Biochemical Systematics and Ecology 43 (2012) 214–225

Contents lists available at SciVerse ScienceDirect

Biochemical Systematics and Ecology

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

Genetic structure of an endangered endemic fish (*Gobiocypris rarus*) in the upper Yangtze River

Yongfeng He [a], Jianwei Wang [a,*], Simon Blanchet [b,c], Sovan Lek [b]

[a] Institute of Hydrobiology, Chinese Academy of Sciences, 7# Donghu South Road, Wuhan, Hubei 430072, China
[b] Université de Toulouse, Lab. Evolution & Diversité Biologique, UMR 5174, CNRS – 118, route de Narbonne, 31062 Toulouse, Cedex 4, France
[c] Station d'Ecologie Expérimental du CNRS à Moulis, Lab. USR 2936, 09 100 Moulis, France

ARTICLE INFO

Article history:
Received 28 September 2011
Accepted 10 March 2012
Available online

Keywords:
Genetic differentiation
Gene flow
Gobiocypris rarus
Conservation implications

ABSTRACT

The goal of this study was to examine the genetic diversity and structure of wild populations of rare minnow (*Gobiocypris rarus*) in the upper Yangtze River. Individuals from nine locations were sampled and genotyped at eight microsatellite markers. Populations differed from each other on allelic frequency, although there were no significant differences on other genetic diversity indices. Our results further suggested low to moderate levels of population differentiation (F_{ST}: 0.0130–0.1537). Specifically, two distinct genetic clusters (*C*1 and *C*2) were found, which may be highly correlated with the evolution of water systems. In addition, a weak but significant isolation-by-distance pattern was best explained by riparian distance through man-made channels. It suggested that fish might use man-made channels for gene flow in history, although no significant recent migration event was found. Therefore, this study highlights the need to consider the genetic specificities of *G. rarus* for sustaining long-term survival of this species.

1. Introduction

Understanding the genetic structure of threatened species is a key point for its effective conservation. In order to determine the overall genetic structure of populations, it is necessary to know the level of both intra- and inter- population variation, and to understand the underlying processes driving this variation (Frankham et al., 2002). Usually, environmental barriers, historical processes, life histories (e.g. mating systems) and even geographical isolation may all, to some extent, shape the genetic structure of populations (Johnson, 2000; Balloux and Lugon-Moulin, 2002; Björklund et al., 2007; Blanchet et al., 2010).

Actually, these factors are usually related to evolutionary processes, i.e., gene flow, genetic drift, selection and mutation (DiBattista, 2008). Understanding gene flow and its effects is central to population genetics, since genetic structuring reflects the number of alleles exchanged between populations. Then the exchange of genes between populations determines the relative effects of selection and genetic drift (Balloux and Lugon-Moulin, 2002). The relationships between gene flow and population genetic structure have been the subject of many studies during several decades (Slatkin, 1985; Bohonak, 1999; Fraser et al., 2004; Crispo et al., 2006; Raeymaekers et al., 2008; Blanchet et al., 2010). To infer individual movements between populations, and predict their consequences on population genetic structures and evolutionary potential are one of the most important issues in the management of species (Crandall et al., 2000; Fraser and Bernatchez, 2001; Frankham et al., 2002).

Analytical approaches for genetic markers such as microsatellites available for inferring population structure have become highly developed in recent years (Beaumont and Rannala, 2004; Pearse and Crandall, 2004). Commonly, a conventional

* Corresponding author. Tel.: +86 27 68780033; fax: +86 27 68780065.
E-mail address: wangjw@ihb.ac.cn (J. Wang).

0305-1978/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bse.2012.03.010

approach such as Wright's *F*-statistics has been used to infer population structure (Wright, 1978). In recent years, individual-based Bayesian approaches such as the software STRUCTURE have been developed as a complementary approach due to their greater precision (Pritchard et al., 2000; Beaumont and Rannala, 2004; Pearse and Crandall, 2004; Evanno et al., 2005; Aspi et al., 2009). In addition, instead of direct methods such as capture–mark–recapture, indirect methods (e.g. Bayesian approaches and the isolation-by-distance framework) have been employed and applied extensively to estimate recent migration and gene flow from molecular data by overcoming the limitation of study areas and samplings (Wilson and Rannala, 2003; Piry et al., 2004; Blanchet et al., 2010).

Rare minnow, *Gobiocypris rarus* Ye *et* Fu, is an endemic cyprinid fish in China, which has been extensively used as an aquatic laboratory animal (Ye and Fu, 1983; Wang et al., 1994, 1998; Wang and Cao, 1997; Chen, 1998; Zhong et al., 2005). Meanwhile, rare minnow is an endangered fish species (Yue and Chen, 1998; Wang and Xie, 2004). It lives mainly in small water systems, such as paddyfields and ditches. Remnant populations have narrow distributions in the upper Yangtze River, which were only identified in the west region of Sichuan Province of China (Ding, 1994; Wang and Cao, 1997; Chen, 1998; Yue and Chen, 1998). They are suffering from anthropogenic disturbance including pollution, channelized habitat and disordered water diversion (Yue and Chen, 1998; Wang et al., 1998; Li et al., 2004; Wang and Xie, 2004; Xiong et al., 2009).

In this study, we used microsatellite markers to quantify genetic diversity and assess genetic structure of wild populations of *G. rarus*. Gene flow and recent migration events within these populations were then estimated in order to probe into the evolutionary process of genetic structure. It is expected to table a proposal about the conservation of *G. rarus*.

2. Material and methods

2.1. Sampling location and tissue collection

We investigated all the known distribution sites and its neighboring areas in April, 2008. Then we found fourteen wild populations in all, which mainly located at the edge of the west and northwest area of the Sichuan Basin. They belonged to four river sub-basins of the upper Yangtze River: the downstream of the Dadu River, the middle and downstream of the Qingyi River, the middle stream of the Minjiang River, and the upstream of the Tuojiang River. Among them, nine populations containing enough samples for genetic population studies were chosen in the present study (Fig. 1).

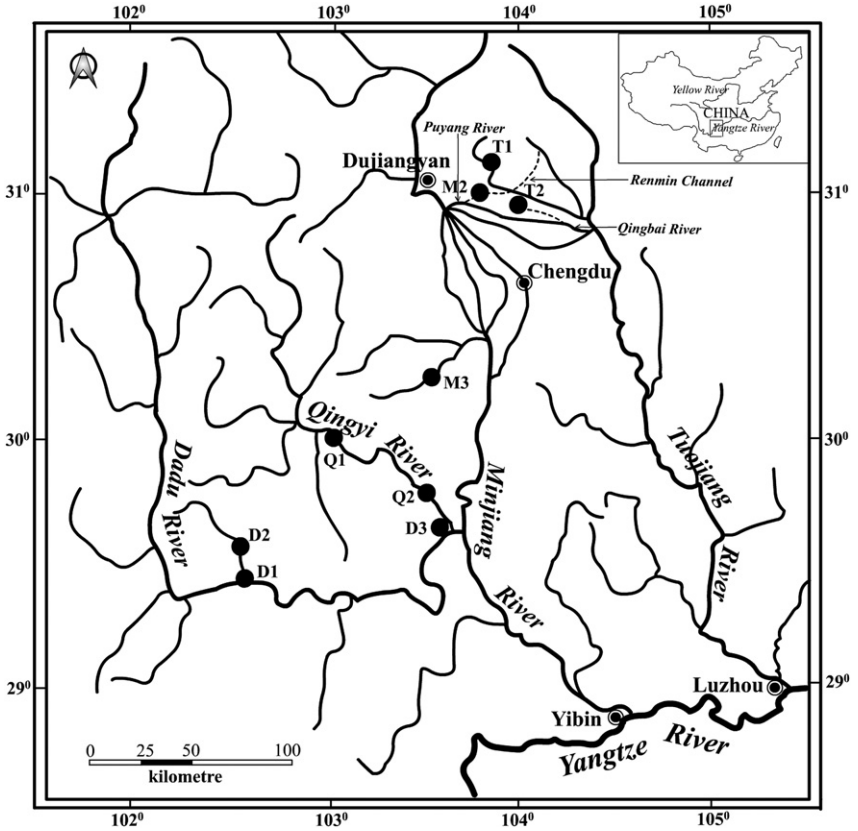

Fig. 1. Map of the sampling locations for *G. rarus*. The dotted line means that channels are invisible on the present scale. Three man-made channels (Renmin Channel, Puyang River, Qingbai River) are shown by arrows.

A total of 416 individuals sampled by nets and electricity equipment were used in this study, i.e., 30–50 individuals per site (Table 1). After capture, the fish were placed into a water vat to keep them alive during transportation. Then the fish were anaesthetized by MS222 and stored in 95% ethanol at −20 °C until further analysis. To obtain genetic material, a piece of fin was removed and used for genomic DNA extraction.

2.2. Microsatellite analyses

Genomic DNA of each sample was extracted using a salt extraction protocol outlined by Aljanabi and Martinez (1997). Microsatellite primers were synthesized as described in Liao et al. (2007) and He and Wang (2010). We screened 15 potential microsatellite primer pairs and found 8 polymorphic loci (GR08, GR22, GR29, Gra02, Gra04, Gra16, Gra25, and Gra30) to be with distinct bands. These eight loci were chosen in the present study. The polymerase chain reaction was carried out according to the procedure as the author described in He and Wang (2010).

2.3. Genetic diversity

Observed and expected heterozygosities for each microsatellite locus and each location were calculated using MSA 4.05 (Dieringer and Schlotterer, 2003). Allelic richness, the observed number of alleles, departures from Hardy–Weinberg equilibrium, and genotypic linkage disequilibrium were calculated using FSTAT v.2.9.3 (Goudet, 2001). Significance levels were adjusted for multiple comparisons using the standard Bonferroni procedure (Rice, 1989).

2.4. Genetic differentiation among populations

The difference of allelic richness and expected heterozygosity among locations were estimated using the Kruskall–Wallis test. Their multiple comparison tests were done using the *pgirmess* library of the *R* software (Ihaka and Gentlemen, 1996). Allelic frequencies and their differences among locations were calculated by Fisher's exact test using Genepop v. 4.0 (Raymond and Rousset, 1995). These differences over genetic diversity indices among locations were part of population genetic structures of *G. rarus*.

Besides, four complementary approaches were also used to measure the population genetic structure of *G. rarus*. Firstly, F_{ST} between pairs of population were calculated using MSA. The significant level of pairwise F_{ST} was adjusted by the Bonferroni procedure ($\alpha = 0.0014$). Then, we used MSA to estimate G'_{ST}, a standardized measure of global genetic differentiation that is independent of the amount of genetic variation observed at the examined loci, to facilitate comparisons with other studies (Hedrick, 2005). The statistical significance of pairwise F_{ST} and G'_{ST} was tested by 10,000 permutations.

Secondly, as an alternative to traditional F_{ST} methods, the Bayesian clustering method in STRUCTURE ver. 2 (Pritchard et al., 2000; Falush et al., 2003) was used to measure the population genetic structure. By using the admixture model and correlated allele frequency parameters, ten replicates of each run from $K = 1$ to $K = 9$ (K is a number of genetic clusters) were performed. Each replicate was run for 20,000 Markov chain Monte Carlo (MCMC) generations with an initial burn-in of 20,000 generations. Both $L(K)$ and ΔK (Evanno et al., 2005) were used to select the most likely number of clusters in the present study. Then we calculated the fractional membership (q) of each individual in each cluster (Pritchard et al., 2000) in order to establish the status of each location. Locations that have a q higher than 75% for a given cluster were considered to be strongly attached with one cluster, while locations that have no q higher than 75% were considered as shared membership between clusters.

Thirdly, AMOVA in ARLEQUIN 2.000 was used to partition the total variance among groups, among populations within groups or within populations (Excoffier et al., 1992; Schneider et al., 2000). Three different group divisions were considered in the present study: 1) four groups according to river basins: Group M (M2, M3), Group T (T1, T2), Group Q (Q1, Q2) and Group D (D1, D2 and D3); 2) two groups according to the formation of alluvial plains: Group MT (M2, M3, T1 and T2) and Group QD (Q1, Q2, D1, D2 and D3); 3) groups according to the outputs of the STRUCTURE analysis described above.

Table 1
Sample location information, including GPS coordinates, altitude and sample size.

Populations	Attributes	GPS locations		Altitude (m)	Sample size
		Latitude	Longitude		
T1	Tuojiang River	31°08′00.6″	103°50′58.3″	792	50
T2	Tuojiang River	30°58′53.1″	103°59′45.8″	566	50
M2	Minjiang River	30°58′46.3″	103°50′02.9″	627	35
M3	Minjiang River	30°26′09.1″	103°19′29.6″	513	50
D1	Dadu River	29°20′18.6″	102°40′21.7″	764	31
D2	Dadu River	29°28′37.3″	102°37′35.4″	939	50
D3	Dadu River	29°34′10.1″	103°40′18.0″	412	50
Q1	Qingyi River	29°59′12.6″	103°04′10.7″	545	50
Q2	Qingyi River	29°40′56.0″	103°34′33.3″	387	50

Finally, a phylogenetic (Neighbor-Joining) tree was built to visualize grouping patterns among locations in the software POPULATIONS version 1.2.30 (Langella, 1999) by using the Nei's genetic distance and 10,000 bootstraps on individuals (Nei et al., 1983; Takezaki and Nei, 2008). This phylogenetic tree was displayed using TreeView (Page, 1996).

2.5. Gene flow and migration rate among populations

We investigated gene flow among locations by using two approaches. Firstly, we measured the relationship between the genetic distance (Slatkin, 1985) and the geographical distance among locations by using a Mantel (1967) permutation procedure in the *vegan* library (Oksanen et al., 2008) of the *R* software. Three different geographical distance measurements were used, respectively: straight line distance (SLD), water course distance (WCD) and riparian distance (RD). SLD was calculated through the online distance calculator between each two GPS coordinates, while RD was calculated along the closest connected water system (e.g. through man-made channels between the Minjiang and Tuojiang Rivers, Fig. 1) in the Google Earth version 5.0. WCD was calculated along the natural river networks (e.g. through the mouth of the Minjiang River and the Tuojiang River, but not man-made channels, Fig. 1) in the Google Earth version 5.0. In order to differentiate between the multiple regression models, Akaike Information Criterion (AIC; Burnham and Anderson, 2002) and Δ_i ($\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$) were calculated through *glm* function (Dobson, 1990) in the *R* software.

Secondly, the migration rates between locations within the past few generations were estimated using BAYESASS 1.3 (Wilson and Rannala, 2003). A total of 3×10^6 iterations (discarding the first 10^6 iterations as burn-in) and a sampling frequency of 2000 were used to estimate the proportion of immigrants into a population, the mean migration rate and their confidence intervals (CI). In this study, delta values for migration rate (m), allele frequencies (P), and inbreeding values (F) were set at 0.15, 0.15 and 0.15. These values achieved the recommended acceptance rates of changes of 40–60% (Wilson and Rannala, 2003).

3. Results

3.1. Genetic diversity

After adjusting for multiple comparisons, no significant linkage disequilibrium was found in any pairs of loci, while significant departures from Hardy–Weinberg equilibrium were found for several loci in some locations (Table 2). However, those loci did not show consistent deviations across all locations. Therefore, we assumed that processes causing this non-equilibrium were specific to those locations, and continued to include those loci in subsequent analyses.

The number of alleles detected at a locus ranged from 4 (locus GR22, Gra04 and Gra25) to 18 (locus GR29), averaging 7.9 over all loci. A total of 63 microsatellite alleles were observed among all geographic locations for rare minnow across eight loci, ranging from a maximum of 61 alleles in T2 to a minimum of 41 alleles in Q1. Allelic richness across all loci varied from 4.8 (Q1) to 7.3 (T2). The observed heterozygosity ranged from 0.485 to 0.675, with a mean of 0.593 across all loci and locations. The expected heterozygosity ranged between 0.678 and 0.782, with a mean of 0.730 over all loci and locations (Table 2).

3.2. Genetic structure

There were no significant differences in allelic richness and expected heterozygosity between locations ($P > 0.05$). However, Fisher's exact test revealed significant allelic frequency differentiations between locations after Bonferroni correction ($P < 0.001$). Based on F_{ST} estimates, significant structure also existed among *G. rarus* locations (Table 3). Pairwise F_{ST} values ranged from 0.013 to 0.154 among all nine geographic locations, with an average $F_{ST} = 0.061$ (Table 3). These values represented low to moderate levels of population differentiation. An overall randomization test of population differentiation was significant for each pair of locations after Bonferroni correction. The largest genetic differentiation was between T1 and Q1 ($F_{ST} = 0.154$), indicating that they were quite isolated populations. The standardized global genetic differentiation measure $G'_{ST} = 0.26$ ($p < 0.001$) also indicated significant levels of genetic differentiation among nine wild locations of *G. rarus*.

Table 2
Genetic diversity of each wild population of *G. rarus* was revealed by eight microsatellite loci, including average allele numbers per locus (A), average allelic richness (AR), average observed heterozygosity (H_O), average expected heterozygosity (H_E), inbreeding coefficients (F_{IS}) and loci with significant departures from Hardy–Weinberg (HW) proportions. The standard error is given in parentheses.

Populations	A	AR	H_O	H_E	F_{IS}	HW disequilibrium
T1	6.50 (1.57)	6.13 (1.43)	0.573 (0.094)	0.685 (0.076)	0.166	GR22/Gra16/Gra25
T2	7.63 (1.64)	7.30 (1.53)	0.564 (0.026)	0.761 (0.035)	0.260	GR29/Gra02/Gra04
M2	7.00 (1.36)	6.83 (1.29)	0.625 (0.054)	0.764 (0.036)	0.185	–
M3	6.63 (0.84)	6.45 (0.82)	0.649 (0.037)	0.781 (0.027)	0.170	–
D1	5.50 (0.95)	5.47 (0.93)	0.536 (0.058)	0.681 (0.041)	0.215	–
D2	6.00 (0.89)	5.74 (0.83)	0.604 (0.068)	0.730 (0.031)	0.174	Gra16
D3	6.63 (1.22)	6.44 (1.17)	0.625 (0.054)	0.724 (0.043)	0.138	–
Q1	5.13 (0.40)	4.79 (0.31)	0.485 (0.056)	0.676 (0.024)	0.285	GR22/GR29/Gra04
Q2	7.00 (1.55)	6.74 (1.45)	0.675 (0.052)	0.756 (0.035)	0.108	–

Table 3
Matrix of pairwise F_{ST} (below diagonal) and riparian geographical distance (km, above diagonal) between populations of *G. rarus* is listed.

Populations	T1	T2	M2	M3	D1	D2	D3	Q1	Q2
T1	0.0000	138.41	163.12	332.43	605.27	624.45	398.56	504.7	416.30
T2	0.0656[a]	0.0000	62.90	239.16	531.18	512.00	305.29	411.43	323.03
M2	0.0580[a]	0.0205[a]	0.0000	181.05	453.89	473.07	247.18	353.31	264.91
M3	0.0881[a]	0.0404[a]	0.0130[b]	0.0000	425.64	444.82	218.93	325.07	236.67
D1	0.0872[a]	0.0652[a]	0.0614[a]	0.0678[a]	0.0000	19.18	217.29	323.42	235.02
D2	0.0752[a]	0.0564[a]	0.0465[a]	0.0344[a]	0.0249[b]	0.0000	236.47	342.60	254.20
D3	0.1214[a]	0.0576[a]	0.0471[a]	0.0433[a]	0.0688[a]	0.0495[a]	0.0000	116.71	28.31
Q1	0.1537[a]	0.0753[a]	0.0650[a]	0.0575[a]	0.0462[a]	0.0722[a]	0.0559[a]	0.0000	88.40
Q2	0.0971[a]	0.0351[a]	0.0268[a]	0.0272[a]	0.0778[a]	0.0465[a]	0.0394[a]	0.0809[a]	0.0000

[a] Represents $P < 0.01$.
[b] Represents $P < 0.05$.

Examination of $L(K)$ values from the STRUCTURE for successive K values showed a maximum likelihood value at $K = 6$, the steepest increase and lowest standard deviation between $K = 1$ and $K = 2$ (Fig. 2). Calculation of ΔK (Evanno et al., 2005) produced a modal value of the statistic at $K = 2$ (Fig. 2). While for the value of ΔK, there was a second smaller mode at $K = 6$. Evanno et al. (2005) revealed that the height of the modal values of ΔK indicated the strength of the population subdivision signal, suggesting deep subdivision at $K = 2$, and less pronounced differentiation at $K = 6$ in the present study. However, $K = 2$ appeared to be the most optimal subdivision for its high cluster membership q values (>60%). For $K = 6$, most of the locations have very low cluster membership ($q < 50\%$). Thus, the subdivision $K = 2$ suggested that the uppermost level of hierarchical genetic structure has two distinct clusters *C*1 and *C*2 (see Fig. 3). Most individuals of M2, M3, D1–D3 and Q2 showed a shared

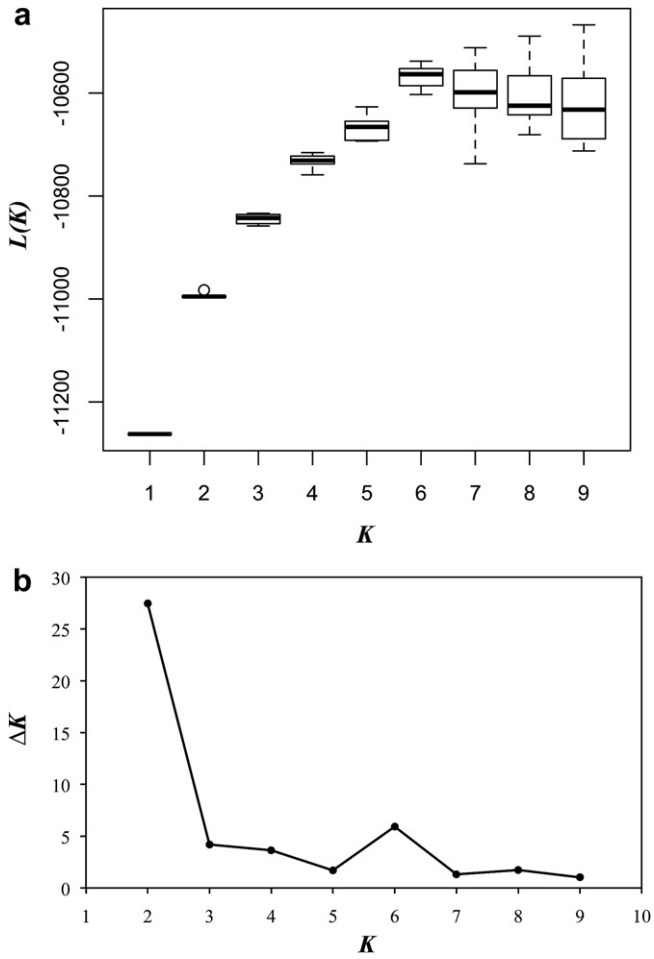

Fig. 2. Different K values obtained by STRUCTURE. (a) Mean (±SD) of $L(K)$ over 10 STRUCTURE runs for successive K values on the overall data set. (b) ΔK as calculated by Evanno et al. (2005): the modal value (here for $K = 2$) shows the uppermost level of genetic structure.

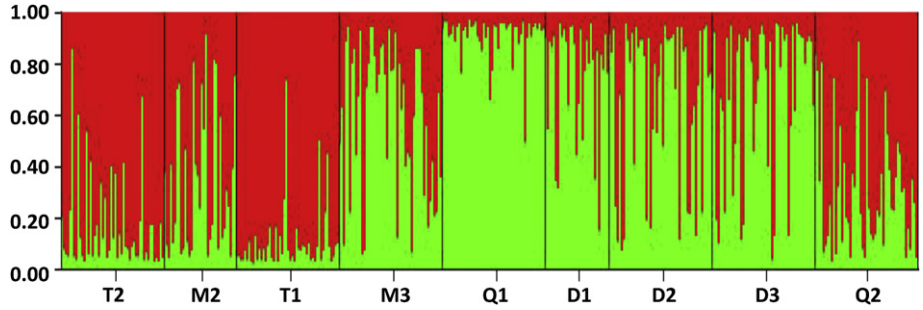

Fig. 3. Assignment of individuals of *G. rarus* using STRUCTURE based on sample locations and $K = 2$. Colors correspond to each cluster. Each bar represents a single individual sample and present in groups based on sampling location.

genetic pattern while most individuals of T1 and Q1 exhibited dominance of an alternate pattern. Each of the nine geographic locations was assigned to one cluster based on their proportion of membership to both clusters, the right one of which was that with the highest probability of membership. Clusters *C*1 and *C*2 consisted of four (T1, T2, Q2 and M2) and five (Q1, D1, D2, D3 and M3) geographic locations, respectively. Four of the nine geographic locations were strongly assigned to one cluster (2/

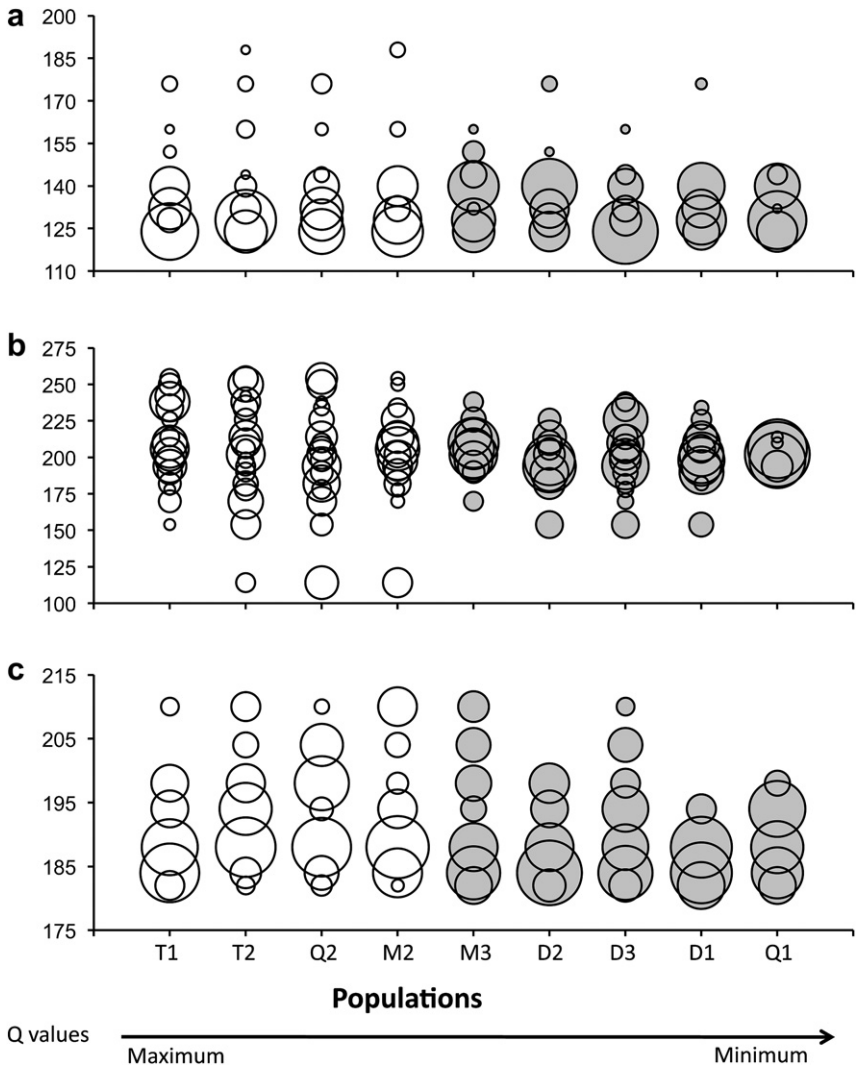

Fig. 4. Schematic illustration of relative allelic frequency of three out of eight loci (a – GR08, b – GR29, c – Gra30) in nine wild populations that showed the highest difference between two STRUCTURE clusters (*C*1 in white circles and *C*2 in gray circles). The populations were ordered from the maximum to minimum of the *q* values (assigned to *C*1).

Table 4
Hierarchical analysis of molecular variance (AMOVA) in the wild populations of *G. rarus*. Three different group divisions were used: four river basin groups, two alluvial plain groups, and two STRUCTURE clusters.

Source of variation	d.f.	Variance components	Percent variation	Fixation index (*F*)
Four river basin groups				
Among groups	3	0.031	0.98	0.010
Among populations within groups	5	0.170	5.45	0.055[a]
Within populations	813	2.920	93.58	0.064[a]
Two alluvial plain groups				
Among groups	1	0.040	1.28	0.013
Among populations within groups	7	0.173	5.53	0.056[a]
Within populations	813	2.920	93.19	0.068[a]
Two STRUCTURE clusters				
Among groups	1	0.049	1.55	0.016[b]
Among populations within groups	7	0.168	5.37	0.055[a]
Within populations	813	2.920	93.08	0.069[a]

[a] Significant at $P < 0.001$.
[b] Significant at $P < 0.05$.

4 for *C*1, 2/5 for *C*2, threshold = 75%). The remaining five locations shared the membership between clusters and had a *q* value lower than 70%. All locations had a proportion of membership in one cluster of at least 60%.

Such a last result suggests mixing to some extent between the two clusters (*C*1 and *C*2) described above. Accordingly, a comparison of allelic frequencies between the two clusters revealed significant differences across all the loci (Fig. 4). Indeed, some loci (GR08, GR29 and Gra30) showed remarkable difference, with each cluster being characterized by specific patterns of allelic frequency and a tendency of changing allelic frequency between the two clusters (Fig. 4). For example, some large alleles with low frequencies in loci GR08 and Gra30 disappeared gradually from *C*1 to *C*2, and middle-sized alleles with high frequencies in locus GR29 gradually dominated in *C*2 (Fig. 4).

Without group information in the data set, hierarchical AMOVA detected significant levels of structure among all the locations. Specifically, 6.3% of the variation was from the variation among the locations ($F_{ST} = 0.063$). With group information included in the data set, hierarchical AMOVA did not detect significant levels of structure either among four river basin groups or between two alluvial plain groups ($P > 0.05$, Table 4). However, a significant difference between two STRUCTURE clusters was detected ($P < 0.05$, Table 4).

According to the phylogenetic tree built by using Neighbor-Joining (NJ) methods, two locations (T1 and Q1) with the largest distance were identified (Fig. 5). This topology also indicated a grouping in accordance with the geographical location of the populations in a degree.

3.3. Gene flow

Results from the Mantel test showed a weak but significant relationship between $F_{ST}/(1 - F_{ST})$ and geographical distance ($P < 0.05$; Table 5; Fig. 6), indicating the existence of a geographic isolation among the wild populations of *G. rarus*. The

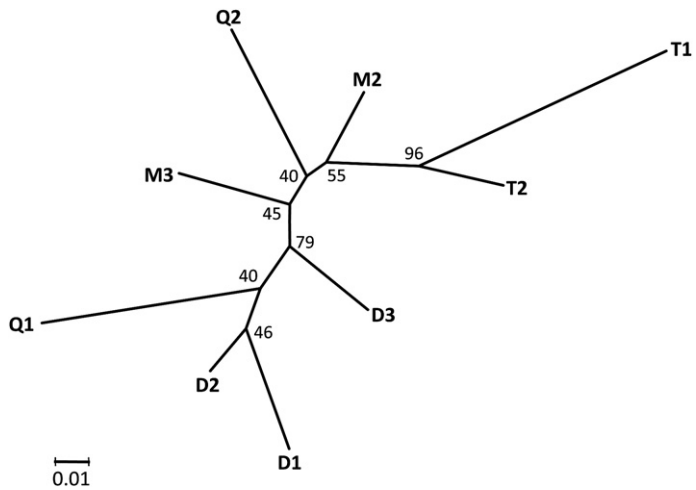

Fig. 5. A phylogenetic (neighbor-joining) tree based on Nei's genetic distance (*Da*, 1983) of 9 wild populations of *G. rarus*. The numbers above the lines are the proportion of similar replicates supporting each node based on 10,000 bootstrap simulations.

Table 5
Information about the regression of genetic differentiation, measured as $F_{ST}/(1-F_{ST})$, on the log of the geographical distance between population pairs. Three different kinds of geographical distance measurement (straight line distance – SLD; water course distance – WCD; riparian distance – RD) were used in the present study. *r* – correlation coefficient; *P* – significant value; AIC – Akaike's Information Criterion.

Geographic distance	$F_{ST}/(1-F_{ST})$		
	r	*P*	AIC
SLD	0.3365	0.0015	85.90478
WCD	0.3215	0.0306	108.0814
RD	0.4188	0.0012	82.54997

highest correlation coefficient was obtained between $F_{ST}/(1-F_{ST})$ and riparian distance, which is shown in Fig. 6. This model also gave the lowest AIC value (Table 5), and all the Δ_i values for it were large than 2.

Based on estimates from BAYESASS, there were no instances of significant immigration rates among all the sample locations, because all the mean estimated migration rates fell within the confidence intervals expected in cases of insufficient signal in the data (95% CI: 4.53×10^{-10}, 0.126; Table 6). Furthermore, recent immigration rates among the majority of sampled locations were quite low ($m < 0.01$) with a high proportion of individuals derived from their own location (>0.90), suggesting

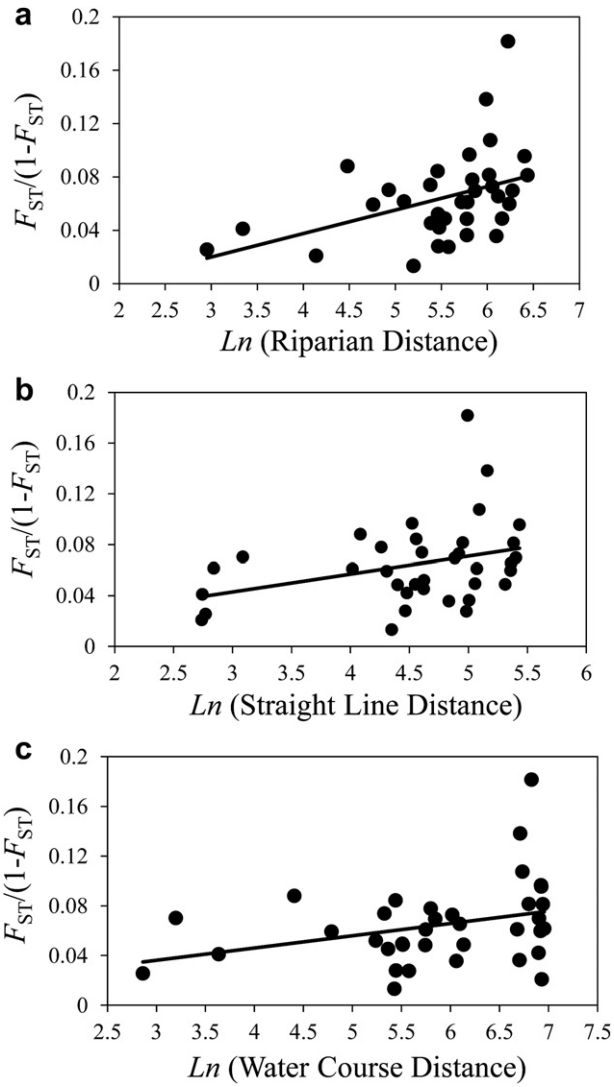

Fig. 6. The correlation model between $F_{ST}/(1-F_{ST})$ and ln(geographical distance, km) of *G. rarus* wild populations is shown, indicating an isolation-by-distance pattern. Three different geographical distances: Riparian distance (a), Straight line distance (b), and Water course distance (c), were used.

Table 6
Bayesian estimates of recent migration rates among wild populations of *G. rarus* using the program BAYESASS. Values shown are the mean migration rate into each population and their respective 95% confidence intervals in parentheses. Values along the diagonal (in bold) are the proportion of individuals derived from the source population for each generation.

		Migration into								
		T1	T2	M2	M3	D1	D2	D3	Q1	Q2
Migration from	T1	**0.991 (0.967–1.000)**	0.015 (0.000–0.058)	0.011 (0.000–0.055)	0.003 (0.000–0.022)	0.005 (0.000–0.032)	0.005 (0.000–0.028)	0.009 (0.000–0.038)	0.001 (0.000–0.009)	0.002 (0.000–0.014)
	T2	0.001 (0.000–0.009)	**0.934 (0.852–0.992)**	0.013 (0.000–0.059)	0.003 (0.000–0.022)	0.005 (0.000–0.032)	0.003 (0.000–0.022)	0.019 (0.000–0.064)	0.001 (0.000–0.007)	0.013 (0.000–0.057)
	M2	0.001 (0.000–0.009)	0.009 (0.000–0.051)	**0.783 (0.700–0.879)**	0.003 (0.000–0.021)	0.005 (0.000–0.032)	0.004 (0.000–0.023)	0.008 (0.000–0.038)	0.001 (0.000–0.008)	0.004 (0.000–0.025)
	M3	0.001 (0.000–0.009)	0.004 (0.000–0.026)	0.112 (0.024–0.210)	**0.965 (0.891–0.999)**	0.005 (0.000–0.029)	0.005 (0.000–0.033)	0.041 (0.001–0.111)	0.001 (0.000–0.008)	0.004 (0.000–0.024)
	D1	0.001 (0.000–0.010)	0.007 (0.000–0.043)	0.006 (0.000–0.035)	0.004 (0.000–0.027)	**0.850 (0.709–0.984)**	0.010 (0.000–0.065)	0.029 (0.000–0.099)	0.001 (0.000–0.010)	0.006 (0.000–0.043)
	D2	0.001 (0.000–0.008)	0.007 (0.000–0.030)	0.016 (0.000–0.062)	0.004 (0.000–0.031)	0.075 (0.000–0.213)	**0.959 (0.877–0.999)**	0.048 (0.000–0.158)	0.001 (0.000–0.010)	0.010 (0.000–0.047)
	D3	0.001 (0.000–0.009)	0.007 (0.000–0.032)	0.012 (0.000–0.060)	0.004 (0.000–0.027)	0.007 (0.000–0.038)	0.003 (0.000–0.024)	**0.801 (0.714–0.898)**	0.001 (0.000–0.010)	0.006 (0.000–0.036)
	Q1	0.001 (0.000–0.009)	0.005 (0.000–0.031)	0.030 (0.000–0.085)	0.004 (0.000–0.032)	0.041 (0.000–0.117)	0.008 (0.000–0.046)	0.037 (0.000–0.105)	**0.991 (0.970–1.000)**	0.004 (0.000–0.023)
	Q2	0.001 (0.000–0.009)	0.013 (0.000–0.058)	0.016 (0.000–0.069)	0.008 (0.000–0.060)	0.006 (0.000–0.032)	0.002 (0.000–0.017)	0.010 (0.000–0.048)	0.001 (0.000–0.008)	**0.950 (0.876–0.994)**

that most areas are isolated from each other, at least with respect to first- and second-generation immigrants. However, we detected one relatively high proportion of immigrants ($m = 0.112$) from M3 into M2, which was close to the upper level of expected values ($m = 0.126$).

4. Discussion

4.1. Genetic structure

The present study revealed a low to moderate level of genetic differentiation among wild populations of *G. rarus*. Similarly, a distinct genetic structure was also detected between clusters by using AMOVA. Moreover, STRUCTURE and phylogenetic analysis implied that this clustering pattern of *G. rarus* may be highly correlated with the structure of the water systems.

As we described in the Material and methods, the sampling sites in the present study almost covered all the distribution range of *G. rarus*, which were located in the Chengdu Plain and its neighboring area. In the view of the water system structure, all the sampling sites were from four river basins (Minjiang River, Tuojiang River, Qingyi River, Dadu River). While in the view of the formation of the Chengdu Plain, they were from two alluvial plains (Qian and Tang, 1997). One was from the alluvial plain of Minjiang and Tuojiang Rivers (MT), and another was from the alluvial plain of the Qingyi and Dadu Rivers (QD). Actually, there were no significant genetic differences either among four river basin groups or between two alluvial plain groups. However, significant genetic differences were found between two STRUCTURE clusters, which were confirmed by the analysis of the frequency distribution of alleles. Compared with the component of two alluvial plain groups, the component of two STRUCTURE clusters was similar except for Q2 and M3.

The histories of river evolution usually act as an important force in the biogeography or distribution of many freshwater fish species. We deduced that the clustering structure in the present study was mainly correlated to the evolution of the Qingyi River. Yuan and Tao (2008) showed that the drainage evolution of the Qingyi River consisted mainly of four stages. During the Middle Pleistocene Epoch, the Qingyi River flowed northwards into the Minjiang River at Xinjin County, where Qionglai River (M3) joined. However, along with the arrival of the Late Pleistocene Epoch, the Qingyi River went southeastwards through Jiajiang County (Q2) into the Dadu River at Leshan City (Li et al., 2006; Li and Guo, 2008; Yuan and Tao, 2008). It was deduced that Q2 did not belong to the Qingyi River basin until the middle stage of the Late Pleistocene Epoch (about 50 thousand years ago), and M3 might have belonged to the Qingyi River basin around the late stage of the Middle Pleistocene Epoch (about 0.15 million years ago). Therefore, the population genetic structure pattern of *G. rarus* could be highly correlated with the structure of the water systems and its evolution history.

4.2. Isolation by distance and distribution pattern

The present study revealed that the best fitted regression model was between F_{ST} and the riparian distance for its lowest AIC value and Δ_i value (Burnham and Anderson, 2004). The most likely dispersal route of *G. rarus* in history might through the man-made channels (e.g., Renmin Channel, Puyang River, Qingbai River, Fig. 1) resulting from the Dujiangyan Irrigation Project being built around 2300 years ago by Li Bing and his son, or through other similar channels resulting from river floods or water system changes (Liu et al., 2006). Therefore, it is easy to understand why no rare minnow was sampled in the river mouth of the Minjiang River at Yibin City and the river mouth of the Tuojiang River at Luzhou City up to now, where high intensities of investigation on fish catches were usually carried out. Because rare minnow didn't migrate through this route. There is need to state that modern man-made channels being channelized or polluted didn't play an important role in the migration of *G. rarus* because very limited recent gene flow among populations was revealed in the present study.

In view of the present distribution and differentiation of *G. rarus*, we can surmise that rare minnow had been widely dispersed in the Chengdu Plain and its neighboring area along with the evolution of the Minjiang and Tuojiang Rivers in history. However, due to urbanization, pollution and the construction of water diversion projects, a large number of rare minnow habitats were lost. Then the remnant populations only survive in narrow areas with fewer anthropogenic disturbances, especially at the edge of the Chengdu Plain at present. These populations usually have dozens to hundreds of kilometers away from each other, exhibiting a discontinuous or spotted distribution pattern.

4.3. Conservation implication

The present study showed significant differentiations among wild populations of *G. rarus*. No matter what is the level of genetic differentiation, conservation measures should be taken for all populations (Wright, 1978; Charlesworth, 1998; Nagylaki, 1998; Hedrick, 1999; Balloux and Lugon-Moulin, 2002). It is especially important for fish species like *G. rarus* with narrow and spotted distribution. However, in fact an important principle for species conservation is to select prior populations for preserving because of limited natural resources, which could be implemented from the determination of Evolutionarily Significant Units (ESUs) and Management Units (MUs). In conservation genetics, ESUs are one of most important conservation units for endangered species (Waples, 1991; Parker et al., 1999; Hedrick et al., 2001; Holycross and Douglas, 2007; Krabbenhoft et al., 2008; Morgan et al., 2008). Moritz (1994) defined ESUs as being reciprocally monophyletic for mitochondrial DNA (mtDNA) alleles and show significant divergence of allele frequencies at nuclear loci. Except for the present study by using microsatellites, we also had sequenced some cyt*b* and D-loop genes in mtDNA of rare minnow

wild populations. However, they owned less polymorphism, which was not suitable for determining ESUs. Therefore, further studies should be carried out to determine the ESUs of *G. rarus*.

According to historical surveys and the present study, all the wild populations of *G. rarus* are facing the threats from anthropogenic disturbance. Among them, some populations (e.g. M2, Q1, D1 and D3) in both clusters *C*1 and *C*2 are experiencing major environmental threats, where conservation measures could not be carried out effectively. However, other populations such as T1, T2 and Q2 in cluster *C*1, M3 and D2 in cluster *C*2 exhibiting large population size, high genetic diversity, extensive allele distribution, favorable habitats and less anthropogenic disturbance should be in prior conservation. For the sake of desired conservation effect, it is necessary to establish the natural reserve and to restore the habitat of *G. rarus*.

Acknowledgments

The authors are most grateful to F. Wang for his technical assistance. We also thank Dr. E. Quéméré and Prof. O. Langella for their kind help in using the software STRUCTURE and POPULATIONS. This research was supported by funds from the National Natural Science Foundation of China (NSFC, 30670292), the Chinese Academy of Sciences (CZBZX-1), the Ministry of Environmental Protection of China (2009ZX07527-005) and the French embassy by providing PhD grants for the co-supervision of the PhD research.

References

Aljanabi, S.M., Martinez, I., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25, 4692–4693.
Aspi, J., Roininen, E., Kiiskila, J., Ruokonen, M., Kojola, I., Bljudnik, L., Danilov, P., Heikkinen, S., Pulliainen, E., 2009. Genetic structure of the northwestern Russian wolf populations and gene flow between Russia and Finland. Conserv. Genet. 10, 815–826.
Balloux, F., Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11, 155–165.
Beaumont, M.A., Rannala, B., 2004. The Bayesian revolution in genetics. Nat. Rev. Genet. 5, 251–261.
Björklund, M., Aho, T., Larsson, L.C., 2007. Genetic differentiation in pikeperch (*Sander lucioperca*): the relative importance of gene flow, drift and common history. J. Fish Biol. 71, 264–278.
Blanchet, S., Rey, O., Etienne, R., Lek, S., Loot, G., 2010. Species-specific responses to landscape fragmentation: implications for management strategies. Evol. Appl. 3 (3), 291–304.
Bohonak, A.J., 1999. Dispersal, gene flow, and population structure. Q. Rev. Biol. 74, 21–45.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference – understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.
Charlesworth, B., 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15, 538–543.
Chen, Y.Y., 1998. Fauna Sinica, Osteichthyes, Cypriniformes II. Science Press, Beijing, pp. 51–52 (in Chinese).
Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M., Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15, 290–295.
Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T., Hendry, A.P., 2006. The relative influence of natural selection and geography on gene flow in guppies. Mol. Ecol. 15, 49–62.
DiBattista, J.D., 2008. Patterns of genetic variation in anthropogenically impacted populations. Conserv. Genet. 9, 141–156.
Dieringer, D., Schlotterer, C., 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes 3, 167–169.
Ding, R.H., 1994. The Fishes of Sichuan, China. Sichuan Science and Technology Press, Chengdu (in Chinese).
Dobson, A.J., 1990. An Introduction to Generalized Linear Models. Chapman and Hall, London.
Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.
Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial-DNA restriction data. Genetics 131, 479–491.
Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.
Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to Conservation Genetics. Cambridge University Press, New York.
Fraser, D.J., Bernatchez, L., 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol. Ecol. 10, 2741–2752.
Fraser, D.J., Lippe, C., Bernatchez, L., 2004. Consequences of unequal population size, asymmetric gene flow and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). Mol. Ecol. 13, 67–80.
Goudet, J., 2001. FSTAT, A Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3). Available from: http://www.unil.ch/izea/softwares/fstat.html.
He, Y.F., Wang, J.W., 2010. Temporal variation in genetic structure of the Chinese rare minnow (*Gobiocypris rarus*) in its type locality revealed by microsatellite markers. Biochem. Genet. 48, 312–325.
Hedrick, P.W., 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53, 313–318.
Hedrick, P.W., 2005. A standardized genetic differentiation measure. Evolution 59, 1633–1638.
Hedrick, P.W., Parker, K.M., Lee, R.N., 2001. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. Mol. Ecol. 10, 1399–1412.
Holycross, A.T., Douglas, M.E., 2007. Geographic isolation, genetic divergence, and ecological non-exchangeability define ESUs in a threatened sky-island rattlesnake. Biol. Conserv. 134, 142–154.
Ihaka, R., Gentlemen, R., 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. 5, 299–314.
Johnson, M.S., 2000. Measuring and interpreting genetic structure to minimize the genetic risks of translocations. Aquacult. Res. 31, 133–143.
Krabbenhoft, T.J., Rohde, F.C., Leibman, A.N., Quattro, J.M., 2008. Concordant mitochondrial and nuclear DNA partitions define evolutionarily significant units in the imperiled pinewoods darter, *Etheostoma mariae* (Pisces: Percidae). Copeia 4, 909–915.
Langella, O., 1999. Populations Version 1.2.30. Distributed by the author, CNRS UPR9034, France.
Li, Y.Z., Guo, B., 2008. Cenozoic tectonics of Chengdu Plain, Sichuan, China. J. Chengdu Univ. Technol. (Sci. Technol. Ed.) 35 (4), 371–376 (in Chinese).
Li, L., Ma, T.W., Wu, Z.B., 2004. Toxic effect of domestic sewage on rare minnow (*Gobiocypris rarus*). Acta Hydrobiol. Sin. 28 (1), 40–44 (in Chinese).
Li, Y., Li, B., Steffen, D., Densmore, A.L., Richardson, N.J., Zhou, R.J., Ellis, M.A., Zhang, Y., 2006. Provenance analysis and drainage evolution in Late Cenozoic Chengdu Basin on eastern margin of Tibetan Plateau. Acta Sedimentol. Sin. 24 (3), 309–320 (in Chinese).

Liao, X.L., Wang, D., Yu, X.M., Li, W.T., Cheng, L., Wang, J.W., Tong, J.G., 2007. Characterization of novel microsatellite loci in rare minnow (*Gobiocypris rarus*) and amplification in closely related species in Gobioninae. Conserv. Genet. 8, 1003–1007.
Liu, X.L., Su, C.J., Xu, Y., Zhang, J.Y., 2006. Water resource problem and sustainable utilization in the upper reaches of Minjiang River. Res. Soil Water Conserv. 13 (3), 189–191 (in Chinese).
Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.
Morgan, M.J., Hunter, D., Pietsch, R., Osborne, W., Keogh, J.S., 2008. Assessment of genetic diversity in the critically endangered Australian corroboree frogs, *Pseudophryne corroboree* and *Pseudophryne pengilleyi*, identifies four evolutionarily significant units for conservation. Mol. Ecol. 17, 3448–3463.
Moritz, C., 1994. Defining evolutionarily-significant-units for conservation. Trends Ecol. Evol. 9, 373–375.
Nagylaki, T., 1998. Fixation indices in subdivided populations. Genetics 148, 1325–1332.
Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19, 153–170.
Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2008. Vegan: Community Ecology Package. R Package Version 1.15-0. http://cran.r-project.org/. http://vegan.r-forge.r-project.org/.
Page, R.D.M., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Parker, K.M., Sheffer, R.J., Hedrick, P.W., 1999. Molecular variation and evolutionarily significant units in the endangered Gila topminnow. Conserv. Biol. 13, 108–116.
Pearse, D.E., Crandall, K.A., 2004. Beyond F_{ST}: analysis of population genetic data for conservation. Conserv. Genet. 5, 585–602.
Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., Estoup, A., 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. J. Hered. 95, 536–539.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Qian, H., Tang, R.C., 1997. On the formation and evolution of the Chengdu Plain. Earthq. Res. Sichuan 3, 1–7 (in Chinese).
Raeymaekers, J.A.M., Maes, G.E., Geldof, S., Hontis, I., Nackaerts, K., Volckaert, F.A.M., 2008. Modeling genetic connectivity in sticklebacks as a guideline for river restoration. Evol. Appl. 1, 475–488.
Raymond, M., Rousset, F., 1995. GENEPOP (Version-1.2) – population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225.
Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin ver.2.000: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Slatkin, M., 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16, 393–430.
Takezaki, N., Nei, M., 2008. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. Genetics 178, 385–392.
Wang, J.W., Cao, W.X., 1997. *Gobiocypris rarus* and fishes as laboratory animals. Trans. Chin. Ichthyol. Soc. 6, 144–152 (in Chinese).
Wang, S., Xie, Y. (Eds.), 2004. China Species Red List. Red List, vol. 1. Higher Education Press, Beijing, pp. 154 (in Chinese).
Wang, T.H., Liu, P.L., Chen, H.X., Liu, H.Q., Yi, Y.L., Guo, W., 1994. Preliminary study on the susceptibility of *Gobiocypris rarus* to Hemorrhagic Virus of Grass Carp (GCHV). Acta Hydrobiol. Sin. 18 (2), 144–149 (in Chinese).
Wang, Z.H., Yin, Y.W., Xu, Z.N., Zhou, J., Zhang, Q., Zhang, D.P., 1998. Acute and subchronic toxicity of pyrethroid insecticides to *Gobiocypris rarus*. Chin. Appl. Environ. Biol. 4 (4), 379–382 (in Chinese).
Waples, R.S., 1991. Pacific Salmon, *Oncorhynchus* spp. and the definition of species under the endangered species act. Mar. Fish. Rev. 53, 11–22.
Wilson, G.A., Rannala, B., 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163, 1177–1191.
Wright, S., 1978. Evolution and the Genetics of Population, Variability Within and Among Natural Populations. The University of Chicago Press, Chicago.
Xiong, D.M., Xie, C.X., Xia, L., 2009. Threatened fishes of the world: *Gobiocypris rarus* Ye and Fu, 1983 (Cyprinidae). Environ. Biol. Fish. 86, 107–108.
Ye, M.R., Fu, T.Y., 1983. Description of a new genus and species of Danioninae from China (Cypriniformes: Cyprinidae). Acta Zootaxon. Sin. 8 (4), 434–437 (in Chinese).
Yuan, J.J., Tao, X.F., 2008. The features of gravel bed and drainage evolution in the Qingyi River valley in the Mingshan-Danling Region, Sichuan. Acta Geol. Sin. 28 (1), 6–12 (in Chinese).
Yue, P.Q., Chen, Y.Y., 1998. China Red Data Book of Endangered Animals: Pisces. Science Press, Beijing, pp. 170–172 (in Chinese).
Zhong, X.P., Xu, Y., Liang, Y., Liao, T., Wang, J.W., 2005. The Chinese rare minnow (*Gobiocypris rarus*) as an in vivo model for endocrine disruption in freshwater teleosts: a full life-cycle test with diethylstilbestrol. Aquat. Toxicol. 71, 85–95.