

Genetic diversity of mitochondrial control region (D-Loop) polymorphisms in *Coilia ectenes taihuensis* inhabiting Taihu Lake, China

J.J. Zhang¹, J.R. Duan², Y.F. Zhou², J.Y. Peng¹ and D.A. Fang^{1,2}

¹Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, Jiangsu, China

²Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi, Jiangsu, China

Corresponding author: D.A. Fang
E-mail: fangda@ffrc.cn

Genet. Mol. Res. 16 (1): gmr16019457

Received October 19, 2016

Accepted December 9, 2016

Published March 16, 2017

DOI <http://dx.doi.org/10.4238/gmr16019457>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. *Coilia ectenes* is a commercially important fishery species in China. *C. ectenes taihuensis* is an endemic and dominant species found in Taihu Lake of China. When compared with *C. ectenes*, *C. ectenes taihuensis* lacks anadromous behavior, and can independently grow and reproduce in Taihu Lake. In this study, the mitochondrial DNA control region (D-loop) sequences were employed to investigate the genetic diversity and population structure of *C. ectenes taihuensis*. Sixty-eight individuals collected from 4 localities in Taihu Lake were examined. Results indicated that in the 887-bp D-loop region, seventy-seven (8.68%) sites were variant, contributing to 53 distinct haplotypes. Although the population haplotype diversity ($H_d = 0.971$ to 1.000) was generally high, the nucleotide diversity ($\pi = 0.616$ to 0.731%) was relatively low among the 4 populations. Additionally, the genetic distances ranged from 0.62 to 0.74% within the populations and from 0.67 to 0.74% between the populations. The neighbor-joining tree indicated that a distinct distribution of phylogenetic structure existed among haplotypes. Analysis of molecular variance and F_{ST}

statistics suggested that a divergence existed among populations in 4 localities, indicating that gene communication might have occurred among those populations. Furthermore, neutral tests and analysis of mismatch distribution reflected that *C. ectenes taihuensis* might have undergone a population expansion during the evolution process. Our study showed the population genetic diversity and structure of *C. ectenes taihuensis*. Results from this study might be helpful in the development and protection of fishery resource within the localities in Taihu Lake in future.

Key words: *C. ectenes taihuensis*; Mitochondrial DNA; Genetic diversity; Aquaculture

INTRODUCTION

Coilia ectenes taihuensis is one of the ecotypes of *C. ectenes*, which belongs to the order Clupeiformes of the family Engraulidae (Ma et al., 2010). It is an endemic ecological species with some distinctive characters. It is mainly found in Taihu Lake, possess a short body length, attains sexual maturity at an earlier age, and subsequently losses its anadromous behavior (Tang et al., 2007). In recent years, many studies have been conducted on *C. ectenes taihuensis* through the use of mitochondrial DNA D-loop control region (mtDNA D-loop).

The mtDNA D-loop is the only non-coding segment which exists in the vertebrate mitochondrial genome (Zardoya and Meyer, 1997). This segment is the most variable part of the mtDNA, and evolves three to five times faster than the rest of the mitochondrial genome (Li et al., 2015). Therefore, it is widely used as a genetic marker to assess the origin, phylogenesis, and intraspecific genetic differentiation of animal species (Rasmussen and Morrissey, 2008). For example, it has previously been used to identify a number of fish species, such as *Acipenser transmontanus* (Brown et al., 1992), *Gadus morhua* (Arnason and Rand, 1992), *Dicentrarchus labrax* (Cecconi et al., 1995), *Alosa sapidissima* (Bentzen et al., 1988), and *Coregonus lavaretus* (Brzuzan, 2000), and offers a lot of genetic information.

In the past few years, new developments have taken place in the field of genetic diversity and population structure of *C. ectenes taihuensis*. The genetic diversity and differentiation between *C. ectenes* and *C. ectenes taihuensis* have been investigated at the mtDNA level (Cheng et al., 2006). Previous studies have demonstrated that the genetic parameters and heredity diversities were relatively lower in *C. ectenes taihuensis* than *C. ectenes* (Xu et al., 2012). They also showed that *C. ectenes taihuensis* is a geographical population of *C. ectenes*, and their differences have not risen to a subspecies level (Xu et al., 2011). However, another study indicated that *C. ectenes taihuensis* has length variations among individuals, and it should be considered a subspecies of *C. ectenes* (Tang et al., 2007). Besides, their haplotype and nucleotide diversities were calculated, and their genetic diversity was thought to be a result of their hybridization and variation (Li et al., 2015).

Fisheries survey results indicated that *C. ectenes taihuensis* has become an absolutely dominant population in Taihu Lake. Another major problem that affects the genetic diversity of *C. ectenes taihuensis* is the fishing pressure since it hinders its natural development and evolution. Therefore, it is important to analyze the genetic diversity and structure of *C. ectenes taihuensis* through the use of D-loop genetic marker since it is very effective.

In this study, 4 populations of *C. ectenes taihuensis* were sampled from four different localities within Taihu Lake. The mtDNA D-loop was used as molecular marker to assess the levels of genetic diversity and differentiation within and between 4 populations. The results of this study might be helpful to the resource protection and sustainable utilization of *C. ectenes taihuensis*.

MATERIAL AND METHODS

Sample collection

Using drift net, 68 individuals of *C. ectenes taihuensis* were collected from four different geographical locations in Taihu Lake: Dalei (DL), Gonghu (GH), Manshan (MS), and Meiliangwan (MLW) (Figure 1). The tail fins of all individuals were transported to the laboratory in 95% ethanol for DNA extraction.

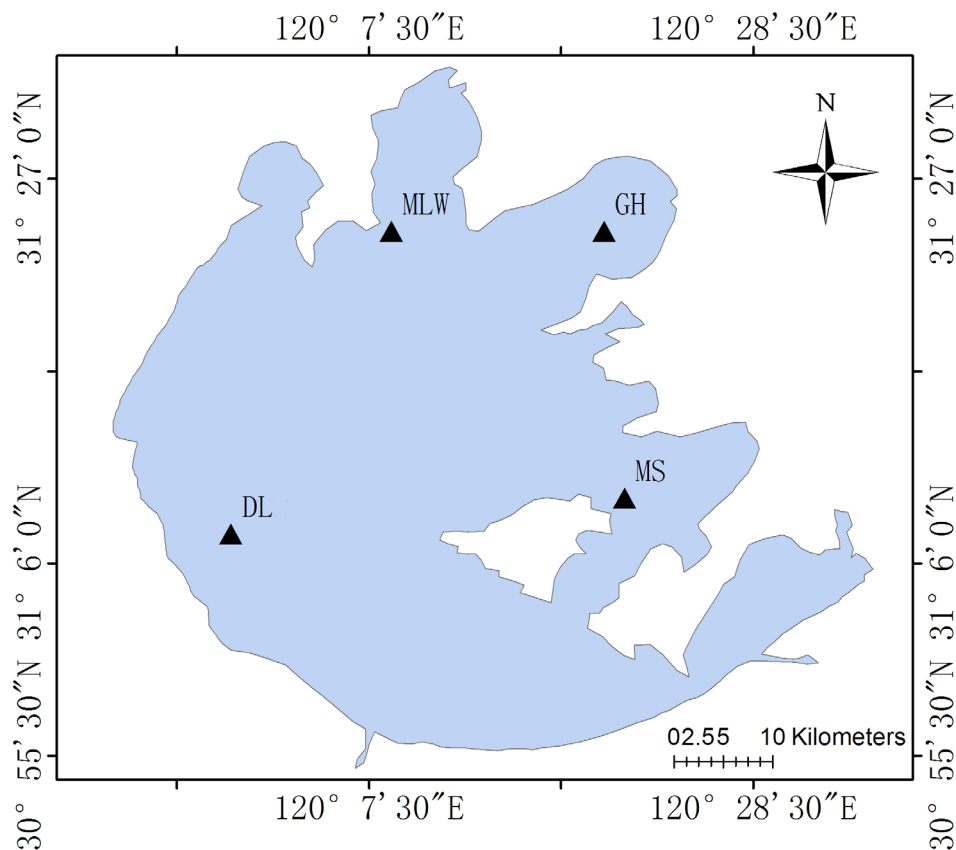


Figure 1. Sampling sites of *Coilia ectenes taihuensis* in Lake Taihu. DL = Dalei; GH = Gonghu; MS = Manshan; and MLW = Meiliangwan.

DNA extraction, amplification, and sequencing

Total genomic DNA was extracted from tissues (tailfins) using the E.Z.N.A. SE Blood DNA Kit (Omega Biotek Store, USA) according to the manufacturer's instructions. The D-loop region was amplified using the primers D-Loop-F (5'-CTAACTCCCAAAGCTAGAATTCT-3') and D-Loop-R (5'-ATCTTAGCATCTTCAGTG-3'). Polymerase chain reaction (PCR) was performed in a 20- μ L reaction volume comprising 10 μ L Premix Taq (1.25 U rTaq polymerase, 0.4 mM each dNTP, 4 mM Mg^{2+} ; TaKaRa, Dalian, China), 0.5 μ L each primer, 2 μ L volume containing 50 ng genomic DNA as template and 7 μ L of sterilized water. PCR amplifications were performed under the following conditions: initial denaturation at 98°C for 1 min, followed by 30 cycles of denaturation at 95°C for 10 s, annealing at 55°C for 15 s, and extension at 72°C for 6 min. Verification of the amplified fragment length was done with a DNA marker DL2000 (TaKaRa) by electrophoresing 3 μ L of each PCR product on 1.2% agarose gel containing Andy Safe DNA Gel Stain (TaKaRa). Both D-Loop-F and D-Loop-R primers were used to sequence the amplified D-loop region.

Data analysis

The D-loop sequences were edited and aligned using Clustal X 1.81 software. The haplotype diversity (H_d), nucleotide diversity (π), fixation index (F_{ST}), Tajima's D, and Fu's F_s test values (Fu, 1997) were calculated followed by the determination of Nei values (Saitou and Nei, 1987) using DnaSP5.10 software (Tajima, 1989). Phylogenetic relationships among haplotypes were reconstructed using the neighbor-joining (NJ) method (Saitou and Nei, 1987). Genetic distances were generated for the phylogenetic reconstruction using models of substitution by MEGA 5.0. A hierarchical analysis of molecular variance (AMOVA) was carried out using the Arlequin 3.0 software (Schneider and Excoffier, 1999). According to different water areas (DL, GH, MS, and MLW), AMOVA was carried out by partitioning the molecular variance into two levels: (1) between groups, and (2) within groups.

Method limitations

During the experiment, it was noted that in the DNA extraction process, all tailfins samples should be completely digested before proceeding to the next extraction step. Care should be observed when adding genomic DNA template into PCR amplification mixtures when the amount of DNA template is limited. In addition, the suitability of the primer must be determined after the PCR amplification. Correct technique should be utilized to attain accurate results.

RESULTS AND DISCUSSION

Sequence variation

The 887-bp partial D-loop sequences of 68 *C. ectenes taihuensis* individuals collected from 4 localities in Taihu Lake were analyzed. The average basic nucleotide compositions of all sequences were as follows: 31.5% T, 15.5% C, 31.2% A, and 21.8% G. In all sequences, the A + T content (62.7%) was found to be relatively higher than the C + G content (37.3%), confirming that the D-loop region is an A-T rich region of the mitochondrial genome (Brown

et al., 1986), and the statistics of this study was in accordance with other fish species analysis. Besides, there were 77 variant sites, 27 of which were parsimony-informative sites for species identification.

Genetic diversity

A total of 53 haplotypes were identified in the 68 samples (Tables 1 and 2). Among these, 45 haplotypes were unique and 8 haplotypes were shared by two or more populations. Among the four studied populations, the MS population showed the highest haplotype number (16). Besides, the lowest haplotype number was observed in the GH population (14). The haplotypes H8 and H27 were the most common since they were found in four individuals, which accounted for 5.88% of all individuals. The other common haplotypes shared between two or more populations were H6 (4.41%), H10 (2.94%), H16 (2.94%), H17 (2.94%), H25 (4.41%), and H38 (2.94%) as shown in Table 1. The haplotype H25 was found in 3 populations, and was likely to be an ancestral haplotype.

The haplotype diversity was very high for all populations, with values ranging from 0.971 (GH) to 1.000 (MS), whereas the nucleotide diversity was relatively low, ranging from 0.00616 (GH) to 0.00731 (MLW). The MLW population showed the highest nucleotide diversity level (0.00731). According to Table 1, the GH population shared its haplotypes with the DL, MS, and MLW populations, whereas the MS population shared its haplotypes with the DL, GH, and MLW populations. This result might have arisen on the basis of the geographical distribution of populations, given the average distance between the GH/MS population and the other 3 populations were relatively shorter, which might offer more opportunities for gene interaction with the other populations (Figure 1). Besides, the haplotype diversity was generally high for all four *C. ectenes taihuensis* populations, while the values of nucleotide diversity were relatively low. The nucleotide diversity might be influenced by the other factors such as ecological habits and life history of *C. ectenes taihuensis*; it likes clustering and is good at population diffusion. Because of these characteristics, *C. ectenes taihuensis* is able to expand its new habitats more easily and form populations more quickly, leading to gene communication within individuals and creation of new haplotypes. Conversely, an increase in nucleotide diversity requires a longer time for accumulation, whereas strong influences will result in sudden changes (Li et al., 2015).

Phylogenetic analysis and population structure

The genetic distance ranged from 0.62 (GH) to 0.74% (MS) within the populations and from 0.67 to 0.74% between populations (Table 3). The genetic variation within the population was the lowest in GH and the highest in MS population. These results provided strong evidence to support that the GH population showed the lowest and the MS population showed the highest nucleotide diversity value. Furthermore, the values of genetic distance between the populations showed that the DL population significantly differentiated from the other populations. As seen on the map (Figure 1), the DL population is far from the other three populations. When comparing the level of genetic variations among the four populations, it was obvious that the individuals from MLW population showed relatively higher genetic varieties. Besides, most pairwise F_{ST} estimates between the populations were high, indicating the variations among the four populations (Table 3). These findings suggested that the genetic differentiation was distributed in *C. ectenes taihuensis* populations.

Table 1. Distribution of mtDNA D-Loop haplotypes of *Coilia ectenes taihuensis*.

Haplotypes	DL	GH	MS	MLW	Total
Hap1	1	-	-	-	1
Hap2	1	-	-	-	1
Hap3	1	-	-	-	1
Hap4	1	-	-	-	1
Hap5	1	-	-	-	1
Hap6	2	-	-	1	3
Hap7	1	-	-	-	1
Hap8	1	3	-	-	4
Hap9	1	-	-	-	1
Hap10	1	-	-	1	2
Hap11	1	-	-	-	1
Hap12	1	-	-	-	1
Hap13	1	-	-	-	1
Hap14	1	-	-	-	1
Hap15	1	-	-	-	1
Hap16	-	1	-	1	2
Hap17	-	1	-	1	2
Hap18	-	1	-	-	1
Hap19	-	1	-	-	1
Hap20	-	1	-	-	1
Hap21	-	1	-	-	1
Hap22	-	1	-	-	1
Hap23	-	1	-	-	1
Hap24	-	1	-	-	1
Hap25	-	1	1	1	3
Hap26	-	1	-	-	1
Hap27	-	2	-	2	4
Hap28	-	1	-	-	1
Hap29	-	-	1	-	1
Hap30	-	-	1	-	1
Hap31	-	-	1	-	1
Hap32	-	-	1	-	1
Hap33	-	-	1	-	1
Hap34	-	-	1	-	1
Hap35	-	-	1	-	1
Hap36	-	-	1	-	1
Hap37	-	-	1	-	1
Hap38	-	-	1	1	2
Hap39	-	-	1	-	1
Hap40	-	-	1	-	1
Hap41	-	-	1	-	1
Hap42	-	-	1	-	1
Hap43	-	-	1	-	1
Hap44	-	-	-	1	1
Hap45	-	-	-	1	1
Hap46	-	-	-	1	1
Hap47	-	-	-	1	1
Hap48	-	-	-	1	1
Hap49	-	-	-	1	1
Hap50	-	-	-	1	1
Hap51	-	-	-	1	1
Hap52	-	-	-	2	2
Hap53	-	-	-	1	1
Numbers	16	17	16	19	68

Table 2. *Coilia ectenes taihuensis* sample site, size, and genetic diversity inferred from the D-loop sequence.

Population	Location	N	N hap	Hd	π	Tajima's D	Fu's FS
DL	Dalei	16	15	0.992	0.00699	-1.29989	-8.656**
GH	Gonghu	17	14	0.971	0.00616	-1.87370*	-6.300**
MS	Manshan	16	16	1.000	0.00711	-1.34594	-11.536**
MLW	Meiliangwan	19	17	0.988	0.00731	-1.24889	-9.327**

N, sample size; N hap, number of haplotypes; Hd, haplotype diversity; π , nucleotide diversity. *P < 0.05, **P < 0.01.

Table 3. Genetic distance between (upper diagonal) and within (diagonal) populations, and the fixation index considering genetic distances (F_{ST}) between populations (lower diagonal).

Population	DL	GH	MS	MLW
DL	0.0070	0.0068 (0.0587)	0.0073 (0.2163)	0.0074 (0.0871)
GH	0.0307 (0.0631)	0.0062	0.0067 (0.1154)	0.0070 (0.1413)
MS	0.0196 (0.1351)	0.0090 (0.3603)	0.0072	0.0074 (0.1691)
MLW	0.0210 (0.1441)	0.0267 (0.1621)	0.0192 (0.1712)	0.0074

DL = Dalei; GH = Gonghu; MS = Manshan; MLW = Meiliangwan. P value is showed in the bracket, which is showed all bigger than 0.05, and there are no significant in genetic distance between and within populations.

Tajima's D showed negative values (Table 2), and Fu's Fs test showed relatively low values for all four populations ($P < 0.01$). These findings suggested that *C. ectenes taihuensis* might have undergone a population expansion during the past evolution process (Tajima, 1989). Further, during the appearance of variations and population expansion, new haplotypes might have formed, leading to an increase in haplotype diversity (Bentzen et al., 1988).

AMOVA identified significant genetic structures in the hierarchical rearrangements examined in this study. The genetic variance within and among populations was significant ($P < 0.05$) (Table 4). Individuals from Taihu Lake showed significant morphological differences with the other *C. ectenes* populations, and they could be easily distinguished from the other species. Moreover, it confirmed that the existence of genetic variability within different population of *C. ectenes*, especially in *C. ectenes taihuensis* (Xie, 2012). The results of AMOVA reflected a similar finding, and provided evidence to support that genetic diversity and differences existed within *C. ectenes taihuensis* populations. The genetic diversity might be influenced by factors such as geographic distance, species life history, and environment. The reasons for genetic divergence might include hydrological barriers and long-term geographical segregation. In addition, the opening of a sluice between the Yangzi River and Taihu Lake every year causes a large amount of water to flow from the Yangzi River to Taihu Lake, which causes *C. ectenes* to flow into the Taihu Lake. Due to the life habits of *C. ectenes taihuensis*, *C. ectenes* settles in Taihu Lake and interacts with *C. ectenes taihuensis* in the different localities of the lake, thereby resulting in high population genetic diversity.

Table 4. Analysis of molecular variance (AMOVA) of 4 populations of *Coilia ectenes taihuensis* based on mtDNA control region sequences.

Populations	Degrees of freedom (d.f.)	Sum of squares	Variance components	Percentage of variation
Among populations	3	12.554	0.6624 (0.04692)*	2.12
Within populations	64	195.887	3.06073*	97.88
Total	67	208.441	3.12697	100

Populations represent the rearrangement tests in the hierarchical analysis with a significance value (P). The asterisk indicates significant difference.

The reconstructed NJ phylogeny tree showed distinct genealogical branches or clusters of all samples. It was found that the haplotype tree had two main distinct branches; one main branch had five distinct secondary branches and the other had eleven distinct secondary branches, and there were many small branches under secondary branches (Figure 2). Moreover, in the tree four secondary branches had many small branches, including 10, 9, 13, and 10 individuals (Figure 2). According to the haplotype distribution in the tree, there were no distinct haplotype groups; all haplotypes of four populations did not cluster into respective branches, indicating that the phylogenetic tree did not show a distinct corresponding relationship between geographic locations and haplotype. Furthermore, it might reflect a high genetic differentiation in the *C. ectenes taihuensis* population. When new individuals settled in the environment, genetic diversity would increase to some extent due to hybridization and variation (Chen and Yan, 2005). Moreover, other factors, such as population structures, inbreeding waters, and high ancestral haplotype frequency, could also affect the genetic distribution (Schneider and Excoffier, 1999).

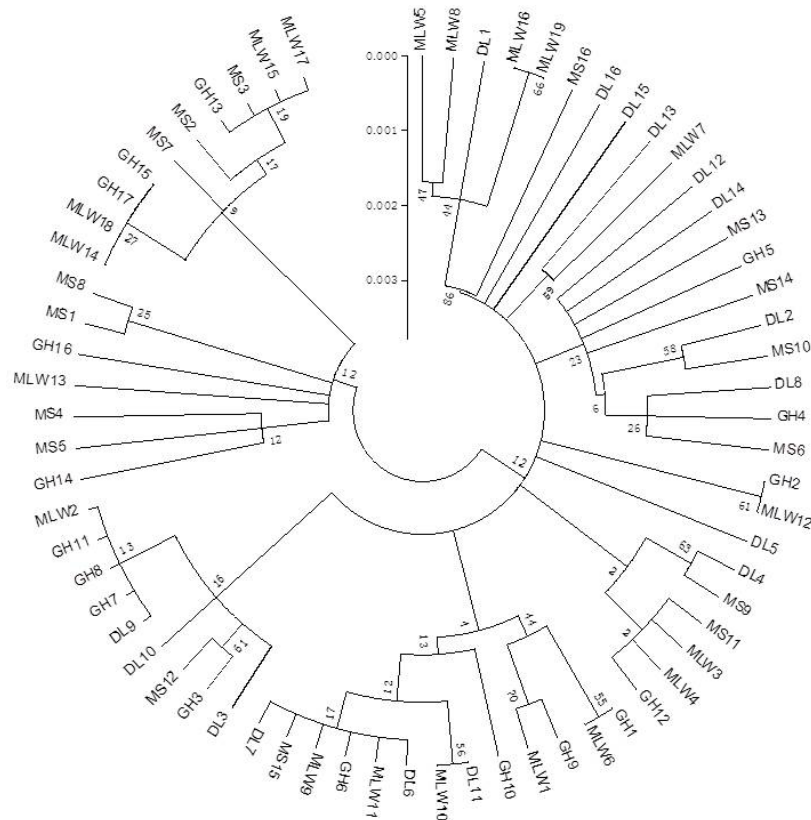


Figure 2. Neighbor-joining (NJ) tree for control region haplotypes of *Coilia ectenes taihuensis*. Bootstrap supports of 50% in 1000 replicates are shown. Numbers at the tips of branches correspond to one haplotype. Bars represent the frequency and geographic distribution of each haplotype.

CONCLUSION

Our results showed that *C. ectenes taihuensis* presented high genetic diversity and frequent gene communication. Based on the findings of the previous studies and the data obtained in the present study, it was concluded that the genetic diversity of *C. ectenes taihuensis* was relatively high (Ma et al., 2012). Our results confirmed that it is still important to make specific and effective plans to protect and utilize this economically and ecologically important species. Besides, we should specially focus on the resource fluctuation and sustainable utilization of *C. ectenes taihuensis*. Findings of the present study can also be used for supporting the development of *C. ectenes taihuensis*.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Special Fund for Agro-Scientific Research in the Public Interest (#201303056-2) and the Fundamental Research Funds from the Freshwater Fisheries Research Center (FFRC), Chinese Academy of Fishery Sciences, China (#2015JBFM15).

REFERENCES

- Arnason E and Rand DM (1992). Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Genetics* 132: 211-220.
- Bentzen P, Leggett WC and Brown GG (1988). Length and restriction site heteroplasmy in the mitochondrial DNA of american shad (*alosa sapidissima*). *Genetics* 118: 509-518.
- Brown GG, Gadaleta G, Pepe G, Saccone C, et al. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.* 192: 503-511. [http://dx.doi.org/10.1016/0022-2836\(86\)90272-X](http://dx.doi.org/10.1016/0022-2836(86)90272-X)
- Brown JR, Beckenbach AT and Smith MJ (1992). Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser transmontanus*). *Genetics* 132: 221-228.
- Brzuzan P (2000). Tandemly repeated sequences in mtDNA control region of whitefish, *Coregonus lavaretus*. *Genome* 43: 584-587. <http://dx.doi.org/10.1139/g00-001>
- Cecconi F, Giorgi M and Mariottini P (1995). Unique features in the mitochondrial D-loop region of the European seabass *Dicentrarchus labrax*. *Gene* 160: 149-155. [http://dx.doi.org/10.1016/0378-1119\(95\)00232-U](http://dx.doi.org/10.1016/0378-1119(95)00232-U)
- Chen YF and Yan YZ (2005). Evolutionary biology of invasions. *Acta Hydrobiol. Sin.* 29: 220-224.
- Cheng Q, Wen J, Wang YL, Han J, et al. (2006). Genetic diversity and genetic differentiation between *Coilia ectenes* and *Coilia ectenes taihuensis* inferred from cytochrome b gene segment sequence of mitochondrial DNA. *Hupo Kexue* 18: 425-430. <http://dx.doi.org/10.18307/2006.0416>
- Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.
- Li XQ, Liu F, Leng CM, Dong GC, et al. (2015). Genetic structure and diffusion of population of *Coilia ectenes taihuensis* in Lake Nansi inferred from the mitochondrial control region. *Hupo Kexue* 27: 686-692. <http://dx.doi.org/10.18307/2015.0417>
- Ma C, Cheng Q, Zhang Q, Zhuang P, et al. (2010). Genetic variation of *Coilia ectenes* (Clupeiformes: Engraulidae) revealed by the complete cytochrome b sequences of mitochondrial DNA. *J. Exp. Mar. Biol. Ecol.* 385: 14-19. <http://dx.doi.org/10.1016/j.jembe.2010.01.015>
- Ma C, Cheng Q and Zhang Q (2012). Genetic diversity and demographical history of *Coilia ectenes* (Clupeiformes: Engraulidae) inferred from the complete control region sequences of mitochondrial DNA. *Mitochondrial DNA* 23:

- 396-404. <http://dx.doi.org/10.3109/19401736.2012.710202>
- Rasmussen RS and Morrissey MT (2008). DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. *Compr. Rev. Food Sci. Food Saf.* 7: 280-295. <http://dx.doi.org/10.1111/j.1541-4337.2008.00046.x>
- Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Schneider S and Excoffier L (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152: 1079-1089.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.
- Tang W, Hu X and Yang J (2007). Species validities of *Coilia brachygnathus* and *C. nasus taihuensis* based on sequence variations of complete mtDNA control region. *Biodivers. Sci.* 15: 224-231. <http://dx.doi.org/10.1360/biodiv.060263>
- Xie JY (2012). Genetic and morphological variation of three freshwater lake populations of *Coilia ectenes* (Engraulidae). *Genetika* 48: 1200-1206.
- Xu GC, Xu P, Gu RB, Zhang CX, et al. (2011). Feeding habits and growth characteristics of pond-cultured *Coilia nasus* fingerlings. *Shengtaixue Zazhi* 30: 2014-2018.
- Xu GC, Wei GL, Li JL, Xu P, et al. (2012). The genetic diversity of farmed tapertail anchovy *Coilia nasus* and *Coilia nasus taihuensis* by mitochondrial D-loop genes analysis. *J. Dalian Ocean Univ.* 27: 448-452.
- Zardoya R and Meyer A (1997). The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). *Genetics* 146: 995-1010.