Pakistan J. Zool., pp 1-10, 2024. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20240603083501

Complete Mitochondrial Genome Sequence of *Nannostomus beckfordi* and Molecular Phylogenetic Analysis

Xian-Ru Li, Yi-Jing Zhan, Kai-Rui Zhang and Cheng-He Sun*

College of Life Sciences, Nanjing Forestry University, Nanjing, 210037, China

ABSTRACT

Nannostomus beckfordi belonging to the order Characiformes is an economically valuable fish. To better understand the origin, evolution, and phylogenetic status of *N. beckfordi*, the mitochondrial genome was sequenced and characterized. The total genome length was 16,742 bp, including 13 protein-coding genes, 22 tRNA-coding genes, two rRNA-coding genes, and one control region (D-loop), and the base composition exhibited a clear AT bias. Except for *ND6* in the L-chain, all the other protein-coding genes were located in the H-chain. The *N. beckfordi* mitochondrial genome had 12 intergenic intervals of 1–29 bp (with a total length of 64 bp). There were six regions of gene overlap (33 bp in total ranging from 1 to 10 bp each). The start codon for protein-coding genes was ATG, except for *COXI*, which used GTG; in terms of termination codon usage, *ATPase6, COXIII, ND4, ND2, COXII,* and *ND3* had incomplete termination codons TA or T, *COXI* had AGG, and the remaining genes contained typical termination codons, TAG or TAA. A phylogenetic tree was constructed using the maximum likelihood method based on the full-length mitochondrial genomes of 72 Characiformes species. Consistent relationships were obtained based on morphological and molecular data, indicating that *N. beckfordi* is most closely related to *Lebiasina astrigata*. This study lays a molecular genetic foundation for the scientific conservation of this species, helps judge its invasion risk, and provides reference for the prevention and control measures against it at the molecular level.

Article Information
Received 03 June 2024
Revised 10 June 2024
Accepted 19 June 2024
Available online 17 October 2024
(early access)

Authors' Contribution
X-RL: Formal analysis, methodology, writing - original draft, writing review and editing. Y-JZ and K-RZ: Formal analysis, writing review and editing. C-HS: Conceptualization, Funding acquisition, software, writing review and editing.

Key words
Fish, Comparative genomics, Genome structure, Alien species

INTRODUCTION

The pencil fish *Nannostomus beckfordi* belongs to the order Characiformes and family Lebiasinidae. Amazonian fish in the family Lebiasinidae stand out for their ornamental potential; they are distributed from Central America to South America. It is characterized by a cylindrical body, red coloration and a long black stripe on its body, and a superior mouth. It lives on shoals, feeds on microcrustaceans or periphytons, and decomposes organic matter (Weitzman and Weitzman, 2003). *N. beckfordi* as an aquarium fish has value in the international market (Allpondsolutions, 2019; Prang, 2008).

The mitochondrial genome, characterized by a simple structure, compact gene arrangement, and strict maternal inheritance, has been used extensively in molecular phylogenetics and population genetics research on metazoans over the past decade (Dellaporta *et al*., 2006; Helfenbein *et al*., 2004; Boore and Brown, 1998). Owing to the developments in DNA sequencing technology, it has become an ideal material for studying the origin, evolution, phylogeny, and population genetics of various species.

Alien species (as opposed to native species) are those whose presence in a region is attributable to human actions, deliberate or inadvertent, that enabled them to overcome biogeographical barriers (Richardson *et al*., 2000, 2011; Pyšek *et al*., 2004; Essl *et al*., 2018). *N. beckfordi* has high ornamental value and economic value in China. But because they are distributed from Central America to South America, it is an alien species. Full-length *N. beckfordi* mitochondrial genome sequences are limited, and little is known about the evolution of the species. In this study, direct sequencing was used to obtain the complete *N. beckfordi* mitochondrial genome sequence for comprehensive and in-depth bioinformatics, evolutionary, and phylogenetic analyses, laying a molecular genetic foundation for the scientific conservation of this species. It also helps judge its invasion risk, and provides reference for the prevention and control measures against it at the molecular level.

* Corresponding author: sch543116411@163.com
0030-9923/2024/0001-0001 $ 9.00/0

MATERIALS AND METHODS

Experimental materials

N. beckfordi was collected in Sep 2022 at the Fuzimiao Flower, Bird, Fish and Insect Market in Qinhuai District, Nanjing City, Jiangsu Province (118°50′29.92″E, 32°0′21.02″N). After preliminary morphological identification, the tail fin tissue was taken and stored in a 1.5 ml centrifuge tube at -80°C for future use.

DNA extraction

All samples were cut into small pieces, and the genomic DNA was extracted using DNAiso Reagent (company address Caf. #) according to the manufacturer's instructions. The ethanol precipitated genomie DNA was suspended in sterilized water, and its quality detected by using 1% agarose gel electrophoresis. DNA purity and concentration were determined using an ultra-micro spectrophotometer and DNA barcoding technology.

Sequencing and annotation

A Covaris ultrasonicator was used to break the genomic DNA into approximately 350 bp fragments, followed by end-repair and the addition of an A base and sequencing adapter at the 3' end. PCR amplification was performed on the products, with recovery and purification using magnetic beads to construct a library. After quality inspection of the DNA library, the Illumina HiSeq high-throughput sequencing platform was used for paired-end sequencing with a sequencing data volume of no less than 6 Gb for each sample. Sequencing was conducted by Shanghai Parsenor Biotechnology Co., Ltd. Low-quality reads and splice sequences were filtered. NOVOPlasty (Dierkxsens *et al*., 2017) was used for de novo assembly, referring to the published mitochondrial genome sequence of *Lebiasina astrigata* MH921292, and BioEdit 7 2.5 (Hall, 1999) was used to proofread the spliced sequences. ITOS 2 (Eliade and Fernández, 1968) (http://mitos.bioinf.unileipzig.de/index.py) and MitoFish (Iwasaki *et al*., 2013) were used for gene annotation. The complete mitogenome sequence of *N. beckfordi* was submitted to the NCBI (National Center for Biotechnology Information) database (Supplementary Table 1).

Phylogenetic analysis

The mitochondrial whole genome sequences of 72 fish were downloaded from NBCI (Table I). PhyloSuite v1.2.1 (Zhang *et al*., 2020) was used for phylogenetic analyses using the maximum likelihood (ML) method. In particular, the ML tree was inferred using IQ TREE (Nguyen *et al*., 2015) under the HKY+I+G4+F model with 50000 ultrafast bootstrap replicates (Minh *et al*., 2013) and

Table I. *Nannostomus beckfordi* mitochondrial genome composition.

Gene	Position		Size/ bp	Intergenic nucleo-tides	Codon		Str-and
	From	To			Start	Stop	
tRNA-Phe	1	70	70	0			H
12S rRNA	71	1025	955	0			H
tRNA-Val	1026	1097	72	0			H
16S rRNA	1098	2795	1698	0			H
tRNA-Leu	2796	2871	76	0			H
ND1	2872	3846	975	0	ATG	TAG	H
tRNA-Ile	3851	3922	72	4			H
tRNA-GIn	3921	3991	71	-2			L
tRNA-Met	3992	4061	70	0			H
ND2	4062	5106	1045	0	ATG	T	H
tRNA-Trp	5107	5178	72	0			H
tRNA-Ala	5182	5250	69	3			L
tRNA-Asn	5252	5324	73	1			L
tRNA-Cys	5354	5420	67	29			L
tRNA-Tyr	5420	5490	71	-1			L
COXI	5492	7048	1557	1	GTG	AGG	H
tRNA-Ser	7040	7110	71	-9			L
tRNA-Asp	7115	7184	70	4			H
COXII	7198	7888	691	13	ATG	T	H
tRNA-Lys	7889	7962	74	0			H
ATPase8	7964	8131	168	1	ATG	TAA	H
ATPase6	8122	8804	683	-10	ATG	TA	H
COXIII	8805	9589	785	0	ATG	TA	H
tRNA-Gly	9590	9662	73	0			H
ND3	9663	10008	346	0	ATG	T	H
tRNA-Arg	10009	10079	71	0			H
ND4L	10080	10376	297	0	ATG	TAA	H
ND4	10370	11751	1382	-7	ATG	TA	H
tRNA-His	11752	11821	70	0			H
tRNA-Ser	11822	11889	68	0			H
tRNA-Leu	11891	11963	73	1			H
ND5	11964	13802	1839	0	ATG	TAA	H
ND6	13799	14317	519	-4	ATG	TAG	L
tRNA-Glu	14318	14386	69	0			L
Cytb	14392	15534	1143	5	ATG	TAA	H
tRNA-Thr	15536	15607	72	1			H
tRNA-Pro	15609	15678	70	1			L
D-loop	15679	16742	1063	0			H

the Shimodaira-Hasegawa-like approximate likelihood ratio test to evaluate branch support (Guindon *et al*., 2010).

RESULTS

Mitochondrial genome structure

The total length of the *N. beckfordi* mitochondrial genome was 16742 bp (Fig. 1 and Table I). The genome included 13 protein-coding genes (PCGs), 22 tRNA genes, two rRNA genes, and one control region (D-loop). The *N. beckfordi* genome had 12 spacer regions (ND1 and tRNA-Ile; tRNA-Trp and tRNA-Ala; tRNA-Ala and tRNA-Asn; tRNA-Asn and tRNA-Cys; tRNA-Tyr and COXI): tRNA-Ser and tRNA-Asp, tRNA-Asp and COXII; tRNA-Lys and ATPase-8; tRNA-Ser and tRNA-Leu; tRNA-Glu and Cytb; Cytb and tRNA-Thr; tRNA-Thr and tRNA-Pro). The lengths of these 12 intervals ranged from 1 to 29 bp, with a total length of 64 bp, and the maximum interval was detected between tRNA-Asn and tRNA-Cys. There were six overlapping gene regions (tRNA-Ile and tRNA-Gln; tRNA-Cys and tRNA-Tyr; COXI and tRNA-Ser; ATPase-8 and ATPase-6; ND4L and ND4; and ND5 and ND6), with a total of 33 bp of overlap. These six overlapping regions ranged in length from 1 to 10 bp, and gene overlap was the longest between ATPase-8 and ATPase-6. Additionally, 19 tightly arranged gene pairs did not overlap or contained gaps.

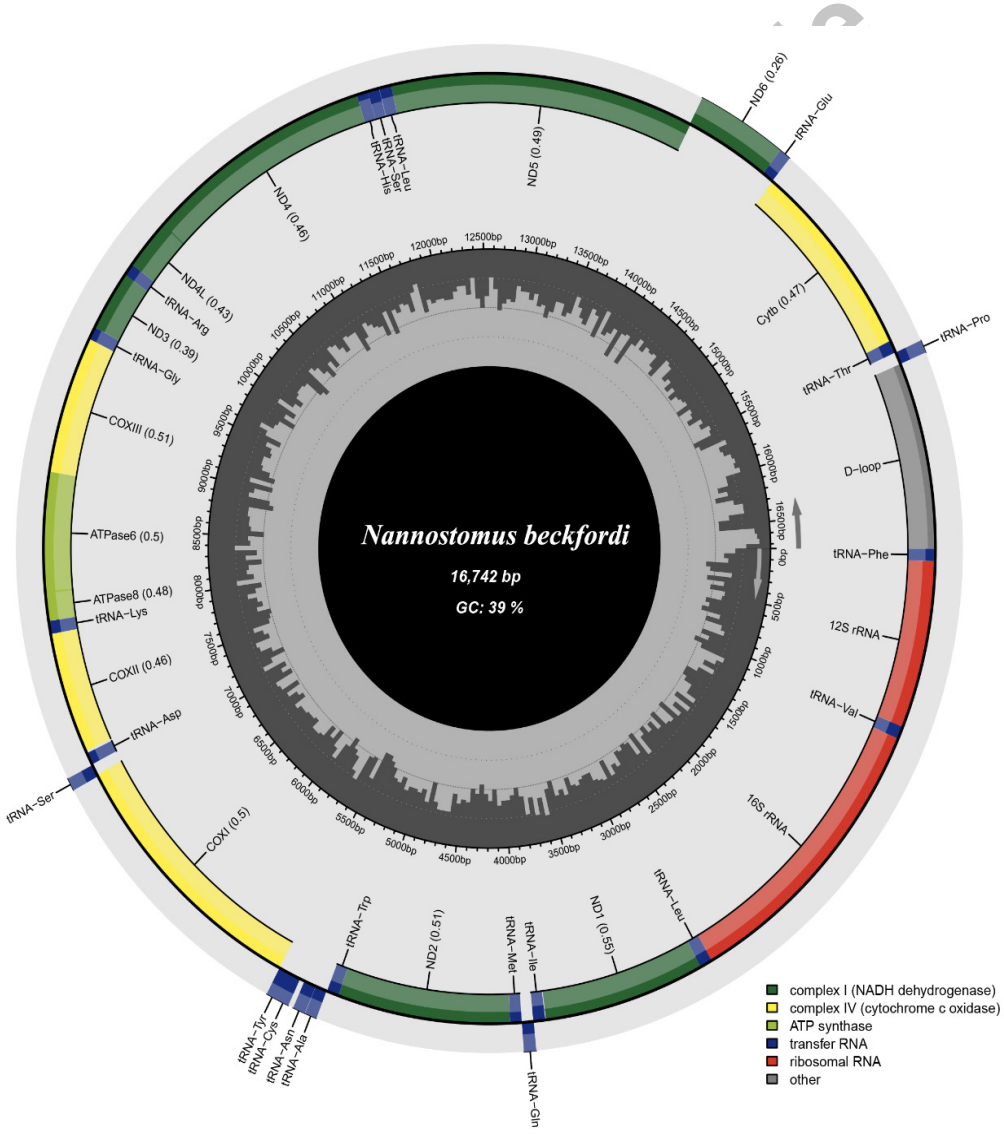

Fig. 1. *Nannostomus beckfordi* mitochondrial genome structure.

Table II. Base composition of the whole *Nannostomus beckfordi* mitochondrial genome.

Regions	Strand	Size (bp)	T (%)	C (%)	A (%)	G (%)	AT (%)	GC (%)
PCGs	+	10902	30.8	25.3	30.1	13.8	60.9	39.1
PCGs	-	519	42.6	11	15.2	31.2	57.8	42.2
rRNAs	+	2653	22.8	22.9	35.2	19.1	58	42
tRNAs	+	1003	27.7	20.6	32.9	18.7	60.6	39.3
tRNAs	-	561	31.6	16.2	23.7	28.5	55.3	44.7
Full genome	+	16742	28.9	24.4	32	14.8	60.9	39.2

PhyloSuite v1.2.1 was used to analyze the base composition of the mitochondrial genome of *N. beckfordi* (Table II). The base distribution in the complete mitochondrial genome sequence was as follows: 28.9% (T), 24.4% (C), 32% (A), and 14.8% (G). The A+T and G+C contents were 60.9% and 39.2%, respectively. There was a clear AT bias. The A+T contents were higher than the G+C contents in H-chain PCGs, L-chain PCGs, H-chain rRNAs, H-chain tRNAs, and L-chain tRNAs. The A+T content of H-chain PCGs was the highest at 60.9%, which was equal to the A+T content of the mitochondrial genome sequence.

Protein-coding genes

The *N. beckfordi* mitochondrial genome contained 13 PCGs with a length of 11421 bp, accounting for 68.2% of the entire genome. In the *N. beckfordi* mitochondrial genome, PCGs were present in both L and H-chains. Except for one NADH reductase complex subunit (*ND6*) on the L-chain, all other PCGs were distributed on the H-chain, including one cytochrome b (*Cytb*), two ATP synthase subunits (*ATPase-6* and *ATPase-8*), three cytochrome oxidase subunits (*COXI, COXII*, and *CO III*), and seven NADH reductase complex subunit-coding genes (*ND1-6* and *ND4L*) (Table I). The start and stop codons of the 13 PCGs are listed in Table I. The most frequent start codon was ATG, followed by GTG. The 12 PCGs with ATG as the start codon were *ND1, ND2, COXII, ATPase8, ATPase6, COXIII, ND3, ND4L, ND4, ND5, ND6*, and *Cytb*. The only PCG with GTG as the start codon was *COXI*. The termination codons of *N. beckfordi* were mainly TAA, followed by the incomplete termination codons T and TA. Four PCGs, *ATPase-8, ND4L, ND5*, and *Cytb*, had TAA as the termination codon, while the only PCG (*COXI*) had AGG as the termination codon. The only PCGs with TAG as the termination codon were *ND1* and *ND6*. The termination codons of the remaining six genes were incomplete, with *ND2, COXII*, and *ND3* using T as the termination codon and *ATPase6, COXIII*, and *ND4* using TA.

An analysis of the base composition of 13 PCGs in the entire mitochondrial genome of *N. beckfordi* is summarized in Table III. The A + T content was higher than the G + C content in 1st, 2nd, and 3rd codons of the H-chain and the 2nd and 3rd codons of the L-chain. Only for the 1st codon of the L-chain was the A + T content slightly lower than the G + C content, indicating a clear AT bias. The third codon of the H-chain had the highest A + T%, whereas the first codon of the L-chain had a low A + T content. The A + T content of the PCG sequences in the H or L-chain increased gradually from the first to the third codon. Moreover, at the same position in the codon, the A + T content of the H-chain codons was higher than that of the L-chain codons.

Table III. Base content at different codon positions of protein-coding genes.

Regions	Strand	Size (bp)	U (%)	C (%)	A (%)	G (%)	AT (%)	GC (%)
1st codon position	+	3634	23.9	24.5	27.8	23.8	51.7	48.3
	-	173	39.3	7.5	10.4	42.8	49.7	50.3
2nd codon position	+	3634	40.7	27.3	18.8	13.2	59.5	40.5
	-	173	44.5	19.7	9.8	26	54.3	45.7
3rd codon position	+	3634	27.9	24.1	43.5	4.5	71.4	28.6
	-	173	43.9	5.8	25.4	24.9	69.3	30.7

The frequency of amino acid usage for each PCG in the mitochondrial genome is shown in Figure 2. The molar mass percentage (mol%) of the amino acid Leu1 (10.74%) was greater than 10%, making it the most frequently used amino acid among all PCGs. Lys (2.21%), Asp (1.89%), Glu (2.47%), Cys (0.71%), Arg (1.95%), and Ser1 (1.55%) had a mol% of ≤ 2.5%, making them the least frequently used amino acids among all PCGs.

The frequency and relative usage of each codon of the *N. beckfordi* mitochondrial genes are shown in Table IV. Each of the approximately 29 codons had a RESU value greater than 1, indicating that these are the preferred codons of *N. beckfordi* mitochondrial genes. Among the above codons, most ended with either A or U bases. Codons ending with C bases were less frequent, with RSCU values of less than 1. However, codons ending in G were least frequent, and all RSCU values were less than 1, indicating the use or avoidance of codons in *N. beckfordi* mitochondrial genes. The *N. beckfordi* mitochondrial genes preferred TAA as a termination codon.

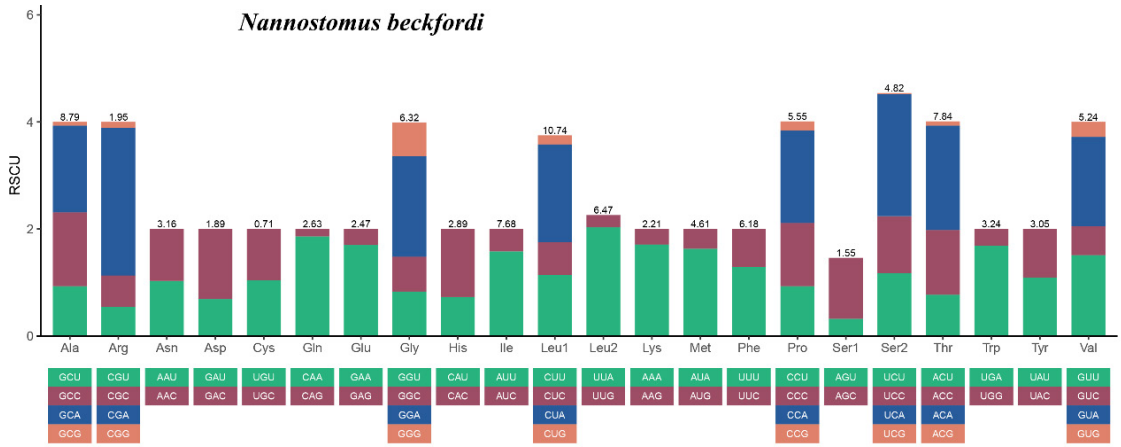

Fig. 2. RSCU values of amino acid and codon usage bias in protein-coding genes.

Table IV. RSCU values for the codons of genes.

AA	Codon	Count	RSCU	AA	Codon	Count	RSCU
Phe	UUU	152	1.29	Ala	GCA	135	1.62
	UUC	83	0.71		GCG	6	0.07
Leu2	UUA	221	2.03	Tyr	UAU	63	1.09
	UUG	25	0.23		UAC	53	0.91
Leu1	CUU	124	1.14	His	CAU	40	0.73
	CUC	67	0.61		CAC	70	1.27
	CUA	199	1.83	Gln	CAA	93	1.86
	CUG	18	0.17		CAG	7	0.14
Ile	AUU	230	1.58	Asn	AAU	62	1.03
	AUC	62	0.42		AAC	58	0.97
Met	AUA	143	1.63	Lys	AAA	72	1.71
	AUG	32	0.37		AAG	12	0.29
Val	GUU	75	1.51	Asp	GAU	25	0.69
	GUC	27	0.54		GAC	47	1.31
	GUA	83	1.67	Glu	GAA	80	1.7
	GUG	14	0.28		GAG	14	0.3
Ser2	UCU	47	1.17	Cys	UGU	14	1.04
	UCC	43	1.07		UGC	13	0.96
	UCA	92	2.28	Trp	UGA	104	1.69
	UCG	1	0.02		UGG	19	0.31
Pro	CCU	49	0.93	Arg	CGU	10	0.54
	CCC	62	1.18		CGC	11	0.59
	CCA	91	1.73		CGA	51	2.76
	CCG	9	0.17		CGG	2	0.11
Thr	ACU	57	0.77	Ser1	AGU	13	0.32
	ACC	90	1.21		AGC	46	1.14
	ACA	145	1.95	Gly	GGU	50	0.83
	ACG	6	0.08		GGC	39	0.65
Ala	GCU	78	0.93		GGA	113	1.88
	GCC	115	1.38		GGG	38	0.63

tRNA genes, rRNA genes, and D-loop

There were 22 tRNA genes in the *N. beckfordi* mitochondrial genome, with tRNA gene sizes ranging from 67 to 76 bp and an overall length of 1564 bp, accounting for 9.34% of the entire genome. Eight tRNA genes were present on the L-chain, namely *tRNA-Gin, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Ser, tRNA-Glu*, and *tRNA-Pro*. The base composition of these genes included A, G, T, and C contents of 23.7%, 28.5%, 31.6%, and 16.2%, respectively. The A+T content was 55.3%, AT skewness was -0.142, and GC skewness was 0.275 Fourteen tRNA genes were present on the H-chain, namely *tRNA-Phe, tRNA-Val, tRNA-Leu, tRNA-Ile, tRNA-Met, tRNA-Trp, tRNA-Asp, tRNA-Lys, tRNA-Gly, tRNA-Arg, tRNA-His, tRNA-Ser, tRNA-Leu*, and *tRNA-Thr*. The A+T content was 60.6%, and the A, G, T, and C contents of tRNA were 32.9%, 18.7%, 27.7%, and 20.6%. AT skewness was 0.086, and GC skewness was -0.048. Detailed information on the base composition of tRNA genes in *N. beckfordi* mitochondrial genes is shown in Table V.

The *N. beckfordi* mitochondrial genome contains two rRNA genes, both located on the H-chain, with an overall length of 2653 bp and an A+T content of 58%. AT skewness was 0.214 and GC skewness was -0.09. The ribosomal small subunit 12S rRNA gene of mitochondrial DNA was located between the *tRNA-Phe* and *tRNA-Val* genes, whereas the ribosomal large subunit 16S rRNA gene was located between the tRNA-Val and tRNA-Leu genes. The length of the *12S rRNA* gene was 955 bp, and the length of the *16S rRNA* gene was 1698 bp. The two genes were separated from each other by the *tRNA-Val* gene at a distance of 73 bp. The base composition of the *12S rRNA* gene was A, G, T, and C contents of 33.1%, 20.7%, 21.8%, and 24.4%, respectively. The A+T content was 54.9%, with an AT bias of 0.206 and a GC bias of

Table V. Base composition of tRNA and rRNA genes in the whole mitochondrial genome of *Nannostomus beckfordi.*

Regions	Strand	Size (bp)	T (%)	C (%)	A (%)	G (%)	AT (%)	GC (%)	AT skew	GC skew
16S rRNA	+	1698	23.4	22	36.4	18.2	59.8	40.2	0.218	-0.095
12S rRNA	+	955	21.8	24.4	33.1	20.7	54.9	45.1	0.206	-0.081
rRNAs	+	2653	22.8	22.9	35.2	19.1	58	42	0.214	-0.09
tRNAs	+	1003	27.7	20.6	32.9	18.7	60.6	39.3	0.086	-0.048
tRNAs	-	561	31.6	16.2	23.7	28.5	55.3	44.7	-0.142	0.275

-0.081. The base composition of the *16S rRNA* genes included A, G, T, and C contents of 36.4%, 18.2%, 23.4%, and 22%, respectively. The A+T content was 59.8%, with an AT bias of 0.218 and a GC bias of -0.095. Detailed information on the base composition of the rRNA genes in the *N. beckfordi* mitochondrial genome is shown in Table V. The *N. beckfordi* mitochondrial genome contained a D-Loop region with a length of 1063 bp, located between tRNA-Pro and tRNA-Phe.

Phylogenetic relationships

Based on whole mitochondrial genome sequences of *N. beckfordi* and 72 other species in Characiformes, a phylogenetic tree was constructed using PhyloSuite v1.2.1 and the ML method.

In the ML tree, the order Characiformes formed a monophyletic group, separate from the representative species in the order Siluriformes used as an outgroup. Species in the order Characiformes were divided into two branches with high support. *Crenuchus spilurus* in the family Crenuchidae formed a branch that diverged earliest from the Characiformes lineage. The other branch was composed of 15 families (Alestidae, Anostomidae, Bryconidae, Chalceidae, Childontidae, Ctenoluciidae, Curimatidae, Erythrinidae, Gasteropelecidae, Hemiodontidae, Hepsetidae, Lebiasinidae, Parodontidae, Prochilodontidae, and Serrasalmidae). This branch also included a pair of sisters group branches.

In particular, one branch was formed by *Phenacogrammus interruptus* in the family Alestidae, *Abrametes hypelonotis*, *Leporinus affinis*, *Megaporinus elongatus*, and *Megaporinus piavussu* in the family Anostomidae, *Hydrolycus scomberoides*in the family Chalceidae, *Childus punctatus* in the family Childontidae, *Curimata mivartii*, and *Curiatopsis evelynae* in the family Curimatidae, *Hoplias intermedius* and *Hoplias malabaricus* in the family Erythrinidae, *Hemiodopsis gracilis* in the family Hemiodontidae, *Hepsetus odoe* of the family Hepsetidae, *Apariodon affinis* in the family Parodontidae, *Ichthyocephalus longirostris*, *Prochilodus argenteus*, *Prochilodus costatus*, *Prochilodus harttii*, *Prochilodus lineatus*, and *Prochilodus vimboides* in the family Prochilodontidae, as well as *Colossoma macropomum*, *Metynnis hypsauchen*, *Myloplus rubripinnis*, *Piaractus brachypomus*, and *Piaractus mesopotamicus* in the family Characin. The other lineage was formed by members of the families Lebiasinidae (*Lebiasina astrigata* and *Nannostomus beckfordi*), Gasteropelecidae (*Carnegiella strigata*), Ctenoluciidae (*Boulengerella machulata* and *Ctenolucius hujeta*), Chalceidae (*Chalceus macroepidotus*, *Aphyocharax rathbuni*, *Astyanax aeneus*, *Astyanax altiparanae*, *Astyanax lacustris*, *Astyanax mexico*, *Deuterodon giton*, *Gephrocharax atracaudatus*, *Grundulus bogotensis*, *Gymnocorymbus ternetzi*, *Hemigrammus armstrongi*, *Hemigrammus bleheri*, *Hemigrammus erythrozonus*, *Hemigrammus ocellifer*, *Hyphessobrycon amandae*, *Hyphessobrycon amapaensis*, *Hyphessobrycon anisitsi*, *Hyphessobrycon elachys*, *Hyphessobrycon flammeus*, *Hyphessobrycon herbertaxelrodi*, *Hyphessobrycon megalopterus*, *Hyphessobrycon pulchripinnis*, *Hyphessobrycon roseus*, *Hyphessobrycon socolofi*, *Hyphessobrycon sweglesi*, *Impaichthys kerri*, *Knodus borki*, *Moenkhausia costae*, *Moenkhausia sanctaefilomenae*, *Nematobrycon palmeri*, *Oligosarcus argenteus*, *Paracheirodon axelrodi*, *Paracheirodon innesi*, *Pristella maximalis*, *Psalidodon fasciatus*, *Psalidodon paranae*, *Psalidodon rivularis*, *Thayeria boehlkei*, and *Brycon henni*), and Bryconidae (*Brycon nattereri* and *Brycon orbityanus* and *Salmininae*).

One branch formed by *C. strigata* of the family Gasteropelecidae and *B. maculata* of the family Ctenoluciidae, and another branch included *L. astrigata* and *N. beckfordi* of the family Lebiasinidae, which together formed a monophyletic group with a node self-expansion support rate of 100%. In the family Lebiasinidae, if two Lebiasinidae lineages were clustered together, the morphological and molecular relationships were consistent, indicating the close genetic relationship. The phylogenetic positions of each species in the ML tree based on rRNA and PCG data are shown in Figure 3.

DISCUSSION

The *N. beckfordi* mitochondrial genome sequence was 16742 bp, including 13 PCGs, 22 tRNA genes,

two rRNA genes, and one control region (D-loop). Fish share gene sequence similarity with other vertebrates, indicating high conservation (Meyer, 1994). Compared with single-copy nuclear genes, the rate of evolution of the mitochondrial genome is faster, and this can be explained by a higher frequency of mitochondrial genome mutations. In particular, the D-loop region has the fastest rate of evolution and is generally used for intraspecific evolutionary analyses (i.e., comparisons between populations); the rate of change of rRNA genes is relatively slow and these regions are commonly used for species- or family-level analyses (Milinkovitch *et al*., 1993). The moderate rates of evolution of protein-coding and tRNA genes make them suitable for both intra- and interspecific analyses. The most commonly used genes are cytochrome b (Cyt b) and NADH dehydrogenase subunit (ND) (Zardoya and Meyer, 1996).

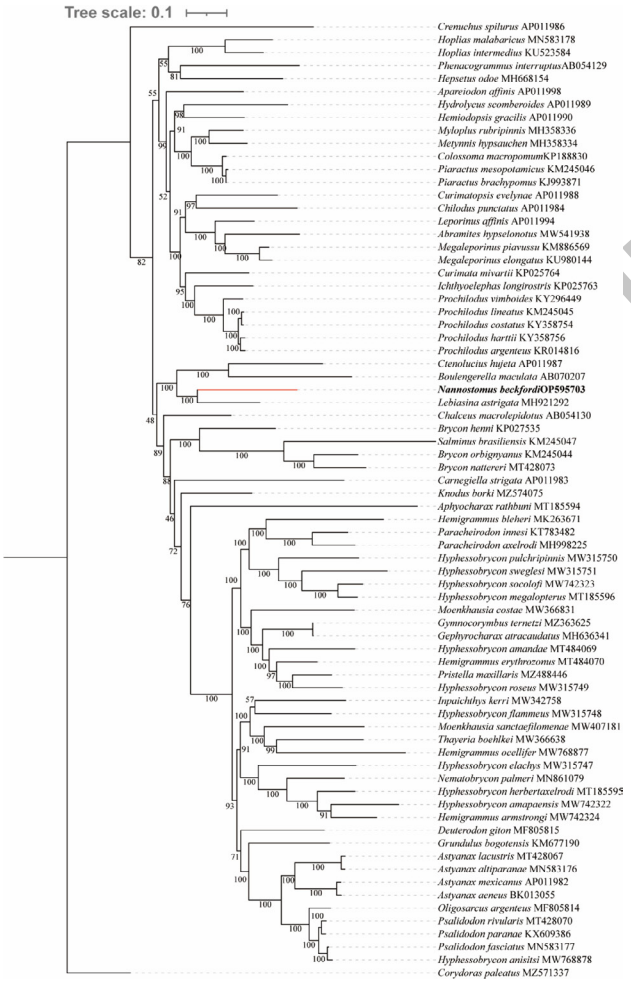

Fig. 3. Maximum likelihood tree based on the PCG dataset (numbers at nodes represent the self-expanding support rate).

The *N. beckfordi* mitochondrial genes are similar to those of other fish in terms of codon usage (Wang *et al*., 2008, 2011; Wu *et al*., 2004). ATG is the most common start codon. There are two common types of termination codons, the complete termination codons TAG and TAA and the incomplete termination of codons TA and T. From this, we can infer that the start and stop codons in the mitochondrial genome may have a special significance in evolution. In the *N. beckfordi* mitochondrial PCGs, TAA was the most common termination codon. However, there were six incomplete termination codons.

The *N. Beckfordi* codons of the PCGs in the entire mitochondrial genome mostly ended with A or U bases, whereas codons ending with C bases appeared less frequently, with most RSCU values less than 1. Codons ending with G were least frequent, and all RSCU values were less than 1, indicating the use or avoidance of codons in the *N. beckfordi* mitochondrial genes. Generally speaking, this preference for codons is the result of mutations and selection, which has significance for the study of the origin and evolution of species, as demonstrated in *Burkholderia anthracis* (Zhao *et al*., 2007).

There were 22 tRNA genes in the *N. beckfordi* mitochondrial genome, with tRNA gene sizes ranging from 67 to 76 bp and an overall length of 1564 bp, accounting for 9.34% of the entire genome. The tRNA genes encoded by the H-chain were scattered between the protein and rRNA genes, with adjacent genes closely connected or even overlapping at intervals of 0–29 bases. Generally, *tRNA-Met, tRNA-His*, and *tRNA-Leu* show the highest conservation, whereas *tRNA-Ser* shows the greatest variation (Tzeng *et al*., 1992). The *12S rRNA* gene was located between the *tRNA-Phe* and *tRNA-Val* genes, whereas the *16S rRNA* gene was located between the *tRNA-Val* and *tRNA-Leu* genes. The length of the *12S rRNA* genes was 955 bp, and the length of the *16S rRNA* gene was 1698 bp. The two genes are separated from each other by the *tRNA-Val* gene at a distance of 73 bp. The rates of evolution of *16 srRNA* and *12 srRNA* are the slowest among mitochondrial genomes, indicating high conservation (Hickson *et al*., 1996; Flook and Rowell, 1997; Baker, 2000; Page, 2000; Misof *et al*., 2002; Page *et al*., 2002; Yoshizawa and Johnson, 2003). The *N. beckfordi* mitochondrial genome contained a D-Loop region with a length of 1063 bp, located between *tRNA-Pro* and *tRNA-Phe*.

A phylogenetic tree of *N. beckfordi* and 72 other Characiformes species was constructed using the ML method. Relationships between *N. beckfordi* and *L. astrigata* based on morphological and molecular data were consistent, indicating a close genetic relationship. This study provided the complete mitochondrial genome sequence obtained using direct sequencing for *N. beckfordi*,

and in-depth bioinformatics analyses will further provide further insight into the genetic characteristics, origin, and evolution of the species. The analysis of phylogenetic relationships within the order Characiformes provide a theoretical basis for the protection and utilization of *N. beckfordi* genetic resources. It also helps judge its invasion risk, and provides reference for the prevention and control measures against it at the molecular level.

DECLARATIONS

Acknowledgments

We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Funding

The present study was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Ethical statement

All specimens in this study were collected in accordance with Chinese laws. The collection and sampling of the specimens were reviewed and approved by the Animal Ethics Committee of Nanjing Forestry University. All experiments were conducted with respect to animal welfare and care. The study complied with CBD and Nagoya protocols and with the ARRIVE guidelines (https://arriveguidelines.org).

Data availability statement

The complete mitochondrial genome sequence and annotations of *Nannostomus beckfordi* is available in the GenBank and accession numbers OP595703.

Supplementary material

There is supplementary material associated with this article. Access the material online at: https://dx.doi.org/10.17582/journal.pjz/20240603083501

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Allpondsolutions, 2019. Https://www.allpondsolutions.co.uk/gold-pencilfish/. Accessed in 06/03/2019.

Baker, R.J., 2000. *Systematics of bats of the family Phyllostomidae based on RAG2 DNA sequences.* https://doi.org/10.5962/bhl.title.147024

Boore, J.L. and Brown, W.M., 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.*, **8**: 668-674. https://doi.org/10.1016/S0959-437X(98)80035-X

Dellaporta, S.L., Xu, A., Sagasser, S., Jakob, W., Moreno. M.A., Buss, L.W. and Schierwater, B., 2006. Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proc. natl. Acad. Sci. India B.*, **103**: 8751-8756. https://doi.org/10.1073/pnas.0602076103

Dierckxsens, N., Mardulyn, P. and Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucl. Acids Res.*, **45**: e18-e18. https://doi.org/10.1093/nar/gkw955

Eliade, M. and Fernández, L.G., 1968. *Mitoyrealidad*. Guadarrama, Madrid, Vol. 25.

Essl, F., Bacher, S., Genovesi, P., Hulme, P.E., Jeschke, J.M., Katsanevakis, S., Kowarik, I., Kühn, I., Pyšek, P., Rabitsch, W., Schindler, S., Van Kleunen, K.M., Vilà, M., Wilson, J.R.U. and Richardson, D.M., 2018. Which taxa are alien? Criteria, applications, and uncertainties. *Bioscience*, **68**: 496-509. https://doi.org/10.1093/biosci/biy057

Flook, P.K. and Rowell, C.H.F., 1997. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Mol. Phylogenet. Evol.*, **8**: 177-192. https://doi.org/10.1006/mpev.1997.0425

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of Phyml 3.0. *Syst. Biol.,* **59**: 307-321. https://doi.org/10.1093/sysbio/syq010

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*, **41**: 95-98.

Helfenbein, K.G., Fourcade, H.M., Vanjani, R.G. and Boore, J.L., 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc. natl. Acad. Sci. India B.*, **101**: 10639-10643. https://doi.org/10.1073/pnas.0400941101

Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J. and Penny, D., 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S

rRNA. *Mol. Biol. Evol.*, **13**: 150-169. https://doi.org/10.1093/oxfordjournals.molbev.a025552

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M. and Nishida, M., 2013. Mito Fish and Mito Annotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.*, **30**: 2531-2540. https://doi.org/10.1093/molbev/mst141

Meyer, A., 1994. 5.1 molecular phylogenetic studies of fish. *Genet. Evol. Aquat. Organ.,* **219**.

Milinkovitch, M.C., Orti, G. and Meyer, A., 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. *Nature*, **361**: 346-348. https://doi.org/10.1038/361346a0

Minh, B.Q., Nguyen, M.A.T. and Von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.*, **30**: 1188-1195. https://doi.org/10.1093/molbev/mst024

Misof, B., Anderson, C.L., Buckley, T.R., Erpenbeck, D., Rickert, A. and Misof, K., 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: Site-specific rate variation is clustered and frequently detected. *J. mol. Evol.*, **55**: 460-469. https://doi.org/10.1007/s00239-002-2341-1

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. and Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**: 268-274. https://doi.org/10.1093/molbev/msu300

Page, R.D., 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. *Nucl. Acids Res.*, **28**: 3839-3845. https://doi.org/10.1093/nar/28.20.3839

Page, R.D.M., Cruickshank, R. and Johnson, K.P., 2002. Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable. *Insect mol. Biol.*, **11**: 361-369. https://doi.org/10.1046/j.1365-2583.2002.00346.x

Prang, G., 2008. An industry analysis of the freshwater ornamental fishery with particular reference to the supply of Brazilian freshwater ornamentals to the UK market. *Sci. Mag. Uakari*, **3**: 7-52. https://doi.org/10.31420/uakari.v3i1.18

Pyšek, P., Richardson, D.M., Rejmánek, M., Webster, G.L., Williamson, M. and Kirschner, J., 2004. Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon*, **53**: 131-143. https://doi.org/10.2307/4135498

Richardson, D.M., Pyšek, P. and Carlton, J.T., 2011. A compendium of essential concepts and terminology in invasion ecology. *Fifty Years Invasion Ecol. Legacy Charles Elton*, **1**: 409-420. https://doi.org/10.1002/9781444329988.ch30

Richardson, D.M., Pyšek, P., Rejmanek, M., Barbour, M.G., Panetta, F.D., West, C.J., 2000. Naturalization and invasion of alien plants: Concepts and definitions. *Divers. Distrib.*, **6**: 93-107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

Tzeng, C.S., Hui, C.F., Shen, S.C. and Huang, P.C., 1992. The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: Conservation and variations among vertebrates. *Nucl. Acids Res.*, **20**: 4853-4858. https://doi.org/10.1093/nar/20.18.4853

Wang, C., Chen, Q., Lu, G., Xu, J., Yang, Q. and Li, S., 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. *Gene*, **424**: 96-101. https://doi.org/10.1016/j.gene.2008.07.011

Wang, Z., Yonezawa, T., Liu, B., Ma, T., Shen, X., Su, J., Guo, S. and Liu, J., 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. *Mol. Biol. Evol.*, **28**: 1553-1556. https://doi.org/10.1093/molbev/msq336

Weitzman, M. and Weitzman, S.H., 2003. Family Lebiasinidae (pencil fishes). In: *Check list of the freshwater fishes of South and Central America* (eds. R.E. Reis, S.O. Kullander, C.J. Ferraris Jr). Edipucrs, Porto Alegre, Brazil. pp. 241–250.

Wu, X.Y., Wang, L., Chen, S.Y., Zan, R.G., Xiao, H. and Zhang, Y.P., 2004. The complete mitochondrial genomes of two species from *Sinocyclocheilus* (Cypriniformes: Cyprinidae) and a phylogenetic analysis within Cyprininae. *Mol. Phylogenet. Evol.*, **37**: 263-227.

Yoshizawa, K. and Johnson, K.P., 2003. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. *Mol. Phylogenet. Evol.*, **29**: 102-114. https://doi.org/10.1016/S1055-7903(03)00073-3

Zardoya, R. and Meyer, A., 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Phylogenet Evol.*, **13**: 933-942. https://doi.org/10.1093/oxfordjournals.molbev.a025661

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W.X. and Wang, G.T., 2020. PhyloSuite:

An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.*, **20**: 348-355. https://doi.org/10.1111/1755-0998.13096

Zhao, S., Zhang, Q., Chen, Z., Zhao, Y. and Zhong, J., 2007. The factors sha synonymous codon usage in the genome of *Burkholderia mallei*. *J. Genet. Genom.*, **34**: 362-372. https://doi.org/10.1016/S1673-8527(07)60039-3

Pakistan J. Zool., pp 1-10, 2024. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20240603083501

Supplementary Material

Complete Mitochondrial Genome Sequence of *Nannostomus beckfordi* and Molecular Phylogenetic Analysis

Xian-Ru Li, Yi-Jing Zhan, Kai-Rui Zhang and Cheng-He Sun*

College of Life Sciences, Nanjing Forestry University, Nanjing, 210037, China

Supplementary Table I. Mitochondrial genomes used in the phylogenetic analysis.

Order/ Family	Organism	Length	AT%	ID
Order: Characiformes				
Family: Alestidae				
	Phenacogrammus interruptus	16,652	55.3	AB054129
Family: Anostomidae				
	Abramites hypselonotus	16,685	57.9	MW541938
	Leporinus affinis	16,259	54.9	AP011994
	Megaleporinus elongatus	16,774	57.0	KU980144
	Megaleporinus piavussu	16,682	57.0	KM886569
Family: Bryconidae				
	Brycon henni	16,885	55.5	KP027535
	Brycon nattereri	16,837	58.1	MT428073
	Brycon orbignyanus	16,800	57.2	KM245044
	Salminus brasiliensis	17,721	55.8	KM245047
Family: Chalceidae				
	Chalceus macrolepidotus	16,850	55.8	AB054130
	Aphyocharax rathbuni	16,678	59.6	MT185594
	Astyanax aeneus	16,769	58.8	BK013055
	Astyanax altiparanae	16,730	58.3	MN583176
	Astyanax lacustris	16,763	58.2	MT428067
	Astyanax mexicanus	16,682	58.8	AP011982
	Deuterodon giton	16,643	59.2	MF805815
	Gephyrocharax atracaudatus	17,049	58.5	MH636341
	Grundulus bogotensis	17,123	60.1	KM677190
	Gymnocorymbus ternetzi	17,999	58.3	MZ363625
	Hemigrammus armstrongi	16,789	58.4	MW742324
	Hemigrammus bleheri	17,021	58.4	MK263671
	Hemigrammus erythrozonus	16,710	57.5	MT484070
	Hemigrammus ocellifer	18,141	60.3	MW768877
	Hydrolycus scomberoides	16,548	52.4	AP011989
	Hyphessobrycon amandae	16,701	57.2	MT484069
	Hyphessobrycon amapaensis	17,824	59.5	MW742322
	Hyphessobrycon anisitsi	16,920	57.4	MW768878
	Hyphessobrycon elachys	17,224	59.3	MW315747
	Hyphessobrycon flammeus	16,008	59.7	MW315748
	Hyphessobrycon herbertaxelrodi	16,841	59.4	MT185595
	Hyphessobrycon megalopterus	16,773	59.5	MT185596
	Hyphessobrycon pulchripinnis	17,020	57.4	MW315750
	Hyphessobrycon roseus	17,046	56.9	MW315749
	Hyphessobrycon socolofi	17,132	58.6	MW742323
	Hyphessobrycon sweglesi	16,080	56.0	MW315751
	Inpaichthys kerri	17,032	60.4	MW342758
	Knodus borki	16,837	58.1	MZ574075
	Moenkhausia costae	15,811	54.7	MW366831

Table continued on next column.....

Table continued on next page.....

* Corresponding author: sch543116411@163.com
0030-9923/2024/0001-0001 $ 9.00/0

Order/ Family	Organism	Length	AT%	ID
	Moenkhausia sanctaefilomenae	18,437	60.0	MW407181
	Nematobrycon palmeri	17,340	61.2	MN861079
	Oligosarcus argenteus	16,711	57.6	MF805814
	Paracheirodon axelrodi	17,100	59.0	MH998225
	Paracheirodon innesi	16,962	58.5	KT783482
	Pristella maxillaris	16,753	57.4	MZ488446
	Psalidodon fasciatus	16,400	57.6	MN583177
	Psalidodon paranae	16,707	57.1	KX609386
	Psalidodon rivularis	16,812	57.2	MT428070
	Thayeria boehlkei	16,624	57.9	MW366638
Family: Chilodontidae				
	Chilodus punctatus	16,869	59.9	AP011984
Family: Crenuchidae				
	Crenuchus spilurus	16,361	61.9	AP011986
Family: Ctenoluciidae				
	Boulengerella maculata	16,446	55.5	AB070207
	Ctenolucius hujeta	16,599	57.9	AP011987
Family: Curimatidae				
	Curimata mivartii	16,705	56.8	KP025764
	Curimatopsis evelynae	16,779	56.4	AP011988
Family: Erythrinidae				
	Hoplias intermedius	16,629	56.0	KU523584
	Hoplias malabaricus	16,602	56.3	MN583178
Family: Gasteropelecidae				
	Carnegiella strigata	17,852	64.5	AP011983
Family: Hemiodontidae				
	Hemiodopsis gracilis	16,731	54.9	AP011990
Family: Hepsetidae				
	Hepsetus odoe	16,802	53.1	MH668154
Family: Lebiasinidae				
	Lebiasina astrigata	16,899	57.4	MH921292
	Nannostomus beckfordi	16,742	60.9	OP595703
Family: Parodontidae				
	Apareiodon affinis	16,679	56.6	AP011998
Family: Prochilodontidae				
	Ichthyoelephas longirostris	16,840	54.5	KP025763
	Prochilodus argenteus	16,697	55.5	KR014816
	Prochilodus costatus	16,699	55.5	KY358754
	Prochilodus harttii	16,697	55.6	KY358756
	Prochilodus lineatus	16,699	55.5	KM245045
	Prochilodus vimboides	16,696	55.4	KY296449
Family: Characin				
	Colossoma macropomum	16,703	54.5	KP188830
	Metynnis hypsauchen	16,737	52.9	MH358334
	Myloplus rubripinnis	16,662	52.7	MH358336
	Piaractus brachypomus	16,722	54.8	KJ993871
	Piaractus mesopotamicus	16,722	54.8	KM245046
Order: Siluriformes				
Family: Callichthyidae				
	Corydoras paleatus	16,593	58.2	MZ571337

Table continued on next column.....