Microsatellites assessment of Chinese sturgeon (Acipenser sinensis Gray) genetic variability

By N. Zhao¹,², W. Ai¹, Z. Shao¹, B. Zhu¹, S. Brosse³ and J. Chang¹

¹Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China; ²Marine and Estuarine Fisheries (Ministry of Agriculture), Shanghai, China; ³LADYBIO-CNRS-University Paul Sabatier, Toulouse Cedex, France

Summary

Four microsatellites were used to examine the genetic variability of the spawning stocks of Chinese sturgeon, Acipenser sinensis, from the Yangtze River sampled over a 3-year period (1999–2001). Within 60 individuals, a total of 28 alleles were detected over four polymorphic microsatellite loci. The number of alleles per locus ranged from 4 to 15, with an average allele number of 7. The number of genotypes per locus ranged from 6 to 41. The genetic diversity of four microsatellite loci varied from 0.34 to 0.67, with an average value of 0.54. For the four microsatellite loci, the deviation from the Hardy–Weinberg equilibrium was mainly due to null alleles. The mean number of alleles per locus and the mean heterozygosity were lower than the average values known for anadromous fishes. Fish were clustered according to their microsatellite characteristics using an unsupervised ‘Artificial Neural Networks’ method entitled ‘Self-organizing Map’. The results revealed no significant genetic differentiation considering genetic distance among samples collected during different years. Lack of heterogeneity among different annual groups of spawning stocks was explained by the complex age structure (from 8 to 27 years for males and 12 to 35 years for females) of Chinese sturgeon, leading to formulate an hypothesis about the maintenance of genetic diversity and stability in long-lived animals.

Introduction

During the last century, the distribution and population sizes for many species of plants and animals have been heavily affected by humans. To save populations from extinction, artificial propagation and/or the transfer of individuals have become a common practice. Little is known about how this practice has affected the distribution of genetic variation in a population because, for most species, no genetic data existed prior to the time the population structure was affected by human activities. The genetic data retrieved from the researched samples in this paper offer an opportunity for studies of the genetic population structure of Chinese sturgeon before it was affected by human activities and therefore act as baseline studies in future studies of effects of human activities.

Chinese sturgeon is an anadromous fish, migrating in the upper reaches of large rivers to spawn. From a historical point of view, Chinese sturgeon only colonizes two river basins: the Yangtze and Pearl rivers; most of the individuals (if not all) now reproduce in the Yangtze River. However, this last population has been under threat since 1981 by the establishment of a large dam (Gezhouba Dam), blocking upstream reproductive migration. This led to a drastic decrease in the suitable spawning areas, as only one spawning ground of the 16 known before the dam construction is still available to the fish (Wei et al., 1997). This last operating spawning ground is located just below the dam and covers an area of approximately 5 km² (Deng et al., 1991; Kynard et al., 1995). This surface is not large enough to permit all mature individuals to spawn (Kynard et al., 1995; Chang, 1999). Aiming to compensate recruitment reduction because of the lack of spawning grounds, artificially hatched larvae have been released since 1984 (Fu et al., 1985; Xiao et al., 1999). However, recent studies suggest that artificial propagation is inefficient and that the natural reproduction still plays a major role in sturgeon population survival (Zhu, 2003). Indeed, artificially propagated individuals account for only 5–10% of the juvenile recruitment (Zhu et al., 2002). Artificial propagation should therefore be increased to ensure sufficient recruitment to sustain the sturgeon population; however, intensive artificial propagation may lead to genetic homogenization, and it is therefore crucial to understand the genetic composition of the natural population.

To reach this aim, microsatellite DNA markers can be used because they possess high levels of polymorphism (O'Connell and Wright, 1997); moreover, this method is able to detect fine genetic variations (Wright and Bentzen, 1994). However, most of the mathematical developments used to quantify population differentiation assume disomic inheritance and Hardy–Weinberg equilibrium (HWE) genotypic frequencies (Rodzen and May, 2002). There is therefore a need to reveal the inheritance for some loci and acquire disomic loci (Pyatkowit et al., 2001; Rodzen and May, 2002). Ludwig et al. (2001) and Zhu et al. (2002) suggested that Chinese sturgeon with approximately 250 chromosomes are functionally tetraploid. Therefore, owing to lack of adequate disomic loci, through the direct use of the distribution of the alleles and genotypes in tetraploid inheritance it is possible to investigate population structure and genetic diversity. In this study, we first described the inheritance of some microsatellite loci and then used these microsatellites to quantify genetic variations in the last natural reproductive population of Chinese sturgeon.

Materials and methods

Sampling

Fish sampling was achieved on the spawning ground during the spawning period (October and November). Sampling was repeated once a year from 1999 to 2001. A total of 60 mature individuals was caught (21 in 1999, 26 in 2000 and 13 in 2001).
Fin clips (approximately 0.5 cm³) were obtained from each individual and stored individually in 100% ethanol until genetic analysis. The notation used to identify each sample was the following: ‘99par-m01’ where 99 represents the year, par is the parent, m is male and 01 the serial number.

Genetical analysis

Genomic DNA was extracted according to a modified Cetyltrimethyl-ammonium-bromide (CTAB) protocol (Saghai-Maroo et al., 1984). DNA quality controls were performed using DU® spectrophotometer (Beckman Instruments, Inc., Fullerton, CA, USA) and agarose gel electrophoresis.

Microsatellite primers examined were Afu-19, Afu-39, Afu-54 (May et al., 1997) and As-100 (Shao et al., 2002). As-100 is redesigned from Spl-106, which was developed for shovelnose sturgeon (Scaphirhynchus platorynchus) (McQuown et al. 2000). Two microsatellites (Afu-19, Afu-39) are trinucleotide repeats and the others (Afu-54, As-100) exhibit a compound motif (Table 1).

Amplification reactions were performed in a 20-µl total volume containing 1 U of Taq DNA polymerase (Biorstar), 0.6 mM of each primer, 175 µM dNTP, 10–15 ng of template DNA, 1.5 mM MgCl₂ and 2-µl of 10X reaction buffer. All reactions were performed in a T1 Thermocycler (Biometra). Initial denaturation was achieved at 94°C for 4 min followed by 35 cycles of denaturation at 94°C for 55 s, annealing at 55–60°C for 55–65 s, extension at 72°C for 1 min. The final step was extended to 6 min at 72°C. The annealing parameter was optimized for each microsatellite loci (see also Table 1). The polymerase chain reaction (PCR) products were electrophoresed in 8% polyacrylamide gels and visualized with silver-staining protocol (Bassam et al., 1991). The polymerase chain reaction (PCR) products were electrophoresed in 8% polyacrylamide gels and visualized with silver-staining protocol (Bassam et al., 1991). The polymerase chain reaction (PCR) products were electrophoresed in 8% polyacrylamide gels and visualized with silver-staining protocol (Bassam et al., 1991).

For a tetrasomic locus, the alleles frequency fills the equation: \( p + q + r + s + \ldots = 1 \). Where \( p, q, r, s \ldots \) are the different alleles in this locus (Li, 1955). Therefore, gene diversity at a tetrasomic locus was estimated using the following formula, which was adapted from the original formula for small sample size with diploid data (Nei, 1987):

\[
\text{Gene diversity} = \frac{n}{n-1} \left(1 - \sum_{i=1}^{q} P_{ij}^4\right)
\]

where \( n \) is the number of gene copies in the sample, \( \sum \) stands for the summation over all the alleles and \( P_{ij} \) is the frequency of the \( i \)th allele and the number of alleles at the \( i \)th locus, respectively. We considered that gene diversity equals the expected heterozygosity (Nei, 1987). The observed heterozygosity was obtained using observed heterozygous divided by the total sample. The deviation from HWE was tested with a chi-square goodness-of-fit test. Due to the small number of samples and to the multiple alleles at the same locus, some \( Ei \) categories were low. We therefore pooled categories to agree with chi-square test requirements. Loci with three or more alleles were considered, using the most common allele as \( p \)-value and grouping the other alleles as \( q = (1 - p) \).

Based on genotype data from 60 individuals, fish were patterned using a Self-organizing Map (SOM) (Kohonen, 1982). The SOM is an ordination method that can achieve the same task as principal component analysis, but which is not biased by rare individuals or non-linearity (Brosse et al., 2001). This artificial intelligence procedure was applied to the 60 fish samples in order to build an objective image of the genetic diversity patterns measured at the three sampling occasions. The SOM is an unsupervised learning algorithm that approximates the probability density function of the input variables through a learning process. Input variables were alleles. For each fish, each allele was coded by the number of dosage present (0–4): 0 if the allele is absent, 1 if present in one dosage and 2 if present in two dosages. The SOM performs a non-linear projection of the multivariate data into a lower dimension (Kohonen, 2001). Formally, the SOM consists of two types of units (or neurones): input and output layers, including connection intensities called weights between the two layers. The array of input units operates as a flow-through layer for the input vectors, whereas the output layer consists of a two-dimensional network of neurones arranged on a hexagonal lattice. A hexagonal lattice is preferred, because it does not favour horizontal or vertical directions (Kohonen, 2001). Each neurone is connected to its nearest neighbours on the grid, and stores a set of connection intensities. During the learning process, the network computes the distance between weight vector \( W \) and input vector \( X \) (alleles of the 60 fish). In the output layer, the best matching unit (BMU) is selected by the algorithm; it corresponds to the neurone (or cell in the grid) having the minimum distance (c). Whereas the weight vectors of other neurones are not changed, for the BMU and its neighbourhood neurones the new weight vectors are updated with the following equation:

\[
w_i(t+1) = w_i(t) + h_{ij}(t)[x(t) - w_i(t)]
\]

where \( t \) is the iteration time, and \( h_{ij}(t) \) is a neighbourhood function and a smoothing kernel for location vectors of neurone \( c \) and \( i \) defined over the lattice of the output layer.

### Table 1

Characterization of the four microsatellite loci

<table>
<thead>
<tr>
<th>Loci</th>
<th>Repeat motif in original clone</th>
<th>( T_a ) (°C)</th>
<th>Time (s)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afu-19</td>
<td>(TTG)(_n)</td>
<td>58</td>
<td>65</td>
<td>May et al. (1997)</td>
</tr>
<tr>
<td>Afu-39</td>
<td>(GGT)(_n)</td>
<td>58</td>
<td>65</td>
<td>May et al. (1997)</td>
</tr>
<tr>
<td>Afu-54</td>
<td>(GATA)(_n)(GACA)(_n)</td>
<td>58</td>
<td>65</td>
<td>May et al. (1997)</td>
</tr>
<tr>
<td>As-100</td>
<td>(AAC)(_n)(TTAA)(_n)</td>
<td>60</td>
<td>60</td>
<td>McQuown et al. (2000), Shao et al. (2002)</td>
</tr>
</tbody>
</table>

\( T_a \) (°C), annealing temperature; time (s), annealing time.
This learning process results in training the network to pattern the input vectors and preserves the connection intensities in the weight vectors.

After the SOM has been trained, it is important to know whether or not it has been properly trained. To do this, we computed a topographic error as indicator of topology preservation of the map. This value represents the proportion of all data vectors for which first and second BMUs are not adjacent for the measurement of topology preservation (Kiviluoto, 1996). Thus, this error value is used as an indicator of the accuracy of the mapping in the preserving topology (Kohonen, 2001). To distinguish boundaries on the resulting SOM map, units were subdivided into different groups according to the similarity of the weight vectors produced by the output neurons. In this study, a hierarchical cluster analysis (Ward distance) was used to find the cluster boundaries on the SOM map. Based on the distribution pattern of the fish and associated boundaries, we identified patterns of genetic diversity of Chinese sturgeon and used the ANOVA to compare the differences of genetic variations between clusters defined on the SOM map.

Results

Band pattern variation

The three (Afu-19, Afu-54 and As-100) microsatellite systems yielded more than two bands for some fish, but never more than four bands for any of the samples. This indicated the tetraploid derivative nature of the Acipenser sinensis genome. For the Afu39 system, one or two bands were produced in all individuals. The banding pattern with two bands exhibited asymmetry in band intensity, also reflecting the polyploid nature. The multiple band patterns included 4, 3 : 1, 2 : 2, 2 : 1 : 1 or 1 : 1 : 1 : 1 dose proportion. No stutter band was observed because of the selection for primers and reaction condition optimization allowed by previous works on this species (Zhu et al., 1999; Shao et al., 2002). Although the strength of amplification depends not only on the number of DNA copies in the genome but also on the fragment length (Ludwig et al., 2001), the potential biases were lowered by no more than four bands and a relevant amplification.

Genetic variations among the different year-class samples

For the total of 60 individuals, the four microsatellite loci examined were polymorphic (P < 99%) and a total of 28 different alleles were detected. Variations were maximal in As-100 with 15 alleles. The Afu-19, -39, and -54 had four, four, and five alleles, respectively. Allele size ranged from 90 to 217 bp (Table 2). The number of genotypes per loci ranged from 6 in Afu-39 to 41 genotypes in As-100. Low-frequency (<1%) alleles were observed in all four loci. Table 2 shows allelic variations at the four microsatellite loci in the 3 different year-class samples. The frequency of the most common allele was similar for the 3 years (Fig. 1). For example, the most common allele at Afu-19 is allele 3, which has 80.0%, 77.9% and 80.8% frequency in each sampling year, respectively. Similar trends were found for genotypes; the most common genotypes were also the same for the three sampling periods: 3333 at locus Afu-19 and 2222 at locus Afu-54 (Fig. 2).

![Fig. 1. Allele frequencies at the four microsatellite loci for the 3 year-class samples.](image)

<table>
<thead>
<tr>
<th>Annual groups</th>
<th>Afu-19</th>
<th>Afu-39</th>
<th>Afu-54</th>
<th>As-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>A</td>
<td>L</td>
<td>M</td>
<td>F (%)</td>
</tr>
<tr>
<td>2000</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>77.9</td>
</tr>
<tr>
<td>2001</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>80.8</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>79.2</td>
</tr>
</tbody>
</table>

A, the number of alleles per loci; L, the number of low-frequency alleles; M, the rank of the most common allele; F%, the frequency of the most common allele; G, the number of the detected genotypes/the sample size.

Data given in the ‘total’ line was based on the 3 years samples without temporal distinction.

Table 2

<table>
<thead>
<tr>
<th>Annual groups</th>
<th>Afu-19</th>
<th>Afu-39</th>
<th>Afu-54</th>
<th>As-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>A</td>
<td>L</td>
<td>M</td>
<td>F (%)</td>
</tr>
<tr>
<td>2000</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>77.9</td>
</tr>
<tr>
<td>2001</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>80.8</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>79.2</td>
</tr>
</tbody>
</table>
The observed heterozygosities ($H_o$) and expected heterozygosities ($H_e$) were calculated on the different year samples as well as the pooled 60 samples without temporal distinction (Table 3). The mean $H_o$ based on the different year-class samples were approximated. The chi-square goodness-of-fit test detected a significant difference ($P < 0.01$) between observed heterozygosity and HWE at two microsatellite loci (Table 3). For both cases, deviation from HWE was due to a heterozygosity deficit. Furthermore, the frequency of certain homozygous genotypes strongly deviated from expectation. For example, the expected number of genotypes of 4444 at locus As-100 was 0.14, whereas observed value was 5.00.

**Fish genetic diversity patterns**

The grid size of the SOM was 35 (5 × 7) units. Such architecture provided the most relevant results, as the topographical error remained close to nil. On the SOM map, Ward clustering algorithm permitted to distinguish three clusters of the tributaries according to the distribution characteristics of alleles (Fig. 3a,b). The results showed that Chinese sturgeon individual clustering was not influenced by the date of sampling, but by the genetic characteristics of the fish. The number of the detected alleles did not differ significantly among the three clusters (ANOVA, $P > 0.05$). In addition, the ranks of the most common allele at the three loci (Afu-19, -39 and -54) were similar for the three clusters.

**Discussion**

Although multiloci markers such as random amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) could reveal more genetic variation than single-locus markers such as microsatellite DNA, this technique presents the advantage of permitting analyses and to precisely define the genotype at specific locus and to quantify allele frequency. Zhu et al. (2002) and Rodzen and May (2002) successfully used this method to study microsatellite inheritance in sturgeons. Most heterozygous individuals were typically represented by multiple bands with different intensities, supporting the tetraploid nature of sturgeon genome (Ludwig et al., 2001; Zhu et al., 2002). This provides the feasibility to analyse genetic data of Chinese sturgeon.

Although a larger amount of samples could increase the relevance of the results, it was not possible to collect more samples, as the Chinese sturgeon is a threatened species protected by Chinese law. However, according to O’Connell and Wright (1997), a sample size of 50 individuals with loci

---

**Table 3**

Heterozygosities at the four microsatellite loci for three annual groups of *Acipenser sinensis*

<table>
<thead>
<tr>
<th>Annual groups</th>
<th>Afu-19</th>
<th>Afu-39</th>
<th>Afu-54</th>
<th>As-100</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>0.45/0.60</td>
<td>0.72/0.82</td>
<td>0.45/0.57</td>
<td>0.59/1.00</td>
<td>0.55/0.75</td>
</tr>
<tr>
<td>2000</td>
<td>0.54/0.64</td>
<td>0.62/0.67</td>
<td>0.19/0.34</td>
<td>0.73/0.73</td>
<td>0.52/0.59</td>
</tr>
<tr>
<td>2001</td>
<td>0.54/0.58</td>
<td>0.54/0.54</td>
<td>0.46/0.54</td>
<td>0.75/0.95</td>
<td>0.572/0.65</td>
</tr>
<tr>
<td>Total</td>
<td>0.51/0.61</td>
<td>0.63/0.70</td>
<td>0.34/0.47</td>
<td>0.67/1.00</td>
<td>0.54/0.69</td>
</tr>
</tbody>
</table>

The data were recorded as $H_o/H_e$. $H_o$, observed heterozygosity; $H_e$, expected heterozygosity.
having between 5 and 10 alleles is sufficient to provide reliable results. Although sample size could artificially reduce the total number of alleles identified per locus as well as the range of allele size, it should not affect the most common allele and its frequency (Xu et al., 2001). Therefore, despite the sample size differences between years, the estimation of the most common allele and the allelic frequencies provide a reliable measure of genetic changes, as they are seldom affected by the sample size. In this way, the method we used allowed reliable identification of genetic differences. These two parameters did not vary much among the three sampling years in this work. Such genetic stability can be explained by the sturgeon’s life history. The Chinese sturgeon is a long-lived animal that attains sexual maturity between 8 and 17 years of age for males, and between 13 and 26 years for females (Deng et al., 1985; Chang, 1999). This discrepancy between age of first sexual maturity in males and females reduces the risk of inbreeding. In the same way for a given sex, fish from the same cohort acquire sexual maturity at different ages, which also contributes to genetic exchanges among the stocks. These two processes sustain genetic diversity of the Chinese sturgeon, although the spawning stock is now limited to <2500 individuals (Chang, 1999).

As expected, levels of genetic variability revealed by microsatellite loci were much higher than the results obtained by Zhang (1998) on allozyme and RAPD loci. For example, the polymorphic loci frequency of allozyme and RAPD is 3.8% and 11.1%, respectively (Zhang, 1998). In the present study, all microsatellite loci tested were polymorphic. Afu-19 loci polymorphism obtained here differed from the result given by Zhu et al. (2002) on the same sturgeon population. This was

Fig. 3. Classification of samples based on the Self-organizing Map (SOM) map. (a) The U-matrix algorithm was applied to set boundaries on the SOM map. Roman numerals indicate the three clusters, and the codes in each unit represent the fish samples (see Material and Methods for details on the codes). (b) Hierarchical classification of the SOM map. The numbers 1–35 correspond to those in the units of (a).
probably caused by experimental differences such as the resolving power difference between gels. Moreover, although only four loci were considered in our study, these results are consistent with Huang et al. (2002) and confirm that compound microsatellites produce a large number of alleles. Microsatellite As-100 possessing trimer repeat sequences following by two back-to-back tetramer repeats had the higher allele diversity, whereas the loci Afu-19 and Afu-39 characterized by pure trimer repeat motifs had a lower allele diversity.

We found an overall deficiency of heterozygotes. Although not all deviation from HWE was a deficit of heterozygosity (Xu et al., 2001), the observed heterozygosity in these loci were lower than the expected values. In addition, the excess in certain genotypes was limited to homozygous genotypes. Many hypotheses could explain such results. First, it could be due to a methodological bias called ‘stutter-related scoring errors’, but stutter band was avoided by preliminary experiments. Secondly, such results could come from null alleles, leading the homozygos to replace heterozygos. Indeed, Jarne and Lagoda (1996) reported that null allele in fish might be quite common, and Pyatskowit et al. (2001) and Rodzen and May (2002) also testified to the influence of null alleles, errors could be due to a methodological bias called ‘stutter-related scoring errors’, but stutter band was avoided by preliminary experiments. Alternatively, the deficit of heterozygosity could be ascribed to population substructure. The causes of deviations from HWE need further studies.

According to Zhang et al. (1999), the reproductive population of A. sinensis has a low heterozygosity (H = 0.039) and a low percentage of allozyme polymorphic loci (P = 0.040). Both parameters produced lower values than those measured for teleost fishes (H = 0.0513 ± 0.034 and P = 0.152 ± 0.098; Nevo, 1978). Our results based on different techniques (i.e. microsatellite markers) confirmed and added weight to these previous statements, as the Chinese sturgeon was found to have low genetic variation. DeWoody and Avise (2000) reported that anadromous fish displayed high levels of genetic variation within populations [mean heterozygosity, H = 0.68, and mean number of alleles per microsatellite locus, A = 11.3, higher than those of Chinese sturgeon (H = 0.54 and A = 7.0, this paper)]. Considering the SOM result and the minor changes in allele number and frequencies we found among the three SOM clusters, there were no strong gradients in genetic diversity. In this study, sampling occurred 17–19 years after construction of the Gezhouba Dam. Thus, a large part of the fish sampled was born before the dam was constructed, allowing us to conclude that the lower levels of genetic variability in the Chinese sturgeon population already existed prior to the beginning of the dam when the population size of the Yangtze sturgeon was large, with annual commercial captures of 20–25 tonnes. Thus, the drastic population size reduction probably did not reduce genetic diversity. Management priorities set on the conservation of this species are therefore useful because the Yangtze sturgeon population has the potential to recover as a viable, sustainable population.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (NSFC 30490234), the Chinese Academy of Sciences (KSCX2-1-03), Executive Office of the State Council Three Gorges Project Construction Committee of China (SX 2003–008) and the Chinese-French advanced research program (PRA E-02/03).

References


Author’s address: Jianbo Chang, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China 430072. E-mail: jbchang@ihb.ac.cn