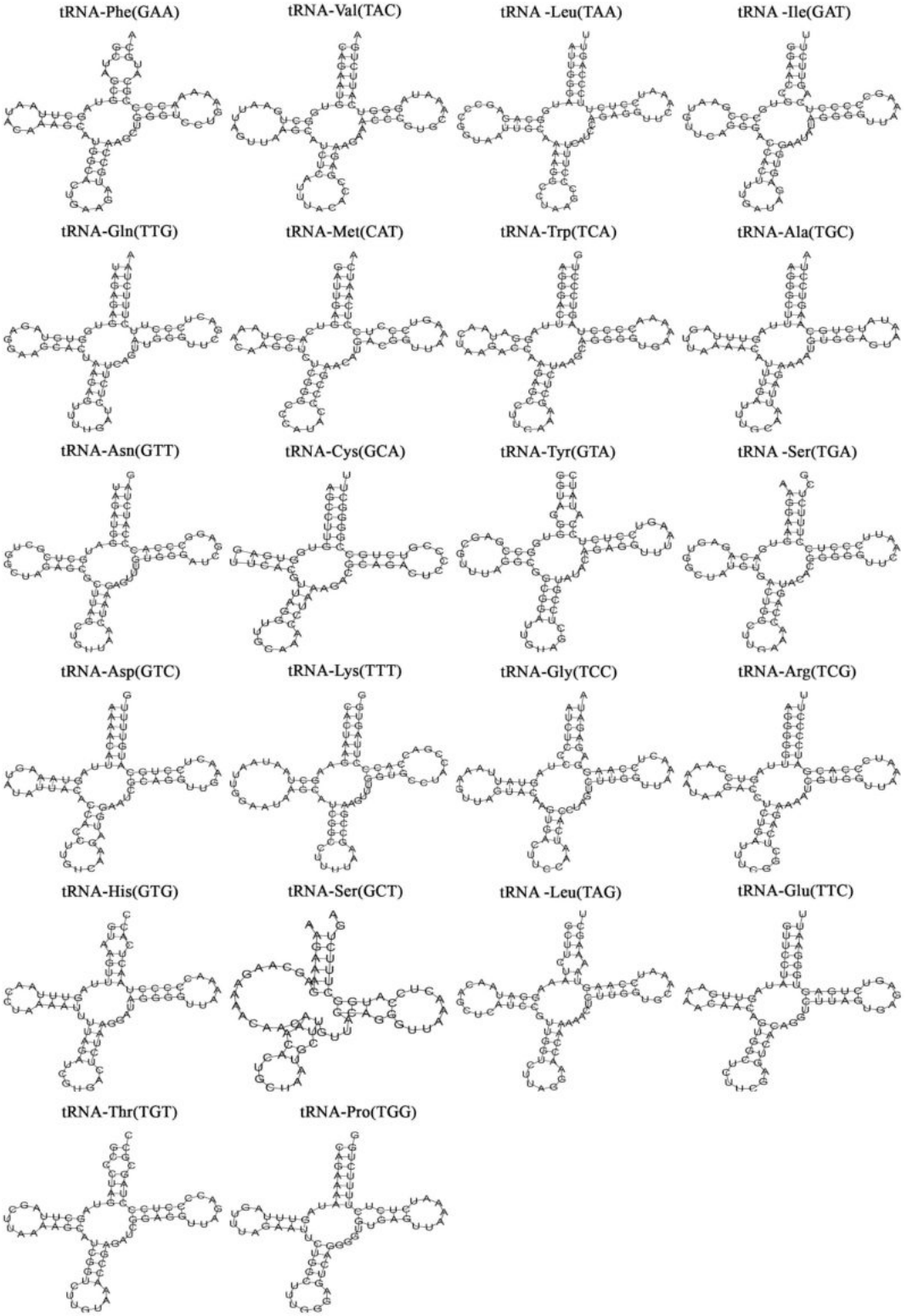

Fig. 5. Secondary structure of the transfer-RNAs (t-RNA) in the mitochondrial genome of *B. malabaricus*.

t- RNA Pro-15621 TTTAGGGCTATAATATGCTGAGAGCCCGCA
CCACAGGATATGTTTACTTACACATACTATGGTTTTACCCATTCT
TAATACATACTATGTATGTATTAGGACATACTTCCACTATCTTCC
TGACAAGCCACTAACATGAAGTACATATTATGATTAATAACCAT
TCATCTATTTTAACCATAAAGCAGGTACTAACTATTTTTAATTCA
CCATTTAAGAGAGAGACCACCAACCAGTTTAAGCAGGAATAC
AAAATCCATGATAGAACCAGGGACATATATACCAAGGGTCATTA
TTAGTGAATTATTCCTGGTATCTGATTCGGAATCTCAGGTCCATC
GCTAAAAGAACCACTAACAATCTAGTAGTAAACGGCATCTGAT
TAGTCAGATGTGTTAACCATTCGATTCATTACCCCACATGCCGA
GCGTTCTTTTATATGCAAGGGGTGAATTTTTATTGGTTTCCTTTT
CACCAACATTTCAGAGTGCAAGCTAAAATGTTAATTAAGGTTG
TACATTTTCCTTGTATGTGATAAATTATATTAATTATCGGAAGACA
TAACTGAAGACTCATTCACTTTTAAGTCATGTACATAACGTATCT
GCACTTTCCTCTATTATTCCTGTTTTCACCCCGGTTTTCGCGCGA
CAAACCCCCTTACCCCCTTAACGCCCCGGAAGTCCTGTTTATCCT
TGTCAAACCCCGAAACCAAGGAGGACCTAAGAACGTGTTAAAC
CAAGCTAGTTCAGATATTGGCTATTGGGTATGCATTATATATATATA
TATATATATATATATATGTACAATTCCAAAATTTTAACAAATTTTGGCA
ACCTTATAAACCCCTGCCAGAATTTAATAAAATAAATTCGGCCCC
GAAAGTCCTAATCTTTTTTGA 16519-t-RNA Phe

Fig. 6. Compositional features of the control region of the *B. malabaricus* mitochondrial genome. Palindromic motif sequence "TACAT' and 'ATGTA' are marked in red and green respectively, and the TA repeats are marked in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

AT content of 54.9% (Table 1). The AT skew value was positive (0.266) while GC skew value was negative (−0.049) which revealed more adenine and cytosine nucleotide prevalence among rRNAs (Table 1). As observed in other fishes and vertebrates, the 12S rRNA and 16S rRNA were clamped by tRNA-Phe and tRNA-Leu with tRNA-Val positioned in between the two rRNA subunits. The 22 tRNA genes contributed a total nucleotide length of 2623 bp to the mitochondrial genome. The AT.% of the tRNAs was found to be higher (54.7%) compared to the GC% (45.3-%). Both the AT and GC skew values were positively biased (0.04 and 0.041 respectively) in the tRNA genes. Among the 22 tRNAs which codes for 20 amino acids, there were two types of tRNAs for serine having sequence TGA and GCT at the anticodon site. The tRNA size ranged from 66 bp of the cysteine to 76 bp of the lysine. Like the case in other fishes [8,9] all the tRNAs displayed a typical clover leaf structure except tRNA Cysteine-GCA and tRNA Serine-GCT in which the D arm stem sequence was not seen to make the ideal Watson-Crick wobble pairing due to non-complementary nucleotide positioning (Fig. 5). Besides, the 12S rRNA and 16S rRNA revealed 92% and 89% similarity between *B. malabaricus* and *B. bendelisis* respectively.

3.4. Control region

The control region of the mitochondrial genome spanned 899 bp length which was flanked by tRNA-Pro and tRNA-Phe on either side (Fig. 2). As observed in other vertebrates and fishes [5,30], it contained the highest AT percentage (64.2%) compared with the other mitochondrial regions. Both the AT and GC skew values were positive, which suggest that the number of adenine and guanine nucleotides are higher than thymine and cytosine (Table 1). The palindromic sequence motifs 'TACAT' and 'ATGTA' which are associated with the termination of heavy strand replication were found in 4 and 3 copies respectively (Fig. 6) as reported in other Cyprinids [5,30]. In addition, a microsatellite like repeat (TA) was observed in the 3′- end of the control region which repeated 13 times. The TA repeat has also been observed in the *B.bendelisis* but slightly in higher number (17 copies) than that of *B. malabaricus*.

3.5. Phylogenetic analysis

As there was only *B.bendelisis* mitochondrial genome available for comparative analysis from the genus *Barilius*, we selected 14 closely related genera (one species per genus) of the subfamily Danioninae for the phylogenetic analysis. The phylogenetic tree was constructed using 15,655 bp sequence after removing highly divergent regions. The phylogenetic tree showed multiple clades and the *B. malabaricus* was placed within a clade containing *B.bendelisis, Raiamas guttatus, Opsaridium ubangiense,* and *Leptocypris sps.* Further, it showed close genetic relationship of *B. malabaricus* and *B. bendelisis* (Fig. 7A). The findings in addition to identification, validated the taxonomical status of *B. malabaricus* as established based on morphological features by previous studies [1,2].

The genus *Barilius* includes about 32 species in which 23 species are present in Indian water bodies [12], however the exact composition of this genus is still unclear. Howes [31,32] suggested that the genus *Barilius* should be restricted to species from South and Southeast Asia, and recognised two groups within the genus. Followed by, Rainboth [33] restricted the genus *Barilius* to species from South Asia, with the Southeast Asian species being placed under *Opsarius*. Molecular phylogenetic studies by Tang et al. [6] and Liao et al. [11] restricted *Barilius* to five species; *B. barila*, *B. evezardi*, *B. modestus*, *B. vagra* and *B. mesopotamicus*, with all other species being moved to *Opsarius*. Thus, we further used the phylogenetically important mtDNA marker genes such as COI, cytochrome *b* and 16S rRNA to understand the phylogenetic relationship of *B. malabaricus* with other *Barilius* species. We also included the sequences of closely related genus *Opsarius* to understand the genetic variability between the genera *Barilius* and *Opsarius*. The Bayesian tree was constructed for each gene fragment separately (Fig. 7B-D). The COI and 16S rRNA genes validated the sequencing results of the present study which showed > 97% homology between our sequence and previously reported *B. malabaricus* sequences (KX461921 & KX461924). All the three markers resulted in similar tree topology and revealed that *B. malabaricus* is genetically closer to *B. canarensis* and *B. bakeri*. Notably, the Western Ghats species, *B. malabaricus, B. canarensis, B. bakeri,* and *B. ardens* were consistently recovered as a distinct clade from other *Barilius* species. Although there were a few other prominent clades, this clade was well supported by the posterior probability for all the three markers. To check the significance of this sub-structuring, AMOVA analysis was performed which explained 26%, 26% and 32% variation between these four *Barilius* species and other *Barilius* species for COI, cytochrome *b* and 16S rRNA genes respectively. The polyphyletic nature of the *Barilius* genus have been also reported in the previous studies [6,11] where *O. bakeri* and *O. canarensis* were recovered in a clade distinct from other species. We further observed that, the following COI gene sequences (*B. bakeri* and *O.bakeri; B. bendelisis and O. bendelisis*; *B.telio* and *O.telio; B.barna* and *O.barna;* Table 3) referred to *Barilius* and *Opsarius* species were clustered together in the phylogenetic tree (Fig. 7B). Nei's genetic distance between these pairs of sequences were also found to be insignificant (Table 3). It suggests that, both *Barilius/Opsarius* samples refer to the same species but due to ambiguity in their taxonomy they have been mentioned as *Barilius* by some authors and *Opsarius* by others. Therefore, further studies using whole mitochondrial or whole genome are called for to clarify the ambiguity existing between the genera *Barilius* and *Opsarius*. We also recommend that the systematics of the bariliine

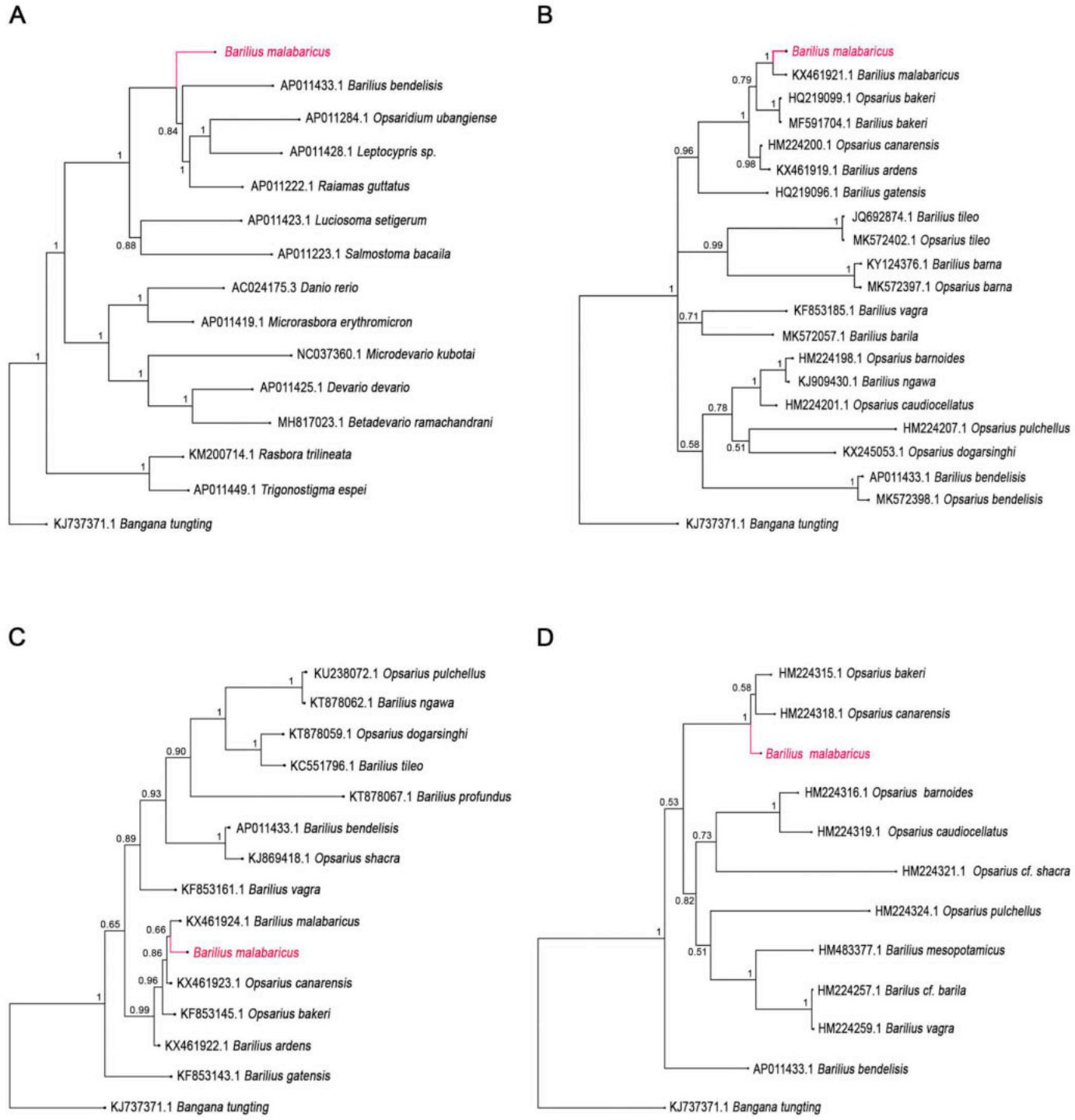

Fig. 7. The phylogenetic relationship of *B. malabaricus* with other closely related species. A) The phylogenetic tree of mitochondrial genome of 14 species belonging to the subfamily Danioninae. B). The phylogenetic tree of COI gene. C). The phylogenetic tree of 16S rRNA gene. D). The phylogenetic tree of cytochrome *b* gene. Bayesian posterior probabilities are shown next to corresponding nodes. The mitochondrial genome, COI, 16S rRNA and cytochrome *b* gene sequences of the various species were downloaded from GenBank. The GenBank accession numbers are shown along with the species names.

Table 3
Nei's genetic distance between *Barilius/Opsarius* species. The Nei's genetic distance was calculated using 655 bp COI gene sequences (MN650735, KX461921, HQ219099, MF591704, HM224200, KX461919, HQ219096, JQ692874, MK572402, KY124376, MK572397, KF853185, MK572057, HM224198, KJ909430, HM224201, HM224207, KX245053, AP011433, MK572398).

Species	*B. malabaricus* (This study)	*B.malab-aricus*	*O. canarensis*	*O. bakeri*	*B. bakeri*	*B. ardens*	*B. gatensis*	*O. caudiocellatus*	*B. ngawa*	*O. barnoides*
B. malabaricus (This study)	0.000									
B.malabaricus	0.033	0.000								
O. canarensis	0.054	0.056	0.000							
O. bakeri	0.065	0.061	0.048	0.000						
B. bakeri	0.065	0.061	0.048	0.000	0.000					
B. ardens	0.059	0.069	0.012	0.061	0.061	0.000				
B. gatensis	0.137	0.138	0.124	0.140	0.140	0.135	0.000			
O. caudiocellatus	0.144	0.149	0.140	0.149	0.149	0.150	0.140	0.000		
B. ngawa	0.146	0.158	0.145	0.168	0.168	0.154	0.152	0.048	0.000	
O. barnoides	0.149	0.160	0.151	0.180	0.180	0.161	0.159	0.054	0.013	0.000
B. barila	0.157	0.162	0.123	0.139	0.139	0.138	0.134	0.149	0.163	0.174
O. dogarsinghi	0.160	0.162	0.160	0.165	0.165	0.169	0.170	0.118	0.116	0.110
O. pulchellus	0.161	0.163	0.148	0.143	0.143	0.152	0.169	0.153	0.171	0.176
B. vagra	0.172	0.174	0.167	0.171	0.171	0.180	0.152	0.157	0.167	0.169
O. barna	0.181	0.179	0.164	0.176	0.176	0.176	0.178	0.192	0.198	0.202
B. barna	0.186	0.179	0.174	0.178	0.178	0.186	0.174	0.197	0.203	0.208
O.bendelisis	0.183	0.180	0.173	0.171	0.171	0.182	0.191	0.170	0.193	0.193
B. bendelisis	0.183	0.183	0.168	0.169	0.169	0.178	0.188	0.175	0.192	0.192
O. tileo	0.204	0.204	0.177	0.199	0.199	0.187	0.194	0.195	0.196	0.195
B. tileo	0.204	0.204	0.177	0.199	0.199	0.187	0.194	0.195	0.196	0.195

Species	*B. barila*	*O. dogarsinghi*	*O. pulchellus*	*B. vagra*	*O. barna*	*B. barna*	*O.bendelisis*	B.bendelisis	*O. tileo*	*B. tileo*
B. malabaricus (This study)										
B.malabaricus										
O. canarensis										
O. bakeri										
B. bakeri										
B. ardens										
B. gatensis										
O. caudiocellatus										
B. ngawa										
O. barnoides										
B. barila	0.000									
O. dogarsinghi	0.188	0.000								
O. pulchellus	0.174	0.171	0.000							
B. vagra	0.139	0.183	0.208	0.000						
O. barna	0.170	0.209	0.213	0.170	0.000					
B. barna	0.177	0.219	0.211	0.175	0.016	0.000				
O.bendelisis	0.186	0.171	0.194	0.183	0.204	0.207	0.000			
B. bendelisis	0.181	0.171	0.199	0.178	0.199	0.197	0.014	0.000		
O. tileo	0.179	0.190	0.208	0.187	0.163	0.175	0.217	0.212	0.000	
B. tileo	0.179	0.190	0.208	0.187	0.163	0.175	0.217	0.212	0.000	0.000

fish need a comprehensive study using morphological as well as molecular analysis for proper conservation and management.

4. Conclusion

We have sequenced the complete mitochondrial genome of *B. malabaricus* for the first time and compared it with closely related species of the subfamily Danioninae. The mitochondrial genome structural features were similar to *B. bendelisis*. In the phylogenetic analysis of the COI, cytochrome *b* and 16S rRNA gene sequences, the species of *Barilius* from west flowing rivers in the Western Ghats were consistently recovered as a distinct clade from other *Barilius* species. We therefore suggest to retain the genus name *Barilius* for the species of the Western Ghats until detailed analyses are conducted on the species now variously placed in the genera *Barilius* and *Opsarius* using both morphological and molecular markers. The genus name *Pteropsarion* (Gunther, 1868), type species *Barilius bakeri*, is also available for *Barilius* species of the Western Ghats.

Declaration of Competing Interest

The authors declare that they have no competing interest.

Acknowledgement

We are thankful to the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, New Delhi (SR/FT/LS-40/2010) for financial assistance. The first author is grateful to the DST-INSPIRE (IF160266) for the financial support in the form of research fellowship.

References

[1] J.D.M. Knight, A. Rai, R.K.P. D'Souza, B. Vijaykrishnan, *Barilius ardens* (Teleostei: Cyprinidae), a new species from the Western Ghats, India, with redescription of *B. malabaricus* and *B. canarensis*, Zootaxa 3926 (2015) 396–412.
[2] T.C. Jerdon, On the freshwater fishes of southern India, Madras Journal of Literature and Science 15 (1849) 302–346.
[3] F. Day, On the fishes of Cochin, on the Malabar coast of India, Proc. Zool. Soc.London 33 (1865) 286–318.
[4] F. Day, The Fishes of INDIA; Being a Natural History of the Fishes Known to Inhabit the Seas and Fresh Waters of India, Burma and Ceylon, London: Quaritsch, London, (1878), pp. 1–168.
[5] C. Wang, Q. Chen, G. Lu, J. Xu, Q. Yang, S. Li, Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): insight into its phylogenic position within Cyprinidae, Gene 424 (2008) 96–101.
[6] K.L. Tang, M.K. Agnew, W.J. Chen, M.V. Hirt, T. Sado, L.M. Schneider, J. Freyhof, Z. Sulaiman, E. Swartz, C. Vid-thayanon, M. Miya, K. Saitoh, A.M. Simons, R.M. Wood, R.L. Mayden, Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae), Mol. Phylogenet. Evol. 57 (2010) 189–214.
[7] M. Nagarajan, M. Raja, P. Vikram, Genetic characterization of *Bagarius* species using cytochrome c oxidase I and cytochrome b genes, Mitochondrial DNA Part A 27 (2016) 381–383.
[8] D.X. Chen, W.Y. Chu, X.L. Liu, X.X. Nong, Y.L. Li, S.J. Du, J.S. Zhang, Phylogenetic studies of three sinipercid fishes (Perciformes: Sinipercidae) based on complete mitochondrial DNA sequences, Mitochondrial DNA 23 (2012) 70–76.
[9] A. Pavan-Kumar, S. Raman, P.G. Koringa, N. Patel, T. Shah, R.K. Singh, G. Krishna, C.G. Joshi, P. Gireesh-Babu, A. Chaudhari, Complete mitochondrial genome of threatened mahseer *Tor tor* (Hamilton 1822) and its phylogenetic relationship within Cyprinidae family, J. Genet. 95 (2016) 853–863.
[10] S. Sah, A. Barat, V. Pande, J. Sati, C. Goel, Population structure of Indian Hill Trout (*Barilius bendelisis*) inferred from variation in mitochondrial DNA sequences, Adv. Biol. Res. 5 (2011) 93–98.
[11] T.Y. Liao, E. Ünlü, S.O. Kullander, Western boundary of the subfamily Danioninae in Asia (Telesotei: Cyprinidae): derived from the systematic position of *Barilius mesopotamicus* based on molecular and morphological data, Zootaxa 2880 (2011) 31–40.
[12] A.K. Singh, R. Kumar, M. Singh, A.K. Mishra, U.K. Chauhan, V.S. Baisvar, R. Verma, N.S. Nagpure, B. Kushwaha, Mitochondrial 16S rRNA gene-based evolutionary divergence and molecular phylogeny of *Barilius* spp, Mitochondrial DNA 26 (2015) 41–47.
[13] A.K. Singh, R. Kumar, A.K. Mishra, M. Singh, V.S. Baisvar, U.K. Chauhan, B. Kushwaha, N.S. Nagpure, Authentication of five *Barilius* species from Indian waters using DNA barcoding, Russ. J. Genet. 52 (2016) 840–846.
[14] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. J. 17 (1) (2011) 10.
[15] N.A. Joshi, J.N. Fass, Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33), [Software] (2011).
[16] H. Xu, X. Luo, J. Qian, X. Pang, J. Song, G. Qian, J. Chen, S. Chen, FastUniq: a fast de novo duplicates removal tool for paired short reads, PLoS One 7 (2012) e52249.
[17] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: Improved de novo metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2013) 313–319.
[18] S. Greiner, P. Lehwark, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, Nucleic Acids Res. 47 (2019) W59–W64.
[19] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[20] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J. Mol. Evol. 41 (1995) 353–358.
[21] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows, Mol. Ecol. Resour. 10 (2010) 564–567.
[22] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.
[23] S.J. Wei, M. Shi, X.X. Chen, M.J. Sharkey, C. van. Achterberg, G.Y. Ye, J.H. He, New views on strand asymmetry in insect mitochondrial genomes, PLoS One 5 (2010) e12708.
[24] S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. De Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, Nature 290 (1981) 457.
[25] V.R. Prabhu, M.S. Arjun, K. Bhavana, R. Kamalakkannan, M. Nagarajan, Complete mitochondrial genome of Indian mithun, *Bos frontalis* and its phylogenetic implications, Mol. Biol. Rep. 46 (2019) 2561–2566.
[26] L. Gong, W. Shi, Z.M. Wang, X.G. Miao, X.Y. Kong, Control region translocation and a tRNA gene inversion in the mitogenome of *Paraplagusia japonica* (Pleuronectiformes: Cynoglossidae), Mitochondrial DNA 24 (2013) 671–673.
[27] R. Li, Y. Gao, Y. Hou, S. Ye, L. Wang, J. Sun, Q. Li, Mitochondrial genome sequencing and analysis of scuticociliates (*Uronema marinum*) isolated from *Takifugu rubripes*, Mitochondrial DNA Part B 3 (2018) 736–737.
[28] J.C.B. Nunez, L.M. Biancani, P.A. Flight, D.E. Nacci, D.M. Rand, D.L. Crawford, M.F. Oleksiak, Stable genetic structure and connectivity in pollution-adapted and nearby pollution-sensitive populations of *Fundulus heteroclitus*, R. Soc. Open Sci. 5 (2018) 171532.
[29] M.J. Alam, J.H. Kim, S. Andriyono, J.H. Lee, S.R. Lee, H. Park, H.W. Kim, Characterization of complete mitochondrial genome and gene organization of sharp-spined notothenia, *Trematomus pennellii* (Perciformes: Nototheniidae), Mitochondrial DNA Part B 4 (2019) 648–649.
[30] X. Wang, J. Wang, S. He, R.L. Mayden, The complete mitochondrial genome of the Chinese hook snout carp *Opsariichthys bidens* (Actinopterygii: Cypriniformes) and an alternative pattern of mitogenomic evolution in vertebrate, Gene 399 (2007) 11–19.
[31] G.J. Howes, The anatomy, phylogeny and classification of bariliine cyprinid fishes, Bull. Brit. Mus. (Nat. Hist.), Zool. 37 (1980) 129–198.
[32] G.J. Howes, Additional notes on bariliine cyprinid fishes, Bull. Brit. Mus. (Nat. Hist.), Zool. 45 (1983) 95–101.
[33] W.J. Rainboth, Cyprinids of South East Asia, in: I. Winfield, J. Nelson (Eds.), Cyprinid Fishes: Fish & Fisheries Series 3, Springer, Dordrecht, 1991, pp. 156–210.