High genetic diversity in cryptic populations of the migratory sutchi catfish *Pangasianodon hypophthalmus* in the Mekong River

N So[1,2], GE Maes[1] and FAM Volckaert[1]

[1]*Laboratory of Aquatic Ecology, Katholieke Universiteit Leuven, Ch. Deberiotstraat 32, Leuven B-3000, Belgium;* [2]*Department of Fisheries, Ministry of Agriculture, Forestry and Fisheries, 186, Norodom Blvd., Phnom Penh, Cambodia*

The detection and conservation of spawning units is of crucial importance in highly migratory species. The sutchi catfish *Pangasianodon hypophthalmus* (Pangasiidae; Teleostei) is a common large-sized tropical fish, which migrates annually to several upstream spawning sites on the Lower Mekong River and feeds on the huge floodplain of the Lower Mekong and Tonle Sap for the other half of the year. We hypothesised that because of the relative size of the feeding and spawning habitat, genetic variability would be high and homogeneous in foraging populations, but that spawning stocks would be distinct in space and time. To test these predictions, 567 individuals from 10 geographic locations separated by up to 1230 km along the Lower Mekong River were genotyped at seven microsatellite loci. The level of genetic diversity was much higher than other freshwater fish and reached values comparable to marine species (mean $H_e = 0.757$). All samples collected at the potential spawning sites deviated from Hardy–Weinberg expectations, suggesting admixture. Individual-based clustering methods revealed genetic heterogeneity and enabled the detection of three genetically distinct sympatric populations. There was no evidence of recent reduction in effective population size in any population. Contrasting with the vast extent of the feeding grounds, the shortage of spawning grounds seems to have moved sutchi catfish towards diachronous spawning. Hence the sustainable exploitation of this natural resource hinges on the conservation of the limited spawning grounds and open migration routes between the spawning and feeding grounds.
Heredity (2006) **96,** 166–174. doi:10.1038/sj.hdy.6800781; published online 21 December 2005

Introduction

Freshwater fish tend to show a lower genetic diversity than marine and anadromous species. In contrast, they typically exhibit greater among population genetic differentiation than their estuarine and marine conspecifics (Ward *et al*, 1994; DeWoody and Avise, 2000). Their postlarval life history can be separated into two phases: a temporally restricted spawning period and a long feeding period. Freshwater fish have constrained gene flow due to drainage structure over short to moderate evolutionary times. Hence, differences in the distribution of genetic variability between freshwater and marine fishes have been attributed to differences in population size and to geographical barriers (to gene flow) among freshwater localities, which isolate populations (Ward *et al*, 1994).

Large tropical rivers, such as the Mekong, Amazon and Congo, draw special attention; the huge discharge volumes, vast length, width and depth, and long geological evolution have imposed special adaptations on their biota. These features might mean that the constraints mentioned above might not be so critical for freshwater fish in these rivers. Consequently, genetic diversity, effective population size and structure may carry some of the signatures of marine populations. However, a critical difference from marine species might be the availability and extent of spawning areas. Several taxa in the Amazon (eg *Arapaima gigas* and pimelodid catfishes such as *Brachyplatystoma flavicans*; Barthem and Goulding, 1997) and Mekong (pangasiid catfishes and cyprinids; Poulsen and Valbo-Jørgensen, 2001; Baird and Flaherty, 2004) make vast migrations between the low to middle basin feeding grounds and upper basin spawning grounds, often located in upland middle-sized river branches. Similarly, the sutchi catfish, *Pangasianodon hypophthalmus* (Sauvage, 1878) (Pangasiidae, Teleostei), a long-distance migratory catfish occurring in large rivers of the Mekong and Chao Phraya basins, exhibits an annual pattern of migration between feeding and spawning grounds. It spawns in the upper stretches of the Cambodian Mekong River between the Khone Falls on the Cambodian/Lao border and the town of Kratie (Figure 1) from May to August (Poulsen and Valbo-Jørgensen, 2001). The discontinuous spawning habitat consists of rapids and sand banks interspersed with deep rocky channels and pools (Van Zalinge *et al*, 2002). It covers about 1% of the feeding area (52 500 km²), which is located in the huge floodplain of Tonle Sap, central and southern Cambodian lower Mekong and the Vietnamese Mekong delta (N So, personal observation). After
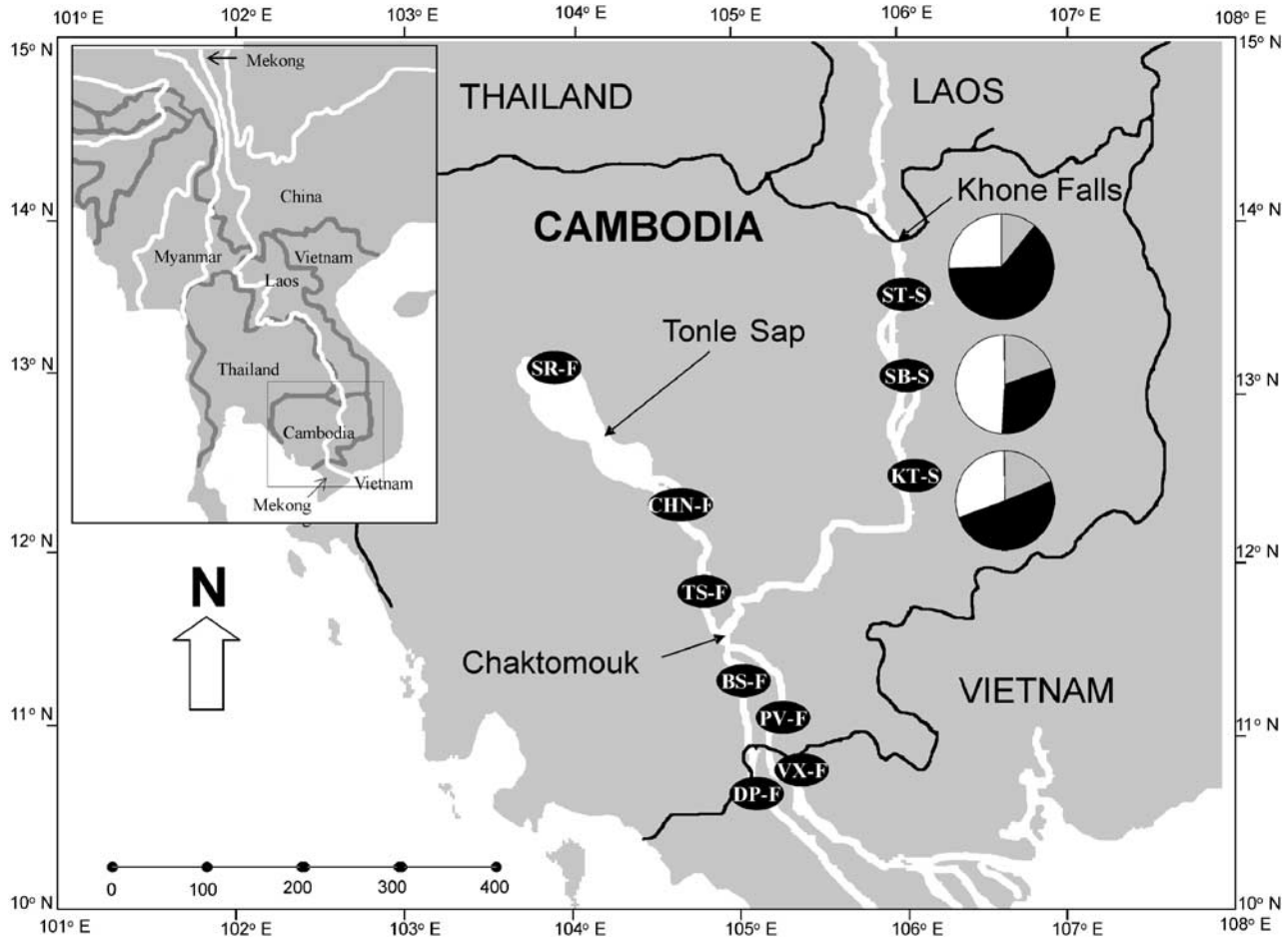
Figure 1 Sampling map of *P. hypophthalmus* in the Lower Mekong River basin showing sample locations at the spawning sites (ST-S, SB-S, KT-S). The newly defined populations are plotted in the frequency pies (grey: Population 1, black: Population 2; and white: Population 3). For sample abbreviations see Table 1.

spawning, adult fish migrate back to the feeding grounds (consisting of inundated forests and other vegetation) and larvae drift to the nursery floodplains located close to the feeding grounds. From this point of view, the river basin constitutes one ecological unit interconnecting upstream spawning habitats with downstream rearing habitats.

P. hypophthalmus is omnivorous and feeds on plants/fruits (vegetable debris), crustaceans and small fish (Rainboth, 1996). Fish mature at a minimal weight of 4 kg and total length of 54 cm in nature, at an age of 3 + years or more in captivity (Van Zalinge *et al*, 2002) and have an average fecundity of 200 000 eggs kg^{-1} (PV Khanh and NT Tung, personal communication). It is one of the largest catfish in the Mekong and can grow up to a maximal weight of 70 kg in nature (IG Baird, personal communication). It is very common, commercially important and intensively exploited in the Lower Mekong River and has been widely introduced for aquaculture in Southeast Asia (Roberts and Vidthayanon, 1991; Rainboth, 1996). Sutchi catfish contributed 10–15% to the total annual capture fisheries production of over 400 000 tons during 1994–1999 in Cambodia (So and Nao, 1999). The estimated census population size is approximately 20–30 × 10^6 (So and Nao, 1999). Several kinds of legal and illegal fishing gear (Van Zalinge *et al*, 2002) are

used to target the migratory sutchi catfish along more than 1000 km of the Lower Mekong River during the spawning migration period (September/October to April/May). The high fishing pressure combined with other anthropogenic activities such as habitat modification, deforestation, deterioration of water quality, water capture for hydropower and irrigation may put the genetic integrity of this species at risk.

In a preliminary study based on mitochondrial DNA-RFLP markers, sutchi catfish showed a moderate genetic diversity, with a suggestion of a historical population bottleneck, and a lack of population genetic structure (So *et al*, 2006). In this study, we performed a genetic diversity analysis with high-resolution microsatellite markers using larger sample sizes and covering its entire range in the Lower Mekong basin. Based on the relative size and distribution of the feeding and spawning habitats, we hypothesised that genetic variability would be high and homogeneous in foraging populations, but that spawning stocks would be genetically distinct in space and time. To test this (1) we analyse overall and population specific levels of genetic diversity in samples collected at the feeding and spawning sites; (2) we test for a recent reduction in effective population size, taking into account historical and contemporary features of the population to compare with a recent mitochondrial DNA

168

study; and (3) we search for the presence of cryptic populations spawning diachronically, which might be a consequence of size constraints on the spawning habitat.

Materials and methods

Samples

We collected tissue samples from the caudal fin of 567 *P. hypophthalmus* (Roberts and Vidthayanon, 1991) at 10 sites along an estimated 1230 km of the Mekong basin in Cambodia and Vietnam between January and August 2002 (Table 1 and Figure 1). The minimum sample size was 50 individuals per site, including three life history stages: adult mature breeders (>50 cm total length, at least 3 years old; 35.5% of the individuals), subadult immature fish (15–50 cm; 1–2 years old; 55.7%) and juvenile fish (<15 cm; 0+ years; 8.8% in the DP-F sample). Three samples originate from the putative spawning grounds (Stung Treng – ST-S, Sambo – SB-S and Kratie – KT-S) and seven samples from well-known feeding grounds (Siem Reap – SR-F, Kompong Chnang – CHN-F, Tonle Sap – TS-F, Prey Veng – PV-F, Bassac – BS-F, Da Phuoc – DP-F and Vinh Xuong – VX-F) (Van Zalinge *et al*, 2002). Adult mature fish collected at the spawning sites may represent various spawning aggregations; hence, subadult individuals were excluded during cluster analysis and subsequent population structure analysis. Similarly, as the feeding ground samples may contain mixtures of various spawning populations, they were only used to assess the genetic variability of foraging populations and to estimate the contribution of each discrete spawning population.

PCR amplification and microsatellite typing

About 100 mg of fin tissue was digested with proteinase K (10 mg ml^{-1}) in 200 µl of 10 × reaction buffer (100 mM (NH$_4$)$_2$SO$_4$, 670 mM Tris-HCl at pH = 8.8 and 0.1% Tween-20) and DTT buffer (1 M 1,4 dithio-DL-threitol and 0.08 M sodiumacetate at pH = 5.2). This solution was incubated at 55°C, vortexed after 3–4 h and left overnight. The digested solution was boiled for 10 min and after cooling (20–30 min) centrifuged at 13 000 rpm for 15 min.

We PCR amplified seven microsatellite loci with some modifications to the original protocols; three loci (Phy01-KUL, Phy03-KUL and Phy05-KUL) contained dinucleotide repeats and were isolated from *P. hypophthalmus* (Volckaert *et al*, 1999; GenBank accession numbers AJ131380-82); two loci (PSP-G 505 and PSP-G 579) contained dinucleotide repeats and two loci (PSP-G 509 and PSP-G 576) contained tetranucleotide repeats and were isolated from a genomic bank including three pangasiid catfishes, but not *P. hypophthalmus* (Hogan and May, 2002; GenBank accession numbers AF378275 and AF378285).

Each optimised PCR reaction consisted of 1 µl (10–100 ng) of total genomic DNA, 1 × reaction buffer (16 mM (NH$_4$)$_2$SO$_4$, 67 mM Tris-HCl, pH 8.8 at 25°C and 0.01% Tween-20), 200 µM dNTPs, MgCl$_2$ at a concentration of either 1 mM (Phy01-KUL, Phy03-KUL and Phy05-KUL), 1.25 mM (PSP-G 505, PSP-G 576 and PSP-G 579) or 1.75 mM (PSP-G 509), each of the forward and reverse primer at a concentration of either 0.2 µM (Phy01-KUL, Phy03-KUL, Phy05-KUL and PSP-G 576) or 0.4 µM (PSP-G 505, PSP-G 509 and PSP-G 579), 0.5 U of SilverStar *Taq* polymerase (Eurogentec, Seraing, Belgium) and deionised water up to either 10 µl (Phy01-KUL, Phy03-KUL and Phy05-KUL) or 25 µl total volume (PSP-G 505, PSP-G 509, PSP-G 576 and PSP-G 579). The amplification cycle of the three loci of Volckaert *et al* (1999) consisted of a 3 min denaturation cycle at 95°C followed by 25 cycles of a 95°C denaturation step for 30 s, an annealing step for 30 s (Phy01-KUL at 65°C, Phy03-KUL at 54°C and Phy05-KUL at 60°C) and a 72°C extension cycle for 30 s. Cycling concluded with a 10-min extension at 72°C. The optimised PCR conditions of the four loci of Hogan and May (2002) were as follows: a 'hot' start and a 2 min denaturation cycle at 94°C followed by 25 cycles of 94°C, a denaturation step for 30 s, an annealing step for 30 s (PSP-G 505 and PSP-G 509 at 63°C and a duplex of PSP-G 576/PSP-G 579 at 58°C) and a 72°C extension cycle for 1 min. Cycling concluded with a 5 min extension at 72°C in thermal cyclers (T1 thermocycler and Tgradient; Biometra, Göttingen, Germany) configured with heated lids. The forward primer for all PCRs was dye-labelled and the resulting amplified fragments were run on an automated sequencer (LI-COR, model 4200; Lincoln, NE, USA) with the appropriate size standards using the software E-seq v. 2.0 (LI-COR Inc., 2001). Products were scored using the software Gene ImagIR V. 4.03 (Scanalytics Inc., 2001,

Table 1 Collection sites (with abbreviation and coordinates), habitat type (including upstream and downstream migration cycle), sample date and size, and proportion of adult fish (breeders) of *P. hypophthalmus* surveyed in the Mekong River basin

General collection site	Abbr.	Coordinate	Habitat type (migration cycle)	Sample date	Sample size	Percentage of adults
Stung Treng, Cambodia	ST-S	13′32.43°N; 105′57.50°E	Spawn (down)	15-05-02	60	90
Kratie, Kratie, Cambodia	KT-S	12′37.00°N; 106′01.28°E	Spawn (down)	01-06-02	60	75
Sambo, Kratie, Cambodia	SB-S	13′06.00°N; 106′05.49°E	Spawn (down)	25-07-02	60	100
Siem Reap, Cambodia	SR-F	13′12.61°N; 103′51.79°E	Feed (up)	02-05-02	50	0
Kg. Chnang, Cambodia	CHN-F	12′22.52°N; 104′36.42°E	Feed (up)	15-03-02	50	0
Tonle Sap, Cambodia	TS-F	11′36.88°N; 104′55.47°E	Feed (up)	01-01-02	60	90
Prey Veng, Cambodia	PV-F	11′07.88°N; 105′14.35°E	Feed (up)	10-04-02	59	0
Bassac, Cambodia	BS-F	11′24.08°N; 105′00.45°E	Feed (up)	29-05-02	58	0
Da Phuoc, Vietnam	DP-F	10′43.16°N; 105′07.48°E	Feed (up)	15-08-02	50	0
Vinh Xuong, Vietnam	VX-F	10′54.45°N; 105′11.54°E	Feed (up)	27-05-02	60	0
Total					567	35.5

Migration cycle: down = downstream migration to feeding grounds; up = upstream migration to spawning grounds; S = spawning ground samples; F = feeding ground samples.

Fairfax, VA, USA). All microsatellite fragment sizes were determined to the nearest 0.4 base pairs (bp) and rounded to the nearest whole repeat number.

Statistical analysis

Genetic diversity was characterised by the observed heterozygosity (H_o), expected heterozygosity (H_e), range of allele sizes (R) and the number of unique alleles (uA) per locus in each sample and over all samples, and per sample over all loci using GENETIX v. 4.01 (Belkhir, 2000). Allelic richness (Ar) was calculated using FSTAT v. 2.9.3 (Goudet, 2001). Allelic richness, rather than number of alleles, was computed to allow for a comparison among samples of different sizes (El Mousadik and Petit, 1996). Differences in Ar and H_e among spawning and foraging samples were tested. For each group, the significance of any difference between weighted group averages of the chosen statistic was tested by 1000 iterations of a permutation scheme implemented in FSTAT.

The observed genotype and allele frequencies were tested for consistency with Hardy–Weinberg equilibrium (HWE) and linkage equilibrium expectation using randomisation tests (ie a permutation method with 5000 iterations estimating levels of significance) implemented in GENETIX. F_{IS}, the inbreeding coefficient within samples, values were computed using GENETIX to measure a potential deficit of heterozygotes caused by inbreeding, null alleles and/or population substructuring (Wahlund effect).

Levels of genetic differentiation were quantified over all spawning samples and for each pairwise sample comparison using Wright's F_{ST} statistic estimated by θ according to Weir and Cockerham (1984) as implemented in GENETIX. The significance levels were computed using a permutation approach (1000 iterations) in GENETIX. The program STRUCTURE v. 2.0 (Pritchard *et al*, 2000) was used to estimate the most likely number of populations represented by all original samples as well as only the samples collected at the spawning sites. STRUCTURE uses a model-based full Bayesian, Markov chain Monte Carlo approach that clusters individuals to minimize Hardy–Weinberg disequilibrium and gametic phase disequilibrium between loci within groups. The number of populations represented in our samples was estimated by pooling all samples (ie all 10 spawning and feeding samples or only the three spawning samples) and calculating the probability of the data, assuming that they originated from one to 10 populations ($K = 1\text{--}10$) in the study area, as described in Pritchard *et al* (2000). Each run consisted of a burn-in period of 2×10^4 steps followed by 10^5 MCMC iterations, and this replicated three times per K. Assignment scores of each individual to the most likely cluster were then analysed. All three samples collected at the spawning grounds (ST-S, SB-S and KT-S) were assigned to the redefined populations (ie baseline samples) using individual-based assignment test as implemented in STRUCTURE.

Furthermore, we recalculated genetic diversity (ie Ar, uA, H_e and H_o) and F_{IS} values within as well as pairwise F_{ST} values between the reassigned populations. Bottlenecked populations may exhibit gametic disequilibrium (Waples, 2002), reduced genetic diversity (particularly reduced allelic diversity, loss of rare or unique alleles;

Allendorf, 1986), increased heterozygosity relative to that expected at mutation-drift equilibrium (Cornuet and Luikart, 1996) and a reduced value of the *M*-statistic (Garza and Williamson, 2001). The former two genetic parameters have been tested as described above and the latter two were tested as follows. Heterozygote excess relative to a nonbottlenecked population in mutation-drift equilibrium having the same number of alleles was tested using BOTTLENECK v. 1.2.02 (Cornuet and Luikart, 1996). As recommended by Piry *et al* (1999), we assumed the two-phase mutation model for this analysis with 95% contribution from the stepwise mutation model (or strict single-step mutations) and 5% multiple-step mutations. Significance of heterozygote excess over all loci was assessed with a Wilcoxon signed-rank test (Cornuet and Luikart, 1996; Piry *et al*, 1999). Furthermore, the *M*-statistic values were calculated according to Garza and Williamson (2001) to detect recent reductions in effective population size. The method does not rely on any mutation model or population size assumption because $M = k/r$ simply measures the ratio of the number of alleles k to the range in allele size r across microsatellite loci. When a population is reduced in size, genetic drift is enhanced and alleles will eventually be lost. However, because the loss of any allele will contribute to a reduction in k, but only a loss of the largest or the smallest allele will contribute to a reduction in r, k is expected to decrease more rapidly than r (Garza and Williamson, 2001). Thus, the M value is smaller in recently reduced populations than in equilibrium populations. Significance was assessed by comparison of the mean value M across all loci and the more conservative rule that $M < 0.680$ suggests a bottleneck effect (Garza and Williamson, 2001).

In all cases, significance levels were adjusted for multiple comparisons using a conservative sequential Bonferroni correction (Rice, 1989).

Results

Genetic diversity within samples

All seven loci surveyed were highly polymorphic; a total of 114 alleles were found in *P. hypophthalmus* (mean 16.3). On average, the spawning grounds (ST-S, SB-S and KT-S) (Ar = 10.0) displayed a significantly ($P < 0.05$) higher level of allelic richness than the feeding grounds (SR-F, CHN-F, TS-F, PV-F, VX-F, BS-F and DP-F) (Ar = 9.0) (Table 2, Appendix). Overall, the spawning grounds (uA = 2.2, $H_e = 0.758$) and the feeding grounds (uA = 1.8, $H_e = 0.756$) showed equivalent amounts of unique alleles and levels of expected heterozygosities.

HWE and linkage disequilibrium

Randomisation tests showed that 15 (21%) of the 70 single-locus permutation tests were inconsistent with HWE expectations (Appendix A). In all, 13 tests showed deficits of heterozygotes: three at locus Phy01-KUL, four at Phy03-KUL, one at Phy05-KUL, two at PSP-G 505 and three at PSP-G 579 (Appendix A). Loci PSP-G 509 and PSP-G 579 each showed excess of heterozygotes in one sample, while locus PSP-G 576 was in agreement with HWE expectation. When results were combined across

Table 2 *P. hypophthalmus*. Population-specific microsatellite diversity, probability values for heterozygote excess (*P*) according to Cornuet and Luikart (1996) and mean *M*-statistic values (*M*) according to Garza and Williamson (2001)

Sample/population	N	Ar	uA	H_e	H_o	r	F_{IS}	P	M
Original sample									
ST-S	57.3	9.8	2.1	0.752	0.720	0.018	0.051*	NA	NA
SB-S	52.7	10.4	2.1	0.765	0.725	0.023	0.059**	NA	NA
KT-S	47.0	9.9	2.3	0.757	0.710	0.026	0.085**	NA	NA
SR-F	49.6	9.0	2.4	0.764	0.729	0.020	0.053*	NA	NA
CHN-F	48.9	8.7	1.1	0.760	0.740	0.011	0.037*	NA	NA
TS-F	56.0	7.9	0.6	0.758	0.728	0.017	0.048*	NA	NA
PV-F	59.4	9.1	2.1	0.772	0.759	0.007	0.025	NA	NA
VX-F	59.9	9.0	2.0	0.743	0.733	0.006	0.022	NA	NA
BS-F	57.0	10.3	2.1	0.768	0.735	0.019	0.052*	NA	NA
DP-F	49.9	8.7	2.4	0.729	0.765	−0.021	−0.055*	NA	NA
Sample average	53.8	9.9	1.9	0.757	0.734	0.013	0.041***	NA	NA
Assigned (newly defined) population									
Pop 1	21.1	11.6	3.7	0.862	0.842	0.010	0.035*	0.711	0.791
Pop 2	68.9	7.1	1.7	0.715	0.731	−0.009	−0.024	0.981	0.774
Pop 3	51.0	6.9	0.9	0.741	0.740	0.000	0.011	0.852	0.746
Population average	47.0	8.5	2.1	0.773	0.771	0.001	0.028*	0.998[a]	0.809[a]

N = mean number of genotyped individuals; Ar = allelic richness; uA = mean number of unique alleles; H_e = mean expected heterozygosity; H_o = mean observed heterozygosity; r = frequency of null alleles; F_{IS} = inbreeding coefficient; NA, not applicable.
Significance levels: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
[a]Probability values for heterozygote excess according to Cornuet and Luikart (1996) and *M*-statistic values according to Garza and Williamson (2001) were obtained when all populations were pooled.

all loci, eight out of 10 samples were inconsistent with HWE (Table 2).

The occurrence of null alleles has been regularly recorded as a major explanation for the observed deficit of heterozygotes (discussed in Brookfield, 1996). The estimated frequency of null alleles (see Brookfield, 1996) per locus from each sample was $r > 0.05$ for loci Phy01-KUL, Phy03-KUL, Phy05-KUL, PSP-G 505 and PSP-G 579 (Appendix A), which corresponded to 6, 6, 0, 5 and 0 expected homozygote null alleles (3%), respectively (Van Oosterhout *et al*, 2004). Nevertheless, it is unlikely that five out of seven loci show null alleles simultaneously. Inbreeding is also an unlikely explanation for observed deviations from HWE because of high levels of genetic diversity at all sampling locations (Table 2) and a large census population size (eg $N > 20 \times 10^6$; So and Nao, 1999). A mixing of different populations (Wahlund effect) seems a more likely explanation and was further tested.

Linkage disequilibrium among pairs of loci was tested for 210 combinations over 10 samples. Of 210 pairwise tests between loci, 40 (19%) were in linkage disequilibrium. Population TS-F showed strong linkage disequilibrium (16 out of 21 significant deviations after Bonferroni correction). After pooling all populations, 17 out of 21 comparisons were highly significant (only four after Bonferroni correction). The Wahlund effect due to subpopulation admixture might explain such pattern; it is less likely that the loci are tightly linked.

Genetic differentiation among samples

The overall genetic differentiation (fixation index, F_{ST}) among all spawning samples was significant ($P < 0.05$, permutation test on 1000 replicates), although the value was very low ($F_{ST} = 0.0052$). The two loci, Phy03-KUL

Table 3 *P. hypophthalmus*. Pairwise matrix of F_{ST} values (Weir and Cockerham, 1984) between the original spawning samples (below diagonal) and the assigned (newly defined) populations (above diagonal), with associated levels of significance

	Pop 1	Pop 2	Pop 3	
ST-S	—	0.0751***	0.0589***	Pop 1
SB-S	0.0067*	—	0.0372***	Pop 2
KT-S	0.0006	0.0065*	—	Pop 3
	ST-S	SB-S	KT-S	

*$P < 0.05$; ***$P < 0.001$.
Sample abbreviations are as in Table 1.

($F_{ST} = 0.0061$, $P = 0.041$) and PSP-G 579 ($F_{ST} = 0.0177$, $P = 0.007$), were the principal contributors to the genetic differentiation among all 10 samples. In addition, levels of genetic differentiation between spawning sample pairs were very low, and only two out of the three pairs were statistically different after Bonferroni correction (Table 3). Considering the significant F_{IS} values within and very low F_{ST} values between spawning samples, there is evidence that the spawning samples may represent a mix of genetically distinct populations.

These observations are consistent with the results of the model-based clustering method, revealing that the spawning samples most likely consisted of several subgroups. The most probable number of populations present in our total sample (ie all spawning and foraging individuals) was three (posterior probability ≈1). Similarly, the highest posterior probability for the presence of three populations was also detected by exclusively analysing our 'spawning' sample. They conformed to HWE expectations after sequential Bonferroni correction

(Table 2). The three original samples captured from the upstream spawning grounds were composed of different proportions of the three newly defined (assigned) populations; Stung Treng was composed of 10.9% fish of Population 1, 63.6% of Population 2 and 25.5% of Population 3; Sambo 20.0% of Population 1, 30.9% of Population 2 and 49.1% of Population 3; and Kratie 19.2% of Population 1, 50.0% of Population 2 and 30.8% of Population 3 (Figure 1).

Overall, the original spawning samples and newly defined populations exhibited similar amounts of genetic diversity for all indices (ie A, Ar, uA and H_e) (Table 2), but pairwise genetic differentiation increased by a factor 10 (Table 3).

Bottleneck effects

All tests for heterozygote excess were nonsignificant (Table 2). When all samples were pooled, the *M*-statistic value ($M = 0.801 > 0.680$) also suggested no recent bottleneck or founder effect. In addition, the three reassigned populations (ie biological populations) did not show evidence of recent reductions in effective population size (Table 2).

Discussion

This study is unique in providing a comprehensive contemporaneous genetic variability survey of a tropical migratory river catfish using high-resolution genetic markers. Our aims were (1) to analyse population-based genetic diversity; (2) to test for a recent reduction in effective population size; and (3) to explore the presence of cryptic populations spawning diachronically. We demonstrate that sutchi catfish show a very high genetic diversity, that spawning populations can be separated into cryptic populations, that population genetic diversity is not reduced and that microsatellite-based results differ from mitochondrial haplotype results.

High genetic diversity of a river catfish in a biodiversity hotspot region

High levels of genetic diversity of sutchi catfish were detected with microsatellite DNA loci (mean $H_e = 0.757$) compared to mitochondrial DNA (mtDNA) haplotypes ($h = 0.499$) (So *et al*, 2006). The discrepancy between both molecular markers can be partially explained by a weak correlation between mitochondrial and nuclear diversity, based on independent demographic processes imposed by historical events (eg Brunner *et al*, 1998; Bernatchez *et al*, 2002). A nonexclusive second explanation for these discrepancies may be related to different mutation rates. They are expected to be higher in microsatellite DNA (eg Moritz *et al*, 1987; Weber and Wong, 1993), promoting the establishment of a higher effective number of alleles at mutation-drift equilibrium for a given population size.

The average expected heterozygosity at seven microsatellite DNA loci is not concordant with the lower mean heterozygosity (0.540) reported in most freshwater fish species (13 species; DeWoody and Avise, 2000; but see, Waters *et al*, 2000 with $H_o = 0.78$ for *Alosa sapidissima*) and much higher than in another Southeast Asian catfish species (Usmani *et al*, 2003). Interestingly, the heterozygosity value of sutchi catfish even approaches the average value of many marine fishes (12 species; DeWoody and Avise, 2000; king mackerel, Broughton *et al*, 2002; plaice, Hoarau *et al*, 2002; Atlantic herring, McPherson *et al*, 2003; Atlantic cod, Knutsen *et al*, 2003). The above average level of heterozygosity of sutchi catfish might mirror a stable Holocene history (Rainboth, 1996) combined with the huge ecosystem, which hosts vast populations susceptible to a high gene flow in the absence of any natural and physical barrier. The huge and ecologically rich floodplain of the Lower Mekong basin harbours vast feeding areas (52 500 km²) that support a huge annual catch (So and Nao, 1999). In contrast, the shortage of spawning area (600 km²) might be an important factor constraining the effective population size and influencing the presence of spatially and/or temporally distinct spawning populations.

Is there evidence for cryptic population structure in the Lower Mekong basin?

Strong heterozygote deficiencies have been reported for many fish species (Waldman and McKinnon, 1993; Castric *et al*, 2001; Hoarau *et al*, 2002). They may be due to the occurrence of null alleles, inbreeding and/or population admixture (Wahlund effect). As heterozygote deficiency was similar at each locus and as the population size of sutchi catfish was considered to be high, we believe that null alleles and inbreeding are unlikely to cause a heterozygote deficiency at several loci simultaneously. Therefore, another explanation must be sought, that is, a Wahlund effect. The three samples collected at the spawning sites (ST-S, SB-S and KT-S) showed heterozygote deficiencies, implying that they could be composed of a mixture of sutchi catfish from differentiated gene pools. Such Wahlund effect has been reported in the brook charr *Salvelinus fontinalis* (Castric *et al*, 2001), where sympatric forms, reinforced by secondary contact between genetically and morphologically differentiated glacial races, may coexist. Dodson *et al* (1995) and McConnell (2004) reported extensive admixture of intraspecific genetic groups of other Southeast Asian freshwater fish such as the catfish *Hemibagrus nemurus* and the cyprinid *Barbodes gonionotus* during Pleistocene low sea levels. In addition, divergent haplotype groups have been detected in sutchi catfish (So *et al*, 2006); they are thought to relate to the Pleistocene climate fluctuations. Sympatric populations may appear in sutchi catfish as a consequence of vast contrasts in life-history specific habitat needs, that is, the reduced availability of spawning area and the huge feeding area. Moreover, mild genetic differentiation was detected among sutchi catfish; all pairwise F_{ST} estimates between original samples collected at the spawning sites were very low (0.0006–0.0067). This may be explained by the presence of cryptic population structure (eg Fillatre *et al*, 2003). To detect such cryptic population structure, we used an individual-based clustering method aiming at providing an answer to the above explanations as the sole source of deficits for all spawning samples (see below). We realise that our argument depends on the accuracy of the assignment and that there is no alternative proof of our findings at the moment.

We found strong evidence for sympatric population structure. Three newly defined populations with a very

high posterior probability were detected among the three original samples collected at the spawning sites. Such sympatric populations may reflect the differential annual spawning and feeding migrations of discrete groups, the nature and size of the spawning and feeding habitats, and the spawning time of the adults (Van Zalinge et al, 2002). Evidence from mitochondrial haplotype distribution and composition also pointed to heterogeneity within populations, although less obvious than at nuclear markers (So et al, 2006). Fillatre et al (2003) suggested that sockeye salmon (Oncorhynchus nerka) return to the Klukshu River, Yukon Canada into two genetically distinct runs to spawn in June–August (early run) and August–September (late run). Hendry and Day (2005) introduce the concept of 'isolation by time', where genetic differences between early and late producers point at least partially to restricted gene flow. They cite many cases in Pacific salmon, birds and plants where narrow-sense heritabilities for reproductive timing have been measured. Sutchi catfish are thus believed to migrate upstream as minimally three genetically distinct populations to spawn diachronically between May/June and July/August at three major (presumed) sites. Our samples thus seem to consist of incompletely assorted populations, as visible from the various proportions of each population.

Is there evidence for a reduced genetic diversity in sutchi catfish?

It is crucial to identify populations that have undergone ancient and recent bottlenecks, because they may have been affected by the small population size through demographic stochasticity, inbreeding or fixation of deleterious alleles, possibly leading to a reduced evolutionary potential and increased probability of extinction. The distribution profile of the microsatellite alleles, heterozygosities, heterozygosity excess (Cornuet and Luikart, 1996) and M-statistic values (Garza and Williamson, 2001) does not show evidence for recent declines in population size. Surprisingly, such evidence contradicts the hypothesis proposed by So et al, 2006) based on mtDNA-RFLP polymorphisms that sutchi catfish show indications of an ancient population bottleneck (ie during the Pleistocene when the course of the Mekong changed dramatically). More likely, the different outcomes may be attributed to the nature of the genetic markers used, which stresses the importance of limiting such inference to the genetic marker used. In general, mtDNA markers may be better in detecting ancient or historical bottlenecks than microsatellites (Koskinen et al, 2002; Avise, 2004). Such ancient declines will have little or no effect on the distribution of microsatellite genetic diversity, resulting from a more recent population expansion, probably during the mid-Holocene. Only 5720 years ago the Mekong basin evolved into its current configuration with the formation of the Tonle Sap river branch (Rainboth, 1996). Thus, both mtDNA-RFLP and microsatellite markers provide complementary, and not contradictory, information (Rogers, 1995). Although fisheries statistics (eg Van Zalinge et al, 2002) do not have the resolution to judge any historical evolution in fisheries mortality, the Cambodian capture fisheries ($N = 20$–30×10^6 for 1994–1999; So and Nao, 1999) fit with the findings

of this study. Although, the microsatellite data do not appear to detect recent and acute bottlenecks, this does not mean that the severity and timing of fishing and other anthropogenic influences have not yet depleted the stocks of sutchi catfish. Overfishing in vulnerable areas, especially at the few spawning sites, will lead to a reduction in effective population size and yield, and ultimately the loss of genetic diversity and population viability.

In conclusion, sutchi catfish are characterised by a substantial genetic diversity, possibly three sympatric populations and no detectable recent reductions in effective population size, despite heavy fishing pressure. Clearly, several issues remain to be addressed to fully understand the dynamics of sutchi catfish. They include a better understanding of the basic biology, population dynamics, ecology and population genetics in order to make the link between spawning and feeding grounds, and to evaluate inter- and intra-annual genetic variation (eg spawning waves and associated larval peaks). This should facilitate an analysis of adaptive traits in order to ensure the confident placement of populations into specific genetically structured hierarchies.

Acknowledgements

References

Allendorf FW (1986). Genetic drift and the loss of alleles versus heterozygosity. Zoo Biol 5: 181–190.

Avise JC (2004). Molecular Markers, Natural History and Evolution. Sinauer Associates: Sunderland, 684pp.

Baird IG, Flaherty MS (2004). Beyond national borders: important Mekong River medium sized migratory carps (Cyprinidae) and fisheries in Laos and Cambodia. Asian Fish Sci 17: 279–298.

Barthem R, Goulding M (1997). The Catfish Connection. Columbia University Press: New York, 144pp.

Belkhir K (2000). GENETIX, version 4.01. Laboratoire Genome, Populations, Interactions, CNRS UPR 9060: Montpellier, France.

Bernatchez L, Rhydderch JG, Kircheis FW (2002). Microsatellite gene diversity analysis in landlocked Arctic char from Maine. Trans Am Fish Soc 131: 1106–1118.

Brookfield JFY (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453–455.

Broughton RE, Stewart LB, Gold JR (2002). Microsatellite variation suggests substantial gene flow between king

mackerel (*Scomberomorus cavalla*) in the western Atlantic Ocean and Gulf of Mexico. *Fish Res* **54**: 305–316.

Brunner PC, Douglas MR, Bernatchez L (1998). Microsatellite and mitochondrial DNA assessment of population structure and stocking effect in Artic char *Salvelinus alpinus* (Teleostei: Salmonidae) from central alpine lakes. *Mol Ecol* **7**: 209–223.

Castric V, Bonney F, Bernatchez L (2001). Landscape structure and hierachical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* **55**: 1016–1028.

Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.

DeWoody JA, Avise JC (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol* **56**: 461–473.

Dodson JJ, Colombani F, Ng PKL (1995). Phylogeographic structure in mitochondrial DNA of a South-East Asian freshwater fish, *Hemibagrus nemurus* (Siluroidei, Bagridae) and Pleistocene sea-level changes on the Sunda shelf. *Mol Ecol* **4**: 331–346.

El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree (*Atgania spinosa* (L. Skeels)) endemic to Morocco. *Theor Appl Genet* **92**: 832–839.

Fillatre EK, Etherton P, Heath DD (2003). Bimodal run distribution in a northern population of sockeye salmon (*Oncorhynchus nerka*): life history and genetic analysis on a temporal scale. *Mol Ecol* **12**: 1793–1805.

Garza JC, Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**: 305–318.

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.

Hendry MA, Day T (2005). Population structure attributable to reproductive time: isolation by time and adaptation by time. *Mol Ecol* **14**: 901–906.

Hoarau G, Rijnsdorp AD, Van der Veer HW, Stam WT, Olsen JL (2002). Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. *Mol Ecol* **11**: 1165–1176.

Hogan ZS, May BP (2002). Twenty-seven new microsatellites for the migratory Asian catfish family Pangasiidae. *Mol Ecol Notes* **2**: 38–41.

Knutsen H, Jorde PE, Andre C, Stenseth NC (2003). Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol* **12**: 385–394.

Koskinen MT, Knizhin I, Primmer CR, Schlötterer C, Weiss S (2002). Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. grayling) provides evidence of ice-age mediated environmental perturbations in the world oldest body of fresh water, Lake Baikal. *Mol Ecol* **11**: 2599–2611.

McConnell SKJ (2004). Mapping aquatic faunal exchanges across the Sunda shelf, South-East Asia, using distributional and genetic data sets from the cyprinid fish *Barbodes gonionotus* (Bleeker, 1850). *J Nat Hist* **38**: 651–670.

McPherson AA, Stephenson RL, Taggart CT (2003). Genetically different Atlantic herring *Clupea harengus* spawning waves. *Mar Ecol Prog Ser* **247**: 303–309.

Moritz C, Dowling TE, Brown WM (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* **18**: 269–292.

Piry S, Luikart G, Cornuet JM (1999). Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* **90**: 502–503.

Poulsen AT, Valbo-Jørgensen J (2001). Fish migration and spawning habits in the Mekong mainstream: a survey using local knowledge (basin-wide). MRC Technical Report, CD ROM version 1. Mekong River Commission: Phnom Penh, Cambodia.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rainboth WJ (1996). *Fishes of the Cambodian Mekong*. FAO identification sheets for fishery purposes. Food and Agriculture Organization, Rome, 265pp.

Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Roberts TR, Vidthayanon C (1991). Systematic revision of the Asian catfish family Pangasiidae, with biological observations and descriptions of 3 new species. *Proc Acad Nat Sci Philos* **143**: 97–144.

Rogers AR (1995). Genetic evidence for a Pleistocene population explosion. *Evolution* **49**: 608–615.

So N, Nao T (1999). *National Aquaculture Development Review (1984–1999) and Aquaculture Development Plan (2000–2020)*. Department of Fisheries, Ministry of Agriculture, Forestry and Fisheries: Phnom Penh, Cambodia, 74pp.

So N, Van Houdt JKJ, Volckaert FAM (2006). Genetic diversity and population history of the migratory catfishes *Pangasianodon hypophthalmus* and *Pangasius bocourti* in the Cambodian Mekong River. *Fish Sci* **72** (in press).

Usmani S, Tan SG, Siraj SS, Yusoff K (2003). Population structure of the Southeast Asian river catfish *Mystus nemurus*. *Anim Genet* **34**: 462–464.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley PF (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.

Van Zalinge NP, Lieng S, Ngor PB, Heng K, Valbo-Jørgensen J (2002). Status of the Mekong *Pangasianodon hypophthalmus* recourses, with special reference to the stocks shared between Cambodia and Vietnam. MRC Technical Paper No. 1, Mekong River Commission, Phnom Penh. ISSN: 1683-1489, 29pp.

Volckaert FAM, Hellemans B, Pouyaud L (1999). Nine polymorphic microsatellite markers in the SE Asian catfishes *Pangasius hypophthalmus* and *Clarias batrachus*. *Anim Genet* **30**: 383–384.

Waldman B, McKinnon JS (1993). Inbreeding and outbreeding in fishes, amphibians and reptiles. In: Thornbel NW (ed) *The Natural History of Inbreeding and Outbreeding*. University of Chicago Press: Chicago. pp 250–282.

Waples RS (2002). Definition and estimation of effective population size in the conservation of endangered species. In: Beissinger SR, McCullough DR (eds) *Population Viability Analysis*. University of Chicago Press: Chicago. pp 147–168.

Ward RD, Woodwark M, Skibinski DOF (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *J Fish Biol* **44**: 213–232.

Waters JM, Epifanio JM, Gunter T, Brown BL (2000). Homing behaviour facilitates subtle genetic differentiation among river populations of *Alosa sapidissima*: microsatellites and mtDNA. *J Fish Biol* **56**: 622–636.

Weber JL, Wong C (1993). Mutation of human short tandem repeats. *Hum Mol Genet* **2**: 1123–1128.

Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

174

Appendix A

Summary statistics for seven microsatellite loci in ten sutchi catfish samples

Locus parameter	Collection site										
	ST-S	*SB-S*	*KT-S*	*SR-F*	*CHN-F*	*TS-F*	*PV-F*	*VX-F*	*BS-F*	*DP-F*	*Total*
Phy01-KUL											
N	60	56	55	50	49	57	59	60	60	49	555
R	144–170	146–172	160–170	156–170	156–170	154–184	154–170	144–170	152–174	156–180	144–184
Ar	7.6	8.2	5.5	7.0	6.0	7.8	6.7	6.6	8.9	7.6	7.5
uA	2	3	2	0	0	0	1	1	2	4	4
H_e	0.557	0.537	0.522	0.705	0.600	0.715	0.680	0.508	0.607	0.669	0.610
H_o	0.583	0.500	0.491	0.600	0.429	0.614	0.644	0.467	0.633	0.653	0.561
r	−0.017	0.024	0.020	0.062	0.107	0.059	0.022	0.027	−0.017	0.009	0.030
F_{IS}	−0.040	0.078	0.068	0.159*	0.296***	0.149*	0.062	0.090	−0.036	0.034	0.084***
Phy03-KUL											
N	55	55	47	50	49	52	59	60	59	50	536
R	1333–183	139–181	137–183	133–185	139–185	141–187	137–181	141–185	139–179	137–177	133–187
Ar	17.3	15.0	14.6	15.5	13.4	12.6	14.8	15.3	16.7	15.0	16.7
uA	7	4	4	8	4	2	7	5	4	5	12
H_e	0.903	0.883	0.896	0.843	0.846	0.893	0.837	0.851	0.913	0.846	0.871
H_o	0.800	0.818	0.851	0.800	0.796	0.731	0.898	0.867	0.797	0.900	0.826
r	0.054	0.034	0.024	0.023	0.027	0.086	−0.033	−0.008	0.061	−0.029	0.024
F_{IS}	0.123***	0.082*	0.061	0.061	0.069	0.191***	−0.065	−0.010	0.136**	−0.054	0.061***
Phy05-KUL											
N	57	42	46	50	49	58	58	60	60	50	530
R	194–204	194–202	190–208	194–204	194–202	194–204	194–202	1954–202	192–208	194–202	190–208
Ar	5.9	5.0	9.7	5.8	5.0	5.9	4.9	5.0	7.2	5.0	6.3
uA	0	1	3	1	0	0	0	0	1	0	3
H_e	0.704	0.674	0.775	0.709	0.698	0.724	0.681	0.729	0.700	0.664	0.706
H_o	0.597	0.691	0.717	0.720	0.796	0.655	0.638	0.750	0.650	0.740	0.695
r	0.063	−0.010	0.032	−0.006	−0.058	0.040	0.025	−0.012	0.029	−0.046	0.006
F_{IS}	0.161*	−0.012	0.085	−0.005	−0.130	0.103	0.071	−0.021	0.079	−0.105	0.029
PSP-G 505											
N	55	54	45	50	49	49	60	60	51	50	523
R	102–196	102–198	102–198	102–198	102–196	102–194	102–198	102–198	102–206	102–198	102–206
Ar	8.7	10.7	10.8	8.0	9.0	6.7	10.5	8.6	9.8	8.8	9.9
uA	1	1	1	0	0	2	1	1	1	1	3
H_e	0.713	0.797	0.722	0.773	0.798	0.680	0.792	0.744	0.754	0.757	0.753
H_o	0.655	0.778	0.578	0.740	0.776	0.694	0.733	0.667	0.765	0.780	0.716
r	0.034	0.011	0.084	0.019	0.013	−0.008	0.033	0.044	−0.006	−0.013	0.021
F_{IS}	0.092	0.033	0.21**	0.053	0.038	−0.011	0.082	0.112*	−0.004	−0.020	0.058**
PSP-G 509											
N	54	53	45	49	48	58	60	59	59	50	535
R	172–208	172–208	172–200	172–208	172–208	172–208	172–208	172–212	172–208	172–208	172–212
Ar	9.5	10.0	8.0	7.8	10.0	9.0	9.6	10.4	9.7	8.7	9.6
uA	2	0	0	1	0	0	3	2	0	2	4
H_e	0.836	0.840	0.820	0.787	0.846	0.851	0.847	0.852	0.839	0.782	0.830
H_o	0.889	0.943	0.867	0.755	0.833	0.793	0.867	0.797	0.780	0.820	0.834
r	−0.029	-0.056	−0.026	0.018	0.007	0.031	−0.011	0.030	0.032	−0.021	−0.002
F_{IS}	−0.054	−0.114**	−0.046	0.051	0.025	0.076	−0.015	0.074	0.079	−0.039	0.007
PSP-G 576											
N	60	52	44	49	49	59	60	60	57	50	540
R	153–201	153–197	153–197	153–197	153–189	153–193	153–197	153–201	153–197	153–189	153–201
Ar	11.3	9.8	11.9	9.7	9.9	8.6	10.4	11.3	11.4	9.7	10.5
uA	1	2	3	2	1	1	3	2	1	1	2
H_e	0.851	0.801	0.832	0.805	0.838	0.795	0.848	0.840	0.845	0.775	0.823
H_o	0.817	0.789	0.864	0.816	0.816	0.814	0.800	0.850	0.877	0.860	0.830
r	0.018	0.007	−0.017	−0.006	0.012	−0.011	0.026	−0.005	−0.018	−0.048	−0.004
F_{IS}	0.048	0.025	−0.027	−0.003	0.036	−0.016	0.065	−0.003	−0.029	−0.099	0.002
PSP-G 579											
N	60	57	47	49	49	59	60	60	53	50	544
R	191–227	189–261	169–213	167–207	167–207	191–205	167–207	167–207	169–207	187–207	167–261
Ar	8.2	14.0	8.7	9.4	7.6	5.0	7.1	6.1	8.5	6.5	9.0
uA	2	3	6	4	3	0	3	3	2	3	9
H_e	0.700	0.824	0.728	0.722	0.695	0.649	0.721	0.676	0.721	0.610	0.705
H_o	0.700	0.754	0.681	0.633	0.735	0.797	0.733	0.733	0.642	0.680	0.709
r	0.000	0.038	0.027	0.052	−0.024	−0.090	−0.007	−0.034	0.046	−0.044	−0.002
F_{IS}	0.009	0.094*	0.076	0.134*	−0.047	−0.220***	−0.008	−0.077	0.120*	−0.106	0.001

Information includes: number of fish genotyped (N), range of alleles (R) in base pair, allelic richness (Ar), number of unique alleles (uA), observed (H_o) and expected heterozygosity (H_e) (Nei, 1978), frequency of null alleles (r) and average inbreeding coefficient (F_{IS}) (Weir and Cockerham, 1984) for each locus in each sample. Sampling abbreviations are given in Table 1.
*$P < 0.05$, **$P < 0.01$; ***$P < 0.001$.