ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 28/04/2025

DOI: https://doi.org/10.3897/arphapreprints.e156693

The complete mitochondrial genomes of *Acheilognathus mengyangensis* (Cypriniformes, Cyprinidae, Acheilognathinae): characterization and phylogenetic analysis

Zhengran LI, Chuanjiang Zhou, Yu Jinhui

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 28/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e156693

The complete mitochondrial genomes of *Acheilognathus mengyangensis* (Cypriniformes, Cyprinidae, Acheilognathinae): characterization and phylogenetic analysis

Zhengran Li[1], Jinhui Yu[1], Chuanjiang Zhou[1]*

1. The Observation and Research Field Station of Taihang Mountain Forest Ecosystems of Henan Province, College of Life Sciences, Henan Normal University, Xinxiang 453007, People's Republic of China.

* Corresponding authors: College of Life Sciences, Henan Normal University, Xinxiang 453007, Henan, China. E-mail addresses: chuanjiang88@163.com (C. Zhou).

Abstract

In this paper, we analyzed the mitochondrial genome data of *Acheilognathus mengyangensis* and used the multi-gene tandem method to elucidate its taxonomic status. The total length of the mitochondrial genome is 16,779 bp, including 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes, and D-loop region. Overall, there was strong AT bias and anti-G bias; Different protein-coding genes exhibit different degrees of codon preference. The rest of the amino acids except $tRNA^{Ser\ (GCT)}$ of the 22 tRNAs were in one form, and the secondary structure was incomplete, and all the other tRNAs folded to show the secondary structure of a typical clover. 13 protein-coding genes of *Acheilognathus* and *Rhodeus* were concatenated to explore the phylogenetic relationship with *Pseudorasbora parva* as the outgroup. The results showed that *A. mengyangensis, A. chankaensis, A. macropterus* is the most closely related. This research supplements Acheilognathus mitochondrial genome data, serving as a molecular basis for safeguarding species, facilitating genetic classification, and investigating Acheilognathinae phylogeny.

Keywords: *Acheilognathus mengyangensis*, mitochondrial genome, phylogenetic relationships

Introduction

Members of the subfamily Acheilognathinae primarily inhabit freshwater environments such as rivers and streams, and prefer areas with slower currents and clear water. This subfamily includes 72 species (Tang et al. 2024) and 6 recognized genera: *Rhodeus* Agassiz (Agassiz 1832), *Acheilognathus* Bleeker, see (Bleeker 1860), *Tanakia* Jordan & Thompson (Jordan et al. 1914), *Paratanakia* Chang (Chang et al. 2014), *Pseudorhodeus* Chang (Chang et al. 2014), and *Sinorhodeus* Li (Li et al. 2017). *Acheilognathus* is widely distributed across major river systems in China, such as the Yangtze River, Pearl River, Amur River, and Yellow River. In the breeding phase, female fish develop an elongated oviduct and deposit their eggs into the gill chambers of *Unio douglasiae Griffith et Pidgeon*. Then, sperm from the male fish enters the *Unio douglasiae Griffith et Pidgeon*'s inhalant siphon and gills to facilitate fertilization. The fertilized eggs subsequently hatch and develop within the mussel's gills until the juveniles are capable of swimming independently (Reichard et al. 2010, Yi et al. 2024).

Progress in genomic technologies has facilitated the characterization of diverse mitochondrial genomes (mtDNA), providing critical insights into fish evolutionary history, population dynamics, and conservation prioritization. (Guo et al. 2004, Chen 2011). The mitochondrial genome of *A. mengyangensis* was fully annotated (length, type, PCGs, non-coding and RNA elements) and comparatively analyzed.

*Acheilognathus mengyangensi*s is a small-sized fish species typically found in slow-moving streams and rivers (Figure 1). The body length of *A. mengyangensis* is generally about 5-8 cm, the body is flattened on the side, spindle-shaped, the body color is usually silvery-white or pale yellow, the dorsal and caudal fins are more developed, and the caudal fin is forked. *A. mengyangensis* mainly feeds on aquatic insects, algae, plankton, etc. This species deposits its eggs in the gills of *Unio douglasiae Griffith et Pidgeon*, To date, specimens have only been collected from the Mengyang River in Mengyang Town, Pengzhou City, and the tributaries of the Baitiao River in Tianma Town, Dujiangyan City, both located in Chengdu, Sichuan Province. Mitochondrial genome sequences of Acheilognathus and Rhodeus species were acquired from the NCBI database. With *Pseudorasbora parva* selected as the outgroup, we concatenated 13 protein-coding genes (PCGs) and built ML and Bayes inference trees, applying the best-fit substitution model for nucleotide sequences and the optimal partitioning strategy correspondingly. The findings presented in this paper are expected to provide valuable insights and theoretical foundations for future research on the evolutionary development of the Acheilognathinae subfamily.

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 28/04/2025. DOI: https://doi.org/10.3897/arphapreprints.e156693

Figure 1. Morphological appearance of *A. mengyangensis*. The photo was taken by Jinhui Yu.

1 Material and methods

1.1 Sample collection and raw data generation

Samples of *A. mengyangensis* were collected from Mengyang Town, Chengdu City, Sichuan Province, China. The specimens were preserved in absolute ethanol and stored at the College of Life Sciences, Henan Normal University. Georeferencing data were unavailable due to incomplete site documentation during sampling. DNA extraction from *A. mengyangensis* was performed using the phenol-chloroform method (Sambrook& Russell 2001). Muscle tissue is dissolved with proteinase K to release DNA, followed by the addition of a phenol-chloroform mixture. After centrifugation, the aqueous layer is collected and DNA is precipitated using isopropanol or ethanol. The DNA pellet is then washed with ethanol and dried. The extracted DNA was dissolved in 40 µl double-distilled water and stored at -20°C.

Through a comparative analysis of the complete mitochondrial genome sequence of *A. mengyangensis* and *A. tonkinensis*, we employed MitoZ to generate the complete mitochondrial genome in GenBank format and to identify the start and end positions of each gene.

1.2 Annotation of the Mitochondrial Genome

Mitogenome reconstruction employed reference-guided assembly against existing Acheilognathinae sequences in NCBI. The starting position of the transfer RNA ($tRNA^{Phe}$) was identified using the MITOS website (Bernt et al. 2013). Codon-based preprocessing of mitochondrial genes followed the standard teleost genetic code, and the circular nature of the mitochondrial genome was verified. The Mitofish platform was utilized to determine the locations of the 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs), 2 ribosomal RNAs (rRNAs), control regions. This process resulted in the generation of a gene map for the *A. mengyangensis* mitochondrial genome. To annotate tRNA genes, we utilized the tRNAscan-SE algorithm (Lowe& Chan 2016), which validated gene boundaries and generated cloverleaf secondary structure models. For PCG translation, nucleotide

sequences were aligned against the vertebrate mitochondrial genetic code in MEGA7.0 (Kumar et al. 2016), with manual verification of start/stop codons. Base composition and relative synonymous codon usage (RSCU) were calculated using MEGA7.0. AT:GC skew was calculated using Perna's formula (Perna & Kocher 1995), where AT Skew = (A - T)/(A + T) and GC Skew = (G - C)/(G + C). CodonW 1.4.2 software (Peden 2000) was used to derive the codon adaptation index (CAI) (Jin et al. 1832), effective number of codons (ENC), and GC and GC3 content.

1.3 Phylogenetic assessment

To investigate the phylogenetic position of *A. mengyangensis*, the dataset comprised 23 mitogenomes (18 Acheilognathus, 5 Rhodeus), curated from NCBI based on assembly completeness (>95% coverage). Taxon inclusion followed the most recent molecular phylogeny confirming Gobioninae-Acheilognathae sisterhood, thereby minimizing outgroup selection artifacts (Tang et al. 2024), *Pseudorasbora parva* (GenBank accession: NC_028016) was designated as the outgroup based on its well-established phylogenetic position as a sister taxon to Acheilognathinae. The final dataset integrated 23 mitogenomes, comprising 18 ingroup and 5 outgroup species, to ensure topological stability. Table 1 provides the GenBank accession numbers and corresponding species names. Two partitioned phylogenetic analysis methods were employed: one based on Bayes principle (Ronquist & Huelsenbeck 2003) and the other on the Maximum Likelihood (ML) partitioned analysis method (Stamatakis 2006). Using Phylosuite software (Zhang et al. 2020), we extracted the 13 protein-coding genes (PCGs), 2 ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes from each species. The extracted PCGs were then aligned individually using ClustalW (Peden 2000).

The sequences were aligned based on the mitochondrial genome using Phylosiute (Zhang et al. 2020) to concatenation the protein-coding genes (PCGs) into a 13 PCG dataset. The optimal partitioning scheme and corresponding nucleotide substitution model were determined using ModelFinder (Kalyaanamoorthy et al. 2017) based on the Akaike Information Criterion (AIC). A maximum likelihood (ML) tree was constructed using IQ-TREE with an edge-linked partitioning model and 5000 ultrafast bootstrap replicates (Stamatakis 2006, Lanfear et al. 2016). For Bayesian inference analyses, PartitionFinder (Lanfear et al. 2016) was employed to select the best partition and associated optimal nucleotide substitution model based on the Bayesian Information Criterion (BIC) (Huelsenbeck & Ronquist 2001, Drummond & Rambaut 2007). Bayesian phylogenetic trees were inferred using MrBayes 3.2.6 (Ronquist & Huelsenbeck 2003) under the N/A model, with two parallel runs of 2,000,000 generations. The initial 25% of sampled trees were discarded as burn-in.

Table 1. Genome sequences from NCBI in this study

GenBank Accession	Speices	Length	AT (%)
NC007885	*Rhodeus uyekii*	16817	55
NC008668	*Acheilognathus typus*	16778	57

NC013704	*Tanakia koreensis*	16563	54
NC013709	*Rhodeus suigensis*	16733	55.1
NC013711	*Acheilognathus macropterus*	16774	57
NC013712	*Acheilognathus yamatsutae*	16703	56.7
NC022690	*Rhodeus shitaiensis*	16774	55
NC023101	*Acheilognathus chankaensis*	16774	58
NC024566	*Tanakia lanceolata*	16607	54
NC025515	*Tanakia limbata*	16565	54
NC026872	*Acheilognathus barbatus*	16770	56
NC027437	*Rhodeus fangi*	16733	55.3
NC028416	*Acheilognathus gracilis*	16988	57
NC028433	*Acheilognathus rhombeus*	16780	56.9
NC028736	*Acheilognathus majusculus*	17155	57
NC029718	*Rhodeus notatus*	16735	55.3
NC031152	*Acheilognathus meridianus*	16563	57.9
NC037404	*Acheilognathus omeiensis*	16774	56.7
NC039820	*Tanakia latimarginata*	16588	54
NC042407	*Acheilognathus tonkinensis*	16767	56.5
NC042717	*Sinorhodeus microlepis*	16591	57
NC066655	*Paratanakia chii*	16575	57
NC015614	*Pseudorasbora parva*	16600	58.9

2 Results

2.1 Genome size and organization

Assembly of the *A. mengyangensis* mitogenome yielded a circular 16,779 bp molecule (Figure 2), exhibiting the conserved vertebrate gene repertoire: 13 PCGs (ranging from 165 bp for ATP8 to 1,836 bp for ND5), 2 rRNAs, and 22 tRNAs. Notably, COX1 (1,581 bp) and CYTB (1,141 bp) displayed the highest sequence conservation (>90% identity) with other Cyprinidae species, 2 ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes, totaling 37 genes. Gene distribution analysis revealed pronounced strand asymmetry: while 28 genes (including 12 PCGs and 16 RNAs) were encoded on the heavy strand (H-strand), only a minor cluster of 8 tRNAs and ND6 (522 bp) resided on the light strand (L-strand). The gene arrangement is consistent with the typical genetic organization observed in teleost fish (Figure 2, Table 2).

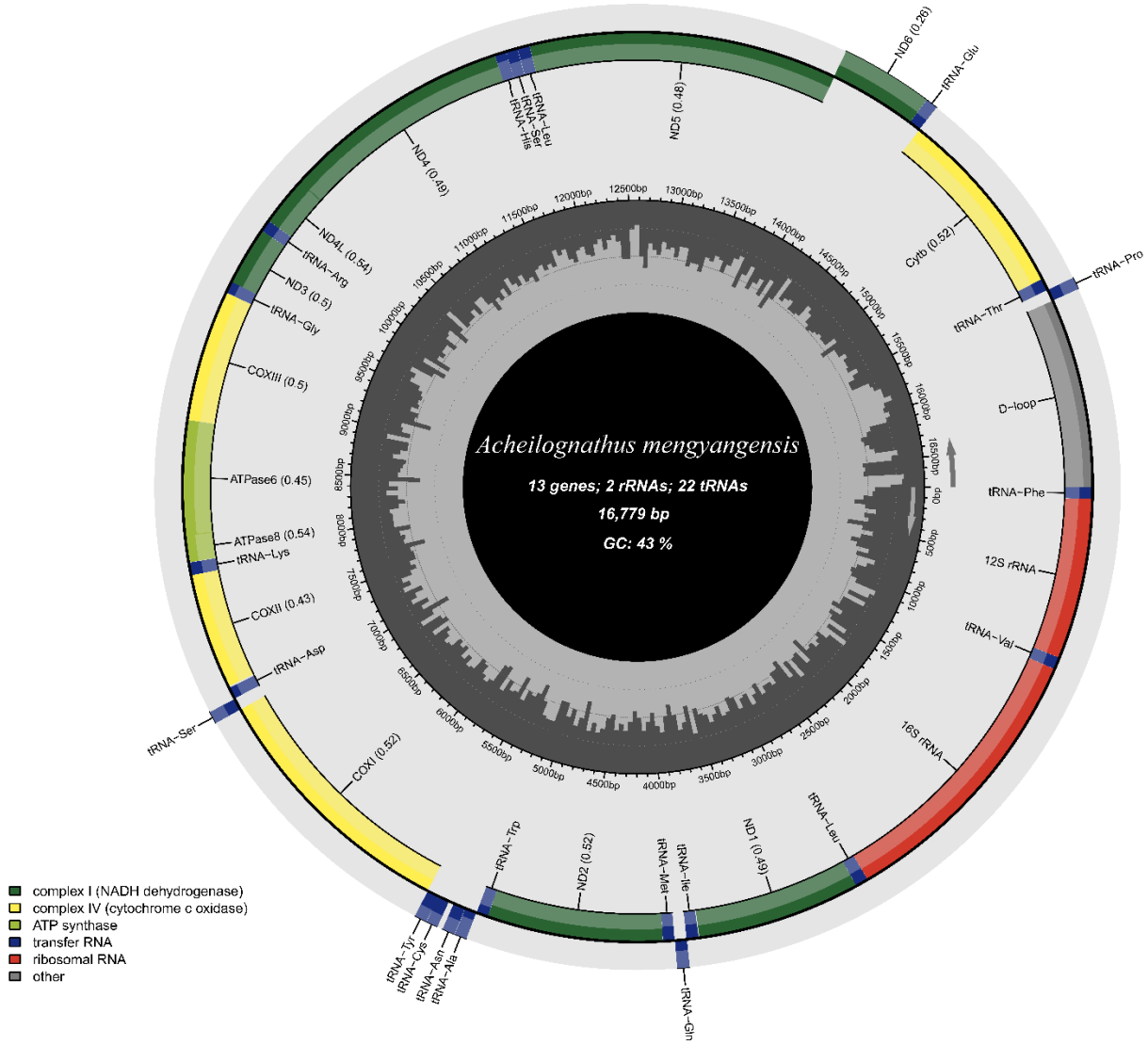

Figure 2. Gene map of the *A. mengyangensis* mitochondrial genomes.

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 28/04/2025 DOI: https://doi.org/10.3897/arphapreprints.e156693

Table 2. Mitochondrial genes and associated features of *A. mengyangensis*

	Type	One-letter code	Strand	Amino acids	Position			Codon		
					Start	Stop	Length (bp)	Start	Stop	Anti-condon
$tRNA^{Phe}$	tRNA	F	H		1	69	69			GAA
12S rRNA	rRNA		H		70	1025	956			
$tRNA^{Val}$	tRNA	V	H		1027	1098	72			TAC
16S rRNA	rRNA		H		1115	2774	1660			
$tRNA^{Leu}$	tRNA	L2	H		2775	2850	76			TAA
ND1	Protein-coding		H	324	2851	3825	975	ATG	TAG	
$tRNA^{Ile}$	tRNA	I	H		3830	3901	72			GAT
$tRNA^{Gln}$	tRNA	Q	L		3900	3970	71			TTG
$tRNA^{Met}$	tRNA	M	H		3972	4040	69			CAT
ND2	Protein-coding		H	348	4041	5087	1047	ATG	T(AA)	
$tRNA^{Trp}$	tRNA	W	H		5086	5155	70			TCA
$tRNA^{Ala}$	tRNA	A	L		5157	5225	69			TGC
$tRNA^{Asn}$	tRNA	N	L		5227	5299	73			GTT
$tRNA^{Cys}$	tRNA	C	L		5331	5398	68			GCA
$tRNA^{Tyr}$	tRNA	Y	L		5399	5469	71			GTA
COX1	Protein-coding		H	516	5471	7021	1551	GTG	TAA	
$tRNA^{Ser}$	tRNA	S2	L		7022	7092	71			TGA
$tRNA^{Asp}$	tRNA	D	H		7095	7165	71			GTC
COX2	Protein-coding		H	230	7173	7863	691	ATG	T(AA)	
$tRNA^{Lys}$	tRNA	K	H		7864	7939	76			TTT
ATP8	Protein-coding		H	54	7941	8105	165	ATG	TAA	

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 28/04/2025 DOI: https://doi.org/10.3897/arphapreprints.e156693

ATP6	Protein-coding		H	227	8099	8782	684	ATG	TA(A)	
COX3	Protein-coding		H	261	8782	9566	785	ATG	T(AA)	
$tRNA^{Gly}$	tRNA	G	H		9566	9636	71			TCC
ND3	Protein-coding		H	116	9637	9987	351	ATG	T(AA)	
$tRNA^{Arg}$	tRNA	R	H		9986	10054	69			TCG
ND4L	Protein-coding		H	98	10055	10351	297	ATG	TAA	
ND4	Protein-coding		H	460	10345	11725	1381	ATG	TA(A)	
$tRNA^{His}$	tRNA	H	H		11726	11794	69			GTG
$tRNA^{Ser}$	tRNA	S1	H		11795	11863	69			GCT
$tRNA^{Leu}$	tRNA	LI	H		11865	11937	73			TAG
ND5	Protein-coding		H	611	11938	13773	1836	ATG	TAG	
ND6	Protein-coding		L	173	13770	14291	522	ATG	TAA	
$tRNA^{Glu}$	tRNA	E	L		14292	14360	69			TTC
CYTB	Protein-coding		H	380	14363	15503	1141	ATG	T(AA)	
$tRNA^{Thr}$	tRNA	T	H		15504	15576	73			TGT
$tRNA^{Pro}$	tRNA	P	L		15576	15645	70			TGG
D-loop	Non-coding		H		16052	16630	579			

2.2 Base composition

In *A. mengyangensis*, the highest base content was observed for A at 29.52%, followed by T at 27.21%, C at 26.28%, and G at 16.99%. A strong AT preference (56.7%) and guanine depletion (16.9%) were observed (Table 3), reflecting the selection of energy-favorable base pairs in the mitochondrial transcripts. This compositional bias is similar to that found in most sequenced hard bone mitotic genomes (Perna & Kocher 1995). Mitogenome-wide nucleotide analysis revealed consistent AT-richness (D-loop:72.1%, tRNAs:61.3%, PCGs:56.7%) with strand-specific biases: 12/13 PCGs showed anti-A bias (except ATP8) and 11/13 anti-G bias (except ND2/ND3), while rRNAs exhibited extreme A-bias (AT-skew>0.25) contrasting with tRNA's unique G-bias. Notably, ND6 shared tRNA-like G/T-bias patterns, and PCGs demonstrated stronger anti-G bias than D-loop or rRNAs, suggesting transcriptional strand asymmetry governs regional composition differences.

2.3 Protein-coding genes

The 13 protein-coding genes in *A. mengyangensis* mitogenome span 11,421 bp, accounting for over two-thirds (68.07%) of its total genomic content.

Table 3 Base composition analysis of *A. mengyangensis* mitogenome regions.

Region	Base composition (%)							
	Total	T	C	A	G	AT (%)	AT-skew	CG-skew
ATP6	683	30.89	27.53	27.23	14.35	58.13	-0.06	-0.31
ATP8	165	24.85	25.45	38.79	10.91	63.64	0.22	-0.40
COI	1551	31.21	25.34	25.21	18.25	56.42	-0.11	-0.16
COII	691	28.65	25.33	28.36	17.66	57.02	-0.01	-0.18
COIII	784	28.83	27.55	25.38	18.24	54.21	-0.06	-0.20
CYTB	1141	29.97	27.17	26.99	15.86	56.97	-0.05	-0.26
ND1	975	28.62	28.51	27.08	15.79	55.69	-0.03	-0.29
ND2	1045	26.12	31.20	28.90	13.78	55.02	0.05	-0.39
ND3	349	29.23	29.51	26.07	15.19	55.30	-0.06	-0.32
ND4	1382	28.29	28.22	28.00	15.48	56.30	-0.01	-0.29
ND4L	297	29.97	27.95	24.92	17.17	54.88	-0.09	-0.24
ND5	1836	29.52	26.85	28.81	14.81	58.33	-0.01	-0.29
ND6	522	36.97	14.37	17.24	31.42	54.21	-0.36	0.37
PCGs	11421	29.52	26.90	26.98	16.61	56.49	-0.04	-0.24
tRNAs	1562	27.53	20.74	28.94	22.79	56.47	0.02	0.05
D-loop	1127	30.79	19.88	35.85	13.49	66.64	0.08	-0.19

12S rRNA	958	19.00	27.14	31.11	22.76	50.10	0.24	-0.09
16S rRNA	1680	21.31	22.68	34.70	21.31	56.01	0.24	-0.03
Complete genome	16779	27.21	26.28	29.52	16.99	56.73	0.04	-0.21

2.3.1 The use of amino acids and codons

The 13 PCGs encode 3,798 amino acids, with Leu (18.2%), Ser (12.1%), Pro (10.7%), and Thr (9.8%) dominating the composition (Table 4). Codon usage analysis (Fig. 3) showed strong AT preference: high-frequency codons (AAA-Lys, AAU-Asn, UAA-Stop) contained 65-72% A/U, whereas C/G-rich codons (GGG-Gly, GGC-Gly, CGC-Arg) accounted for <5% of occurrences. Based on RSCU comparisons, the most frequently used codons in *A. mengyangensis* were GCC (1.74%), UAA (1.56%), CAA (1.38%), and UUA (1.36%), encode the amino acids Ala, Tyr, Glu and Leu, respectively, Conversely, the least frequently used codons were those encoding alanine (GCG, 0.33%), serine (UCG, 0.45%), and threonine (ACG, 0.45%).

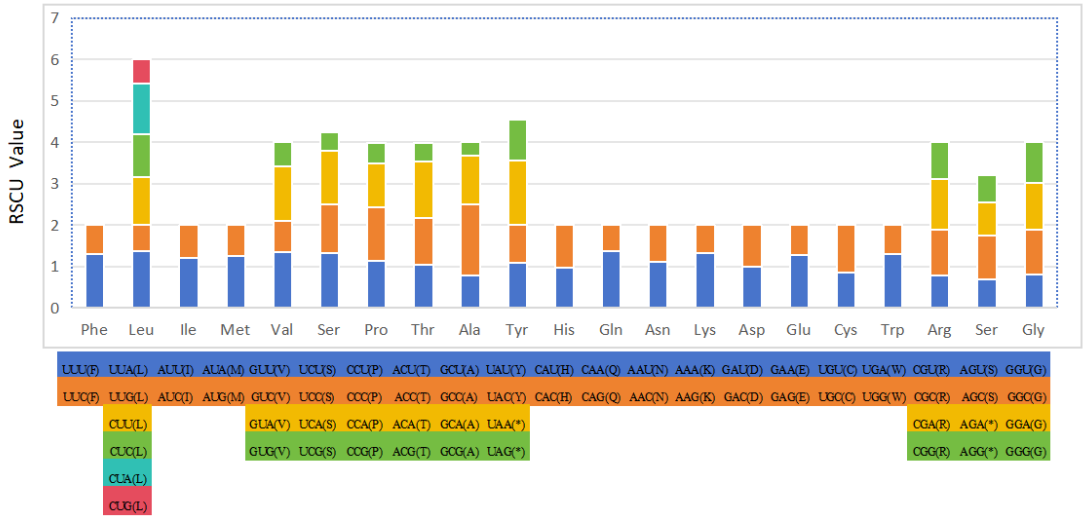

Figure 3. Results from analysis of Relative Synonymous Codon Usage (RSCU) of the mitochondrial genome of *A. mengyangensis*.

Table 4. Results from the Relative Synonymous Codon Usage (RSCU) analysis for the PCGs of the mitochondrial genome of *A. mengyangensis*. * denotes stop codon.

AA	Codon	Count	RSCU	AA	Codon	Count	RSCU
Phe	UUU(F)	172	1.29	Tyr	UAU(Y)	116	1.09
	UUC(F)	94	0.71		UAC(Y)	97	0.91
Leu	UUA(L)	144	1.36		UAA(*)	133	1.56
	UUG(L)	68	0.64		UAG(*)	85	1

Ile	CUU(L)	124	1.17	His	CAU(H)	101	0.98
	CUC(L)	109	1.03		CAC(H)	106	1.02
	CUA(L)	128	1.21	Gln	CAA(Q)	140	1.38
	CUG(L)	63	0.59		CAG(Q)	63	0.62
	AUU(I)	154	1.21	Asn	AAU(N)	144	1.1
	AUC(I)	100	0.79		AAC(N)	117	0.9
Met	AUA(M)	110	1.26	Lys	AAA(K)	169	1.33
	AUG(M)	65	0.74		AAG(K)	85	0.67
Val	GUU(V)	68	1.34	Asp	GAU(D)	70	0.99
	GUC(V)	38	0.75		GAC(D)	71	1.01
	GUA(V)	67	1.32	Glu	GAA(E)	93	1.27
	GUG(V)	30	0.59		GAG(E)	53	0.73
Ser	UCU(S)	103	1.32	Cys	UGU(C)	45	0.84
	UCC(S)	93	1.19		UGC(C)	62	1.16
	UCA(S)	100	1.28	Trp	UGA(W)	107	1.31
	UCG(S)	35	0.45		UGG(W)	56	0.69
Pro	CCU(P)	135	1.13	Arg	CGU(R)	30	0.77
	CCC(P)	155	1.3		CGC(R)	43	1.11
	CCA(P)	126	1.06		CGA(R)	48	1.24
	CCG(P)	60	0.5		CGG(R)	34	0.88
Thr	ACU(T)	101	1.04	Ser	AGU(S)	53	0.68
	ACC(T)	110	1.14		AGC(S)	83	1.07
	ACA(T)	132	1.36		AGA(*)	68	0.8
	ACG(T)	44	0.45		AGG(*)	55	0.65
Ala	GCU(A)	54	0.77	Gly	GGU(G)	52	0.81
	GCC(A)	122	1.74		GGC(G)	69	1.07
	GCA(A)	82	1.17		GGA(G)	73	1.14
	GCG(A)	23	0.33		GGG(G)	63	0.98

2.3.2 Start codon and stop codon

In *A. mengyangensis*, most of the PCGs initiate translation with the start codon ATG. The only exception is the COX1 gene, which uses GTG as its start codon. For stop codons, ATP8, COX1, ND6, and ND4L utilize TAA as the complete stop codon. ND1 and ND5 employ TAG as their complete stop codon. In contrast, COX2, COX3, ND2, ND3, ND4, ATP6, and CYTB use incomplete stop codons, specifically TA- and T--.

2.3.3 Synonymous codon preference

Synonymous codon preference (SCP) arises from the differential selection of translationally equivalent codons, driven by mutational drift and functional optimization

pressures. Generally, higher gene expression levels are associated with stronger SCP. To explore codon usage preferences in the mitochondrial genome of *A. mengyangensis*, we calculated the codon adaptation index (CAI), effective number of codons (ENC), GC content, and GC content at the third codon position (GC3) using CodonW 1.4.2 software. These metrics provide insights into the evolutionary and functional constraints shaping codon usage in this species.

The effective number of codons (ENC) quantifies the extent to which codon usage deviates from random selection, with values typically ranging from 20 to 61. In the mitochondrial genome of *A. mengyangensis*, the ENC values for PCGs ranged from 36.93 to 57.39 (Table 5), with most values falling within the 45–50 range. This indicates a moderate degree of codon usage preference in the mitochondrial genome of *A. mengyangensis*. Codon adaptation index (Jin et al. 1832) values ranged from 0 to 1, with high values indicating high gene expression levels and pronounced SCP (Sharp & Li 1987). Among the protein-coding genes (PCGs) in *A. mengyangensis*, distinct codon usage patterns were observed among PCGs: while ATP8 and cytochrome oxidase subunits (COX1-3) displayed elevated synonymous codon preference (SCP >1.8) and high expression, ATP6 exhibited minimal bias (SCP <0.5) and lowest transcriptional activity (Table 5). Examination of third-position GC content (GC3) across all codons - excluding those encoding methionine, tryptophan, and stop signals - demonstrated consistently depressed GC3 levels (mean 22.4±3.1%) in all 13 PCGs, with G/C nucleotides occurring 2.3-fold less frequently than expected at wobble positions ($p<0.01$). This finding further supports the strong AT preference observed in the *A. mengyangensis* PCGs (Table 3).

2.4 tRNA and rRNA

The mitochondrial genome encodes two ribosomal RNA components: a highly conserved 16S rRNA (large subunit) and a more variable 12S rRNA (small subunit), with structural conservation exceeding sequence similarity. The 958 bp 12S rRNA locus, positioned between $tRNA^{Phe}$ and $tRNA^{Val}$, constitutes 5.71% of the mitogenome and displays balanced nucleotide composition (AT=50.57%).

In A. mengyangensis, there are 22 transfer RNA (tRNA) genes, with eight located on the light strand (L-strand) and 14 on the heavy strand (H-strand). The lengths of these tRNA genes range from 69 to 77 bp. Except for Leu and Ser, which each have two types of tRNAs, the remaining 18 tRNAs are represented by a single type. Among these, the tRNA with the anticodon $tRNA^{Ser\ (GCT)}$ (located on the L-strand) lacks the D-arm,

resulting in an incomplete secondary structure. However, the other 21 tRNAs can fold into the canonical cloverleaf secondary structure (Figure 4), a feature commonly observed in the mitochondrial genomes of other fish species (Broughton et al. 2001, Hwang et al. 2013). For tRNAs with atypical structures, such as those lacking the D-arm, normal functionality may depend on co-evolutionary interaction factors or post-transcriptional RNA editing processes. These mechanisms help ensure proper tRNA function despite structural deviations. (Masta& Boore 2004).

Table 5. Preference for codon usage of genes encoding proteins in mitochondrial genome of *A. mengyangensis*

	CAI	ENC	GC	GC3
ATP6	0.118	48.99	0.421	0.318
ATP8	0.191	55.09	0.374	0.313
COI	0.171	46.12	0.440	0.363
COII	0.186	57.39	0.433	0.368
COIII	0.191	45.60	0.463	0.407
CYTB	0.169	42.36	0.434	0.408
ND1	0.130	45.82	0.446	0.374
ND2	0.125	41.17	0.454	0.422
ND3	0.126	36.93	0.453	0.445
ND4	0.127	45.69	0.441	0.372
ND4L	0.131	48.74	0.456	0.375
ND5	0.158	51.06	0.418	0.403
ND6	0.132	50.06	0.462	0.421
PCGs	0.165	57.17	0.447	0.439

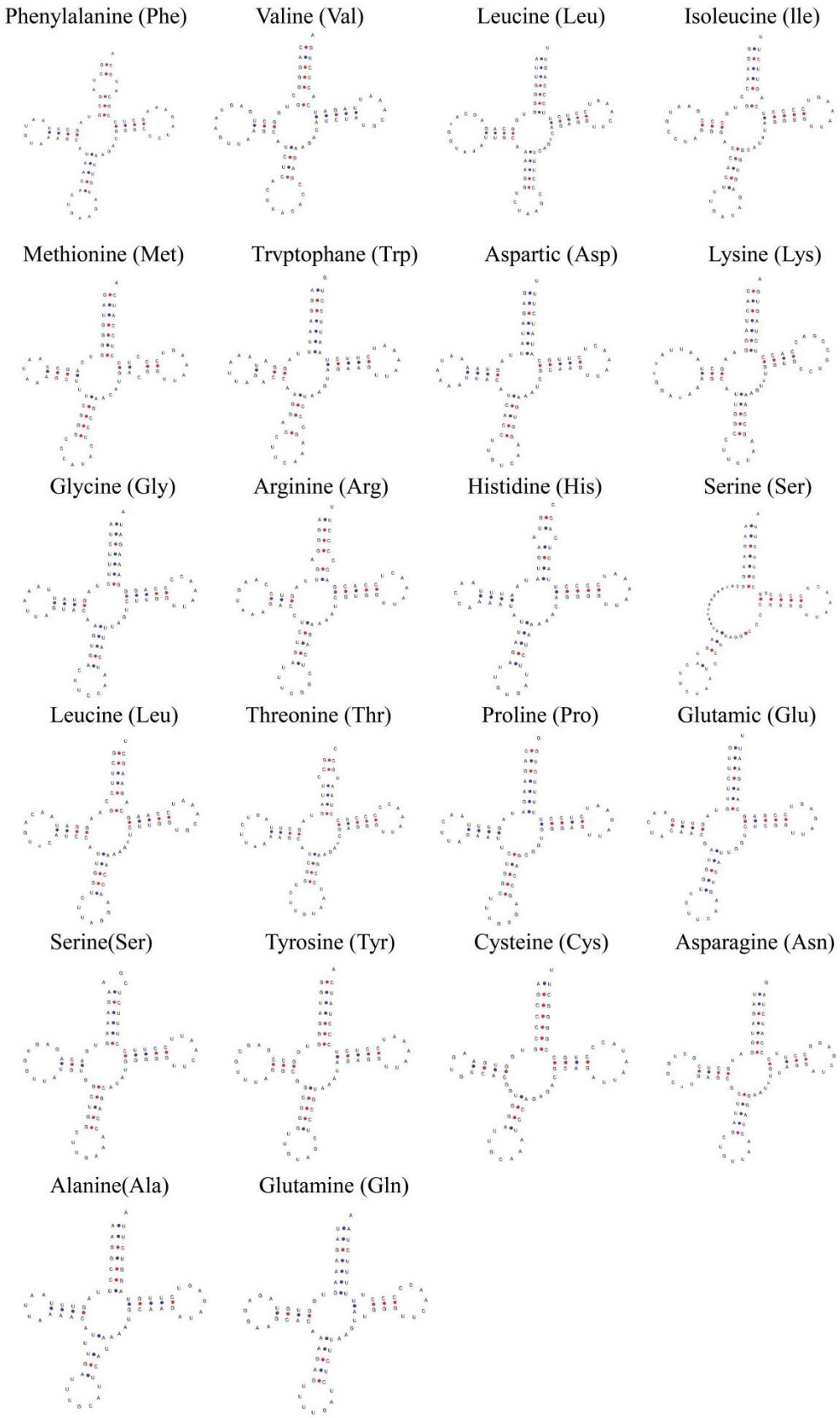

Figure 4 Secondary structure of the 22 tRNA genes of the mitochondrial genome of *A. mengyangensis*

2.5 Phylogeny and systematic of *A. mengyangensis*

In this study, the phylogenetic relationships of *A. mengyangensis* were investigated by constructing Maximum Likelihood (ML) and Bayesian (MrBayes) trees. These trees were built by concatenating 13 protein-coding genes (PCGs) and using *Pseudorasbora parva* as the outgroup. The best partitioning models for ML and Bayesian analyses were determined based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), respectively, using the most suitable nucleotide substitution models for 34 concatenated PCG sequences.

For the ML analysis, the optimal partitioning models were as follows:

HKY+F+I for cytb, cox1, cox2, and cox3;

GTR+F+I+G4 for atp6, atp8, nad1, nad2, nad3, nad4, nad4L, and nad6;

GTR+F+G4 for nad5.

The optimal partitioning models for the Bayesian analysis were slightly different, leading to minor variations in node support values. While the overall topological structure of the trees was consistent, some nodes showed differences in support rates or posterior probabilities. These discrepancies may arise from the differing algorithms used in ML and Bayesian analyses, which can influence the resulting phylogenetic trees.

The results support the conclusion that *Acheilognathus* and *Rhodeus* form monophyletic groups (Yang 2010). Specifically, *A. mengyangensis* was found to be most closely related to Acheilognathus chankaensis and *Acheilognathus macropterus*. Both the ML and MrBayes trees showed high node support rates for these relationships (70/0.761) (Figures 4 and 5, respectively).

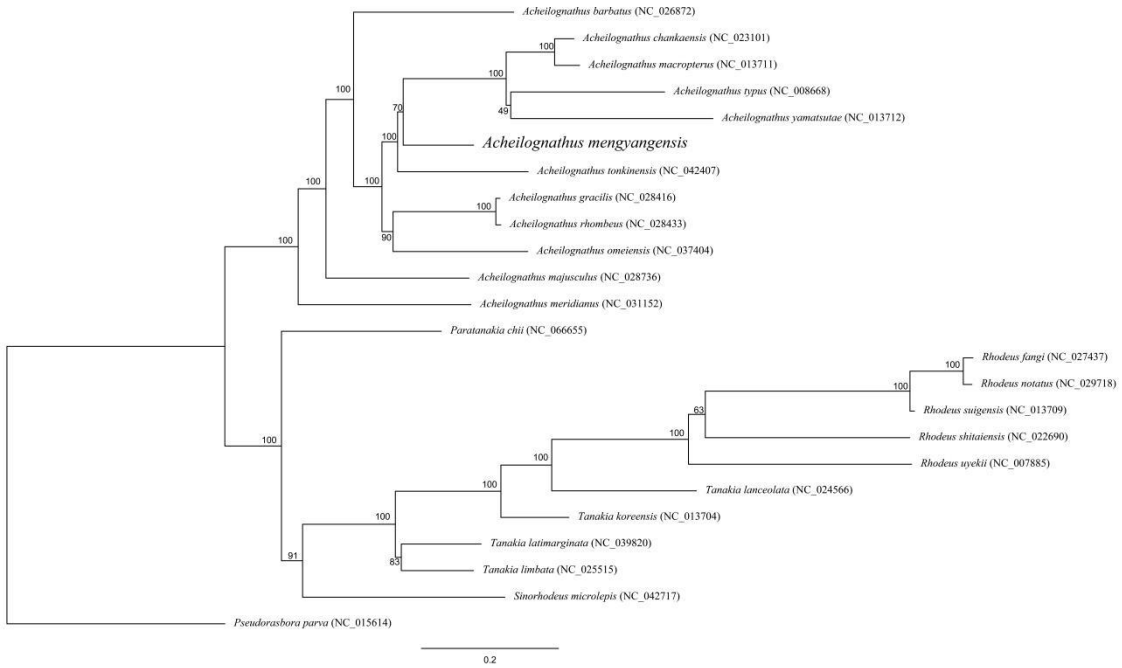

Figure 5. Phylogenetic trees derived from the Bayes approaches

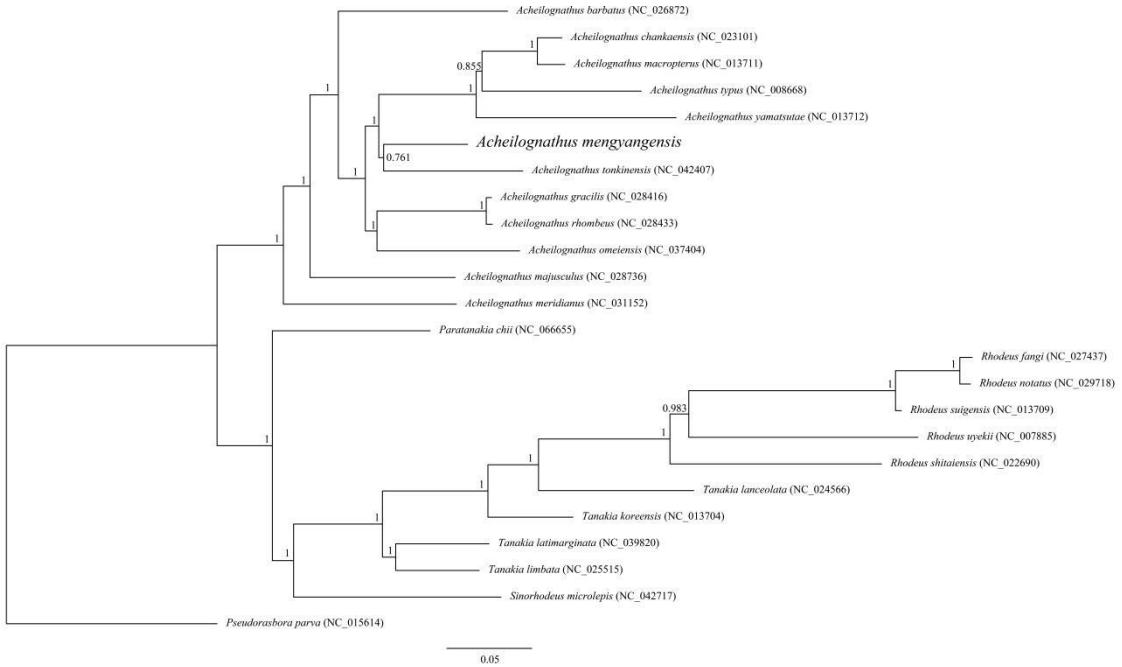

Figure 6 Phylogenetic trees derived from the Maximum-Likelihood (ML) approaches.

3 Discussion and Conclusions

We assembled and annotated the complete mitochondrial genome of *A. mengyangensis*, revealing a 16,779 bp circular DNA molecule - approximately the length reported by other Acheilognathinae species, such as *A. chankaensis* (16,774 bp), A. rhombeus (16,780 bp), and *A. omeiensis* (16,774 bp). Divergence in genome sizes likely reflects copy number variation of repetitive elements in regulatory regions. (Wang et al. 2020). Consistent with the typical mitochondrial genome structure of other teleost fish,

the *A. mengyangensis* mitochondrial genome comprises 13 protein-coding genes (PCGs), 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, a non-coding control region (D-loop), and an origin of light-strand replication (OL). The majority of these genes are located on the heavy strand (H-strand), with only ND6 and 8 tRNAs found on the light strand (L-strand). The gene arrangement aligns with the typical genetic organization observed in the Acheilognathinae subfamily (Hwang et al. 2012, Hwang et al. 2013, Zhu et al. 2021). Base composition analysis revealed significant nucleotide bias in *A. mengyangensis* mitogenome: AT content (56.7%) substantially exceeded GC content (43.3%), with particularly low guanine representation (16.9%). (Table 3). This pattern is consistent with the base composition observed in the mitochondrial genomes of most other teleost fishes. (Perna & Kocher 1995).

Based on the analysis of synonymous codon usage frequency in the *A. mengyangensis* mitochondrial genome, we observed that the relative synonymous codon usage (RSCU) values for UUU, AUU, and UUA were greater than 1 (Figure 3). This indicates a preference for these codons in the mitochondrial genome of *A. mengyangensis*.

In this study, we reconstructed the phylogenetic relationships of *A. mengyangensis* by concatenating 13 PCGs. The phylogenetic results revealed topological differences compared to previous studies, which can be attributed to variations in the choice of outgroups, comparative species, molecular markers, and individual gene sequences. These factors highlight the complexity of phylogenetic analysis and the need for comprehensive approaches to achieve more accurate and consistent results (Yang 2010, Chang et al. 2014, Cheng et al. 2014, Miyake et al. 2021). Due to the limited research currently available on *A. mengyangensis*, it is challenging to conduct a more comprehensive comparative analysis of this species. For *A. mengyangensis* itself, the results of the reconstructed phylogenetic tree indicate that it is most closely related to *Acheilognathus chankaensis* and *Acheilognathus macropterus*. The node support rates for this relationship, as determined by Maximum Likelihood (ML) and MrBayes analyses, are 0.761 and 70, respectively (Figure 5 and 6).

Compared to previous studies, this research utilized protein-coding genes (PCGs) from the mitochondrial genome to construct a phylogenetic tree, providing a more comprehensive set of gene loci and a gene tree that more accurately reflects the species tree. However, as technology continues to advance, discrepancies between gene trees and species trees have become increasingly apparent. These discrepancies may arise from

factors such as incomplete lineage sorting, hybridization, and gene flow, which can complicate the interpretation of phylogenetic relationships (Zhang et al. 2024). To establish more precise and reliable phylogenetic relationships, it is essential to incorporate supplementary data sources, such as next-generation sequencing (NGS) and reduced-representation genomic data. These advanced approaches can provide a broader and more detailed genetic perspective, helping to address discrepancies between gene trees and species trees and improving the accuracy of phylogenetic inferences.

We present the complete mitochondrial genome sequence and associated characteristics of *A. mengyangensis*. A detailed analysis was conducted on the gene structure, RNA secondary structure, D-loop region, and base composition. These findings not only enrich the mitochondrial genome database for *Acheilognathus* but also provide valuable molecular and genetic insights for species conservation, molecular identification, and evolutionary studies within the Acheilognathinae subfamily.

Data resources

The genome sequence data that support the findings of this study are openly available in (https://db.cngb.org/search/?q=CNP0006862) under the CNSA project accession CNP0006862.

Ethical approval

All experimental procedures received prior approval from the Institutional Animal Ethics Committee of Henan Normal University (Approval No.: HNAREC-202X-XXX) and were conducted in strict accordance with:

National Standard GB/T 35892-2018 requirements

ARRIVE 2.0 reporting guidelines

International Council for Laboratory Animal Science (ICLAS) ethical benchmarks

Continuous monitoring ensured full compliance throughout the study duration.

Disclosure statement

The authors state that there are no potential conflicts of interest.

Author Contributions

ZHOU CJ and YU JH: led the conceptualization and workflow design efforts; LI ZR and YU JH: spearheaded data acquisition and analysis. All authors collaborated in drafting the paper and approving the final version, with contributions ranked in order of significance.

Funding

This work was supported by the following funding bodies: the Henan Province Department of Science and Technology (182102110046, 222102110294), and the Training Plan of Young Excellent Teachers in Colleges and Universities of Henan Province (2019GGJS063).

ORCID

Chuanjiang Zhou: https://orcid.org/0000-0002-6433-737X

References

Agassiz L. 1832. Untersuchungen über die fossilen Süsswasser-Fische der tertiären Formationen.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler P. F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69(2): 313-319. doi:10.1016/j.ympev.2012.08.023

Bleeker P. 1860. Conspectus systematis cyprinorum. Nat. Tijdschr. Ned. Ind. 20: 421-441.

Broughton R. E., Milam J. E., Roe B. A. J. G. r. 2001. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Research 11(11): 1958-1967. doi:10.1101/gr.156801

Chang C.H., F. Li, Shao K.T., Lin Y.S., Morosawa T., Kim S., Koo H., Kim W., Lee J.S., He S. J. M. P., 2014. Phylogenetic relationships of Acheilognathidae (Cypriniformes: Cyprinoidea) as revealed from evidence of both nuclear and mitochondrial gene sequence variation: evidence for necessary taxonomic revision in the family and the identification of cryptic species. Molecular Phylogenetics and Evolution 81: 182-194. doi:10.1016/j.ympev.2014.08.026

Cheng P., Yu D., Liu S., Tang Q., Liu H. J. Z. s. 2014. Molecular phylogeny and conservation priorities of the subfamily Acheilognathinae (Teleostei: Cyprinidae). Zoolog Sci 31(5): 300-308. doi:10.2108/zs130069

Chen X 2011. The study of genetic diversity and molecular phylogenetic of Acheilognathinae. Doctoral dissertation Henan Normal University

Drummond A. J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7(1): 214. doi:10.1186/1471-2148-7-214

Guo X.H., Liu S.J., Liu Q., Liu Y. J. Y. C. x. b. A. G. S. 2004. New progresses on mitochondrial DNA in fish. Yi Chuan Xue Bao 31(9): 983-1000.ISBN: 0379-4172

Huelsenbeck J. P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8): 754-755. doi:10.1093/bioinformatics/17.8.754

Hwang D.S., Byeon H.K., Lee J.S. 2013. Complete mitochondrial genome of the freshwater sculpin *Cottus koreanus* (Scorpaeniformes, Cottidae). Mitochondrial DNA 24(5): 490-491. doi:10.3109/19401736.2013.770500

Hwang D. Y., Koim S. K., Ryu S. H., Hwang U. W. J. M. D. 2012. Complete mitochondrial genome of *Acheilognathus signifer* (Cypriniformes, Cyprinidae): Comparison of light-strand replication origins from six Acheilognathus species. Mitochondrial DNA 23(4): 306-307. doi:10.3109/19401736.2012.683180

Jin X., Cheng H., Li M., Zou H., Cai J., Amoah K., Li W., Wang G. J. P. I. 2024. Description of three new species of *Gyrodactylus* von Nordmann, 1832 (Monogenea: Gyrodactylidae) on bitterling fishes (Acheilognathinae) from China. Parasitology International 101(5):102893 doi: 10.1016/j.parint.2024.102893

Jordan D. S., Thompson W. F., Carnegie M. 1914. Record of the fishes obtained in Japan in 1911. Pittsburgh, Published by the authority of the Board of Trustees of the Carnegie Institute.

Kalyaanamoorthy S., Minh B. Q., Wong T. K. F., von Haeseler A., Jermiin L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14(6): 587-589. doi:10.1038/nmeth.4285

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33(7): 1870-1874. doi:10.1093/molbev/msw054

Lanfear R., Frandsen P. B., Wright A. M., Senfeld T., Calcott B. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Molecular Biology and Evolution 34(3): 772-773. doi:10.1093/molbev/msw260

Li F., Liao T.Y., Arai R., Zhao L. J. Z. 2017. *Sinorhodeus microlepis*, a new genus and species of bitterling from China (Teleostei: Cyprinidae: Acheilognathinae). Zootaxa 4353(1): 69–88-69–88. doi:10.11646/zootaxa.4353.1.4

Lowe T. M., Chan P. P. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44(W1): W54-W57. doi:10.1093/nar/gkw413

Masta S. E., Boore J. L. 2004. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. Molecular Biology and Evolution 21(5): 893-902. doi:10.1093/molbev/msh096

Miyake T., Nakajima J., Umemura K., Onikura N., Ueda T., Smith C., Kawamura K. 2021. Genetic diversification of the Kanehira bitterling inferred from mitochondrial DNA, with comments on the phylogenetic relationship with its sister species. J Fish Biol 99(5): 1677-1695. doi:10.1111/jfb.14876

Peden, John F. (2000). Analysis of codon usage, University of Nottingham, 226 pp.

Perna N. T., Kocher T. D. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41(3): 353-358. doi:10.1007/BF01215182

Reichard M., Polačik M., Tarkan A. S., Spence R., Gaygusuz Ö., Ercan E., Ondračková M., Smith C. (2010). The bitterling–mussel coevolutionary relationship in areas of recent and ancient sympatry. Evolution 64(10): 3047-3056. doi:10.1111/j.1558-5646.2010.01032.x

Ronquist F., Huelsenbeck J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19(12): 1572-1574. doi:10.1093/bioinformatics/btg180

Sambrook J., Russell D. W. 2001. Molecular Cloning: Ch. 8. In Vitro amplification of DNA by the polymerase chain reaction, Cold Spring Harbor Laboratory Press.

Sharp P. M., Li W.H. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research 15(3): 1281-1295. doi:10.1093/nar/15.3.1281

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21): 2688-2690. doi:10.1093/bioinformatics/btl446

Tang Y., Ma W., Chen X., Nie G. and Zhou C. J. (2024). Four new complete mitochondrial genomes of Gobioninae fishes (Teleostei: Cyprinidae) and their phylogenetic implications. PeerJ 12: e16632. doi:10.7717/peerj.16632

Wang, I.C., Lin H.D., Liang C.M., Huang C.C., Wang R.D., Yang J.Q., Wang W.K. J. Z. 2020. Complete mitochondrial genome of the freshwater fish *Onychostoma lepturum* (Teleostei, Cyprinidae): genome characterization and phylogenetic analysis. Zookeys 1005: 57. doi:10.3897/zookeys.1005.57592

Yang Q 2010. Taxonomic review and molecular phylogenetic analysis of the subfamily Acheilognathinae (Teleostei: Cypriniformes). Doctoral dissertation Huazhong Agricultural University

Yi, W., Reichard M., Rücklin M., Richardson M. (2024). Parasitic fish embryos do a "front-flip" on the yolk to resist expulsion from the host. Proceedings of the National Academy of Sciences 121(9): e2310082121. doi: 10.1073/pnas.2310082121

Zhang D., Gao F., Jakovlić I., Zou H., Zhang J., Li W. X., 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(1): 348-355. doi:10.1111/1755-0998.13096

Zhang Z., Liu G., Li M. 2024. Incomplete lineage sorting and gene flow within *Allium* (Amayllidaceae). Molecular Phylogenetics and Evolution 195: 108054. doi:10.1016/j.ympev.2024.108054

Zhu L., Che X., Liu X., Wang X., Wang J., Cheng G., Chen X., Chen X. 2021. Complete mitochondrial genome of *Acheilognathus hypselonotus* Bleekers (Cypriniformes: Cyprinidae) in China's Dianshan Lake. Mitochondrial DNA Part B 6(12): 3367-3368. doi:10.1080/23802359.2021.1997109