GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic diversity of *Setipinna taty* (Engraulidae) populations from the China Sea based on mitochondrial DNA control region sequences

H.Y. Li, T.J. Xu, Y.Z. Cheng, D.Q. Sun and R.X. Wang

Laboratory for Marine Living Resources and Molecular Engineering,
College of Marine Science, Zhejiang Ocean University, Zhoushan, P.R. China

Corresponding author: T.J. Xu / R.X. Wang
E-mail: tianjunxu@163.com / wangrixin1123@126.com

Genet. Mol. Res. 11 (2): 1230-1237 (2012)
Received January 18, 2011
Accepted October 11, 2011
Published May 9, 2012
DOI http://dx.doi.org/10.4238/2012.May.9.1

ABSTRACT. The genetic diversity of *Setipinna taty*, which is commercially fished in the China Sea, was studied based on mitochondrial DNA control region sequences. PCR was used to amplify the control region fragment in 100 individuals of *S. taty* collected from Weihai (WH), Yantai (YT), Zhoushan (ZS), Xiangshan (XS), and Ninghai (NH) in China. A control region fragment of 656 bp was successfully sequenced in these 100 individuals. The A+T content of this *S. taty* control region was 71.7%; 172 variable sites and 62 haplotypes were found. Nucleotide diversity in the WH, YT, ZS, XS, and NH groups was 0.0228, 0.0247, 0.0441, 0.0126, and 0.0238, respectively. The haplotype diversity was 0.984, 0.911, 0.989, 0.926, and 0.979, respectively. Analysis of molecular variance showed that 97.95% of genetic variation was within populations, and only 2.05% among populations. The neighbor-joining phylogenetic tree obtained based on genetic distance showed that no significant genealogical structure exists throughout this range of *S. taty*. These results indicate no apparent geographical differentiation in the comparison of Yellow Sea and East China Sea populations of *S. taty*.

Within the control region, we identified an extended termination-associated sequence domain, a central conserved sequence block domain and a conserved sequence block domain; insertions of short tandem repeat sequence segments were found at the 5' end of the control region.

Key words: *Setipinna taty*; Mitochondrial DNA; Control region; Genetic diversity; Genetic structure

INTRODUCTION

The half-fin anchovy *Setipinna taty* belongs to the Engraulidae, and is distributed in the western part of the Indian Ocean and the Pacific Ocean. Generally, the *S. taty* attains a length of 15 cm and the fishing seasons are in spring and autumn. Spawning occurs in the Yellow Sea from February to April, and in the north of the East China Sea from May to June. *S. taty* spawns pelagic eggs and it is a carnivorous fish that eats zooplankton. The *S. taty* is an important commercial fish because of its delicious taste.

S. taty is a highly valued marine fish species; however, resources of *S. taty* have been reduced due to overfishing and water pollution. A number of studies on *S. taty* have been carried out, but most were related to ecological aspects, temporal and spatial distribution, community structure, and broodstock biological characteristics of *S. taty* (Sun and Ren, 2003; Zhang et al. 2004; Liu Y et al., 2004; Liu et al., 2006; Xiong et al., 2009; Guo et al., 2010). To date, little is known about the genetic variation of *S. taty.* As an important commercial fish, the lack of appropriate polymorphic markers has limited the study of the phylogenetics and population genetics of this species. Among various molecular markers, the sequences of mitochondrial DNA (mtDNA) are one of the most commonly used molecular markers, which have been successfully used in revealing population genetic diversity because they have several intrinsic characteristics, such as rapid evolution rate and relatively short generation time (Yang et al., 2008; Peng et al., 2010).

The mtDNA control region, which is also called the displacement-loop (D-loop) region, is located between $tRNA^{Pro}$ and $tRNA^{Phe}$ in mtDNA. The control region is the non-coding region of mitochondrial DNA, and it has been proven to be an ideal marker for assessing genetic structure of recently diverged or closely related populations or species (Avise, 1994; Bremer et al., 1996; Iguchi et al., 1999; Rand, 2000; Tabata and Taniguchi, 2000; Ishikawa et al., 2001). This region contains the DNA sequence associated with replication and transcription. Structurally, it is divided into three domains, including the termination-associated sequence domain, the central conserved sequence block domain and the conserved sequence block domain (Southern et al., 1988; Sbisa et al., 1997; Guo et al., 2004). The structure of the control region in some fishes has also been reported (Broughton et al., 1994; Lee et al., 1995; Zeng and Liu, 2001; Liu 2002; Guo et al., 2003; Zhang et al., 2003). However, there has been no reported study about the complete or partial sequence, conserved domains, and functional motifs of the control region of the *S. taty*. For these reasons, the objective of this research was to assess the genetic diversity of *S. taty* in these regions and analyze the structure of its control region. This information will in turn provide an important theoretical basis for the protection and sustainable utilization of this species.

MATERIAL AND METHODS

Sample collection and genomic DNA extraction

A total of 100 samples of *S. taty* were collected from five geographic locations, Weihai (WH), Yantai (YT), Zhoushan (ZS), Xiangshan (XS), and Ninghai (NH) in China. Twenty individuals were selected for per population. A small portion of the caudal fin was sampled and preserved in 95% ethanol. Total genomic DNA was isolated from the caudal fin using the standard phenol-chloroform method with some modification, and checked under UV light after 1.0% agarose gel electrophoresis, which was subsequently dissolved in 100 µL TE buffer and kept at 4°C for later use.

PCR amplification and sequencing

A primer pair was used to amplify the control region of *S. taty.* The two primer sequences in this procedure were MitDL-F (5'-CACCCYTRRCTCCCAAAGCYA-3') and MitDL-R (5'-GGTGCGGRKACTTGCATGTRTAA-3') (Huang et al., 2009). The primers used in this study were specifically designed to amplify the nearly complete control region of fish mtDNA. The primer pair has been successfully used in the amplification of the control region of nearly 20 fish species. The length of amplified products in this study was 1261 bp, which is longer than that reported by Huang et al. (2009), due to the insertions of tandem repeats. To analyze the genetic diversity of populations of *S. taty*, we examined the first hypervariable region of the control region; to obtain the nearly complete sequence of the control region, the amplified products were sequenced by using the pair of primers, and the sequence fragments were then edited with the Seqman program (DNAstar) for a contig assembly. The structural recognition was made by comparison with the key conserved sequence reported in the other fishes. The PCR amplifications were carried out in a 50-µL volume, which contained 35.5 µL sterilized water, 5 µL 10X PCR buffer (containing 1.5 mM Mg^{2+}), 4 µL 2.5 mM dNTPs, 2 µL of the forward and reverse primers, 1 µL diluted DNA template, and 2.5 U Taq DNA polymerase (Tiangen). The PCRs were performed using the following schedule: initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min, with a final extension at 72°C for 10 min and then holding at 4°C. PCR amplification was performed with an ABI 9700. The PCR products were checked on 1.5% agarose gel and then purified for direct sequencing by an automatic DNA sequencer (ABI 3730).

Data analyses

MEGA 4.0 was used to calculate statistical values such as the base composition of gene sequences and number of polymorphic sites. Meanwhile, multiple alignments of the sequences and the construction of a neighbor-joining dendrogram were also performed by MEGA 4.0. Haplotypic diversity (h), nucleotide diversity (P_i), number of haplotypes and gene flow (N_m) were calculated using DNAsp 4.0, which was used to estimate the genetic variability of *S. taty* populations. Analysis of molecular variance (AMOVA) was performed using the Arlequin 3.11 software to estimate how variation was partitioned among populations.

RESULTS AND DISCUSSION

Genetic variation

All sequences of the control region were edited and aligned using the MAGE 4.0 software. A total of 656 bp of the control region fragment were successfully sequenced for 100 individuals from five populations of *S. taty*. When all the DNA sequences were aligned, 172 variable sequence sites in the mtDNA control region were found among 100 individuals. The nucleotide composition of the mtDNA control region of the *S. taty* was calculated, in which the average A, T, C, and G contents were 41.0, 30.7, 19.3, and 9.0%, respectively, and the base distribution showed no apparent difference among populations. The results showed that the G bases were relatively low. The A + T content was higher than the G + C content among the sequences examined, which was consistent with previous findings that the control region is an A + T-rich region of the mitochondrial genome (Brown et al., 1986; Saccone et al., 1987, Cheng et al., 2010). For the five populations, the average transition/transversion ratio was 2.456, which suggested that the transition was higher than transversion, which is consistent with conclusions of other authors (Liu HY et al., 2004; Yang et al., 2008; Peng et al., 2010).

Among the 100 individuals, 172 variable sites (118 parsimony informative sites) and 62 haplotypes were found within the mtDNA control region. In general, three haplotypes (Hap1, Hap2, and Hap7) were shared by the five populations in this study, and four haplotypes (Hap9, Hap25, Hap30, and Hap31) were found in two populations, while the rest were all exclusively found in single populations. The haplotypes resulted from mutations, which demonstrated that the genetic diversity of *S. taty* is relatively rich. These yielded 62 haplotypes at the population level, which suggested that high genetic diversity exists in this species.

The number of haplotypes, the values of P_i, and h within each population are presented in Table 1. The average h and P_i among the five geographic groups were 0.958 and 0.0255, respectively. In addition, the haplotype diversity and nucleotide diversity showed little difference in the five populations, where the ZS population showed the highest haplotype diversity and nucleotide diversity, which suggested that the genetic diversity of the ZS population is rich. The results revealed abundant nucleotide diversity in *S. taty*.

Table 1. The parameters of genetic diversity of *Setipinna taty*.

Population	Number of haplotypes	Haplotype diversity (h)	Nucleotide diversity (P_i)
S. taty (WH)	18	0.984	0.0228
S. taty (YT)	13	0.911	0.0247
S. taty (ZS)	18	0.989	0.0441
S. taty (XS)	12	0.926	0.0126
S. taty (NH)	17	0.979	0.0238
Average value	62	0.958	0.0255

WH = Weihai; YT = Yantai; ZS = Zhoushan; XS = Xiangshan; NH = Ninghai.

The genetic distance between the ZS and YT populations was the highest, up to 0.045, while that between the XS and WH populations was the lowest at 0.020 (Table 2). Among the five populations, the average genetic distance was 0.031. The genetic distances within groups were 0.026, 0.029, 0.062, 0.014, and 0.027, respectively. The genetic distance between populations reflected genetic relationships between the populations. In the five experimental

populations, the XS and WH populations showed the lowest genetic distance, which might have been due to the time of population differentiation being short.

Table 2. Genetic distance (below the diagonal) between *Setipinna taty* stocks.

Population	WH	YT	ZS	XS
WH	-	-	-	-
YT	0.027	-	-	-
ZS	0.043	0.045	-	-
XS	0.020	0.021	0.037	-
NH	0.026	0.028	0.044	0.021

For population abbreviations, see Table 1.

Gene differentiation (F_{ST}) and gene flow (N_m) are important indices evaluating the population genetic structure (Zhang et al., 2007). F_{ST} and N_m were also calculated to characterize the gene diversity and the distribution of the variation using the DNAsp software. The results showed that the pairwise genetic differentiation index F_{ST} showed no significant genetic differentiation between each population and revealed that the majority of genetic differentiation resided within populations. N_m was low among the five populations, and the results suggested little genetic differentiation among the populations. This might have been caused by fish migration and the short distances between the populations.

The genetic variation based on AMOVA showed that 97.95% of genetic variation resided within populations, while between populations it was only 2.05% (Table 3). This result suggested that less significant geographical division was present, which may suggest a weak and unstable regional genetic structure in this species.

Table 3. Analysis of molecular variance (AMOVA) for populations of *Setipinna taty* stocks.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	F_{ST}	P
Between populations	4	77.550	0.2865	2.05	0.0206	0.0850
Within populations	95	1297.400	13.6568	97.95		
Total	99	1374.950	13.9434			

d.f. = degrees of freedom.

A neighbor-joining tree was constructed based on Kimura 2-paramter genetic distances of the *S. taty* mtDNA control region sequences (Figure 1). The results indicated that no significant genealogical structure existed throughout the examined range of *S. taty*. In Figure 1, we can see that there is some crossing between the Yellow Sea and East China Sea populations, which also shows no obvious geographical isolation with respect to Yellow Sea and East China Sea populations.

Structure of control region

The control region of *S. taty* was determined to be 1261 bp long, and had an overall base composition that was rich in A and T (A + T = 71.7%). Within this sequence, an extent termination-associated sequence (ETAS), central conserved sequence block domain (CSB-D, CSB-E and CSB-F) and conserved sequence block domain (CSB-1, CSB-2 and CSB-3) can be easily identified in *S. taty* mtDNA (Figure 2), which are the characteristics of the control region. The conserved ETAS sequence is "TACATACTATGCATTATAT" and the motif sequence of

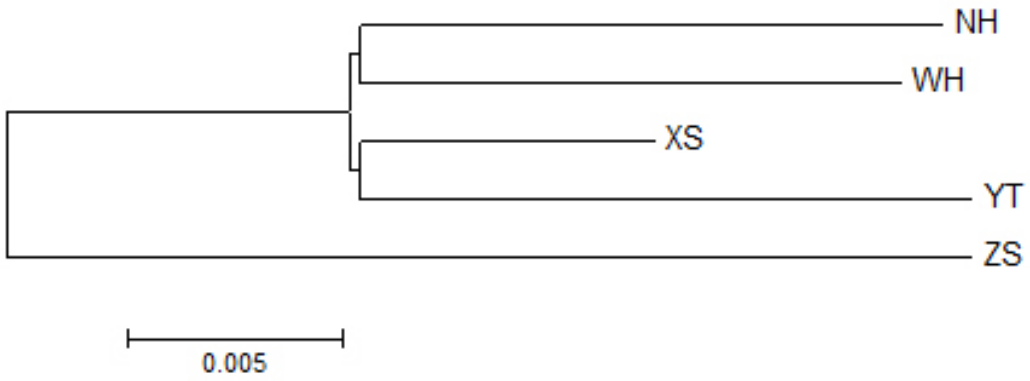

Figure 1. Molecular phylogenetic tree of five stocks of *Setipinna taty*. For population abbreviations, see Table 1.

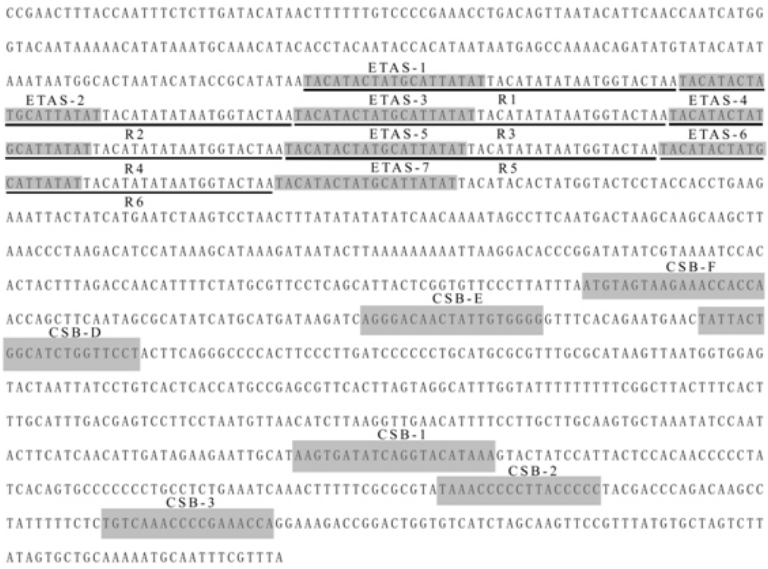

Figure 2. Sequence and structure of the control region of *Setipinna taty*. The tandem repeats were underlined and marked with R1, R2, R3, R4, R5, and R6, respectively. The conserved domains were shadowed and marked (ETAS, CSB-F, CSB-E, CSB-D, CSB-1, CSB-2, and CSB-3).

ETAS is TACAT with one palindromic sequence, ATGCA, which was detected in *S. taty*. The same phenomenon was reported in *Coilia* fishes (Zhu et al., 2008). On the other hand, the ETAS sequence (TACATACTATGCATTATAT) of 19 bp was identified, where the motif was repeated six times. The mechanism of slippage and mispairing during replication of the mitogenome could explain the tandem repeats in the control region (Broughton and Dowling, 1997). Several fragments comparable to termination-associated sequences were identified in ETAS, which are thought to act as a signal for the termination of H-strand elongation, and this domain is a hyper-

variable domain that may be useful in analyzing interspecies variation within *S. taty* populations. Insertions of short tandem repeat sequence segments were found at the 5' end of the control region, and a tandem repeated sequence of 39 bp was identified that was repeated six times, which are the characteristics of this region. The short tandem repeat sequence segments in the control region of *S. taty* may be caused by slide mismatches, which exhibited length variability between individuals. In the central conserved sequence block domain, we identified the CSB-F, CSB-E, and CSB-D regions. The consensus sequence of CSB-F is "ATGTACTAAGAAAACCACCA", which serves to differentiate the central conserved sequence block domain from the termination-associated sequence domain. CSB-E is located downstream of CSB-F, and its consensus sequence is "AGGGACAACTATTGTGGGG", characterized by the GTGGGG box. CSB-E was followed by CSB-D (consensus sequence "TATTACTGGCATCTGGTTCCT"). The consensus sequences of CSB-F, CSB-E, and CSB-D in *S. taty* were highly conserved and consistent with those described in other fishes. In addition, the sequences of conserved sequence blocks CSB-1 (ATAAGTGATATCAGGTACATAAA), CSB-2 (TAAACCCCCTTACCCCC) and CSB-3 (TGTCAAACCCCGAAACCA) were conserved, which is consistent with the CSB-2 and CSB-3 identified in other fishes (Liu and Cui, 2009), and they are thought to be involved in positioning RNA polymerase both for transcription and for priming replication (Clayton, 1991; Shadel and Clayton, 1997). These features suggest that the asymmetrical replication mechanism revealed for mammalian mitochondrial DNA also operates in *S. taty*.

ACKNOWLEDGMENTS

Research supported by the Seed Grant Programs of Science and Technology Commission Foundation of the Zhejiang Province and the Open Foundation from Ocean Fishery Science and Technology in the Most Important Subjects of Zhejiang (#20110206).

REFERENCES

Avise JC (1994). Molecular Markers, Natural History and Evolution. Sinauer Associates, New York.

Bremer JRA, Mejuto J, Greig TW and Ely B (1996). Global population structure of the swordfish (*Xiphias gladius* L.) as revealed by analysis of the mitochondrial DNA control region. *J. Exp. Mar. Biol. Ecol.* 197: 295-310.

Broughton RE and Dowling TE (1994). Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics* 138: 179-190.

Broughton RE and Dowling TE (1997). Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella spiloptera*. *Mol. Biol. Evol.* 14: 1187-1196.

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.

Cheng Y, Xu T, Shi G and Wang R (2010). Complete mitochondrial genome of the miiuy croaker *Miichthys miiuy* (Perciformes, Sciaenidae) with phylogenetic consideration. *Mar. Genomics* 3: 201-209.

Clayton DA (1991). Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends Biochem. Sci.* 16: 107-111.

Guo B, Zhang B, Dai FQ and Jin XS (2010). Diet composition and ontogenetic variation in feeding habits of juvenile *Setipinna taty* in the Haizhou bay. *J. Fish. China* 34: 741-747.

Guo X, Liu S and Liu Y (2003). Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. *Aquaculture* 224: 25-38.

Guo XH, Liu SJ, Liu Q and Liu Y (2004). New progresses on mitochondrial DNA in fish. *Yi Chuan Xue Bao* 31: 983-1000.

Huang ZJ, Xu XP, Tang JJ, Zhang JQ, et al. (2009). Application and primer design of mitochondrial DNA D-loop of freshwater fishes. *Acta Sci. Nat. Univ. Sunyatseni.* DOI [CNKI:SUN:ZSDZ.0.2009-04-017].

Iguchi K, Tanimura Y, Takeshima H and Nishida M (1999). Genetic variation and geographic population structure of

amphidromous ayu *Plecoglossus altivelis* as examined by mitochondrial DNA sequencing. *Fish. Sci.* 65: 63-67.
Ishikawa S, Aoyama J, Tsukamoto K and Nishida M (2001). Population structure of the Japanese Eel, *Anguilla japonica* as examined by mitochondrial DNA sequencing. *Fish. Sci.* 67: 246-253.
Lee WJ, Conroy J, Howell WH and Kocher TD (1995). Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol.* 41: 54-66.
Liu HY, Jing SG, Su TF and Gong SY (2004). Polymorphism study of the mitochondrial DNA D-loop gene sequences from *Sparus latus*. *J. Fish. China* 28: 371-374.
Liu HZ (2002). The structure and evolution of the mtDNA control region in fish: taking example for Acheilognathinae. *Prog. Nat. Sci.* 12: 266-270.
Liu Y, Cheng J and Li S (2004). A study on the distribution of *Sefipinna taty* in the East China Sea. *Mar. Fish.* 26: 255-260.
Liu Y, Cheng JH and Chen XG (2006). Studies on the seasonal distribution of *Setipinna taty* in the East China Sea. *Mar. Fish. Res.* 27: 1-6.
Liu Y and Cui Z (2009). The complete mitochondrial genome sequence of the cutlassfish *Trichiurus japonicus* (Perciformes: Trichiuridae): Genome characterization and phylogenetic considerations. *Mar. Genomics* 2: 133-142.
Peng SM, Shi ZH and Hou J (2010). Comparative analysis on the genetic diversity of cultured and wild silver pomfret populations based on mtD-loop and COI gene. *J. Fish. China* 34: 19-24.
Rand E (2000). Mitochondrial DNA. Blackwell, Oxford.
Saccone C, Attimonelli M and Sbisa E (1987). Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. *J. Mol. Evol.* 26: 205-211.
Sbisa E, Tanzariello F, Reyes A, Pesole G, et al. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205: 125-140.
Shadel GS and Clayton DA (1997). Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* 66: 409-435.
Southern SO, Southern PJ and Dizon AE (1988). Molecular characterization of a cloned dolphin mitochondrial genome. *J. Mol. Evol.* 28: 32-42.
Sun SD and Ren YP (2003). Study on the fishery biology of *Setipinna taty* in the southern yellow sea. *Trans. Oceanol. Limnol.* 1: 62-65.
Tabata K and Taniguchi N (2000). Differences between *Pargus major* and *Pagrus auratus* through mainly mtDNA control region analysis. *Fish. Sci.* 66: 9-18.
Xiong Y, Tang JH, Liu PT, Zhang XM, et al. (2009). Resource estimate on *Setipinna taty* in the southern yellow sea. *Oceanol. Limnol.* 40: 500-505.
Yang B, Chen XY and Yang JX (2008). Structure of the mitochondrial DNA control region and population genetic diversity analysis of *Anabarilius grahami* (Regan). *Zool. Res.* 29: 379-385.
Zeng QL and Liu HZ (2001). Study on mitochondrial DNA control region of the *Ictiobus cypriellus*. *J. Hubei Univ.* 23: 261-264.
Zhang C, Chen X, He T, Liu X, et al. (2007). Genetic structure of *Malus sieversii* population from Xinjiang, China, revealed by SSR markers. *J. Genet. Genomics* 34: 947-955
Zhang MH, Wang Y and Zhang J (2004). Studies on the growth and death character of *Setipinna taty* in the South of Bohai Sea. *J. Zhejiang Ocean Univ.* 23: 31-36.
Zhang Y, Zhang E and He SP (2003). Studies on the structure of the control region of the bagridae in China and its phylogenetic significance. *Acta Hydrobiol. Sin.* 27: 463-467.
Zhu TJ, Yang JQ and Tang WQ (2008). MtDNA control region sequence structure of the genus *Coilia* in Yangtze river estuary. *J. Shanghai Fish. Univ.* 17: 152-157.