Microsatellite DNA diversity of *Coilia mystus* (Clupeiformes: Engraulidae) in three Chinese estuaries

Fang Chen\(^1,2\); Qiqun Cheng\(^1,*\)

Received: July 2011    Accepted: November 2011

Abstract

An analysis of eight microsatellite loci in 90 individuals was performed to define the genetic structure and variability of three estuarine populations of *Coilia mystus*: 30 individuals each collected from ChangJiang River (Yangtze River) estuary-CJ, MinJiang River estuary-MJ, and ZhuJiang River (Pearl River) estuary-ZJ. The results showed polymorphic information contents (*PIC*) of 0.78, 0.77, and 0.64 in CJ, MJ, and ZJ, respectively. The mean observed number of alleles (*N_a*) ranged from 7.38 to 11.88. The mean observed and expected heterozygosities ranged from 0.09 to 0.21 and from 0.68 to 0.81, respectively. The ZJ population was least polymorphic. Highly significant deviations from Hardy-Weinberg equilibrium, mostly due to deficits of heterozygote, were found in these three populations. Pairwise *F_{ST}* and Reynolds’ distance indicated that the three estuarine populations were genetically distinct, in accordance with the principal component analysis (PCA) and the Bayesian model-based clustering algorithm analysis. The results showed that pairwise genetic differentiation among these three estuarine populations were relatively high, with possible divergence to subspecies level. This study provides microsatellite DNA evidence for assessing the genetic distinctness of *C. mystus* populations and will benefit fishery resource management and sustainable utilization of *C. mystus*.

Keywords: *Coilia mystus*, Microsatellite, Genetic diversity, Population structure, Genetic divergence

---

1- Key Laboratory of East China Sea and Oceanic Fishery Resources Exploitation, Ministry of Agriculture, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China.

2- College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China.

*Corresponding author’s email: qiquncheng@gmail.com
Introduction

Coilia mystus (Linnaeus 1758) is a small to moderate-sized anchovy in the genus Coilia, family Engraulidae and order Clupeiformes. It is distributed widely in Chinese estuaries, and is one of the most important species to China’s estuarine fishery economy. C. mystus is a short-distance anadromous fish, which migrates from waters near the ocean to brackish water areas every year during the spawning season. Mature C. mystus migrates upriver, spawns in the brackish waters, and never enters pure freshwater areas (East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences and Shanghai Fisheries Research Institute, 1990). The ChangJiang River (Yangtze River), MinJiang River, and ZhuJiang River (Pearl River) are three important rivers, which are separated from each other by long distances. Therefore, local adaptation, overfishing, and environmental conditions may contribute to genetic divergence of the three populations.

Previously, we have studied the divergence of these three populations by using morphological and mitochondrial indices. We measured morphological differences between these three populations of C. mystus, which were difficult to differentiate using the meristic index, and found great morphological differences between them which can be discriminated with synthetical indices. The Minjiang population was not closely clustered with either the Changjiang or the Zhujiang populations (Cheng et al., 2005; Cheng, 2010). We also compared the partial cytochrome b and 16S rRNA gene sequences of these three populations, and hypothesized that they were diverged at the subspecies level (Cheng et al., 2006, 2008). The significant morphological differences between populations of C. mystus were in concordance with genetic divergence of mitochondrial DNA data.

Microsatellites markers, which are neutral genetic markers, abundant in fish genomes, with high levels of polymorphism (Tautz, 1989), are simple tandemly repeated sequence motifs consisting of repeat units of 1-6 bp. Microsatellite markers are very useful for molecular ecology and population genetic studies of fishery resources. They have been widely used in many species for evaluating population genetic diversity, constructing genetic maps, and determining pedigrees (McDonald et al., 2004; Guyomar et al., 2006; Pettay and Lajeunesse, 2009). We have developed several microsatellite markers in C. ectenes, a closely related species of C. mystus, and found these markers are suitable for cross-species amplification in C. mystus (Ma et al., 2011).

The research of genetic diversity of species can help us understand the genetic structure of species and population history and then comprehend their evolution for a conservation strategy. In the present study, we applied eight microsatellite primer pairs to study genetic diversity and variability of
three estuarine populations of *C. mystus*. The present study provides microsatellite DNA evidence for assessing the genetic distinctness of *C. mystus* populations and will benefit fishery resource management and sustainable utilization of *C. mystus*.

**Materials and methods**

*Sample collection and DNA extraction*

A total of 90 *C. mystus* were obtained from Changjiang (CJ), Minjiang (MJ), and Zhujiang (ZJ) river estuaries, with 30 individuals of each (Table 1 and Fig. 1). All individuals were transferred to the laboratory in dry ice and then stored at -80°C before DNA extraction. Genomic DNA was extracted from muscle tissues. Muscle tissues were dissected and then digested by proteinase K, followed by phenol-chloroform extraction and 100% ethanol precipitation (Sambrook et al., 1989).

*Microsatellite analysis*

We genotyped all 90 individuals at 8 microsatellite loci (D13, D55, D58, D63, D81, D95, D104, D114) following our established methods (Ma et al., 2011). PCR reactions were carried out in a 25µL volume comprising 100 ng of genomic DNA, 10×PCR buffer (Mg²⁺), 10 µM of each primer, 2.5 mM of dNTP mix, and 2.5 U of *Taq* polymerase (Takara). The thermocycling conditions were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, annealing temperature (see Table 1 in Ma et al., 2011) for 30 s, and 72°C for 1 min, and then a final extension of 72°C for 7 min. PCR products were separated in 3% agarose gels. The range of allele sizes was determined by referring to the *Mse*I marker (TianGen Biotech) and molecular tools in Smartview Analysis Software (FuRi Tech).

*Statistical analysis*

For each marker and population, POPGENE 1.31 software (Yeh et al., 1999) was used to estimate allele frequencies (*P*), observed number of alleles (*Nₐ*), effective number of alleles (*Nₑ*), observed heterozygosity (*Hₒ*), expected heterozygosity (*Hₑ*), Polymorphism information content (*PIC*) was computed according to the following formula (Botstein et al., 1980):

\[
PIC = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} 2P_i^2P_j^2
\]

where *Pᵢ* and *Pⱼ* are the frequencies of the *i*th and *j*th alleles at one locus, and *n* is the number of alleles at one locus.

GENEPOP 1.2 software (Raymond and Rousset, 1995) was used to estimate departures from Hardy-Weinberg equilibrium at each locus, which implements Fisher’s exact tests for multiple alleles (Guo and Thompson, 1992). The null allele frequency was estimated by MICRO-CHECKER version 2.2.3 (Van Oosterhout et al., 2004).

Genetic differences between populations were evaluated by calculating pairwise *Fₛₜ* values and Reynold’s genetic
distance using ARLEQUIN 3.0 (Excoffier et al., 2005). As an alternative approach to represent the genetic relationship among the *C. mystus* populations, a Principal Component Analysis (PCA) was also performed based on gene frequencies of all the variable loci, using the GENETIX software package (Belkhir et al., 2000). A Bayesian model-based clustering algorithm implemented in the computer program STRUCTURE 2.0 (Pritchard et al., 2000) was also applied. This model-based algorithm places the given individuals into *K* clusters, where *K* is chosen in advance but can be varied across independent runs of the algorithm. All runs were carried out under “admixture model”, with a burn-in period of 10,000 iterations and a period of data collection of 10,000 iterations.

**Results**

*Within population variation*

All the eight microsatellite markers were successfully amplified in all 90 individuals of these three populations and were found to have high polymorphic levels. Alleles ranged in size from 100 to 310 bp over the eight loci. In these three populations, the observed number of alleles across loci in the populations ranged from 2 alleles in ZJ to 20 alleles in CJ. The mean observed number of alleles (*N*<sub>o</sub>) ranged from 7.38 to 11.88 across populations. The mean effective alleles (*N*<sub>e</sub>) per population ranged from 3.97 to 7.68. Polymorphism information contents (*PIC*) were between 0.22 and 0.92, and the averages were 0.78, 0.77, and 0.64 in CJ, MJ, and ZJ, respectively. The mean observed and expected heterozygosities for the three populations ranged from 0.09 to 0.21 and from 0.68 to 0.81, respectively (Table 1). The average observed heterozygosity was lower than the expected heterozygosity at all the loci. These parameters of genetic diversity indicated that there were moderate-to-high levels of polymorphism within these three populations. An exact *P*-value test indicated that all loci significantly deviated from HWE and was associated with heterozygote deficiencies (Table 1). Based on average *F*<sub>IS</sub> values, it can be seen that the pattern of heterozygote deficiencies was pronounced in all populations.

*Genetic differentiation and relationships between populations*

The pairwise *F*<sub>ST</sub> and Reynolds’ distance values for the three populations are shown in Table 2. The three populations were well differentiated from each other (*F*<sub>ST</sub> = 0.116-0.161). MJ and ZJ was the most divergent (*F*<sub>ST</sub> = 0.161) followed by CJ and ZJ (*F*<sub>ST</sub> = 0.150). The divergence between the CJ and MJ was lowest (*F*<sub>ST</sub> = 0.116). The pairwise *F*<sub>ST</sub> comparisons revealed a relatively high degree of genetic differentiation between these three populations. The Reynolds’ distances ranged from 0.156 to 0.231. The genetic distance tended to be the least (0.156) between CJ and MJ and the widest (0.231) between CJ
and ZJ. PCA based on the correlation matrix obtained from the allele frequencies was also performed (Fig. 2). The first axis explained 29.5% of the inertia and distinguished ZJ from the rest of the populations. The second axis contributed 17.7% of the inertia and separated CJ from MJ. The result indicated that these three populations could be separated clearly by these eight microsatellite markers. The STRUCTURE 2.0 software was used to evaluate the assignments of individuals to 1-3 assumed populations with three simulations (Fig. 3). The three populations were separated at $K = 3$, which represented the membership of each individual to the three clusters. It matches the result of the Principal Component Analysis.

Figure 1: Locations of sampling sites (numbers). 1. ChangJiang River (Yangtze River) estuary, Shanghai; 2. MinJiang River estuary, Fuzhou; 3. ZhuJiang River (Pearl River) estuary, Guangzhou

Figure 2: Scattergram showing relative positions of three *Coilia mystus* populations defined by PC1 vs PC2 of eight microsatellite loci
Figure 3: Clustering assignment of the three populations provided by STRUCTURE analyses

Table 1: Estimation of genetic variability of three Coilia mystus populations at 8 microsatellite loci

<table>
<thead>
<tr>
<th>Pop/Locus</th>
<th>D13</th>
<th>D55</th>
<th>D58</th>
<th>D63</th>
<th>D81</th>
<th>D95</th>
<th>D104</th>
<th>D114</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>15</td>
<td>16</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>4</td>
<td>20</td>
<td>3</td>
<td>11.75</td>
</tr>
<tr>
<td>Ne</td>
<td>10.58</td>
<td>8.85</td>
<td>4.84</td>
<td>8.84</td>
<td>4.59</td>
<td>2.94</td>
<td>10.45</td>
<td>2.65</td>
<td>6.72</td>
</tr>
<tr>
<td>Ho</td>
<td>0.10'</td>
<td>0.07'</td>
<td>0.00'</td>
<td>0.26'</td>
<td>0.93'</td>
<td>0.00'</td>
<td>0.32'</td>
<td>0.00'</td>
<td>0.21</td>
</tr>
<tr>
<td>He</td>
<td>0.91</td>
<td>0.89</td>
<td>0.79</td>
<td>0.89</td>
<td>0.78</td>
<td>0.66</td>
<td>0.9</td>
<td>0.62</td>
<td>0.81</td>
</tr>
<tr>
<td>PIC</td>
<td>0.9</td>
<td>0.88</td>
<td>0.78</td>
<td>0.88</td>
<td>0.76</td>
<td>0.59</td>
<td>0.9</td>
<td>0.55</td>
<td>0.78</td>
</tr>
<tr>
<td>MJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>18</td>
<td>17</td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>17</td>
<td>10</td>
<td>11.88</td>
</tr>
<tr>
<td>Ne</td>
<td>13.24</td>
<td>11.2</td>
<td>9.89</td>
<td>7.5</td>
<td>3.91</td>
<td>1.77</td>
<td>11.11</td>
<td>2.82</td>
<td>7.68</td>
</tr>
<tr>
<td>Ho</td>
<td>0.24'</td>
<td>0.25'</td>
<td>0.00'</td>
<td>0.11'</td>
<td>0.00'</td>
<td>0.00'</td>
<td>0.13'</td>
<td>0.00'</td>
<td>0.09</td>
</tr>
<tr>
<td>He</td>
<td>0.92</td>
<td>0.91</td>
<td>0.9</td>
<td>0.87</td>
<td>0.74</td>
<td>0.44</td>
<td>0.91</td>
<td>0.64</td>
<td>0.79</td>
</tr>
<tr>
<td>PIC</td>
<td>0.92</td>
<td>0.9</td>
<td>0.89</td>
<td>0.85</td>
<td>0.7</td>
<td>0.37</td>
<td>0.9</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>ZJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>7</td>
<td>7.38</td>
</tr>
<tr>
<td>Ne</td>
<td>6.48</td>
<td>6.28</td>
<td>3.58</td>
<td>2.41</td>
<td>2.78</td>
<td>1.34</td>
<td>4.89</td>
<td>4.04</td>
<td>3.97</td>
</tr>
<tr>
<td>Ho</td>
<td>0.22'</td>
<td>0.24'</td>
<td>0.14'</td>
<td>0.21'</td>
<td>0.00'</td>
<td>0.00'</td>
<td>0.13'</td>
<td>0.00'</td>
<td>0.12</td>
</tr>
<tr>
<td>He</td>
<td>0.85</td>
<td>0.84</td>
<td>0.72</td>
<td>0.59</td>
<td>0.64</td>
<td>0.25</td>
<td>0.8</td>
<td>0.75</td>
<td>0.68</td>
</tr>
<tr>
<td>PIC</td>
<td>0.83</td>
<td>0.82</td>
<td>0.67</td>
<td>0.55</td>
<td>0.58</td>
<td>0.22</td>
<td>0.77</td>
<td>0.72</td>
<td>0.64</td>
</tr>
</tbody>
</table>

*: Significant deviation from Hardy–Weinberg equilibrium

Table 2: Genetic distances (Reynold’s distance, above diagonal) and differentiation (F_{ST}, below diagonal) between Coilia mystus populations

<table>
<thead>
<tr>
<th></th>
<th>CJ</th>
<th>MJ</th>
<th>ZJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ</td>
<td>0.79</td>
<td>0.79</td>
<td>0.94</td>
</tr>
<tr>
<td>MJ</td>
<td>0.83</td>
<td>0.82</td>
<td>0.67</td>
</tr>
<tr>
<td>ZJ</td>
<td>0.79</td>
<td>0.79</td>
<td>0.94</td>
</tr>
</tbody>
</table>

*: Significant deviation from Hardy–Weinberg equilibrium
Discussion

Genetic diversity within populations

Biological population diversity is an important foundation for evaluating species resources. It is true that the species to adapt to different environments. It is a precondition to use the species resource persistently, to keep the highest level of genetic variation. $N_a$, $H_o$, $H_e$ and $PIC$ are all parameters for measuring population genetic variations.

On the basis of the present study, analyses of eight microsatellite loci revealed relatively high levels of genetic variation. The mean effective alleles and mean $PIC$ per population was from 3.97 to 7.68 and from 0.64 to 0.78, respectively. Bostein et al. (1980) indicated that a locus was considered highly polymorphic when $PIC \geq 0.50$, low polymorphic when $PIC \leq 0.25$, and moderate polymorphic when the $PIC$ value was between 0.25 and 0.50. The $PIC$ here ranged from 0.64 to 0.78, which means there is a high level of polymorphism in these three estuarine populations and it is suitable for estimating genetic diversity of $C. mystus$ populations. The observed allele numbers and allele ranges here were relatively higher than that observed in a previous study of $C. mystus$ based on 12 microsatellite loci (Ma et al., 2011). The mean observed and expected heterozygosities were from 0.09 to 0.21 and from 0.68 to 0.81, respectively. It can be speculated that these discrepancies may be due to the sharp decline in number of $C. mystus$ in the recent years, which may result in low observed genetic heterozygosities in populations. This may indicate loss of genetic diversity and the simpler genetic structure in the three estuarine populations, which would reduce the environmental adaptability.

Cheng et al. (2008) studied the partial mitochondrial cytochrome $b$ and 16S rRNA gene sequences of these three populations, and indicated that genetic diversity was lowest within the ZJ population, which is in concordance with the present microsatellite analysis. Human activity may be influencing genetic diversity within the three populations. Guangzhou is highly developed economically with human activities, which may be lowering the genetic diversity of $C. mystus$ populations through overharvesting.

This study showed that all loci significantly deviated from the Hardy-Weinberg equilibrium. Biased samples and small sample size may lead to this phenomenon. However, these two reasons can be excluded in this study. $C. mystus$ is one of the main economic fish in the Changjiang, Minjiang, and Zhujiang river estuaries and the samples were collected randomly. Barker et al. (1993) proposed that at least 25 individuals should be analyzed for each species/population for an acceptable statistical resolution. In this study, 30 individuals were analysed for each population, and they were therefore representative of the population genetic
Generally, heterozygote deficit may be linked to multiple causes: population substructure, inbreeding, null alleles, the Wahlund effect, and natural selection (Ardren et al., 1999; Lade et al., 1996; Castric et al., 2002; Morand et al., 2002). *C. mystus* is inhabited in Chinese estuary and no obvious geographical barriers separate migration and gene flow between individuals, so population substructure may not exist. Fixed factor $F_{IS}$ is an important parameter to evaluate the occurrence of inbreeding in populations. $F_{IS}$ values for each population are relatively high, suggesting that inbreeding exists in each population. Over-fishing and environmental pollution could provide a reasonable explanation for inbreeding within population. Null alleles within population may be another factor to account for the increased homozygosity (Pemberton et al., 1995) as has been observed in related species due to the sequence mutation of microsatellite flanking (Harper et al., 2003). However, in all samples, though each locus deviated from the Hardy-Weinberg equilibrium, at least one allele can be stably amplified, so the frequency of null alleles was not high enough to affect the results of this study significantly (Goodman et al., 2001; Abbott and Double, 2003; Harper et al., 2003; Lucchini et al., 2004). The Wahlund effect may also be excluded, because no significant hidden substructure was detected in the *C. mystus* population. Natural selection, which usually acts on long-term evolutionary processes, was also considered as a possible cause of heterozygote deficiency. The different environmental factors such as water temperature, water current, salinity, food availability, and fishing intensity in these three rivers may influence populations throughout the sampling period.

**Relationships and genetic differentiations between populations**

The values of $F_{ST}$ and Reynolds’ distance suggested that the three populations have evolved independently and have developed significant genetic structure with possible divergence to subspecies level. This result was in concordance with morphological differences (Cheng, 2010) and genetic divergence of mitochondrial DNA data (Cheng et al., 2006, 2008).

According to this research, both $F_{ST}$ and Reynolds’ distance values showed that MJ and CJ were closely related while CJ and ZJ populations were farthest. The molecular results matched the results of the previous morphological study. Cheng et al. (2005) also found that there were great morphological differences between the Changjiang and the Zhujiang populations. The most likely reason for this finding could be due to geographic proximity. The long distance separating Guangzhou and Shanghai (1,689 km), coupled with the short-distance migrating habit of *C. mystus*, affords little chance for gene flow between the ZJ and CJ populations.
PCA and Structure analyses clearly revealed the genetic structuring of the investigated population into three isolated clusters with restricted gene flow among them. The potential capacity of populations to adapt and evolve as independent biological entities in different environmental conditions is restricted by the exchange of individuals between populations. A sufficient degree of isolation may result in notable phenotypic and genetic differentiation among fish populations within species, which may be recognizable as a basis for separation and management of distinct populations (Turian, 2004). Geographic isolation, living environment, population genetic bottleneck, gene flow, and selection have large effects on the genetic construction of populations. Heredity and variation are necessary to adapt to the changing environment. Population genetic research on a wide distributing species can elucidate population genetic variability and structuring.

In summary, the three estuarine populations of *C. mystus* maintain a moderate-to-high level of genetic diversity within populations and higher genetic differentiation among populations. The status of genetic diversity in these three estuary populations may affect the management and utilization of this species. Some necessary measures should be implemented to decrease harmful human activities, such as overfishing, environmental pollution, and irrigation projects. Only when the pressures on these populations are reduced, can the genetic diversity of *C. mystus* in these waters be maintained and restored.

**Acknowledgements**

The author would like to thank Mrs. Jindi Han and Mr. Qizhe Liu for assistance in sample collection. This work was supported by the Special Research Fund for national non-profit institutes (East China Sea Fisheries Research Institute, Project No.2008Z02) and the National Natural Science Foundation of China (No.30600456).

**References**


Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F., 2000. GENETIX 4.01, logiciel sous Windows™ pour la gé­né´tique des populations. Laboratoire Ge­nome, Populations, CNRS UPR 9060.Interactions, Université de Montpellier II, Montpellier, France.


East China Sea Fisheries Research Institute, Chinese Academy of Fisheries Science and Shanghai Fisheries Research Institute, 1990. The fishes of Shanghai Area. Shanghai Scientific and Technical Publishers, Shanghai.


http://www.ualberta.ca/~fyeh/.