Pakistan J. Zool., vol. 56(3), pp 1389-1401, 2024 **DOI**: https://dx.doi.org/10.17582/journal.pjz/20230330030335

Complete Mitochondrial Genome Sequence of Ruffe *Acerina cernua* (Perciformes, Percidae): Mitogenomic Characteristics and Phylogenetic Implications

Yuping Liu, Sige Wang and Tianyan Yang*

Fishery College, Zhejiang Ocean University, Zhoushan, Zhejiang, China, 316022

Article Information
Received 30 March 2023
Revised 25 April 2023
Accepted 16 May 2023
Available online 21 July 2023
(early access)
Published 02 May 2024

Authors' Contribution
YL and TY presented the concept of the study and wrote the manuscript. TY, YL and SW designed research, conducted experiments, explained research results, and wrote and revised manuscripts. TY helped write and revise the manuscript.

Key words
***Acerina cernua*, Mitogenome, Sequence analysis, Phylogeny**

ABSTRACT

Acerina cernua is mainly distributed in Europe and Asia, but has recently invaded in North America and new areas of Europe. In China, it is native to the Irtysh River and Ulungur Lake, and considered to be an indigenous fish endemic to Xinjiang. In this paper, the complete mitochondrial genome (16607 bp) of *A.cernua* was sequenced by high-throughput sequencing technology, including 13 protein-coding genes (PCGs), 22 tRNA genes, 2 rRNA genes, a control region (CR) and a L-strand replication origin (O_L), and its gene composition and order were similar to those of most teleosts. The A + T content of the whole mtDNA was 55.39 %, suggesting an obvious anti-G bias (16.64 %). The positive AT-skew (0.01) and negative GC-skew (-0.25) were revealed. The analysis of codon usage showed that NNA-type codons were used most frequently, which was consistent with the A bias at the third codon positions in PCGs. Three base mismatches were detected in the secondary structures of tRNAs, namely A-C, U-U and A-A, which mainly occurred in the amino acid receptor arm and TψC arm. The CR contained three different domains: extended termination-associated sequences, central conserved region (CSB-F, CSB-E and CSB-D) and conserved sequence region (CSB1, CSB2 and CSB3). Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses were performed based on 12 PCGs. The similar topological structures confirmed that the relationship between *Acerina* and *Sander* was the closest.

INTRODUCTION

Mitochondria (mt) are important functional organelles in eukaryotic cells and mitochondrial DNA (mtDNA) is a small closed circular molecule, which is independent of the nuclear genome with a length of 15-20 kb. It normally contains 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, a L-strand replication origin (O_L), as well as a large non-coding region, namely control region (CR) or D-loop region (Boore, 1999). Because of the advantages of maternal inheritance, simple structure, independent replication, high mutation rate and relatively stable probability of variation, mtDNA has been widely applied in genetic structure, phylogeny and phylogeography at various taxonomic levels, and has become a powerful tool to solve both the interspecific genetic differentiation and intraspecific genetic variation (Inoue *et al*., 2001; Kawahara *et al*., 2008; Perea *et al*., 2010; Imoto *et al*., 2013). In recent years, the advancement of DNA sequencing technology has provided great convenience for the rapid and accurate acquisition of fish mitogenome data. More and more complete mitochondrial DNA sequences of teleosts have been reported, which intensively accelerate the taxonomic and phylogenetic researches of bony fishes.

The ruffe *Acerina cernua* (Linnaeus, 1758), also known as *Gymnocephalus cernua* (Linnaeus, 1758), belongs to the family Percidae and order Perciformes. This small spiny freshwater fish is originated in the Northern Europe and northwest Asia, but inadvertently introduced into North America and Western Europe in the 1980s. Nowadays, it has expanded its natural distribution areas and widely spreads in the Great Lakes and major rivers in Europe as an invasive species (Guo, 2012; Kottelat and Freyhof, 2007). In China, it is an aboriginal fish that naturally occurs in Irtysh River and Ulungur Lake of Xinjiang (Zhu *et al*., 2014). Enormous amounts of researches centering on physiology, ecology, biology and population dynamics

* Corresponding author: hellojelly1130@163.com
0030-9923/2024/0003-1389 $ 9.00/0

of this fish have been carried out until now (Van and Van, 1994; Zhu *et al*., 2014). However, the investigations into its genetic background are very limited, which seriously hinder the rational utilization of the germplasm resources of exotic species (Nesbø *et al*., 1998; Zhang *et al*., 2017; Li *et al*., 2019). In this study, high-throughput sequencing (HTS) technology was used to obtain the complete mtDNA sequence of *A. cernua* collected from Irtysh River in Xinjiang, China. By comparing with the mitogenomes of 15 species of Percidae, the molecular phylogenetic relationship of Percidae was further discussed. The results of this study are expected to provide reference information for germplasm identification and genetic diversity evaluation of *A. cernua*.

MATERIALS AND METHODS

Experimental sample and genomic DNA extraction

In this study, the samples of *A. cernua* were collected from Irtysh River in Xinjiang, China in September 2017. About 2g fresh muscle tissue from the dorsal fin base was cut into 1.5 mL EP centrifuge tubes, immersed in 95% alcohol for 48 hours, and then frozen at −20 °C. Genomic DNA was extracted by traditional chloroform-Tris saturated phenol method (Maniatis *et al*., 1985). The concentration of genomic DNA was determined by NanoDrop ND-1000 ultramicro spectrophotometer, and DNA integrity was detected by 1 % agarose gel electrophoresis.

High-throughput sequencing and gene annotation

The qualified DNA samples were randomly broken into fragments with the length of about 300 bp by Covaris ultrasonic disruptor to construct a small fragment genomic DNA library of *A. cernua*. Sequencing was commissioned to Hengchuang Gene Technology Co., Ltd (Shenzhen, China) based on Illumina Hiseq 2500 high-throughput sequencing platform. The measured clean reads were assembled by SOAPdenovo 2.04 software (http://soap.genomics.org.cn/soapdenovo.html) (Li *et al*., 2010), and the local assembly results were optimized based on the paired-end and overlap relationships of reads. GapCloser 1.12 software (http://soap.genomics.org.cn/soapdenovo.html) (Zhao *et al*., 2011) was used to make up and repair the gaps introduced in the process of splicing scaffold. Ultimately, the redundant segment sequences were removed to obtain the final assembly result. The mitogenome was extracted and assembled using NOVO Plasty software (Dierckxsens *et al*., 2017). The complete sequence obtained by splicing was uploaded to the MITOS web server (http://mitofish.aori.u-tokyo.ac.jp/) for annotation of protein-coding genes (PCGs), RNAs and non-coding regions, and then supplemented by manual alignment correction to finally determine the location and length of each gene.

Data analysis

The secondary structure of transfer RNA (tRNA) gene was predicted by online software tRNAscan-SE (http://trna.ucsc.edu/tRNAscan-SE/) (Lowe and Chan, 2016). Ribosomal RNA (rRNA) secondary structure was mapped using the Visusalise RNA 2D Structure (R2DT) tool on the RNAcentral website (https://rnacentral.org/) (The RNA central Consortium, 2019), and the free energy of different rRNA gene was calculated by RNAfold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). The nucleotide composition and codon usage were analyzed by MEGA-X software (Kumar *et al*., 2008). The AT-skew and GC-skew values were calculated and analyzed by Microsoft Excel software. The complete mitogenome sequences of 15 Percidae species were downloaded from GenBank database (https://www.ncbi.nlm.nih.gov/) (Table I). Using *Lateolabrax maculatus* (GenBank accession number: KT852999) and *Micropterus salmoides* (GenBank accession number: DQ536425) as outgroups, the phylogenetic tree was constructed based on concatenated sequences of 12 PCGs (except *ND6* gene encoded by L-strand) after removing stop codons by using maximum likelihood (ML) (Guindon and Gascuel, 2003) and Bayesian inference (BI) (Huelsenbeck and Ronquist, 2001) methods, respectively.

Table I. List of 15 Percidae species and two outgroups used in this paper.

Species	A+T (%)	Length (bp)	GenBank accession No.
Sander vitreus	55.23	16858	KP013098
Sander lucioperca	55.63	16542	KP125333
Percina macrolepida	53.25	16602	DQ536430
Percina kusha	53.65	16626	OP238461
Percina freemanorum	53.64	16578	OP326604
Percina copelandi	53.13	16607	MW856850
Percina brevicauda	53.02	16608	MK778456
Perca schrenkii	54.93	16536	KR905930
Perca fluviatilis	55.10	16537	KM410088
Perca flavescens	55.38	16537	JX629442
Gymnocephalus cernua	55.45	16614	KM978956
Etheostoma spectabile	52.42	16539	MK243404
Etheostoma jessiae	52.20	16600	KY965072
Etheostoma chuckwachatte	54.00	16603	KY965071
Etheostoma caeruleum	52.90	16588	KY660678
Acerina cernua	55.39	16607	This study
Micropterus salmoides (outgroup)	53.48	16484	DQ536425
Lateolabrax maculatus (outgroup)	53.13	16723	KT852999

RESULTS AND DISCUSSION

General features of the mt genome of A. cernua

The complete mitogenome of *A. cernua* was 16607 bp in length, which was in the length range (16484-16858 bp) of the mitogenome of 15 reported Percidae fish (Table I). The mtDNA structure of *A. cernua* was highly conserved and conforms to the typical mitogenome structure of vertebrates (Miya *et al*., 2003), which contained 13 PCGs, 22 tRNA genes, 2 rRNA genes, and 2 major non-coding regions (Fig. 1, Table II). Among them, *ND6* gene and 8 tRNAs (tRNA-Gln, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Ser^{UGA}, tRNA-Glu, and tRNA-Pro) were encoded on the light-strand (L-strand), and the remainders were encoded by the heavy-strand (H-strand). Two non-coding regions included a light-strand replication origin (O_L) (38 bp) between tRNA-Asn and tRNA-Cys, and a control region (CR) (944 bp) between tRNA-Pro and tRNA-Phe, respectively.

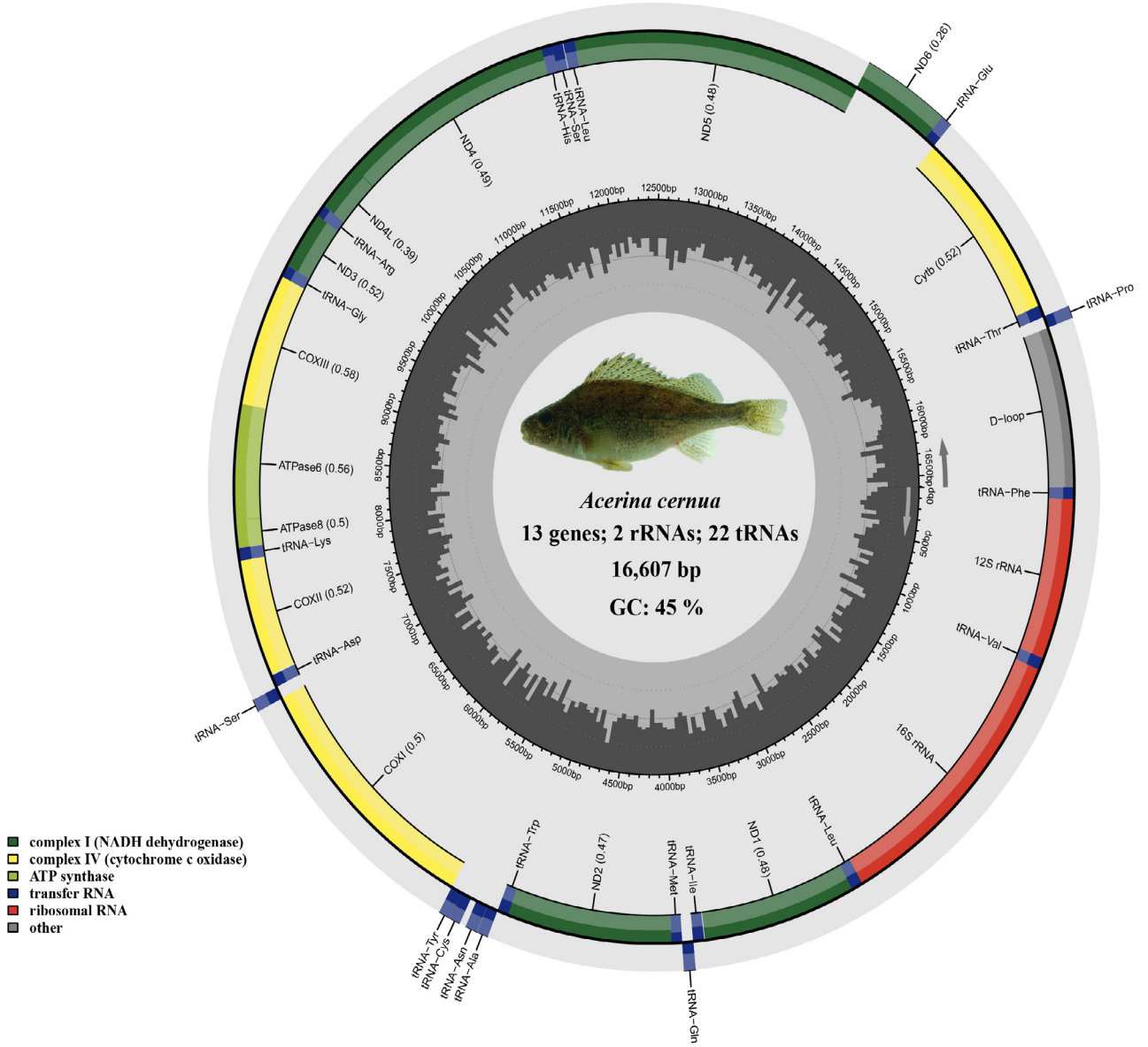

Fig. 1. Gene map of the *A. cernua* mitogenome.

Table II. Features of the mitogenome of *A. cernua.*

Gene	Coding strand	Start	End	Length/ bp	Space or overlap/bp	Amino acid (aa)	Star codon	Stop codon	Antico-don	Anticodon site
tRNA-Phe (F)	H	1	68	68	0				GAA	31-33
12S rRNA	H	69	1017	949	0					
tRNA-Val (V)	H	1018	1089	72	0				TAC	34-36
16S rRNA	H	1090	2781	1692	0					
tRNA-Leu^{UAA} (L1)	H	2782	2855	74	0				TAA	36-38
ND1	H	2856	3830	975	5	324	ATG	TAA		
tRNA-Ile (I)	H	3836	3905	70	-1				GAT	31-33
tRNA-Gln (Q)	L	3905	3975	71	-1				TTG	33-35
tRNA-Met (M)	H	3975	4043	69	0				CAT	31-33
ND2	H	4044	5089	1046	0	348	ATG	TA		
tRNA-Trp (W)	H	5090	5160	71	1				TCA	33-35
tRNA-Ala (A)	L	5162	5230	69	1				TGC	31-33
tRNA-Asn (N)	L	5232	5304	73	0				GTT	34-36
O_L	L	5305	5342	38	0					
tRNA-Cys (C)	L	5343	5408	66	0				GCA	29-31
tRNA-Tyr (Y)	L	5409	5479	71	1				GTA	33-35
COI	H	5481	7031	1551	0	516	GTG	TAA		
tRNA-Ser^{UGA} (S1)	L	7032	7102	71	3				TGA	33-35
tRNA-Asp (D)	H	7106	7177	72	6				GTC	34-36
COII	H	7184	7874	691	0	230	ATG	T		
tRNA-Lys (K)	H	7875	7948	74	1				TTT	35-37
ATP8	H	7950	8117	168	-10	55	ATG	TAA		
ATP6	H	8108	8790	683	0	227	ATG	TA		
COIII	H	8791	9575	785	0	261	ATG	TA		
tRNA-Gly (G)	H	9576	9646	71	0				TCC	33-35
ND3	H	9647	9995	349	0	116	ATG	T		
tRNA-Arg (R)	H	9996	10064	69	0				TCG	31-33
ND4L	H	10065	10361	297	-7	98	ATG	TAA		
ND4	H	10355	11735	1381	0	460	ATG	T		
tRNA-His (H)	H	11736	11804	69	0				GTG	31-33
tRNA-Ser^{GCU} (S2)	H	11805	11872	68	5				GCT	28-30
tRNA-Leu^{UAG} (L2)	H	11878	11950	73	0				TAG	34-36
ND5	H	11951	13789	1839	-4	612	ATG	TAA		
ND6	L	13786	14307	522	0	173	ATG	TAG		
tRNA-Glu (E)	L	14308	14376	69	5				TTC	31-33
Cyt b	H	14382	15522	1141	0	380	ATG	T		
tRNA-Thr (T)	H	15523	15594	72	-1				TGT	33-35
tRNA-Pro (P)	L	15594	15663	70	0				TGG	32-34
D-loop	H	15664	16607	944	0					

The mitogenome of *A. cernua* had 9 internal spacer regions and 6 overlapping regions (Table II). The total length of the former was 28 bp, accounting for 0.169 % of the whole mitogenome, and the interval spacer region between tRNA-Asp and *COII* was the largest (6 bp). The total length of the latter was 24 bp, with the largest overlapping region detected between *ATP8* and *ATP6* (10 bp). The overall base composition was listed as: 27.83 % A, 27.98 % C, 27.55 % T, 16.64 % G (Table III). The percentages of A, T and C were basically not much different from those of published Percidae fishes, but the content of guanine was the least (Zhang *et al*., 2017; Li *et al*., 2019). The mitogenome of *A. cernua* owned a higher proportion of A+T (55.39 %), showing the obvious AT bias. This phenomenon also existed in other Percidae fishes, but the content varied slightly from species to species. The reason for this base preference might be related to natural mutations and selection pressures during the process of replication and transcription (Zhong *et al*., 2002).

Protein-coding genes (PCGs) and the codon usage bias

Base composition analysis indicated that the total sequence length of 13 PCGs was 11428 bp, occupying 68.81 % of the whole mitogenome, apparently showing an AT bias (55.11 %). The number of encoded amino acid residues amounted to 3800. Among all the PCGs, *ND5* had the longest base sequence (1839 bp), encoding 612 amino acids. While, *ATP8* was the shortest (168 bp), encoding only 55 amino acids. Excepting for *COI* starting with GTG, the other genes all used ATG as the initiation codon. The incomplete codon phenomenon was common in metazoan mitogenomes. In the present study, five genes (*ND1*, *COI*, *ATP8*, *ND4L* and *ND5*) ended with TAA, *ND6* was terminated by TAG, and the remaining genes were incomplete codons (TA- in *ND2*, *ATP6* and *COIII*; T-- in *COII*, *ND3*, *ND4* and Cyt *b*) (Table II). Metazoan mt genomes usually present a clear strand bias in nucleotide composition, which can be measured as AT- and GC-skews (Perna and Kocher, 1995; Hassanin *et al*., 2005). The AT-skew and GC-skew values are important yardsticks

Table III. Composition and skewness of *A. cernua* mitogenome.

Gene name/ Codon site	Length/ bp	Base composition/ %						AT-skew	GC-skew
		T	C	A	G	A+T	C+G		
ND1	975	31.49	28.00	24.92	15.59	56.41	43.59	-0.12	-0.28
ND2	1046	27.72	33.56	26.20	12.52	53.92	46.08	-0.03	-0.46
COI	1551	30.75	27.27	23.40	18.57	54.16	45.84	-0.14	-0.19
COII	691	29.23	26.34	27.79	16.64	57.02	42.98	-0.03	-0.23
ATP8	168	25.00	33.33	30.95	10.71	55.95	44.05	0.11	-0.51
ATP6	683	30.60	32.50	23.13	13.76	53.73	46.27	-0.14	-0.41
COIII	785	29.04	29.04	25.99	15.92	55.03	44.97	-0.06	-0.29
ND3	349	32.38	31.23	19.77	16.62	52.15	47.85	-0.24	-0.31
ND4L	297	28.62	32.66	22.56	16.16	51.18	48.82	-0.12	-0.34
ND4	1381	28.02	30.56	26.57	14.84	54.60	45.40	-0.03	-0.35
ND5	1839	28.87	28.98	28.93	13.21	57.80	42.20	0.00	-0.37
ND6	522	37.93	14.37	16.48	31.23	54.41	45.59	-0.39	0.37
Cyt b	1141	29.54	30.50	24.98	14.99	54.51	45.49	-0.08	-0.34
1st codon	-	21.09	27.56	25.23	26.12	46.32	53.68	0.09	-0.03
2nd codon	-	40.35	27.78	18.22	13.65	58.57	41.43	-0.38	-0.34
3rd codon	-	27.98	31.79	32.48	7.75	60.46	39.54	0.07	-0.61
PCGs	11428	29.80	29.04	25.31	15.85	55.11	44.89	-0.08	-0.29
rRNAs	2641	22.38	24.95	31.81	20.86	54.18	45.82	0.17	-0.09
tRNAs	1552	26.68	21.52	28.48	23.32	55.15	44.85	0.03	0.04
O_L	38	21.05	31.58	10.53	36.84	31.58	68.42	-0.33	0.08
D-loop	944	32.20	19.17	31.57	17.06	63.77	36.23	-0.01	-0.06
mtDNA	16607	27.55	27.98	27.83	16.64	55.39	44.61	0.01	-0.25

for nucleotide differences between the heavy chain and the light chain. The larger the absolute values, the more noteworthy differences between the two (Perna and Kocher, 1995). In PCGs, the AT-skews were from -0.39 (*ND6*) to 0.11 (*ATP8*), and all the GC-skews were negative except *ND6* (Table III). As the conventional preference in most mitogenomes, AT-skews are positive and GC-skews are negative, besides, the former (absolute value) is generally lower in magnitude than the latter (Fonseca *et al*., 2014). However, the AT-skews and GC-skews of most PCGs were negative in this study, indicating that the chain asymmetry in the nucleotide composition and the excess of cytosine and thymine in the H-strand (Table III, Fig. 2).

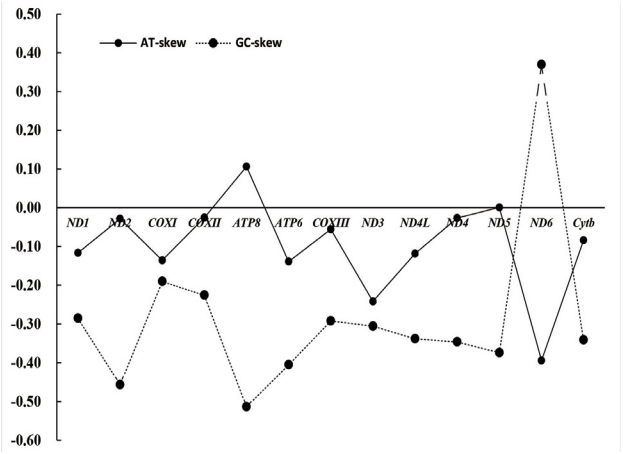

Fig. 2. AT-skews and GC-skews of protein coding genes in mitogenome of *A. cernua*.

The amino acid composition and the relative synonymous codon usage (RSCU) were preliminarily analyzed. The results showed that among the 3795 amino acids encoded, Leu2 was the most frequently used one (14.45 %). It was speculated that this result was related to the fact that the transmembrane protein encoded by the mitochondrial gene was mainly composed of hydrophobic amino acids (Zou, 2008). The percentage of Cys was the least, only accounting for 0.61 % (Fig. 3). RSCU means the preference for the use of synonymous codon, and it is defined as the ratio of the observed value of the synonymous codon number to the expected value of codon occurrence (Behura and Severson, 2013). There were 30 preferred codons (RSCU ≥ 1) in 13 PCGs, in which the highest RSCU value (2.18) was the codon CGA encoding Arg. The codons AGA and AGG were not used, so the codons with base A at the third site except UUA, CCA and AGA were all preferred codons, which coincided with A bias at the third codon position in the PCGs. UAA and UAG were stop codons and didn't encode any amino acids at all (Table IV).

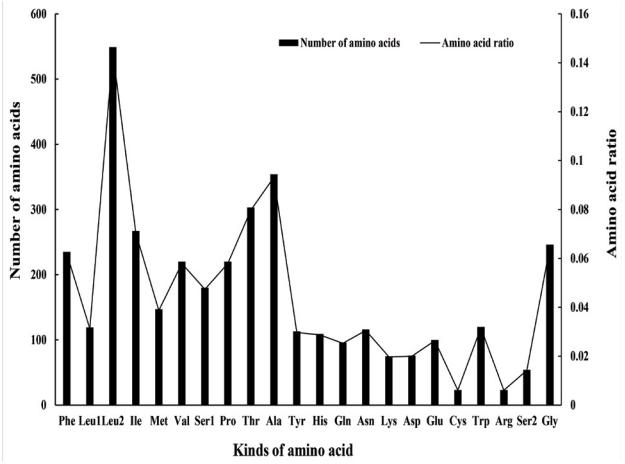

Fig. 3. Quantities of amino acids encoded by PCGs.

Transfer and ribosomal RNA genes

The typical 22 tRNA genes were identified in the mt genome of *A.cernua* ranging from 66 bp (tRNA-Cys) to 74 bp (tRNA-Lys). The base contents of tRNAs were 28.48% A, 21.52% C, 26.68% T, and 23.32% G, respectively (Table III). All 21 tRNAs could fold into typical clover secondary structure, but only tRNA-Ser ^{GCU} lacked dihydrouracil (DHU) stem (Fig. 4). Three base mismatches (A-C, U-U and A-A) were detected in tRNAs. Except for mismatches in the anticodon arms of tRNA-Trp and tRNA-Ser ^{GCU} as well as the DHU stem of tRNA-Arg, the others mainly occurred on the amino acid receptor arms and the TψC arms. However, partial mismatches of these tRNA genes can be corrected by later RNA editing without causing amino acid transport disorders, and these mismatches help eliminating deleterious mutations (Tomita *et al*., 1996; Lynch, 1997). The length of DHU stems was almost 4 bp excluding tRNA-Val, tRNA-Tyr and tRNA-Ser ^{UGA} (3 bp). Only the TψC stems of tRNA-Phe, tRNA-Lys and tRNA-Gln were 4 bp in length. The number of bases in the TψC loop and anticodon loop is mostly 7, but it varied greatly in DHU loop, ranging from 5 to10.

The full lengths of 12S rRNA and 16S rRNA genes of *A. cernua* mitogenome were 949 bp and 1692 bp, respectively. Both of them were located regularly between tRNA-Phe and tRNA-Leu ^{UUA}, and separated by tRNA-Val (Table II). The AT content (54.18%) was higher than GC content (45.82%), and the A base accounted for the largest proportion (31.81%) in the rRNAs (Table III). The secondary structures of rRNAs gene were relatively conserved, forming multiple stem-loop structures of different sizes. The predicted free energy of 12S rRNA gene was -260.32 kcal/ mol, and the free energy of 16S rRNA gene was -475.35 kcal/ mol.

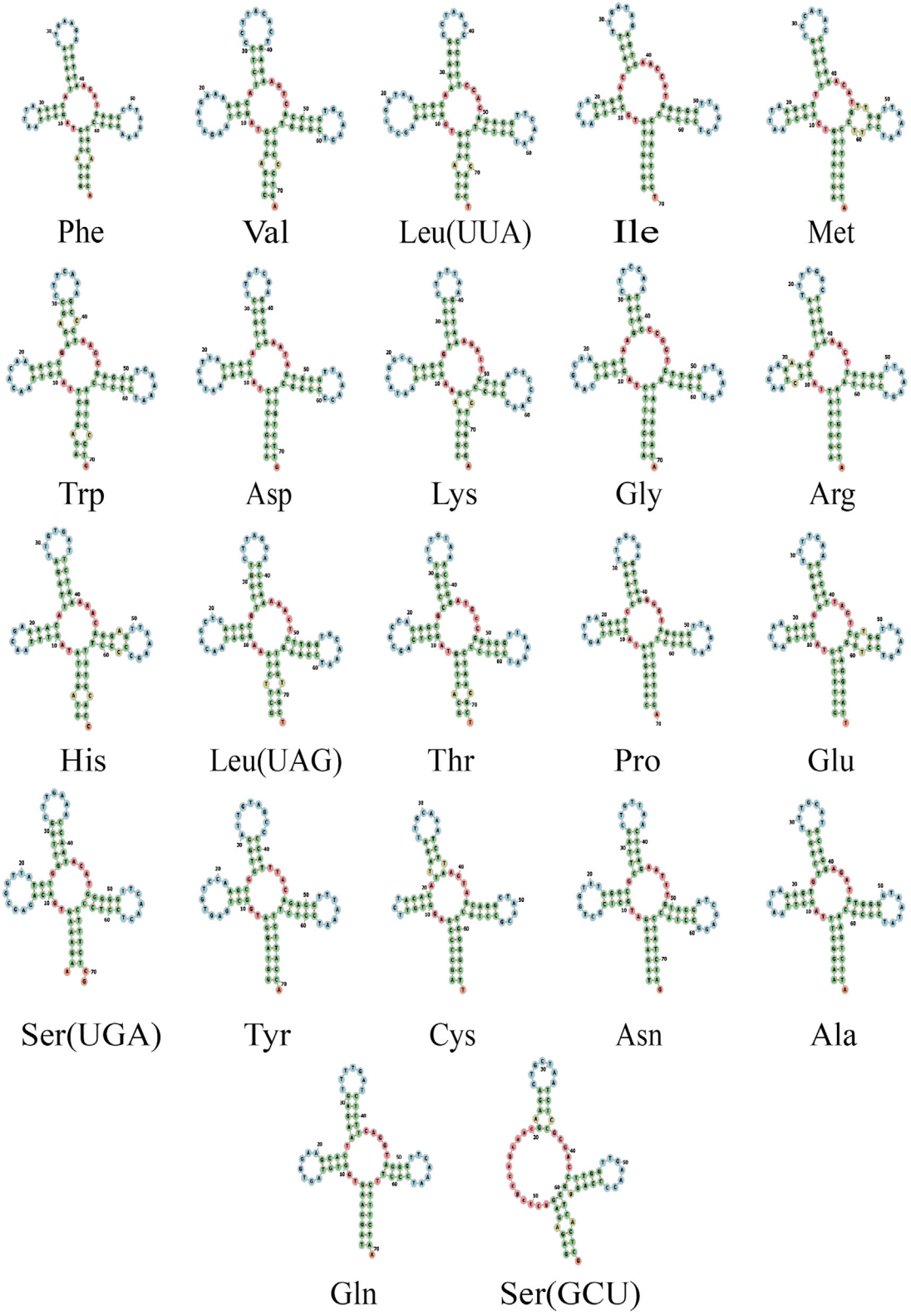

Fig. 4. The presumptive secondary structures of 22 tRNAs in mitochondrion of *A. cernua*.

Table IV. The number of codons and RSCU in *A. cernua.*

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	128	1.09	UCU(S1)	44	1.13	UAU(Y)	52	0.92	UGU(C)	7	0.61
UUC(F)	107	0.91	UCC(S1)	70	1.79	UAC(Y)	61	1.08	UGC(C)	16	1.39
UUA(L1)	110	0.99	UCA(S1)	57	1.46	UAA(*)	0	0	UGA(W)	105	1.75
UUG(L1)	9	0.08	UCG(S1)	9	0.23	UAG(*)	0	0	UGG(W)	15	0.25
CUU(L2)	180	1.62	CCU(P)	85	1.55	CAU(H)	35	0.64	CGU(R)	15	0.78
CUC(L2)	139	1.25	CCC(P)	83	1.51	CAC(H)	74	1.36	CGC(R)	11	0.57
CUA(L2)	179	1.61	CCA(P)	42	0.76	CAA(Q)	82	1.71	CGA(R)	42	2.18
CUG(L2)	51	0.46	CCG(P)	10	0.18	CAG(Q)	14	0.29	CGG(R)	9	0.47
AUU(I)	181	1.36	ACU(T)	60	0.79	AAU(N)	39	0.67	AGU(S2)	14	0.36
AUC(I)	86	0.64	ACC(T)	126	1.66	AAC(N)	77	1.33	AGC(S2)	40	1.03
AUA(M)	91	1.24	ACA(T)	110	1.45	AAA(K)	66	1.76	AGA(*)	0	0
AUG(M)	56	0.76	ACG(T)	7	0.09	AAG(K)	9	0.24	AGG(*)	0	0
GUU(V)	80	1.45	GCU(A)	79	0.89	GAU(D)	28	0.74	GGU(G)	37	0.60
GUC(V)	49	0.89	GCC(A)	143	1.62	GAC(D)	48	1.26	GGC(G)	80	1.30
GUA(V)	71	1.29	GCA(A)	120	1.36	GAA(E)	71	1.42	GGA(G)	85	1.38
GUG(V)	20	0.36	GCG(A)	12	0.14	GAG(E)	29	0.58	GGG(G)	44	0.72

Commonly, the lower absolute value of the free energy means the more stable the molecular structure (Zuker and Stiegler, 1981). It indicated that the 12S rRNA was more conservative than 16S rRNA. Because the evolutionary rate of 12S rRNA was relatively low in the mitogenome, it was suitable for the study of the evolutionary relationship of the family and above classification system (Song *et al*., 2008).

The rRNA secondary structures were predicted refering to the *Salmo salar* mitochondrial 12S rRNA (*b.16.m. S. salar*) file provided by CRW (Cannone *et al*., 2002) and *Homo sapiens* mitochondrial 16S rRNA (mHS_LSU_3D) file provided by Ribo Vision (Bernier *et al*., 2014), respectively. The 12S rRNA secondary structure of *A. cernua* mitochondrial DNA contained four domains (Fig. 5A), of which domain I and domain II were variable regions, and domain III and domain IV were conserved regions. The secondary structure of 16S rRNA comprised six domains (Fig. 5B), of which the I − III and VI domains were variable regions, and IV and V domains were conserved regions. In contrast with *S. salar*, there were base substitutions in both four regions of 12S rRNA, but base insertion and reposition were just observed in domain I and domain IV. Compared with *H. sapiens*, due to the tremendous differences between two species, the base substitution, insertion and reposition of *A. cernua* 16S rRNA appeared more frequently. The reposition and insertion were only discovered in the loop region. Therefore, the sequence of the rRNA stem region was considered to be relatively conservative than the sequence of the loop region.

Non-coding regions

The O_L was located inside the WANCY tRNA cluster (tRNA-Trp, tRNA-Ala, tRNA-Asn, tRNA-Cys and tRNA-Tyr) with a length of 38 bp. The AT content of O_L was 31.58 %, versus 68.42 % of the GC content, showing obvious GC preference. Just like most vertebrates, this small DNA fragment could be folded into a stable stem-loop structure with the stem and loop lengths of 24 bp and 14 bp, respectively (Macey *et al*., 1997; Zhang *et al*., 2014; Shi *et al*., 2015). The highly conserved block (5 ‘ -CCCCGG-3 ‘) was also found in tRNA-Cys gene after this sequence, suggesting that it was involved in regulating genome replication and transcription (Boore, 1999; Hixson *et al*., 1986).

Another non-coding region CR is proved to be the fastest evolutionary sequence in mtDNA and has been widely used in population genetics and molecular systematics researches (Xiao and Zhang, 2000; Liu *et al*., 2008). The CR was located between tRNA-Pro and tRNA-Phe, with a length of 931 bp and the AT content of 63.77 %. The mitochondrial CR sequence can be divided into three distinct domains: extended termination-associated sequences (ETAS), central conserved domain (CD), and conserved sequence block (CSB). Among them, the central conserved domain can be further organized into CSB-F, CSB-E, CSB-D, CSB-C and CSB-B regions, and the conserved sequence block also can fall into CSB1, CSB2 and CSB3 blocks (Lee *et al*., 1995; Li *et al*., 1996; Liu *et al*., 2008).

Reference to a variety of CR sequences in the order Perciformes, the length of ETAS was identified to be 364 bp in *A. cernua*. It contained two repetitive motif “TACAT”

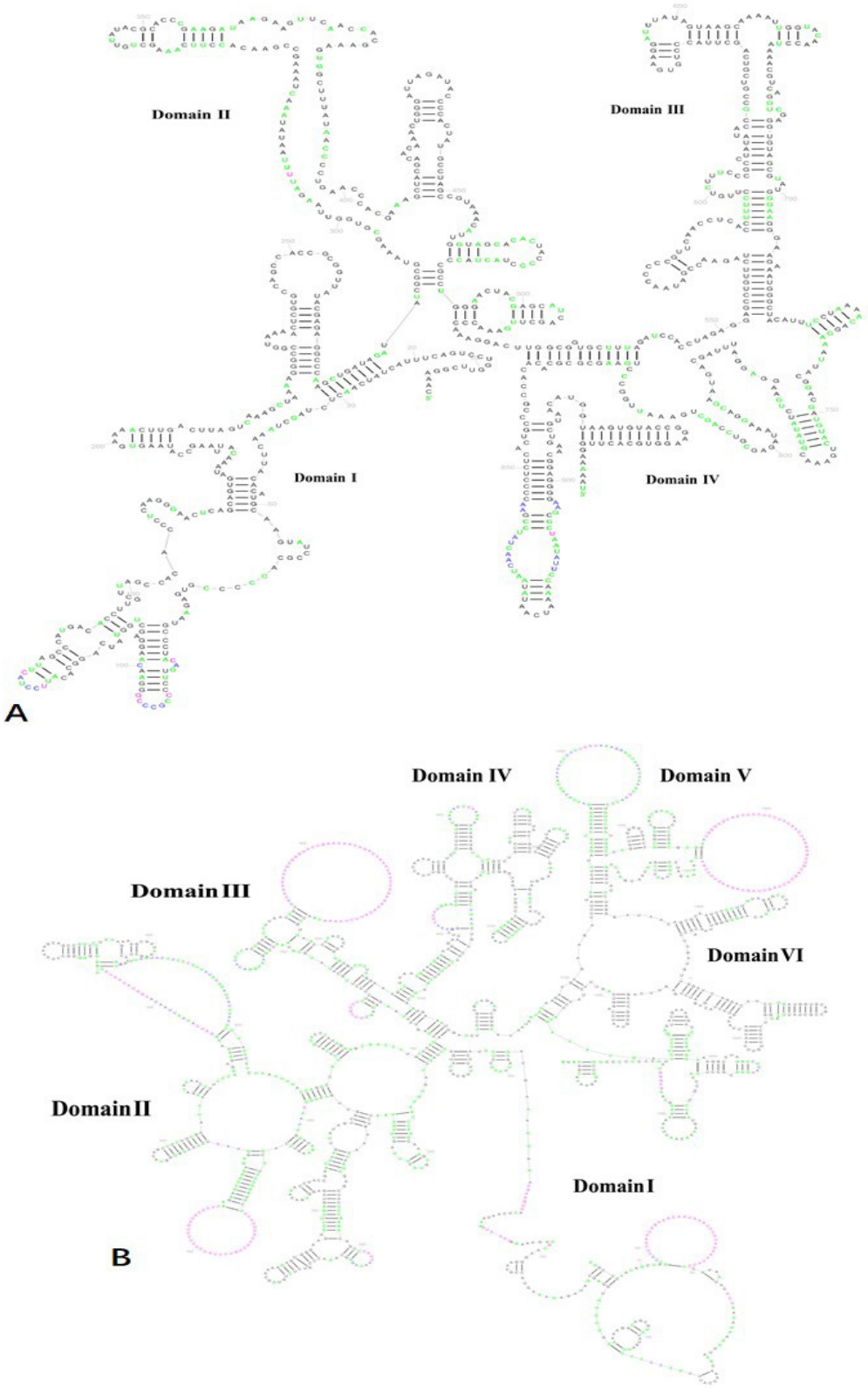

Fig. 5. A: Predicted secondary structure of 12S rRNA (A) and 16S rRNA (B) of *A. cernua*.

and an inverse complementary sequence "ATGTA", which could form a hairpin structure. A tandem repeat sequence "CAAGT ATTTG" was found in ETAS, and it also appeared in sturgeons (Zhang *et al*., 2000), *Coilia* (Tang *et al*., 2007), pikeperches (Faber and Stepien, 1998) and other fishes. The similar conserved sequences, corresponding to CSB-F, CSB-E and CSB-D were defined in CD, in which CSB-E could be identified by "GTGGG"-box sequence. It is reported that CSB-E and CSB-D are very similar and highly conserved among various Perciformes fishes (Wang, 2008; Zhao *et al*., 2016). Here, a highly conserved sequence "TCTTT TTTTT TTTTT TTCCT TTC" was also recognized, which was found in many fish control regions and was similar to a conserved fragment (CSB-B) in mammalian control regions (Southern *et al*., 1988). The sequence length of CSB (including CSB1, CSB2 and CSB3) was 220 bp. Among them, CSB1 is related to the initiation of mitochondrial DNA replication, so the conserved sequence region is considered to be the most critical part of the entire CR (Su *et al*., 2012) (Fig. 6).

Fig. 6. Mitochondrial control region structure of *A. cernua*.

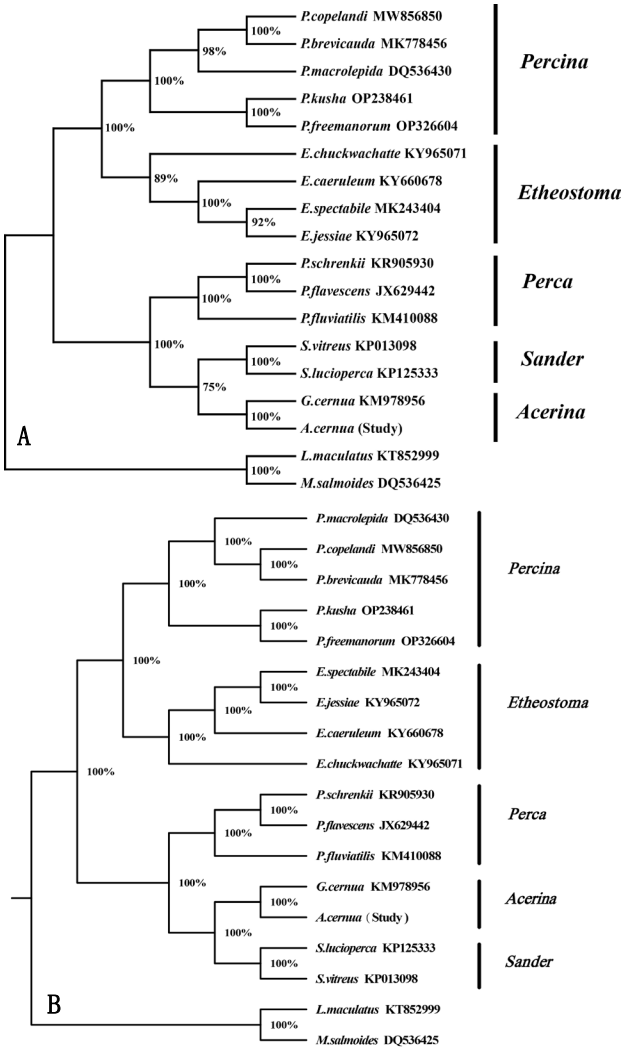

Fig. 7. Molecular phylogenetic trees constructed by ML (A) and BI (B) methods based on 12 PCGs concatenated sequences.

Phylogenetic analysis

Considering that only the *ND6* gene was located on the L-strand and the heterogeneity of base composition was likely to lead to poor phylogenetic analysis (Miya *et al*., 2003), therefore, 12 PCGs sequences encoded by H-strand were adopted to explore the phylogenetic relationships among Perciformes species. The topological trees were separately constructed by ML and BI methods (Fig. 7). The results showed that two trees possessed the similar topological structures, and species from the same genus clustered into the one evolutionary branch. Because *G. cernua* was regarded as synonyms of *A. cernua*, they were clustered into a separate branch both in the ML tree and BI tree. The genera *Acerina* and *Sander* first

gathered into a small branch, and then converged together with genus *Perca*. Therefore, *Acerina* and *Sander* had the closest genetic relationship, followed by *Perca*. By contrast, *Acerina* and *Percina*, *Acerina* and *Etheostama* were distantly related. Our result was consistent with the phylogenetic relationship constructed by Li *et al*. (2019) using the Neighbour-joining (NJ) method.

CONCLUSION

In summary, the complete mitogenome of *A. cernua* was determined by using High-throughput sequencing. The mitogenome (16607 bp) had similar gene order and orientation to those of other Percidae fishes. Based on 12 PCGs sequences, two kinds of molecular evolutionary trees were constructed and the phylogenetic status of *A. cernua* were also discussed, which provided new clues and references for the study of species evolution, classification and genetic diversity of this fish group.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Peng Chen in Xinjiang Fisheries Research Institute for his help in sample collecting.

Funding

This study was supported by the National Innovation and Entrepreneurship Training Program for College Students (No. 202210340023); Science and Technology Innovation Project of College Students in Zhejiang Province (No. 2023R411006).

IRB approval and ethical statement

The animal study protocol was approved by the Ethics Committee of Zhejiang Ocean University (protocol code: ZJOU-ECAE20211876, date of approval: 16 December 2021).

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Behura, S.K. and Severson, D.W., 2013. Codon usage bias: causative factors, quantification methods and genome-wide patterns with emphasis on insect genomes. *Biol. Rev. Cambridge Philos. Soc.*, **88**: 49–61. https://doi.org/10.1111/j.1469-185X.2012.00242.x

Bernier, C.R., Petrov, A.S., Waterbury, C.C., Jett, J., Li, F., Freil, L.E., Xiong, X., Wang, L., Migliozzi, B.L., Hershkovits, E., Xue, Y., Hsiao, C., Bowman, J.C., Harvey, S.C., Grover, M.A., Wartell, Z.J. and Williams, L.D., 2014. RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.*, **169**: 195–207. https://doi.org/10.1039/C3FD00126A

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

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N. and Gutell, R.R., 2002. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf.*, **3**: 2. https://doi.org/10.1186/1471-2105-3-15

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

Faber, J.E. and Stepien, C.A., 1998. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the Pike-Perches *Stizostedion*. *Mol. Phylogenet. Evol.*, **10**: 310–322. https://doi.org/10.1006/mpev.1998.0530

Fonseca, M.M., Harris, D.J. and Posada, D., 2014. The inversion of the Control Region in three mitogenomes provides further evidence for an asymmetric model of vertebrate mtDNA replication. *PLoS One*, **9**: e106654. https://doi.org/10.1371/journal.pone.0106654

Guindon, S. and Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systemat. Biol.*, **52**: 696–704. https://doi.org/10.1080/10635150390235520

Guo, Y., 2012. *Xinjiang ichthyology*. Xinjiang Science and Technology Press. Urumqi, Xingjing, China.

Hassanin, A., Léger, N. and Deutsch, J., 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Syst. Biol.*, **54**: 277–298. https://doi.org/10.1080/10635150590947843

Hixson, J.E., Wong, T.W. and Clayton, D.A., 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J. biol. Chem.*, **261**: 2384–2390. https://doi.org/10.1016/S0021-9258(17)35948-3

Huelsenbeck, J.P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees.

Bioinformatics (Oxford, England), **17**: 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Imoto, J.M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y.P., Miya, M., Nishida, M. and Hanzawa, N., 2013. Phylogeny and biogeography of highly diverged freshwater fish species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. *Gene*, **514**: 112–124. https://doi.org/10.1016/j.gene.2012.10.019

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M., 2001. A mitogenomic perspective on the basal teleostean phylogeny: Resolving higher-level relationships with longer DNA sequences. *Mol. phylogenet. Evol.*, **20**: 275–285. https://doi.org/10.1006/mpev.2001.0970

Kawahara, R., Miya, M., Mabuchi, K., Lavoué, S., Inoue, J.G., Satoh, T.P., Kawaguchi, A. and Nishida, M., 2008. Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): A new perspective based on whole mitogenome sequences from 75 higher teleosts. *Mol. phylogenet. Evol.*, **46**: 224–236. https://doi.org/10.1016/j.ympev.2007.07.009

Kottelat, M. and Freyhof, J., 2007. *Handbook of European freshwater fishes*. Publications Kottelat, Cornol, CH. pp. 646.

Kumar, S., Nei, M., Dudley, J. and Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings Bioinf.*, **9**: 299–306. https://doi.org/10.1093/bib/bbn017

Lee, W.J., Conroy, J., Howell, W.H. and Kocher, T.D., 1995. Structure and evolution of teleost mitochondrial control regions. *J. mol. Evol.*, **41**: 54–66. https://doi.org/10.1007/BF00174041

Li, R.Q., Zhu, H.M., Ruan, J., Qian, W.B., Fang, X.D., Shi, Z.B., Li, Y.R., Li, S.T., Shan, G., Kristiansen, K., Li, S.G., Yang, H.M., Wang, J. and Wang, J., 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.*, **20**: 265–272. https://doi.org/10.1101/gr.097261.109

Li, Y., 1996. Study on the structure of mtDNA control region in fish. *J. Qujing Normal Coll.*, **15**: 36-39.

Li, Y.Y., Liu, L.X., Zhang, R.H., Ma, C. and Liu, Y.G., 2019. The complete mitochondrial DNA sequence of ruffe *Acerina cernua*. *Mitochondrial DNA Part B.*, **4**: 1590-1591. https://doi.org/10.1080/23802359.2019.1602005

Liu, H.Y., Yu, L.N. and Zhang, F.R., 2008. Molecular structure and application progress of fish mitochondrial DNA control region. *Reservoir Fish.*, **28**: 4-8.

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**: W54–W57. https://doi.org/10.1093/nar/gkw413

Lynch, M., 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol. Biol. Evol.*, **14**: 914–925. https://doi.org/10.1093/oxfordjournals.molbev.a025834

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z. and Papenfuss, T.J., 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.*, **14**: 91–104. https://doi.org/10.1093/oxfordjournals.molbev.a025706

Miniatis, T., Fritsch, E.F., Sambrook, J. and Engel, J., 1985. Molecular cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory. 1982, 545 S., 42 $. *Acta Biotechnol.*, **5**: 104-104. https://doi.org/10.1002/abio.370050118

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M. and Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. phylogenet. Evol.*, **26**: 121–138. https://doi.org/10.1016/S1055-7903(02)00332-9

Nesbø, C.L., Arab, M.O. and Jakobsen, K.S., 1998. Heteroplasmy, length and sequence variation in the mtDNA control regions of three percid fish species (*Perca fluviatilis*, *Acerina cernua*, *Stizostedion lucioperca*). *Genetics*, **148**: 1907–1919. https://doi.org/10.1093/genetics/148.4.1907

Perea, S., Böhme, M., Zupancic, P., Freyhof, J., Sanda, R., Ozuluğ, M., Abdoli, A. and Doadrio, I., 2010. Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evol. Biol.*, **10**: 265. https://doi.org/10.1186/1471-2148-10-265

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

Shi, W., Gong, L., Wang, S.Y., Miao, X.G. and Kong, X.Y., 2015. Tandem duplication and random loss for mitogenome rearrangement in *Symphurus* (Teleost: Pleuronectiformes). *BMC Genom.*, **16**: 355. https://doi.org/10.1186/s12864-015-1581-6

Song, J., Zhang, X.F., Diao, X.T. and Jin, M., 2008. Research progress of mitochondrial DNA, 12S rRNA and 16S rRNA. *Anhui agric. Sci. Bull.*, **14**: 42-43, 37.

Southern, S.O., Southern, P.J. and Dizon, A.E., 1988. Molecular characterization of a cloned dolphin mitochondrial genome. *J. mol. Evol.*, **28**: 32–42. https://doi.org/10.1007/BF02143495

Su, T.F., Jiang, S.G., Ma, Z.M. and Li, X., 2012. Structure and evolution of mitochondrial control regions in *Evynnis cardinalis*. *S. China Fish. Sci.*, **8**: 1-8.

Tang, W.Q., Hu, X.L. and Yang, J.Q., 2007. Species validities of *Coilia brachygnathus* and *C. nasus* taihuensis based on sequence variations of complete mtDNA control region. *Biodiv. Sci.*, **15**: 224-231. https://doi.org/10.1360/biodiv.060263

The RNAcentral Consortium, 2019. RNAcentral: A hub of information for non-coding RNA sequences. *Nucl. Acids Res.*, **47**: D221–D229. https://doi.org/10.1093/nar/gky1034

Tomita, K., Ueda, T. and Watanabe, K., 1996. RNA editing in the acceptor stem of squid mitochondrial tRNA(Tyr). *Nucl. Acids Res.*, **24**: 4987–4991. https://doi.org/10.1093/nar/24.24.4987

Van, N.S.M. and Van, M.J.T.P.W., 1994. Mechanophysiological properties of the supraorbital lateral line canal in ruffe (*Acerina cernua* L.). *Proc. biol. Sci.,* **256**: 239–246. https://doi.org/10.1098/rspb.1994.0076

Wang, D.Y., 2008. *Length and sequence variation in mitochondrial DNA control region and genetic structure of* Bostrichthy sinensis. PhD thesis, Xiamen University, Xiamen, China.

Xiao, W.H. and Zhang, Y.P., 2000. Genetics and evolution of mitochondrial DNA in fish. *Acta Hydrobiol. Sin.*, **24**: 384-391.

Zhang, B., Sun, Y. and Shi, G., 2014. The complete mitochondrial genome of the four finger threadfin *Eleutheronema tetradactylum* (Perciforms: Polynemidae) and comparison of light strand replication origin within Percoidei. *Mitochondr. DNA*, **25**: 411–413. https://doi.org/10.3109/19401736.2013.809433

Zhang, S.M., Wu, Q.J. and Zhang, Y.P., 2000. Tandem repeats of Chinese sturgeon (*Acipenser sinensis*) and related species and its significance in evolution. *Chin. J. Biochem. mol. Biol.*, **16**: 458-461.

Zhang, Y., Li, Y.F., Shen, Z.W., Ru, H.J., Wu, X.X., Liu, X.J. and Ni, Z.H., 2017. Mitochondrial DNA sequence of ruffe (*Gymnocephalus cernua*). *Mitochondr. DNA A, DNA Mapp., Sequenc. Anal.*, **28**: 299–300. https://doi.org/10.3109/19401736.2014.1003888

Zhao, L.L., Bi, X.X., Song, L. and Gao, T.X., 2016. Analysis of the structure of mitochondrial DNA control region and the genetic diversity of *Trachidermus fasciatus* in different populations. *Acta Hydrobiol. Sin.*, **40**:35-41.

Zhao, Q.Y., Wang, Y., Kong, Y.M., Luo, D., Li, X. and Hao, P., 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinf.*, **12**: S2. https://doi.org/10.1186/1471-2105-12-S14-S2

Zhong, D., Zhao, G.J., Zhang, Z.S. and Xun, A.L., 2002. Advance in the entire balance and local unbalance of base distribution in genome. *Hereditas*, **24**: 351–355.

Zhu, M.Y., Wang, X., Zhou, Y., Song, Y. and Qin, L., 2014. Population dynamics of diplostomum infection in *Acerina cernua* in Ergis River. *J. Hydroecol.*, **35**: 78-81.

Zou, L.Y., 2008. *Research on relevant problems of discriminating sequences and structures of outer membrane proteins*. PhD thesis, National University of Defense Technology, Hunan, China.

Zuker, M. and Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.*, **9**: 133–148. https://doi.org/10.1093/nar/9.1.133