Pakistan J. Zool., vol. 50(6), pp 2141-2151, 2018. **DOI**: http://dx.doi.org/10.17582/journal.pjz/2018.50.6.2141.2151

Complete Mitochondrial DNA Genome Sequences for Two Lineages in *Coilia mystus* (Clupeiformes: Engraulididae): Mitogenomic Perspective on the Phylogenetic Relationships of Genus *Coilia*

Ai Guo[1,2,3], Jiaguang Xiao[4], Binbin Shan[4], Tianxiang Gao[5] and Yongdong Zhou[3,*]

[1]*College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China*
[2]*State Key Laboratory of Satellite Ocean Environment Dynamics, The Second Institute of Oceanography, State Oceanic Administration, Hangzhou 310012, China*
[3]*Marine Fisheries Research Institute of Zhejiang Province, Zhoushan 316021, China*
[4]*Fishery College, Ocean University of China, Qingdao 266003, China*
[5]*Fishery College, Zhejiang Ocean University, Zhoushan 316022, China*

Ai Guo and Jiaguang Xiao contributed equally to this study.

Article Information
Received 17 June 2017
Revised 24 August 2017
Accepted 02 October 2017
Available online 06 September 2018

Authors' Contribution
AG, JX, TG and YZ conceived and designed the experiments. JX and BS performed the experiments and analyzed the data. AG, TG and YZ contributed reagents/materials. AG, JX and BS wrote the article. TG and YZ proofread the manuscript and approved the final version.

Key words
***Coilia mystus*, Mitochondrial genome, Genome structure, Genetic divergence, Phylogeny.**

ABSTRACT

To better understand the genetic diversity and phylogeny of *Coilia*, the complete mitochondrial genomes of two lineages in *Coilia mystus* were compared. They were all typical circular double stranded DNA molecule with 17075 bp in *C. mystus* N and 16964 bp in *C. mystus* S, respectively, containing the standard metazoan set of 22 transfer RNA genes, 2 ribosomal RNA genes, 13 protein-coding genes and non-coding regions. The mitogenomes of *C. mystus* N and *C. mystus* S shared the identical structural organization and gene arrangement with those of other *Coilia* fishes. Both lineages of *C. mystus* showed similar features in not only the strand-specific asymmetry of nucleotide composition, but also the codon usage of genes. Whereas a significant variation among *Coilia* species was detected in length of the control region, mainly caused by the variable number of tandem repeats. Phylogenetic analysis was performed based on 13 concatenated mitochondrial protein-coding genes from 8 *Coilia* mitochondrial genomes. The results supported that *C. lindmani* at first clustered with *C. nasus* and *C. grayii* which had close relationships (d=0.028), then clustered with *C. mystus* which exhibited obvious genetic differentiation between *C. mystus* N and *C. mystus* S (d=0.083). *C. reynaldi* was at the basal part of the trees, and showed obvious genetic differentiations with other *Coilia* species (d>0.19). Our results suggested that the north and south lineages of *C. mystus* could be genetically distinct as different species.

INTRODUCTION

As an important functional eukaryotic organelle, mitochondria become highly economized and conserved (Boore, 1999). Most genes of the original endosymbiont are lost or have been transferred to the nucleus after the Genome Reductive Evolution (GRE) process (Andersson and Kurland, 1998; Khachane *et al*., 2007; Ghiselli *et al*., 2013; Kannan *et al*., 2014). In animals, mitochondrial DNA (mtDNA) is a compact double-stranded circular molecule that typically contains 13 protein-coding genes (PCGs), 2 ribosomal RNA genes and 22 transfer RNAs. Moreover, a large noncoding control region (CR) commonly related to the initiation of transcription and replication sequences usually presents in CRs (Boore 1999; Breton *et al*., 2014). The mtDNA genomes have played a significant role in the studies of population genetics and reconstruction of phylogeny due to their intrinsic properties (*i.e.,* fast evolutionary rate, high information content, maternal inheritance and lack of recombination) (Simon *et al*., 2006).

Fish of the genus *Coilia* are small to moderate in size and primarily inhabit in coastal waters or estuaries in the Indo-West Pacific region, and some of them can tolerate low salinities in freshwater (Wongratana, 1980). Taxonomical debates and genetic divergence of *Coilia*, especially *Coilia nasus* and *Coilia mystus* in China, are always the hot topics of research. At present, it is

* Corresponding author: zyd511@126.com
0030-9923/2018/0006-2141 $ 9.00/0

commonly accepted that *Coilia brachygnathus* and *Coilia nasus taihuensis* are synonym of *C. nasus* (Yang *et al*., 2010), because small genetic divergence has been found between diadromous and freshwater ecotypes of *C. nasus* (Cheng and Lu, 2005; Yang *et al*., 2010). *C. mystus* which is widely distributed along the coast of China is a short distance migration fish (Whitehead, 1985). Initially, three clades of *C. mystus* were described as the Yangtze River group, the Minjiang River group and the Pearl River group, respectively (Cheng *et al*., 2008; Yan *et al*., 2009). Recently, the updated conclusion revealed *C. mystus* should be divided into two clades: *C. mystus* Northern populations (N) and *C. mystus* Southern populations (S) (Yang, 2012). The *C. mystus* (N) which was distributed in the north of the coast of Ningbo in China, was just the above-mentioned Yangtze River group; the *C. mystus* (S) which widely lived in the rest of southern coast of China consist of the Minjiang River group and the Pearl River group as previously described (Yang, 2012). Actually, there is no significant differentiation in morphological characters between *C. mystus* (N) and *C. mystus* (S), excepting on the number of gill rakers and vertebra (Yang, 2012).

In this study, we compared the complete mitogenomes of five *Coilia* species to explore the divergence among different species and lineages of *C. mystus*. Phylogenetic analysis was conducted based on the protein coding genes of the mitogenomes using the Maximum Likelihood (ML) and the Bayesian Inference (BI) methods to gain insight on the genetic diversity between two lineages of *C. mystus* and its phylogenetic status in genus *Coilia*.

MATERIALS AND METHODS

Samples collection and sequencing

Two specimens of *C. mystus*, namely, *C. mystus* Northern populations (N) and *C. mystus* Southern populations (S), collected from Shanghai and Daya Bay in the coast of China, respectively, were sequenced in this study. The identifications of *C. mystus* N and *C. mystus* S were based on the results of Yang (2012). Moreover, six mitochondrial genome sequences were downloaded from GenBank for genetic comparation and phylogenetic reconstruction, including *Coilia grayii*, *Coilia lindmani*, *Coilia reynaldi* and three specimens of *C. nasus* (Table II, Lavoue *et al*., 2010; Zhang *et al*., 2016, a, b; Zhao *et al*., 2016). Total genomic DNA was isolated from the muscle tissue by proteinase K digestion followed by the standard phenol/chloroform method (Sambrook and Russell, 2001).

The complete mitogenomes of *C. mystus* were amplified using a long-PCR technique (Chang *et al*., 1994; Miya and Nishida, 1999) and primer-walking method. Both PCR primers were designed referring to congeneric mitogenome sequence available in GenBank and implemented in Primer Premier 5.0 software (PRIMER Biosoft International). Contiguous segments overlapped by at least 50 bp to ascertain the accuracy of sequencing. Long-PCR and normal PCR reactions were performed in a TAKARA thermal cycler following the standard procedures (Cheng *et al*., 2012). All fragments were sequenced on ABI Prism 3730 from both strands after purification.

Sequence annotation and analysis

Sequences were edited and aligned using DNASTAR software (DNASTAR, Madison, WI, USA) with default parameters, and refined manually. Locations for protein-coding genes and rRNAs were identified by DOGMA (Wyman *et al*., 2004). Gene predictions were further improved by comparing DNA sequences with those of *C. nasus*, *C. lindmani* and *C. reynaldi*. The base composition and codon usage of the 13 protein-coding genes were analyzed with Mega 5.0 (Tamura *et al*., 2011). Nucleotide composition skew analysis was carried out with the formulas AT-skew = [A - T] / [A + T] and GC-skew = [G - C] / [G + C], respectively (Perna and Kocher, 1995). Most tRNA genes were identified by their proposed cloverleaf secondary structures using web-based tRNAscan-SE software (http://lowelab.ucsc.edu/tRNAscan-SE/). The remaining tRNA genes unidentified were determined by inspecting sequences for tRNA-like secondary structures and anticodons. Control region was identified by comparing with the homologous sequences. Variable number of tandem repeats (VNTRs) were detected using Tandem Repeats Finder (Benson, 1999).

Phylogenetic reconstruction

Phylogenetic analysis was performed based on 13 concatenated mitochondrial PCGs from 8 *Coilia* mitochondrial genomes, *Engraulis japonicas* (NC_003097) (Inoue *et al*., 2001a) and *Engraulis encrasicolus* (NC_009581) (Lavoue *et al*., 2007) were used as the out-groups. Nucleotide sequences from all 10 mitochondrial PCGs were edited and aligned using ClustalX 1.83 under default settings (Thompson *et al*., 1997), stop codons and gaps were removed and finally concatenated into a sequence matrix (11,391 sites in length).

The phylogenetic trees were built using two approaches including maximum-likelihood (ML) analysis by PAUP* 4.0 (Swofford, 2002) and a partitioned Bayesian inference (BI) analysis by Mrbayes 3.12 (Huelsenbeck and Ronquist, 2001). Substitution model selection was conducted by a comparison of Akaike Information Criterion (AIC) scores (Akaike, 1974) with jModelTest 2 (Darriba *et al*., 2012). GTR + G model was chosen as the best-fitting model for

ML analyses and the node reliability was estimated after 1000 bootstrap replicates. For the Bayesian procedure, a set of optimal models was selected for different positions (GTR+I+G for the 1st and the 3rd positions, GTR+I for the 2nd position). Four Markov chains were run for 1,000,000 generations by sampling the trees every 1000 generations. After the first 2500 trees (25%) were discarded as burn-in, the 50% majority rule consensus tree and the Bayesian posterior probabilities (BPP) were estimated using the remaining 7500 sampled trees.

Table I.- Characteristics of the mitochondrial genomes of *C. mystus* N and *C. mystus* S.

Gene/Region	Position		Size (bp) Nucleotide (N/S)	Amino acid	Gap[b] (N/S)	Codon		Strand
	From(N/S)	To (N/S)				Start	Stop[a]	
$tRNA^{Phe}$	1	69	69		0			H
12S rRNA	70	1022	953		0			H
$tRNA^{Val}$	1023	1094	72		0			H
16S rRNA	1095	2786	1692		0			H
$tRNA^{Leu}$(UUR)	2787	2861	75		0			H
ND1	2862	3836	975	324	1	ATG	TAA	H
$tRNA^{Ile}$	3838	3909	72		-1			H
$tRNA^{Gln}$	3909	3979	71		-1			L
$tRNA^{Met}$	3979	4047	69		0			H
ND2	4048	5093	1046	348	0	ATG	TA	H
$tRNA^{Trp}$	5094	5163	70		2			H
$tRNA^{Ala}$	5166	5234	69		1			L
$tRNA^{Asn}$	5236	5308	73		0			L
O_L	5309	5342/5341	34/33		-3			H
$tRNA^{Cys}$	5340/5339	5405/5404	66		0			L
$tRNA^{Tyr}$	5406/5405	5476/5475	71		1			L
CO I	5478/5477	7022/7021	1545	514	0	GTG	TAA	H
$tRNA^{Ser}$(UCN)	7023/7022	7093/7092	71		5			L
$tRNA^{Asp}$	7099/7098	7166/7165	68		11			H
CO II	7178/7177	7868/7867	691	230	0	ATG	T	H
$tRNA^{Lys}$	7869/7868	7941/7940	73		1			H
ATPase8	7943/7942	8110/8109	168	55	-10	ATG	TAA	H
ATPase6	8101/8100	8783/8782	683	227	0	ATG	TA	H
CO III	8784/8783	9568/9567	785	261	0	ATG	TA	H
$tRNA^{Gly}$	9569/9568	9639/9638	71		0			H
ND3	9640/9639	9989/9988	350	116	-1	ATG	TA	H
$tRNA^{Arg}$	9989/9988	10057/10056	69		0			H
ND4L	10058/10057	10354/10353	297	98	-7	ATG	TAA	H
ND4	10348/10347	11728/11727	1381	460	0	ATG	T	H
$tRNA^{His}$	11729/11728	11797/11796	69		1			H
$tRNA^{Ser}$(AGY)	11799/11798	11865/11864	67		0			H
$tRNA^{Leu}$(CUN)	11866/11865	11937/11936	72		0			H
ND5	11938/11937	13773/13772	1836	611	-4	ATG	TAA	H
ND6	13770/13769	14291/14290	522	173	1	ATG	TAA	L
$tRNA^{Glu}$	14293/14292	14361/14360	69		5/4			L
Cyt *b*	14367/14365	15507/15505	1141	380	0	ATG	T	H
$tRNA^{Thr}$	15508/15506	15577/15575	70		-1			H
$tRNA^{Pro}$	15577/15575	15647/15645	71		0			L
Control region	15648/15646	17075/16964	1428/1319					H

a, TA and T represent incomplete stop codons; b, positive numbers correspond to the nucleotides separating adjacent genes, negative numbers indicate overlapping nucleotides.

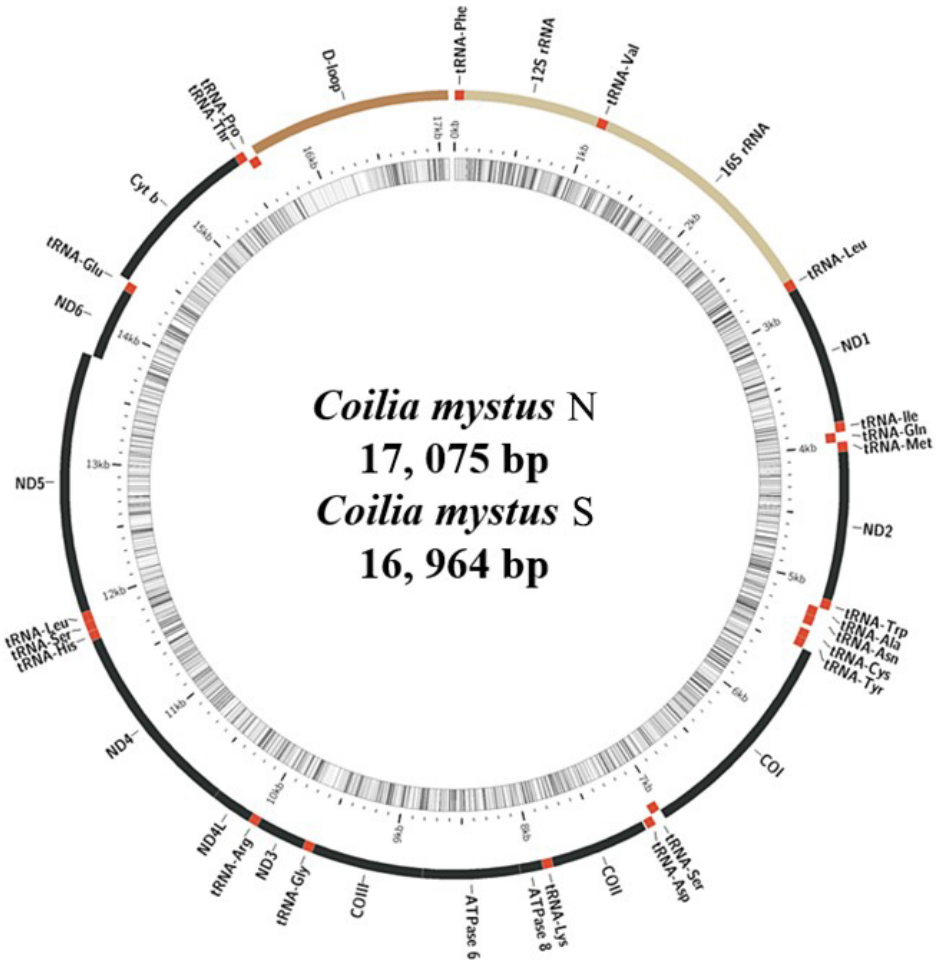

Fig. 1. Gene organization for the mitochondrial genomes of the *Coilia mystus* N and *Coilia mystus* S. (*C. mystus* N and *C. mystus* S shared the same structural organization and gene arrangement, so only gene organization of *C. mystus* N were shown).

RESULTS AND DISCUSSION

General features of the mitogenomes

The complete mitogenome sequences of *C. mystus* N and *C. mystus* S were 17, 075 bp and 16,964 bp in length (Fig. 1, Table I). Actually, *C. mystus* N had the longest mitogenome and the shortest was that of *C. lindmani* (16835 bp) in all sequenced *Coilia* species (Table II). Length differences were primarily the result of variation in intergenic nucleotides and the control region, predominately, variable number tandem repeats detected in control regions of all eight mitogenomes of *Coilia* fishes. The mitogenomes of *C. mystus* N and *C. mystus* S had the same structural organization and gene arrangement with those of other *Coilia* fishes (Fig. 1). Both *C. mystus* N and *C. mystus* S exhibited a clear strand-specific bias in composition, most of the genes were encoded on the heavy stand (H-strand) except for ND6 and 8 tRNAs (Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu and Pro), which were oriented to on the light strand (L-strand). These features also could be found in other vertebrates (Lavoue *et al*., 2010, 2013; Li *et al*., 2013; Liu *et al*., 2017; Shan *et al*., 2014; Teacher *et al*., 2012).

The AT content of mitogenomes varied among *Coilia* taxa (Table II) from 56.3% (*C. reynaldi*) to 57.8% (*C. mystus* N). Overall structural compositions of *Coilia*, this specific bias in nucleotide composition commonly exhibited excepting the first codon positions of PCGs like other bony fishes (Mabuchi *et al*., 2007; Cheng *et al*., 2012; Xiao, 2015). For GC- / AT-skews analysis, all *Coilia* species were displayed a positive AT-skew from 0.069 (*C. mystus* N) to 0.094 (*C. reynaldi*), especially in rRNA genes (≥0.255), and a strong negative GC-skew from -0.257 (*C. mystus* N) to -0.271 (*C. nasus* PYH and *C. grayii*), especially in PCGs (≤-0.274) (Supplementary Table I).

Table II.- Genomic characteristics of eight *Coilia* mitochondrial genomes.

Species	GenBank accession	Genome		13 protein-coding genes					2 rRNA		22 tRNA		CR	
		L (bp)	A+T (%)	L (bp)*	A+T (%)				L (bp)	A+T (%)	L (bp)	A+T (%)	L (bp)	A+T (%)
					All pos.	1st codon pos.	2nd codon pos.	3rd codon pos.						
C. mystus S	MF363002	16964	57.6	11391	57.5	49.7	58.5	64.4	2645	54.0	1548	55.7	1319	67.6
C. mystus N	MF363003	17075	57.8	11391	57.9	49.7	58.4	65.6	2645	53.7	1548	55.0	1428	67.4
C. nasus AS	KM257636	16900	57.5	11391	57.5	49.5	58.5	64.6	2641	54.0	1549	55.3	1252	66.7
C. nasus NB	KM363243	16896	57.5	11391	57.6	49.5	58.5	64.7	2641	54.0	1549	55.3	1252	66.3
C. nasus PYH	KM276661	16896	57.5	11391	57.5	49.5	58.5	64.6	2641	54.1	1549	55.5	1252	67.1
C. reynaldi	NC014276	17064	56.3	11391	55.6	48.5	58.4	59.8	2641	53.7	1550	54.5	1419	69.2
C. grayii	KP317088	16851	57.2	11391	57.4	49.3	58.4	64.5	2640	53.7	1549	55.4	1208	65.2
C. lindmani	NC014271	16835	56.7	11391	56.9	49.0	58.4	63.2	2640	53.8	1550	55.1	1191	63.7

*excluding the stop codons; L, length; pos., position.

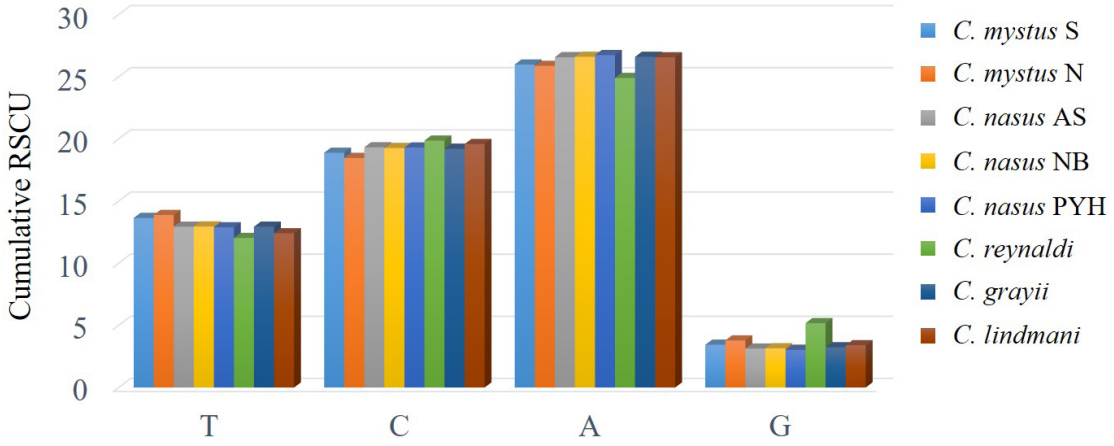

Fig. 2. H-strand comparison of frequencies of codons ending with the same nucleotide. Values on the y-axis represent the sum of relative synonymous codon usage (RSCU) values of codons ending with the same nucleotide across all codon families (x-axis).

Protein-coding genes and Codon usage

All 13 protein-coding genes found in other vertebrates were also present in *C. mystus* N and *C. mystus* S as well as other *Coilia* species, including three subunits of the cytochrome c oxidase (COI-III), seven subunits of the NADH ubiquinone oxidoreductase complex (ND1-6, ND4L), one subunit of the ubiquinol cytochrome *b* oxidoreductase complex (Cyt *b*), and two subunit of ATP synthases (ATP6 and ATP8) (Fig. 1, Table I). Without regard to the stop codons, the length of 13 protein-coding genes were exactly same (11391 bp) among all the *Coilia* species (Table II). The mitogenome of *C. mystus* N and *C. mystus* S exhibited a canonical genetic code shared by most vertebrates (Inoue *et al*., 2001b; Ramakodi *et al*., 2015). An orthodox initiation codon ATG was used for all protein-coding genes except for COI starting with GTG (Table I). A diverse pattern of codon usage within stop codons consisting complete stop codon and incomplete stop codon was showed in Table I, which seems to be a common tendency in fish mitogenomes (Cheng *et al*., 2012; Li *et al*., 2013).

The mitochondrial genomes of *C. mystus* N and *C. mystus* S consisted of 3797 codons excluding stop codons. Codons for Leucine possessed the highest percentage value of 15.39% and 16.01%, which may be related to the function of chondriosome of encoding many transmembrane proteins (Gillespie *et al*., 2006), while those for Cysteine were the least represented with a percentage value of 0.82% and 0.81%, as observed in other *Coilia* fishes (Supplementary Fig. S1; Zhang, 2015). Comparing the relative synonymous codon usage (RSCU) of *Coilia*, it showed the similar tendency in H-strand that A-terminal was in great abundance and G-terminal was extremely destitute (Fig. 2). The underlying mechanism

responsible for the strand bias has been generally interpreted as evidence of an asymmetrical directional mutation pressure associated with replication processes when one strand remains transiently in a single-stranded state, making it more vulnerable to DNA damage (Perna and Kocher, 1995).

Transfer and ribosomal RNA genes

The complete set of 22 tRNA genes which were usually found in metazoans was present in *C. mystus* N and *C. mystus* S mitogenome. In addition, 14 tRNA genes were transcribed on the H-strand, whereas other 8 tRNA genes were oriented to the L-strand (Table I). Although G-U wobbles and other atypical pairings were constantly detected, typical cloverleaf secondary structures were shown for 21 tRNA genes with the exception of tRNASer (AGY) who lacked the recognizable DHU stem found in almost all vertebrate mitogenomes (Li, 2014; Miya and Nishida, 1999; Zhang *et al*., 2016a). Stem mismatches were common for mitochondrial tRNA genes and were probably repaired via a post-transcriptional editing process (Lavrov *et al*., 2000).

The two rRNA genes which were identified in *C. mystus* N and *C. mystus* S with no length variation, a small (12S) subunit of rRNA comprising 953 bp long and a large (16S) subunit of 1692 bp, were also similar size to their counterparts in other *Coilia* mitogenomes (Table II). But subtle differences also existed, 16S rRNA in *C. nasus* and *C. reynaldi* were 1688 bp, 16S rRNA in *C. grayii* and C. *lindmani* were 1687 bp (Zhang, 2015).

Control region and sequence repeats

Mitochondrial control region is the major non-coding segment in the vertebrate mitogenome which is AT-rich and highly variable by a faster rate of evolution (Sbisa *et al*., 1997). This region frequently lead to the length variation in the whole mitogenome, whereas its control elements related to regulatory functions are known to be highly conserved (Arnason and Rand, 1992; Broughton and Dowling, 1994; Lunt *et al*., 1998). The control region of both *C. mystus* was located between the $tRNA^{Pro}$ and $tRNA^{Phe}$ genes, and determined to be 1428 bp in *C. mystus* N and 1319 bp in *C. mystus* S. Actually, *C. mystus* N had the longest control region and the shortest was that of *C. lindmani* (1191 bp) in all sequenced *Coilia* species (Table II). The variable number tandem repeats (VNTRs) which was common in genus *Coilia* fishes (Zhu *et al*., 2008), was the main reason for different length among *Coilia*. The VNTRs were found always presenting after about 200bp at the start of control region and having different type among *Coilia* species. It is worth mentioning that it showed different type and size repeats between two linages of *C. mystus*, however it shared the identical repeat and times among three *C. nasus* (Fig. 3, Table III). The different type and size of repeat unit between two lineages of *C. mystus* provided strong evidence to support that the two clades may represent different species. The VNTRs were believed resulting from illegitimate elongation model (Buroker *et al*., 1990).

The structures of control region for *Coilia* species were identical except for the length change. By comparing with the recognition sites among *Coilia* species, three conservative domains were detected: the extended termination associated sequence domain (TAS), the central conserved sequence block domain and the conserved sequence block domain (Fig. 3). The motif-TACAT in TAS was easily found, and so was the complementary TAS (cTAS) motif-ATGTA in the extended termination associated sequence domain which may function as a recognition site for terminating synthesis of heavy strand (Clayton, 1991). Moreover, the VNTRs were in this domain. In the central conserved sequence block domain, CSB-F which distinguished the central conserved sequence block domain from the termination associated sequence domain, CSB-E which was always characterized by the GTGGG box and CSB-D were recognized. The conserved sequence block domain which was thought to be involved in positioning RNA polymerase both for transcription and for priming replication (Shadel and Clayton, 1997) was characterized by CSB1, CSB2 and CSB3.

Table III.- The variable number tandem repeats (VNTRs) among genus *Coilia*.

Species	Repeat unit	Repeat unit length (bp)	Repeat time
C. mystus S	ATATTATGCATTATATTACATATATATTATGGTATAGTAC	40	6
C. mystus N	GTACATACTATGCATTATATTACATATATTATGGTATA	38	9
C. nasus	ATATTACATATATTATGGTATAGTACATACTATGTATT	38	5
C. reynaldi	TACATATATGATATAGTACATACTATGCATTATATTACA	39	12
C. grayii	TATATTACATATATTATGGTATAGTACATACTATGTAT	38	4
C. lindmani	CATATTATGTATTATATTACATATATTATGGTATAGTA	38	3

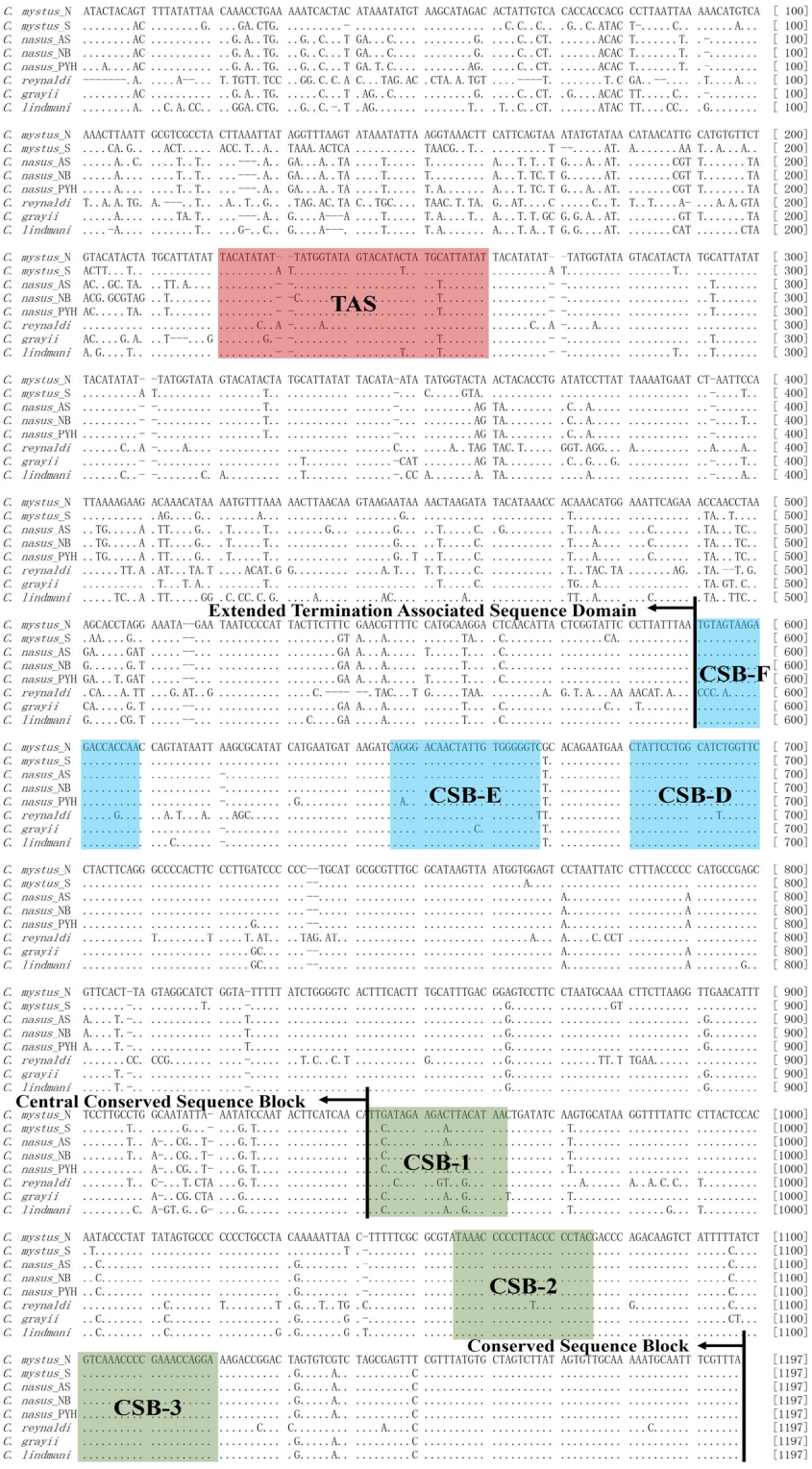

Fig. 3. Complete nucleotide sequences of mitochondrial control region of *Coilia*. Termination associated sequence (TAS), central conserved sequences (CSB-F, -E, -D) and sequence blocks (CSB-1, -2, -3). VNTRs were embellished in the same times for sequences alignment.

Table IV.- Matrix of net average genetic distances based on 13 protein-coding genes sequences among genus *Coilia*.

	C. mystus S	*C. mystus* N	*C. nasus*	*C. grayii*	*C. lindmani*
C. mystus N	0.083				
C. nasus	0.083	0.080			
C. grayii	0.086	0.082	0.028		
C. lindmani	0.095	0.093	0.051	0.055	
C. reynaldi	0.195	0.194	0.195	0.197	0.198

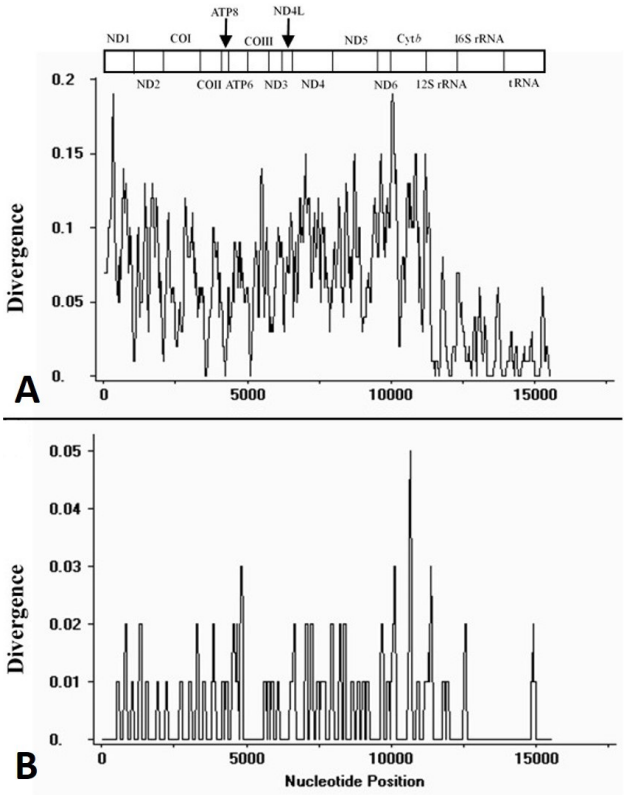

Fig. 4. Plot of divergences among mtDNA sequences excluding non-coding regions. The sliding window analysis calculates the divergence of *C. mystus* N and *C. mystus* S (A), *C. mystus* N and *C. mystus* (KJ710625) (B), respectively. The bar at the top illustrates the position of protein-coding genes, rRNAs and tRNAs are represented as black boxes.

Divergence of the two C. mystus *mitogenomes*

A sliding window analysis was used to quantify genome-wide nucleotide variability between *C. mystus* N and *C. mystus* S, contrastively, a same analysis between *C. mystus* N and *C. mystus* (a downloaded sequence from GenBank, KJ710625) was also performed (Fig. 4). The divergences between *C. mystus* N and *C. mystus* S were much higher (Fig. 4A) when compared with the nucleotide divergences between the *C. mystus* N and *C. mystus* (KJ710625) which was also collected from Yangtze estuary (<0.05, Fig. 4B). The higher divergences were showed in ND1, ND4, ND5, ND6 and Cyt *b*, while lower divergences were observed in tRNA or rRNA. From the nucleotide divergence analysis, it could demonstrate that there were obvious genetic divergences between north lineage and south lineage in *C. mystus*.

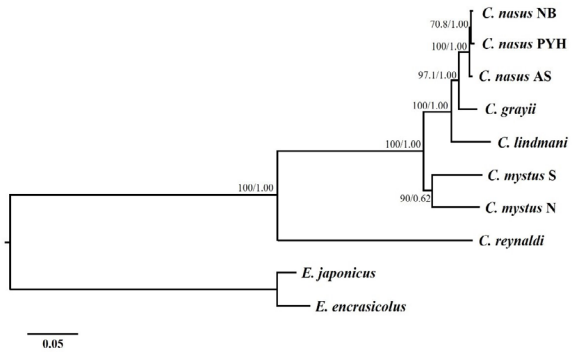

Fig. 5. Inferred phylogenetic relationships among *Coilia* based on the concatenated nucleotide sequences of 13 mitochondrial protein-coding genes using maximum likelihood (ML) and Bayesian Inference (BI). Numbers on branches are bootstrap percentages and Bayesian posterior probabilities. *Engraulis japonicas* (NC_003097) and *Engraulis encrasicolus* (NC_009581) are used as outgroups.

Phylogenetic analysis

ML and BI analyses are done with the concatenated nucleotide data containing 8 *Coilia* sequences and two outgroup taxa. The topological relationships of two phylogenetic analyses remained consistent, and all analyses provided high bootstrap support values for all internodes (Fig. 5). The resultant topology showed *C. lindmani* at first clustered with *C. nasus* and *C. grayii* which had close relationships. Then they clustered with the clade *C. mystus* which exhibited obvious genetic differentiation between *C. mystus* N and *C. mystus* S. *C. reynaldi* was at the basal part of the trees, and was showed obvious genetic differentiations with other *Coilia* species (Table IV). In accordance with the phylogenetic analyses,

the genetic distance between *C. mystus* N and *C. mystus* S also revealed the obvious genetic differentiation (0.083), even higher than the genetic distance among *C. nasus*, *C. grayii* and *C. lindmani* (0.028~0.055). In conclusion, the mitogenomic data supported the deep intraspecific differentiations in *C. mystus*, revealing that *C. mystus* N and *C. mystus* S might be the different species.

For the first time, our study investigates phylogenetic relationships within genus *Coilia* based on the complete mitogenome sequences. The tree topologies obtained in the present study were identical regardless of the analytic method used, and were statistically well supported by high bootstrap and posterior probability values. Therefore, the result suggested that the north and south lineages of *C. mystus* could be genetically distinct as different species. The divergences of north and south lineages of *C. mystus* was probably on account of the drastic changes of sea level on ice age of Pleistocene (Yang, 2012; Liu *et al*., 2012). The *C. mystus* was separated in two shelters. Moreover, *C. mystus* was short-distance migration species (Whitehead, 1985). This life habit might have great influences on the divergences of north and south lineages of *C. mystus*.

ACKNOWLEDGMENTS

We are grateful to Ms. Nan Zhang for her constructive suggestions on the manuscript and technical assistance and we gratefully thank Cunyin Dou, Yuting Wang and Yancui Chen for collecting the specimens. This work was supported by Special Fund for Agro-scientific Research in the Public Interest (No. 201303048 and No. 201303050) and the Scientific Startup Foundation of Zhejiang Ocean University (Q1505). We declare that there are no conflicts of interests regarding the publication of this article and authors are solely responsible for the contents and writing of the paper.

Supplementary material

There is supplementary material associated with this article. Access the material online at: http://dx.doi.org/10.17582/journal.pjz/2018.50.6.2141.2151

Statement of conflict of interest

Authors have declared no conflict of interest.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Contr*., **19**: 716-723. https://doi.org/10.1109/TAC.1974.1100705

Andersson, S.G. and Kurland, C.G., 1998. Reductive evolution of resident genomes. *Trends Microbiol*., **6**: 263-268. https://doi.org/10.1016/S0966-842X(98)01312-2

Arnason, E. and Rand, D.M., 1992. Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic Cod, *Gadus morhus*. *Genetics*, **132**: 211-200.

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

Benson, G., 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucl. Acids Res*., **27**: 573-580. https://doi.org/10.1093/nar/27.2.573

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T. and Passamonti, M., 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet*., **30**: 555-564. https://doi.org/10.1016/j.tig.2014.09.002

Broughton, R.E. and Dowling, T.E., 1994. Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics*, **138**: 179-190.

Buroker, N.E., Brown, J.R. and Gilbert, T.A., 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics*, **124**: 157-163.

Chang, Y.C., Huang, F.L. and Lo, T.B., 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. *J. mol. Evol*., **38**: 138-155. https://doi.org/10.1007/BF00166161

Cheng, J., Ma, G.Q., Song, N. and Gao, T.X., 2012. Complete mitochondrial genome sequence of bighead croaker *Collichthys niveatus* (Perciformes, Sciaenidae): A mitogenomic perspective on the phylogenetic relationships of Pseudosciaeniae. *Gene*, **491**: 210-223. https://doi.org/10.1016/j.gene.2011.09.020

Cheng, Q.Q. and Lu, D.R., 2005. PCR-RFLP analysis of cytochrome *b* gene does not support *Coilia ectenes taihuensis* being a subspecies of *Coilia ectenes*. *J. Genet*., **84**: 307-310. https://doi.org/10.1007/BF02715801

Cheng, Q.Q., Ma, C.Y., Zhuang, P., Sha, Z.X., Lu, X. and Miao, J., 2008. Genetic structure and evolution characters in three populations of *Coilia mystus* based on cytochrome *b* gene segment sequence of mitochondrial DNA. *J. Fish. China*, **32**: 1-7.

Clayton, D.A., 1991. Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends Biochem. Sci*., **16**: 107-111. https://doi.org/10.1016/0968-0004(91)90043-U

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods*, **9**: 772. https://doi.org/10.1038/nmeth.2109

Ghiselli, F., Milani, L., Guerra, D., Chang, P.L., Breton, S., Nuzhdin, S.V. and Passamonti, M., 2013. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biol. Evol.*, **5**: 1535-1554. https://doi.org/10.1093/gbe/evt112

Gillespie, J.J., Johnston, J.S., Cannone, J.J., Gutell, R.R., 2006. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization, and retrotransposable elements. *Insect mol. Biol.*, **15**: 657-686. https://doi.org/10.1111/j.1365-2583.2006.00689.x

Huelsenbeck, J.P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**: 754-755. https://doi.org/10.1093/bioinformatics/17.8.754

Kannan, S., Rogozin, I.B. and Koonin, E.V., 2014. MitoCOGs: Clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. *BMC Evol. Biol.*, **14**: 237. https://doi.org/10.1186/s12862-014-0237-5

Khachane, A.N., Timmis, K.N. and Martins dos Santos V.A., 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Mol. Biol. Evol.*, **24**: 449-456. https://doi.org/10.1093/molbev/msl174

Lavoue, S., Miya, M., Saitoh, K., Ishiguro, N.B. and Nishida, M., 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Mol. phylogenet. Evol.*, **43**: 1096-1105. https://doi.org/10.1016/j.ympev.2006.09.018

Lavoue, S., Miya, M. and Nishida, M., 2010. Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. *Mol. phylogenet. Evol.*, **56**: 480-485. https://doi.org/10.1016/j.ympev.2009.11.022

Lavoue, S., Miya, M., Musikasinthorn, P., Chen, W.J. and Nishida, M., 2013. Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). *PLoS One*, **8**: e56485. https://doi.org/10.1371/journal.pone.0056485

Lavrov, D.V., Brown, W.M. and Boore, J.L., 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc. natl. Acad. Sci. U.S.A.*, **97**: 13738-13742. https://doi.org/10.1073/pnas.250402997

Li, N., Zhang, Z.H., Zhao, L.L. and Gao, T.X., 2013. Complete mitochondrial DNA sequence of the Pacific sand lance *Ammodytes hexapterus* (Perciformes: Ammodytidae): Mitogenomic erspective on the distinction of *Ammodytes hexapterus* and *Ammodytes personatus*. *Mitochond. DNA*, **24**: 463-465. https://doi.org/10.3109/19401736.2013.766180

Li, N., 2014. *Molecular phylogeography of sand lance and the red stingray in the Northwestern Pacific*. PhD thesis, Ocean University of China, Qingdao, China.

Liu, M., Lin, L.S., Gao, T.X. and Grant, W.S., 2012. What maintains the central north Pacific genetic discontinuity in Pacific herring? *PLoS One*, **7**: e50340. https://doi.org/10.1371/journal.pone.0050340

Liu, Z.S., Song, N., Yanagimoto, T., Han, Z.Q., Shui, B.N. and Gao, T.X., 2017. Complete mitochondrial genome of three fish species (Perciformes: Amblyopinae): genome description and phylogenetic relationships. *Pakistan J. Zool.,* **49**: 111-120. http://doi.org/10.17582/journal.pjz/2017.49.1.111.120

Lunt, D.H., Whipple, L.E. and Hyman, B.C., 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): Utility and problems in molecular ecology. *Mol. Ecol.*, **7**: 1441-1455. https://doi.org/10.1046/j.1365-294x.1998.00495.x

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M., 2001a. Complete mitochondrial DNA sequence of the Japanese anchovy *Engraulis japonicas*. *Fish. Sci.*, **67**: 828-835. https://doi.org/10.1046/j.1444-2906.2001.00329.x

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M., 2001b. 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

Mabuchi, K., Miya, M., Azuma, Y. and Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evol. Biol.*, **7**: 12. https://doi.org/10.1186/1471-2148-7-10

Miya, M. and Nishida, M., 1999. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): First example of transfer RNA gene rearrangements in bony fishes. *Mar. Biotechnol.*, **1**: 416-426. https://doi.org/10.1007/PL00011798

Perna, N.T. and Kocher, T.D., 1995. Patterns of nucleotide composition at four folded generate sites of animal mitochondrial genomes. *J. mol. Evol.*, **41**: 353-358. https://doi.org/10.1007/BF00186547

Ramakodi, M.P., Singh, B., Wells, J.D., Guerrero, F. and Ray, D.A., 2015. A 454 sequencing approach to dipteran mitochondrial genome research. *Genomics*, **105**: 53-60. https://doi.org/10.1016/j.ygeno.2014.10.014

Sambrook, J. and Russell, D.W., 2001. *Molecular cloning: A laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, New York, US.

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

Shadel, G.S. and Clayton, D.A., 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem*., **66**: 409-435. https://doi.org/10.1146/annurev.biochem.66.1.409

Shan, B.B., Zhao, L.L., Gao, T.X. and Lu, H.S., 2014. The complete mitochondrial genome of *Nibea coibor* (Perciformes, Sciaenidae). *Mitochond. DNA*, **17**: 1-2. https://doi.org/10.3109/19401736.2014.958726

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B. and Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst*., **37**: 545-579. https://doi.org/10.1146/annurev.ecolsys.37.091305.110018

Swofford, D.L., 2002. *PAUP* 4.0: Phylogenetic analysis using parsimony*. Sinauer Associates, Inc., Sunderland, MA.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol*., **28**: 2731-2739. https://doi.org/10.1093/molbev/msr121

Teacher, A.G., Andre, C., Merila, J. and Wheat, C.W., 2012. Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evol. Biol*., **12**: 248. https://doi.org/10.1186/1471-2148-12-248

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignments aided by quality analysis tools. *Nucl. Acids Res*., **24**: 4876-4882. https://doi.org/10.1093/nar/25.24.4876

Whitehead, P.J.P., 1985. FAO Species Catalogue, Vol. 7. Clupeoid Fishes of the World (Suborder Clupeoidei). Part 1: Chirocentridae, Clupeidae and Pristigasteridae. *FAO Fish. Synop*., **125**: 305-579.

Wongratana, T., 1980. *Systematics of clupeoid fishes of the Indo-Pacific region*. PhD thesis, University of London, London, UK.

Wyman, S.K., Jansen, R.K. and Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**: 3252-3255. https://doi.org/10.1093/bioinformatics/bth352

Xiao, J.G., 2015. *The complete mitochondrial genomes and phylogenetic analysis of Sillago species*. Master thesis, Ocean University of China, Qingdao, China.

Yan, X.L., Tang, W.Q. and Yang, J.Q., 2009. Population genetic structure of tapertail anchovy (*Coilia mystus*) in coastal waters of southeast China based on mtDNA control region sequences. *Biodiv. Sci*., **17**: 143-150. https://doi.org/10.3724/SP.J.1003.2009.08286

Yang, Q.L., Han, Z.Q., Sun, D.R., Xie, S.G. and Lin, L.S., 2010. Genetics and phylogeny of genus *Coilia*, in China based on AFLP markers. *Chin. J. Oceanol. Limnol*., **28**: 795-801. https://doi.org/10.1007/s00343-010-9093-3

Yang, Q.L., 2012. *Phylogenetic analysis of genus* Coilia *in China and molecular phylogeography of* C. nasus *and* C. mystus. PhD thesis, Ocean University of China, Qingdao, China.

Zhang, N., 2015. *Studies on complete mitochondrial genomes of genus* Coilia *and* Lota lota. Master thesis, Ocean University of China, Qingdao, China.

Zhang, Z.H., Zhang, N., Liu, M. and Gao, T.X., 2016. The complete mitochondrial genome of *Coilia grayii* (Clupeiformes: Engraulidae). *Mitochond. DNA*, **27**: 3175. https://doi.org/10.3109/19401736.2014.926508

Zhang, N., Song, N., Han, Z.Q. and Gao, T.X., 2016a. The complete mitochondrial genome of *Coilia nasus* (Clupeiformes: Engraulidae) from the coast of Ningbo in China. *Mitochond. DNA*, **27**: 1. https://doi.org/10.1080/24701394.2016.1248429

Zhang, N., Song, N. and Gao, T.X., 2016b. The complete mitochondrial genome of *Coilia nasus* (Clupeiformes: Engraulidae) from Ariake Sea. *Mitochond. DNA*, **27**: 1518. https://doi.org/10.3109/19401736.2014.926508

Zhao, L.L., Zhao, Y.H., Zhang, N., Gao, T.X. and Zhang, Z.H., 2016. The complete mitogenome of *Coilia nasus* (Clupeiformes: Engraulidae) from Poyang Lake. *Mitochond. DNA*, **27**: 1608.

Zhu, T.J., Yang, J.Q. and Tang, W.Q., 2008. MtDNA control region sequence structure of the genus *Coilia* in Yangtze River estuary. *J. Shanghai Ocean Univ*., **17**: 152-157. https://doi.org/10.1007/s11741-008-0213-1