Documenting the advantages and limitations of different classes of molecular markers in a well-established phylogeographic context: lessons from the Iberian endemic Golden-striped salamander, *Chioglossa lusitanica* (Caudata: Salamandridae)

FERNANDO SEQUEIRA¹²*, JOÃO ALEXANDRINO³, STEVEN WEISS⁴ and NUNO FERRAND¹²

¹CIBIOUP, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal
²Departamento de Zoologia e Antropologia, Faculdade de Ciências, Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal
³Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista, 13506-900 Rio Claro, São Paulo, Brazil
⁴Karl-Franzens University Graz, Institute of Zoology, Universitätsplatz, A-8010 Graz, Austria

Received 5 November 2007; accepted for publication 2 January 2008

Previous analyses of mitochondrial (mt)DNA and allozymes covering the range of the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*, suggested a Pleistocene split of the historical species distribution into two population units (north and south of the Mondego river), postglacial expansion into the northernmost extant range, and secondary contact with neutral diffusion of genes close to the Mondego river. We extended analysis of molecular variation over the species range using seven microsatellite loci and the nuclear β-fibrinogen intron 7 (β-fibint7). Both microsatellites and β-fibint7 showed moderate to high levels of population structure, concordant with patterns detected with mtDNA and allozymes; and a general pattern of isolation-by-distance, contrasting the marked differentiation of two population groups suggested by mtDNA and allozymes. Bayesian multilocus analyses showed contrasting results as populations north and south of the Douro river were clearly differentiated based on microsatellites, whereas allozymes revealed differentiation north and south of the Mondego river. Additionally, decreased microsatellite variability in the north supported the hypothesis of postglacial colonization of this region. The well-documented evolutionary history of *C. lusitanica* provides an excellent framework within which the advantages and limitations of different classes of markers can be evaluated in defining patterns of population substructure and inferring evolutionary processes across distinct spatio-temporal scales. The present study serves as a cautionary note for investigations that rely on a single type of molecular marker, especially when the organism under study exhibits a widespread distribution and complex natural history. © 2008 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2008, 95, 371–387.


INTRODUCTION

In recent decades, phylogeographical studies have flourished, aided by the development of new molecular techniques combined with advances in population genetic theory and the implementation of new analytical tools (Avise, 2000; Weiss & Ferrand, 2007). Although the use of uniparentally inherited single locus markers (chloroplast and mtDNA) is still the method of choice for most initial phylogeographic

*Corresponding author. E-mail: fsequeira@mail.icav.up.pt
surveys (Moore, 1995; Avise, 1998; Zhang & Hewitt, 2003; Weiss & Ferrand, 2007), it has become increasingly clear that such data sets may provide a very limited view of population history, or even be wholly misleading (Pamilo & Nei, 1988; Palumbi & Baker, 1994; Wu, 1991; Hare & Avise, 1998; Alves et al., 2006). Although many factors are at play, the stochastic nature of the coalescence process alone is thought to be the largest obstacle to obtaining an accurate reconstruction of population history from a single genetic marker (Edwards & Beerli, 2000; Broughton & Harrison, 2003; Jennings & Edwards, 2005).

Reliable inferences on population structure and evolutionary history should therefore be drawn from the analyses of several independent loci (Slatkin & Maddison, 1990; Karl & Avise, 1992; Hare, 2001). Although multi-locus analysis is rare and still wrought with analytical hurdles (Hey & Machado, 2003; Wakeley, 2004; Jennings & Edwards, 2005), one reasonable way to assess the reliability of inferences based on mitochondrial (mt)DNA is to compare phylogeographical patterns of multiple classes of markers such as allozymes, nuclear gene sequences and microsatellites. Relative to less variable nuclear and mitochondrial markers, microsatellites (i.e. simple-sequence repeat nuclear loci; Amos et al., 1996; Jarne & Lagoda, 1996) evolve rapidly and may be more useful in revealing fine-scale population structure and/or inferring evolutionary processes on smaller temporal scales (Estoup et al., 1998; Palo et al., 2004; Neumann et al., 2005). We carried out a multi-locus analysis of an organism with a confined distribution and a relatively simple albeit varied, historical biogeography, the golden-striped salamander, *Chioglossa lusitanica* (Bocage, 1864), to investigate the ability of different molecular markers to detect genetic structure at distinct spatial and temporal scales, as well as their degree of concordance in supporting an inferred evolutionary scenario.

*Chioglossa lusitanica* is a peculiar streamside salamander endemic to the northwestern Iberian Peninsula (Arntzen, 1999; Teixeira, Ferrand & Arntzen, 2001), where it is restricted to the banks of small temperate streams in hilly landscapes historically covered by deciduous forests (Arntzen, 1981; Zagwijn, 1992; Teixeira et al., 2001). Previous analyses of mitochondrial cytochrome *b* gene sequences and seven polymorphic allozyme loci demonstrated strong phylogeographic structure, revealing the existence of two genetically divergent population groups on either side of the Mondego River, in central Portugal (Alexandrino, Arntzen & Ferrand, 2002; Alexandrino et al., 2000). Approximate molecular clock calibrations for allozymes and the salamandrid cytochrome *b* gene suggested a Late Pliocene/Early Pleistocene genetic isolation. Vicariance across the Mondego River valley is supported by bioclimatic modelling (Teixeira et al., 2001; Teixeira & Arntzen, 2002). The data additionally suggested that the northernmost range of the distribution of *C. lusitanica* resulted from postglacial expansion. Presently, the two population groups coincide in areas adjacent to the Mondego River (Alexandrino et al., 2000; Alexandrino et al., 2002) with a resulting hybrid zone shaped by neutral admixture and, possibly, recurrent contacts throughout the Pleistocene (Sequeira et al., 2005a).

In the present study, we extend analyses of molecular variation in *C. lusitanica* to one single-copy nuclear intron and seven microsatellite loci (Sequeira et al., 2005b) to investigate the consistency of the following previously published inferences (Alexandrino et al., 2000, 2002): (1) subdivision of the historical range of the species into two genetically divergent population units; (2) occurrence of a contact zone between these two lineages; and (3) recent range expansion as a consequence of postglacial colonization. We expect this comparative analysis of multiple genetic markers to provide: (1) further insights into the relative roles of historical and contemporary processes in determining the present phylogeographic patterns of *C. lusitanica* and (2) a more generally applicable lesson concerning our ability to infer evolutionary scenarios of an organism with a complex history from a limited number of genetic markers.

**MATERIAL AND METHODS**

**COLLECTION OF SAMPLES AND MARKERS**

A total of 286 salamanders from 13 locations (15 to 27 individuals/location) were sampled between 1997 and 2003 from throughout the species distribution (Fig. 1). Tissue samples (tail tips) were preserved at −70 °C or in 70% ethanol. Whole genomic DNA extraction was conducted in accordance with previously described procedures (Sambrook, Fritsch & Maniatis, 1989).

Seven microsatellite loci (four tetra-, two tri-, and one dinucleotide) described by Sequeira et al. (2005b) and sequences of the entire intron 7 (716 bp) of the β-fibrinogen gene (β-fibint7) were chosen for analysis. Microsatellites were amplified with polymerase chain reaction (PCR) conditions described by Sequeira et al. (2005b). Most microsatellite genotyping was accomplished with fluorescently labelled primers, and an automated sequencer (ABI 3100; Perkin Elmer) using an internal size standard (Rox – genescan 350; Applied Biosystems), and analysed using GENESCAN software (Applied Biosystems), as described by Sequeira et al. (2005b). Genotyping of some samples was performed on 6% polyacrylamide gels and visualized with standard silver staining protocols. To ensure no bias in allele detection or sizing, several
individuals from each population were scored with both techniques.

The entire $\beta$-fibint7 (716 bp) was PCR amplified using a two-step protocol and two pairs of primers as described by Sequeira, Ferrand & Harris (2006). PCRs were performed in 10-µl reactions with Taq polymerase (Ecogen) using the conditions: initial denaturation of 94 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 40 s, annealing at 56 °C for 1 min, extension at 72 °C for 1:30 min, and a final extension of 72 °C (7 min). Amplified fragments were sequenced in both directions using the primers BFXF and BFXR on a 310 or 3100 Applied Biosystem DNA genotyper, following the ABI PRISM BigDye Terminator Cycle sequencing protocols. The sequences were aligned using BioEdit, version 5.0.9 (Hall, 1999). Gametic phase of the double-stranded sequences was either inferred through haplotype subtraction (Clark, 1990), or by the Amplification Refractory Mutation System (Newton et al., 1989). For example, because we were unable to determine the haplotypes from sequences heterozygous in both positions 38 and 394, a site-specific PCR primer (BFARM 5′-CTG CAC TGA CAT TCT ACC TCC ATG CTT GAG-3′) was designed to selectively amplify the haplotype with the nucleotide G in position 38. This site-specific primer was used for sequencing the products of the second PCR reaction, following the procedure described above.

DATA ANALYSES

Genetic diversity
For microsatellites, the mean number of alleles per locus, allele frequencies, allelic richness, observed ($H_0$) and expected ($H_E$) heterozygosities (Nei, 1978) were determined using the software FSTAT, version 2.9.3 (Goudet, 2001). For $\beta$-fibint7, the nucleotide diversity ($\pi$) haplotype diversity ($h$), and the average number of nucleotide differences among sequences ($K$) were calculated with DnaSP 4.1 software (Rozas et al., 2003).

Population and geographic structure
For comparisons among multiple markers, previously published allozyme and mtDNA data (Alexandrino
zyme data consisted of allele frequencies (total of 28 alleles) at seven polymorphic loci (PGM-1, PEP-A, PEP-B, PEP-C, PEP-D, ADH, and PGD) scored for 432 salamanders from 17 localities (Fig. 1) distributed across the entire species’ range (Alexandrino et al., 2000). mtDNA data include 124 sequences (30 distinct haplotypes) of a cytochrome b fragment, 700-bp long, sampled from the same 17 localities of mtDNA analysis (Alexandrino et al., 2002).

For microsatellites, deviation from Hardy–Weinberg equilibrium, genotypic linkage disequilibrium and pairwise genetic differentiation among population samples were analysed by Fisher’s exact test, implemented in GENEPOP 3.3 (Raymond & Rousset, 1995). All probability tests were based on Markov chain Monte Carlo (MCMC) simulations (Guo & Thompson, 1992; Raymond & Rousset, 1995) using default values and adjusting the critical probability for each test with a sequential Bonferroni correction (Rice, 1989). Population genetic differentiation as revealed by microsatellite loci was quantified by estimating \( F_{ST} \) (Weir & Cockerman, 1984) in GENETIX, version 4.01 (Belkhir et al., 2000), as opposed to the microsatellite-specific \( R_{ST} \) measure (Slatkin, 1995), because \( F_{ST} \) provides a more accurate estimation of population structure when both sample size and number of loci are relatively small (Gaggiotti et al., 1999). Statistical significance of \( F_{ST} \) values was tested using 1000 iterations, and 95% bootstrapped confidence intervals (CI) were used. For \( \beta \)-fibint7, we estimated \( \Phi_{ST} \) (Excoffier, Smouse & Quattro, 1992; Michealakis & Excoffier, 1996) from the pairwise nucleotide differences between haplotypes using ARLEQUIN, version 2.0 (Schneider, Roessli & Excoffier, 2000). \( \beta \)-fibint7 haplotype data were also coded as genotypic data to calculate \( F_{ST} \) estimates (Weir & Cockerman, 1984) in GENETIX.

Population structure as revealed by both microsatellite and allozyme loci was investigated using the Bayesian multilocus methodology implemented in STRUCTURE, version 2.0 (Pritchard, Stephens & Donnelly, 2000). The number of population clusters \( K \) was estimated assuming no a priori assignment of individuals to populations (USEPOPINFO = 0), and using the admixture model (NOADMIX = 0). Runs with \( K = 1-13 \) (microsatellites) and \( K = 1-17 \) (allozymes) were carried out with \( 5 \times 10^4 \) MCMC repetitions, discarding the first \( 5 \times 10^4 \) iterations as burn in, and repeated five times to assess the consistency of the results. The inference of optimal clustering was based on comparing, for each \( K \) simulated clusters, (1) the log likelihood \( \ln Pr (X|K) \) and (2) on the approach proposed by Evanno, Regnaut & Goudet (2005), a method that searches for a mode in the distribution of \( \Delta K \), a quantity related to the second order of change of the log likelihood of the data. The program DISTRACT (Rosenberg, 2002) was used to display individuals’ membership coefficients for each cluster.

Hierarchical analysis of molecular variance (AMOVA) was performed using ARLEQUIN to evaluate the merits of several hypotheses of population structure derived from: (1) previous mtDNA analyses (Alexandrino et al., 2002) and (2) our own Bayesian analyses of microsatellites and allozymes. AMOVAs were performed using the haplotype distances calculated as the number of restriction site differences between haplotypes for mtDNA data, and using allele frequencies, for microsatellites and allozymes.

Isolation-by-distance was examined through linear regression of pairwise population genetic distances against geographical distance with a Mantel non-parametric permutation test (Mantel, 1967) as implemented in GENETIX. For microsatellites and allozymes, the association between pairwise \( F_{ST} / (1 – F_{ST}) \) and the natural logarithm of the geographical distance were used, as suggested by Rousset & Raymond (1997). For \( \beta \)-fibint7 and mtDNA, genetic distances were derived from haplotype data \( (D_h, \text{net nucleotide divergence per site between populations, Nei, 1987}) \), using DnaSP 4.1 software. The hypothesis of decreased genetic diversity in northern populations was examined through linear regression of allelic richness, for microsatellites and allozymes, and of haplotype diversity, for \( \beta \)-fibint7 and mtDNA, positioned along a south–north axis. Computations were carried out using STATISTICA for windows, version 6.0 (StatSoft, 2001).

Pairwise population genetic distances, Nei’s modified chord distance \( D_h \) (Nei, Tajima & Tateno, 1983) and microsatellite-specific \( \delta^2 \) (Goldstein et al., 1995) genetic distances, for microsatellites, were used to construct Neighbour-joining (NJ) (Saitou & Nei, 1987) population trees with the software POPULATIONS, version 1.2.28 (http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php). Support for the tree topology was evaluated by bootstrapping with 1000 iterations over loci. The trees were viewed in the program TREEVIEW, version 1.6.6. (Page, 1996).

\( \beta \)-fibint7 genealogy

The minimum number of recombination events in the history of the sample, \( Rm \) (Hudson & Kaplan, 1985), and the number of pairs of sites showing four gametic types in the sampled sequences was estimated using the four-gamete test as implemented in the DnaSP 4.1 software. Tajima’s \( D \) (Tajima, 1989) and Fu’s \( F_{s} \) (Fu, 1997) tests were used to evaluate whether the \( \beta \)-fibint7 polymorphism conforms to a neutral model of evolution. These tests were performed either with all sequences, or excluding the putative recombinant haplotypes (see Results). Statistical significance was determined by comparing estimated values against a
distribution generated from 10,000 random samples under the hypothesis of selective neutrality and population equilibrium, with no recombination (Hudson, 1990), using the coalescent simulator in DnaSP.

Genealogical relationships among haplotypes were analysed by constructing a network with the Network 3.1 software (Röhl, 2000). Inference of the ancestral haplotype was qualitatively made through the use of sequences from a sister-species of *C. lusitanica*, *Mertensiella caucasica* (Veith et al., 1998; Sequeira et al., 2006). Additionally, we estimated the unrooted genealogical tree probabilities with the infinite-sites model (Griffiths & Tavaré, 1995) using the software GENETREE, version 9.0 (R.C. Griffiths; http://www.Stats.ox.ac.uk/~griff/software.html). Because the data did not allow differentiation between recombination and recurrent mutation at the origin of two haplotypes (see Results), two independent analyses in which we included each one of those haplotypes at a time were performed.

**RESULTS**

**GENETIC DIVERSITY**

*Microsatellites*

There were five to 24 alleles per locus (Table 1; see also Supporting Information, Fig. S1). The uncorrected mean number of alleles per population ranged from 2.9 (GE and SAL) to 8.4 (LO). Lower levels of allelic diversity adjusted for sample size (allelic richness) and expected heterozygosity were encountered in populations north of Douro River compared to more southern populations. Allelic richness ranged from 2.7 (SAL) to 7.5 (LO), whereas expected heterozygosity ranged from 0.273 (PO) to 0.756 (LO). Five populations (AC, CP, LO, VA and SAI) out of thirteen showed private alleles, with a mean frequency ranging from 0.019 at locus Cl17 (VA) to 0.119 at locus Cl19 (VA).

The number of private alleles per locus ranged from one (loci Cl17 and Cl145) to six (locus Cl19) (Table 1).

*β-fibint7*

A total of 636 bp were analyzed in 62 individuals (124 haplotype sequences), revealing three variable positions defining five distinct haplotypes (Fig. 2A, B, Table 2). Although the number of pairs of sites showing all four gametic types was one, and *Rm*, the minimum number of recombination events in the history of the sample, was also one, the low variability revealed at *β-fibint7* did not allow us to discriminate between those alternative hypotheses. An additional possibility is that both h4 and h5 result from the recurrence of the 38 C→G mutation in the background haplotypes h1 and h2, respectively (Fig. 2C). However, both their restricted geographical distribution and the fact that position 38 is not part of a hypermutable CpG motif (Cooper & Krawezak, 1993) make this hypothesis highly unlikely.

Overall haplotype diversity varied among populations, ranging from 0 in MO and SAL, to 0.71 in VA and BU populations. The overall nucleotide diversity was fairly low (π = 0.0015), ranging from 0 (MO and SAL) to 0.0017 (SAI and PO) (Table 3). Although Tajima’s *D* and Fu’s *F* indices were positive considering all sequences (1.268 and 0.548, respectively), and excluding the putative recombinant haplotype h4 (0.860 and 1.045, respectively) or h5 (1.076 and 1.296, respectively), the values did not differ significantly from the neutral expectation (*P* > 0.10).

**POPULATION AND GEOGRAPHIC STRUCTURE**

*Microsatellites*

There were no significant deviations from Hardy–Weinberg or linkage equilibria. All populations were significantly differentiated from each other (*P* < 0.001), with the exception of comparisons between GE, PO, and SAL populations. Pairwise *F*~ST~ was in the range 0–0.477 with an overall differentiation among populations of *F*~ST~ = 0.268 (95% CI = 0.147–0.418).

The Bayesian structure analysis did not reveal an unequivocal number of genetic clusters as Ln Pr (*s*) increased from *K* = 2 to *K* = 13 (Fig. 3A), and the application of the method described by Evanno et al. (2005) did not provide any obvious mode (results not shown). Although the highest value of Δ*K* was found at *K* = 2, it was clear from the analyses that biologically meaningful clusters were observed at higher values of *K*. Hence, we analysed the patterns of cluster composition examining the mean confidence assignment of all individuals (q) to their most probable cluster (an indicative measure of the assignment robustness) and the across population average of the maximum proportion of assignment of individuals from a breeding population to any simulated cluster (*Q*). Average *Q* is used here as a measure of homogeneity both within breeding populations and within simulated clusters. We examined clustering levels using cut-off points of average *Q* > 0.9 and average *Q* > 0.8. Considering *Q* > 0.9, we accepted clustering at *K* = 2 and *K* = 3 (Figs 3A, 4A), levels of *K* in which
Ln Pr (X/K) reached its maximum rate of increase, and the mean individual confidence assignment was \( q \approx 0.95 \). At \( K = 2 \), the two clusters separated the populations located north (VAL, GE, PO, and SAL) and south (MU, AC, CP, LO, VA, BU, SAI, CO, and MO) of the Douro river, respectively. At \( K = 3 \), populations south of Mondego River clustered, whereas the other two clusters corresponded to populations north and south of the Douro River. If we relaxed the cut-off point to average \( Q > 0.8 \), additionally structure units ranging from \( K = 4 \) through \( K = 7 \) were accepted (Fig. 3A). At \( K = 4 \), CO and MO populations formed a separate population cluster. At \( K = 5 \), some individuals from VAL were assigned with a probability higher than 0.9 to the CO population. At \( K = 6 \), populations south of Mondego separated into two clusters, one including CP and LO and the other populations located north of the Douro.

### Table 1. Genetic variation at seven microsatellite loci in 13 Chioglossa lusitanica populations

<table>
<thead>
<tr>
<th>Locus</th>
<th>Salo</th>
<th>PO</th>
<th>GE</th>
<th>VAL</th>
<th>MO</th>
<th>CO</th>
<th>SAI</th>
<th>BU</th>
<th>VA</th>
<th>LO</th>
<th>CP</th>
<th>AC</th>
<th>MU</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>25</td>
<td>21</td>
<td>24</td>
<td>22</td>
<td>23</td>
<td>22</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1.9</td>
<td>1.7</td>
<td>7.2</td>
<td>5.5</td>
<td>11.0</td>
<td>6.5</td>
</tr>
<tr>
<td>CI6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>20</td>
<td>25</td>
<td>21</td>
<td>25</td>
<td>22</td>
<td>25</td>
<td>22</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>5 (1)</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>3.7</td>
<td>5.6</td>
<td>4.3</td>
<td>4.3</td>
<td>6.7</td>
<td>6.2</td>
<td>7.4</td>
<td>4.7</td>
<td>3.0</td>
<td>4.5</td>
</tr>
<tr>
<td>CI17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>16</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>24</td>
<td>21</td>
<td>27</td>
<td>23</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>10 (1)</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>R</td>
<td>2.9</td>
<td>4.0</td>
<td>3.0</td>
<td>6.6</td>
<td>4.6</td>
<td>4.8</td>
<td>6.6</td>
<td>5.9</td>
<td>9.4</td>
<td>4.7</td>
<td>4.9</td>
<td>3.7</td>
<td>2.0</td>
</tr>
<tr>
<td>CI19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>21</td>
<td>16</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>24</td>
<td>21</td>
<td>26</td>
<td>22</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>14</td>
<td>14 (1)</td>
<td>13</td>
<td>15 (5)</td>
<td>15</td>
<td>14</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>R</td>
<td>4.5</td>
<td>6</td>
<td>4.8</td>
<td>9.3</td>
<td>7.2</td>
<td>12.5</td>
<td>13.0</td>
<td>11.6</td>
<td>13.5</td>
<td>12.9</td>
<td>12.6</td>
<td>10.7</td>
<td>9.6</td>
</tr>
<tr>
<td>CI39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>21</td>
<td>16</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>21 (1)</td>
<td>25</td>
<td>22</td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3 (1)</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.8</td>
<td>2.9</td>
<td>2.0</td>
<td>2.6</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>CI136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>24</td>
<td>21</td>
<td>26</td>
<td>22</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>11</td>
<td>12</td>
<td>12 (1)</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>R</td>
<td>4.9</td>
<td>7.0</td>
<td>6.0</td>
<td>5.7</td>
<td>10.5</td>
<td>10.5</td>
<td>11.4</td>
<td>9.6</td>
<td>8.5</td>
<td>10.6</td>
<td>10.2</td>
<td>10.0</td>
<td>4.7</td>
</tr>
<tr>
<td>CI145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>23</td>
<td>21</td>
<td>25</td>
<td>22</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>n</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>9 (1)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>R</td>
<td>3.7</td>
<td>4.0</td>
<td>3.0</td>
<td>4.0</td>
<td>5.8</td>
<td>6.5</td>
<td>7.6</td>
<td>6.7</td>
<td>4.9</td>
<td>7.4</td>
<td>8.2</td>
<td>6.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>

| AR    | 2.7  | 3.6 | 2.8 | 4.5 | 5.2 | 5.9 | 6.8 | 6.5 | 6.5 | 7.5 | 6.9 | 6.6 | 4.8 |
| A     | 2.9  | 3.6 | 2.9 | 4.7 | 5.6 | 6.6 | 7.3 | 6.9 | 7.0 | 8.4 | 7.4 | 6.9 | 5.1 |
| \( H_0 \) | 0.306 | 0.295 | 0.286 | 0.461 | 0.586 | 0.571 | 0.531 | 0.622 | 0.587 | 0.753 | 0.724 | 0.692 | 0.633 |
| \( H_E \) | 0.286 | 0.273 | 0.300 | 0.474 | 0.553 | 0.569 | 0.578 | 0.642 | 0.594 | 0.756 | 0.712 | 0.640 | 0.579 |

\( N \), sample size; \( n \), number of alleles detected (with the indication of private alleles in parenthesis); \( R \), allelic richness per locus; \( AR \), allelic richness per population; \( A \), mean number of alleles. \( H_0 \) and \( H_E \) are the observed and expected heterozygosity, respectively, across all loci. Population localization according to geographical region is indicated. For sampling locations, see Fig. 1.
Figure 2. Parsimony networks representing the possible phylogenetic relationships between the five $\beta$-fibint7 haplotypes. A, considering h5 as recombinant. B, considering h4 as recombinant. C, considering that h4 or h5 originated by a recurrent mutation. Circle size is proportional to the frequency of each haplotype in the total sample. The putative ancestral allele is indicated by an arrow. Dashed lines represent loops that may result from recombination. D, distribution of haplotypes observed in the 12 C. lusitanica populations studied. Pie charts represent the frequencies of $\beta$-fibint7 haplotypes in each population. For sampling locations, see Fig. 1.

Table 2. Variable positions in the 636 bp segment of the $\beta$-fibint7 defining five haplotypes in Chioglossa lusitanica, and correspondent base composition in its sister-species, Mertensiella caucasica

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Variable sites</th>
<th>h5 as recombinant</th>
<th>h4 as recombinant</th>
<th>GenBank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38</td>
<td>394</td>
<td>395</td>
<td>ML (SD)</td>
</tr>
<tr>
<td>h1</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>$1.12 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.27 $\times 10^{-7}$)</td>
</tr>
<tr>
<td>h2</td>
<td>.</td>
<td>A</td>
<td>.</td>
<td>$1.40 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7.82 $\times 10^{-7}$)</td>
</tr>
<tr>
<td>h3</td>
<td>.</td>
<td>A</td>
<td>A</td>
<td>$1.09 \times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.49 $\times 10^{-9}$)</td>
</tr>
<tr>
<td>h4</td>
<td>G</td>
<td>.</td>
<td>.</td>
<td>$4.93 \times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.98 $\times 10^{-8}$)</td>
</tr>
<tr>
<td>h5</td>
<td>G</td>
<td>A</td>
<td>.</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.94 $\times 10^{-8}$)</td>
</tr>
<tr>
<td>Mertensiella caucasica</td>
<td>T</td>
<td>A</td>
<td>.</td>
<td>-</td>
</tr>
</tbody>
</table>

Maximum likelihood (ML, with standard deviation) and relative likelihood (RL) estimates of the most likely ancestral haplotype based on unrooted genealogical tree probabilities, either considering the haplotype h5 or h4 as recombinant, and GenBank accession numbers for the new haplotypes obtained in the present study. *Sequeira et al. (2006).
composed by the more southern AC and MU populations. At $K = 7$, SAI population defined a separate cluster, and some individuals from BU were assigned with a probability higher than 0.8 to this cluster. AMOVA performed with the detected groups of populations corresponding to values of $K = 2$ and $K = 7$ were statistically significant for all parameters (Table 4). At $K = 7$, differences among groups (25.5%) and differences among populations within groups (3.2%) were higher and lower, respectively, compared to the same parameters at $K$ levels of 3–6 (data not shown).

The NJ tree constructed from $D_A$ distances (Fig. 5) showed three clusters supported by bootstrap values $> 90%$: (1) AC and MU populations; (2) CP, LO, VA, BU, SAI, CO, and MO populations; and (3) VAL, GE, PO, and SAL populations. These three clusters are almost completely concordant with the STRUCTURE analysis at $K = 3$. The NJ tree constructed from $D_A$ distances showed an unresolved tree topology with bootstrap values $< 70%$ (data not shown). A significantly positive relationship was observed between pairwise population $F_{ST}$ and geographic distances ($r = 0.678$, $P < 0.001$), whereas significantly negative relationship was detected between allelic richness and distance along a south–north axis ($r^2 = 0.823$, $P < 0.001$).

$\beta$-fibint7

Significant levels of genetic substructuring were found as revealed by both average $F_{ST}$ and $\Phi_{ST}$ values (0.479 and 0.223, respectively). Haplotypes h1 and h2 differed by a single point mutation and were found to be the most frequent, occurring predominantly to the north and south of the Mondego River, while haplotypes h4 and h5 were essentially confined to the hybrid zone. Additionally, haplotype h3 was irregularly distributed across most of $C. lusitanica$’s range (Fig. 2D). A significantly positive relationship between $D_A$ and geographic distances ($r = 0.288$, $P < 0.05$) was detected, but there was no association between haplotype diversity and distance along a south–north axis ($r^2 = 0.189$, $P > 0.05$).

Allozymes

By contrast to the structure analysis with microsatellites, a disproportionally high $\Delta K$ was seen at $K = 2$, and no further substructuring was accepted (using $Q > 0.9$) (Fig. 3B). At $K = 2$, Ln Pr ($X/K$) reached its maximum rate of increase and the mean

<table>
<thead>
<tr>
<th>Populations</th>
<th>$n$</th>
<th>$H$</th>
<th>$\Pi$</th>
<th>$K$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>North of Douro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salas</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pontevedra</td>
<td>10</td>
<td>0.53</td>
<td>0.0017</td>
<td>1.07</td>
<td>2</td>
</tr>
<tr>
<td>Gerês</td>
<td>8</td>
<td>0.25</td>
<td>0.0008</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>Valongo</td>
<td>10</td>
<td>0.60</td>
<td>0.0012</td>
<td>0.73</td>
<td>3</td>
</tr>
<tr>
<td>Between rivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montemuro</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Covelo</td>
<td>10</td>
<td>0.20</td>
<td>0.0006</td>
<td>0.40</td>
<td>2</td>
</tr>
<tr>
<td>Saide</td>
<td>10</td>
<td>0.53</td>
<td>0.0017</td>
<td>1.07</td>
<td>2</td>
</tr>
<tr>
<td>Buçaco</td>
<td>12</td>
<td>0.71</td>
<td>0.0014</td>
<td>0.89</td>
<td>4</td>
</tr>
<tr>
<td>Várzea</td>
<td>12</td>
<td>0.71</td>
<td>0.0015</td>
<td>0.94</td>
<td>4</td>
</tr>
<tr>
<td>South of Mondego</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lousã</td>
<td>12</td>
<td>0.17</td>
<td>0.0005</td>
<td>0.33</td>
<td>2</td>
</tr>
<tr>
<td>Açor</td>
<td>10</td>
<td>0.53</td>
<td>0.0008</td>
<td>0.53</td>
<td>2</td>
</tr>
<tr>
<td>Muradal</td>
<td>10</td>
<td>0.20</td>
<td>0.0003</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>0.71</td>
<td>0.0015</td>
<td>0.96</td>
<td>5</td>
</tr>
</tbody>
</table>

$n$, number of sequences in each population; $h$, haplotype diversity; $\pi$, nucleotide diversity; $K$, average number of nucleotide differences; $S$, number of haplotypes.

Table 3. Measures of $\beta$-fibint7 diversity

Figure 3. Bayesian analysis of population structure in $Chioglossa lusitanica$ performed with seven microsatellite loci (A) and with seven allozyme loci (B) using the software STRUCTURE. Depicted by solid circles, is the log likelihood Ln Pr($X/K$) for each $K$ simulated clusters. Depicted by solid triangles, is the across population average of the maximum proportion of assignment of individuals from a breeding population to any simulated cluster ($Q$), for each $K$ simulated clusters (shown with standard deviation).
Figure 4. Bayesian clustering results of STRUCTURE analysis (A) for microsatellites ($K = 2$–$7$) and (B) for allozymes ($K = 2$–$3$). Each individual is represented as a vertical line partitioned into $K$ coloured segments, whose length is proportional to the individual’s estimated membership coefficient. A black line separates individuals of different populations. These are labelled below the figure and are sorted from north to south (from the left to the right of the figure). Population localization according to geographical region is indicated on the top of the figure. The mean confidence assignment of all individuals ($q$) to their most probable cluster for each respective $K$ is indicated in parenthesis.

Table 4. Analysis of molecular variance results according to the partitioning of genetic diversity revealed by different markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Among groups</th>
<th>Among populations within groups</th>
<th>Within populations</th>
<th>Fixation indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Among groups</td>
<td>Total variance</td>
<td>% of Total</td>
<td>Total variance</td>
</tr>
<tr>
<td>Microsatellites</td>
<td>K2</td>
<td>20.9</td>
<td>0.588</td>
<td>14.6</td>
</tr>
<tr>
<td></td>
<td>K7</td>
<td>25.5</td>
<td>0.648</td>
<td>3.2</td>
</tr>
<tr>
<td>Allozymes</td>
<td>K2</td>
<td>67.4</td>
<td>0.951</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>K3</td>
<td>60.9</td>
<td>0.627</td>
<td>10.9</td>
</tr>
<tr>
<td>mtDNA</td>
<td>G2</td>
<td>78.0</td>
<td>3.415</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>G6</td>
<td>85.0</td>
<td>2.846</td>
<td>1.94</td>
</tr>
</tbody>
</table>

The groups correspond to those defined by STRUCTURE analysis at different levels of $K$. G2 and G6 correspond to two scenarios of population subdivision revealed by nested clade analysis of mitochondrial DNA variation (Alexandrino et al., 2002) indicating two (populations north and south of Buçaco) and six groups (1: SAL, FS, PO, GE, BA, BJ, VAL, and CO; 2: TA and MO; 3: SAI and BU; 4: BU and VA; 5: CP and LO; 6: MU and AC) (for sampling locations, see Fig. 1.). All fixation indices present $P < 0.001$. © 2008 The Linnean Society of London, Biological Journal of the Linnean Society, 2008, 95, 371–387
For populations exhibiting the northern mtDNA lineage: (1) there was no association between pairwise population genetic and geographic distances ($r = 0.141$, $P > 0.05$) but (2) a negative association between haplotype diversity and distance along a south–north axis ($r^2 = 0.76$; $P < 0.05$). As with allozymes, populations showing the southern mtDNA lineage were too few and thus associations were not tested.

**DISCUSSION**

The comparative analysis of both new (microsatellites and a nuclear intron) and existing (allozymes and mtDNA) data increased resolution and support for some prior inferences on the evolutionary history of *C. lusitanica*, but also illuminated both the strengths and limitations of each set of markers: (1) both microsatellites and the nuclear intron showed moderate to high levels of population structure, in agreement with patterns detected with mtDNA and allozymes; (2) both microsatellites and the nuclear intron showed a general pattern of isolation-by-distance, in contrast to the marked differentiation of two population groups suggested by mtDNA and allozymes; and (3) Bayesian analysis of microsatellites clearly separated populations to the north and south of the Douro river, while allozymes supported a similar distinction of populations north and south of Mondego River.

**PATTERNS OF POPULATION STRUCTURE**

Microsatellite data showed significant levels of population substructure. The level of genetic differentiation observed in this study (global $F_{ST} = 0.27$) is approximately equal or higher than the values found in other amphibian species (0.16–0.24) over large geographical scales (Monsen & Blouin, 2003; Burns, Eldridge & Houlden, 2004; Palo et al., 2004). The high population differentiation found in *C. lusitanica* appears to be associated with its low mobility and narrow ecophysiological requirements, meaning a strong dependence on brooks or other moist habitats that determines a limited and highly directional dispersal ability (Arntzen, 1981, 1994). Similar results were found among other amphibian species with low vagility (Palo et al., 2004; Spear et al., 2005).

Despite high levels of population differentiation, Bayesian multilocus analysis of microsatellite variation was not able to unequivocally support a specific number of population groups. This may be due to the strong pattern of isolation by distance observed in *C. lusitanica*, which limits the ability of the program’s algorithm to resolve population clusters (Pritchard & Wen, 2002; Worley et al., 2004). An important first result derived from our Bayesian clustering analysis was the unexpected spatial pattern of two clusters
geographically distributed to the north and to the south of the Douro river (Fig. 4A), which stands in stark contrast with the results obtained with allozyme data or with mtDNA evidence (Alexandrino et al., 2002). Another notable result was concerned with the magnitude of genetic structure. A spatial pattern of seven genetic clusters was suggested for microsatellites, whereas no subsequent identification of further clusters after $K=3$ was achieved with allozymes, demonstrating a more limited power of this type of markers compared to microsatellites in revealing patterns of genetic substructure (Figs 4A, 4B). Further population subdivision revealed within the seven defined groups, albeit with lower individual assignment confidence, may reflect mixed membership in multiple clusters, likely resulting from finer-scale isolation by distance (Pritchard & Wen, 2002; Worley et al., 2004). Partitioning into seven groups of populations is also supported by the AMOVA analysis, as shown by the highest contribution from differences among groups (approximately 26%) and the lowest contribution from differences among populations within groups (approximately 3%), to the overall genetic variation. The seven groups of populations were generally concordant with those revealed from previous analyses of mtDNA data (Alexandrino et al., 2002), with the exception of an additional group suggested by microsatellites, which supported the distinction of the VAL and CO populations.

Bootstrap supported (i.e. > 90%) population structure from the $D_\Lambda$ distance based tree is similar to population structure revealed by Bayesian analysis at $K=3$, with the exception of the separation between populations from the Lousã mountains (LO and CP) and populations from the mountains of Açor and Muradal (AC and MU), which is indeed observed in our Bayesian analysis, but only at $K=6$–7. The $D_\Lambda$ tree differs considerably, however, from the unresolved topology (i.e. bootstrap values all below 70%) demonstrated by the $\delta_\mu^2$ distance based tree. Although both measures yielded similar topologies in some studies (Roberts, Schwartz & Karl, 2004), detecting phylogenetic population relationships with microsatellite-specific distance measures has been problematic (Estoup et al., 1995; Angers & Bernatchez, 1998; Queney et al., 2001). The relative performance of distinct microsatellite genetic distance measures in describing population relationships has been a matter of controversy (Angers & Bernatchez, 1998; Goldstein & Schlötterer, 1999), partly related to assumptions concerning the fit of data to an evolutionary model, such as the the infinite allele model (Kimura & Crow, 1964) or a stepwise mutation model (SMM; Kimura & Ohta, 1978). In the present study, the non-microsatellite-specific $D_\Lambda$ distance recovers population relationships that are more similar to the structure revealed by Bayesian analysis, as well as to previously inferred relationships based on mtDNA and allozymes (Alexandrino et al., 2000), than that recovered by the microsatellite-specific $\delta_\mu^2$ distances. This suggests that $\delta_\mu^2$ assumes an inadequate model of evolution for our data.

Levels of polymorphism of $\beta$-fibint7 in $C. lusitanica$ were very low, with only three observed segregating sites in twelve populations across the species range. On average, $\beta$-fibint7 variation in $C. lusitanica$ was much lower than that found in $Lacerta schreiberi$ (Godinho et al., 2006) as well as other Iberian amphibians such as Lissotriton boscai (Teixeira, 2007) and Alytes obstetricians (Gonçalves, 2007). Nonetheless, $\beta$-fibint7 showed significant levels of genetic substructuring in $C. lusitanica$ ($\Phi_{ST} = 0.2$ and $F_{ST} = 0.5$ for haplotypic and genotypic data, respectively) but no phylogeographic structure (Fig. 2A, B). Instead, an overall pattern of isolation-by-distance is evident when genetic distance takes into account both allele sharing and haplotype genealogical divergence ($D_\Lambda$). Dissociation of haplotype genealogy with geography in historically isolated populations has been mainly explained through the persistence of ancestral polymorphism in populations or contemporary gene flow between populations (Hare & Avise, 1998; Broughton & Harrison, 2003; Caicedo & Schaal, 2004). For the slowly evolving $\beta$-fibint7, we suggest that ancestral polymorphism more likely explains our observations, given the combined mtDNA and allozyme evidence suggesting that northernmost and southernmost populations of $C. lusitanica$ have not exchanged genes recently (Alexandrino et al., 2000, 2002).

**Population history as inferred from distinct markers**

**Microsatellites**

The historical diversification of two evolutionary lineages of $C. lusitanica$ was inferred from mtDNA and allozymes (Alexandrino et al., 2000, 2002), but was not apparent from microsatellite data (present study). Two non-exclusive hypotheses could explain our observations: (1) higher mutation rates of microsatellites relative to the time of isolation (1.5–3.0 Myr; Alexandrino et al., 2000) between the two population groups may have resulted in extensive homoplasy and (2) contemporary gene flow has hidden the genetic signature of historical isolation across the Mondego River. Given a generation time of approximately 4 years (Lima, Arntzen & Ferrand, 2001; Sequeira, Ferrand & Crespo, 2003), substantial divergence between the two groups would be expected across the whole genome at neutral loci with moderate to fast mutation rates. Although allelic distribution profiles at selected microsatellite loci revealed
appreciable differences between the two groups, extensive allele size overlap occurs at most loci (Fig. S1). Given typical microsatellite mutation rates (Amos et al., 1996), we suggest that allele size homoplasy, through a combination of mutation constraints and back mutations, may account for the observed pattern of relatively homogenized allele size distributions (Nauta & Weissing, 1996; Xu et al., 2000). Similar hypotheses have been proposed for other species showing allopatric divergent mtDNA lineages but with low differentiation at microsatellite loci (Queney et al., 2001; Neumann et al., 2005). Abundant contemporary gene flow could also have contributed to conceal the genetic signature of historical isolation. However, allozyme data support that gene flow between diverged populations is restricted to the extent of the hybrid zone (approximately 10 km) with no evidence of further introgression (Sequeira et al., 2005a).

As expected, microsatellite analysis gives more robust support for events in the more recent past providing additional insights into previously postulated hypothesis of recent range expansion and post-glacial colonization of northern Iberia. For example, microsatellite data showed strongly depleted levels of diversity in populations north of the Douro River compared to more southern localities in the C. lusitanica range. Considering that both allelic richness and heterozygosity were markedly reduced in the north, and that microsatellites are fast-evolving markers that rapidly re-establish previous diversity levels, our data clearly support a recent and strong founder effect associated with the cross of the Douro River and northern colonization. In the south, estimates of genetic diversity are slightly lower in the MU population than in other southern populations, but the differences are not significant, and do not support a southward postglacial colonization, as had been suggested by allozymes and mtDNA (Alexandrino et al., 2000, 2002). However, estimates of allelic richness in populations from the mountains of Açor and Muradal (AC and MU populations) may be biased by the unusually high number of alleles at locus Cl5 in these populations (seven and 12 alleles, respectively; Table 1, Fig. S1) compared to southern localities located near the Mondego River. A visual inspection of patterns of allelic distribution at loci Cl6, Cl17 and Cl136 shows a slight reduction in the number of alleles at the AC and MU populations, suggesting either that recent colonization and/or local population bottlenecks may have occurred.

The study of the C. lusitanica hybrid zone with diagnostic allozyme and mtDNA loci suggested the absence of genetic barriers to the neutral diffusion of alleles across the zone (Sequeira et al., 2005a). Cline shapes at nuclear markers also indicated that secondary contact between the two lineages probably occurred before the Holocene as a result of several instances of isolation and contact, limited by the habitat conditions that prevented or promoted connectivity across the Mondego valley (Teixeira et al., 2001; Teixeira & Arntzen, 2002). Microsatellites showed that populations located south and north of the Mondego are differentiated and form distinct population groups (K > 2), with two moderately informative loci between the two groups (Cl5 and Cl6; Fig. S1). Additionally, the presence of some rare microsatellite alleles in populations south and north of the Mondego (e.g. allele 244 at locus Cl19; Fig. S1) suggests recent gene flow between the two diverged groups. However, because the analysis of microsatellite loci could not clearly define two deeply differentiated groups, the hypothesis of secondary contact cannot be examined from these data alone.

**β-fibint7**

The geographical distribution of β-fibint7 haplotypes could approximately distinguish a northern and a southern group of populations characterized, respectively, by the presence of haplotypes h1 and h2 that showed opposite clinal variation along a south–north transect (Fig. 2D). However, the absence of further association between genealogy and geography prevented the clear distinction of the two population groups inferred from mtDNA and allozyme data. This most likely results from a slower lineage sorting of a nuclear gene genealogy compared to mtDNA (Hare, 2001; Palumbi, Cipriano & Hare, 2001; Hudson & Turelli, 2003). However, β-fibint7 variation is also discordant with some nuclear allozyme loci. The pronounced genetic break revealed by allozymes resulted from the analysis of variation observed at seven polymorphisms out of twenty-two screened loci (Alexandrino et al., 1997, 2000). Levels of population subdivision varied according to distinct polymorphic loci in C. lusitanica: only four out of seven showed alternatively fixed alleles for the two evolutionary lineages (Alexandrino et al., 2000). We suggest that the observed differences between both cytoplasmic and nuclear loci, and among distinct nuclear loci, could have resulted from the differential persistence of ancestral polymorphism in populations following independent evolutionary trajectories (Pamilo & Nei, 1988; Moore, 1995). Although natural selection may also have contributed to shape patterns of allozyme polymorphism on C. lusitanica, the stochasticity of the coalescent process is sufficient to explain the observed differences at distinct nuclear loci.

The most diverse populations for the β-fibint7 were found along the Mondego river valley region, whereas a decrease in genetic diversity was detected in both the northernmost and southernmost populations,
characterized by the presence at high frequency of haplotype h1 and h2, respectively. This decrease in genetic diversity is in close agreement with the genetic effects of the postulated postglacial colonization described on the basis of mtDNA, allozymes and microsatellites. Interestingly, haplotypes h4 and h5 were found to occur exclusively in populations close to Mondego River valley. In recent years, several studies have reported high frequency alleles in hybrid zones that are generally absent in the parental populations (Sage & Selander, 1979; Barton & Hewitt, 1985). Although we cannot rule out other causative factors for the observed distribution of h4 and h5 in the hybrid zone, such as a combination of back mutations and genetic drift, it is very likely that these haplotypes reflect the phenomenon of novel hybrid-zone alleles (Bradley et al., 1993), resulting from long-term persistence and admixture of two divergent lineages of *C. lusitanica* in the Mondego valley.

**IMPLICATIONS FOR THE STUDY OF EVOLUTIONARY PROCESSES AT DISTINCT SPATIO-TEMPORAL SCALES**

Comparative analysis of the merits of different types of molecular markers in reconstructing the evolutionary history of an organism exhibiting very predictable ecological characteristics (e.g. strict dependence of mountain streams, low vagility) provided notable insights in the understanding of their usefulness when evolutionary processes are studied at different spatio-temporal scales. A remarkable discordance between markers was observed for allozymes and microsatellites when only two genetic clusters were considered in the STRUCTURE analysis. Although allozymes clearly reflected the historical genetic discontinuity associated with the Mondego River valley, in concordance with mtDNA evidence, microsatellites more clearly revealed increases in genetic distance associated with the presumably more recent expansion across the Douro River, located further north of Mondego valley. This discordance highlights the effectiveness of both allozymes and mtDNA in preserving the genetic signature of an old split, compared to microsatellites. Thus, microsatellites may not be the markers of choice to capture the signal of hybrid zones that resulted from long-term isolation of populations among allopatric refugia, especially in those originated by multiple pulses of expansion and admixture during the Pleistocene. Similar results were reported in recent studies of hybrid zones originated by secondary contact after long-term isolation in a variety of other organisms, such as the European rabbit (Queney et al., 2001), the Iberian midwife toad (Gonçalves, 2007), and the Australian rainforest skink (Phillips, Baird & Moritz, 2004). In each of these studies, a pattern of broad overlap in allele size was found at most loci among highly divergent mtDNA lineages that were differentiated during the Quaternary glaciations.

On the other hand, what is an apparent cline of variation in allele frequencies for allozymes in populations of *C. lusitanica* north of the Douro River is shown with microsatellites as a well-defined cluster resulting from a recent strong discontinuity across a geographical barrier. In this case, allozymes simply lack the appropriate power to detect such a noticeable cluster, whereas the elapsed time is sufficient for the signal of recent historical discontinuity to be preserved in the rapidly mutating microsatellite loci. This confirms the expectation of the utility and robustness of microsatellites in supporting relatively recent events, in particular those associated with the postglacial colonization of previously unsuitable areas (Queney et al., 2001; Palo et al., 2004; Rowe, Harris & Beebee, 2006).

Although the present study only provides analysis of a single nuclear gene genealogy as an alternative to an organelle genealogy, an observation was made that is important to the investigation of hybrid zones. The inference of a recombination event in the hybrid zone of *Chioglossa* is analogous to the inference in Godinho et al. (2006), using the same intron for the hybrid zone analysis of another Iberian endemic. These authors suggested that divergent lineages of Schreiber’s green lizard experienced repeated population expansions and contractions creating multiple opportunities for admixture and the establishment of a long-lived hybrid zone (*sensu* Hewitt, 1996). Although these issues can only be clarified by future detailed studies including the analysis of multiple nuclear gene genealogies, it should be stressed that the emergence and spread of evolutionary novelties go unnoticed when only the traditional genotyping of microsatellites or the sequencing of the nonrecombining mtDNA molecule is performed. Future studies involving the use of multiple compound markers (e.g. SNP-defined haplotypes closely linked to one or more microsatellites) will be powerful tools in the analysis of the admixture dynamics of this type of hybrid zones (Baird, 2006).

Taken together with previously published mtDNA and allozyme data, the present study of a set of microsatellites and a nuclear gene genealogy in a species with a restricted geographical distribution and a relatively well-known natural history revealed the complementary merits and limitations of the different types of molecular markers in enriching our understanding of the evolutionary history of *C. lusitanica*. Additionally, the present study may serve as a cautionary note to those investigations that rely on only one or two types of molecular markers, especially when the organism under study exhibits a widespread distribution and a complex natural history.
ACKNOWLEDGEMENTS

This work was partially financed by Fundação para a Ciência e a Tecnologia (FCT), through research projects POCTI/BSE/40987/2001 and POCTI/BSE/41395/2001, and SFRH/BPD/3365/2000 PhD grant and by a post-doctoral grant (SFRH/BPD/27134/2006) to FS. We are also grateful to Raquel Godinho, Sara Rocha, Catarina Pinho, Rui Faria, Albano Beja-Pereira and Pedro Esteves, as well as Stephan Koblmüller, Nina Duftner, and Simona Susnik from the Institute of Zoology in Graz, for their help with the laboratory work.

REFERENCES


Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.


Palumbi SR, Cipriano F, Hare MP. 2001. Predicting


Rosenberg NA. 2002. *DISTRUCTIONS: a program for the graphical display of structure results*. Available at http://www.cmb.usc.edu/~noahr/distruct.html


Schneider S, Roessli D, Excoffier L. 2000. *ARLEQUIN (version 2.0): a software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Allelic frequencies at seven microsatellite loci for sampled populations. Circles represent different alleles, and their surfaces are directly proportional to their frequencies. A, *Cl5*; B, *Cl6*; C, *Cl45*; D, *Cl17*; E, *Cl136*; F, *Cl39*; G, *Cl19*. Alleles sizes (in bp) are shown on the x-axis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.