Studies on genetic diversity of Chinese mitten crab Eriocheir sinensis of Yangtze River system based on mitochondrial DNA control region

Haihua Wang¹, *, Guangpeng Feng², b and Yanping Zhang¹, *

¹Jiangxi Institute for Fisheries Sciences, Nanchang, China
²East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai, China

*Corresponding author e-mail: 775069405@qq.com, *jxswhh@163.com, bcoolwindfgp@qq.com

Abstract. We obtained mtDNA control region sequences of 102 E. sinensis individuals of 7 populations in Yangtze River system (Adults of E. sinensis in Huangmai, Poyang Lake, Pengze, Anqing, Zhenjiang and the Yangtze estuary, juveniles of E. sinensis in Zhenjiang and the Yangtze estuary , abbreviated as HM, PY, PP, AQ, ZC, ZY and CY) and analysed their genetic diversity. The resulted showed that the complete sequences of control region was 515 bp. Total of 65 variation sites including 31 singleton polymorphic sites and 34 parsimony informative polymorphic sites were defined in the 102 individuals sampled of 7 populations, in which 60 haplotypes were detected. The haplotype diversity (Hd) was found to be 0.919±0.023, nucleotide diversity 0.01007±0.00081, the average nucleotide difference (K) 5.106. Genetic differentiation index (Fst) among populations was -0.0098 to 0.0983; genetic distance was 0.0067 to 0.0122; the genetic variation among populations accounts for only 3.3% of total variation. Mismatch distribution suggested that E. sinensis in the Yangtze River system had not experienced a rapid population expansion in the recent past. The results showed that genetic diversity level of E. sinensis in the Yangtze River system was high. The seven populations had not been formed geographic differentiation obviously and apomixes would occur between them.

1. Introduction

With characteristics like matrilineal inheritance, stable molecular structure and high evolutionary rate [1], mitochondrial DNA becomes an effective molecular marker to study origin of species and phylogeny. Mitochondrial DNA control region sequence does not need to encode protein, so it possesses a higher evolutionary rate when compared with genes in other parts of mitochondria. Hence it is extensively applied to studies on genetic structure diversity and population history of species or populations with a close genetic relationship. In this paper, through the exploration on hyper variable region sequence in mitochondrial DNA control region of E. sinensis, an analysis was made on genetic diversity and population genetic structure of 7 wild populations collected from Huangmei, Poyang Lake, Pengze, Anqing, Zhenjiang and Shanghai SanJia Harbor in Yangtze River System. This study aimed to know about variation characteristics of the natural population of E. sinensis in different river
sections of Yangtze River System after multiplication release as well as the genetic differentiation and evolution relation between different groups. A scientific basis was provided for protection, management and sustainable utilization of E. sinensis in Yangtze River System.

2. Material and Methods

2.1. Materials
E. sinensis samples were collected from Huangmei (HM) (15), Poyang Lake (PY) (14), Pengze (PP) (19), Anqing (AH) (13), Zhenjiang (17 adult crabs (ZC) and 13 young crabs (ZY)), and San Jiang Harbor (11 young crabs (CY)) from Aug. to Nov. 2015. There were 102 samples in total. The sampling site was from the upstream of Yangtze Estuary to Hubei Huangmei, stretching across five provinces and cities including Hubei, Jiangxi, Anhui, Jiangsu and Shanghai, as shown in Fig. 1. Catching in natural water areas or catching with cage net was adopted as the sampling mode. One leg was taken from every E. sinensis, and put into the Eppendorf tube in 95% ethyl alcohol. The samples were sent to Jiangxi Fisheries Science Research Institute and stored at -80°C until used for DNA extraction.

2.2. Methods

2.2.1. DNA extraction. The cheliped muscle tissues of E. sinensis were cut into pieces of 100 mg, and put into the Eppendorf tube of 1.5 mL. 400 L of SET buffer solution was added, and then 40 μL of 10% SDS and 10 μL (20 mg/mL) of protease K were added successively. The mixture stayed overnight for digestion in the water bath of 55°C, followed by phenol-chloroform extraction and 100% ethanol precipitation [2]. Extracted genomic DNA was checked using 1.0% agarose gel electrophoresis, then diluted to the appropriate concentration (about 100 ng per ml) for PCR amplification.

2.2.2. PCR amplification and sequencing. Primer YL-F and YL-R were used to amplify the hyper variable region sequence in control region, and the sequences of primer are as follows: YL-F: 5'-ACGTAACTGAAATGCTGTTC-3'; YL-R: 5'-ACCCGTTTCCCCCTTAGAGGA-3'. The primers were synthesized by Sangon Inc. (Shanghai, China). The amplification reaction of the control region was in 50 μL volumes. Amplification reaction mix was composed of 100 ng DNA template, 2 μL of 2.5 mM dNTPs, 2 μL of each primer (10 mM), 5.0 μL of 10× reaction buffer, 0.25 μL of Taq DNA-polymerase (5 units/μL; TaRaKa), and sterile water up to a final volume of 50 μL. The PCR profile consisted of an initial denaturation at 94°C for 4 min, followed by 35 cycles of denaturation (94°C for 30 s), annealing (54°C for 45 s), and extension (72°C for 1 min), then a final extension step of 72°C for 10 min. The PCR product amplified from D-loop mtDNA was detected by electrophoresis on a 1.0% agarose gel, purified with the PCR-MTM Clean Up System (Viogene), and were sequenced.
using the forward and reverse primers in conjunction with the BigDyeTM Terminator Kit (ABI PRISM BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit, Applied Biosystems). The cycle sequencing products were purified by ethanol precipitation, separated using a POP-7TM polymer (Applied Bio systems) and detected using an ABI 3730 DNA Analyzer.

2.3. Data analysis
The DNA sequences were aligned through multiple sequence alignment with the Clustal X 1.83 program [3]. All E. sinensis population genetic analyses were performed using Dna SP 5.0 [4]. The extent of intra-population variation was estimated by calculating haplotype diversity (Hd) and nucleotide diversity (π). Gene flow (Nm) and Fst were also calculated using DnaSP 5.0. Phylogenetic relationships among haplotypes were estimated using Neighbor–Joining (NJ) based on a matrix of Kimura 2 parameter (K-2-P) distances method and calculated with parsimony analysis in MEGA4.0 [5], and the confidence level in the generated tree was obtained using 1000 bootstraps.

The population structure of 7 populations, genetic distance, migration rate among them and hierarchical analysis of molecular variance (AMOVA) were performed by calculating the Fst between populations and testing their significance whit 1000 permutations. The mismatch distributions were also examined using the Arlequin version 3.1.1 program [6]. Tajima’s D [7,8] and Fu's Fs [9] neutrality test were conducted using DnaSP and Arlequin3.11 software. A median-joining network of all haplotypes was established using Network4.6.1.0 software [10].

3. Results

3.1. Mitochondrial D-loop sequence characters of E. sinensis
The segment of the control region was amplified and sequenced from 102 individuals of E. sinensis. The complete sequences of 515 sites were aligned, and average nucleotide frequencies of adenine (A), thymine (T), guanine (G), and cytosine(C) were 42.3%, 36.2%, 10% and 11.4%, respectively (Tab.1). A+T base content was 78.5%, which was greater than the content of C+G (21.5%). The results showed that D-loop sequences of E. sinensis were strong base composition bias, the content of G was obviously lower than contents of the other 3 bases.

| Population | Base |
|------------|------|
| A          | 42.4 |
| G          | 10   |
| C          | 11.3 |
| T          | 36.4 |
| HM         |      |
| PP         |      |
| AQ         |      |
| ZY         |      |
| CY         |      |

3.2. Genetic diversity and phylogeny of E. sinensis
A total of 102 mt DNA control region sequences was analyzed and aligned, 65 sites were polymorphic (S) (Fig.2), containing 31 singleton variable sites and 34 parsimony-informative sites. 60 haplotypes were identified among 102 individuals (occupying 58.8% of the total sample size), including 53 unique haplotypes and 7 shared haplotypes. Among the shared haplotypes, 5 haplotypes were shared by more than two populations. As the most extensively distributed haplotype, Hap_16 was shared by 28 individuals of 7 population s (taking up 27.4% of the total sample size). PP population had the highest number of haplotypes and unique haplotypes, 18 and 14 respectively. ZY and CY populations presented the lowest number of haplotypes, which was 8; they had 5 and 6 unique haplotypes.
respectively (Tab.2). Haplotype frequency of PP was the highest, and ZC was the lowest. Unique haplotype frequency of HM populations was the highest, while AQ was the lowest (Tab.2).

Genetic diversity with seven populations can be reflected through haplotype diversity, nucleotide diversity and average nucleotide differences (Tab.2). Haplotype diversity (Hd) of seven populations was 0.919±0.023, nucleotide diversity (π) was 0.01007±0.00081, average nucleotide differences (K) was 5.106. The lowest values of both diversities were found in the ZC population, and the highest values of both diversities were found in the PP population, indicating that PP population was highly polymorphic. According to haplotype diversity, nucleotide diversity and average nucleotide differences, the genetic diversity order of 7 populations from high to low was PP>HM>PY>AQ>CY>ZY>ZC.

Figure 2. Variable is of D-Loop gene segment sequences in the Eriocheir sinensis.
Table 2. The sampling sites and genetic diversity in of the seven populations of *Eriocheir sinensis*.

| Population | N  | Nh  | U   | F(%) | UF(%) | hd         | π             | K   |
|------------|----|-----|-----|------|-------|------------|----------------|-----|
| HM         | 15 | 12  | 11  | 80   | 91.7  | 0.943±0.054 | 0.01179±0.00302 | 6.0 |
| PY         | 14 | 9   | 6   | 64.3 | 66.7  | 0.879±0.079 | 0.01172±0.00327 | 5.967 |
| PP         | 19 | 18  | 14  | 94.7 | 77.8  | 0.994±0.019 | 0.01387±0.00319 | 7.047 |
| AQ         | 13 | 10  | 6   | 76.9 | 60.0  | 0.962±0.041 | 0.01063±0.00297 | 5.410 |
| ZY         | 13 | 8   | 5   | 61.5 | 62.5  | 0.859±0.089 | 0.00840±0.00261 | 4.282 |
| ZC         | 17 | 10  | 7   | 58.8 | 70.0  | 0.794±0.103 | 0.00601±0.00246 | 3.074 |
| CY         | 11 | 8   | 6   | 72.7 | 75.0  | 0.927±0.066 | 0.00986±0.00285 | 5.018 |
| Total      | 102| 60  |     |      |       | 0.919±0.023 | 0.01007±0.00081 | 5.106 |

Note: N-Number of samples; Nh-Number of haplotypes; U-Unique of haplotypes; F-Frequency of haplotypes; UF-Frequency of Unique haplotypes; hd-Haplootype diversity; π-Nucleotide diversity; K-number of nucleotide differences.

3.3. Genetic distance and population genetic structure

Genetic distances within the seven populations were between 0.59% and 1.31%, and among populations were from 0.67% to 1.22%. The largest genetic distance was 1.22% between PP and CY populations, and the smallest was 0.67% between ZC and ZY populations (Tab. 3). The fixation index (Fst) among the seven populations varies from -0.0098 to 0.0983 (Tab. 3). Gene flow (Nm) existed among various geographical groups, and ranged between 4.013 and infinity (Tab. 7).

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

| Population | CY   | ZY   | ZC | AQ  | PP  | PY  | HM   |
|------------|------|------|----|-----|-----|-----|------|
| CY         | 0.0097 | 0.0408* | -0.0041 | 0.0393 | 0.0708 | 0.0343 | 0.0983* |
| ZY         | 0.0084 | 0.0074 | -0.0026 | 0.0267 | 0.0518 | -0.0098 | 0.0710* |
| ZC         | 0.0079 | 0.0067 | 0.0059 | 0.0372 | 0.1108** | 0.0290 | 0.0738* |
| AQ         | 0.0103 | 0.0092 | 0.0084 | 0.0103 | -0.0222 | -0.0261 | -0.005 |
| PP         | 0.0122 | 0.0109 | 0.0107 | 0.0114 | 0.0131 | -0.0049 | 0.0185 |
| PY         | 0.0109 | 0.0095 | 0.0090 | 0.0107 | 0.0122 | 0.0113 | 0.0138 |
| HM         | 0.0113 | 0.0098 | 0.0089 | 0.0104 | 0.0121 | 0.0113 | 0.0113 |

Note: Average pairwise difference between populations is below diagonal. Fixation index is above diagonal, Average pairwise difference within populations is along diagonal. * means significant difference at 0.05 level (P<0.05); ** means significant difference at 0.01 level (P<0.01).

3.4. AMOVA analysis and phylogenetic analysis

An analysis of molecular variance (AMOVA) was conducted to describe the variance components of the F index, and showed that most variance of *E. sinensis* occurred within the populations in geographic and phylogenetic clade structures (P<0.001), suggesting that this is main source of total variance (Tab. 4). These results suggested that the seven populations of *E. sinensis* in Yangtze River System have not developed significant genetic structure.

The phylogenetic relationship among haplotypes was constructed based on Kimura-2-Parameter distances. In the tree, most of the haplotypes clustered weakly (<50% bootstrap support) (Fig. 3a). It meant that all haplotypes could not be unbiasedly assigned into a clade, which may be explained by the less variation among these haplotypes (Tab. 4). Network analysis results among various groups of *E. sinensis* in Yangtze River System are presented in Fig. 3b. 60 haplotypes showed an asteroid network graph; there was no obvious branch among various populations of *E. sinensis*; branch corresponding to the sampling site was not detected and there was no outstanding pedigree structure.
The haplotype H_12 was located in the network center. This haplotype might be the original one and other haplotypes derived from it.

**Table 4.** Analysis of molecular variance (AMOVA) of seven populations of *Eriocheir sinensis*

| Source of variation | d.f. | Sum of squares | Variance components | Percentage of variation/% | global divergence |
|---------------------|------|----------------|---------------------|--------------------------|------------------|
| Among populations   | 6    | 22.244         | 0.08463Va**         | 3.30                     |                  |
| Within population   | 95   | 235.619        | 2.48020Vb           | 96.70                    | 0.0330**         |
| Total               | 101  | 257.863        | 2.56483             |                          |                  |

3.5. **Tests of neutrality and estimates of population expansion**

Tajima's D [6, 7] and Fu's Fs [8] tests were carried out to examine deviations from neutrality, which would be expected with population expansion. The neutrality tests and mismatch analysis were performed using Arlequin3.11 software (Tab.5 and Tab.6). The results showed that *E. sinensis* in Yangtze River System refused the population expansion model (Tajima's D = -1.118, P > 0.05; Fu's Fs = -3.808, P > 0.05). The neutrality tests for various geographical populations showed that ZC populations...
(Tajima's $D=-1.068$, $P<0.05$; Fu's $F_s=-3.28$; $P<0.05$) is accepted as the population expansion model, while the other 6 populations are refused the population expansion model. Among various populations, $\tau$ value of AQ population was the smallest, and ZY was the highest (Tab.6). According to mismatch analysis on D-loop sequence of 7 E. sinensis populations, expected values of base bifurcation point distribution presented typical multimodal distribution, showing that E. sinensis population in Yangtze River System did not experience population expansion in history (Fig.4 and Fig.5). Raggedness statistics ($R_g$) of the 7 populations ranged from 0.37 to 0.96, not reaching the significance level ($P>0.05$). This further supported the proposition that E. sinensis in Yangtze River System did not experience population expansion. SSD between observed distribution and expected distribution among the 7 populations did not reach the significance level, indicating that E. sinensis population in Yangtze River System did not experience the expansion period (Tab.6).

Figure 4. Mismatch distributions of mitochondrial D-loop of seven populations of E. sinensis.

Figure 5. Phylogentic analysis of seven populations of E. sinensis in Yangtze River
Table 5. Neutrality tests for *E. sinensis*

|          | AQ    | ZC    | ZY    | CY    | PP    | PY    | HM    | Total samples |
|----------|-------|-------|-------|-------|-------|-------|-------|---------------|
| Tajima's $D$ | -1.040 | -1.608* | -0.964 | -0.747 | -1.052 | -1.450 | -0.970 | -1.118        |
| p-value  | 0.151 | 0.039 | 0.183 | 0.245 | 0.139 | 0.062 | 0.162 | 0.140         |
| Fu's $FS$ | -2.916 | -3.528* | -1.558 | -1.552 | -11.828** | -1.067 | -4.206* | -3.808        |
| p-value  | 0.061 | 0.026 | 0.157 | 0.170 | 0.000 | 0.026 | 0.015 | 0.101         |

Notes: * significant (P<0.05) ** extremely significant (P<0.01)

Table 6. Mismatch distribution for *E. sinensis*

| Mismatch distribution | AQ | ZC | ZY | CY | PP | PY | HM |
|-----------------------|----|----|----|----|----|----|----|
| $\tau$                | 1.437 | 4.046 | 7.419 | 2.851 | 5.053 | 7.176 | 5.101 |
| 00                    | 4.797 | 2.055 | 3.228 | 2.612 | 3.164 | 5.456 | 1.363 |
| 01                    | 99999 | 4.533 | 4.225 | 13.159 | 24.001 | 8.267 | 24.104 |
| SSD                   | 0.008 | 0.042 | 0.054 | 0.027 | 0.027 | 0.011 | 0.020 | 0.017 |

| Goodness-of-fit tests | Model(SSD)p-value | Raggedness indexp-value | Raggedness p-value |
|-----------------------|-------------------|------------------------|-------------------|
|                       | 0.80              | 0.0176                 | 0.96              |
|                       | 0.26              | 0.114                  | 0.58              |
|                       | 0.18              | 0.104                  | 0.37              |
|                       | 0.42              | 0.0575                 | 0.59              |
|                       | 0.13              | 0.0127                 | 0.84              |
|                       | 0.61              | 0.0364                 | 0.95              |
|                       | 0.56              | 0.0515                 | 0.51              |

Table 7. Population gene flow for comparisons between *E. sinensis* populations

|          | AQ     | ZC     | ZY     | CY     | PP     | PY     | HM     |
|----------|--------|--------|--------|--------|--------|--------|--------|
| AQ       | 12.934 |        |        |        |        |        |        |
| ZC       | 18.232 | Inf    |        |        |        |        |        |
| ZY       |        | 11.747 | Inf    |        |        |        |        |
| CY       | 12.220 |        | 4.013  | 9.151  | 6.564  |        |        |
| PP       |        | Inf    | 16.749 | Inf    | 14.062 | Inf    |        |
| PY       | Inf    |        |        |        |        |        |        |
| HM       | Inf    |        |        |        |        |        |        |

4. Conclusion
In conclusion, genetic diversity of HM, PY, PP, AQ and CY populations in Yangtze River System was at a high level. Genetic diversity of ZC population and ZY population were slightly lower than that of the other five populations. Hence effective measures should be taken to protect the wild resources, so as to realize reasonable development and utilization of recourses. In addition, genetic differentiation of a certain degree existed between *E. sinensis* populations in different river sections. This reminds us that monitoring and evaluation on genetic diversity of wild resources in Yangtze River System should be strengthened, and the resource management force must be enhanced, so as to prevent germplasm degeneration of *E. sinensis in* Yangtze River System caused by artificial introduction and multiplication release.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (No.31972795).

References
[1] Sui L, Zhang F, Wang X, et al. Genetic diversity and population structure of the Chinese mitten crab *Eriocheir sinensis* in its native range [J]. Marine Biology, 2009, 156 (8): 1573 - 1583.
[2] Sambrook, J., E. F. Fritsch, T. Maniatis, 1989. Molecular Cloning: A laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[3] Thompson J D, Gibson T J, Plewniak F, et al. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J] . Nucleic
Acids Research, 1997, 25 (24): 4876 - 4882.

[4] Rozas J, Sanchez-DelBarrio J C, Messeguer X, et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods [J]. Bioinformatics, 2003, 19 (18): 2496 - 2497.

[5] Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596 - 1599.

[6] Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 2005, 1: 47 - 50.

[7] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism [J]. Genetics, 1989, 123 (3): 585 – 595.

[8] Tajima F. The effect of change in population size on DNA polymorphism [J]. Genetics, 1989, 123 (3): 597 - 601.

[9] Fu Y X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection [J]. Genetics, 1997, 147 (2): 915 - 925.

[10] Morrison D A. Networks in phylogenetic analysis: new tools for population biology [J]. International Journal for Parasitology, 2005, 35 (5): 567 - 582.