Biochemical Systematics and Ecology 54 (2014) 103–112

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Population structure of *Coilia nasus* in the Yangtze River revealed by insertion of short interspersed elements

Dong Liu, Yingying Li, Wenqiao Tang*, Jinquan Yang, Hongyi Guo, Guoli Zhu, Huihua Li

Laboratory of Ichthyology, Shanghai Ocean University, Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 13 June 2013
Accepted 21 December 2013
Available online 8 February 2014

Keywords:
Fish
Engraulidae
Population
Retrotransposon
SINE

ABSTRACT

Coilia nasus is found in the Yangtze River and the coastal waters of China, Korea, and Japan. Two ecotypes (anadromous and freshwater-resident populations) are distributed throughout the Yangtze River basin based on their ecology and behavior, but relatively little is known about the population structure of this species. Analysis of short interspersed element (SINE) insertions, which vary among individuals, has been acknowledged to provide a unique way to study population divergence. SINEs isolated from *C. nasus* were characterized, and this enabled analysis of the SINE insertion pattern in six populations distributed throughout the Yangtze River basin. In all populations, four SINE loci displayed individual polymorphism, and two SINE loci showed a stochastic loss in all individuals of two resident populations. The correlation between genetic and geographic populations indicated a degree of genetic isolation in this species. In contrast with *Coilia grayii* and *Coilia mystus*, two SINE loci appeared only in *C. nasus*. Sequencing analysis indicated that the high insertion variability of SINEs was attributed mainly to the tails, which contained various repeat copies. The results in this study will be useful for sustainable management of fishery resources and conservation of this species.

1. Introduction

The grenadier anchovy *Coilia nasus* in the family Engraulidae is a coastal, estuarine, and freshwater species. This species is distributed widely throughout the western coastal waters of Korea and the Ariake Sound of southwestern Japan, as well as the middle and lower reaches of the Yangtze River, the East Sea, and the Yellow Sea of China (Whitehead et al., 1988). Within its distribution, *C. nasus* exhibits remarkable diversity in terms of morphology, ecology, and behavior. Two ecotypes of *C. nasus* have been found in the middle and lower reaches of the Yangtze River basin: a resident population and an anadromous population (Zhang, 2001). The former population is composed of a freshwater-resident population, *Coilia brachygnathus*, and a landlocked population, *C. nasus taihuensis*. The various geographic populations are formed through species expansion and adaptive radiation, with an estimated age of 0.17–0.13 Myr (Yang et al., 2008).

The freshwater-resident population inhabits only fresh water and spawns and lives in the middle and lower reaches of the Yangtze River basin. The landlocked population is found only in Taihu Lake, a short distance from the Yangtze River. Taihu Lake

E-mail address: wqtang@shou.edu.cn (W. Tang).

0305-1978
http://dx.doi.org/10.1016/j.bse.2013.12.022

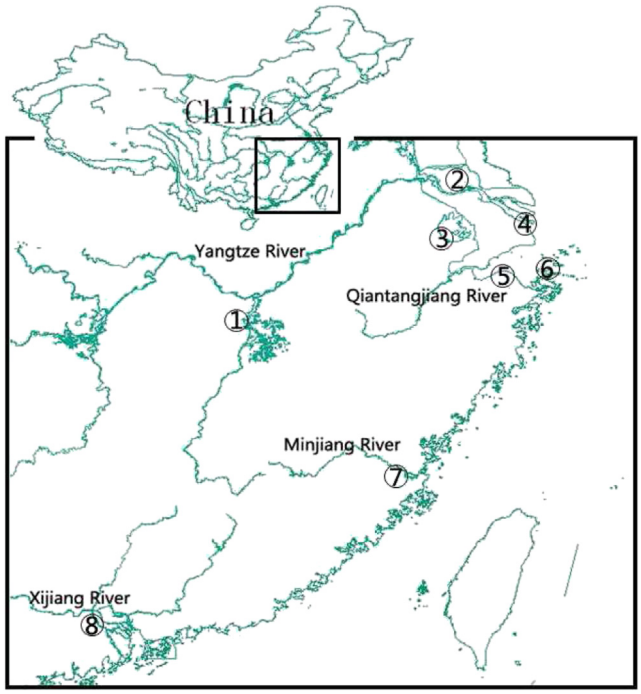

Fig. 1. The eight sampling sites for *C. nasus*, *C. mystus*, and *C. grayii*. The collection sites (circles) correspond to locations given in the text: ① Poyang Lake, ② Jingjiang, ③ Taihu Lake, ④ Jiu Duan Sha, ⑤ Qiantang Jiang, ⑥ Zhoushan, ⑦ Mingjiang, and ⑧ Xijiang.

is the third largest freshwater lake in China, with an area of 2250 square kilometers (Chang, 1996). The anadromous population migrates for spawning from its oceanic habitat to the Yangtze River during its reproductive season from March to August, swimming 1400 km up the Yangtze River. From September to November, after reproduction, these fish and their progeny migrate back to the ocean (Zhang, 2001). Otolith strontium-to-calcium ratios in the anadromous population have confirmed their life history pattern (Dou et al., 2012). This home-migrating fish is one of the most important fish in China, with high economic value owing to its delicacy and nutritional value, whereas the freshwater-resident fish is of less economic value. However, the anadromous population has significantly decreased as a resource during the past decade due to overfishing, pollution, coastal construction, and other factors (Zhang et al., 2005). Interestingly, catches of the landlocked population continued to increase from 1952 to 2004, and the species is now the dominant fish in Taihu Lake (Liu et al., 2005). Therefore, it is important to understand the population structure of *C. nasus* in the Yangtze River basin to ensure better management and conservation of this species' fishery resources.

Significant genetic diversity, based on mitochondrial DNA control region sequences, has been reported among geographical populations of *C. nasus* in the Yangtze River, the Yalu Jiang River, and the Minjiang River (e.g., Ma et al., 2012). Various geographic populations collected from the Yangtze River basin have also shown genetic differences that were not correlated with geographical locations (Yang et al., 2008), despite prodigious morphological differences (e.g., jaw length, number of vertebrae, and anal fin rays). In previous studies, populations of *C. nasus* distributed in the Yangtze River and Taihu Lake were each treated as closely related to the species *C. brachygnathus* (which has a shorter jaw length than *C. nasus*) and the subspecies *C. nasus taihuensis* (which has fewer vertebrae than *C. nasus*), based on morphometric characteristics and ecological and physiological attributes (Yuan et al., 1980). Species identification of *C. nasus* distributed in the Yangtze River was resolved by exploring various markers such as allozymes and mitochondrial DNA (Liu, 1995; Tang et al., 2007; Yang et al., 2008; Zhou et al., 2010). Closely related species of *C. brachygnathus* and subspecies of *C. nasus taihuensis* regarded as synonymous to *C. nasus* should be geographical/ecological populations of this species (Tang et al., 2007; Zhou et al., 2010). No geographical differentiations have been observed in these populations based on genetic data, although significant morphological differences have been found (Xie, 2012). It is not known whether these population-specific morphometric characters are influenced by environmental or genetic factors. The genetic diversity of *C. nasus* is thought to be closely related to its adaptability, variability, and evolutionary potential, prerequisites for living organisms to cope with uncertainty in the environment (Ma et al., 2012). Some populations have discrete geographic distributions, such as the Taihu Lake population, which, based on mitochondrial data, is not genetically different from other populations (Yang et al., 2008; Ma et al., 2012). Perhaps because the mitochondrial genome is maternally inherited, using it as a genetic marker does not permit the detection of species undergoing population expansion events during short periods. A sensible molecular marker should be identified to detect such events.

Table 1
Cn-SINE selection primers, annealing temperatures, and PCR product size. The forward primer (F) is a Cn-SINE specific primer for each insertion site.

Locus	Sense (F)/ Antisense (R)	Primer sequence (5′–3′)	Annealing temperature (°C)	Product size (bp)
G2	F	CGAATCCCGCCCTACCCAT	57	443
	R	TTTTCCGCCCCTTTGCAAC		
AH2	F	TCTAACTCCACACTGCTCCAG	52	289
	R	TGCGAAACATTTTTGTCTCAT		
AH4	F	TGGTTAGGGATTTGGTCTTGC	58	577
	R	GTCACTGTGGTTTATGTGGGA		
AF1	F	GAGCAAGGCATCTAACCCCAC	56	720
	R	CATCATCGCAACATTCAGCAA		
AF7	F	GCCTCGCAATCGGAAGG	59	661
	R	GCAGGAGACGCAAACGG		
T31	F	AGGGAGTTGGTCTTGGGATCG	58	409
	R	GGTAATTTGCTAAAGGGGCTG		
T32	F	ATTCTGGCAGCTGTGGTCTAGC	58	346
	R	TCCCGTGTCATCAGTGTGTTTA		
T41	F	ACACACACACCAACACCAACG	60	335
	R	CCACACCTCCAGGCAGAAATC		
T49	F	GGTTAGGTAGCTGGTCTTGGGA	56	380
	R	TAATGAGTTTTTAGAGGGGGTC		
T87	F	AGAATGTGAGTGGGGGAGTGA	58	171
	R	TGCTGGCATATTGTAAGGAGG		
T178	F	CAAGGCATCTAACCCCACACT	56	279
	R	ACAACTTCCCAACATCGACAA		
T210	F	AAAAATGGGGATGGGGGAGTA	58	265
	R	GCAACAGAGGAAAATGGAGAC		

Short interspersed elements (SINEs) represent genetic markers and have been widely used as a powerful tool in studying systematic biology and population histories in a variety of taxa (Shedlock and Okada, 2000; Ray, 2007). These elements are reverse transposable elements and multiply in genomes via the reverse transcription of RNA from a parental sequence. SINEs are excellent markers in examining the population processes of taxa because of their random integration sites and the irreversible nature of integration at each site (Takahashi et al., 2001; Ray, 2007). Compared with mitochondrial, microsatellite, and other molecular markers, SINEs offer certain advantages over more commonly used genetic characters. First and foremost, SINE markers are essentially homoplasy-free characters and provide an extremely accurate picture of the evolutionary relationships of taxa (Ray, 2007). As one example, over 10,000 polymorphic loci have been identified in comparing the genomes of two breeds of domestic dog (Wang and Kirkness, 2005). Application of the SINE method to population genetic analyses has not been explored outside of organisms with known genomes, such as humans, primates, whales, and a few fish (Ray et al., 2005; Nikaido et al., 2006). Reasons why SINEs are not widely applied to investigate the population structures of taxa include the difficulty in isolating them in organisms with unknown genomes and identifying large numbers of polymorphic loci within a given species (Ray, 2007).

In a previous study (Liu et al., 2012a), we reported tRNA-derived Cn-SINEs, a SINE family isolated from *C. nasus*. In the present study, we investigate intraspecies variation of Cn-SINE insertions in the genome of *C. nasus* collected from six sites in the Yangtze River basin and from two other species in the genus *Coilia*, *Coilia grayii* and *Coilia mystus*. The aim of this work is to use SINE insertions to evaluate geographical population differentiations of *C. nasus* distributed within the Yangtze River basin, providing valuable information that can be used to help protect and sustain this species' resources.

2. Materials and methods

2.1. Sample collection

Samples of *C. nasus* were collected from six sites: four sites in the middle and lower reaches of the Yangtze River — Poyang Lake (PY), Jingjiang (JJ) in Jiangsu province, Taihu Lake (TH), and the Jiu Duan Sha site at the estuary of the Yangtze River (JD) in Shanghai — and the other two sites in the estuary of the Qiantang Jiang River (QT) and Zhoushan (ZS) in the coastal region of the East Sea in China (Fig. 1). Sampling from these six field sites enabled the inclusion of all ecotypes of this species. Specifically, the JJ and JD samples represent the home-migrating population (collected between May and June), the PY samples represent the freshwater-resident population (collected in December and January), the TH samples represent the landlocked population, and the QT and ZS samples represent the marine population. Eight individuals per site were randomly chosen as representative samples. In addition, samples of *C. grayii* were collected from the Xijiang River in the Guangdong province and samples of *C. mystus* were collected from the Minjiang River in the Fujian province (Fig. 1), with three randomly chosen

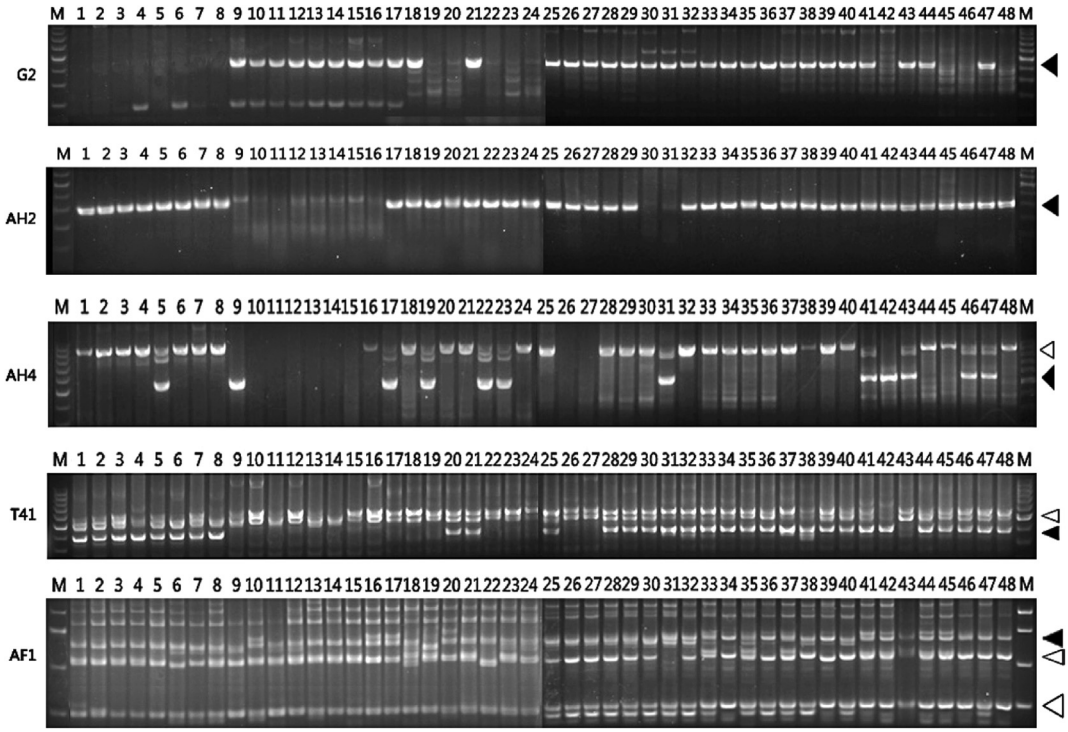

Fig. 2. Results of the PCR amplification of *C. nasus* genome at five loci. The template DNAs are denoted 1–8 for TH, 9–16 for PY, 17–24 for QT, 25–32 for JJ, 33–40 for JD, 41–48 for ZS, and M for a marker. The dark arrowheads indicate the presence of a band of the predicted size, whereas the light arrowheads indicate bands with an unpredicted size.

individuals from each species. Muscle tissue used for the analysis was preserved in 95% ethanol and the specimens were deposited at the Fish Collection of Shanghai Ocean University in China.

2.2. DNA extraction and primer design

Total genomic DNA was extracted from small amounts of ethanol-preserved muscle by proteinase K digestion in lysis buffer at 55 °C for 2–3 h, using the TIANamp Genomic DNA Kit according to the manufacturer's instructions (Tiangen, China). The Cn-SINE insertion and flanking sequences were obtained from an anadromous *C. nasus* individual via the polymerase chain reaction (PCR) and the magnetic bead-mediated target capture approach from one of our previous studies (Liu et al., 2012a). The GenBank accession numbers for the Cn-SINE insertions were JQ083280–JQ083297JQ083280JQ083281JQ083282JQ083283JQ083284JQ083285JQ083286JQ083287JQ083288JQ083289JQ083290J-Q083291JQ083292JQ083293JQ083294JQ083295JQ083296JQ083297. Primer3 software was used to design primers for the Cn-SINE insertion sites (Lalitha, 2000) and amplification of the Cn-SINE loci was performed using a Cn-SINE specific primer and a second primer specific to the genomic region flanking the insertion, as described by Morescalchi et al. (2010) and Weiss et al. (2012). The primer names were prefixed with the Cn-SINE insertion site, described previously by Liu et al. (2012a). All oligonucleotides were synthesized by Sangon Inc. (Shanghai, China). Table 1 lists information for the PCR primers. PCR reactions were carried out as described by Liu et al. (2009) and PCR amplification was carried out in an Eppendorf Mastercycler (Eppendorf, Munich, Germany) at 94 °C for 3 min; 30 cycles of 94 °C for 30 s; 52–60 °C for 30 s, depending on the Cn-SINE; and 72 °C for 1 min, with a final extension at 72 °C for 10 min. Amplification products were separated via 2.0% agarose gel electrophoresis and visualized by ethidium bromide staining.

2.3. Insertion polymorphisms of Cn-SINE and sequencing

Primer pairs that successfully amplified products in *C. nasus* were selected to screen the genomes of three individuals in each of two other species, *C. grayii* and *C. mystus*, to identify the evolutionary origin of Cn-SINEs via the presence or absence of insertion polymorphisms. The predicted size of PCR products from these species indicated the presence of a Cn-SINE unit; the lack of such size products from a certain species could be a consequence of veritable deletions of the SINE insertion locus and/or nucleotide mutations at the primer binding site, as described by Nikaido et al. (2007). When the amplified product sizes varied widely between individuals' SINE insertion loci, representative products were selected for DNA sequence analysis to

identify the source of this disparity. The PCR products were separated, cloned, and sequenced as described by Liu et al. (2009), and these sequences was identified through BLAST search in nucleotide (nr/nt) collections of NCBI Genbank with the basic blastn default parameters (www.ncbi.nlm.nih.gov), and then deposited in Genbank. The DNA sequences were aligned through multiple sequence alignment with the ClustalX2 program, version 2.0 (Chenna et al., 2003). All sequences with repeat elements were identified using RepeatMasker (A.F.A. Smit, R. Hubley, and P. Green, unpublished data; current version open-4.0.1; http://www.repeatmasker.org).

3. Results

3.1. Intraspecific variation of Cn-SINE insertions in C. nasus

In one of our previous studies (Liu et al., 2012a), we isolated 18 loci containing members of the Cn-SINE family from *C. nasus*. The amplification patterns of PCR products obtained from Cn-SINE insertion events at 12 genomic loci with a Cn-SINE specific primer and a second primer specific to the flanking SINE insertion sequence exhibited greater than 50% diversification. The PCR product size for each of seven loci (AF7, T31, T32, T49, T87, T178, and T210) showed a single band, as predicted in Table 1, indicating that these loci should be orthologous in all individuals from the six populations (data not shown). Four Cn-SINE insertion loci at G2, AH2, AH4, and T41 displayed polymorphism in individuals, which is potentially useful for analyzing the population structure of *C. nasus*, as shown in Fig. 1. In the case of locus G2, the predicted products containing a Cn-SINE unit were observed in PY, JJ, and JS, and in some individuals of QT (3/8) and ZS (4/8); no such band was exhibited in the entire TH population. Similarly, at the locus AH2, isolated from the anadromous population of *C. nasus*, the Cn-SINE insertion events were found in TH, QT, JD, and ZS, and in some individuals from JJ (6/8); such insertion events did not occur in the PY population. These results provide evidence that genetic diversity in *C. nasus* emerged via the stochastic loss of the SINE insertion locus in certain populations. It must be noted that locus AH4 also exhibited random insertion events with predicted sizes in some individuals (12/48), but such insertion events did not correlate with specific populations. The variation of insertion events at the five Cn-SINE loci (G2, AH2, AH4, T41, and AF1) indicated that *C. nasus* was highly heterogenous among populations and even within populations, although not every population had a specific Cn-SINE locus insertion or deletion. The large size difference compared with the shorter fragment of locus AH4 was due to a 389-bp fragment inserted exactly in the middle of the shorter one, as determined by sequence analysis (accession number KF007213). Locus T41 yielded the predicted product size in some populations and was absent in PY and individuals from other populations; larger fragments were amplified from the genomes of almost all individuals. The sequence analysis of locus T41 indicated a SINE unit inserted in the microsatellite region of the larger fragment. For example, in an individual from the JJ population (accession number KF007214–KF007215KF007214KF007215), the 5′ end of the Cn-SINE insertion region was followed by a complete repeat $(GACA)_{12}$ region and an incomplete repeat $(TG)_{53}$ region in the larger fragment.

Locus AF1 exhibited 3–7 bands of various lengths in individuals of *C. nasus* (Fig. 2). To confirm the polymorphism resulting from an insertion, only fragments smaller than predicted (720 bp) were cloned and sequenced. Three fragments, sizes 190, 450, and 600 bp, were obtained at this locus (accession number KF007216–KF007217KF007216KF007217; the <200-bp sequence was not deposited into NCBI). Analysis of sequence alignments indicated that all of these sequences harbored part of the Cn-SINE unit sequence, namely, Cn-SINE's body (a primer corresponding to this region) and tail region, but the flanking Cn-SINE sequences were significantly different from each other (Fig. 3). The Cn-SINE element that was inserted in these fragments possessed a 3′ tail sequence of about 60 bp similar to that in long interspersed element (LINE) retrotransposons isolated from eel (Kajikawa and Okada, 2002). Both Cn-SINE and LINE share a (TGTAA) repeat at the 3′ end of the tail (Fig. 3), which is a requirement for the retrotransposition of SINEs (Kajikawa and Okada, 2002). The Cn-SINE elements obtained via the PCR primer pair of locus AF1 appeared to harbor the most (TGTAA) repeats (six) out of all the Cn-SINE

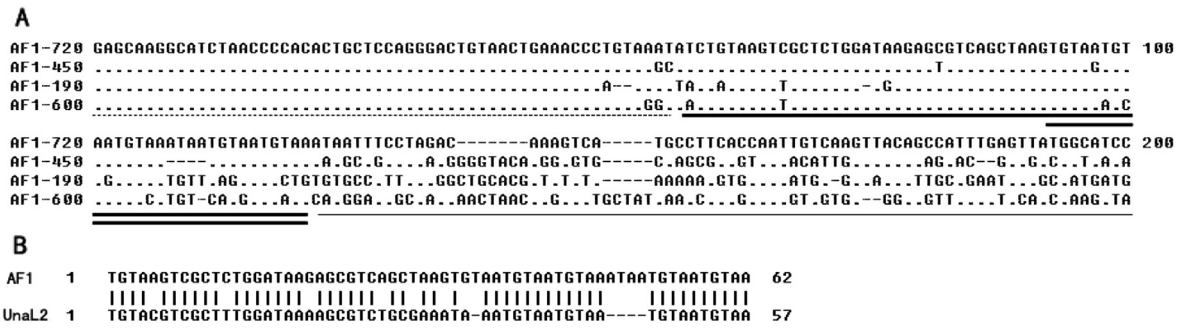

Fig. 3. Alignments of nucleotide sequences of Cn-SINEs from loci AF1 and UnaL2 from eel *Anguilla japonica*. (A) Multiple alignments of the Cn-SINEs from PCR amplification with the primer pair for locus AF1. (B) Alignments of about 60-bp-tails of Cn-SINEs from AF1 and UnaL2 from eel (GenBank number AB1796240). The dots correspond to the same bases, and the dashes denote gaps. The dotted underline marks the partial body of a SINE unit. The bold underline marks the tail of a SINE unit shared by UnaL2 and the double underline marks a (TGTAA) repeat region. The fine underline shows the flanking SINE insertion sequence. Vertical lines indicate identical nucleotides.

members (Liu et al., 2012a), indicating that the heterozygous genotypes in *C. nasus* individuals due to multiple insertion events in the amplified region are not in the same position as reported previously (Ray et al., 2005), although no population-specific fragment was observed in all *C. nasus* individuals.

3.2. Interspecies variation of Cn-SINE

We took a similar approach to examine the distribution of Cn-SINEs in other species of the genus *Coilia*, namely *C. grayii* and *C. mystus*, to determine whether insertions of Cn-SINE were fixed in *C. nasus*. Out of the 12 loci described above, the four orthologous loci, T31, T49, T178, and T210, extensively shared Cn-SINE units, as expected (see Fig. 4, in which T31 is shown as an example). Note that loci G2 and AH2 are present in *C. nasus* and absent in both *C. mystus* and *C. grayii* (Fig. 4, G2 and AH2), indicating that these two loci were species-specifically inserted into *C. nasus*. Locus T87 was species-specifically present in both *C. grayii* and *C. nasus* (Fig. 4, T87), with 99% sequence identification between the species. For the four loci AH4, AF1, AF7, and T41, SINE insertions were polymorphic between the two species and each locus yielded one to four bands. The bands differed in size because the Cn-SINE insertions were not in the exact same positions, which were confirmed by sequencing the PCR bands (example: locus AF7, accession number KF007218–KF007219KF007218KF007219). Such variation in insertions could result from ancestral Cn-SINE polymorphisms followed by incomplete lineage sorting during rapid speciation of *Coilia* species. Locus T32 yielded a larger band in both *C. nasus* and *C. grayii*, whereas a shorter band was present in *C. mystus*, as shown in Fig. 4 (T32). These bands were cloned and sequenced. The sequence alignment showed that the nucleotide identity was higher between *C. nasus* and *C. grayii* (98%) than between *C. nasus* and *C. mystus* (68%) (accession number KF007220–KF007221KF007220KF007221). A 17-bp fragment deleted from the conserved box B region of Cn-SINE was observed in *C. mystus*, providing evidence that heterozygous genotypes among species were generated by the fragment of SINEs deleted during rapid successive speciation.

4. Discussion

4.1. Genetic diversity of C. nasus

The genetic diversity of *C. nasus* at different geographical population levels has been discussed in previous reports. Yang et al. (2008) reported that the genetic distances among the four populations TH, QT, ZS, and JJ (0.010–0.013) are smaller than those between PY and the four populations (0.017–0.018). Individuals in the PY population, based on the mitochondrial DNA complete control region, appear in many distant branches of the phylogenetic tree, indicating that PY is not an isolated population. Ma et al. (2012) also reported that, based on mitochondrial data, the genetic distances (0.0062–0.0073) among three populations distributed in the lower reaches of the Yangtze River (the Anqing in the Anhui province, the Sang Jia Gang in

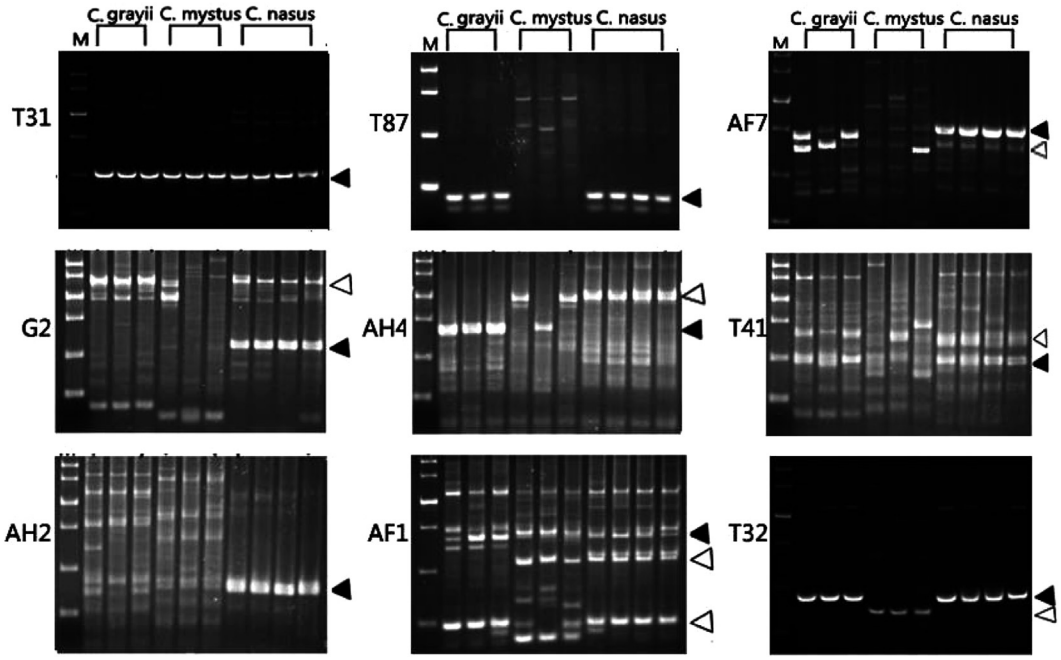

Fig. 4. Electrophoretic gels of PCR products for nine representative Cn-SINE loci. The locus names and template DNA are indicated. M denotes a marker. The dark arrowheads indicate bands of the predicted size listed in Table 1, whereas the open arrowheads show bands with sizes that were not predicted.

Shanghai, and TH) were smaller than those between PY and each of the three populations (0.0087–0.0094). Although PY showed genetic differences from other populations of *C. nasus*, mitochondrial markers were not able to distinguish PY individuals from other populations. In addition, the population living in Lake Poyang coexists with individuals from other populations due to migration (Tao et al., 2013), which makes PY individuals difficult to identify via mitochondrial markers. Our recent study (Liu et al., 2012b) confirmed that gene flow, as measured by nuclear markers, occurs specifically within geographic populations. Based on mitochondrial and microsatellite data, no geographical differentiation has been observed among these populations to date (Ma et al., 2012; Yang et al., 2012). In contrast, analysis of the morphological characteristics of maxillary length and vertebrae number has shown significant differences among the JJ, JD, TH, and PY populations of *C. nasus*, except between the JJ and JD populations (Cheng and Tang, 2011). The JJ and JD populations are considered an anadromous ecotype of *C. nasus*, whereas TH and PY populations are considered a resident ecotype of *C. nasus*, as determined by their behavior (Fig. 5; Zhang, 2001).

In the present work, we successfully identified two Cn-SINE loci, G2 and AH2, through their species-specific insertion in *C. nasus*. Interestingly, locus G2 is absent in the TH population, as is locus AH2 from the PY population, indicating that the *C. nasus* genome is indeed heterozygous to a certain extent among different populations. Our results provide an important clue in distinguishing the different ecotypes of *C. nasus* distributed throughout the Yangtze River basin. The absence of insertion at locus AH2 supported PY as a geographic population according to morphological characteristics that differed from those of other populations at the molecular level. Morphologically, the PY population's resident ecotype exhibited a shorter up-jaw length than that of other populations, and PY individuals could be distinguished from other populations based on this criterion. Similarly, loci G2 and T41 unambiguously supported TH as a geographic population, although the latter locus was also absent in some individuals of QT (5/8; see Fig. 2). *C. nasus* exhibited remarkable morphological and molecular differences, perhaps acquired through independent and explosive adaptive radiation. Polymorphic insertion events of Cn-SINEs in *C. nasus* may be the evolutionary reservoirs of different geographic populations. Locus AH4 showed a high degree of insertion polymorphism in populations of *C. nasus* (Fig. 2) and provided evidence of evolutionary reservoirs, although this insertion event was not specifically related to geographic populations. Takahashi et al. (2001) proved that the diversity of cichlid fish in Lake Tanganyika most likely stemmed from a period of adaptive radiation and showed extensive incomplete lineage sorting of alleles for the presence or absence of SINEs during successive speciation events. We found that locus AF1 shows multiple insertions in the amplified region, indicating that such a SINE unit may be active in the genetic diversity of *C. nasus*. To our knowledge, this phenomenon has not been previously reported for other fishes.

4.2. Cn-SINE-induced genetic heterogeneity

Application of the SINE method to population genetic analyses has not been well explored outside of humans and a few taxa with known genomes. In chum salmon and pink salmon, six and four loci, respectively, were fixed at the species-specific level (Takasaki et al., 1997). The genomes of salmon species were shaped and reshaped by the amplification and dispersion of different SINE families (Kido et al., 1991). Due to the way the presence/absence of SINE insertions has contributed to the descriptions of phylogenetic character types, SINE loci have been widely used to investigate the phylogeny of taxa, and primers used for the PCR detection strategy were designed from the flanking SINE sequences (Ray et al., 2005; Nikaido et al., 2007). Therefore, some species may not give rise to amplified products because of the large degree of changes in the flanking SINE sequences.

In this study, we used primers corresponding to the Cn-SINE regions of loci and a second primer specific to the genomic region flanking the insertion (Table 1). The conserved Cn-SINE unit allowed for efficient amplification of the sequence diversity of intra- and interspecies of *C. nasus*, as described by Morescalchi et al. (2010). We can draw several important conclusions from our results about the genetic heterogeneity of *Coilia* species. (i) The Cn-SINEs isolated from loci G2, AH2, AH4, AF, and T41 share the same tail with LINEs and showed polymorphisms (either presence or absence) in the population structures of *C. nasus;* in addition, the observed diversity of insertion events was produced by multiple retrotranspositions of such Cn-SINEs via LINEs, resulting in the genetic heterogeneity of *C. nasus*. Analysis of previously obtained sequence combinations (Liu et al., 2012a) indicated that members of the Cn-SINE family are divided into two types: one harboring a tail at its 3′ end similar to LINEs and the other without such a tail. The SINEs and LINEs isolated from eel shared a 3′ tail with a repeat

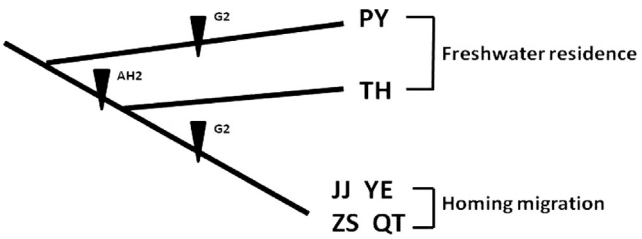

Fig. 5. Dendrogram of four populations of *C. nasus* obtained from analyses of morphology (Cheng and Tang, 2011). Vertical arrowheads denote insertions of Cn-SINE into each lineage based on Cn-SINE insertion data (Locus G2 and AH2).

(TGTAA) structure, which is a requirement for the retrotransposition of SINEs when the repeat is complete (Kajikawa and Okada, 2002).

(ii) Locus AH4 showed that the insertion diversity of *C. nasus* was due to a fragment insertion in the shorter band that yielded the larger band, resulting in genetic heterogeneity among and within *C. nasus*. Such an event is a so-called secondary insertion event in the amplified region, as previously reported by Ray et al. (2005).
(iii) The Cn-SINE inserted into microsatellite regions induced the genetic diversity of *C. nasus* but did not disturb the gene functions, so that the retrotransposons coexisted with their host genome. Sequence analysis of locus T42 showed the SINE insertion followed by a microsatellite-rich region, and the difference in microsatellite numbers led to the population diversity of *C. nasus*. This phenomenon was previously reported in the rice genome (Akagi et al., 2001).
(iv) Deletion mutations accumulated in specific species, indicating that genetic diversity at the species level was itself due to the sequence variety of Cn-SINE, except for insertion/absence events. In one of our previous studies (Liu et al., 2012a), a Cn-SINE unit consisted of three regions: the tRNA original 5′ terminal head, the body, and the 3′ terminal tail. The head of the Cn-SINE harbored two conserved boxes, A and B, in which polymerase III transcription for DNA templates carrying SINE originates (Kajikawa and Okada, 2002). In this study, we found that locus T32 showed a deletion of box B in *C. mystus*, compared with *C. nasus* and *C. grayii*. Such a mutation may contribute to retrotransposons being inactive and becoming new genetic material for a species (Feschotte and Pritham, 2007). The results of our preliminary study of *C. nasus* and other *Coilia* species provide direct evidence of these species' genetic diversity.

4.3. Evolutionary origin of ancestral Cn-SINE

A SINE is believed to generate from the genome of one individual and spread into the population through sexual reproduction. If two species diverged before a locus became fixed for SINE insertion, such a locus might be polymorphic between the resulting species (Takasaki et al., 1997). To investigate the specific differentiation of Cn-SINE insertions in *Coilia*, we used three species of this genus, *C. mystus*, *C. grayii*, and *C. nasus*, for comparative analysis. These species are distributed throughout the southern and northern coasts of the China Sea; *C. grayii* is north of Fuzhou in the Fujian province and *C. mystus* is spread throughout the coastal waters of the China Sea (Zhang, 2001). *Coilia dussumieri*, another species of *Coilia* restricted to Hong Kong, was not examined in our study due to difficulties in collecting individuals of this species. The phylogenetic relationships among the three studied species have been suggested in previous studies, as shown in Fig. 6 (Tang et al., 2007; Yang et al., 2010; Zhou et al., 2010).

We employed the phylogenetic relationships of these three species to further assess the evolutionary origin of Cn-SINE insertions. In this study, four loci (T31, T49, T178, and T210) invariably revealed the presence of a parallel insertion in all three species – *C. mystus*, *C. grayii*, and *C. nasus* – indicating the insertions at these loci occurred prior to species differentiation. One locus (T87) showed a single band pattern of PCR amplification in *C. nasus* and *C. grayii*, and such a band pattern was lacking in *C. mystus*. This finding shows that locus T87 was not fixed before *C. mystus* differentiated. If this locus had been cleanly removed from *C. mystus*, it should be present in other species affiliated with Engraulidae, but we did not observe such insertion events in the species we examined (data not shown). Two loci, AH4 and AF7, showed intraspecific polymorphism in two species, *C. mystus* and *C. grayii*. Interestingly, AH4 continued to have insertion polymorphisms, whereas locus AF7 showed only one band pattern in all populations of *C. nasus*, suggesting that the fixing of Cn-SINE insertions occurred after these

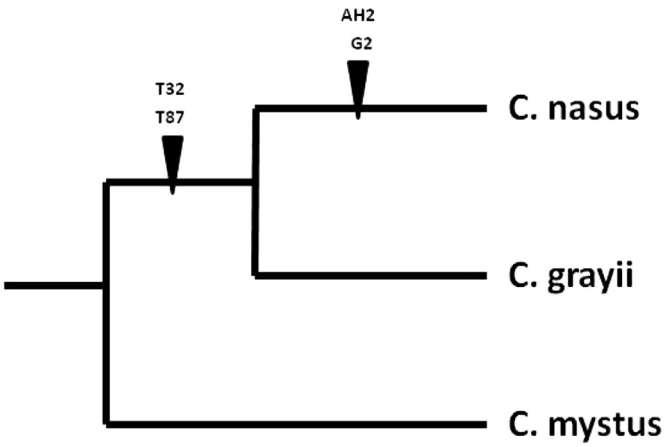

Fig. 6. Possible phylogenetic relationships among *C. nasus*, *C. grayii*, and *C. mystus*. The relationships were obtained from control region sequences (Tang et al., 2007), COI (Zhou et al., 2010) in mitochondrial DNA, and restriction fragment length polymorphism analyses (Yang et al., 2010). Vertical arrowheads denote insertions of Cn-SINE into each lineage based on Cn-SINE insertion data (Loci T32, T87, AH2 and G2).

species' differentiation. Such insertion polymorphisms of SINEs are very rare, and most SINEs are fixed in all populations of a given species (Takasaki et al., 1994). It may be possible that the time since the divergence of *Coilia* species has not been long enough for most SINEs to become fixed among these species, except for a few loci. For example, out of 12 loci examined, only two loci, AH2 and G2, were species-specifically fixed in *C. nasus*. In turn, the amplification pattern at loci T32 and T87 indicated a closer relationship between *C. nasus* and *C. grayii* than that between *C. nasus* and *C. mystus*. Cheng et al. (2008) hypothesized that *C. nasus* and *C. mystus* were closely related and that their differences may be considered subspecies based on the combined cytochrome *b*, 12S rRNA, and 16S rRNA gene sequences. Our results supported phylogenetic relationships of three species of *Coilia* from the genetic analyses of mitochondrial sequences (Fig. 6; Tang et al., 2007; Zhou et al., 2010) and AFLP (Yang et al., 2010). In the present study, we found that loci AH2 and G2 were specifically fixed in *C. nasus* and were polymorphic among populations of this species, which can be treated as a useful and convenient molecular marker, allowing population structure and genetic diversity of *C. nasus* to be estimated. Therefore, dynamics monitoring of *C. nasus* populations will be established, facilitating the management of fisheries resources and the development of sustainable fishing strategies to protect anadromous population of *C. nasus*. It should be noted that the anadromous population has significantly decreased over the past several decades, and appropriate measures have to be taken to prevent the decline of fishery resources and to decrease pollution of its aquatic habitats. Based on SINE insertion data, genetic difference between populations were observed, indicating that the genetic diversity of *C. nasus* should be protected.

Acknowledgments

We gratefully acknowledge grants from the Innovation Program of Shanghai Municipal Education Commission (No. 12YZ132), the National Natural Science Foundation of China (No. 31172407), the Ministry of Education's Doctoral Discipline Foundation in Higher Educational Institutions (No. 20123104110006), Shanghai Universities First-class Disciplines Project of Fisheries, and the Ministry of Education's Doctoral Discipline Foundation for Young Teachers in Higher Educational Institutions (No. 20103104120001).

References

Akagi, H., Yokozeki, Y., Inagaki, A., Mori, K., Fujimura, T., 2001. Micron, a microsatellite-targeting transposable element in the rice genome. Mol. Genet. Genomics 266, 471–480.
Chang, W., 1996. Major environmental changes since 1950 and the onset of accelerated eutrophication in Taihu Lake, China. Acta Palaeontol. Sin. 35, 155–174.
Cheng, Q., Cheng, H., Wang, L., Zhong, Y., Lu, D., 2008. A preliminary genetic distinctness of four *Coilia* fishes (Clupeiformes: Engraulidae) inferred from mitochondrial DNA sequences. Russ. J. Genet. 44, 339–343.
Cheng, W., Tang, W., 2011. Some phenotypic varieties between different ecotypes of *Coilia nasus* in Yangtze River. Chin. J. Zool. 46, 33–40.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res. 31, 3497–3500.
Dou, S., Kazuki, Y., Yu, X., Cao, L., Mari, K., Tsuguo, O., Katsumi, T., 2012. The migratory history of anadromous and non-anadromous tapertail anchovy *Coilia nasus* in the Yangtze River estuary revealed by the otolith Sr:Ca ratio. Environ. Biol. Fish. 95, 481–490.
Feschotte, C., Pritham, E.J., 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331–368.
Kajikawa, M., Okada, N., 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. Cell 111, 433–444.
Kido, Y., Aono, M., Yamaki, T., Matsumoto, K.I., Murata, S., Saneyoshi, M., Okada, N., 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. U.S.A. 88, 2326–2330.
Lalitha, S., 2000. Primer premier 5. Biotech. Softw. Internet Rep. 1, 270–272.
Liu, D., You, C., Liu, S., Liu, L., Duan, W., Chen, S., Yan, J., Liu, Y., 2009. Characterization of a novel Tc1-like transposon from bream (Cyprinidae, Megalobrama) and its genetic variation in the polyploidy progeny of bream–red crucian carp crosses. J. Mol. Evol. 69, 395–403.
Liu, D., Zhu, G., Tang, W., Yang, J., Guo, H., 2012a. PCR and magnetic bead-mediated target capture for the isolation of short interspersed nucleotide elements in fishes. Int. J. Mol. Sci. 13, 2048–2062.
Liu, D., Guo, H.Y., Tang, W.Q., Yang, J.Q., 2012b. Comparative evolution of S7 intron 1 and ribosomal internal transcribed spacer in *Coilia nasus* (Clupeiformes: Engraulidae). Int. J. Mol. Sci. 13, 3085–3100.
Liu, E., Liu, Z., Chen, W., Chen, K., 2005. Changes in the yield and composition of the fish catches and their relation to the environmental factors in Lake Taihu. J. Lake Sci. 17, 251–255.
Liu, W., 1995. Biochemical and morphological comparison and interspecific relationships of four species of the genus *Coilia* in China. Oceanol. Linn. Sin. 26, 558–565.
Ma, C., Cheng, Q., Zhang, Q., 2012. Genetic diversity and demographical history of *Coilia ectenes* (Clupeiformes: Engraulidae) inferred from the complete control region sequences of mitochondrial DNA. Mitochond. DNA 23, 396–404.
Morescalchi, M.A., Barucca, M., Stingo, V., Capriglione, T., 2010. Polypteridae (Actinopterygii: Cladistia) and DANA-SINEs insertions. Mar. Genom. 3, 79–84.
Nikaido, M., Hamilton, H., Makino, H., Sasaki, T., Takahashi, K., Goto, M., Kanda, N., Pastene, L.A., Okada, N., 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. Mol. Biol. Evol. 23, 866–873.
Nikaido, M., Piskurek, O., Okada, N., 2007. Toothed whale monophyly reassessed by SINE insertion analysis: the absence of lineage sorting effects suggests a small population of a common ancestral species. Mol. Phylogenet. Evol. 43, 216–224.
Ray, D.A., 2007. SINEs of progress: Mobile element applications to molecular ecology. Mol. Ecol. 16, 19–33.
Ray, D.A., Xing, J., Hedges, D.J., Hall, M.A., Laborde, M.E., Anders, B.A., White, B.R., Stoilova, N., Fowlkes, J.D., Landry, K.E., 2005. Alu insertion loci and platyrrhine primate phylogeny. Mol. Phylogenet. Evol. 35, 117–126.
Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22, 148–160.
Takahashi, K., Terai, Y., Nishida, M., Okada, N., 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol. Biol. Evol. 18, 2057–2066.
Takasaki, N., Murata, S., Saitoh, M., Kobayashi, T., Park, L., Okada, N., 1994. Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: a process of parasitization of entire genomes during the evolution of salmonids. Proc. Natl. Acad. Sci. U.S.A. 91, 10153–10157.
Takasaki, N., Yamaki, T., Hamada, M., Park, L., Okada, N., 1997. The salmon SmaI family of short interspersed repetitive elements (SINEs): Interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. Genetics 146, 369–380.

Tang, W.Q., Hu, X., 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.
Tao, J., Zhou, X.Q., Liu, H.B., Liu, H.Z., Yang, J., 2013. Two microchemistry patterns in otoliths of *Coilia nasus* from Poyang Lake China. J. Fish. China 37, 239–244.
Wang, W., Kirkness, E.F., 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res. 15, 1798–1808.
Weiss, J., Mallona, I., Gomez-di-Marco, P., Fernández-Valera, J.M., Egea-Cortines, M., 2012. Genotyping antirrhinum commercial varieties using miniature inverted-repeat transposable elements (MITEs). Sci. Hortic. 144, 161–167.
Whitehead, P., Nelson, G., Wongratana, T., 1988. FAO Species Catalogue. Clupeoid Fishes of the World (Suborder Clupeoidei). Part 2. Engraulididae. In: FAO Fisheries Synopsis, vol. 125, pp. 460–475.
Xie, J., 2012. Genetic and morphological variation of three freshwater lake populations of *Coilia ectenes* (Engraulidae). Russ. J. Genet. 48, 1029–1034.
Yang, J.Q., Hu, X.L., Tang, W.Q., 2008. Genetic structure and population history of *Coilia nasus* in Yangtze River and its south adjacent waters. J. Shanghai Fish Univ. 5, 513–519.
Yang, J.Q., Liu, D., Liu, Z.-Z., Wang, L.-J., Tang, W.-Q., 2012. Isolation and characterization of polymorphic microsatellite markers from Japanese grenadier anchovy (*Coilia nasus*). Conserv. Genet. Resour. 4, 519–521.
Yang, Q., Han, Z., Sun, D., Xie, S., Lin, L., Gao, T., 2010. Genetics and phylogeny of genus *Coilia* in China based on AFLP markers. Chin. J. Oceanol. Limnol. 28, 795–801.
Yuan, C., Qing, A., Liu, R., Lin, J., 1980. Discussion on Subspecific Taxonomy of the Genus *Coilia* in Middle and Lower Reaches of Chang Jiang and Southeast Coastal China Sea, vol. 3. Nanjing University, pp. 67–82.
Zhang, M., Xu, D., Liu, K., Shi, W., 2005. Studies on biological characteristics and change of resource of *Coilia nasus Schlegel* in the lower reaches of the Yangtze River. Resour. Environ. Yangtze Basin 14, 693–698.
Zhang, S.Y. (Ed.), 2001. Fauna Sinica-osteichthyes: Acipenseriformes, Elopiformes, Clupeiformes, Gonorhynchiformes. Science Press, Beijing, pp. 148–156.
Zhou, X., Yang, J., Tang, W., Liu, D., 2010. Species validities analyses of Chinese *Coilia* fishes based on mtDNA COI barcoding. Acta Zoo. Taxon. Sin. 35, 819–826.