npg

Intron polymorphism (EPIC-PCR) reveals phylogeographic structure of *Zacco platypus* in China: a possible target for aquaculture development

P Berrebi[1], E Boissin[1], F Fang[2] and G Cattaneo-Berrebi[1]

[1]*Laboratoire Ecosystèmes Lagunaires, UMR 5119, cc093, University Montpellier 2, place Bataillon, 34095 Montpellier, cedex 05, France;*
[2]*Department of Vertebrate Zoology, Swedish Museum of Natural History, POB 50007, SE 104 05 Stockholm, Sweden*

The present study aims at a phylogeographic description of *Zacco platypus* from southeast China, in order to detect subdivisions within the nominal species. Two main basins were sampled: the Chang Jiang (Yangstze River) in central and east China (Hunan and Sichuan provinces) and the Xi Jiang, the more southern main tributary of the Zhu Jiang (Pearl River, Guangxi province). A total of 27 intron systems were tested, five of them were informative and gave 12 interpretable and polymorphic loci. Within the diversity of *Z. platypus*, four genetic groups were identified by multi-dimensional (FCA) analyses, corresponding to distinct genetic pools. The geographical distribution of the genetic groups corresponds neither with the drainage structure, nor the geographic distances between samples. It follows that isolation by distance and limited migration are insufficient to explain this geographic structure. The history of the river network therefore appears to have played an important role.
Heredity (2005) **94,** 589–598. doi:10.1038/sj.hdy.6800660
Published online 27 April 2005

Introduction

The general organisation of the Asian ichthyofauna has been investigated by various authors (Banarescu and Coad, 1991; Rainboth, 1991; Banarescu, 1992) and the Chinese ichthyofauna has recently been inventoried. The Chang Jiang (Yangtze River) basin harbours 361 species, of which 177 are endemic (Fu *et al*, 2003) and the Zhu Jiang (Pearl River) 133 freshwater fish species, of which about 33 are endemic (Anonymous, 1981). This ichthyo-fauna, considered to be very rich (Banarescu, 1992) and underestimated (Perdices *et al*, 2003), is threatened by hydrological alterations, mainly the Three Gorges Dam (and other dams) especially in the upper reaches. Overfishing has led to a decrease in fish size, mainly in lakes (Cao *et al*, 1991; Xie *et al*, 2001), and the biodiversity is thought to be decreasing (Fu *et al*, 2003). Faced with this danger, fish farming is a solution for biodiversity conservation (Brown and Day, 2002). For this strategy, the use of detailed methods for discriminating intraspe-cific geographic diversification is necessary. Here, we give an example of a phylogeographic survey of a possible candidate for domestication using methods adapted to difficulties in the field.

The family Cyprinidae is the most widespread and species-rich freshwater fish family in the world. Un-fortunately, very little is known about Cyprinidae in China. One of the most common and widespread genera in this family is *Zacco*, an Asian representative. Together with the genus *Opsariichthys*, it can be found from the Amur River in Russia to northern Vietnam (Banarescu and Coad, 1991). According to FISHBASE (Froese and Pauly, 2003), the species *Z. chengtui*, *Z. taiwanensis* and *Z. platypus* are also native to China.

The systematics within the genus *Zacco* is still rapidly changing and may be assisted by genetic studies. For instance, the well-known Japanese dark chub *Z. tem-minckii* has been divided into two taxa using genetic data: two mainly sympatric types described using 27 allozyme loci. Since no hybrids have been found at the 13 diagnostic loci, Okazaki *et al* (1991) proposed that the two types of dark chub are indeed two distinct species, *Z. temminckii* and *Z. sieboldi*.

Howes (1980) described significant morphological differences between the genera *Zacco* and *Opsariichthys*, and suggested that *Z. pachycephalus* is a member of *Opsariichthys* rather than *Zacco*. Ashiwa and Hosoya (1998) also supported Howes' conclusion. A new key for this genus has recently been produced for Japanese species (Hosoya *et al*, 2003). Biogeographical research has also been conducted in Taiwan on *Z. pachycephalus* using 26 allozymic loci. Several highly differentiated groups of populations were discovered, and attributed to separa-tion into different refuges during the most recent glacial period (Wang *et al*, 1999).

All these examples highlight the necessity for improv-ing the systematics of these species, especially before starting a domestication programme. A multidisciplinary

Correspondence: *P Berrebi, Laboratoire Ecosystèmes Lagunaires, UMR 5119, cc093, University Montpellier 2, place Bataillon, 34095 Montpellier, cedex 05, France. E-mail: berrebi@univ-montp2.fr*

group of Chinese and European scientists, funded by the European Community, have investigated a dozen target species that seem to be promising for fish farming in China. The genus *Zacco*, and especially the pale chub *Z. platypus*, is considered to be a target species. It is abundant and widespread in Chinese waters, but otherwise relatively little studied.

On the basis of morphology, Chinese populations of *Z. platypus* have been considered to form a single species, despite its huge range. However, Perdices *et al* (2003) presented a different picture in a recent paper written as part of the same scientific network and using most of the range sampled here. The surprising results, based on cytochrome *b* sequences, indicated that there are at least four undescribed species in this region. *Zacco A* occupies most of the Hunan-Guangxi regions, in the tributaries of two basins, the Chang Jiang and the Zhu Jiang catchments. *Zacco B* was recorded in the south-east of the Sichuan province. *Zacco C* is situated in the Li Shui river, in the Chang Jiang basin (not sampled in this survey). Finally, *Zacco D* is composed of two samples called here A25 and A37, in the southern tributaries of the Chang Jiang basin.

The aim of this study is the phylogeographic description of *Z. platypus* in south-west China in order to check this published interpretation, which parallels one or several in Japan and Taiwan in similar species complexes. The clear description of the genetic subdivisions will be of importance in a future programme assessing the comparative performance of populations and for constituting domestic strains. Nuclear markers are used in order to provide estimates of polymorphism, which are not given in the Perdices *et al* paper (2003) as it was based on only a few (3–11) individuals in each sample.

Since the nuclear DNA of this genus has never been analysed, introns were chosen, as many universal primers are available in the literature. The EPIC PCR (Exon-primed Intro-Crossed) technique was used. Otherwise, without any previous knowledge, allozymes would have been the only other markers able to give information on the nuclear genome without long and expensive preparation (ie genomic bank constitution). Unfortunately, keeping samples frozen in the field was impossible because the infrastructure was not available to consistently supply liquid nitrogen in southern China, and allozyme analysis was abandoned.

This survey can also be considered to test the capacities of a little-used nuclear marker class (introns) to describe geographic structure of a nearly unknown species.

Materials and methods

Sampling

Sampling trips were organised in three Chinese regions: Hunan Province in March–April 2001 (five sites, A25 to A39), Sichuan Province in August 2002 (three sites, B52 to B54) and Guangxi Province in March 2003 (sites C04 to C55b). The location of the sampling sites is given in Figure 1 and the characteristics of all sampling sites in Table 1. A total of 264 specimens of *Z. platypus* were collected from 15 localities. Another cyprinid species was used as an outgroup for tree polarisation: a sample of 16 specimens of *Opsariichthys bidens*. The two species share similar external morphological characters, and both have a vast distribution. However, osteological characters separate them well (Howes, 1980).

The collecting methods are shown in Table 1: when the sample was caught in the wild, it was considered as a part of a natural population; when the sample was
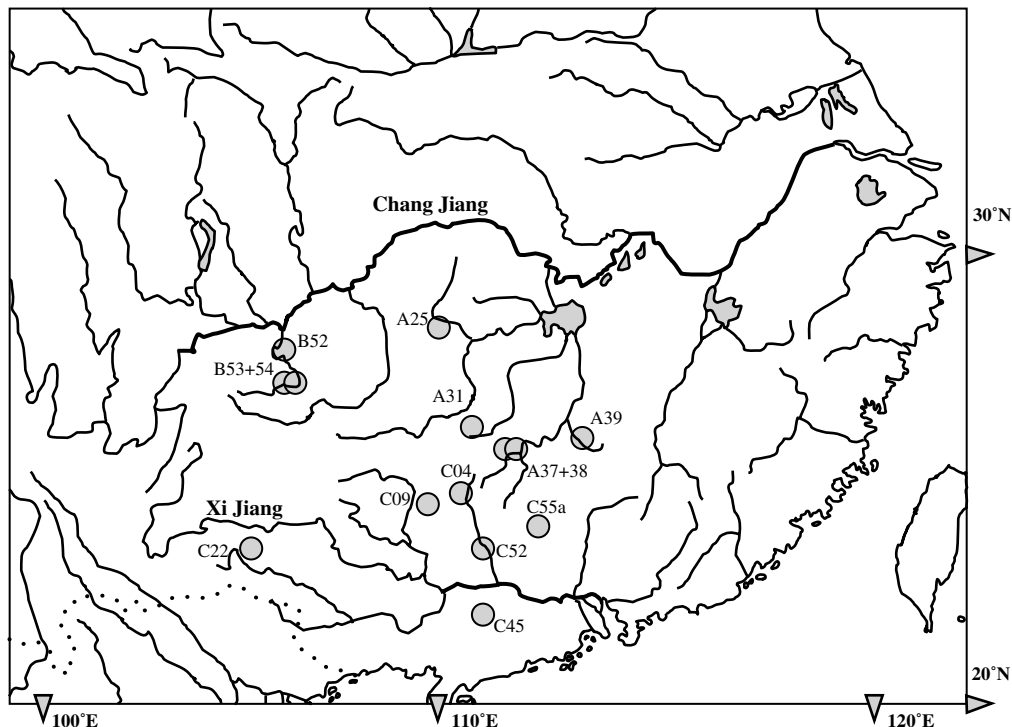

Figure 1 Map of sampling sites in central west and south China.

Table 1 Sampling details for *Z. platypus*

City/market	Basin	Stations EC number	Site codes	Coordinates	Sample sizes	Dates	Gear
Gantan He R.	Chang Jiang	EC-2002-006	A06	28°48′58″N; 111°28′56″E	21	18 March 2002	Seine
Near Baojing County	Chang Jiang	EC-2002-025	A25	28°44′45″N; 109°33′11″E	19	28 March 2002	Seine
Near Gulou Xiang	Chang Jiang	EC-2002-031	A31	27°04′33″N; 110°30′14″E	19	1 April 2002	Seine
Qiyang County market	Chang Jiang	EC-2002-037	A37	26°34′59″N; 111°50′59″E	5	7 April 2002	—
Near Baipaohui Cun	Chang Jiang	EC-2002-038	A38	26°16′08″N; 112°04′43″E	14	7 April 2002	Seine
At Pengtang Cun	Chang Jiang	EC-2002-039	A39	27°00′08″N; 113°01′16″E	12	8 April 2002	Seine
Near Xunchang	Chang Jiang	EC-2002-052	B52	28°28′53″N; 104°40′53″E	10	8 August 2002	Seine
Near Xinglong	Chang Jiang	EC-2002-053	B53	28°15′21″N; 105°21′34″E	12	8 August 2002	Cast net
Near Xuyong	Chang Jiang	EC-2002-054	B54	28°09′01″N; 105°24′54″E	31	9 August 2002	Hand net
Sanjie Zhen market	Xi Jiang	EC-2003-004	C04	25°28′12″N; 110°23′01″E	24	3 March 2003	—
Beijiang market	Xi Jiang	EC-2003-009	C09	24°59′02″N; 109°18′38″E	21	6 & 7 March 2003	—
At Baiyao Cun	Xi Jiang	EC-2003-022	C22	23°49′36″N; 106°58′40″E	29	11 March 2003	Seine
You Jiang tributary	Xi Jiang	EC-2003-045	C45	22°47′35″N; 110°36′07″E	28	21 March 2003	Seine
Zhaoping market	Xi Jiang	EC-2003-052	C52	24°10′41″N; 110°48′27″E	10	25 March 2003	—
Fuyang Jiang R.	Xi Jiang	EC-2003-055b	C55b	24°50′16″N; 111°15′58″E	30	26 March 2003	Electrofishing

In all, 16 further specimens of the species *O. bidens* were obtained from the stations A06 (10 fish), A37 (one), A38 (one) and A39 (four).

Table 2 Informative (polymorphic) intron system characteristics

System name	Intron position	Abbrev.	Primer name	Annealing temp.	Primer sequence	Reference
Actin 1	1	Act-1	PmActI-F	60°	CCATACCTTCTACAATGAGCTCCG	Bierne *et al* (2000)
			PmActI-R		GACCAGAGGCGTAGAGGAGAGC	
Actin 2	2	Act-2	Act-2-F	54°	GCTATAACCCTCGTAGATGGGCAC	Atarhouch *et al* (2003)
			Act-2-R		ATCTGGCACCACACCTTCTACAA	
Aldolase B	2	Aldo-B2	Aldo-B2-F	44°	TCAGGGCATTGTCGTCGG	Hassan *et al* (2002)
			Aldexon-3-R		CTGTTCCATTTAGACCAGC	
Aldolase C	1	Aldo-C1	Aldo-C1-F	58°	CCTGGCTGCGGACGAGTCTGTGGG	Atarhouch *et al* (2003)
			Aldo-C1-R		GGCGGTACTGTCTGCGGTTCTCC	
Myosin light chain B	2	Mlc-2	Mlc-2-F	54°	CTCGGGGTCCAGGACCTTGAA	Atarhouch *et al* (2003)
			Mlc-2-R		CTTCACCGTCGACCTCACCATG	

bought at a fish market, the exact origin was not certain because the fish could have come from more than one population.

Two main basins were sampled: the Chang Jiang (Yangstze River) in central and east China and the Xi Jiang, the more southerly main tributary of the Zhu Jiang (Pearl River). Specimens were identified morphologically in the field and identification was confirmed at the Swedish Museum of Natural History. All fish analysed were considered to belong to the species *Z. platypus* (except for the 16 specimens of *O. bidens*).

Molecular analyses

DNA was extracted from fin tissues using the phenol: chloroform method. PCR reactions were carried out in a total volume of 10 μl, containing 1 μl of 10× buffer (Promega), 2.5 mM $MgCl_2$ 0.2 mM of each dNTP (Invitrogen), 0.5 μM of each primer (MWG-Biotech AG, labelled with CY5 or Fluorescein), 0.3 U of *Taq* polymerase (Sigma) and 1 μl of DNA template (at about 150 μg/ml).

Thermo-cycling conditions in an Eppendorf Mastercycler® consisted of an initial denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at an appropriate temperature (see Table 2) for 1 min, extension at 72°C for 1 min 20 s, and a final extension at 72°C for 10 min.

These intragene sequences are highly polymorphic noncoding regions surrounded by relatively conserved exon sequences from which the primers have been designed. When these universal primers are used, several loci can be amplified, corresponding to duplica-

tions of the DNA through ancient polyploidisations, tandem repeats or pseudogenes (see Hassan *et al*, 2002; Atarhouch *et al*, 2003). Since several loci can be amplified by a single PCR reaction, it is necessary to verify that the primers actually annealed to their target DNA sequence. The appropriate conditions were assessed in a preliminary experiment using a thermal gradient of annealing temperatures produced by an Eppendorf Mastercycler Gradient® to determine the maximum annealing temperature allowing amplification. When several loci were amplified, we concluded that targeted sequences are the same and show a close match with the complementary universal primers. We call the set of loci amplified by the same pair of primers the 'system'. Only the length polymorphism of introns amplification was analysed (Figure 2), which means that a part of the variation, such as single base-pair substitutions, is not detected.

In total, 1 μl of PCR mixture from each fish was loaded onto an 8% denaturing polyacrylamide gel (Biorad). The PCR products were visualised with an FMBIO® fluorescent imaging system (Hitachi). Allele sizes were determined using a fluorescently labelled ladder of known size (Promega) with the FMBIO ANALYSIS 8.0 image analyser program.

Statistical analyses

Most statistical analyses were performed using the program GENETIX (Belkhir *et al*, 1998). This analysis was performed in four steps:

First, a multidimensional analysis was used to obtain the overall structure of the samples. A large part of the

Figure 2 Mlc2-c intron locus polymorphism showing several diagnostic alleles separating populations of geographic samples of the same nominal *Z. platypus* species.

distribution range of this species was sampled, so a genetic structure, due at least to isolation between basins, was expected. The best way to determine the overall structure is to keep the individual fish as the unit of information, irrespective of the sample origin. Among the various possible statistical methods, a multidimensional analysis, well suited to binary genotypic data, was used: Factorial Correspondence Analysis (FCA) first proposed by Benzécri (1973). The input matrix must be disjunct and the values discrete. The coding method has been described by She *et al* (1987). FCA does not make use of data describing the origin of the individuals and can, therefore, demonstrate the existence of unexpected taxa or subgroups, for example, a given sample can consist of several sympatric subgroups. Thus, this stage of analysis is a prerequisite for other calculations.

Second, depending on the samples and possible subsamples, standard parameters were estimated, that is, the allele frequencies, the observed (H_o) and unbiased expected heterozygosity (H_{nb}) of Nei (1978) and the F_{is} using the estimator f of Weir and Cockerham (1984).

Null alleles (no amplification) were observed at the Aldo-B2-A, Aldo-C1-A and Mlc-2-b loci. They are labelled '001' in Table 3. The first two loci showed intrasample polymorphism. Since heterozygote genotypes cannot be recognised (null/active heterozygotes appear as active/active homozygotes), we made a back-calculation assuming that the population is in Hardy–Weinberg equilibrium, which is not always the case (this must be considered as an approximation). Starting from the only certain genotype frequency (null/null homozygote), we estimated the null allele frequency to be the square root of the genotype frequency. The recalculated frequencies are indicated in bold characters in Table 3. Note that the three recalculated loci were not used in the calculation of the H and F_{is} estimations.

Third, to fully investigate the population structure, an AMOVA was performed to test the hierarchical stream structure, using only allele frequencies. For this, the ARLEQUIN software (Schneider *et al*, 2000) was used to describe the distribution of the genetic diversity among different levels of organisation: the basin (here two basins were sampled, the Chang Jiang basin whose samples have been labelled A, and the Xi Jiang basin labelled C) and the subbasins (two subbasins have been considered in the Chang Jiang basin, constituted, respectively, by the samples A25 + A31 and A37 + A38 + A39; four subbasins were considered in the Xi Jiang basin, constituted by the samples C55b, C04 + C52, C09 and C45).

F_{st} values were estimated between samples to illustrate the third and last level of differentiation. For this, the estimator θ of Weir and Cockerham (1984) was calculated and the significance estimated by resampling (5000 permutations) using GENETIX software. Since they had null alleles, the Aldo-B2-A, Aldo-C1-A and Mlc-2-b loci were not included.

Fourth, a phylogenetic tree was constructed based on the Nei's (1978) genetic distances between samples. This distance calculation takes into account the effects of small sample sizes. The Neighbour Joining (NJ) algorithm of the PHYLIP 3.5c software (Felsenstein, 1993) was implemented and trees were displayed with the TREEVIEW 1.40 program for tree construction (Page, 1996). The robustness of the branches was tested by the bootstrap method, using a data set built with the SEQBOOT program from PHYLIP.

Results

Subdivisions within the species *Z. platypus*

A total of 27 intron systems were tested. Table 2 shows the data for the five systems which were informative (ie including one or more polymorphic loci) and that gave easily interpretable and polymorphic patterns (Figure 2). In total, 12 polymorphic loci were scored and their allele frequencies calculated. The heterozygosity parameters given in Table 3 must be interpreted with caution, as they do not represent the estimated polymorphism of the species, but rather the relative polymorphism of samples.

The FCA analyses (Figure 3a and b) clearly discriminate intraspecific subdivisions. The initial overall analysis (14 *Z. platypus* samples, plus one *Opsariichthys bidens* outgroup sample) indicates that five discrete genetic groups can be identified, one of them being the outgroup (Figure 3a). The percentage of inertia explained by the three first axes are 19.6, 17.2 and 8.8%, respectively. Within the diversity of *Z. platypus*, four groups are isolated on the diagram, which correspond to distinct genetic pools, each showing intragroup similarities:

Group 1 contains 69% of the total sample. Its geographical distribution ranges from the south of the middle Chang Jiang tributaries up to the middle and low part of the Xi Jiang basin (samples A25, A31, A37 to 39, plus C04, C09, C45, C52 and C55a). No allele specificity has been detected for basins.

A close but distinct group (number 2) is constituted by sample C22 alone. This is the most western sample from the Xi Jiang and its limited representation is probably

Table 3 Allele frequencies measured in 12 intron loci. Allele 001 is a null allele (no amplification)

	A25	A31	A37	A38	A39	B52	B53	B54	C04	C09	C22	C45	C52	C55b	Opsa
Act-1-A	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**27**	**16**
500	0	0.13	0	0.04	0	1	0	0.00	0.73	1	0	0.14	0	0.17	1
510	1	0.55	1	0.93	0.71	0	0	0.10	0.25	0	1	0.86	1	0.76	0
520	0	0.32	0	0.04	0.29	0	1	0.68	0.02	0	0	0	0	0.07	0
530	0	0	0	0	0	0	0	0.23	0	0	0	0	0	0	0
Act-1-B	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**27**	**9**	**26**	**14**
660	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
680	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0
690	0	0	0	0	0	0	0	0	0.04	0.29	0	0.22	0.17	0	0
700	0.03	0.24	0	0.21	0	0	0	0	0.02	0	0	0.07	0	0.15	0
710	0	0	0	0	0	0	0	0	0	0	0	0	0	0.10	0
720	0.08	0	0	0	0	0	0	0	0.04	0.36	0.57	0	0	0	0
730	0.16	0.03	0	0	0	0	0	0	0.23	0.36	0.02	0.07	0	0	0.04
740	0.13	0.11	0	0.18	0	0	0.46	0.03	0.29	0	0.16	0.59	0.17	0	0.11
750	0	0.42	0.40	0.14	0.17	0	0.08	0.10	0.17	0	0.02	0.04	0.50	0.10	0.11
760	0.34	0.11	0.40	0.36	0.33	0.50	0.46	0.24	0.10	0	0.12	0	0.17	0.23	0.18
770	0.13	0.05	0.20	0.07	0.50	0.50	0	0.40	0	0	0.10	0	0	0.19	0.18
780	0.03	0.05	0	0.04	0	0	0	0.18	0.04	0	0.02	0	0	0.15	0.11
790	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0.02	0.14
800	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0.06	0.07
820	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04
830	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04
Act-2-A	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
370	0	0	0	0	0	0	0.04	0.55	0	0	0	0	0	0	0
380	0.16	0.11	0.10	0.07	0.17	0.15	0	0	0.06	0	0	0	0	0.03	0.44
390	0.84	0.89	0.90	0.93	0.83	0.10	0.96	0.45	0.94	1	1	1	1	0.97	0.13
400	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0
410	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0.44
Act-2-B	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
580	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
590	1	0.92	1	0.93	0.83	0	0	0.08	0.98	1	1	1	1	1	0
600	0	0.08	0	0.07	0.17	0	1	0.92	0.02	0	0	0	0	0	0
Aldo-B2-A	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**27**	**28**	**10**	**30**	**16**
001	**0.76**	0	**0.77**	**0.26**	0	0	0	0	0	0	0	0	0	0	0
250	0	0	0	0	0	0.85	0.29	0	0	0	0	0	0	0	0.47
270	**0.24**	1	**0.23**	**0.74**	1	0.15	0.58	0.90	1	1	0	1	1	1	0.53
300	0	0	0	0	0	0	0.13	0.10	0	0	1	0	0	0	0
Aldo-B2-B	**19**	**19**	**4**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
275	0.05	1	0.25	1	1	0	1	1	1	1	1	1	1	1	0
290	0.95	0	0.75	0	0	0	0	0	0	0	0	0	0	0	1
300	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Aldo-B2-C	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
210	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0
220	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1
Aldo-C1-A	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**10**	**16**
001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
550	1	1	1	1	1	0.85	0	0.05	1	1	0	1	1	1	0
570	0	0	0	0	0	0.15	1	0.95	0	0	1	0	0	0	0
Aldo-C1-B	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
001	**0.94**	0	**0.77**	0	0	1	1	1	**0.20**	**0.37**	1	**0.42**	**0.45**	**0.48**	1
600	**0.06**	1	**0.23**	1	1	0	0	0	**0.80**	**0.63**	0	**0.58**	**0.55**	**0.52**	0
Aldo-C1-C	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
185	1	1	1	0.96	1	0.85	0	0.05	1	1	1	1	1	1	1
190	0	0	0	0.04	0	0.15	1	0.95	0	0	0	0	0	0	0
Mlc-2-b	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**24**	**21**	**29**	**28**	**10**	**30**	**16**
001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
420	1	1	1	1	1	0	0	0	1	1	0	1	1	1	0
425	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0
Mlc-2-c	**19**	**19**	**5**	**14**	**12**	**10**	**12**	**31**	**23**	**21**	**29**	**28**	**10**	**30**	**16**
39	0.03	0	0.40	0	0.08	0	0	0	0.04	0	1	0	0	0	0

Table 3 Continued

	A25	*A31*	*A37*	*A38*	*A39*	*B52*	*B53*	*B54*	*C04*	*C09*	*C22*	*C45*	*C52*	*C55b*	*Opsa*
400	0.92	0.61	0.60	0.50	0.42	0	1	1	0.30	0	0	0	0	0.35	0
410	0.05	0.26	0	0.46	0.50	0	0	0	0.65	1	0	1	1	0.60	0
420	0	0.13	0	0.04	0	1	0	0	0	0	0	0	0	0.05	0
430	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
H_{nb}	**0.15**	**0.25**	**0.21**	**0.20**	**0.25**	**0.16**	**0.08**	**0.22**	**0.21**	**0.08**	**0.07**	**0.09**	**0.08**	**0.20**	**0.17**
H_o	0.12	0.18	0.07	0.17	0.10	0.10	0.06	0.12	0.11	0.04	0.07	0.02	0.02	0.13	0.16
Fishing	S	S	S	M	S	S	S	S	S	M	M	S	S	M	X
F_{is}	0.19	0.28	0.71	0.15	0.60	0.27	−0.06	0.39	0.47	0.45	0.03	0.74	0.70	0.35	—
P	**	***	**	NS	***	*	NS	***	***	**	NS	***	**	***	—

The first row of numbers, in italics, for all loci gives the sample size.
The frequencies in bold have been recalculated (see 'Results'). Unbiased (H_{nb}) and observed (H_o) heterozygosities were calculated according to Nei (1978). The sampling method is given (S: scientific net sampling; M: bought at the market; X: several samples mixed). The Weir and Cockerham (1984) estimator f of F_{is} was calculated and the significance (P) of its departure from the value zero obtained by 5000 permutations (NS, nonsignificant; *, significant at 5% level; **, 1%; ***, 0.1%).

due to the sampling range that was chosen for this project. One individual from group 1 (station A39) was close to this second group.

The third group is constituted by samples B53 and B54, from the upper Chang Jiang basin.

The fourth group, is composed of the B52 sample. Its position on axis 3 (not shown) is far towards the positive coordinates, whereas the other groups are near the origin of this third axis. Although genetically clearly distinct, groups 3 and 4 are geographically very close (less than 20 km), on the Chishui He subbasin.

Finally, the outgroup sample is clearly distinct.

Figure 3b represents another FCA involving group 1 alone. The three first axes account for 13.6, 9 and 7.8%, respectively, of the total inertia, indicating a weaker structure than in the former FCA. The first axis is preponderant, explaining the main significant structure detected: samples A25 and A37 form a subgroup, which is only slightly overlapping with the remaining individuals of this first group. The differentiation within group 1 is at a lower scale than the intergroup differentiation.

Several levels of organisation have been taken into account. (i) Analysing interbasins differentiation, with AMOVA, the genetic diversity resides mostly within samples (57.7%), rather than between samples of each basin (28.6%) or between basins (13.7%). Between basins, $F_{ct} = 0.13$ ($P < 0.001$ ***) indicates a significant differentiation. (ii) Analysing subbasin structure, no significant F_{ct} values have been observed in each basin (AMOVA). In Chang Jiang basin, between subbasins, $F_{ct} = −0.09$ ($P = 0.21$, not significant). In Xi Jiang basin, between subbasins, $F_{ct} = −0.22$ ($P = 0.60$, not significant). (iii) Intersamples differentiation has been tested through F_{st} calculations (Genetix software). High values in pair-wise comparisons probably reflect ascertainment bias due to selecting only polymorphic loci. All these values were considered to be significantly different from zero using permutation tests, except the three comparisons between samples A38, A39 and C55b, representing the most similar samples.

Phylogenetic structure

Genetic distances summarise the amount of differentiation among samples or taxa. The calculation is not based on all the observed loci but only on polymorphic ones. Consequently, the distances can only be used for this study and not compared to other examples in the literature.

The overall intergeneric Nei (1978) distances is $D = 1.18$. The intergroup distances range from $D = 0.57$ (groups 1 and 2) up to $D = 1.31$ (groups 1 and 4). Note that some intraspecific distances exceed the intergeneric distance.

To visualise the general organisation of the nominal species *Z. platypus*, an NJ tree was constructed using Nei (1978) distances. Figure 4 is in agreement with the FCA observations. The outgroup is well separated from the genus *Zacco*. The bootstrap values confirm the significance of most clusters. We consider nodes supported by more than 70% of bootstrap replicates to be highly robust (Zharkikh and Li, 1992; Hillis and Bull, 1993; Lecointre *et al*, 1994). However, we also consider groups that are supported by more than 50% of bootstrap replicates. In this way, the division *Opsariichthys* – group $1+2$ – group 3 – group 4 – is very robust. The division between groups 1 and 2 is less well supported, which is also reflected in their proximity in the FCA analysis.

Population characteristics

Table 3 shows the allele frequencies observed at the 12 polymorphic loci. A total of 52 alleles were scored. Null alleles were recorded at the Aldo-B2-A, Aldo-C1-A and Mlc-2-B loci. When recalculated, the frequencies are indicated in bold characters.

Unbiased heterozygosity ranges from 0.07 to 0.25. As only polymorphic loci were taken into account and also because no data are available on intron polymorphism in the literature, we do not know if these values are high. Of the previously defined genetic groups, group 1 is the most polymorphic with a mean $H_{nb} = 0.17$ (10 samples); group 2 $H_{nb} = 0.07$ (one sample); group 3 $H_{nb} = 0.15$ (2 samples) and group 4 $H_{nb} = 0.16$ (1 sample).

Within group 1, the Hunan samples ('A') are more polymorphic with $H_{nb} = 0.21$ while the Guangxi samples ('C') show an $H_{nb} = 0.13$. Within group 1, the diversity is similar between samples caught by scientists or bought at the market (0.18 and 0.16, respectively).

a

b
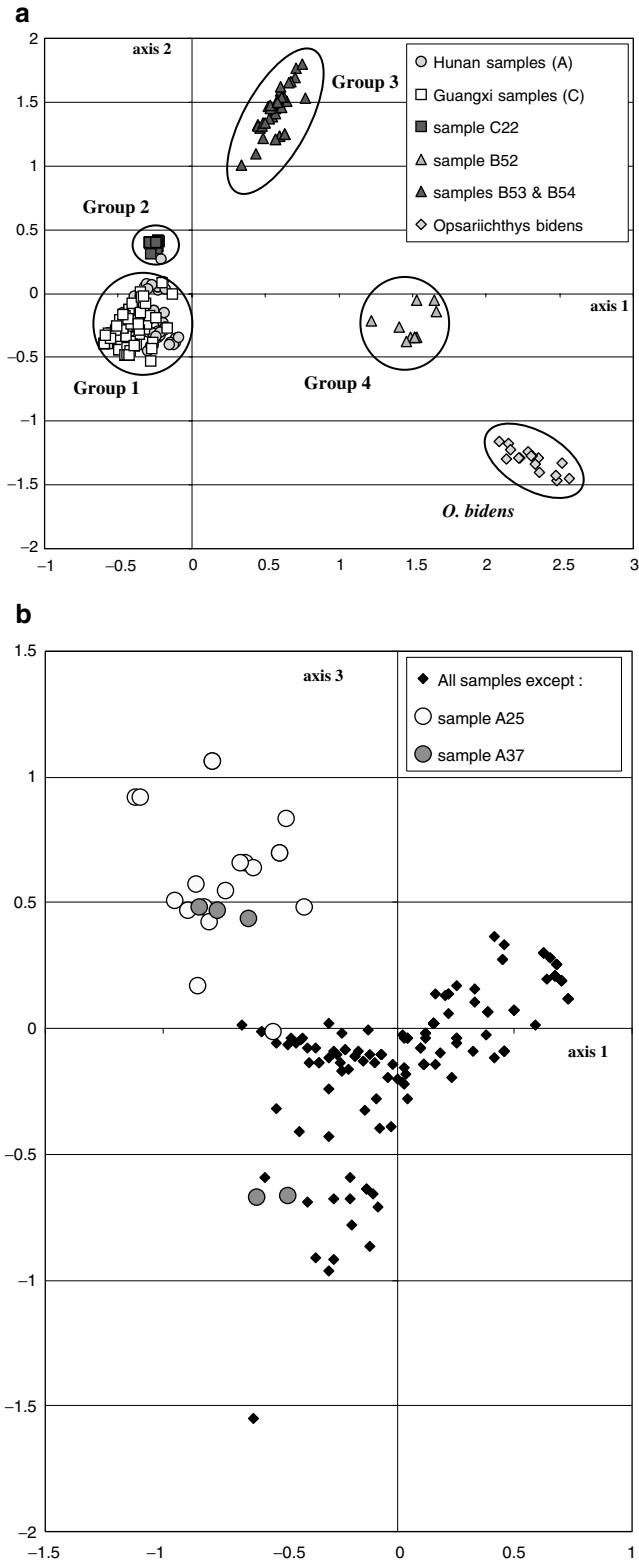
Figure 3 Factorial Correspondence Analysis plot of the first two axes. (**a**) Analysis of the whole sample; (**b**) analysis of group 1 only.
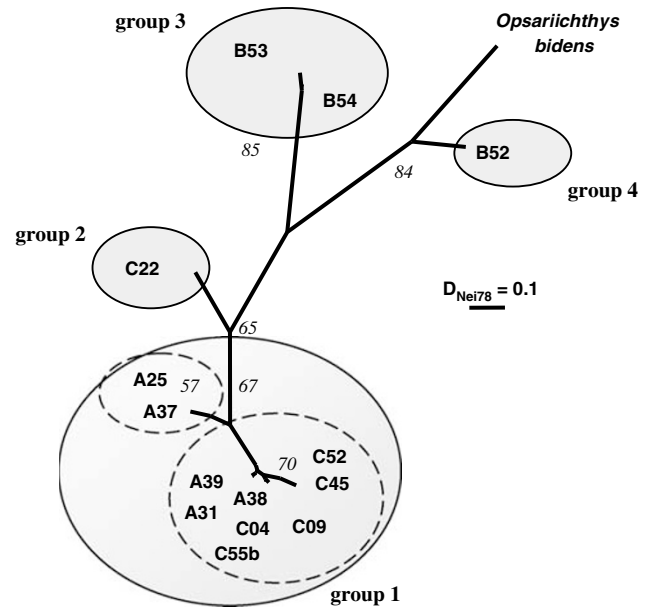
Figure 4 NJ tree based on Nei's (1978) distances between *Z. platypus* samples. *O. bidens* is used as outgroup. In italics are percentage bootstrap values.

fishing/market origin or the genetic group to which the fish belong.

Discussion

Phylogeographic structure of *Z. platypus*

A large part of the Chinese distribution of *Z. platypus* was sampled and analysed. The nominal species is divided into at least four genetic groups differentiated by several diagnostic alleles from the intron data. The main group, group 1, occurs in the central part of the sampling area, suggesting that the genetic groups considered as marginal in this study are probably well represented in other regions, to the west and north of the area sampled.

The second main finding concerns the distribution of group 1, which is well represented in the middle tributaries of the two sampled basins with little differentiation between them.

Analysis of molecular variance (AMOVA and F_{st}) indicated that most of the genetic variation in group 1 is among samples, although limited yet significant differentiation was detected between basins. No significant F_{ct} has been observed between subbasins. The extensive range of the group 1 taxon, covering an area of about 800 km and at least over two basins, suggests that the speciation occurred in a former riverine network linking the central part of both basins. The present isolation between samples is independent of distance, the differentiation is not significant between subbasins yet is significant between samples of the same subbasin. These patterns may indicate that gene flow was previously high within basins but has become limited, due to climatic change or human influence.

The overall picture does not show a clear differentiation between basins within group 1. It seems probable that differentiation by vicariance is rather slow in this species, unless the limited interbasins differentiation is

When analysing heterozygote deficiencies (F_{is}), which give an idea of population cohesion and panmixia, it is surprising to see that most samples are not in equilibrium. This general observation does not depend on

due to river captures which would have to have occurred recently. The genetic differentiation between the four groups appears to be rather ancient. However, more information on the drainage development of the three sampled provinces would be required for a more detailed explanation of the phylogeography of aquatic organisms in this region.

The geographic proximity of groups 3 and 4 has not led to a corresponding genetic proximity. This pattern suggests reproductive isolation, probably corresponding to species status. Group 1 shows a slight geographic differentiation. Group 2 (sample C22 only), to the west of the group 1 distribution, could also be considered to be a distant subgroup of the group 1. The NJ tree (and therefore the genetic distances) confirms the genetic proximity of groups 1 and 2. These two groups are, however, separated by two diagnostic loci: Aldo-B2-A and Aldo-C1-A, which is not the case for the differentiation withing group 1.

Perdices *et al* (2003) confirmed the existence of a main genetic group in the sampling area (called 'Zacco A' = our 'group 1') including, among others, our localities A31 and A39. Samples A25 and A37, which are weakly differentiated from the main group according to our nuclear markers (Figure 3b), were strongly divergent according to mtDNA data, forming the 'Zacco D' group (which is part of our 'group 1'). The more clear-cut results may be explained by the greater sensitivity to population isolation of mtDNA, because of its clonal transmission (Avise, 1994). The isolation of populations B53 and B54 was also confirmed by the mtDNA data, forming 'Zacco B' (= 'group 3' here). Perdices *et al* (2003) proposed giving species status to the four taxa identified in their study, given the presence of unique haplotypes in each lineage. We agree with this point of view, mainly because the intertaxon genetic distances (eg groups 1 and 3) are similar to the interspecies distances (*Zacco-Opsariichthys*), although genetic distance can suggest but not establish species status. A morphological analysis of the same fish is currently underway at the Swedish Museum of Natural History to decide the phylogenetic status of the *Zacco* lineages.

Evolutionary history of the *Z. platypus* species complex in China

The geographical distribution pattern of the *Z. platypus* genetic groups is not consistent with the present-day drainage structure. Moreover, the genetic differentiation does not correspond to the geographic distances between samples: genetically very similar samples have been observed at very long distances (samples A31 and C45 for example) and well-differentiated samples (possibly belonging to different species) have been sampled in closely adjacent localities (samples B52 and B53). These observations establish that limited migration and, therefore, isolation by distance, is only a partial structure explanation of the pattern of genetic differentiation.

The genetic structure could also have been shaped by the history of the river network. It is known that the river courses have changed several times in the region (Rainboth, 1991), and especially in the Chang Jiang basin (Xiao *et al*, 2001). Detailed reconstitution of the changes in surface topology and slopes in the recent and ancient past is necessary to understand the possible connections between basins. This type of reconstruction has been done, for example, to explain the history of a European cyprinid (Persat and Berrebi, 1990), but still needs to be done for China.

By comparing the genetic structure of other fish species in the same region, we can try to corroborate these observations. Another study has been conducted on a closely related cyprinid species, *O. bidens* (Berrebi *et al*, in press) sampled in Hunan and Guangxi, using introns. The overall structure of *O. bidens* is similar to that of *Z. platypus*, showing several independent lineages (probably species) generally distributed in different rivers. In *O. bidens* also, genetic structure and river networks are not concordant, indicating that similar forces shaped both species complexes. The *O. bidens* study detected a main taxon inhabiting the area around the junction between south-middle tributaries of Chang Jiang and north-middle tributaries of Xi Jiang basins. This is clearly the mark of an ancient single shared river network, which became divided into two basins after the *Z. platypus* and *O. bidens* main group speciation.

The use of introns in phylogeography

Introns are still infrequently used for population genetics and phylogeography. This kind of marker, the subject of recent publications (Dixon *et al*, 1996; Villablanca *et al*, 1998; Bierne *et al*, 2000; Daguin *et al*, 2001) is promising for intraspecific structure description; however, basic knowledge is still missing and several technical problems must still be solved.

In this study, we propose several improvements: the PCR thermal gradient is a prerequisite for distinguishing priming on target sequences from priming on random incomplete sequences. This method reduces the number of amplified loci, and the remainder are likely to be homologous. Homologous loci should comprise the intron loci included in the active gene and recently derived silent copies.

One of the questions that needs to be solved is the frequent deficit of heterozygotes (end of Table 3). This deficit has been observed in several nuclear markers such as microsatellites (Aurelle and Berrebi, 2002) and several hypotheses can be proposed to explain it. Selection, the Walhund effect, kin-structure and linkage are classical explanations, but the possibility that the deficit is an artefact must be considered first. One of the possible artefacts is intra-PCR competition. Larger alleles may not be detected because of their relatively inefficient amplification. We have found no published reference demonstrating such a phenomenon on introns, but we investigated this possiblity by plotting the F_{is} values (heterozygote deficiencies) against the range of allele sizes. Two tests were performed by calculating the Pearson correlation coefficient (i) between the F_{is} estimations and the allele distribution range (ie the difference between the largest and the shortest alleles) and (ii) between the F_{is} estimation and the number of alleles. Both coefficients were positive (0.263 and 0.274, respectively) but not significant. We, therefore, have no evidence of this kind of artefact.

Population characteristics and strain constitution

In addition to examining the population structure of *Z. platypus*, the relationships of lineages and their history,

this survey also had an applied objective. The strong structure observed within the nominal *Z. platypus* species leads us to suggest that at least four strains should be selected for performance comparisons.

Among the sampled populations of each taxon, only group 1 has been sufficiently sampled to provide alternative strains. Samples A31 and A39 are the most polymorphic and give the best security in terms of adaptability of the strain for rearing and stocking. Because the samples of the two sampled basins (Chang Jiang and Xi Jiang) differ slightly (as reflected in their mean polymorphism and the AMOVA), the comparison of performances should also include population C04 or C55b, which are good candidates with a heterozygosity of 0.20.

Sample A25 should also be tested because of its weak differentiation from group 1 and because mtDNA data indicated a strong and ancient divergence from group 1 (Perdices *et al*, 2003). Groups 2, 3 and 4 should be evaluated after the collection of more samples from neighbouring regions.

Acknowledgements

References

Anonymous (1981). *Freshwater Fishes of Guangxi Province*. Guangxi People's Press: Nanning.

Ashiwa H, Hosoya K (1998). Osteology of *Zacco pachycephalus*, sensu Jordan & Evermann (1903), with special reference to its systematic position. *Envir Biol Fish* **52**: 163–171.

Atarhouch T, Rami M, Cattaneo-Berrebi G, Ibanez C, Augros S, Boissin E *et al* (2003). New primers for EPIC amplification of intron sequences for fish and other vertebrate population genetic studies. *Biotechniques* **35**: 676–682.

Aurelle D, Berrebi P (2002). Natural and artificial secondary contact in brown trout (*Salmo trutta*, L.) in the French western Pyrenees assessed by allozymes and microsatellites. *Heredity* **89**: 171–183.

Avise JC (1994). *Molecular Markers, Natural History and Evolution*. Chapman & Hall: London. 511 pp.

Banarescu P (1992). *Distribution and Dispersal of Freshwater Animals in North America and Eurasia*. Aula-Verlag: Wiesbaden.

Banarescu P, Coad BW (1991). Cyprinids of Eurasia. In: Winfield IJ, Nelson JS (eds) *Cyprinid Fishes, Systematics, Biology and Exploitation*. Chapman & Hall: London. pp 127–155.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1998). GENETIX, logiciel sous WindowsTM pour la génétique des populations http://www.univ-montp2.fr/~genetix/genetix.htm Laboratoire Génome et Populations, CNRS UPR 9060, Université Montpellier II, Montpellier (France).

Benzécri JP (1973). *L'analyse des données*. Dunod: Paris.

Berrebi P, Rétif X, Fang F, Zhang CG. Population structure and systematics of *Opsariichthys bidens* (Osteichthyes: Cyprinidae) in southeast China using a new nuclear marker: the introns (EPIC-PCR). *Biol J Linn Soc* (in press).

Bierne N, Lehnert SA, Bédier E, Bonhomme F, Moore SS (2000). Screening for intron-length polymorphisms in peneid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Mol Ecol* **9**: 233–235.

Brown C, Day RL (2002). The future of stock enhancements: lessons for 239 hatchery practice from conservation biology. *Fish Fisher* **3**: 79–94.

Cao W, Zhang G, Ma J, Yu D (1991). Preliminary studies on the phenomenon of size diminution of the fish resources in Lake Honghu. In: Honghu Research Group, Academia Sinica (ed) *Studies on Comprehensive Exploitation of Aquatic Biological Productivity and Improvement of Ecological Environment in Lake Honghu*. Ocean Press: Beijing, China. pp 148–152 (in Chinese).

Daguin C, Bonhomme F, Borsa P (2001). The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity* **86**: 342–354.

Dixon B, Nagelkerke LAJ, Sibbing FA, Egberts E, Stet RJM (1996). Isolation of major histocompatibility complex class II beta sequences from lake Tana barbel (*Barbus intermedius* complex). *Folia Zool* **45**: 47–54.

Felsenstein J (1993). *PHYLIP 3.5 (Phylogeny Inference Package)*. University of Washington: Seattle, Washington, USA.

Froese R, Pauly D (Eds) (2003). *FishBase*. World Wide Web electronic publication. www.fishbase.org version 10 December 2003.

Fu C, Wu JH, Chen JK, Qu QH, Lei GC (2003). Freshwater fish biodiversity in the Yangtze River basin of China: patterns, threats and conservation. *Biodivers Conserv* **12**: 1649–1685.

Hassan M, Lemaire C, Fauvelot C, Bonhomme F (2002). Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. *Mol Ecol Notes* **2**: 334–340.

Hillis DM, Bull JJ (1993). An empirical test of bootstrapping as a method for accessing confidence in phylogenetic analysis. *Syst Bio* **42**: 182–192.

Hosoya K, Ashiwa H, Watanabe M, Mizuguchi K, Okazaki T (2003). *Zacco sieboldii*, a species distinct from *Zacco temminckii* (Cyprinidae). *Ichthyol Res* **50**: 1–8.

Howes GJ (1980). The anatomy and classification of bariliine cyprinid fishes. *Bull Br Mus Nat Hist (Zool)* **37**: 129–198.

Lecointre G, Philippe H, Lê HLV, Le Guyader H (1994). How many nucleotides are required to resolve a phylogenetic problem? The use of a new statistical method applicable to available sequences. *Mol Phylogenet Evol* **3**: 292–309.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Okazaki T, Watanabe M, Mizuguchi K, Hosoya K (1991). Genetic differentiation between 2 types of dark chub, *Zacco temmincki*, in Japan. *Jpn J Ichthyol* **38**: 133–140.

Page RDM (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357–358.

Perdices A, Cunha C, Coelho MM (2003). Phylogenetic structure of *Zacco platypus* (Teleostei, Cyprinidae) populations on the upper and middle Chang Jiang (=Yangtze) drainage inferred from cytochrome *b* sequences. *Mol Phylogenet Evol* **31**: 192–203.

Persat H, Berrebi P (1990). Relative ages of present populations of *Barbus barbus* and *Barbus meridionalis* (Cyprinidae) in southern France: preliminary considerations. *Aquat Living Resour* **3**: 253–263.

Rainboth WJ (1991). Cyprinids of south east Asia. In: Winfield IJ, Nelson JS (eds) *Cyprinid Fishes, Systematics, Biology and Exploitation*. Chapman & Hall: London. pp 156–210.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

She JX, Autem M, Kotoulas G, Pasteur N, Bonhomme F (1987). Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biol J Linn Soc* **32**: 357–371.

Villablanca F, Roderick G, Palumbi S (1998). Invasion genetics of the mediterranean fruit fly: variation in multiple nuclear introns. *Mol Ecol* **7**: 547–560.

Wang HY, Tsai MP, Yu MJ, Lee SC (1999). Influence of glaciation on divergence patterns of the endemic minnow, *Zacco pachycephalus*, in Taiwan. *Mol Ecol* **8**: 1879–1888.

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

Xiao W, Zhang Y, Liu H (2001). Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in east Asia. *Mol Phylogenet Evol* **18**: 163–173.

Xie S, Cui Y, Li Z (2001). Dietary–morphological relationships of fishes in Liangzi Lake, China. *J Fish Biol* **58**: 1714–1729.

Zharkikh A, Li WH (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. *Mol Biol Evol* **9**: 1119–1147.