Molecular Ecology (2005) **14**, 1939–1954 doi: 10.1111/j.1365-294X.2005.02569.x

Phylogeography of the cyprinid *Squalius aradensis* and implications for conservation of the endemic freshwater fauna of southern Portugal

N. MESQUITA,* B. HÄNFLING,† G. R. CARVALHO† and M. M. COELHO*
Universidade de Lisboa, Faculdade de Ciências, Departamento de Biologia Animal/Centro de Biologia Ambiental, Campo Grande, Bloco C2–3° Piso, 1749–016 Lisboa, Portugal, †Molecular Ecology & Fisheries Genetics Laboratory, Department of Biological Sciences, University of Hull, Hull HU6 7RX, UK

Abstract

The Iberian cyprinid fauna, characterized by the presence of numerous endemic species, has suffered from significant habitat degradation. The critically endangered *Squalius aradensis* is restricted to small drainages of southern Portugal, habitats that typically exhibit a characteristic Mediterranean-type heterogeneous hydrological system throughout the year, including alternation of flooding events during winter and complete drought in large river sections during summer. To assess the effect of historical and recent processes on genetic diversity in *S. aradensis* we examined within- and among-population variability in cytochrome *b* and six polymorphic microsatellite loci. Estimates of genetic diversity in time and space through the combined use of traditional Φ-/*F*-statistics, phylogenetic trees, ordination methods and nested clade analysis indicated significant and congruent structuring among populations. Data suggest that the Arade drainage represent the evolutionary centre of the species, with subsequent allopatric fragmentation across drainages. Factors other than isolation by distance strongly affected the within-drainage genetic differentiation observed in these Mediterranean-type drainages, including recent population expansion from a bottleneck event and restricted gene flow imposed by a long-term barrier (brackish water area). Significant correlation was found between *S. aradensis* allelic diversity and upstream drainage area. The relevance of findings for conservation issues is discussed in relation to local intermittent hydrological conditions, the highly restricted distribution and the critically endangered status of the species.

Keywords: endangered endemic Cyprinidae, gene flow, microsatellites, mitochondrial DNA, phylogeography, *Squalius aradensis*

Received 5 January 2005; revision accepted 1 March 2005

Introduction

The circum-Mediterranean region has been classified by Myers *et al*. (2000) as one of the 25 'biodiversity hotspots for conservation priorities', and as such, is characterized by the presence of an exceptional number of endemic species that are currently endangered by habitat loss. One such species is the endemic and recently described *Squalius aradensis* (Coelho *et al*. 1998), currently included in the Critically Endangered category of the IUCN Red List (Instituto da Conservação da Natureza unpublished data, http://www.icn.pt/documentos/livro_vermelho/Classif_Crit_Peixes.pdf). *Squalius aradensis* is found only in the small Mediterranean-type drainages of Portugal (SMTDP) located in the extreme southwestern region of the Iberian Peninsula (Fig. 1). These drainages, though representing a small hydrological area of the Iberian Peninsula, contain considerable ichthyofaunistic value, with the presence of local endemic species, typically exhibiting a highly restricted distribution (Mesquita & Coelho 2002). The habitats of the SMTDP are characterized by a heterogeneous annual hydrological cycle, typical of the circum-Mediterranean region. In these intermittent systems the occurrence of flooding events during the winter,

Correspondence: N. Mesquita, Fax: + 351 217500028; E-mail: npmesquita@fc.ul.pt

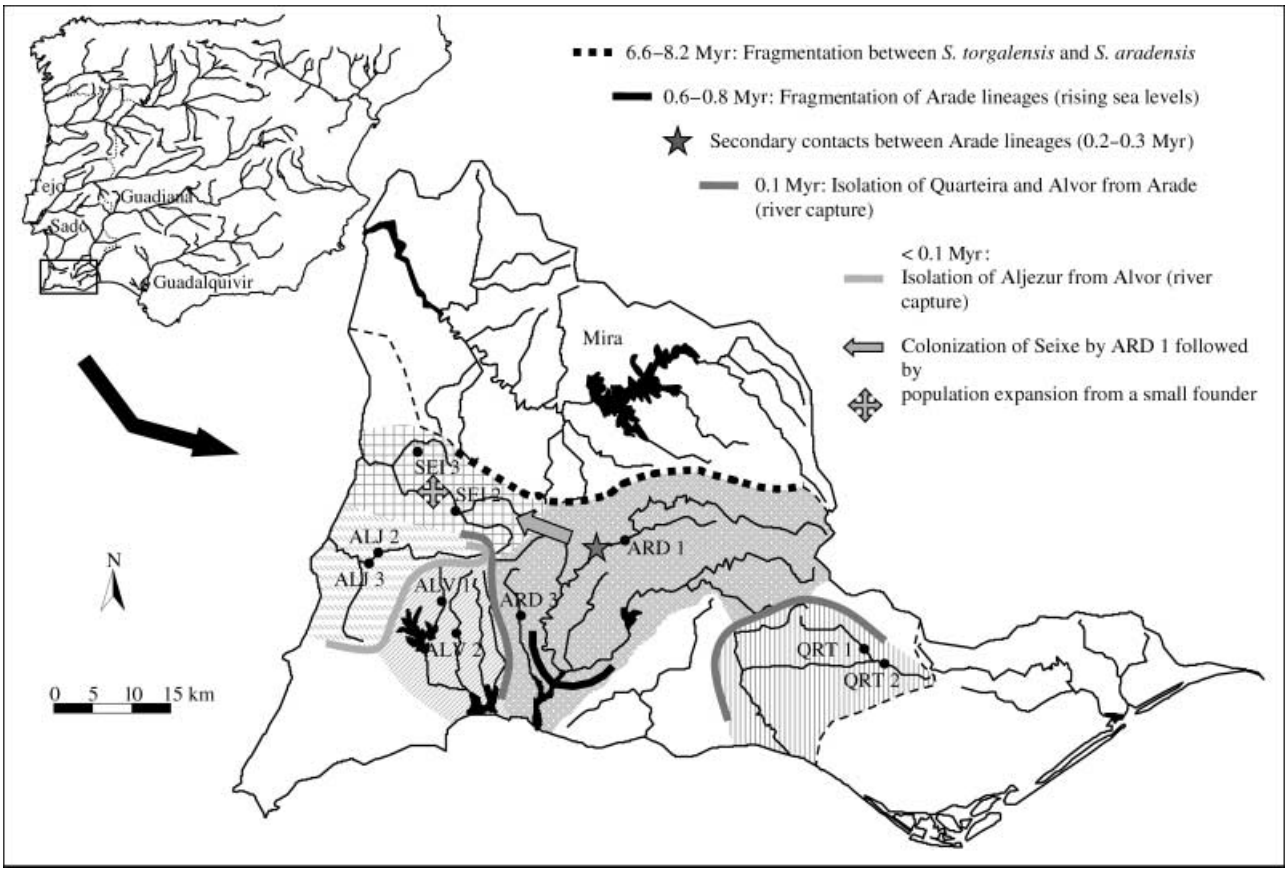

Fig. 1 Sampling sites location and *Squalius aradensis* distribution area (–––, distribution limits), and the historical and recent processes that determine the *S. aradensis* genetic structure (see Discussion for details). Sampled drainages: Seixe (SEI), Aljezur (ALJ), Alvor (ALV), Arade (ARD), Quarteira (QRT). Concluding scenario depicted from mitochondrial and nuclear data and the combination of Φ-/*F*-statistics, traditional phylogeography, ordination methods and NCA. Second possible scenario—fragmentation of Seixe from ARD3 (0.2–0.3 Myr) and recent invasion of ARD1 by Seixe individuals (postglacial, < 0.1 Myr). Differentiation of the sister species *S. torgalensis* and *S. aradensis*.

and in particular, the complete drying up of large river sections during summer, often results in a series of isolated pools, leading to periodically high fish mortality and subsequent populations bottlenecks (Coelho *et al*. 1995; Magalhães 2002). The unique biodiversity together with the intermittent hydrological conditions of these rivers makes it extremely important to preserve habitats that support distinct populations during the droughts, as well as implementing measures to prevent the dispersion of non-native species or populations (Magalhães 2002; Mesquita & Coelho 2002).

Although the primary goal in conservation of freshwater fish diversity necessitates ecological considerations, it is now well recognized that evolutionary processes at the population level are paramount (Moritz *et al*. 2002), including the estimation of within- and among-population genetic diversity (Avise 1995; Carvalho 1998; Hewitt 2001). Such considerations are of particular relevance for threatened species inhabiting heterogeneous habitats, as is the case of most endemic Iberian cyprinids (e.g. Mesquita *et al*. 2001; Salgueiro *et al*. 2003).

Previous genetic studies using both allozymes (Coelho *et al*. 1995, 1997) and mitochondrial DNA (mtDNA; Brito *et al*. 1997; Mesquita *et al*. 2001; Sanjur *et al*. 2003) have contributed to the detection of endemic cryptic taxa of the SMTDP. *Squalius* and *Chondrostoma* populations from the SMTDP, for example, represent distinct evolutionary lineages, with an origin estimated for *Squalius* in lower Pliocene–Messinian [3.6–7.2 million years (Myr); Brito *et al*. 1997; Sanjur *et al*. 2003]. Such molecular evidence has led to the formal description of three new species—*Squalius torgalensis* and *S. aradensis* (Coelho *et al*. (1998), and *Chondrostoma almacai* (Coelho *et al*., unpublished). Previous authors (Coelho *et al*. 1995, 1997; Mesquita *et al*. 2001) have suggested that such high levels of differentiation within a small geographical area may be explained by both historical factors, such as isolation arising from the formation and evolution of the drainage systems, and ecological factors, in particular, local hydrological conditions. However, to date there has been no study focusing on the relative roles of historical and contemporary population processes in shaping within-species

genetic structure that encompass several widely distributed drainage systems in the SMTDP.

Although analysis of mtDNA variability has become a favoured approach used for phylogeographical analysis at the intraspecific level (Avise 1995; Hewitt 2001), rapidly evolving microsatellites often reveal additional insights across recent timescales and among populations exhibiting shallow phylogeographical relationships (Angers & Bernatchez 1998; Hänfling *et al*. 2002). Particularly informative, however, has been the combined use of both mtDNA and microsatellites, allowing the separation of historical and contemporary population processes (e.g. Randi *et al*. 2003; Johnson *et al*. 2004).

Here we explore, using complementary analyses of mtDNA and microsatellite variation, the relative contribution of historical and recent processes in shaping genetic structure in *S. aradensis*. In particular, by examining patterns across several distinct drainages, we assess the degree of concordance in response of fragmented populations to the seasonal heterogeneity typical of SMTDP. Emergent patterns will be especially pertinent to the subsequent management and conservation of fish and associated freshwater fauna in the region.

Materials and methods

Sample sites and DNA extraction

A total of 585 individuals from 10 distinct populations (two from each drainage) were collected by electrofishing (March 2001; 55.9 mm mean fork length; specimens were returned to the stream) in the SMTDP, comprising the species' geographical range [Seixe, Aljezur, Alvor, Arade and Quarteira drainages; Fig. 1; labels as in Mesquita & Coelho (2002)]. Fin clips were preserved in 100% ethanol at 4 °C, and total genomic DNA was extracted following a protocol adapted from Miller *et al*. (1988).

mtDNA sequence analysis

Amplification and sequencing of the complete mtDNA cytochrome *b* (cyt *b*) gene, 1140 bp, were undertaken for a subset of individuals sampled: nine individuals from QRT2, 10 individuals from ALJ3 and 11 individuals from SEI2, SEI3, ALJ2, ALV1, ALV2, ARD1, ARD3 and QRT1 (Fig. 1). Polymerase chain reaction (PCR) conditions and purification of the double-stranded amplification products followed Mesquita *et al*. (2001). Purified amplification products were cycle sequenced in both directions using the PCR primers and screened on an ABI3700 or ABI3730xl DNA Analyser (Applied Biosystems) automatic DNA sequencer. Cyt *b* sequences (EMBL nos. AJ583056–AJ583086 and AJ852428–AJ852503) were aligned using SEQUENCHER version 4.1 (Gene Codes Corporation).

Within-population diversity was quantified by haplotype diversity (*h*; Nei 1987) and nucleotide diversity (π; Nei 1987) using REAP version 4.0 (McElroy *et al*. 1992). Geographic structure was investigated using three different routes of analyses. First, pairwise Φ_{ST} (Φ-statistics; Excoffier *et al*. 1992) values were computed using WINAMOVA version 1.55 (Excoffier *et al*. 1992) and considering the most suitable evolutionary model according to MODELTEST (Posada & Crandall 1998), in order to quantify the relative amount of among-population differentiation.

Second, phylogenetic tree-building algorithms were used, including the calculation of genetic distances among haplotypes. These analyses were performed to identify ancestral and derived haplotypes by outgroup rooting and to calculate divergence times among populations. In order to identify the most appropriate evolutionary model by Akaike information criterion (AIC), the program MODELTEST version 3.5 (Posada & Crandall 1998) was used. Subsequently, the most appropriate model was used to calculate pairwise genetic distances among haplotypes. Phylogenetic trees were constructed considering neighbour-joining (NJ) genetic distance and maximum-parsimony (MP) methods, and using PAUP version 4.0b10 (Swofford 2002). MP analysis was performed using an heuristic search with TBR branch-swapping algorithm and random stepwise sequence addition with 10 replicates. For both NJ and MP analysis the stability of the nodes was tested by 1000 bootstrap replicates. *Squalius torgalensis* (EMBL no. Z75929; Brito *et al*. 1997), the allopatric (Mira drainage restricted; Fig. 1) sister species of *Squalius aradensis* (Brito *et al*. 1997) was used as an outgroup. Divergence times between the major *S. aradensis* lineages were estimated using the molecular clock hypothesis. In order to test for homogeneity of nucleotide substitution rates, the two-cluster test and branch-length test were used according to Takezaki *et al*. (1995) with the computer program made available by the authors, LINTREE. Different but largely congruent cyt *b* molecular clock calibrations using fossil and geological data have been proposed for cyprinids (reviewed in Doadrio & Carmona 2004). Here we applied the most recent and commonly used (Doadrio & Carmona 2004) calibration of Dowling *et al*. (2002), considering the unconstrained and constrained (1.05% and 1.31% divergence per pairwise comparison per Myr, respectively) resulting time range and using the Tamura–Nei (*TN*; Tamura & Nei 1993) distance with gamma corrected for rate variation and 0.3 as alpha value (+ *G*). *TN* + *G* sequence pairwise values were calculated using PAUP version 4.0b10.

Third, network analysis and subsequent nested clade analysis (NCA; Templeton 1998 and references therein) were undertaken in order to test specific biogeographical hypotheses underlying observed haplotype distribution patterns. The estimation of the most parsimonious unrooted haplotype network was carried out using TCS version 1.18 (Clement *et al*. 2000). The resulting cladogram was converted

into a nested clade design following Templeton & Sing (1993). To test for significant associations between clades and geographical locations, the program GEODIS version 2.2 (Posada *et al*. 2000) was used to perform the nested contingency analysis (Templeton & Sing 1993) and the nested geographical distance analysis (Templeton 1998). The geographical measures clade distance (D_c), nested clade distance (D_n) and differences between the clade and nested clade distance of tips and interior clades of a nesting clade (I-T_c and I-T_n), and their statistical significance were calculated for 10 000 random permutations (Templeton 1998; Posada *et al*. 2000). The population pairwise geographical distance matrix included in the input file for GEODIS consisted of the shortest river course distances (km) as measured manually with a map card reader (MR300, Oregon Scientific) on 1 : 25 000 maps. For within-drainage distances the measurements were made along present river courses. For the among-drainage distances, in the absence of information on past connections such as fluvial captures and common ancestral drainage systems, assumptions were made that all neighbour headwaters had temporary connections and the closest one to each pair of sampling points was considered. For non-neighbour drainages, a stepping-stone model consisting of a sequence of connections between neighbour drainages was assumed. To assign the interior/tip status to the two nesting clades in the total cladogram, and to infer the most suitable population structure model and historical scenario for observed geographical associations, an upgraded version of the inference key for the nested haplotype tree analysis of geographical distances from Templeton (1998; 14.07. 2004, http://darwin.uvigo.es/) was used.

Microsatellite analysis

Overall, 585 individuals from 10 distinct *S. aradensis* populations were genotyped for seven polymorphic microsatellite loci, including five species-specific (SarN–primers, Appendix; Mesquita *et al*. 2003) and two cross-species (*Luxilus cornutus*; LCO–primers, Appendix and *LCO8*; Turner *et al*. 2004) sets of primers. PCR followed the conditions used by Mesquita *et al*. (2003). For loci *LCO4* and *LCO8*, 1.5 mM of $MgCl_2$ was used in the PCR, and optimal annealing temperatures were 58 °C and 60 °C, respectively. The cy5-labelled amplified fragments were separated on polyacrylamide gels using an ALFexpress™ (Amersham Pharmacia) automatic sequencer. At least four internal size standards, two of smaller and two of larger size than the alleles size range for each locus, were included and FRAGMENT MANAGER version 1.2 (Amersham Pharmacia) was used to determine allele length. In addition, and in order to make the allele-cross-reference between gels easier and more accurate, the amplification products from the same two specimens were included in all runs for each locus.

MICRO-CHECKER version 2.2 (Van Oosterhout *et al*. 2004), which tests for deviation of observed frequencies of allele size differences among and within genotypes against those expected using a Monte Carlo simulation method, was used to test for errors due to stuttering, large allele dropout and null alleles. Gametic linkage disequilibrium for all pairs of loci and genotype frequencies conformity to Hardy–Weinberg equilibrium (HWE) for each locus and for each population were tested using GENEPOP version 3.3 (Raymond & Rousset 1995).

Genetic diversity was estimated as the average number of alleles per locus (A) and mean expected heterozygosity across all loci (H_E, Nei's unbiased estimate; Nei 1978). As the number of alleles detected in a sample is correlated to sample size (Hauser *et al*. 2002), and sample size varied among samples, we estimated A of each sample at 14 individuals, the size of the smallest sample, to facilitate comparisons across populations (A_{14}). Average values and 95% confidence limits for A_{14} were calculated by resampling 14 individuals 1000 times (Hauser *et al*. 2002) using POPTOOLS version 2.5 (add-in for Microsoft Excel by G. Hood, http://www.cse.csiro.au/poptools/). Average values and confidence limits were achieved by subsampling without and with replacement, respectively (Hauser *et al*. 2002).

At migration–drift–equilibrium genetic diversity at neutral loci is strongly linked to effective population size. Therefore one might predict a correlation between neutral genetic diversity and ecological factors that affect effective population size. Here the drainage area upstream of a sample site was used as a surrogate for habitat size. This measure has been shown to be a better correlation with genetic diversity than complete drainage area for the freshwater fish *Cottus gobio* (Hänfling *et al*. 2002). Linear regression analysis for A_{14} against upstream drainage area and an ANOVA to test its significance were undertaken using SPSS version 10.0.1 (SPSS Inc.).

Diverse methods have been developed to determine if recent reductions in effective population size have occurred, and here three different bottleneck tests were applied (Cornuet & Luikart 1996; Luikart *et al*. 1998; Garza & Williamson 2001). Employing the program BOTTLENECK version 1.2.02 (Piry *et al*. 1999) heterozygosity excess, under the infinite allele model (IAM), the stepwise-mutation model (SMM) and a two-phase model (TPM), was tested using the Wilcoxon signed rank test (recommended when using a small number of microsatellite loci; Cornuet & Luikart 1996), and the distribution of allele frequency classes were examined for a distortion from the normal L-shaped distribution (Luikart *et al*. 1998). Because the two previous tests are particularly sensitive to deviations from HWE, locus *LCO4* was excluded from the analysis. A third method, the Garza & Williamson (2001), was applied using the programs M_P_VAL and CRITICAL_M (Garza & Williamson 2001). Herein, the M value [$M = A/R$, the ratio of the number of alleles (A)

to the microsatellite size range (*R*); Garza & Williamson 2001] was compared within 10 000 simulation replicates, considering different model assumptions (SMM, $p_s = 1.0$ and $\Delta_g = 1.0$; TPM, $p_s = 0.9$ and $\Delta_g = 3.5$; TPM, $p_s = 0.88$ and $\Delta_g = 2.8$; Garza & Williamson 2001) under a range of pre-bottleneck population size (N_e) and mutation rate (μ) combinations ($\theta = 4N_e\mu$ between 0.1 and 50, corresponding to $N_e = 50$–25 000 when $\mu = 5 \times 10^{-4}$/locus/generation and to $N_e = 250$–125 000 when $\mu = 10^{-4}$/locus/generation). The locus *SarN5C12b* was excluded from the *M* ratio analysis, as it did not meet the test requirements of having rare alleles (< 0.1 frequency) or empty allelic states within the range of common alleles (> 0.1 frequency) (Garza & Williamson 2001).

Genetic differentiation based on microsatellite data was analysed using complementary approaches both on the individual and the population level. Individual multilocus genotypes were subjected to a three-dimensional–factorial correspondence analysis (3D–FCA; Benzécri 1973) using GENETIX version 4.01 (Belkhir *et al*. 2000). Such an approach does not make *a priori* assumptions about populations and visualizes both variation within and differentiation among populations. The relative amount of differentiation among populations was quantified computing pairwise F_{ST} (Weir & Cockerham 1984) values using ARLEQUIN version 2.001 (Schneider *et al*. 2000). Pairwise Cavalli-Sforza & Edwards's (1967) chord distances (*CSE*) were calculated and an unrooted NJ tree constructed using the PHYLIP version 3.5c software package (Felsenstein 1993), because *CSE* is particularly suitable for reconstructing trees from microsatellite data (Takezaki & Nei 1996). The stability of the nodes was tested by 1000 bootstrap replicates. In order to test whether genetic differentiation within drainages can be explained simply by geographical distance (isolation by distance), linear regression analysis of $F_{ST}/(1 - F_{ST})$ vs. the river distance (km; as calculated for NCA) was undertaken for each pair of populations from the same drainage, as suggested by Rousset (1997), and using SPSS 10.0.1 (including an ANOVA to test its significance). Note that all the pairwise comparisons represent independent data points, since only one pair of populations from each drainage was used.

Results

The alignment of the 1140 bp sequences was straightforward, with no indels or stop codons detected upon putative translation into amino acids. A total of 1111 positions were constant and 12 were parsimony informative, the third being the most informative codon position (21 variable characters and 10 parsimony informative) followed by the first (five variable characters and two parsimony informative). The average number of transitions (ti) and transversions (tv) were 5.44 and 0.86, respectively, and the average ti/tv ratio was 5.38. A total of 23 distinct cyt *b* haplotypes were identified for the 107 specimens assayed (Table 1).

Table 1 Distribution of *Squalius aradensis* 23 mtDNA haplotypes (Hn) and genetic diversity measures at each sample location (see Fig. 1 for samples abbreviations)

Haplotypes	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	N_H	N_S	h	π
SEI2	8	1	1	1																				4	11	0.468	0.0005
SEI3	11																							1	11	0.000	0.0000
ALJ2					5	3	1	1	1															5	11	0.727	0.0011
ALJ3					4				3	1	1	1												5	10	0.758	0.0015
ALV1						6							4	1										3	11	0.589	0.0008
ALV2						11																		1	11	0.000	0.0000
ARD1	5														2	3	1							4	11	0.710	0.0033
ARD3																		10	1					2	11	0.173	0.0002
QRT1																				1	7	3		3	11	0.537	0.0005
QRT2																					7	1	1	3	9	0.392	0.0006
N_S	24	1	1	1	9	20	1	1	4	1	1	1	4	1	2	3	1	10	1	1	14	4	1		107	0.435*	0.0008*

N_H, total number of haplotypes; N_S, total number of specimens; h, haplotype diversity; π, nucleotide diversity; *average value across samples.

For microsatellites, the analysis with MICRO-CHECKER 2.2 revealed possible scoring errors and/or presence of null alleles for both cross-species loci *LCO4* and *LCO8*. Null alleles and possible scoring errors due to stuttering were evident for almost all populations at locus *LCO8*, causing this locus to be excluded from further analyses. At locus *LCO4*, two populations (ALV2 and QRT1) showed evidence of null alleles. Consequently allele frequencies of these populations were adjusted for the presence of null alleles according to Brookfield (1996), as suggested in MICRO-CHECKER 2.2. Using GENEPOP 3.3, locus *LCO4* exhibited consistent significant deviations from HWE for populations SEI3, ALV1, ALV2 and QRT1 (Appendix). No significant ($P < 0.05$) gametic linkage disequilibrium for any pairs of loci was found. With a few exceptions, all six microsatellite loci were polymorphic for all populations (Appendix), with four to 28 alleles per locus across all samples. 'Private' alleles (Slatkin 1985) were identified in six of the 10 *Squalius aradensis* populations sampled, ranging from an average frequency of 0.0–4.5% private alleles per locus among populations.

Within-population genetic diversity

The levels of within-population diversity at both cyt *b* and microsatellite loci varied markedly across populations. Haplotype diversity (h) ranged from 0.000 to 0.758 and nucleotide diversity (π) from 0.0000 to 0.0033 (Table 1). The values for both h and π were particularly high in population ARD1. Levels of genetic variability at microsatellite loci ranged from $A_{14} = 2.0$ and $H_E = 0.236$ for population ALV1 to $A_{14} = 6.5$ and $H_E = 0.637$ for population ARD1 (Appendix). A significant positive correlation ($y = 0.0759x + 16.413$, $r^2 = 0.651$, $F_{1,8} = 14.825$, $P \leq 0.005$) was observed between A_{14} and upstream drainage area (linear regression analysis plot not shown).

Five populations showed significant ($P < 0.05$) heterozygote excess according to IAM, three populations according to TPM and one according to SMM, and there was a shift in the distribution of allele frequency classes in two populations (Table 2). Most consistently, the two populations from the Seixe drainage exhibited significant bottleneck effects. Though no such effect was detected for ALV1 under any specific mutation model, the low number of polymorphic loci in this population resulted in a low power of the Wilcoxon signed rank test (Table 2). Applying the approach of Garza & Williamson (2001), the M value for each sampling site ranged from 0.45 at QRT2 to 0.89 at ARD1 (Table 2). Seixe, ALV2, ARD3 and Quarteira presented a relatively high number of situations in which its M ratio value was significantly smaller than the equilibrium expectation ($P < 0.05$) across the different combinations of mutation models and prebottleneck θ values considered (Table 2).

Among-populations genetic structure

The global Φ_{ST} value among populations based on cyt *b* sequence data was high (0.829, $P < 0.001$). Pairwise Φ_{ST} values ranged from −0.03 to 1.00 and were significant ($P < 0.05$), except for some within-drainage comparisons (Seixe, Aljezur, Alvor and Quarteira; Table 3). Though significant, more reduced Φ_{ST} values were also observed between ARD1 and Seixe samples (0.21–0.23) and between Aljezur and Alvor samples (0.32–0.41). A considerable level of

Table 2 Bottleneck analysis. Heterozygote excess (significant P values in bold), deviated allele frequency classes and M ratio tests

	Heterozygote excess and deviated allele frequency classes				
Sampling site	IAM*	TPM*	SMM*	L-shape*	M ratio value
SEI2	**0.016**	**0.016**	**0.016**	deviation	0.77‡
SEI3	**0.016**	**0.016**	**0.047**	no deviation	0.69‡
ALJ2	0.562	0.937	0.969	no deviation	0.80
ALJ3	0.844	0.937	0.969	no deviation	0.83
ALV1†	0.125	0.250	0.250	deviation	0.85
ALV2	0.937	0.969	0.969	no deviation	0.69‡
ARD1	**0.047**	0.500	0.922	no deviation	0.89
ARD3	0.594	0.891	0.953	no deviation	0.70‡
QRT1	**0.031**	0.406	0.969	no deviation	0.53‡
QRT2	**0.016**	**0.047**	0.891	no deviation	0.45‡

*Significance of heterozygote excess according to the Wilcoxon signed rank test under the IAM, TPM and SMM for each population, and deviation of allele frequency classes from a normal L-shaped; †note that the tests for ALV1 are only based on two polymorphic loci (Appendix); ‡sampling sites with M ratio value significantly smaller than the equilibrium expectation ($P < 0.05$) for a relative high number (≥ 50%) of situations across the different combinations of mutation models and prebottleneck θ values considered.

Table 3 Estimated pairwise Φ_{ST} (cyt *b* data, below diagonal) and F_{ST} (microsatellites data, above diagonal) values among populations. Φ_{ST} overall value = 0.829 ($P < 0.001$) and F_{ST} overall value = 0.286 ($P < 0.001$)

Populations	SEI2	SEI3	ALJ2	ALJ3	ALV1	ALV2	ARD1	ARD3	QRT1	QRT2
SEI2		0.016	0.399	0.366	0.315	0.380	0.114	0.267	0.160	0.171
SEI3	0.000*		0.379	0.349	0.309	0.384	0.140	0.297	0.198	0.209
ALJ2	0.914	0.939		0.008	0.247	0.197	0.355	0.441	0.444	0.456
ALJ3	0.899	0.923	−0.030*		0.185	0.158	0.334	0.432	0.427	0.434
ALV1	0.928	0.955	0.318	0.347		0.241†	0.315	0.454	0.401	0.378
ALV2	0.971	1.000	0.382	0.413	0.200*†		0.325	0.408	0.408	0.409
ARD1	0.207	0.230	0.677	0.660	0.681	0.720		0.106	0.088	0.111
ARD3	0.892	0.971	0.901	0.881	0.921	0.985	0.439		0.165	0.225
QRT1	0.940	0.967	0.725	0.700	0.744	0.870	0.690	0.938		0.018
QRT2	0.938	0.968	0.708	0.679	0.730	0.872	0.668	0.938	−0.009*	

*No significant ($P > 0.05$) pairwise values within drainage comparisons except for Arade; †distinct levels of cyt *b* sequence and microsatellite loci genetic differentiation between populations within Alvor, with a ALV1–ALV2 Φ_{ST} value considerably smaller than the overall Φ_{ST} value and a ALV1–ALV2 F_{ST} value at the same level of the overall F_{ST} value.

genetic differentiation among populations (global $F_{ST} = 0.286$, $P < 0.001$) was also observed at microsatellite loci, with pairwise F_{ST} values ranging from 0.008 to 0.456, all being significant ($P < 0.05$; Table 3). Levels of genetic differentiation at mtDNA and microsatellite loci were generally concordant when comparing samples within drainages (Table 3). One noteworthy exception was the sample pair from the Alvor drainage, which exhibited an ALV1–ALV2 Φ_{ST} value considerably smaller than the overall Φ_{ST} value, but an ALV1–ALV2 F_{ST} value at the same level of the overall F_{ST} value (the highest F_{ST} value within drainage; Table 3). The Arade samples, ARD1 and ARD3, revealed much higher levels of within-drainage genetic divergence for both data sets in comparison with other drainages (Table 3).

The *TN* model, assuming rate matrix parameters of 1.0000, 18.5576, 1.0000, 1.0000 and 3.7528, and base frequencies of 0.2590 for A, 0.2881 for C, 0. 1709 for G and 0.2820 for T, proved to be the most appropriate evolutionary model for the cyt *b* data set. Phylogenetic clustering of cyt *b* haplotypes using both NJ of *TN*-distance and MP analysis revealed identical trees for all supported nodes and small differences in the level of bootstrap support among methods (Fig. 2a). The haplotype tree showed two well-supported clades in which haplotypes from the Arade drainage occupied basal positions. Haplotypes from the other drainages were grouped into three clades, Seixe + some haplotypes from ARD1, Aljezur + Alvor and Quarteira, respectively, although the bootstrap support for the latter two groups was low (Fig. 2a). Furthermore, within the Aljezur + Alvor group the Aljezur samples occupy mainly derived positions whereas the Alvor samples occupy more basal. A similar picture was revealed by the microsatellite analysis although complete congruence cannot be expected due to the different nature of both data sets (gene tree of maternally inherited mtDNA vs. population tree based on nuclear markers) (Fig. 2). The NJ tree of *CSE* distances showed that Seixe, Aljezur + Alvor and Quarteira are well-supported monophyletic groups (Fig. 2b), which is also apparent from the 3D–FCA (plot not shown). The exact position of populations ARD1 and ARD3 is not clear from the NJ tree of *CSE* distances (Fig. 2b), but both populations occupy a central position in the 3D–FCA (plot not shown) confirming the ancestral status of the Arade populations.

Nested clade analysis (Figs 3 and 4) indicated the presence of two higher-level clades, which are synonymous to the two major clades of the haplotype tree (Fig. 2a). Clade 3-1 included Seixe, ARD1 and ARD3 haplotypes, and clade 3-2 included Aljezur, Alvor, ARD1 and Quarteira haplotypes (Figs 3 and 4). Within both main clades Seixe, Aljezur, Alvor and Quarteira samples comprise haplotypes that occupy mainly tip-derived positions in the cladogram, while both Arade populations include haplotypes with essentially interior-ancestral positions (Fig. 3). Clear association between the two major clades and geographical location was found for both nested contingency test and nested geographical distance analysis ($P < 0.05$; Fig. 4), and allopatric fragmentation was proposed to underlie the differentiation of the main clades (Templeton 1998; upgraded version 14.07. 2004; Fig. 4). Two 2-step clades were observed within clade 3-1 (2-1, Seixe and ARD1; 2-2, ARD3), and three 2-step clades were observed within 3-2 (2-3, Quarteira; 2-4, Aljezur and Alvor; 2-5, ARD1) (Figs 3 and 4). The geographical association ($P < 0.05$) was in both cases significant, and allopatric differentiation was again the most likely cause for the differentiation (Fig. 4). Two-step clades were substructured in 10 1-step clades comprising the 23 haplotypes (Figs 3 and 4). Significant geographical association ($P < 0.05$) among 1-step clades and on the haplotype level was only

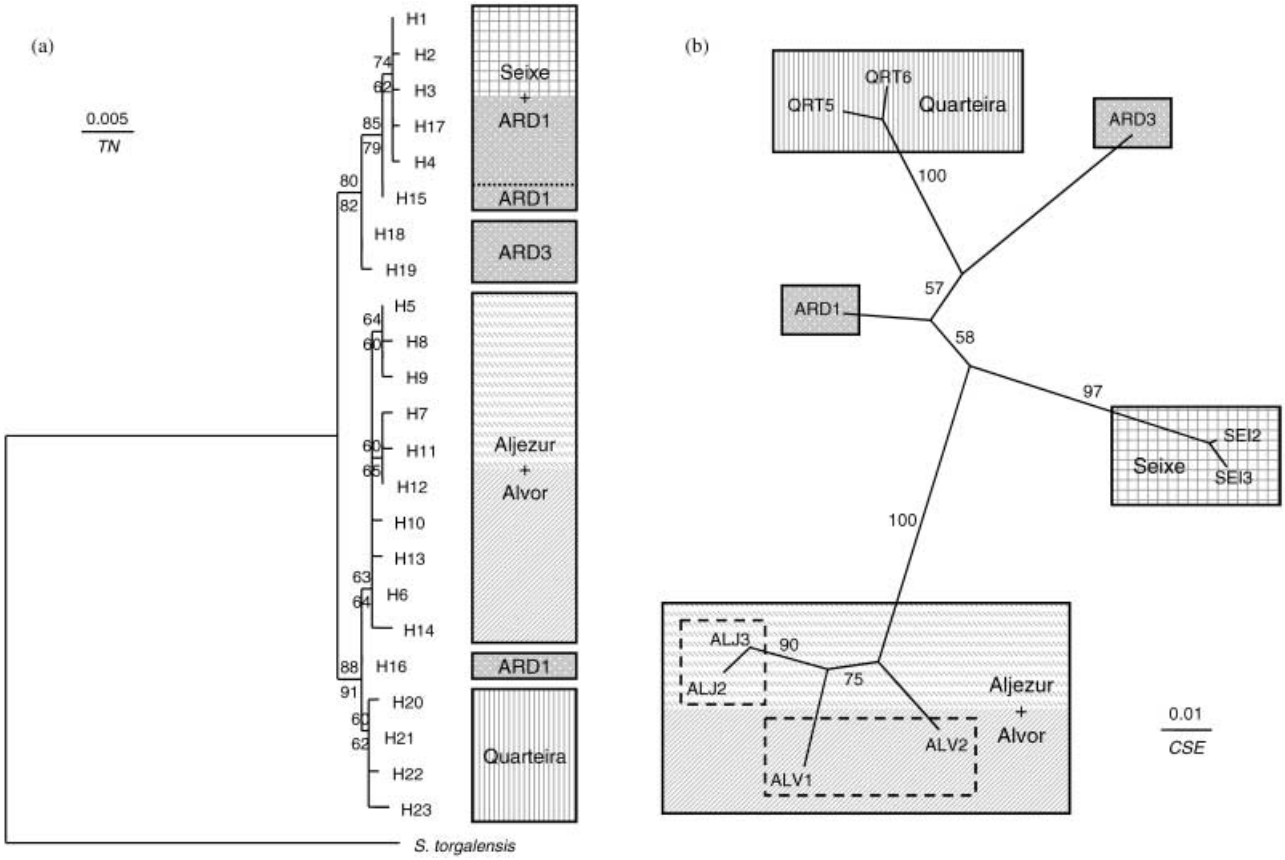

Fig. 2 Neighbour-joining dendrograms inferred from 1140 bp cyt *b* haplotypes using the *TN* distance (a. rooted, using *Squalius torgalensis* as outgroup species) and from a matrix of *CSE* pairwise values based on six microsatellite loci (b. unrooted). For 1000 bootstrap replicates node values (%) of 50% and higher are shown in (a) for both NJ (above) and MP (below) analysis.

found within clades 2-4 and 1-10. Both clades contained a mixture of Alvor and Aljezur haplotypes, and in both cases the inference key indicated that restricted gene flow could have contributed to their differentiation (Fig. 4). Furthermore, within clade 2-4 Aljezur haplotypes present mainly tip-derived positions, whereas Alvor haplotypes occupy primarily interior-ancestral positions (Fig. 3). The clades 2-1 and 1-1, including Seixe and ARD1, showed no significant geographical association ($P > 0.05$; Fig. 4), though comprising distinct isolated drainage systems. All other nonsignificant associations comprised only haplotypes of a single drainage. With a distinct pattern from all the other drainages, the Arade samples, and specifically the ARD1 locality, appears distributed in different 1-, 2- and 3-step clades, including representatives of the two most divergent haplotype lineages (Figs 3 and 4). However, fragmentation events for these clades were proposed by NCA, a situation that might indicate secondary contact (Templeton 2001). The average pairwise distance between geographical centres of the clades found at ARD1 did not approach zero with rising clade level, which is a further indication for secondary contact (Templeton 2001).

The two-cluster and branch-length tests for constancy of evolutionary rates revealed no significant deviation from equality (at 1% level). In order to infer the evolutionary history of the *S. aradensis* populations, several factors were taken into account, including the genetic diversity of the cyt *b* haplotypes found in ARD1, their scattered positions in the phylogenetic trees and NCA cladogram (Figs 2, 3 and 4) and subsequent inferred secondary contact events. To allow for these considerations, the divergence times between Arade and Aljezur + Alvor and between Arade and Quarteira were calculated considering haplotype H16 from ARD1, and divergence times between Arade and Seixe by considering haplotypes H1, H15 and H17 from ARD1 or the haplotypes present in ARD3. Estimates of the divergence times using the molecular clock hypothesis ranged from ≤ 0.1 Myr between ARD1 and the Seixe, Alvor and Quarteira drainages and between Aljezur and Alvor, to 0.6–0.8 Myr between the two major cyt *b* mtDNA lineages (clades 3-1 and 3-2). Divergence times between the sister species *S. torgalensis* in Mira drainage and the Arade samples were 6.6–8.2 Myr.

Linear regression analysis (plot not shown) revealed that there was no general correlation between geographical and

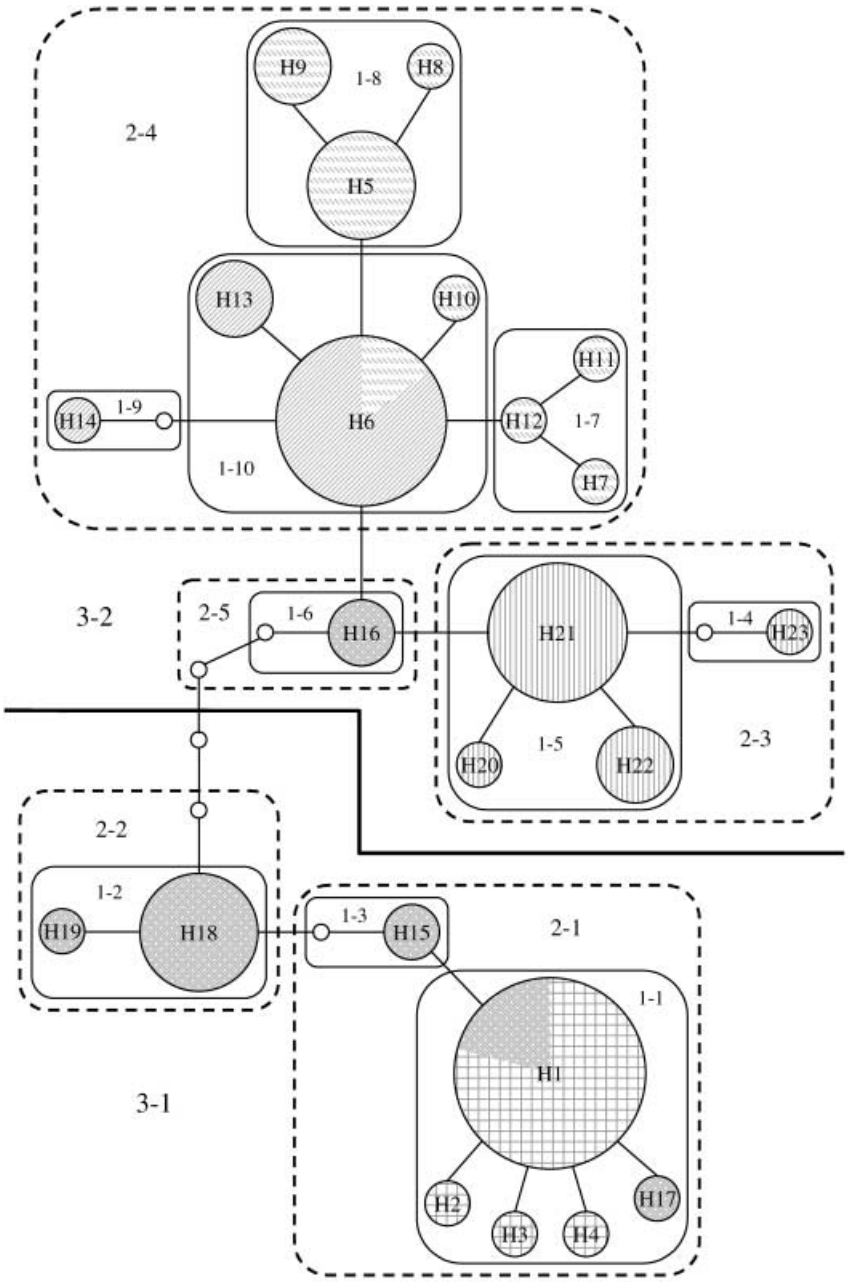

Fig. 3 Unrooted network and nesting design derived from the 23 mtDNA haplotypes (Hn) found in *Squalius aradensis* samples from the five drainages, Seixe (□), Aljezur (□), Alvor (□), Arade (□), Quarteira (□). Each line represents one mutational step connection between two haplotypes (0-step clades) and all are supported as being parsimonious at the 95% level. Narrow-lined polygons define 1-step clades (1-n), dash-lined polygons identify 2-step clades (2-n) and the thick line separates the 3-step clades (3-n). Small empty circles represent missing intermediate haplotype states.

genetic distance [$F_{ST}/(1 - F_{ST})$] within drainages. Genetic distance was particularly large compared to geographical distance among the Alvor sites (ALV1–ALV2), whereas the two sites from Seixe drainage (SEI2–SEI3) exhibited the smallest values of genetic divergence in comparison with geographical distance.

Discussion

Assessment of the population subdivision in *Squalius aradensis* through combined analysis of mitochondrial and nuclear variation, using traditional statistics of population structure (Φ-/*F*-statistics), phylogenetic trees, ordination methods and cladistic nested analysis, was largely congruent and partly complementary. Analyses revealed a detailed pattern of geographical structure on both recent and historical timescales. Findings are especially relevant for endangered species with highly restricted distribution and intermittent habitats such as *S. aradensis*.

Phylogeography and evolutionary history

Our data support the conclusion from previous studies (Brito *et al*. 1997; Sanjur *et al*. 2003), that *S. aradensis* diverged from its

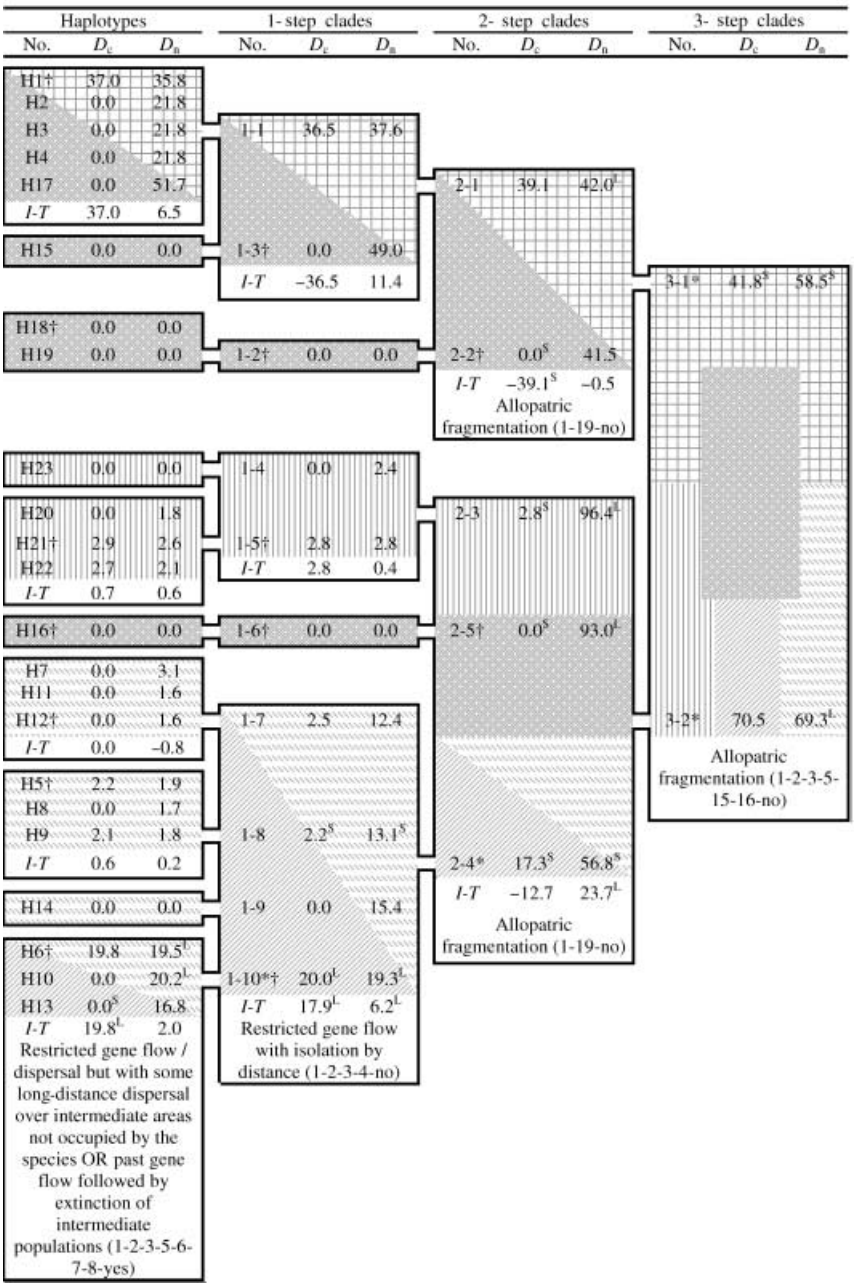

Fig. 4 Nested cladistic analysis of the geographical distribution of *Squalius aradensis* mtDNA haplotypes. See Fig. 4 for nesting design, haplotypes (0-step clades) and clades designation, and drainages shading. Interior clades are signed (†). Significant, at 95% level, nested contingency tests of geographical association for clades with genetic and geographical variation are signed (*). For each clade the hydrographic D_c and D_n distances (see text for definitions and calculation methods) are given in kilometres. For clades including tip and interior nested clades the average difference between interior vs. tip clades for both distance measures is given (*I-T*). Distance measures that were significantly small (S) or large (L) at the 5% level are indicated. For each clade with significant results the biological inference, done following the upgraded version of the inference key from Templeton (1998; 14.07. 2004), is given.

sister species, *Squalius torgalensis*, 6.6–8.2 million years ago (Ma) (Fig. 1), possibly coinciding with the development of the eastern mountain ('Serra do Caldeirão'), where the Mira, Arade and Quarteira drainages are established, in the Lower Pliocene–Upper Miocene (3.6–7.2 Myr; Dias 2001). Compared to this timescale, the differentiation among *S. aradensis* populations is relatively young and was caused by various allopatric fragmentation events during the last 0.8 Myr.

The two Arade populations were never clustered together in the phylogenetic trees and in the NCA network, and indeed represent two of the most divergent lineages. Furthermore Arade populations occupied basal/central positions in all phylogenetic trees and in the 3D–FCA, and the interior-ancestral positions of both clades within the NCA haplotypes network were occupied by Arade haplotypes (H16 in clade 3-2 and H18 in clade 3-1). Such patterns, supported by high levels of nucleotide and microsatellite diversity, indicate that Arade may represent the evolutionary centre of the species, i.e. the distribution of *S. aradensis* was originally restricted to the Arade drainage (or equivalent ancestral single drainage). Accordingly, the two major haplotype lineages (clade 3-1 and clade 3-2) were derived

from an allopatric fragmentation within the Arade drainage 0.6–0.8 Ma (Fig. 1). Biogeographical patterns in freshwater fishes are often determined by events of fluvial captures in the headwaters and of rivers confluence/divergence (due to marine regression/transgression) in the river mouths (e.g. in cyprinids, Durand *et al*. 1999). A putative scenario for the within-Arade fragmentation entails elevated sea levels during the Quaternary times (< 1.8 Myr; occurrence of sequence of several alternated marine regression and transgression events; N. Pimentel, personal communication), resulting in isolation of the various Arade tributaries through a brackish water barrier (0.6–0.8 Myr; Fig. 1). Such a barrier must have disappeared subsequently leading to secondary contact of both clades in ARD1 (H16 of clade 3-2 and H1, H15 and H17 of clade 3-1). However, there are alternative explanations for the occurrence of these two haplotype lineages within site ARD1, as discussed later.

According to both mtDNA and microsatellite data Aljezur + Alvor and Quarteira formed well-supported monophyletic groups, which are derived in respect to the Arade populations. Applying a molecular clock estimate suggests that both the Aljezur + Alvor group and Quarteira would have been isolated from the Arade, from the 3-2 clade, 0.1 Ma (Fig. 1). Alternated marine regression and transgression events (Quaternary, last 1.8 Myr) seem an unlikely explanation for the separation between Arade and the Aljezur + Alvor and between Arade and Quarteira. A connection between rivers via a common estuary would have meant that both of the Arade lineages (which were separated in different tributaries) could have invaded Alvor and Quarteira. However, only haplotypes of lineage 3-2 are found in these drainages, making the alternative explanation of colonization by fluvial captures in the headwater regions of one of the Arade tributaries more plausible. Furthermore, several geological and geomorphological aspects strongly indicate a separation due to fluvial captures and reorganization of the different drainage areas (N. Pimentel, personal communication), namely: (i) geographical proximity of drainage basins, but far distance between river mouths; (ii) hydrographic geometric patterns suggesting fluvial captures (e.g. square angles and aligned valleys draining to opposite directions); and (iii) no record of any Quaternary important regression event until the last glacial (Sicilian, 0.02 Ma, with sea level *c*. –180 m; Dias *et al*. 2000).

Populations from rivers Alvor and Aljezur showed little, but nonetheless, significant differentiation at both microsatellites and mtDNA. According to NCA this differentiation can be most likely explained by isolation by distance with restricted gene flow. We interpret the NCA results as a very recent fragmentation, most likely through a river capture in the headwater region, because it can be assumed that there is no current gene flow between the two drainages. Mitochondrial DNA data support the view that Alvor presents a more interior-ancestral position in the cladogram, probably representing the origin/centre of the Aljezur differentiation (< 0.1 Myr; Fig. 1). The formation of the western mountain ('Serra de Monchique'), where the Seixe, Aljezur, Alvor and Arade drainages are established, is reported to take place during the Upper Pliocene (1.8–3.6 Myr; Dias 2001), thus predating the differentiation of among *S. aradensis* populations. However, this area continues to exhibit considerable seismic activity (N. Pimentel, personal communication), constituting a possible geological event affecting differentiation, such as small fluvial captures among the headwater regions.

Curiously, one of the two Arade samples (ARD1) even contained representatives of the two most divergent haplotype lineages. Neutral theory predicts that lineage sorting would create monophyletic lineages among haplotypes of isolated populations over time. Given the high level of divergence between the lineages found in population ARD1 it appears extremely unlikely that these represent retained ancient polymorphism. This fact and the geographical distribution of these two lineages suggest that ARD1 is in fact a secondary contact zone (Fig. 1). The lack of significant geographical structuring observed for the clades that included Seixe and ARD1, in the presumed absence of gene flow between the two drainages, together with the suggested secondary contact zone in ARD1, suggests two possible scenarios (Fig. 1). First, the secondary contact in ARD1 could be the result of re-established gene flow (0.2–0.3 Myr) between previously isolated lineages within the Arade (0.6–0.8 Myr) and Seixe was colonized from ARD1 haplotypes (< 0.1 Myr; Fig. 1), possibly as recently as after the last glaciation. In this case, the absence of clade 3-2 haplotypes in the Seixe would be explained by a founder effect or by the fact that secondary contact occurred following the separation event. In support of this scenario is the fact that the microsatellite data of Seixe show clear evidence of a genetic bottleneck, and that differentiation between the two Seixe sites is unusually low, suggesting very recent geographical expansion. The second scenario (Fig. 1), would be a differentiation of Seixe from ARD3 through allopatric fragmentation about 0.2–0.3 Ma, and a recent invasion of the ARD1 by Seixe individuals (postglacial, < 0.1 Myr; fluvial captures in the close headwaters or stocking). An argument against would be that there is no obvious geographical connection between ARD3 and Seixe. Moreover, the information about low differentiation and associated population bottleneck in the Seixe populations inferred from microsatellite data is not really congruent with this scenario.

Implications for conservation

Data from the present study can support conservation management decisions by analysing the pattern of geographical distribution of genetic variability within populations and differentiation among populations.

A significant correlation was found between *S. aradensis* allelic diversity and upstream drainage area. Such reduced genetic diversity for smaller upstream drainage areas, may be related to the impact of upstream drainage area on habitat size and effective population size (Hänfling & Brandl 1998; Hänfling *et al*. 2002) and to the heterogeneous nature of local hydrological conditions. River courses with smaller upstream drainage area will tend to experience an enhanced impact of summer drought, with fewer dry-season fish refuges available, resulting in higher mortality, reduced population size and loss of genetic diversity. Such reduced genetic variability reflects not only low and possibly fluctuating population sizes and therefore a demographic vulnerability of upstream populations, but also indicates a genetic vulnerability due to inbreeding. Consequently we recommend that conservation efforts should pay particular attention to isolated upstream populations and smaller drainages such as Seixe, Aljezur and Alvor. Despite the significant levels of genetic differentiation generally observed among populations/drainages for both mtDNA and microsatellites markers, Arade presented considerably higher levels of genetic variability, comprising both major *S. aradensis* evolutionary divergent lineages. As such it likely represents the evolutionary centre of origin of the species, constituting, besides the other populations from the drainage, an important reservoir for the recolonization of other drainages following local extinction events.

In order to optimize the management and conservation strategies to be implemented in the SMTDPs, and in particular for *S. aradensis*, it is necessary to identify distinct population units. However, the identification (and subsequent management) of distinct evolutionary significant units (ESUs; Moritz 1994) within species is far from simple, with emphasis being given to adaptive phenotypic diversity and/or historical isolation processes (reviewed in Moritz *et al*. 2002). Recent studies (e.g. Angers & Bernatchez 1998; Mesquita *et al*. 2001; Salgueiro *et al*. 2003) have used mtDNA and microsatellites information to identify evolutionary distinct groups of populations within a species—identification of ESUs, and of management units (MUs; Moritz 1994). Here we combined the information of both mtDNA and nuclear data in order to propose the most suitable units for the future management of populations of this critically endangered species. Considering the criteria of phylogeographical differentiation and reciprocal monophyly for ESUs, plus the observed significant values of Φ/F_{ST} for both mtDNA and microsatellite data, at least four main ESUs should be considered, Seixe + ARD1, ARD3, Aljezur + Alvor and Quarteira. However it should be taken into consideration, that given the heterogeneous nature of the Arade populations, other ESUs may exist in unsampled tributaries of the Arade drainage.

Freshwater systems in general have been significantly affected by habitat degradation, introduction of alien species, overharvesting and water diversion (reviewed in Cowx 2002). The SMTDP, though representing a very small hydrological area of the Iberian Peninsula, contains considerable ichthyofaunistic value, with the presence of local endemic species, each exhibiting a highly restricted distribution, together with other endangered Iberian endemic cyprinid species (Mesquita & Coelho 2002). The unique genetic diversity, the very restricted distribution and the specific intermittent hydrological conditions of these rivers make it extremely important to preserve habitats that support populations during the droughts, as well as implementing measures to prevent the dispersion of exotic species and obstructions to population connectivity (Magalhães 2002; Mesquita & Coelho 2002).

Final remarks

In conclusion, the Arade drainage most likely represents the evolutionary origin of the species, and relatively recent processes of fluvial capture should explain its distribution in the other drainages. Although there are alternative hypothesis to explain the observed pattern of genetic structuring we consider these less likely. Moreover, levels of within-drainage differentiation varied substantially and were unlikely to be explained by isolation by distance, but more likely by drainage specific processes such as bottlenecks followed by a recent expansion and anthropogenic differentiation.

Acknowledgements

Authors thank 'FCUL-Cyprinids' group for help in fieldwork, Carina Cunha for help with cyt *b* sequence data collection, Cock van Oosterhout and Lorenz Hauser for assisting in analysis, Nuno Pimentel for support in geological data discussion and Africa Gómez for helpful comments. This work was supported by the project FCT/POCTI/BSE/40868/2001 (coparticipated by FEDER) granted to M. M. Coelho, by PRAXISXXI BD/19751/99 to N. Mesquita, and by the INAG/FCUL protocol 'Estudos genéticos da ictiofauna–medidas de monitorização e de compensação da construção da barragem de Odelouca'.

References

Angers B, Bernatchez L (1998) Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, **15**, 143–159.

Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, **9**, 686–690.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) *GENETIX 4.01, logiciel sous Windows™ pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, Université de Montpellier II, Montpellier. (http://www.univ-montp2.fr/~genetix/genetix/genetix.htm).

Benzécri JP (1973) *L'analyse Des Donnèes. Tome II: L'analyse Des Correspondances*. Dunod Press, Paris.

Brito RM, Briolay J, Galtier N, Bouvet Y, Coelho MM (1997) Phylogenetics relationships within genus *Leuciscus* (Pisces, Cyprinidae) in Portuguese freshwaters, based on mitochondrial DNA cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, **8**, 435–442.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Carvalho GR (1998) Molecular ecology: origins and approach. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 1–23. IOS Press, Amsterdam.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimations procedures. *American Journal of Human Genetics*, **19**, 233–257.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660. (http://darwin.uvigo.es/).

Coelho MM, Brito RM, Pacheco TR, Figueiredo D, Pires AM (1995) Genetic variation and divergence of *Leuciscus pyrenaicus* and *L. carolitertii* (Pisces, Cyprinidae). *Journal of Fish Biology*, **47** (Suppl. A), 243–258.

Coelho MM, Alves MJ, Rodrigues E (1997) Patterns of genetic divergence in *Chondrostoma lusitanicum* Collares-Pereira, in intermittent Portuguese rivers. *Fisheries Management and Ecology*, **4**, 223–232.

Coelho MM, Bogutskaya NG, Rodrigues JA, Collares-Pereira MJ (1998) *Leuciscus torgalensis* and *L. aradensis*, two new cyprinids for Portuguese fresh waters. *Journal of Fish Biology*, **52**, 937–950.

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.

Cowx IG (2002) Analysis of threats to freshwater fish conservation: past and present challenges. In: *Conservation of Freshwater Fishes: Options for the Future* (eds Collares-Pereira MJ, Cowx IG, Coelho MM), pp. 201–220. Blackwell Science, Oxford.

Dias RP (2001) *Neotectónica da região do Algarve*. PhD Thesis, University of Lisbon, Portugal.

Dias JMA, Boski T, Rodrigues A, Magalhães F (2000) Coast line evolution in Portugal since the Last Glacial Maximum until present — a synthesis. *Marine Geology*, **170**, 177–186.

Doadrio I, Carmona JA (2004) Phylogenetic relationships and biogeography of the genus *Chondrostoma* inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 802–815.

Dowling TE, Tibbets CA, Minckley WL, Smith GR (2002) Evolutionary relationships of the plagopterins (Teleostei: Cyprinidae) from cytochrome *b* sequences. *Copeia*, **2002**, 665–678.

Durand JD, Templeton AR, Guinand B, Imsiridou A, Bouvet Y (1999) Nested clade and phylogeographic analyses of the chub, *Leuciscus cephalus* (Teleostei, Cyprinidae), in Greece: implications for Balkan Peninsula biogeography. *Molecular Phylogenetics and Evolution*, **13**, 565–580.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Felsenstein J (1993) PHYLIP *(Phylogeny Inference Package) version 3.5c*. Department of Genetics, University of Washington, Seattle. (http://evolution.genetics.washington.edu/phylip.html).

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318. (http://santacruz.nmfs.noaa.gov/staff/carlos_garza/software.php).

Hänfling B, Brandl R (1998) Genetic variability, population size and isolation of distinct populations in the freshwater fish *Cottus gobio* L. *Molecular Ecology*, **7**, 1625–1632.

Hänfling B, Hellemans B, Volckaert FAM, Carvalho GR (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology*, **11**, 1717–1729.

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11742–11747.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Johnson WE, Godoy JA, Palomares F *et al.* (2004) Phylogenetic and phylogeographic analysis of Iberian lynx populations. *Journal of Heredity*, **95**, 19–28.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

Magalhães MF (2002) *Gradientes espaciais, variabilidade ambiental e estrutura de comunidades piscícolas em rios de tipo Mediterrânico*. PhD Thesis, University of Lisbon, Portugal.

McElroy D, Morgan P, Bermingham E, Kornfield I (1992) REAP: an integrated environment for manipulation and phylogenetic analysis of restriction site data. *Journal of Heredity*, **83**, 157–158.

Mesquita N, Coelho MM (2002) The ichthyofauna of the small Mediterranean-type drainages of Portugal: its importance for conservation. In: *Conservation of Freshwater Fishes: Options for the Future* (eds Collares-Pereira MJ, Cowx IG, Coelho MM), pp. 65–71. Blackwell Science, Oxford.

Mesquita N, Carvalho G, Shaw P, Crespo E, Coelho MM (2001) River basin-related genetic structuring in an endangered fish species, *Chondrostoma lusitanicum*, based on mtDNA sequencing and RFLP analysis. *Heredity*, **86**, 253–264.

Mesquita N, Cunha C, Hänfling B *et al.* (2003) Isolation and characterization of polymorphic microsatellite loci in the endangered Portuguese freshwater fish *Squalius aradensis* (Cyprinidae). *Molecular Ecology Notes*, **3**, 572–574.

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, **16**, 1215.

Moritz C (1994) Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Moritz C, McGuigan K, Bernatchez L (2002) Conservation of freshwater fishes: integrating evolution and genetics with ecology. In: *Conservation of Freshwater Fishes: Options for the Future* (eds Collares-Pereira MJ, Cowx IG, Coelho MM), pp. 293–310. Blackwell Science, Oxford.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Jennifer K (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

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

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 248–249. (http://www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html).

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818. (http://darwin.uvigo.es/).

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488. (http://darwin.uvigo.es/).

Randi E, Tabarroni C, Rimondi S, Lucchini V, Sfougaris A (2003) Phylogeography of the rock partridge (*Alectoris graeca*). *Molecular Ecology*, **12**, 2201–2214.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact tests and acumenicism. *Journal of Heredity*, **86**, 248–249. (ftp://isem.isem.univ–montp2.fr/pub/pc/genepop/).

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Salgueiro P, Carvalho G, Collares-Pereira MJ, Coelho MM (2003) Microsatellite analysis of genetic population structure of the endangered cyprinid *Anaecypris hispanica*. Portugal: implications for conservation. *Biological Conservation*, **109**, 47–56.

Sanjur OI, Carmona JA, Doadrio I (2003) Evolutionary and biogeographical patterns within Iberian populations of the genus *Squalius* inferred from molecular data. *Molecular Phylogenetics and Evolution*, **29**, 20–30.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN *Version 2.000: a software for population genetic data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland. (http://lgb.unige.ch/arlequin/software/).

Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution*, **39**, 53–65.

Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, *Version* 4. Sinauer Associates, Sunderland, Massachusetts.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution*, **12**, 823–833. (http://mep.bio.psu.edu/databases.html#LINTREE).

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Templeton AR (1998) Nested clade analysis of phylogeographic data: testing hypothesis about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779–791.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.

Turner TF, Dowling TE, Broughton RE, Gold JR (2004) Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily Leusicinae). *Conservation Genetics*, **5**, 279–281.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538. (http://www.microchecker.hull.ac.uk/).

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

Partnership represents one of the main goals of Hull and Lisbon laboratories, with projects focusing especially on molecular genetic analysis of fish population biodiversity, origin and dynamics of population structuring, and implications for management and conservation. Natacha Mesquita is a PhD student co-supervised at the Lisbon and Hull Universities and this work comprises a chapter of her dissertation, which focuses on relations between historical and contemporary events, populations' genetic structure and conservation genetics of endemic cyprinids. Bernd Hänfling is a University Research Fellow interested in the evolution of freshwater fishes, in particular phylogeography and conservation genetics.

Appendix

Allelic variability at six microsatellite loci in 10 *Squalius aradensis* populations (for each population per locus and across loci; labels according Fig. 1)

Sampling site		*SarN7K4*	*SarN2F11a*	*SarN7F8*	*SarN7G5*	*SarN5C12b*	*LCO4**	Average all loci
SEI2	A	5	6	2	4	2	10	4.8
$N = 59$	A_{14}	4.7	5.0	2	3.9	2	5.4	3.8
	R	137–153	150–162	163–165	113–119	115–119	279–325	—
	S	137	150	163	117	119	309	—
	F	0.30	0.40	0.67	0.43	0.71	0.66	0.53
	H_O	0.746	0.712	0.458	0.661	0.339	0.542	0.576
	H_E	0.771	0.734	0.446	0.655	0.414	0.537	0.593
SEI3	A	5	6	3	5	2	7	4.7
$N = 96$	A_{14}	4.4	4.9	2.1	3.9	2.0	4.3	3.6
	R	137–153	150–162	161–165	113–125	115–119	279–323	—
	S	137	150	163	117	119	309	—
	F	0.41	0.32	0.56	0.52	0.86	0.70	0.56
	H_O	0.656	0.771	0.438	0.615	0.229	**0.458**†	0.528
	H_E	0.723	0.745	0.499	0.623	0.235	0.480	0.551
ALJ2	A	2	6	1	3	2	6	3.3
$N = 83$	A_{14}	1.2	3.3	1	2.2	1.2	4.5	2.2
	R	145–147	152–172	165	113–121	115–119	303–315	—
	S	147	166	165	113	119	303	—
	F	0.99	0.51	1.00	0.56	0.99	0.42	0.75
	H_O	0.012	0.518	0.000	0.530	0.012	0.651	0.287
	H_E	0.012	0.566	0.000	0.501	0.012	0.691	0.297
ALJ3	A	4	5	2	3	1	6	3.5
$N = 63$	A_{14}	1.8	3.3	1.4	2.4	1.0	4.3	2.4
	R	145–151	152–168	163–165	113–119	119	299–311	—
	S	147	152	165	113	119	309	—
	F	0.97	0.56	0.98	0.52	1.00	0.43	0.74
	H_O	0.063	0.540	0.032	0.619	0.000	0.635	0.315
	H_E	0.063	0.543	0.031	0.518	0.000	0.651	0.301
ALV1	A	2	1	1	2	1	7	2.3
$N = 14$	A_{14}	—	—	—	—	—	—	—
	R	147–149	152	165	113–117	119	295–317	—
	S	147	152	165	117	119	309	
	F	0.75	1.00	1.00	0.86	1.00	0.43	0.84
	H_O	0.500	0.000	0.000	0.286	0.000	**0.714**†	0.250
	H_E	0.389	0.000	0.000	0.254	0.000	0.772	0.236
ALV2	A	2	2	3	3	1	6	2.8
$N = 29$	A_{14}	1.5	1.7	2.4	3.0	1.0	4.8	2.4
	R	147–153	152–154	163–167	113–119	119	265–309	—
	S	147	152	165	113	119	301	—
	F	0.98	0.97	0.91	0.62	1.00	0.414	0.82
	H_O	0.034	0.069	0.172	0.59	0.000	**0.690**‡	0.259
	H_E	0.034	0.068	0.163	0.54	0.000	0.761	0.261
ARD1	A	9	12	2	8	3	20	9.0
$N = 59$	A_{14}	6.5	10.0	2.0	5.9	2.5	12.2	6.5
	R	137–153	140–168	163–165	111–127	115–123	275–327	—
	S	145	156	163	113	119	309	—
	F	0.46	0.21	0.83	0.31	0.86	0.16	0.47
	H_O	0.763	0.915	0.271	0.661	0.220	0.864	0.616
	H_E	0.724	0.888	0.284	0.769	0.242	0.915	0.637

Appendix *Continued*

Sampling site		*SarN7K4*	*SarN2F11a*	*SarN7F8*	*SarN7G5*	*SarN5C12b*	*LCO4**	Average all loci
ARD3	*A*	4	7	2	5	3	11	5.3
N = 56	A_{14}	3.4	5.9	1.2	3.5	1.5	6.3	3.6
	R	145–153	140–172	163–165	107–119	111–123	277–313	—
	S	147	156	163	113	119	305	—
	F	0.54	0.36	0.99	0.59	0.98	0.38	0.64
	H_O	0.464	0.750	0.018	0.589	0.036	0.679	0.423
	H_E	0.588	0.769	0.018	0.568	0.036	0.721	0.450
QRT1	*A*	4	12	3	7	3	13	7.0
N = 71	A_{14}	3.6	8.3	2.3	4.6	2.4	7.0	4.7
	R	141–153	140–220	143–165	113–125	111–123	281–331	—
	S	145	140	163	119	119	305	—
	F	0.49	0.27	0.72	0.61	0.77	0.36	0.54
	H_O	0.662	0.845	0.437	0.606	0.352	0.781	0.614
	H_E	0.624	0.848	0.415	0.580	0.358	0.774	0.600
QRT2	*A*	4	10	3	5	2	12	6.0
N = 55	A_{14}	3.2	7.5	3.0	3.6	2.0	7.6	4.5
	R	141–153	140–214	143–165	113–125	119–123	281–331	—
	S	145	152	163	119	119	305	—
	F	0.62	0.32	0.62	0.62	0.65	0.43	0.54
	H_O	0.545	0.836	0.564	0.545	0.418	**0.655‡**	0.594
	H_E	0.531	0.823	0.526	0.529	0.462	0.750	0.604

N, Number of specimens genotyped; *A*, number of alleles; A_{14}, average number of alleles from 1000 subsamples of 14 individuals shuffled (sampling without replacement); *R*, allelic size range in bp; *S*, size in bp of the most common allele; *F*, frequency of the most common allele; H_O, observed heterozygosity; H_E, expected heterozygosity; *after correction for null alleles using MICRO-CHECKER version 2.2; † in bold significant, $P < 0.05$, deviations from HWE using GENEPOP version 3.3; ‡ in bold significant, $P < 0.01$, deviations from HWE using GENEPOP 3.3.