Pakistan J. Zool., pp 1-10, 2023. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20230213080243

The Complete Mitochondrial Genome of a Rare Cavefish (*Sinocyclocheilus cyphotergous*) and Comparative Genomic Analyses in *Sinocyclocheilus*

Xiaoping Gao[1], Yanping Li[2]*, Renyi Zhang[3], Yunyun Lv[2], Yongming Wang[2], Jinrong Shi[2], Jiang Xie[2], Chiping Kong[1] and Lekang Li[1]

[1]*Jiujiang Academy of Agricultural Sciences, Jiujiang, 332101, China*
[2]*Key Laboratory of Sichuan Province for Fishes Conservation and Utilization in the Upper Reaches of the Yangtze River, College of Life Sciences, Neijiang Normal University, Neijiang, 641100, China*
[3]*Key Laboratory of National Forestry and Grassland Administration on Biodiversity Conservation in Karst Mountainous Areas of Southwestern China, School of Life Science, Guizhou Normal University, Guiyang, Guizhou, 550001, China*

Xiaoping Gao and Yanping Li contributed equally to this project.

Article Information
Received 13 February 2023
Revised 20 March 2023
Accepted 09 April 2023
Available online 09 October 2023
(early access)

Authors' Contribution
XG, YL and RZ conceived and designed the experiments, XG, YL, RZ, and YLv analyzed the data. YL wrote the original draft; YW, JS, JX, KC, YLv and LL reviewed and revised the manuscript. All authors have read and approved the publication of this manuscript.

Key words
***Sinocyclocheilus cyphotergous*, Mitogenome assembly, Annotation, Phylogenetic relationship, Evolution analysis**

ABSTRACT

The *Sinocyclocheilus cyphotergous*, belonging to the family Cyprinidae, is endemic to the Karst area of the Yunnan-Guizhou Plateau. Here, we determined the complete mitogenome of *S. cyphotergous* using an Illumina Hiseq X Ten sequencer. This mitogenome's structure is typical circular with 16,611 bp in length, consisting of 13 protein-coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes, and a control region. The overall base composition of *S. cyphotergous* is 31.33% A, 25.89% T, 26.49% C, and 16.29% G with a slight AT bias of 57.22%. Most mitochondrial genes except *ND6* and eight tRNAs were encoded on the heavy strand. All tRNA genes fold into the typical cloverleaf secondary structures, except for tRNA-Ser (AGY) that lacked the dihydrouracil arm. 15 of 22 tRNA genes were found to have 29 G-U mismatches in their secondary structures, which formed a weak bond. In addition, mismatches of A-C, C-C, U-U, and A-A were also found in their tRNA secondary structures. Result of substitution rate estimation among the mitochondrial protein coding genes (PCGs) showed *ATP8* had the largest average *Ka* and *Ka/Ks*, while *COI* had the lowest, which implied that *ATP8* might evolve more quickly than the other mitochondrial protein coding genes. Phylogenetic analyses based on Bayesian inference (BI) and maximum likelihood (ML) revealed all *Sinocyclocheilus* species in this research formed a solid monophyletic group and grouped into two major clades with strong support excluded *S. jii*. Additionally, *S. cyphotergous* in this study was closely related to *S. multipunctatus* and *S. punctatus*. In summary, this study provided novel insights into the phylogeny of the *Sinocyclocheilus* fishes, conducive to the conservation genetics and cave adaptation for *S. cyphotergous*.

INTRODUCTION

The genus *Sinocyclocheilus* (Cypriniformes: Cyprinidae) is monophyletic and endemic to the Karst area of the Yunnan-Guizhou Plateau and its surrounding region Guangxi, southwestern China, being the most diverse genus of cyprinid fishes in China (Zhao and Zhang, 2009). There are more than 60 valid species that have been recognized in this genus, and most of them distribute in the tributaries of Xijiang River, while few are distributed in the Jinsha River and Honghe River. Also, most of the *Sinocyclocheilus* have various cave-dwelling behaviors or underground streams (Zhao and Zhang, 2009).

At present, although there are some studies referring to the phylogenetic relationship of the genus *Sinocyclocheilus* (Li and He, 2009; Liang *et al*., 2011; Luo and Zhang, 2021; Xiao *et al*., 2005; Zhang and Wang, 2018), the phylogenetic relationship among the species

* Corresponding author: liyanping@njtc.edu.cn
0030-9923/2022/0001-0001 $ 9.00/0

of this genus has not been completely solved due to the continuous discovery of new species, and the inconsistency of the phylogenetic relationship between morphological characteristics and molecular data. Additionally, the lack of genetic markers in molecular phylogenetic investigations further complicates determining this genus' evolutionary affiliation. Mitochondrial DNA has many valuable features including relatively conserved gene content and organization, lack of genetic recombination, maternal inheritance, and relatively fast evolutionary rate. Hence, partial or complete mitochondrial genes have been used to determine molecular phylogenetic and evolutionary relationships (Boore, 1999; Lv *et al*., 2018; Zhang *et al*., 2022; Zhang and Wang, 2018). However, the study of mitogenomes is still scarce in the genus of *Sinocyclocheilus* compared to its rich species.

S. cyphotergous is the most representative cave fish in Guizhou province and is named for its unique horn protrusion on its back. It bears vulnerable status on the International Union for the Conservation of Nature (IUCN) Red List, and is listed as national second-grade key protected fish. However, the available genetic data for this species remains scarce. The purposes of this study is shown as: (1) Assemble and annotate the mitochondrial genome of *S. cyphotergous* to determine its genomic organization and structure of the complete mitochondrial sequence; (2) reconstruct the phylogeny of the genus *Sinocyclocheilus* using the mitochondrial genome obtained here and NCBI; (3) determine the substitution rates of protein coding genes based on the complete mitochondrial genome of the genus *Sinocyclocheilus* and to identify signals of positive selection. This study not only improves understanding of genomic and phylogenetic information of *Sinocyclocheilus*, but also provides reference for the conservation genetics of *S. cyphotergous*.

MATERIALS AND METHODS

Sample collection

Samples were collected from Luodian County, Guizhou Province, China (25°34′31″N, 106°50′18″E). Muscle samples were preserved in 95% ethanol, and voucher specimens were deposited at the School of Life Sciences, Guizhou Normal University, Guiyang, Guizhou Province, China. (Voucher no. GZNUSLS202009042). Sampling was performed according to Chinese animal protection laws.

DNA extraction and sequencing

Genomic DNA was extracted from muscle tissue using a DNeasy blood and tissue kit (QIAGEN, Hilden, Germany). DNA integrity, purity and concentration were assessed with an Agilent 5400 fragment analyzer (Agilent Technologies, Santa Clara, CA, U.S.A.). After the DNA sample was qualified, and the template size was 100 ng/ul, it was randomly sheared with a *Covaris ultrasonicator* (Covaris Inc., Woburn, MA, U.S.A.), and then the library was constructed through several steps: End repair and phosphorylation, adding A-tailing, ligating index adapter, purification, denaturing and PCR amplification. After the library was constructed, a Qubit 2.0 (Life Technologies, Singapore) was used to quantify and dilute the library. We then employed an Agilent 2100 Bioanalyzer (Agilent) to detect inserted fragments in the library. Finally, the effective concentration of the library was accurately quantified by q-PCR to ensure the library quality. After that, different libraries were pooled into the flow cell according to the effective concentration and target drop-off data. Illumina paired-end sequencing was conducted with Illumina Hiseq X Ten sequencer (Illumina, San Diego, CA, U.S.A.).

Mitochondrial genome assembly and annotation

The raw data contained adapter information, low-quality bases, and undetected bases (indicated by N), which would interfere with subsequent analysis. We therefore filtered the raw data using the following criteria: (1) filtered out reads containing adapter sequences; (2) removed paired reads, when the content of N in a single-ended sequence exceeded 10%; (3) base with quality no more than 5 was regarded as low-quality base based on phred+33. If more than half were low-quality bases in a sequence, this sequence, along with the paired one was discarded. The reads belonging to the mitochondrial genome of *Sinocyclocheilus cyphotergous* were identified by alignment with the other nine released mitogenomes of congeneric species within *Sinocyclocheilus* (Accession number: MG026730.1; MZ781221.1; NC_058003.1; MT361975.1; KX528071.1; MK387704.1; NC_057312.1; NC_056194.1; NC_056143.1), which was performed by minimap2 version 2.24. Then, the reads within the alignments were extracted and assembled to a complete circle by ABYSS version 2.1.5 with a k-mer length of 27 bp. The assembled mitogenome was annotated using the MitoAnnotator on the homepage (Iwasaki *et al*., 2013).

Mitogenome characteristic analyses

All putative tRNA genes were identified using tRNAscan-SE search server (Lowe and Chan, 2016) and MITOS (Bernt *et al*., 2013). The secondary structures of tRNAs were drawn by MITOS. The base composition, amino acid composition, codon usage and relative synonymous codon usage (RSCU) of 13 protein coding genes (PCGs) were analyzed using MEGA version 7.0 (Kumar *et al*., 2016). Nucleotide compositional skew was

calculated using the formula: AT-skew = (A − T) / (A + T) and GC-skew = (G − C) / (G + C) (Perna and Kocher, 1995).

Table I. The information of the mitochondrial genomes used in this study.

No.	Taxonomic position/ Species	Size (bp)	Accession no.
Cypriniformes, Cyprinidae			
1	*Sinocyclocheilus punctatus*	16582	NC_058003
2	*S. ronganensis*	16587	NC_032385
3	*S. grahami*	16585	NC_013189
4	*S. microphthalmus*	16589	MN145877
5	*S. yishanensis*	16573	MK387704
6	*S. huizeensis*	16585	NC_044072
7	*S. qujingensis*	16588	NC_043910
8	*S. wumengshanensis*	16585	NC_039769
9	*S. tingi*	16584	NC_039594
10	*S. oxycephalus*	16585	NC_037858
11	*S. jii*	16577	NC_037197
12	*S. multipunctatus*	16586	MG026730
13	*S. bicornutus*	17426	NC_031382
14	*S. anophthalmus*	16574	NC_023472
15	*S. furcodorsalis*	16581	NC_019995
16	*S. altishoulderus*	16589	NC_013186
17	*S. longibarbatus*	16787	MW345239
18	*S. lingyunensis*	16572	NC_056143
19	*S. angularis*	16586	MW362289
20	*S. anshuiensis*	16618	NC_027169
21	*S. rhinocerous*	16588	NC_027168
22	*S. cyphotergous*	16611	this study
23	*Cyprinus carpio*	16581	KF856965
24	*Barbus barbus*	16600	NC_008654

Phylogenetic analysis

To investigate the phylogenetic relationships between *S. cyphotergous* and other species of *Sinocyclocheilus* fishes, the phylogenetic analyses were conducted based on the mitochondrial genome sequence of *S. cyphotergous* assembled in this study and other 21 *Sinocyclocheilus* fishes that were downloaded from NCBI GenBank (Table I). We used two species of *Cyprinus carpio* and *Barbus barbus* as outgroups. Therefore, the complete datasets used for phylogenetic analysis included 24 complete mitochondrial genomes (Table I).

All sequences were aligned with MAFFT version 5.0 (Katoh *et al*., 2005), and the phylogenetic tree was inferred by IQ-TREE version 1.6.12 (Lam-Tung *et al*., 2015) with the "-spp" option to allow partition-specific evolution rates. Twenty-four mitochondrial genomes were partitioned to six partitions, and the best model for each partition was detected with IQ-tree Model Finder (Kalyaanamoorthy *et al*., 2017) based on Bayesian information criterion. Maximum likelihood (ML) tree and Bayesian inference (BI) was performed in IQ-tree using fixed models as defined by the best-fit partitioning schemes. Node support of the trees were inferred by conducting non-parametric bootstrap (1000 replicates) and ultrafast bootstrap (5000 replicates) (Hoang *et al*., 2018), respectively. Trees were graphically visualized with FigTree version 1.4.2.

Substitution rate estimation and comparison

The rates of nonsynonymous (*Ka*) and synonymous (*Ks*) substitutions may differ during the evolution process among mitochondrial coding genes. To investigate the evolutionary patterns under different selective pressures among 13 PCGs in *Sinocyclocheilus* fishes, we further calculated the average *Ka* and *Ka*/ *Ks* values among all 13 PCGs of 21 *Sinocyclocheilus* fishes.

Based on the annotations of the 21 *Sinocyclocheilus* mitochondrial genomes, each of the 13 protein coding genes were extracted, and each protein coding gene was aligned separately with the codon-based model in the Muscle module of Mega version 7 (Kumar *et al*., 2016). The pairwise *Ka* and *Ka*/ *Ks* values between each pair of single-gene datasets were calculated in DnaSP version 5.0 (Librado and Rozas, 2009). The average values of *Ka* and *Ka*/ *Ks* were calculated in R, and were used to represent changes of each coding gene's substitution rates.

RESULTS AND DISCUSSION

Mitogenome features and nucleotide composition of S. cyphotergous

A total of 27,099,642,600 raw reads were generated and it has been deposited to NCBI database (see additional details in Data availability statement). After assembly, the complete mitogenome of *S. cyphotergous* was obtained (accession number: OQ319607), with a total length of 16,611 bp. The mitogenome of *S. cyphotergous* also consists of 13 protein-coding genes (*ND1*, *ND2*, *COI*, *COII*, *ATP8*, *ATP6*, *COIII*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*, and *Cyt b*), 22 transfer RNA (tRNA) genes, 2 ribosomal RNA genes, and one non-coding control region (D-loop region) (Fig. 1, Table II). The arrangement of these genes was identical to other previously published *Sinocyclocheilus* mitogenomes (Li and Yang, 2021; Luo and Zhang, 2021; Zhang and Wang, 2018). Most genes were transcribed from the heavy strand (2 rRNAs, 12 protein-coding genes and 14 tRNAs);

only nine genes, including one protein coding gene (*ND6*) and eight tRNAs (tRNA-Gln, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Ser (UCN), tRNA-Glu and tRNA-Pro), were encoded on the light strand.

There were 6 overlapping regions totaling 22 bp and varied in size from 1 to 7 bp; the longest overlapping region was located between ATP8 and ATP6, ND4L and ND4. Likewise, a total of 68 intergenic nucleotides were dispersed in 13 intergenic spacer regions and ranged in length from 1 to 32 bp; the longest gap was located between tRNA-Asn and tRNA-Cys. The overlapping and intergenic regions of *S. cyphotergous* is similar to the most other fish mitochondrial genomes that without gene rearrangement occurred (He et al., 2012, 2016; Luo and Zhang, 2021; Wu et al., 2009; Zhang and Wang, 2018). The 12S and 16S rRNA genes were 955 and 1679 bp, respectively. They were located between tRNA-Phe and tRNA-Leu (UUR) and were separated by tRNA-Val (Table II). The D-loop region was located between tRNA-Pro and tRNA-Phe.

As is the case with other *Sinocyclocheilus* mitogenomes, the overall base composition of *S. cyphotergous* was 31.33% A, 25.89% T, 26.49% C, and 16.29% G, with a slight AT bias of 57.22%, which was in accordance with most other fish mitogenomes. The A+T contents of PCGs, rRNAs, tRNAs, and D-loop region were all above 50% (Table III). The D-loop region, known as A+T rich region, has the highest A+T content (68.83%), which is typical for animal mitochondrial genomes (Guo et al., 2003; Sbisa et al., 1997; Zhang and Wang, 2018; Zou et al., 2017). The skew statistics showed a similar pattern of base composition in *S. cyphotergous* mitogenome, except for the tRNA, which was also similar to other *Sinocyclocheilus* mitogenomes (Zhang and Wang, 2018).

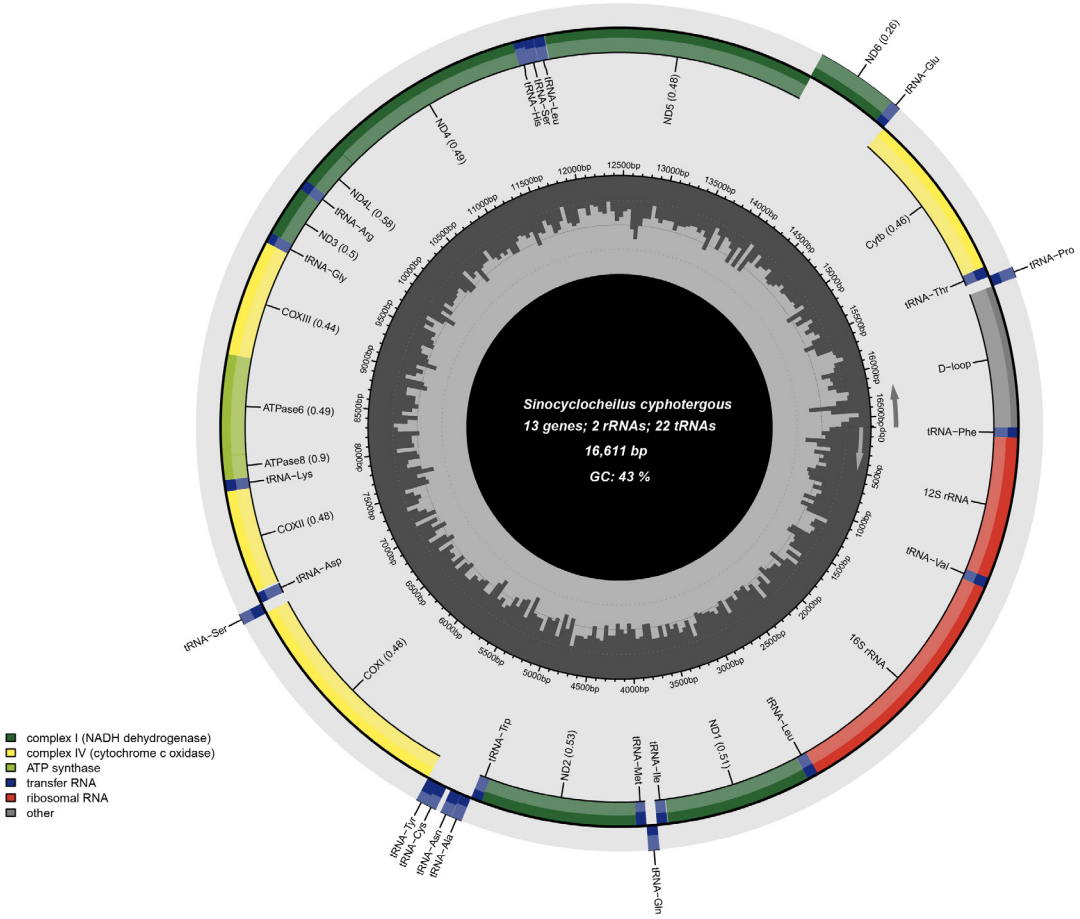

Fig. 1. Circular map of the complete mitochondrial genome of *S. cyphotergous*. Genes encoded on the H-strand and L-strand are shown inside and outside the circular map of the mitogenome. Protein-coding genes of NADH dehydrogenase were colored in green, cytochrome c oxidases were colored in yellow, and ATP synthases were colored in light green. Blue and red colour represented transfer RNA and ribosomal RNA, respectively. Control region of D-loop was colored in gray.

Table II. Mitochondrial genome organization of *Sinocyclocheilus cyphotergous*.

Gene/ Element	Codon					Intergenic nucleotides*	Strand†
	From	To	Length (bp)	Start	Stop		
tRNA-Phe	1	69	69			0	H
12S rRNA	70	1024	955			0	H
tRNA-Val	1025	1096	72			0	H
16S rRNA	1097	2775	1679			0	H
tRNA-Leu (UUR)	2776	2851	76			+1	H
ND1	2853	3827	975	ATG	TAA	+4	H
tRNA- Ile	3832	3903	72			-2	H
RNA- Gln	3902	3972	71			+1	L
tRNA-Met	3974	4042	69			0	H
ND2	4043	5087	1045	ATG	T--	0	H
tRNA-Trp	5088	5160	73			+2	H
tRNA-Ala	5163	5231	69			+1	L
tRNA-Asn	5233	5305	73			+32	L
tRNA-Cys	5338	5404	67			-1	L
Tyr	5404	5474	71			+1	L
COI	5476	7026	1551	GTG	TAA	0	H
tRNA-Ser (UCN)	7027	7097	71			+3	L
tRNA-Asp	7101	7172	72			+13	H
COII	7186	7876	691	ATG	T--	0	H
tRNA-Lys	7877	7952	76			+1	H
ATP8	7954	8118	165	ATG	TAA	-7	H
ATP6	8112	8794	683	ATG	TA-	0	H
COIII	8795	9579	785	ATG	TA-	0	H
tRNA-Gly	9580	9651	72			0	H
ND3	9652	10000	349	ATG	T--	0	H
tRNA-Arg	10001	10070	70			0	H
ND4L	10071	10367	297	ATG	TAA	-7	H
ND4	10361	11741	1381	ATG	T--	0	H
tRNA-His	11742	11811	70			0	H
tRNA-Ser (AGY)	11812	11880	69			+1	H
RNA-Leu (CUN)	11882	11954	73			+3	H
ND5	11958	13781	1824	ATG	TAA	-4	H
ND6	13778	14299	522	ATG	TAA	0	L
tRNA-Glu	14300	14368	69			+5	L
Cyt b	14374	15514	1141	ATG	T--	0	H
RNA-Thr	15515	15586	72			-1	H
tRNA-Pro	15586	15655	70			0	L
D-loop	15656	16611	956			0	H

*Numbers correspond to the nucleotides separating different genes. Negative numbers indicate overlapping nucleotides between adjacent genes.

Protein coding genes and codon usage patterns

Almost all PCGs started with the typical ATG initiation codons whereas *COI* started with GTG. Among these 13 PCGs, six PCGs including *ND1*, *COI*, *ATP8*, *ND4L*, *ND5* and *ND6* were terminated with the typical TAA codons, while other seven PCGs (*ND2*, *COII*, *ATP6*, *COIII*, *ND3*, *ND4* and *Cyt b*) were characterized by incomplete stop codons (T or TA) (Table II). The base compositions of the 13 PCGs were 28.10% for T, 27.01% for C, 29.14% for A and 15.75% for G (Table III).

Excluding the termination codons, a total of 3803 amino acids from 13 PCGs were encoded in the mitogenome of *S. cyphotergous*. The relative synonymous codon usage (RSCU) of the 13 PCGs in *S. cyphotergous* was shown in Figure 2. The most frequently used codon in *S. cyphotergous* was GCC-Ala (1.69%), followed by AAA-Lys (1.51%) and AGG-Arg (1.48%) (Fig. 2).

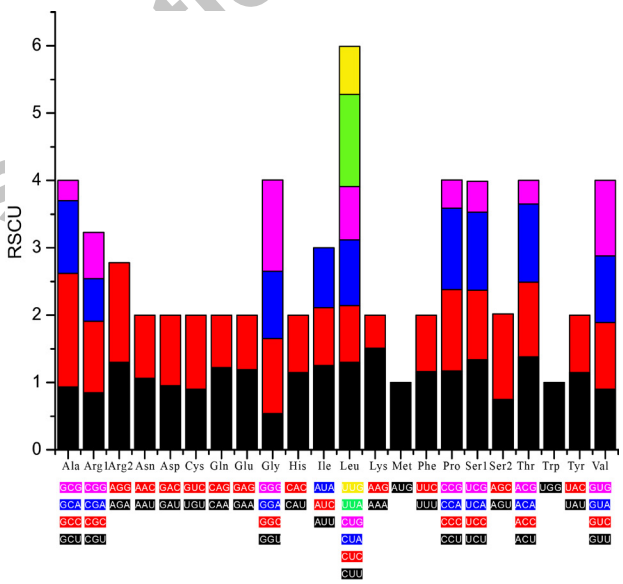

Fig. 2. Relative synonymous codon usage (RSCU) in the *S. cyphotergous* mitogenome. X axis shows the Condon family, and Y axis shows the RSCU values.

Transfer RNAs and ribosomal RNAs

The mitogenome of *S. cyphotergous* was predicted to encode 22 tRNA genes, which were interspersed along the genome, with the length varying from 67 bp (tRNA-Cys) to 76 bp (tRNA-Leu (UUR) and tRNA-Lys) (Table II). All the tRNAs of *S. cyphotergous* were capable of folding into a typical clover-leaf secondary structure except for tRNA-Ser (AGY) with an incomplete dihydrouridine arm (Fig. 3), which is similar to other *Sinocyclocheilus* fishes (Wu et al., 2009; Zhang and Wang, 2018). 15 of 22 tRNA genes, including tRNA-Leu, tRNA-Gln, tRNA-Met, tRNA-Trp,

Table III. Nucleotide composition of the *S. cyphotergous* mitochondrial genome.

	Length (bp)	T%	C%	A%	G%	A+ T%	AT-skew	GC-skew
Genome	16611	25.89	26.49	31.33	16.29	57.22	0.09	-0.24
PCGs	11409	28.10	27.01	29.14	15.75	57.24	0.02	-0.26
1st codon position	3803	28.95	27.77	29.95	13.33	59.00	0.02	-0.35
2nd codon position	3803	27.24	27.03	29.03	16.70	56.27	0.03	-0.24
3rd codon position	3803	28.11	26.24	28.43	17.22	56.54	0.01	-0.21
rRNA	2634	19.70	24.72	34.51	21.07	54.21	0.27	-0.08
tRNA	1566	26.69	21.39	28.87	23.05	55.56	0.04	0.04
D-loop region	956	34.00	18.20	34.83	12.97	68.83	0.01	-0.17

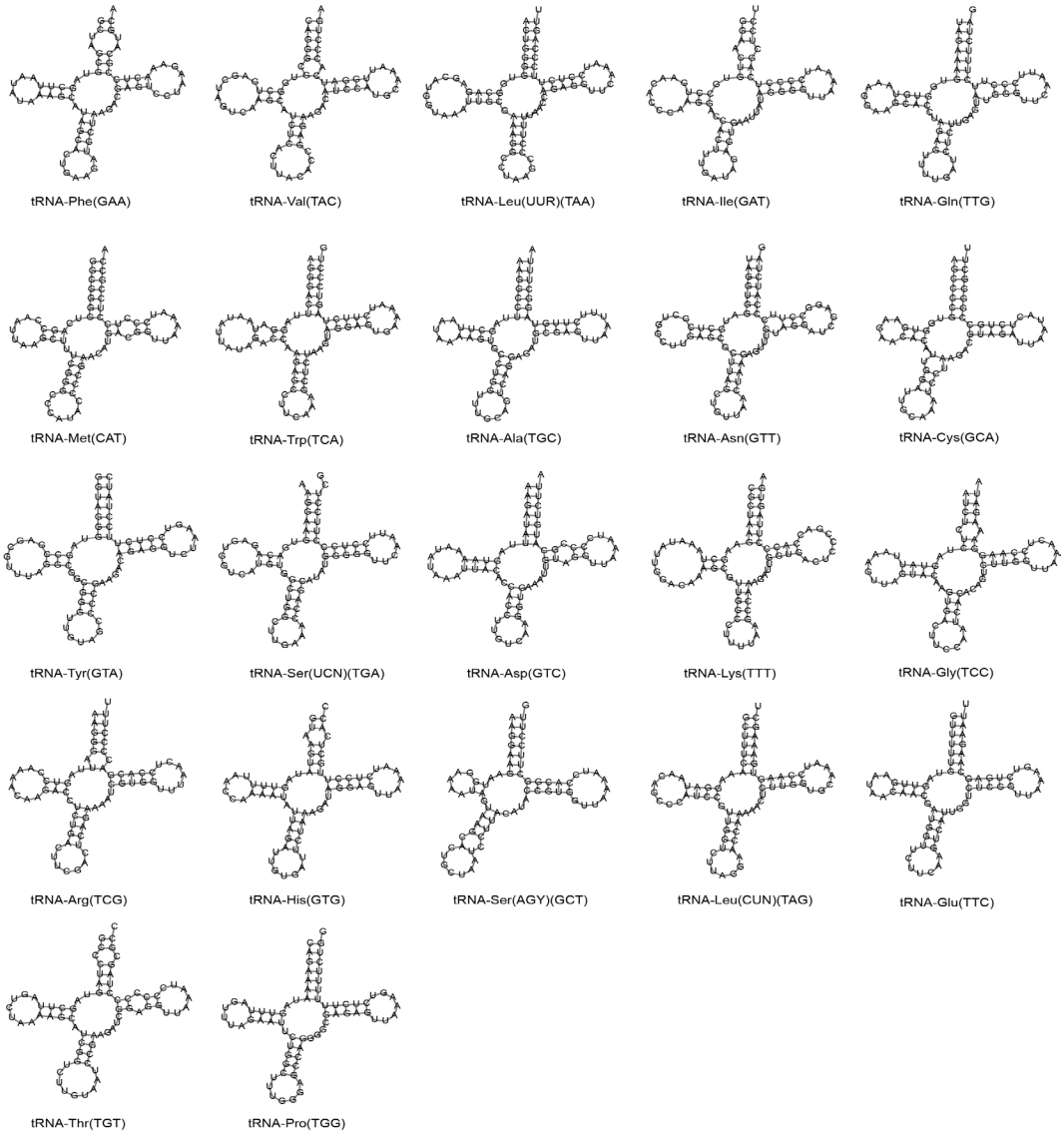

Fig. 3. Putative secondary structures of the 22 tRNA genes identified in the mitochondrial genome of *S. cyphotergous*. The tRNAs are labelled with abbreviations of their corresponding amino acid. The orders based on clockwise from top are amino acid acceptor arm, the TΨC arm, the anticodon arm, and the dihydrouridine arm.

tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Asp, tRNA-Lys, tRNA-Gly, tRNA-His, tRNA-Ser, tRNA-Glu, and tRNA-Pro, were found to have 29 G-U mismatches in their secondary structures, which formed a weak bond. Six tRNA including tRNA-Ile, tRNA-Gly, tRNA-Arg, tRNA-His, tRNA-Asp, and tRNA-Thr, were found to have A-C mismatches. Three C-C mismatches of tRNA-Thr, tRNA-Met, and tRNA-Ser were found in their amino acid acceptor arm, TΨC stem and anticodon arm, respectively. Two U-U mismatches of tRNA-Gln and tRNA-Asn were found in their TΨC stem, and one U-U mismatch of tRNA-Cys was found in the anticodon arm. In addition, an A-A mismatch of tRNA-Trp was found in the dihydrouridine arm (Fig. 3).

As in other fish mitogenome sequences, there were two rRNA genes including 12S rRNA and 16S rRNA in *S. cyphotergous*. The size of 12S rRNA was 955 bp and located between tRNA-Phe and tRNA-Val, and the size of 16S rRNA was 1679 bp and located between tRNA-Val and tRNA-Leu (UUR), respectively. The base composition of the two rRNAs was 19.70% for T, 24.72% for C, 34.51% for A, and 21.07 for G. The A + T content was 54.21%, thus slightly higher than the G + C content (Table III). The length and A + T content of these two rRNAs are similar to the other *Sinocyclocheilus* species mitogenomes (Hao *et al*., 2016; He *et al*., 2012, 2016; Wu *et al*., 2009; Zhang and Wang, 2018).

Phylogenetic analysis within Sinocyclocheilus *fishes*

To date, there are 21 other *Sinocyclocheilus* mitogenomes available on GenBank. Our multiple sequence alignment of 24 complete mitochondrial genomes, including two outgroup species (*Cyprinus carpio* and *Barbus barbu*) and 22 *Sinocyclocheilus* species. The total alignment for phylogenetic analysis was 11,385 bp. Model selection using Partition Finder 2 indicated an optimal partitioning scheme with six partitions, and the first partition was ND4+ATP6+CYTB+ND5 with GTR+F+I+G4 model, the second partition was ND1+ND2 with TIM3+F+I+G4 model, the third partition was ND6 with HKY+F+G4 model, the fourth partition was COIII+COI+COII+ND4L with GTR+F+I+G4 model, the fifth partition was ATP8 with K3Pu+F+I model, and the sixth partition was ND3 with GTR+F+G4 model. The phylogeny trees resulting from BI and ML analyses showed identical topologies, and only slight differences were occurred between the Bayesian posterior probabilities and ML bootstrap values (Fig. 4). Phylogenetic analysis revealed all *Sinocyclocheilus* species formed a solid monophyletic group and grouped into two major clades with strong support. *S. cyphotergous* in this study was clustered together and closely related to *S. multipunctatus* and *S. punctatus*. Additionally, *S. jii* is the sister species to a clade of all other *Sinocyclocheilus* species based on the phylogenetic reconstruction (Fig. 4), which is also supported by the previous research (Li and Yang, 2021). In summary, our study provides a new resource for understanding the whole mitochondrial genome of *S. cyphotergous*, which will promote the molecular study of this genus.

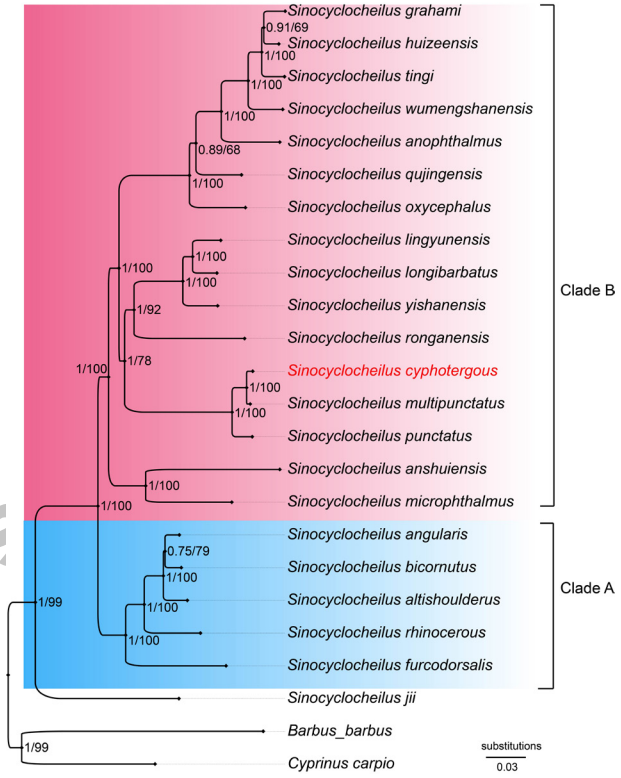

Fig. 4. Phylogenetic relationships of species in genus *Sinocyclocheilus* inferred by Bayesian Inference and Maximum Likelihood analyses, based on the mitochondrial genome. Numbers on the branches from left to right are Bayesian posterior probabilities obtained by BI and ML bootstrap values, respectively.

The nucleotide substitution rate in the Sinocyclocheilus *fishes*

Of all average values of *Ka* and *Ka/Ks* across the 13 PCGs of the *Sinocyclocheilus* fishes, *ATP8* had the largest average *Ka* and *Ka/Ks*, while *COI* had the lowest (Fig. 5), which implies that *ATP8* might evolve more quickly than other mitochondrial protein coding genes. The evolutionary patterns under different selective pressures among 13 PCGs in the *Sinocyclocheilus* fishes was similar to the Bagridae catfishes (Zhang *et al*., 2022) and Glyptosternoid fishes (Lv *et al*., 2018). In addition, the average *Ka/Ks* ratios for all PCGs were far lower than one, which indicates that they were all under strong purifying selection.

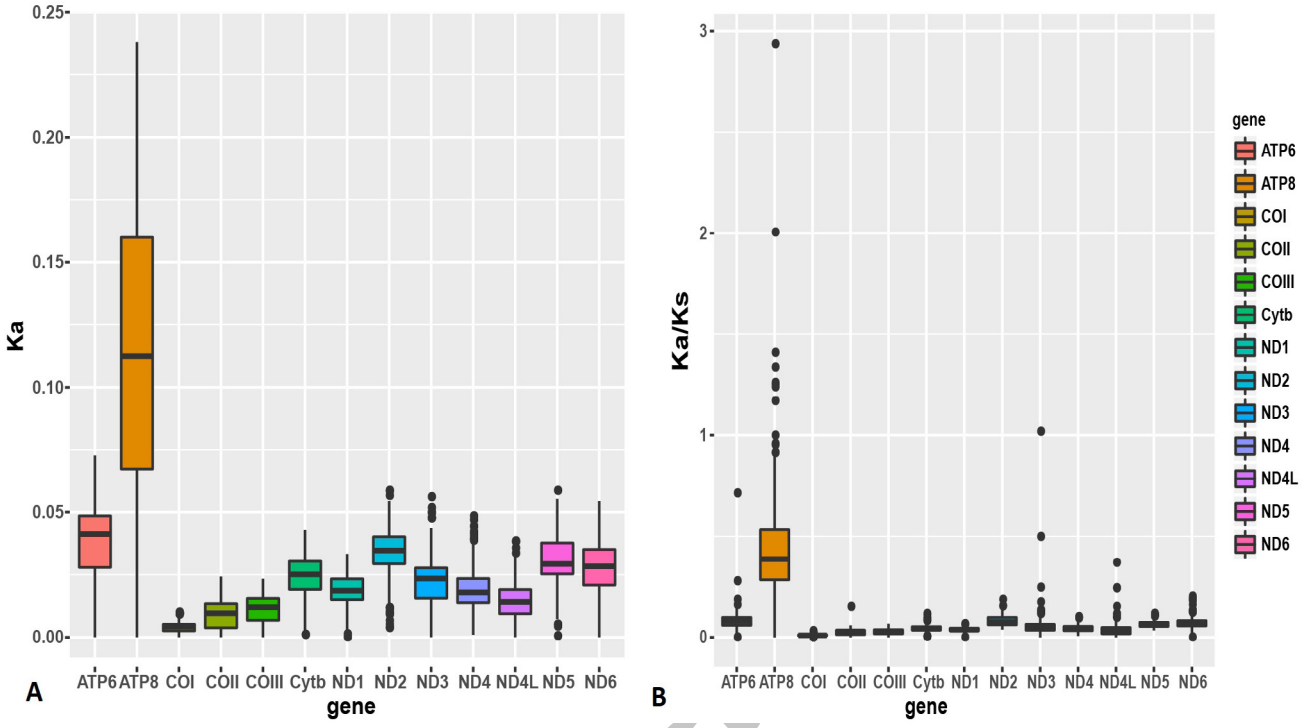

Fig. 5. Boxplots of (a) *Ka* and (b) *Ka/Ks* for the 13 mitochondrial protein-coding genes of the 21 *Sinocyclocheilus* mitogenomes examined. The average *Ka* and *Ka/Ks* were greatest in *ATP8*.

CONCLUSION

In the present study, we first collected a rare cavefish *S. cyphotergous*, and then determined the complete mitochondrial genome using Illumina sequencing. The mitogenome was 16,611 bp in length, which contained 37 genes and one control region, as is typical of teleost mitogenomes. All tRNAs could fold into the typical cloverleaf secondary structures, except for tRNA-Ser (AGY) that lacked the dihydrouracil arm. The ratio of non-synonymous and synonymous substitutions (Ka/Ks) of all the 13 PCGs were less than 1, indicating negative or purifying selection evolved in these genes. The evolutionary rate of *ATP8* was the fastest and *COI* was the slowest. The reconstructed phylogenetic tree based on the 13 PCGs from the mitochondrial genome of 24 species supported that *S. cyphotergous* in this study was closely related to *S. multipunctatus* and *S. punctatus*, and all *Sinocyclocheilus* species in this research formed a solid monophyletic group and grouped into two major clades with strong support excluded *S. jii*. Finally, this study provides novel insights into the phylogeny of the *Sinocyclocheilus* fishes, and offers genetic basis for the conservation of *S. cyphotergous*.

ACKNOWLEDGMENTS

We thank Dr. Yunhai Yi for her valuable and constructive suggestions in revising the manuscript. We also thank three anonymous reviewers for their valuable comments that improved the quality of this article.

Funding

This work was supported by the Opening Foundation of the Key Laboratory of Sichuan Province for Fishes Conservation and Utilization in the Upper Reaches of the Yangtze River (NJTCCJSYSYS14), and the Neijiang Normal University Key Program Fund (2021ZD09).

IRB approval

This study was reviewed and approved by the IRB of Neijiang Normal University for being conducted in accordance with ethical guidelines for animal research.

Ethical statement

The experimental samples used for this study were collected and processed in accordance with the approval of Animal Experiment Ethics Committee in Neijiang Normal University.

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/) under the accession no. OQ319607. The associated BioProject, SRA, and BioSample numbers are PRJNA916069, SRR22894312, and SAMN32411344, respectively.

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. and Stadler, P., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.*, **69**: 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Boore, J.L., 1999. Animal mitochondrial genomes. *Nucl. Acids Res.*, **27**: 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Guo, X., Liu, S. and Liu, Y., 2003. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. *Aquaculture*, **224**: 25–38. https://doi.org/10.1016/S0044-8486(03)00168-6

Hao, Z., Zhang, Q. and Qu, B., 2016. Complete mitochondrial genome of a cavefish, *Sinocyclocheilus anophthalmus* (Cypriniformes: Cyprinidae). *Mitochondrial DNA A*, **27**: 84. https://doi.org/10.3109/19401736.2013.873913

He, S., Liang, X.F., Chu, W.Y. and Chen, D.X., 2012. Complete mitochondrial genome of the blind cave barbel *Sinocyclocheilus furcodorsalis* (Cypriniformes: Cyprinidae). *Mitochondrial DNA*, **23**: 429. https://doi.org/10.3109/19401736.2012.710216

He, S., Lu, J., Jiang, W., Yang, S., Yang, J. and Shi, Q., 2016. The complete mitochondrial genome sequence of a cavefish *Sinocyclocheilus anshuiensis* (Cypriniformes: Cyprinidae). *Mitochondrial DNA A*, **27**: 4256–4258. https://doi.org/10.3109/19401736.2015.1046127

Hoang, D.T., Chernomor, O., Haeseler, A.V., Minh, B.Q. and Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, **35**: 518–522. https://doi.org/10.1093/molbev/msx281

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H. and Miya, M., 2013. MitoFish and MitoAnnotator: 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

Kalyaanamoorthy, S., Minh, B.Q., Wong, T., Haeseler, A.V. and Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, **14**: 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Kuma, K.I., Toh, H. and Miyata, T., 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucl. Acids Res.*, **33**: 511–518. https://doi.org/10.1093/nar/gki198

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, **33**: 1870–1874. https://doi.org/10.1093/molbev/msw054

Lam-Tung, N., Schmidt, H.A., Arndt, V.H. and Quang, M.B., 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

Li, P. and Yang, J., 2021. Characterization of the complete mitochondrial genome of *Sinocyclocheilus lingyunensis* (Cypriniformes: Cyprinidae). *Mitochondrial DNA B*, **6**: 1759–1760. https://doi.org/10.1080/23802359.2021.1914222

Li, Z. and He, S., 2009. Relaxed purifying selection of rhodopsin gene within a Chinese endemic cavefish genus *Sinocyclocheilus* (Pisces: Cypriniformes). *Hydrobiologia*, **624**: 139–149. https://doi.org/10.1007/s10750-008-9688-2

Liang, X., Cao, L. and Zhang, C., 2011. Molecular phylogeny of the *Sinocyclocheilus* (Cypriniformes: Cyprinidae) fishes in northwest part of Guangxi, China. *Environ. Biol. Fish.*, **92**: 371–379. https://doi.org/10.1007/s10641-011-9847-6

Librado and Rozas, 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**: 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lowe, T.M. and Chan, P.P., 2016. tRNAscan-SE Online: Integrating search and context for analysis of transfer RNA genes. *Nucl. Acids Res.*, **44**: 54–57. https://doi.org/10.1093/nar/gkw413

Luo, Q. and Zhang, R., 2021. The complete mitochondrial genome and phylogenetic analysis of *Sinocyclocheilus angularis* (Cypriniformes: Cyprinidae). *Mitochondrial DNA B*, **6**: 3438–3439. https://doi.org/10.1080/23802359.2021.1920862

Lv, Y., Li, Y., Ruan, Z., Bian, C., You, X., Yang, J., Jiang, W. and Shi, Q., 2018. The complete mitochondrial

genome of *Glyptothorax macromaculatus* provides a well-resolved molecular phylogeny of the Chinese sisorid catfishes. *Genes*, **9**: 282. https://doi.org/10.3390/genes9060282

Perna, N.T. and Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. mol. Evol.*, **41**: 353–358. https://doi.org/10.1007/BF01215182

Sbisa, E., Tanzariello, F., Reyes, A., Pesole, G. and Saccone, C., 1997. Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. *Gene*, **205**: 125–140. https://doi.org/10.1016/S0378-1119(97)00404-6

Wu, X., Wang, L., Chen, S., Zan, R., Xiao, H. and Zhang, Y., 2009. The complete mitochondrial genomes of two species from *Sinocyclocheilus* (Cypriniformes: Cyprinidae) and a phylogenetic analysis within Cyprininae. *Mol. Biol. Rep.*, **37**: 2163–2171. https://doi.org/10.1007/s11033-009-9689-x

Xiao, H., Chen, S.Y., Liu, Z.M., Zhang, R.D., Li, W.X., Zan, R.G. and Zhang, Y.P., 2005. Molecular phylogeny of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.*, **36**: 67–77. https://doi.org/10.1016/j.ympev.2004.12.007

Zhang, R. and Wang, X., 2018. Characterization and phylogenetic analysis of the complete mitogenome of a rare cavefish, *Sinocyclocheilus multipunctatus* (Cypriniformes: Cyprinidae). *Genes Genom.*, **40**: 1034–1040. https://doi.org/10.1007/s13258-018-0711-3

Zhang, R., Deng, L., Lv, X. and Tang, Q., 2022. Complete mitochondrial genomes of two catfishes (Siluriformes, Bagridae) and their phylogenetic implications. *ZooKeys*, **1115**. https://doi.org/10.3897/zookeys.1115.85249

Zhao, Y. and Zhang, C., 2009. Endemic fishes of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) in China-species diversity, cave adaptation, systematics and zoogeography. Science Press, Beijing, China. (in Chinese).

Zou, Y., Xie, B., Qin, C., Wang, Y., Yuan, D., Li, R. and Wen, Z., 2017. The complete mitochondrial genome of a threatened loach (*Sinibotia reevesae*) and its phylogeny. *Genes Genom.*, **39**: 767–778. https://doi.org/10.1007/s13258-017-0541-8