

## ORIGINAL ARTICLE

# The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula

C Cunha<sup>1,2</sup>, I Doadrio<sup>2</sup>, J Abrantes<sup>1</sup> and MM Coelho<sup>1</sup><sup>1</sup>Faculdade de Ciências de Lisboa, Centro de Biologia Ambiental, Universidade de Lisboa, Lisbon, Portugal and <sup>2</sup>Museo Nacional de Ciencias Naturales, CSIC, Departamento de Biodiversidad y Biología Evolutiva, C/José Gutiérrez Abascal, Madrid, Spain

Understanding the population structure, population dynamics and processes that give rise to polyploidy and helps to maintain it is central to our knowledge of the evolution of asexual vertebrates. Previous studies revealed high genetic diversity and several reproductive pathways in the southern populations of the *Squalius alburnoides* hybrid complex. In contrast, lower genetic variability and the associated limited chance of introducing new genetic combinations may threaten the survival of the northern Mondego populations. We analysed the genetic diversity and structure of nine populations of *S. alburnoides* in the Iberian Peninsula using microsatellite loci to provide further insights on the evolutionary history of this complex. Special attention was given to the less-studied northern populations (Mondego and Douro basins). Marked population structure, a high frequency of private alleles and a high diversity of some biotypes in the

Douro basin indicate that some northern populations may not be at high risk of extinction, contrary to what was expected. The genetic diversity found in the northern Douro populations contradicts the general trend of remarkable genetic impoverishment northwards that occurs in other species and regions. The results indicate the possible existence of a glacial refugium in the Rabaçal River, corroborating findings in other species of this region. Historical events seem to have affected the geographical patterns of genetic variability found among and within the northern and southern populations of this complex and contributed to different patterns of genome composition. Therefore, historical events might have a major role in the long-term persistence of some polyploid hybrid taxa.

Heredity (2011) 106, 100–112; doi:10.1038/hdy.2010.70; published online 9 June 2010

**Keywords:** hybridization; polyploidy; microsatellites; genetic diversity; population structure; glacial refugium

## Introduction

Understanding the evolution of asexual vertebrates requires knowledge of their genetic population, structure and dynamics. To gain this knowledge, it is necessary to understand the influence of biotype composition, reproductive mechanisms and historical processes in shaping current population patterns. Among the vertebrates, approximately 90 species of fish, amphibians and reptiles are asexual (Janko *et al.*, 2007 and references therein) and many show hybridization and polyploidy (Schultz, 1969). Asexual species have difficulty adapting to environmental change and may become extinct because of a decline in fitness, related to the accumulation of slightly deleterious mutations, the lack of recombination (Muller's ratchet) and/or ecological processes, such as Red-Queen dynamics (Muller, 1932; Konfrashov, 1988; Butlin, 2002; Loewe and Lamatsch, 2008). At least for a short term, some asexual species combat these debilitating effects through (i) multiple hybridization events, (ii) high heterozygosity buffering the effect of gene

redundancy on mutations, (iii) recurrent hybridization involving parental genotypes that increases the genetic diversity of the new species and (iv) recombination (Vrijenhoek, 1994; Comai, 2005; Bi *et al.*, 2008). Moreover, tolerance to increases or decreases in ploidy level (Leitch and Leitch, 2008) allows easy incorporation of an additional dose of proteins provided by a third set of chromosomes (Loewe and Lamatsch, 2008). These occurrences, combined with several types of reproductive mechanisms, may explain the persistence of asexual organisms over time (Bogart *et al.*, 2007; Lampert *et al.*, 2008). However, the existence of these compensatory mechanisms to generate the variation needed for adaptation to changing environments does not mean that asexual taxa will avoid extinction in all cases (Loewe and Lamatsch, 2008).

The *Squalius alburnoides* complex displays a number of characteristics that make it an extraordinary example of viability and evolutionary success among polyploid hybrid taxa.

This Iberian cyprinid fish has descended from an interspecific hybridization between *Squalius pyrenaicus* females (P genome) and males of an extinct species (A genome) that has been placed in a sister taxon of *Anaocypris hispanica* (for example, Crespo-López *et al.*, 2007). Sympatric bisexual *Squalius* species act as sperm donors (*S. pyrenaicus* in the southern basins and

Correspondence: Dr C Cunha, Museo Nacional de Ciencias Naturales, CSIC, Departamento de Biodiversidad y Biología Evolutiva, C/José Gutiérrez Abascal, 2, Madrid 28006, Spain.

E-mail: mcnj676@mncn.csic.es

Received 2 September 2009; revised 28 April 2010; accepted 29 April 2010; published online 9 June 2010

*S. carolitertii* in the northern ones—C genome), contributing new genetic material (Alves *et al.*, 2001). In contrast to other asexual species, the *S. alburnoides* complex is not strictly asexual (Cunha *et al.*, 2008 and references therein). In fact, fertile females and males are found, and often there are functional pathways towards sexual reproduction involving recombination, meiosis and syngamy.

Diploid (for example, CA, PA), triploid (for example, CAA, PAA) and tetraploid (for example, CCAA, PPAA) biotypes are found in *S. alburnoides*, including both females and males that are fertile, with a combination of several reproductive modes. In southern populations, triploid females show meiotic hybridogenesis with random segregation and recombination between non-excluded genomes, producing haploid and, more rarely, diploid gametes. The diploid females always produce diploid hybrid eggs, although a few may develop by gynogenesis (3%). The diploid eggs yield triploid progeny after fertilization (Alves *et al.*, 2001). As in other asexual complexes (*Ambystoma*: Bogart and Licht, 1986; Bogart, 1989; *Rana*: Hotz *et al.*, 1992; *Phoxinus*: Goddard and Schultz, 1993), the *S. alburnoides* complex has regenerated and maintained the extinct parental species genotype (AA, all males) through the fertilization of A oocytes (produced by PAA females) by reduced A sperm (produced by AA males) (Alves *et al.*, 2002). This AA genotype is apparently absent from the northern populations. The reproductive mechanisms described in the northern populations include hybridogenesis in hybrid diploids and triploids from the Douro basin (Carmona *et al.*, 1997) and meiotic hybridogenesis in triploids from the Mondego basin (Pala and Coelho, 2005).

Most analyses of mitochondrial DNA sequences and allozyme markers have indicated multiple, unidirectional, independent hybridization origins of the *S. alburnoides* complex (Carmona *et al.*, 1997; Cunha *et al.*, 2004; Alves *et al.*, 1997a,b); however, results differ regarding the number of hybridization origins. Alves *et al.* (1997b) suggest two hybridization origins (Guadiana-Tagus and Sado), whereas Cunha *et al.* (2004) suggest five historically independent hybridization events in the following areas: the Alagón Tagus tributary-Douro, Mondego-Tagus, Guadiana-Guadalquivir, Sado, and Quarteira, with subsequent dispersal by colonization at the end of the Pleistocene. Recently, however, a study based on partial cytochrome *b* sequences (Sousa-Santos *et al.*, 2007) interpreted haplotype sharing among some populations of *S. alburnoides* as a consequence of a recent single origin for the complex (700 000 years ago) in the Upper Tagus-Guadiana area, followed by dispersion. Therefore, the precise number and time of hybridization origins of the *S. alburnoides* complex is still up for debate.

Previous studies used microsatellites to analyse patterns of genetic diversity and similarity between *S. alburnoides* biotypes with different ploidy levels and their parental species (Pala and Coelho, 2005; Crespo-López *et al.*, 2007). These studies revealed the high genetic diversity of southern Portuguese populations from the Tagus and Guadiana basins and confirmed the intricate pathways of genetic exchange that contribute to the continuous shifting and maintenance of biotype diversity. However, *S. alburnoides* populations from the Mondego River, a small, independent Portuguese basin

situated in the northern range of this complex, showed much lower genetic variability than southern populations. AA males do not exist in the Mondego populations, limiting the chance of introducing new genetic combinations in the A genome. The loss of genetic diversity in these populations may threaten their survival (Pala and Coelho, 2005).

The distinct reproductive mechanisms, biotype population compositions and potentially different hybridization origins within the distribution range of this complex provide an excellent system with which to infer the relative influence of contemporary (for example, genetic drift and genetic flow) and historical factors (for example, glaciations and past fragmentation) on the long-term persistence of *S. alburnoides* in the Iberian Peninsula, as well as on the population structure of polyploid hybrid complexes. Moreover, freshwater systems provide exceptional opportunities to study the distribution of genetic diversity, in part because drainage structure can restrict gene flow among populations in different basins and preserve historical 'prints' of colonization events (Suk and Neff, 2009).

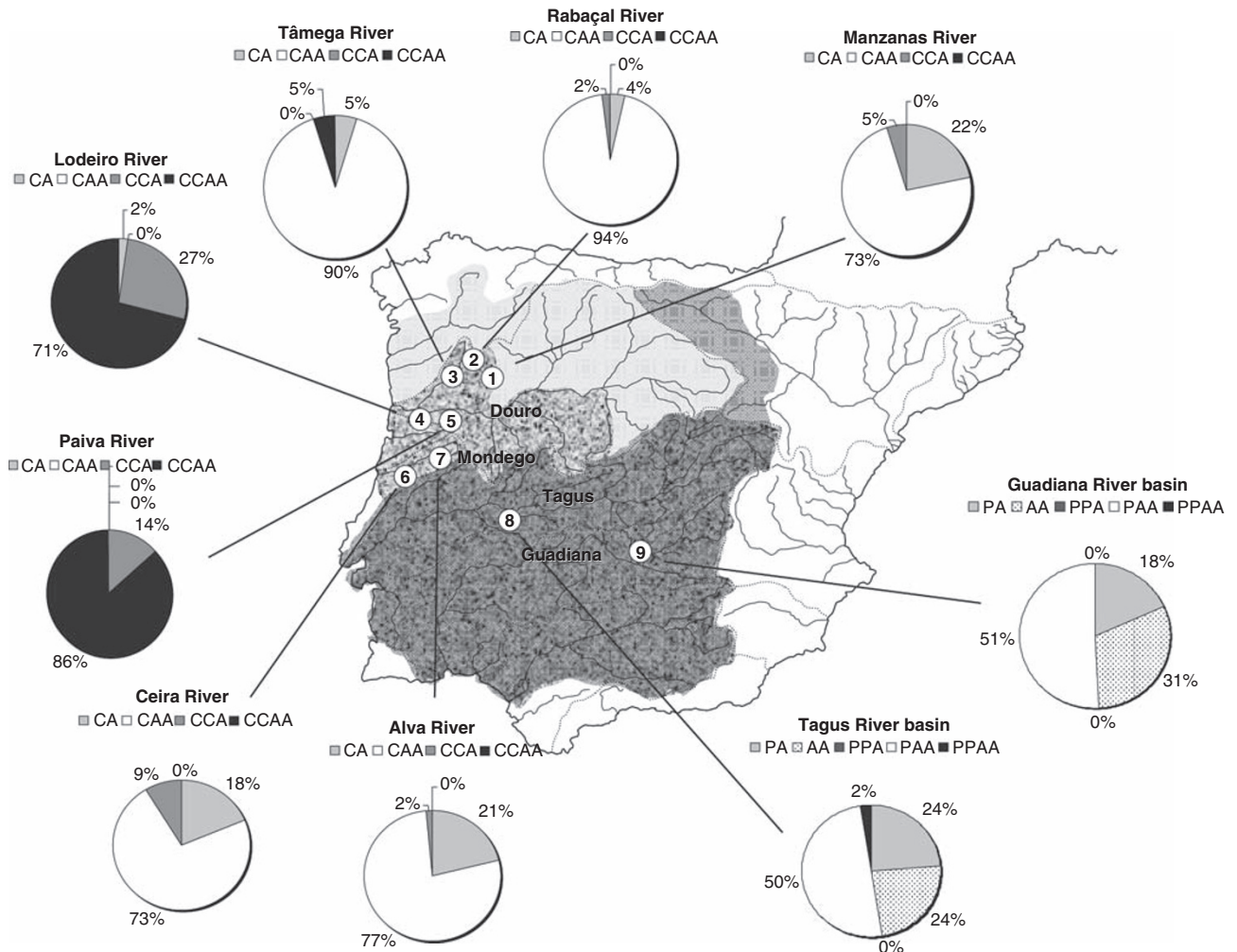
Here, for the first time, we use microsatellite DNA markers to examine the biotype composition, genetic diversity and population structure of *S. alburnoides* from the Douro River basin. The Douro is the largest basin in the Iberian Peninsula and the longest river in the northern range of *S. alburnoides*, where it lives in sympatry with *S. carolitertii*. We also analysed other northern and southern *S. alburnoides* populations, as well as their sperm donors (*S. carolitertii* and *S. pyrenaicus*), to investigate the evolutionary history of polyploid hybrid *S. alburnoides* populations. We aim to understand the effect of contemporary and historical factors on the population genetics of this asexual complex and to tentatively explore its hybrid origins and history of colonization. To achieve these goals, our data were compared with previous analyses based on mitochondrial DNA and microsatellites. For instance, if *S. alburnoides* from the south colonized the northern Iberian Peninsula, we would expect a loss of genetic variability during this process. Moreover, if the colonization of the north were relatively recent, as proposed by some authors, less population structure in the north would be predicted.

## Materials and methods

### Sample collection and laboratory procedures

A total of 701 specimens were obtained from the locations indicated in Figure 1: 83 *S. carolitertii* (9 Manzanás, 24 Rabaçal, 15 Tâmega, 24 Lodeiro, 11 Paiva) and 248 *S. alburnoides* (36 Manzanás, 54 Rabaçal, 39 Tâmega, 96 Lodeiro, 23 Paiva) from the Douro basin; 41 *S. carolitertii* (29 Ceira, 12 Alva) and 157 *S. alburnoides* (92 Ceira, 65 Alva) from the Mondego basin; 20 *S. pyrenaicus* and 43 *S. alburnoides* from the Tagus basin (Almonte); and 32 *S. pyrenaicus* and 48 *S. alburnoides* from the Guadiana basin (Zujar). We also examined a population of *A. hispanica* (29), the closest relation to the paternal ancestor of *S. alburnoides*, from the only basin where this species still exists, Guadiana (Zujar).

The ploidy level determination of all specimens was made using flow cytometry of blood cells as described by



**Figure 1** Map of the Iberian Peninsula showing the collection sites and distribution area of the *S. alburnoides* complex. The textured area represents the distributional range of the complex; light grey indicates the distribution of the northern *S. carolitertii* (sperm donor); dark grey indicates the distribution of the southern *S. pyrenaicus* (sperm donor). The pies illustrate the percentage of each *S. alburnoides* biotype in the populations sampled: (1) Manzanas River, (2) Rabaçal River, (3) Tâmega River, (4) Lodeiro River, (5) Paiva River, (6) Ceira River, (7) Alva River, (8) Almonte River (Tagus) and (9) Zujar River (Guadiana).

Collares-Pereira and Moreira da Costa (1999). Blood samples were stabilized in buffer and immediately frozen at  $-80^{\circ}\text{C}$ . Chicken erythrocytes were used as both an internal and external standard. The sex of each fish was determined by examining the gonads under a dissecting microscope.

For microsatellite analyses, total DNA was extracted from fins preserved in absolute ethanol or from frozen muscle tissue using the standard procedures outlined by Sambrook *et al.* (1989). Eight microsatellites were investigated for *S. alburnoides*: *n7k4*, *n7j4*, *e2f8*, *e1g6* (Mesquita *et al.*, 2003; Pala and Coelho, 2005), *lco1*, *lco3*, *lco4* and *lco5* (Turner *et al.*, 2004). PCR reactions were performed as described by Pala and Coelho (2005) except for *LCO1*, which was amplified in  $10\ \mu\text{l}$  volumes containing 25–50 ng of template DNA, 0.2 mM of each dNTP, 2 mM  $\text{MgCl}_2$ , 1 U Taq DNA polymerase (Invitrogen, Paisley, UK) and  $0.2\ \mu\text{M}$  of each primer. PCR products were electrophoresed in a CEQ 2000XL automatic sequencer (Beckman Coulter, Brea, CA, USA) and the molecular weights of alleles determined using CEQ fragment analysis software.

The information obtained on microsatellite loci and ploidy levels allowed us to identify the genome copy number of the intergeneric hybrid complex (for example, CAA or CCA, PAA or PPA) through the identification of specific alleles from the A, C and P genomes. Previous work has shown this method to be accurate for *S. alburnoides* (Pala and Coelho, 2005; Crespo-López *et al.*, 2007; Cunha *et al.*, 2008) and for other asexual complexes (Christiansen, 2005; Lampert *et al.*, 2006; Ramsden *et al.*, 2006). The assignment of alleles to each species was confirmed with the analysis of hybrids, especially diploids, in which the alleles attributed to each genome (C, P and A) should be present. Data sets from Paiva, Lodeiro and Ceira, Alva populations were compiled from Cunha *et al.* (2008) and Pala and Coelho (2005), respectively. The *LCO1* locus was not available for the Mondego populations and was thus excluded from further analyses. A contingency  $\chi^2$ -test was performed using SPSS Statistics 17.0 software (SPSS Inc., Chicago, IL, USA) to investigate whether biotype frequencies differed among basins.

### Statistical analyses

The genetic diversity of the *S. alburnoides* biotypes was measured with FDASH (Obbard *et al.*, 2006), which is suitable for the analysis of allopolyploids in which allele dosage cannot always be determined. The observed heterozygosity (Ho), number of phenotypes (NoPhen), average number of unshared alleles between each pair of individuals within populations ( $H'_S$ ) and diversity of allelic phenotypes ( $H'^{\text{phen}}$ ) within samples, averaged across samples (Obbard *et al.*, 2006) were estimated.

Possible scoring errors due to stuttering, large allele dropout or the occurrence of null alleles were assessed using Micro-Checker 2.2 (Oosterhout *et al.*, 2004) for *S. carolitertii*, *S. pyrenaicus*, *A. hispanica* and for the reconstituted parental species (AA males).

The following analyses were performed for all progenitor genomes of *S. alburnoides*, the A, C and P genomes. In the allotetraploids, the two genomes were analysed separately because they contain genetic information arising from the two progenitor species. In triploids, only the homospecific genome was analysed. To eliminate the effect of sample size, unbiased allelic richness and private allelic richness were estimated using the rarefaction method implemented in HP-RARE 1.0 (Kalinowski, 2005) for the C, P and A genomes. Intra-population genetic diversity of the three nuclear genomes of *S. alburnoides* was evaluated by estimating gene diversity and observed heterozygosity using GENEPOP 3.4 (Raymond and Rousset, 1995). This software was also used to test for deviations from Hardy–Weinberg equilibrium, applying the Exact test with the default settings of the Markov chain Monte Carlo procedure.

Kruskal–Wallis, Mann–Whitney *U* and Wilcoxon tests were used to determine whether the measures of genetic diversity differed between subpopulations. These analyses were performed with SPSS Statistics 17.0 software (SPSS Inc., Chicago, IL, USA). The modified false discovery rate procedure by Benjamini and Yekutieli (2001) was used to correct multiple comparisons.

To examine the extent of genetic differentiation and structuring, multiple approaches were applied to the A genome because it is the only genome present across the entire distribution of *S. alburnoides*. The A genome has been maintained since the formation of the complex, allowing comparisons between all populations and inferences about historical processes to be made. Genetic differentiation between pairs of populations was quantified through pairwise  $F_{ST}$  estimates based on variance in allelic frequencies using FSTAT 2.9.3 (Goudet, 2001). The interpretation of the  $F_{ST}$  values from multi-allelic data is problematic because their maximum values depend on the amount of within-population variation and even in the absence of any shared allele often fails to reach the theoretical maximum of 1 (Hedrick, 2005; Meirmans, 2006). We therefore calculated pairwise Theta' ( $\Theta'_{ST}$ ) following Meirmans (2006), using the program Recode-Data (Meirmans, 2006), an estimator that is highly comparable with  $G'_{ST}$ . This standardized measure ranges from 0 to 1 and therefore makes interpretation of the degree of subdivision much easier and facilitates comparing results among studies. A Mantel test for matrix correspondence was used to test for correlation between the pairwise  $F_{ST}$  and Theta' estimates (similar to  $G'_{ST}$ ).

Genetic differentiation was also estimated using mutational differences among alleles, assuming the

stepwise mutation model by computing pairwise standardized  $R_{ST}$  estimates with RSTCALC (Goodman, 1997). We examined the effect of allele sizes on the levels of genetic structuring following Hardy *et al.* (2003). This test compares observed estimates of  $R_{ST}$  with simulated values of  $\rho R_{ST}$ , in which allele size classes have been randomized, resulting in measures analogous to  $F_{ST}$ . The tests were performed with 10 000 permutations using Spagedi 1.2 (Hardy and Vekemans, 2002). The value of  $R_{ST}$  is mainly determined by extra mutations that accumulate within each population. The opportunity for such mutations will largely depend on the time since population divergence. The  $R_{ST}$  estimates will generally be larger than  $F_{ST}$  estimates when divergence time has been sufficient for mutations to accumulate. In contrast, both parameters should yield relatively similar values when the divergence time has been too short for neutral mutations to rise to high frequencies. In such cases, genetic drift and gene flow are likely to be more important than mutation in determining the extent of population divergence.

The population structure was further analysed without *a priori* clustering of individuals into populations using the Bayesian program Structure (Pritchard *et al.*, 2000), by varying the number of clusters in the data set (*K*) and assigning admixture proportions of each individual to these clusters. Ten independent runs were carried out for each value of *K* (1 to 9) clusters. Markov Chain Monte Carlo was run for  $10^6$  iterations after burn-in periods of  $10^5$  iterations. The best value of *K* was inferred from  $\ln \text{Pr}[X/K]$  (Pritchard *et al.*, 2000) and the modal value  $\Delta K$  (Evanno *et al.*, 2005). A Bayesian assignment test was performed using GeneClass 2.0 (Piry *et al.*, 2004) software, which gives the probability of each individual belonging to the reference population. The computation followed the Bayesian method of Rannala and Mountain (1997), on a simulation of 10 000 individuals.

To understand the hierarchical partitioning of genetic structure, an analysis of molecular variance was performed using Arlequin 3.0 (Excoffier *et al.*, 2005). Groups were defined to assess population configurations and their geographical distribution.

To estimate isolation by distance, the genetic distances of  $R_{ST}/(1-R_{ST})$  and  $F_{ST}/(1-F_{ST})$  were regressed against the minimum stream distance between populations. Significance of the relationship was evaluated by a Mantel test with 10 000 permutations performed in the Isolde program included in Genepop.

Genetic relationships among polyploid populations and sperm donors, avoiding the effect of different ploidy levels, were determined as  $D_{TL}$  distances (Tomiuk and Loeschcke, 1991) estimated using PopDist (Guldstrandtsen *et al.*, 2000). Genetic relationships among populations of diploid and triploid biotypes and the A genome were inferred from the Edwards Cavalli–Sforza– $D_{CE}$  distance (Cavalli-Sforza and Edwards, 1967) and through shared allele distance, DAS (Sai-Chakraborty and Jin, 1993). These distances were calculated with the Populations 1.2.28 software (Langella, 2002). The neighbour-joining clustering method was used to construct the phylogenetic trees and the tree-topology support was assessed by bootstrapping (1000 iterations).

Recent population bottlenecks were inferred by the program Bottleneck (Piry *et al.*, 1999). As is recommended when analysing a small number of microsatel-

lites, the Wilcoxon rank-sign test was used to test the significance of heterozygote excess fewer than three different models of microsatellite evolution, the infinite allele model, the stepwise mutation model and a two-phase model.

## Results

### Biotype composition

Microsatellites proved to be a powerful tool to identify the various *S. alburnoides* biotypes (Figure 1). Genotypes could be determined for 94% of the individuals. The loci *LCO3*, *LCO4*, *E1G6* and *LCO1* were extremely useful for discriminating the complete genotypes in diploid and polyploid biotypes.

The populations distributed across the geographical ranges of the two sperm donors were generally composed of the same biotypes. However, there were no diploid nuclear non-hybrid males (AA males) in the northern populations, and no tetraploid or diploid hybrid males were found in the Guadiana basin (see Figure 1). Differences in the frequency and distribution of biotypes were also observed between river basins (Figure 1).

According to flow cytometer and microsatellite data, Douro populations (except Lodeiro and Paiva) and Mondego populations were mainly composed of CAA triploid females (>70%, Figure 1). However, the Mondego and Douro basins showed significant differences in population composition (Figure 1) ( $\chi^2 = 90.7$ ,  $P < 0.001$ ). The Douro region had a significantly lower ( $P < 0.05$ ) percentage of diploid forms (mainly males) with only 7%, while the Mondego region had around 20% of diploids (mainly males). The proportion of CCA biotypes in the Douro basin (13%) was significantly larger ( $P < 0.05$ ) than observed in the Mondego basin (6%). However, the most marked difference between the Douro and Mondego basins was the absence of tetraploids from the Mondego, whereas in the Douro, the Lodeiro and Paiva populations were mainly composed of symmetric tetraploid individuals (71 and 86%, respectively).

### Sex ratio

A sex ratio deviation was observed among *S. alburnoides* triploids: CAA/PAA were mostly females and CCA were mostly males (Supplementary Table S1). This finding suggests a relation between genome dosage and sex. However, it is not known how sex is determined in *S. alburnoides*. Previous studies indicated that the mechanism of sex determination cannot be fully explained either by female or male heterogamety (Alves *et al.*, 1998, 1999) apparently being strictly a genetic mechanism (Pala *et al.*, 2009). It seems that sex ratio in triploid forms is influenced by other *Squalius* genomes (C and P), either as a consequence of lower fitness of the AA, CAA/PAA males or CCA/PPA females and/or by the mechanism of sex determination.

### Genetic diversity and population structure

Micro-Checker analyses revealed no evidence of scoring errors in any loci except *E2F8*, which was thus omitted from the analysis. Therefore, a set of six microsatellites was used for further analysis. Overall, the triploids

showed significantly higher phenotype diversity than the hybrid diploids and tetraploids (Mann–Whitney's test,  $P < 0.025$ ). Significantly higher diversity of phenotypes was observed in triploids from Rabaçal (Mann–Whitney test,  $P < 0.041$ ) and Manzanás (Mann–Whitney test,  $P < 0.033$ ) in contrast to hybrid diploids (Supplementary Table S1). In most sampled populations with CAA, this biotype showed the highest  $H^{phen}$  and  $H'_S$  values (Supplementary Table S1).

The observed heterozygosity value of 1.0 per locus indicated a hybrid origin of the biotypes. Values lower than 1.0 were occasionally observed. Since no null alleles were detected and recombination between genomes was unlikely because of the observed reproductive modes, the values below 1.0 were probably because of the presence of shared alleles between genomes.

In all, 110 alleles were found for the A genome: 48 alleles in the northern populations and 79 alleles in the southern populations (with sharing of alleles between populations). Significant differences were found among river basins for all genetic diversity parameters analysed (Kruskal–Wallis test for allelic richness, private allelic richness and gene diversity,  $P < 0.011$  in all cases). The southern river basins showed higher genetic diversity than the northern ones in all parameters analysed (Table 1) and the differences between these regions were significant (Mann–Whitney *U* tests, all  $P < 0.002$ ). For the C and P genomes, 46 and 61 alleles were found in northern and southern populations, respectively. Among the Douro populations, Rabaçal showed significantly higher allelic richness, private alleles and gene diversity in the A genome than all other populations (Wilcoxon signed-rank tests, all  $P < 0.05$ ). Rabaçal also showed significantly higher allelic richness, private alleles and gene diversity in the A genome than the two Mondego populations (Wilcoxon signed-rank tests, both  $P < 0.046$ ), whereas no significant differences were found in the Tagus and Guadiana basins. For the C genome, the genetic diversity of the Rabaçal was significantly higher than the Lodeiro and Paiva (Wilcoxon signed-rank tests,  $P < 0.046$ ), whereas there were no significant differences in the Mondego populations. The Paiva population showed significantly lower genetic diversity values than the other populations for the C and A genomes (Mann–Whitney *U* tests, all  $P < 0.035$ ).

After false discovery rate correction a significant deviation from Hardy–Weinberg equilibrium in the C genome was noted for Alva and Tâmega because of a heterozygote deficit across all loci ( $P < 0.041$ ). For the A genome, Tâmega, Rabaçal, Tagus and Guadiana showed significant deviation from Hardy–Weinberg equilibrium across all loci ( $P < 0.022$ ) as the result of a heterozygote deficit. The heterozygote deficit found in the A genome was probably a consequence of asexual reproductive mechanisms (Menken *et al.*, 1995). The deficit detected across all loci ( $P < 0.001$ ) of the C genome in the Alva and Tâmega populations could have been the consequence of a Wahlund effect.

The pairwise standardized genetic differentiation values (highly comparable to  $G'_{ST}$ ) were in general higher than the  $F_{ST}$  ones; however, they showed the same pattern (Supplementary Table S2). In fact, a significant correlation was found between these two genetic differentiation measures ( $r = 0.925$ ,  $P < 0.001$ ) and thus only  $F_{ST}$  values were used subsequently.

**Table 1** Intra-population genetic diversity of the biotypes in the *S. alburnoides* complex

River basin	Populations	C/P genome		Private allelic richness		Gene diversity (He)	A genome		Private allelic richness		Gene diversity (He)
		Allelic richness (A)		Total	Average across loci		Allelic richness (A)		Total	Average across loci	
		Total	Average across loci				Total	Average across loci			
Douro	Manzanas	9.66	1.61	0.53	0.09	0.27	10.66	1.78	1.69	0.28	0.23
	Rabaçal	13.79	2.30	2.82	0.47	0.44	15.88	2.65	4.45	0.74	0.50
	Tâmega	10.55	1.76	0.85	0.14	0.31	11.87	1.98	1.97	0.33	0.35
	Paiva	7.51	0.91	0.00	0.00	0.10	6.91	1.15	0.21	0.04	0.06
	Lodeiro	8.28	1.39	0.10	0.02	0.15	6.83	1.14	0.12	0.02	0.05
Mondego	Ceira	13.81	2.30	1.43	0.24	0.43	8.14	1.36	0.22	0.04	0.17
	Alva	12.39	2.07	1.35	0.23	0.41	9.78	1.63	0.66	0.11	0.26
Tagus	Almonte	12.03	2.01	5.77	0.96	0.34	17.21	2.87	7.82	1.30	0.57
Guadiana	Zujar	18.75	3.12	12.49	2.08	0.57	21.47	3.58	12.02	2.01	0.67

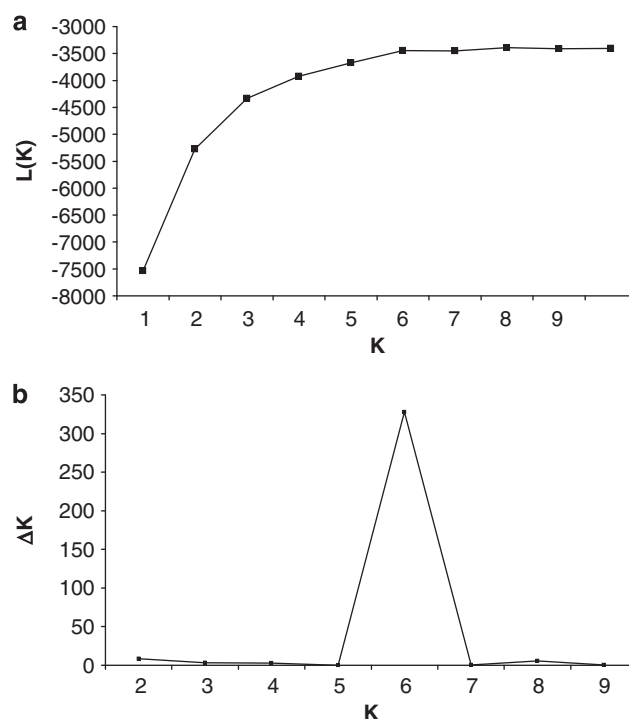
**Table 2** Pairwise genetic distance matrices for the A genome of *S. alburnoides* complex populations based on  $F_{ST}$  (below diagonal) and  $R_{ST}$  (above diagonal) estimators

	Do—Tamega	Do—Rabaçal	Do—Paiva	Do—Manzanas	Do—Lodeiro	Mo—Ceira	Mo—Alva	Tg—Almonte	Gd—Zujar
Do—Tamega		0.047*	0.272*	0.006	0.322*	0.800*	0.829*	0.732*	0.741*
Do—Rabaçal	0.120*		0.165*	0.055*	0.181*	0.446*	0.491*	0.410*	0.524*
Do—Paiva	0.302*	0.359*		0.360*	0.000	0.899*	0.920*	0.812*	0.796*
Do—Manzanas	0.089*	0.184*	0.548*		0.406*	0.798*	0.824*	0.726*	0.731*
Do—Lodeiro	0.416*	0.475*	0.000	0.657*		0.910*	0.930*	0.820*	0.800*
Mo—Ceira	0.679*	0.556*	0.843*	0.747*	0.881*		0.156*	0.776*	0.685*
Mo—Alva	0.630*	0.512*	0.771*	0.689*	0.832*	0.230*		0.779*	0.661*
Tg—Almonte	0.436*	0.352*	0.571*	0.493*	0.704*	0.555*	0.473*		0.438*
Gd—Zujar	0.384*	0.297*	0.504*	0.432*	0.651*	0.490*	0.422*	0.199*	

Do, Douro R. basin; Gd, Guadiana R. basin; Mo, Mondego R. basin; Tg, Tagus R. basin. \* $P < 0.001$ .

The null hypothesis of  $R_{ST} = F_{ST}$  over all loci was rejected ( $R_{ST} = 0.894 > \rho R_{ST} = 0.595$ , 95% confidence interval 0.267–0.863,  $P < 0.001$ ), suggesting that mutations contributed to genetic differentiation and that the mutation process at least partially followed a stepwise mutation model. The significant global  $R_{ST}$  ( $P < 0.0001$ ) over all loci suggested the population structure in *S. alburnoides*. The  $R_{ST}$  estimates were significantly higher than the  $F_{ST}$  estimates among basins (Wilcoxon signed-rank tests,  $P < 0.014$ ) and lower within the Douro and Mondego basins (Table 2). All pairwise comparisons made among Douro populations except the Lodeiro and Paiva showed significant  $F_{ST}$  values. Within the Douro basin, significant  $F_{ST}$  values ranged between 0.089 and 0.657 ( $P < 0.001$ ). The highest  $F_{ST}$  value was between Lodeiro and Manzanas, whereas the lowest differentiation was observed between Manzanas and Tâmega. Lodeiro and Paiva showed no significant differentiation ( $F_{ST} = 0.000$ ,  $P > 0.05$ ).

Bayesian clustering analyses detected the highest likelihood and a modal value of  $\Delta K$  for the model  $K = 6$  (Figure 2), and the probability of individuals' assignment was high ( $P > 0.8$ ). At values of  $K = 6$ , most individuals from a location were clustered into their *a priori* population of origin. The exceptions were three single clusters of Lodeiro and Paiva, Tâmega and Manzanas, and Ceira and Alva. The Bayesian assignment test assigned 67.6% of the individuals to their own source locations and no individuals were assigned to river basins outside their sources.

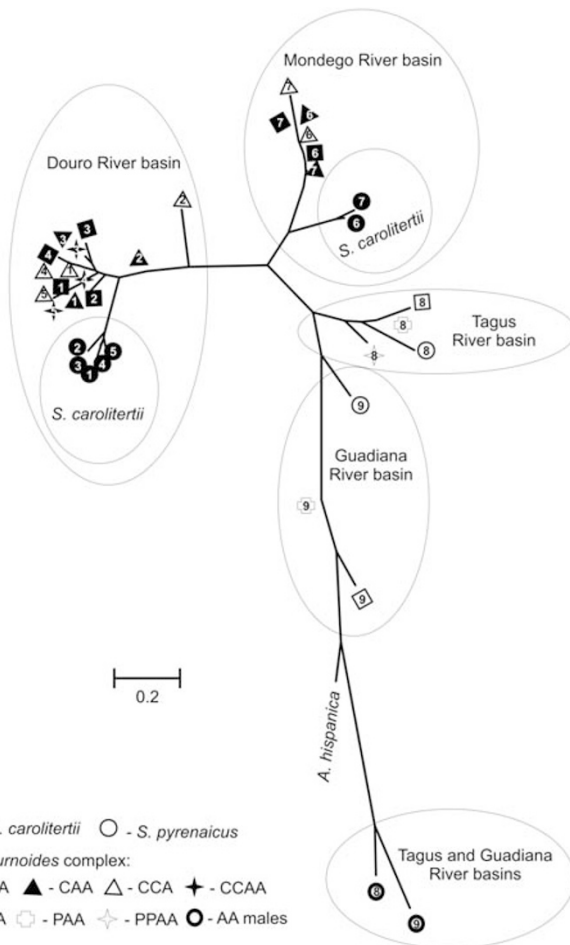


**Figure 2** Bayesian clustering of *S. alburnoides* samples (A genome). (a) Log probability of data,  $L(K)$ . (b) *Ad hoc* statistic  $\Delta K$  based on the rate of change in the log probability, as a function of  $K$  over 20 runs for each  $K$  value (putative different clusters tested).

**Table 3** Hierarchical analysis of molecular variance (AMOVA) for four alternative groupings of populations, based on the *A* genome

Structure tested	Among groups			Within groups			Among populations		
	%	P-value	$F_{CT}$	%	P-value	$F_{SC}$	%	P	$F_{ST}$
K = 6: Rabaçal vs (Tâmega, Manzanias) vs (Lodeiro, Paiva) vs Mondego vs Tagus vs Guadiana Douro vs Mondego vs Tagus vs Guadiana	51.4	<0.001	0.514	5.07	<0.001	0.1043	43.57	<0.001	0.563
Douro vs (Mondego, Tagus) vs Guadiana (Douro, Mondego, Tagus) vs Guadiana North vs South	47.4	<0.001	0.474	14.3	<0.001	0.272	38.3	<0.001	0.617
	40.0	<0.001	0.400	21.8	<0.001	0.364	38.2	<0.001	0.618
	0	>0.05	0	58.9	<0.001	0.559	46.5	<0.001	0.535
	10.5	>0.05	0.105	48.1	<0.001	0.537	41.4	<0.001	0.586

Percentage of total variance explained (%) and fixation indices are given for three hierarchical levels.



**Figure 3** Neighbour-joining tree based on the  $D_{TL}$  distance for diploids and polyploids, their two sympatric sperm donors (*S. carolitertii*, *S. pyrenaicus*) and *A. hispanica*, a sister taxon of the extinct paternal ancestor of *S. alburnoides*.

The hierarchical analysis of molecular variance revealed that the proportion of among-group variance was higher when populations were divided into six and four groups: (1) Rabaçal, (2) Manzanias-Tâmega, (3) Loudeiro-Paiva, (4) Ceira-Alva, (5) Tagus and (6) Guadiana basins; and (1) Douro, (2) Mondego, (3) Tagus and (4) Guadiana basins (Table 3). For the above groups, 51.4 and 47.4% of the variance was distributed within populations, respectively, and only 5.7 and 14.3% among populations, respectively.

The neighbour-joining tree built using the genetic distance  $D_{TL}$  clearly revealed clusters with a geographical pattern for *S. carolitertii*, *S. pyrenaicus* and *S. alburnoides* biotypes. The reconstituted parental species (AA) are in a cluster close to the *A. hispanica* cluster, as was expected, because *A. hispanica* is a sister taxon to the extinct paternal ancestor of *S. alburnoides* (Figure 3).

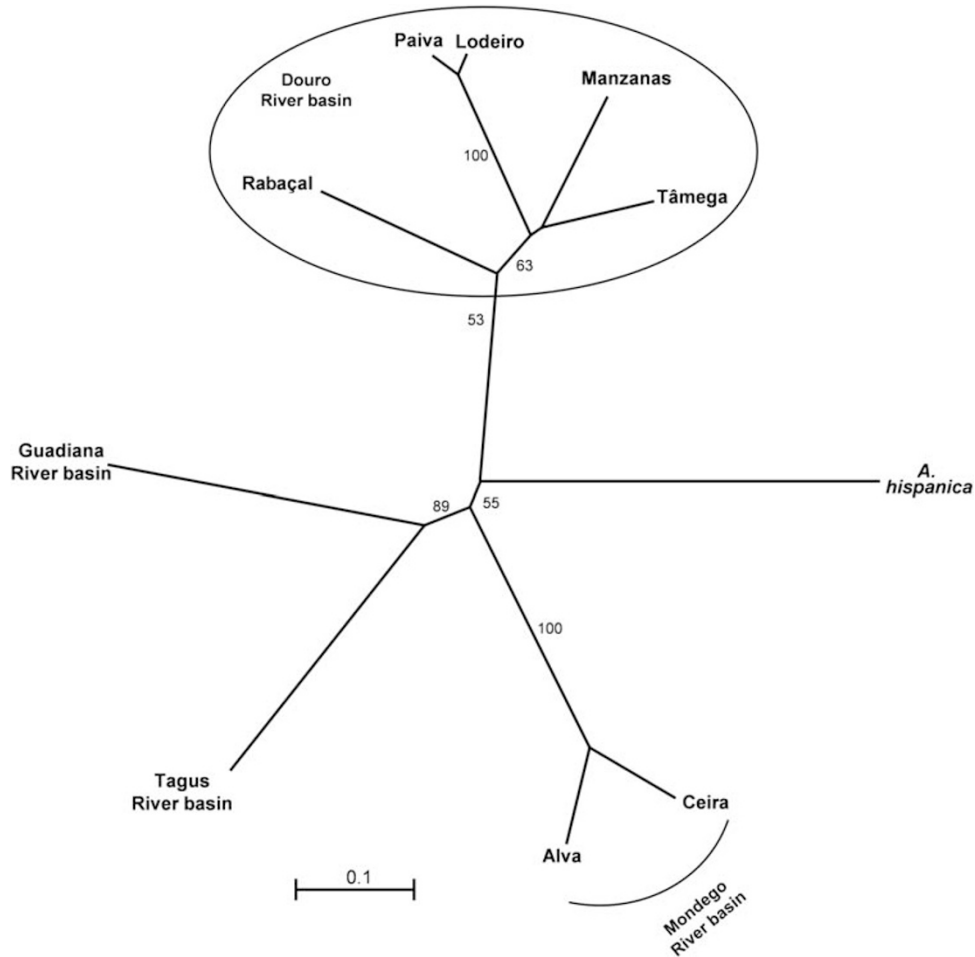
Neighbour-joining trees constructed using  $D_{AS}$  (not shown) and  $D_{CE}$  genetic distances revealed identical topologies for the *A* genome, indicating a geographical structure similar to the one obtained above (Figure 4). The exceptions were trees obtained for triploids, which showed a clear effect of the copy number of each genome instead of a geographical pattern. For instance, CCA biotypes from the Douro and Mondego basins were more closely related than the CCA and CAA biotypes within each basin (Supplementary Figure S2c).

The Mantel test showed no significant correlation between genetic distances ( $R_{ST}$  and  $F_{ST}$ ) and geographical distances [ $R_{ST}$  ( $r=0.140$ ,  $R^2=0.019$ ,  $P=0.075$ );  $F_{ST}$  ( $r=0.046$ ,  $R^2=0.002$ ,  $P=0.196$ )] (Supplementary Figure S4). Bottleneck tests showed no evidence of recent population bottlenecks in any of the populations.

## Discussion

Asexual taxa are supposed to accumulate deleterious mutations, lose uncorrupted genotypes by chance and have limited ability to adapt to rapid environmental changes. However, 'little sex' seems to be sufficient for avoiding the accumulation of mutations and for maintaining genetic variability (Hurst and Peck, 1996). As such, strictly asexual taxa are very rare and most do not completely avoid sexual reproduction, which could be the key to their long persistence. There are many examples of hybrid-polyploid complexes that are not strictly asexual. For instance, *Poecilia formosa* shows paternal leakage of undamaged DNA from its sexual sister species (Loewe and Lamatsch, 2008), and the *Ambystoma* complex maintains various intergenomic exchanges between *A. laterale* and *A. jeffersonianum* genomes in unisexual individuals (Bi and Bogart, 2006; Bi *et al.*, 2007). In the allopolyploid *S. alburnoides* complex there are several reproductive mechanisms that, in association with different ploidy levels, can generate new genetic combinations.

The breeding dynamics of *S. alburnoides* allow bidirectional movement of alleles between different polyploid levels, which avoids the fatal accumulation of deleterious



**Figure 4** Neighbour-joining tree, based on  $D_{CE}$  microsatellite genetic distances for the A genome, among populations of *S. alburnoides* from the Douro, Mondego, Tagus and Guadiana River basins. Branch lengths are proportional to the genetic distances between populations. Numbers indicate nodes with bootstrap support higher than 50% in 1000 replications.

mutations and allows the incorporation of novel genetic variability, making this complex an example of viability and evolutionary success among polyploid hybrid taxa. There are, however, geographical variations in the mechanisms shown by *S. alburnoides* to overcome the disadvantages of asexuality (Alves *et al.*, 1998; Pala and Coelho, 2005; Crespo-López *et al.*, 2006; Cunha *et al.*, 2008). An uncommonly high number of pathways involving recombination, meiosis and gamete syngamy generate new genetic material in the southern populations (reviewed by Alves *et al.*, 2002). Our results confirmed high genetic diversity in the southern populations because of the incorporation of P alleles from the bisexual *S. pyrenaicus*. Also, meiotic hybridogenesis and gametogenesis in AA males, that allow for recombination of A alleles, are highly relevant for the maintenance of genetic variability. The meiotic hybridogenesis has also been observed in other asexual taxa, such as *Rana esculenta* (Graf and Pelaz, 1989) and *Misgurnus anguillicaudatus* (Morishima *et al.*, 2008). It causes the diploidization of triploids, which can explain their persistence in nature (Otto, 2007; Pala *et al.*, 2008).

The success of triploids in nature is mainly related to the effect of heterosis. The maintenance of a permanent heterozygote condition results from the absence of intergenomic recombination and gene redundancy (Comai,

2005), which may be a crucial intermediate step in the formation of even-ploidy sexual lineages (Mable, 2003). The *S. alburnoides* complex, like other hybrid complexes such as *Ambystoma*, *Rana esculenta* and *Phoxinus* (Berger, 1973; Bogart *et al.*, 1985; Goddard *et al.*, 1989), has a prevalence of triploid females in most populations, accompanied by the loss of normal sexual reproduction. Triploidy in *S. alburnoides* was shown to be an intermediate step for the formation of tetraploid lineages (Alves *et al.*, 2004; Cunha *et al.*, 2008) and, consequently, it has evolutionary importance. In fact, the Douro basin is the first place where established populations mainly composed of symmetric tetraploids of both sexes (1:1 sex ratio) were observed (Lodeiro and Paiva, Figure 1, Supplementary Table 1) in a process of speciation adapted to specific habitat conditions (Cunha *et al.*, 2008, 2009). In other Douro populations, such as in Tâmega, the occurrence of symmetrical tetraploid individuals may also indicate the stabilization of new bisexual polyploid lineages. The return to a balanced genome may allow adaptation and long-term evolution of these populations. Tetraploidization has apparently also arisen in the Tagus River basin, either through triploidy or diploidy (Alves *et al.*, 1999, 2004). However, there was no evidence of self-sustaining populations of tetraploids in this southern basin. This is probably



because of environmental conditions that characterize this region, namely a heterogeneous annual hydrological cycle characteristic of the circum-Mediterranean region and habitat loss and hydrological disturbance that increases competition among the *S. alburnoides* biotypes for the sympatric sperm donor, *S. pyrenaicus*. The pathways for evolution and speciation by tetraploidization are also apparently unfeasible in the Mondego and Guadiana Rivers because this ploidy was not found in any individuals from this region. Such findings and future studies will help to further explore the persistence of triploid forms in nature and their importance in the maintenance of polyploid hybrid complexes.

The genetic diversity of the northern populations of *S. alburnoides* results from the incorporation of C alleles arising from the sexual *S. carolitertii* in each generation and the diversity of reproductive pathways. However, the low frequency of hybrid males and the absence of AA males make the northern populations more dependent on the sympatric sexual species, *S. carolitertii*. Thus, the northern hybrid populations have more competition with the sexual species that probably reduces their dispersal potential and settlement, similar to what has been proposed for the spiny loaches (Janko *et al.*, 2005). The absence of AA males and of symmetrical tetraploids from the Mondego populations (Figure 1) reduces the incorporation of new genetic material in the A genome, resulting in decreased genetic diversity compared with the southern populations and perhaps restricting their long term persistence (Pala and Coelho, 2005). In the Douro populations the only way to recruit new genetic combinations for the A genome is through gametogenesis in symmetric tetraploids because the diploid and the triploid hybrid females do not show meiotic hybridogenesis (Supplementary Figure S3) and AA males are absent. Therefore, much lower genetic diversity would be expected than in the southern and Mondego populations, resulting in reduced adaptive potential and viability. However, this trend was not observed, indicating that other processes such as historical factors can also influence the viability of northern populations.

The tertiary orogeny of the Iberian Peninsula and subsequent glaciations have led to important changes in climate and topography (Taberlet *et al.*, 1998; Loidi, 1999; Thompson, 1999). These factors had a significant impact on contemporary distribution patterns and genetic variation of a great number of species (for example, *Lacerta schreiberi*: Paulo *et al.*, 2002; Godinho *et al.*, 2008; *Chioglossa lusitanica*: Alexandrino *et al.*, 2000, 2007). Low mitochondrial DNA diversity has been found in fish populations from the northern Iberian Peninsula, such as *Pseudochondrostoma duriensis* and *S. carolitertii*, and in other northern fish populations that formerly inhabited glaciated regions (Brito *et al.*, 1997; Bernatchez and Wilson, 1998; Anger and Schlosser, 2007; Mesquita *et al.*, 2007; Aboim *et al.*, 2009). Microsatellite data indicate lower genetic variability in the northern *S. carolitertii* (C genome) compared with *S. pyrenaicus* (P genome) from the south, which may reflect the climatic fluctuations felt in the northern Iberian region during the Pleistocene glaciations (Williams *et al.*, 1988; Pérez-Alberti *et al.*, 2004). When temperatures decreased, bottleneck events may have led to high mortality and a drastic decrease of genetic diversity in the *S. carolitertii* populations.

The generation of genetic diversity in *S. alburnoides* can result from the accumulation of genetic differences among hybrids that derived from a common ancestor, or from multiple origins of hybrids arising from the sexual ancestors. Alves *et al.* (1997b) and Cunha *et al.* (2004) suggested multiple hybridization origins for the *S. alburnoides* complex, with subsequent dispersal by colonization at the end of the Pleistocene (Cunha *et al.*, 2004). However, a scenario of recent single origination (700 000 years ago) in the Upper Tagus-Guadiana area followed by dispersion has been recently suggested (Sousa-Santos *et al.*, 2007). These authors also inferred that the Mondego and Douro Rivers could have been colonized from the Tagus river around 0.05 and 0.01 million years ago, respectively. If *S. alburnoides* colonized the north of the Iberian Peninsula from the south, a loss of genetic variability would be expected during this process. Moreover, if the colonization of the north was relatively recent, less population structuring in the north would be expected because of the effects of shared polymorphisms. The geographical distribution of microsatellite allelic richness in *S. alburnoides* found in this study did not entirely fit a simple south-north colonization scenario, because allelic richness was not significantly lower in populations theoretically colonized more recently, as is true of the Douro basin. Moreover, we observed genetic structuring of river basins that was highly supported by overall  $R_{ST}$  and corroborated by Bayesian assignment tests, STRUCTURE and analysis of molecular variance results. Neighbour-joining phylogenies confirmed the phylogeographical structure of *S. alburnoides* obtained from the complete mitochondrial cytochrome *b* (Alves *et al.*, 1997b; Cunha *et al.*, 2004). The Mondego populations fit the trend of lower genetic diversity in the A genome when compared with the southern populations, possibly as a consequence of their colonization from the Tagus, supporting the hybridization origin scenarios suggested by Cunha *et al.* (2004) and Sousa-Santos *et al.* (2007). These populations also display fewer mechanisms for incorporating new genetic combinations in the paternal genome (A genome) than the southern populations. Colonization from the Tagus River would have caused founder effects and genetic drift, which could explain the reduced genetic variability of the Mondego basin populations and their high genetic differentiation from the other basins.

However, our findings at the edge of the northern distribution of the *S. alburnoides* complex (Douro basin) contradict the general trend observed in many species (for example, Hewitt, 1996; Paulo *et al.*, 2001; Horn *et al.*, 2006) of decreased genetic variability towards the north. The unusually high levels of genetic differentiation of Douro populations (except the Rabaçal) from the other basins, when combined with low levels of genetic diversity, are consistent with the idea that these populations were subject to harsh genetic founder effects followed by genetic drift. Nevertheless, it is worth noting that a correlation between genetic differentiation and geographical distance was not observed. This result does not imply, however, that isolation by distance would not be the best model to describe the genetic differentiation associated with the spatial distribution of these populations. The lack of correlation could be an artefact caused by the high genetic differentiation found within the Douro River basin. However, when we tested isolation

by distance between the Rabaçal River (the more genetically diverse population in the Douro basin) and the other basins, this model was again not supported (data not shown). Thus, the results suggest that differences in genetic variation found in *S. alburnoides* derived not only from genetic drift and mutation associated with their population dynamics, but also from historical events associated with range shifts and Pleistocene climatic fluctuations.

The populations within the Douro basin varied the most in allele frequencies (that is, they showed higher  $F_{ST}$  values than  $R_{ST}$ ), and they displayed a paucity of private alleles as they moved away from Rabaçal. The north-western Rabaçal River population showed high genetic diversity ( $A$ , private allelic richness and  $H_e$ ) (Table 1), similar to that shown by the southern populations, and reduced genetic variability was observed as we moved away from Rabaçal. The evolutionary divergence of populations can be influenced by the opposing effects of migration, which tends to homogenize populations, and genetic drift, which leads to increased population differentiation (Allendorf and Luikart, 2007). Our data seem to indicate that the genetic diversity of populations adjacent to Rabaçal, such as Tâmega and Manzanás, is because of the contributions from its genetic pool that decreased during successive founder events. The tetraploid populations of Lodeiro and Paiva showed strong evidence of founder events, including a reduction in the number of alleles and high frequencies of other alleles (Supplementary Figure S1). These populations seem to have been colonized from the neighbouring population of Tâmega, but this colonization was not recent because bottleneck tests did not detect recent founder events. Indeed, despite the high degree of genetic differentiation of these populations from the neighbouring population of Tâmega, the assignment tests indicated that individuals from these two populations could belong to Tâmega.

Higher levels of heterozygosity and a wider range of allele sizes in Rabaçal (Supplementary Figure S1) suggest more stable demographic conditions. Higher  $R_{ST}$  values (compared with the  $F_{ST}$  ones) were recorded for Rabaçal than the southern populations, indicating a sufficiently long period for the divergence of accumulated mutations. Therefore, our results suggest that Rabaçal constituted a northern Pleistocene refugium whose population re-colonized the western populations of the Douro basin after glacial retreat. The possible existence of a Douro refugium is also indicated by *Podarcis bocagei* (Sá-Sousa, 2001), *Iberolacerta* spp. (Godinho, 2003), other freshwater fish (Gómez and Lunt, 2006) and the presence of thermophilic plants in this region (Costa et al., 1999). The Douro region was intensely affected by Pleistocene glaciations, while the Rabaçal region was able to maintain warm conditions in some valleys and tributaries during the Würm glaciation (Costa et al., 1999). We therefore propose that *S. alburnoides* populations arose in Douro before or during the Pleistocene glacial cycles and that during glacial times they remained in a single refugium within the Rabaçal region.

Historical events have not only contributed to clear geographical patterns of genetic variability among and within the northern and southern populations of *S. alburnoides*, but they have also contributed to differential patterns of genome composition. Our findings

indicated the capacity of microsatellite analyses to identify recent historical re-colonization events by an ancestral population from at least one glacial refugium in the northern range of the *S. alburnoides* complex. As in other asexual taxa, the long-term persistence of *S. alburnoides* depends mostly on genetic diversity, and new hybridization events can help stop the decay of genome variability in *S. alburnoides* populations. However, not all populations have the same potential for survival because long-term viability of southern and northern populations depends not only on their genetic legacy, but also on population dynamics that may allow the introduction of fresh genes that can eventually lead to new speciation events. The combination of these processes may allow long-term survival of *S. alburnoides*, and therefore, understanding these contributions is central to our knowledge of this and other hybrid complexes.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We wish to thank A Perdices, H Gante and S Perea for their help with sample collection and A Perdices and L Costa for their help with morphological and sex identification of specimens. We also thank C Luís for constructive comments that greatly helped improve the manuscript. We thank A Burton, E Bailey and S Davis for revising the English. This study was financially supported by grant POCTI/BSE/40868/2001 to MM Coelho, grant REM2001-0662/GLO-07M/0109/2002 to I Doadrio and a PhD grant awarded to C Cunha (FCT-SFRH/BD/8637/2002).

## References

- Aboim MA, Cunha C, Coelho MM (2009). Redistribution of the geographical ranges of the Iberian cyprinid genus *Pseudochondrostoma* based on a phylogenetic analysis: implications for the historical rearrangements of the north-western Iberian drainages. *J Fish Biol* 74: 1337–1346.
- Alexandrino J, Froufe E, Arnetzen JW, Ferrand N (2000). Genetic subdivision, glacial refugia and postglacial recolonization in the golden-striped salamander, *Chioglossa lusitanica* (Amphibia: Urodela). *Mol Ecol* 9: 771–781.
- Alexandrino J, Teixeira J, Arntzen JW, Ferrand N (2007). Historical biogeography and conservation of the golden-striped salamander (*Chioglossa lusitanica*) in northwestern Iberia: integrating ecological, phenotypic and phylogeographic data. In: Weiss S, Ferrand N (eds). *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity*. Springer: Dordrecht, pp 189–205.
- Allendorf FW, Luikart G (2007). *Conservation and the Genetics of Populations*. Blackwell Publishing: Malden, USA.
- Alves MJ, Coelho MM, Collares-Pereira MJ (1997a). The *Rutilus alburnoides* complex (Cyprinidae): evidence for hybrid origin. *J Zool Syst Evol Res* 35: 1–10.
- Alves MJ, Coelho MM, Collares-Pereira MJ (1998). Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. *Mol Biol Evol* 15: 1233–1242.
- Alves MJ, Coelho MM, Collares-Pereira MJ (2001). Evolution in action through hybridization and polyploidy in an Iberian freshwater fish: a genetic review. *Genetica* 111: 375–385.

- Alves MJ, Coelho MM, Collares-Pereira MJ, Dowling TE (1997b). Maternal ancestry of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae) as determined by analysis of cytochrome b sequences. *Evolution* **51**: 1584–1592.
- Alves MJ, Coelho MM, Próspero MI, Collares-Pereira MJ (1999). Production of fertile unreduced sperm by hybrids males of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): an alternative route to genome tetraploidization in unisexuals. *Genetics* **151**: 277–283.
- Alves MJ, Collares-Pereira MJ, Dowling TE, Coelho MM (2002). The genetics of maintenance of an all-male lineage in the *Squalius alburnoides* complex. *J Fish Biol* **60**: 649–662.
- Alves MJ, Gromicho M, Collares-Pereira MJ, Crespo-López E, Coelho MM (2004). Simultaneous production of triploid and haploid eggs by triploid *Squalius alburnoides* (Teleostei: Cyprinidae). *J Exp Zool* **301A**: 552–558.
- Anger B, Schlosser IJ (2007). The origin of *Phoxinus eos-neogaeus* unisexual hybrids. *Mol Ecol* **16**: 4562–4571.
- Benjamini Y, Yekutieli D (2001). The control of the False Discovery Rate in multiple testing under dependency. *Ann Stat* **29**: 1165–1188.
- Berger L (1973). Systematics and hybridization in European green frogs of *Rana esculenta* complex. *J Herpetol* **7**: 1–10.
- Bernatchez L, Wilson CC (1998). Comparative phylogeography of nearctic and paleartic fishes. *Mol Ecol* **7**: 431–452.
- Bi K, Bogart JP (2006). Identification of intergenomic recombinations in unisexual salamanders of the genus *Ambystoma* by genomic *in situ* hybridization. *Cytogenet Genome Res* **112**: 307–312.
- Bi K, Bogart JP, Fu J (2007). Intergenomic translocations in unisexual salamanders of the genus *Ambystoma* (Amphibia, Caudata). *Cytogenet Genome Res* **116**: 289–297.
- Bi K, Bogart JP, Fu J (2008). The prevalence of genome replacement in unisexual salamanders of the genus *Ambystoma* (Amphibia, Caudata) revealed by nuclear gene genealogy. *BMC Evol Biol* **8**: 158.
- Bogart JP (1989). A mechanism for interspecific gene exchange via all-female salamander hybrids. In: Dawley RM, Bogart JP (eds). *Evolution and Ecology of Unisexual Vertebrates*. New York State Museum Bulletin 466: Albany, New York, pp 170–179.
- Bogart JP, Bi K, Fu J, Nobel DWA, Niedzwiecki J (2007). Unisexual salamanders (genus *Ambystoma*) present a new reproductive mode for eukaryotes. *Genome* **50**: 119–136.
- Bogart JP, Licht LE (1986). Reproduction and the origin of polyploids in hybrid salamanders of the genus *Ambystoma*. *Can J Genet Cytol* **28**: 605–617.
- Bogart JP, Litch LE, Oldham MJ, Darbyshire SJ (1985). Electrophoretic identification of *Ambystoma laterale* and *Ambystoma texanum* as well as their diploid and triploid interspecific hybrid (Amphibia:Caudata) on Pelee Island, Ontario. *Can J Zool* **63**: 340–347.
- Brito RM, Briolay J, Galtier N, Bouvet Y, Coelho MM (1997). Phylogenetic relationships within genus *Leuciscus* (Pisces, Cyprinidae) in Portuguese freshwaters based on mitochondrial DNA cytochrome b sequences. *Mol Phylogenet Evol* **8**: 435–442.
- Butlin R (2002). Opinion – evolution of sex: The costs and benefits of sex: new insights from old asexual lineages. *Nat Rev Genet* **3**: 311–317.
- Carmona JA, Sanjur OI, Doadrio I, Machordom A, Vrijenhoek RC (1997). Hybridogenetic reproduction and maternal ancestry of polyploid Iberian fish: the *Tropidophoxinellus alburnoides* complex. *Genetics* **146**: 983–993.
- Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* **19**: 233–257.
- Christiansen DG (2005). A microsatellite-based method for genotyping diploid and triploid water frogs of the *Rana esculenta* hybrid complex. *Mol Ecol Notes* **5**: 190–193.
- Collares-Pereira MJ, Moreira da Costa L (1999). Intraspecific and interspecific genome size variation in Iberian Cyprinidae and the problem of diploidy and polyploidy, with review of genome sizes within the family. *Folia Zool* **48**: 61–76.
- Comai L (2005). The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**: 836–846.
- Costa JC, Aguiar C, Capelo JH, Lousã M, Neto C (1999). Biogeografia de Portugal Continental. *Quercetia* **0**: 5–56.
- Crespo-López ME, Duarte TL, Dowling TE, Coelho MM (2006). Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. *Zoology* **109**: 277–286.
- Crespo-López ME, Pala I, Duarte TL, Dowling TE, Coelho MM (2007). Genetic structure of the diploid-polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. *J Fish Biol* **71**: 423–436.
- Cunha C, Bastir M, Coelho MM, Doadrio I (2009). Body shape evolution among ploidy levels of the *Squalius alburnoides* hybrid complex (Teleostei, Cyprinidae). *J Evol Biol* **22**: 718–728.
- Cunha C, Coelho MM, Carmona JA, Doadrio I (2004). Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridization events. *Mol Ecol* **13**: 2807–2817.
- Cunha C, Doadrio I, Coelho MM (2008). Speciation towards tetraploidisation after intermediate processes of non-sexual reproduction. *Philos Trans Roy Soc B* **363**: 2921–2929.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver.3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 47–50.
- Goddard KA, Dawley RM, Dowling TE (1989). Origin and genetic relationships of diploid, triploid and diploid-triploid mosaic biotypes in the *Phoxinus eos-neogaeus* unisexual complex. In: Dawley RM, Bogart JP (eds). *Evolution and Ecology of Unisexual Vertebrates*. New York State Museum Bulletin 466: Albany, New York, pp 268–280.
- Goddard KA, Schultz RJ (1993). Aclonal reproduction by polyploid members of the clonal hybrid species *Phoxinus eos-neogaeus* (Cyprinidae). *Copeia* **1993**: 650–660.
- Godinho MR (2003). *História evolutiva do lagarto-de-água, Lacerta schreiberi* PhD thesis, University of Lisbon.
- Godinho R, Crespo EG, Ferrand N (2008). The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Mol Ecol* **17**: 4670–4683.
- Gómez A, Lunt DH (2006). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N (eds). *Phylogeography of Southern European Refugia*. Springer-Verlag: Berlin, pp 155–188.
- Goodman SJ (1997). Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and a determining their significance. *Mol Ecol* **6**: 881–885.
- Goudet J (2001). *FSTAT Version 2.9.3, a program to estimate and test gene diversities and fixation indices* Available from <http://www.unil.ch/izea/software/fstat.html>.
- Graf J-D, Pelaz MP (1989). Evolutionary genetics of the *Rana esculenta* complex. In: Dawley RM, Bogart JP (eds). *Evolution and Ecology of Unisexual Vertebrates*. New York State Museum Bulletin 466: Albany, New York, pp 289–302.
- Guldstrandtsen B, Tomiuk J, Loeschcke V (2000). Popdist, version 1.1.1: a program to calculate population genetic distance and identity measures. *J Hered* **91**: 178–179.
- Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**: 1467–1482.
- Hardy OJ, Vekemans X (2002). Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.

- Hedrick P (2005). A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.
- Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc* **58**: 247–276.
- Horn A, Roux-Morabito G, Lieutier F, Kerdelhue C (2006). Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae). *Mol Ecol* **15**: 1603–1615.
- Hotz H, Beerli P, Spolsky C (1992). Mitochondrial DNA reveals formation of nonhybrids frogs by natural matings between hemiclinal hybrids. *Mol Bio Evol* **9**: 610–620.
- Hurst LH, Peck JR (1996). Recent advances in understanding the evolution and maintenance of sex. *Trends Ecol Evol* **11**: A46–A52.
- Janko K, Bohan J, Lamatch D, Flajshans M (2007). The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis: Teleostei), and their ability to establish successful clonal lineages—on the evolution of polyploidy in asexual vertebrates. *Genetica* **131**: 185–194.
- Janko K, Culling MA, Ráb P, Kotlík P (2005). Ice age cloning—comparison of the Quaternary evolutionary histories of sexual and clonal forms of spiny loaches (*Cobitis*; Teleostei) using the analysis of mitochondrial DNA variation. *Mol Ecol* **14**: 2991–3004.
- Kalinowski ST (2005). HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol Ecol Notes* **5**: 187–189.
- Konfrashov AS (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**: 435–440.
- Lampert KP, Lamatsch DK, Fischer P, Scharl M (2008). A tetraploid Amazon molly, *Poecilia formosa*. *J Hered* **99**: 223–226.
- Lampert KP, Lamatsch DK, Schories S, Hopf A, Garcia de León FJ, Scharl M (2006). Microsatellites for the gynogenetic Amazon molly, *Poecilia formosa*: useful tools for detection of mutation rate, ploidy determination and overall genetic diversity. *J Genet* **85**: 67–71.
- Langella O (2002). *Populations 1.2.28. Logiciel de génétique des populations. Laboratoire Populations, Génétique et Évolution, CNRS, Gif-sur-Yvette* Available from <http://www.pge.cnrs-gif.fr/bioinfo/populations/>.
- Leitch AR, Leitch IJ (2008). Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–4883.
- Loewe L, Lamatsch DK (2008). Muller's ratchet may threaten the Amazon molly and other ancient asexuals. *BMC Evol Biol* **8**: 88–108.
- Loidi J (1999). General description of the Iberian Peninsula: substrate and relief. *Itinera Geobot* **13**: 17–40.
- Mable BK (2003). Breaking down taxonomic barriers in polyploidy research. *Trends Plant Sci* **8**: 582–590.
- Meirmans PG (2006). Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* **60**: 2399–2402.
- Menken SB, Smith E, Den Nus HCM (1995). Genetical populations structure in plants: gene flow between diploid sexual and triploid asexual dandelions (*Taraxacum section ruderalia*). *Evolution* **49**: 1108–1118.
- Mesquita N, Cunha C, Carvalho G, Coelho MM (2007). Comparative phylogeography of endemic cyprinids in the south-west Iberian Peninsula: evidence for a new ichthyogeographic area. *J Fish Biol* **71**: 45–75.
- Mesquita N, Cunha C, Hänfling B, Carvalho GR, Ze-Ze L, Tenreiro R et al. (2003). Isolation and characterization of polymorphic microsatellite loci in the endemic Portuguese freshwater fish *Squalius aradensis* (Cyprinidae). *Mol Ecol Notes* **3**: 572–574.
- Morishima K, Yoshikawa H, Arai K (2008). Meiotic hybridogenesis in triploid *Misgurnus* loach derived from a clonal lineage. *Heredity* **100**: 581–586.
- Muller HJ (1932). Some genetic aspects of sex. *Am Nat* **66**: 118–138.
- Obbard DJ, Harris SA, Pannell JR (2006). Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity* **97**: 296–303.
- Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004). Micro-checker: software for identifying and correcting genotype errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.
- Otto SP (2007). The evolutionary consequences of polyploidy. *Cell* **131**: 452–462.
- Pala I, Scharl M, Thorsteinsdóttir S, Coelho MM (2009). Sex determination in the *Squalius alburnoides* complex: an initial characterization of sex cascade elements in the context of a hybrid polyploid genome. *PLoS ONE* **4**: e6401. doi:10.1371/journal.pone.0006401.
- Pala I, Coelho MM (2005). Contrasting views over a hybrid complex: between speciation and evolutionary 'dead – end'. *Gene* **347**: 283–294.
- Pala I, Coelho MM, Scharl M (2008). Dosage compensation by gene-copy silencing in a triploid hybrid fish. *Curr Biol* **18**: 1344–1348.
- Paulo OS, Dias C, Bruford MW, Jordan WC, Nichols RA (2001). The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proc Biol Sci* **268**: 1625–1630.
- Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002). Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard. *Mol Ecol* **11**: 809–819.
- Pérez-Alberti A, Valcárcel-Díaz M, Blanco-Chao R (2004). Pleistocene glaciation in Spain. In: Ehlers J, Gibbard PL (eds). *Quaternary Glaciations – Extent and Chronology. Part I: Europe*. Elsevier: Amsterdam. pp 389–394.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004). GeneClass2: a Software for genetic assignment and first-generation migrant detection. *J Hered* **95**: 536–539.
- Piry S, Luikart G, Cornuet JM (1999). Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* **90**: 502–503.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Ramsden C, Bériault K, Bogart JP (2006). A nonlethal method of identification of *Ambystoma laterale*, *A. jeffersonianum* and sympatric unisexuals. *Mol Ecol Notes* **6**: 261–264.
- Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* **94**: 9197–9201.
- Raymond M, Rousset F (1995). GENPOP (version 3.3): population genetics software for exact test and ecumenism. *J Hered* **86**: 248–249.
- Sai-Chakraborty R, Jin L (1993). A unified approach to study hypervariable polymorphism: statistical considerations of determining relatedness and population distances. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys A (eds). *DNA Fingerprinting*. Birkhauser Verlag: Basel, pp 153–175.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual* 2nd edn. Cold Spring Harbor Laboratory Press: New York.
- Sá-Sousa P (2001). Comparative chorology between *Podarcis bocagei* and *P. carbonellae* (Sauria: Lacertidae) in Portugal. *Rev Esp Herp* **15**: 85–97.
- Schultz RJ (1969). Hybridization, unisexuality, and polyploidy in the teleost *Poeciliopsis* (Poeciliidae) and other vertebrates. *Am Nat* **103**: 605–619.
- Sousa-Santos C, Collares-Pereira MJ, Almada V (2007). Reading the history of a hybrid fish complex from its molecular record. *Mol Phylogenet Evol* **45**: 981–996.
- Suk HY, Neff BD (2009). Microsatellite genetic differentiation among populations of the Trinidadian guppy. *Heredity* **102**: 425–434.

- Taberlet P, Fumagalli L, Wust-Saucy A-G (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* 7: 454–464.
- Thompson JD (1999). Population differentiation in Mediterranean plants: insights into colonization history and the evolution and conservation of endemic species. *Heredity* 82: 229–236.
- Tomiuk J, Loeschcke V (1991). A new measure of genetic identity between populations of asexual and sexual species. *Evolution* 45: 1685–1694.
- Turner TF, Dowling TE, Broughton RE, Gold JR (2004). Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily Leuciscinae). *Conserv Genet* 5: 279–281.
- Vrijenhoek RC (1994). Unisexual fish—model systems for studying ecology and evolution. *Ann Rev Ecol Syst* 25: 71–96.
- Williams M, Dunkerley D, De-Decker P, Kershaw P, Chappel J (1988). *Quaternary Environments*. Arnold Publishers: London.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)