Molecular Ecology (2002) **11**, 809–819

Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard

OCTÁVIO S. PAULO,* WILLIAM C. JORDAN,† MICHAEL W. BRUFORD‡ and RICHARD A. NICHOLS§

**Centro de Biologia Ambiental/Departamento de Zoologia e Antropologia, Faculdade de Ciencias da Universidade de Lisboa, P- 1749–016 Lisboa, Portugal, †Institute of Zoology, Zoological Society of London, London NW1 4RY, UK, ‡Cardiff School of Biosciences, Cardiff University, Cathays Park, Cardiff CF10 3TL, UK, §School of Biological Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK*

Abstract

The distribution of the lizard *Lacerta schreiberi* is likely to have been severely affected by the climatic cycles that have influenced the Iberian Peninsula. Information about the species ecology and Iberian physiogeography was used to generate specific hypotheses about episodes of colonization and subsequent population persistence. These hypotheses generated predictions about the distribution of genetic variation, which were tested using nested clade analysis (NCA) supplemented by analysis of molecular variance (amova). Two predictions were confirmed by NCA; that is those that specified multiple and allopatric refugia. However, the remaining three predictions were not corroborated by the analyses. Firstly, a simple analysis of the distribution of genetic variability failed to detect an expected difference in the pattern of colonization between the inland mountain system and the coastal region. Moreover, while NCA did detect the expected genetic pattern in southern coastal populations, it was explained in terms of long-distance migration, which seems implausible because of the extent of unsuitable habitat. A more likely cause of the pattern is population fragmentation and a reduction in population size caused during the Holocene. Finally, NCA also failed to detect a northwestern population expansion, which is supported by other evidence. We conclude that NCA has a limited ability to detect range expansion led by individuals with more ancestral (interior) haplotypes.

Keywords: Iberian Peninsula, *Lacerta schreiberi*, phylogeography, Pleistocene, Pliocene, range expansion

Received 8 August 2001; revision received 8 January 2002; accepted 8 January 2002

Introduction

Phylogeography, the newest subdiscipline of Biogeography, is undergoing rapid development. One important reason is the refinement of analytical approaches that are designed to deal with specific problems posed by intraspecific DNA sequence data, the raw material of the majority of phylogeographical studies (Avise 2000). Traditional phylogenetic methods are particularly suitable for interspecific analysis but have proved to be less useful at the intraspecific level. Among other developments in population genetics, recent advances in coalescent theory have provided methods for assessing evolutionary and population genetic hypotheses (Posada & Crandall 2001). Nested clade analysis (NCA) is one of the most promising techniques under this new theoretical framework, and makes it possible to disentangle population structure from population history. It achieves this by reconstructing the main sequence of events that have generated the current genetic pattern in a species (Templeton *et al*. 1995; Templeton 1998).

The principal aim of this paper was to use NCA to investigate the current population structure and population history of an Iberian Peninsula lizard, *Lacerta schreiberi*, and to assess several hypotheses about the history of its spread and persistence. A secondary aim was to compare the results

Correspondence: Octávio S. Paulo. Fax: 00 351 217500028; E-mail: octavio.paulo@fc.ul.pt

from NCA with those from a hierarchical analysis of molecular variance.

The contraction-expansion paradigm

The Quaternary period was characterized by a number of climatic oscillations that were responsible for a series of contractions and expansions of species ranges all over the world (see review in Hewitt 2000). Evidence from numerous species suggests that southern parts of Europe functioned as ice age refugia, in particular the peninsulas of Iberia, Italy and the Balkans and possibly areas near the Caucasus and the Caspian Sea (Hewitt 1999). These areas mainly correspond to the remains of arborescent habitats at the peak of the last glaciation (Huntley & Birks 1983; Zagwijn 1992). During interglacial warm periods, colonization of northern Europe, by persistent species from southern refugia, allowed partial replenishment of faunal and floral compositions and led to the formation of suture or hybrid zones, where different genotypes came in contact (Remington 1968; Taberlet *et al*. 1998; Hewitt 1999; 2000). Hewitt (1996; 1999; 2000) has explored the genetic consequences of ice ages, in particular the ways in which populations persisted in southern refugia and the manner of their colonization of northern areas during postglacial events.

The process of colonization

The mode of colonization from refugial populations can be thought of as falling between two extreme modes: 'pioneer' and 'phalanx' (Nichols & Hewitt 1994). Phalanx expansion is a slow process, with a high proportion of individuals dispersing over short distances and consequently, new populations maintain almost all the genetic diversity of the original population. Something resembling phalanx expansion is thought to have happened within southern Europe as populations changed their altitudinal range at different phases of the glacial cycle.

In contrast, the pioneer process is a rapid expansion involving founder events, where long-distance migrants establish populations some distance from their source. This type of expansion may have been responsible for the recolonization of large tracts of northern Europe and it could explain the large areas of reduced genetic diversity seen at higher latitudes.

Ibrahim *et al*. (1996) modelled three different dispersal processes and compared the resulting genetic patterns. The three processes were: leptokurtic dispersal (with a high proportion of long distance migrants); a normal dispersal function and a stepping stone process. Leptokurtic dispersal produces large and expanding patches of homozygosity compared to the other forms of dispersal, resembling the patterns observed in northern Europe. As a consequence, this 'leading edge' process is now often used to explain the way postglacial northward expansion has structured genome distribution. The exact pattern of expansion will depend on the sharpness of the climatic change, the latitude and the topography of the region, and the dispersal and reproductive capabilities of the organism (Nichols & Hewitt 1994).

Dispersal at the leading edge would have taken place when the climate warmed rapidly and long distance dispersal was possible into large areas of newly suitable vacant habitats, where isolated populations could be founded and grow rapidly to a large size before the area became widely colonized. From these areas the new population would have expanded towards newly available free areas. This process may have been repeated several times until all available habitats were occupied, involving successive bottlenecks for the colonizing genome and the concomitant loss of alleles and increase in homozygosity. Once a new population isolate was established, it would be more difficult for other migrants to contribute significantly to the gene pool of the colonisers, and consequently the populations behind the front line of the species range would contribute very little to the expansion process. In this way, most genetic variability would be restricted to pre-expansion populations with much lower levels of genetic diversity in derived populations (Hewitt 1999).

Alternative explanations for the north–south difference could involve differential selection, with selection favouring a common haplotype in northern areas and several haplotypes in southern regions (Excoffier 1990; Ballard & Kreitman 1995; Merilä *et al*. 1997). A third option is that the northern pattern is typical and that southern richness results from the admixture of original populations which previously diverged allopatrically in several local refugia, generating 'artificially' high levels of genetic variability.

The process of persistence

For a species to persist through climatic cycles, it must either track the changing distribution of suitable habitat, or it must adapt. The evidence suggests that during the warm interstadial periods of the climatic oscillations species persisted in southern areas by moving their range to higher altitudes. In addition 'edge' populations expanded toward newly available northern areas**.** Meanwhile, lowland and southern populations became extinct. Conversely, during cold glacial periods most of the populations became extinct in northern areas and at high altitudes while some populations colonized and persisted in warm valleys and southern lowland areas.

So far the empirical evidence supporting this model comes from the current distribution of species, pollen records, and the pattern of genetic diversity apparent in the Alpine, Apennine, Balkan and Iberian mountain ranges. Parapatric

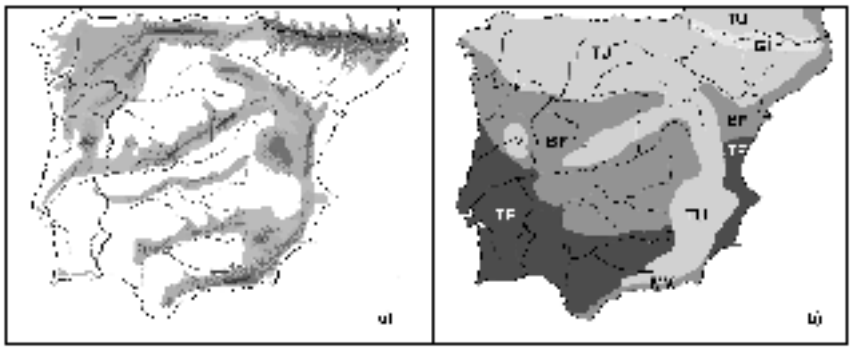

Fig. 1 Maps of the Iberian Peninsula. (a) Schematic representation of the main mountain systems and rivers. Darker shades represent higher altitudes, thin lines represent rivers and thicker lines are political and continental borders. (b) Inferred vegetation pattern of Iberia during the last glacial maximum: TF- temperate mixed forest; MV-Mediterranean vegetation; BF-boreal forest; TU-tundra; GI-glacial ice in the modern day Pyrenees (adapted from Crowley & North 1991).

or allopatric divergence may have generated the genetic patterns over several ice ages, as populations survived by limited movement of their ranges between suitable locations. Hybrid zones may have been produced as the forms expanded to become parapatric (Hewitt 1996). The observed patterns could alternatively be explained by pre-Quaternary divergence, and not be directly associated with the persistence process.

The Iberian Peninsula

The Iberian Peninsula has several mountain and river systems generally running east–west, acting as major barriers to north–south dispersal for many species. Some of these systems are almost parallel to each other (Fig. 1). This topography produces a wide variety of climatic conditions, with northern mountains and river valleys very different from southern mountains and lowlands. In addition, a strong Atlantic influence creates a east–west climatic differentiation and enhances the differences between humid northern and dry, southern mountain faces. This pattern generates a diversity and fragmentation of habitats across Iberia that is likely to have created opportunities for species persistence during Pleistocene climatic cycles.

The last glacial cycle is the best-understood ice age, and at its height (which lasted until 18 000 BP) the Iberian Peninsula, in common with other parts of southern Europe, was much less affected than northern and central regions of the continent. While northern and central parts of Europe were covered by ice, tundra or steppe habitats, Iberia still had temperate mixed forest (chiefly broadleaf); boreal forest; Mediterranean vegetation (mostly coniferous); and some glacial ice and tundra in the northern and eastern areas (Fig. 1) (Huntley & Birks 1983; Crowley & North 1991). Two areas with deciduous forest became known as the Lusitanian and Andalucian refugia and have been assumed to be refugial areas for many species inside Iberia. However, the detailed distribution of the vegetation is still controversial (Zagwijn 1992), largely because there is a lack of extensive fine scale pollen records covering the upper Pleistocene.

Schreiber's green lizard

Schreiber's green lizard, *L. schreiberi*, is an Iberian Peninsula endemic, with a present-day distribution as shown in Fig. 2. The main distribution is in the northwestern part of Iberia, from the Basque Country in northern Spain to the region near Lisbon. It also occupies the Central Mountain system of Iberia and has some small isolated populations in Toledo, Guadalupe in Spain, and the São Mamede and Monchique mountains of southern Portugal. A recent, and probably anthropogenic, fragmentation has led to progressive isolation of some small populations on the west coast of Portugal, i.e. in Sintra, the Montejunto Mountains and the Caldas da Rainha area (Brito *et al*. 1996; 1998) (Fig. 2).

The species has atypical ecological preferences when compared with other Iberian lizards. In particular it displays a preference for habitats with a low number of sunlight hours per year and high annual precipitation (Brito *et al*. 1999). In southern and central isolates, the species is restricted to river corridors, usually in mountains, but sometimes in lowlands areas adjacent to humid mountain faces. Where the climatic conditions allow for the development of deciduous forest or humid mountain habitats, as in northwest Iberia, the species is less restricted and shows a continuous distribution (Riva 1987; Marco 1994). *L. schreiberi* has no recognized subspecies and displays remarkable morphological uniformity (Galan 1984).

The application of models of persistence and colonization to L. schreiberi

Based on Iberian physiography, on the known ecology and distribution of Schreiber's green lizard and on preliminary genetic results (Paulo 2001; Paulo *et al*. 2001), it is possible to derive and test some hypotheses on the processes responsible for shaping genetic variation within this species.

The colonization model predicts two distinct colonization processes. Mass movement up and down mountains is expected to result in only a small reduction of genetic variability. This type of shift may have occurred in the inland Central Mountain areas. The second type of colonization is

Fig. 2 Map of the Iberian Peninsula, with the distribution of Schreiber's green lizard *Lacerta schreiberi* and the name and location of the sampled populations. Different grey tones correspond to the different clades/regions identified by the analytical techniques: CNR-Coastal Northern Region, CSR-Coastal Southern Region, ISR-Inland Southern Region, INRw-Inland Northern Region west, INRe-Inland Northern Region east.

'leading edge' expansion producing long distance range expansion with reduction of genetic variability. This process might have occurred along the coastal populations where there are longer uninterrupted tracts of suitable habitat.

During glacial periods, it is likely that the populations in northern areas and the central mountains declined in distribution and size. Some populations probably went through severe bottlenecks and many are likely to have become extinct. Those that persisted are likely to have found refuges in warm interior river valleys or coastal zones. During interglacial periods, the suitable habitat would have expanded again and the populations expanded with it. As we are currently in an interglacial period, the distribution may have resembled that found today.

Conversely, the current distribution suggests that southern populations would have been fragmented or semi-isolated during interglacials. During the glacial periods they may have expanded in distribution and population size due to the extension of deciduous forest habitat (Fig. 1b). This expanded population would have the opportunity to replenish, by mutation, the lost genetic variability during the interglacial bottleneck period.

Conventional wisdom assumes that there was a single large refuge for lizards on the Iberian Atlantic coast during ice ages (Marco 1994). However the reasoning outlined above suggests that there would have been several additional refugia of variable size. Some may have been strongly affected by bottlenecks or drift, others may not. A similar pattern of multiple refugia is an accepted explanation for plant species persistence in the Iberian Peninsula (Bennett *et al*. 1991; Peñalba 1994; Allen *et al*. 1996; Ramil-Rego *et al*. 1998a; Ramil-Rego *et al*. 1998b). However, the exact location and size of the refugia for lizards, in particular in the central inland region, are still unknown.

Based on the above reasoning, it is possible to derive five predictions that can be assessed by phylogeographical analyses: (i) there will be genetic evidence for multiple inland refugia that were; (ii) separated by allopatric fragmentation; (iii) there will be higher genetic variability in inland mountain areas than in northwestern coastal populations due to 'phalanx' colonization in the mountains and 'leading edge' expansion along the coast; (iv) southern isolated populations will show evidence of recent fragmentation caused by Holocene habitat reduction; (v) northwestern populations will show evidence of recent range expansion.

Materials and methods

Samples and DNA extraction and sequencing

Samples were collected under licence by clipping two or three tail rings from free-living animals that were immediately released back into the wild. Figure 2 shows the location and names of the 18 sample sites. Tissue samples were immediately preserved in 25% w/v DMSO (dimethylsulphoxide) in 6 M NaCl at room temperature. Total DNA extraction was carried out by proteinase K/SDS overnight digestion at 50 °C of 0.5–1 g of tissue and purified by phenol:chloroform extraction and ethanol precipitation according to standard protocols (Sambrook *et al*. 1989). Polymerase chain reaction (PCR) amplification of a fragment of the cytochrome *b* gene of mitochondrial DNA (mtDNA) was carried out using the following published primers (or modified versions): cytochrome *b*, L14841 5'AAA AAG CTT CCA TCC AAC ATC TCA GCA

TGA TGA AA-3′ and H15551 5′-AAA TAG GAA GTA TCA CTC TGG TTT-3′(Kocher *et al*. 1989; Moritz *et al*. 1992). Two internal primers were also designed and used to sequence fragments of the amplified product: a truncated version of primer L14841, 5′-CCA TCC AAC ATC TCA GCA TGA TGA AA-3′ and H15150 5′-CAG AAG GAT ATC TGT CCT C-3′. These primers are named according to their occurrence on the heavy (H) or light (L) strand and their 3′ position in the human mitochondrial DNA sequence (Anderson *et al*. 1981). PCR reactions were carried out in a final volume of 50 µL, with 2 µL of DNA (50–100 ng), 50 pmol of each primer, 1.25 mM of each dNTP, in a reaction buffer of 50 mM KCl, 20 mM Tris-HCl (pH 8.4), 2 mM $MgCl_2$ with 10 µg BSA and 1 unit *Taq* DNA polymerase (GibcoBRL). A programmable thermal cycler (Perkin-Elmer model 9600 and 2400) was used for amplifications. PCR conditions were an initial denaturation step of 94 °C for 5 min, followed by 30 cycles consisting of denaturation at 94 °C for 30 s, annealing at 51 °C for 30 s, extension at 72 °C for 30 s, and a final extension step of 72 °C for 7 min PCR products were then purified and concentrated with Geneclean (Bio 101). An aliquot of 25–50 pg of purified DNA template was used for dRhodamine terminator cycle sequencing (ABI Prism) following the manufacturer's instructions. Both strands were sequenced for all samples, and sequencing products resolved on a semiautomated sequencer (ABI Prism model 377). Sequences were aligned first using SEQUENCHER™ 4.0 software and then adjusted by eye.

Population genetic variability and geographical partition

Measures of genetic diversity (haplotype and nucleotide diversity) were calculated using DnaSP 3.0 (Rozas & Rozas 1999). The geographical pattern of population subdivision and population structure was assessed by analysis of molecular variance, AMOVA (Excoffier *et al*. 1992). The prespecified hierarchy yields three components of genetic variation: within populations relative to the whole species (ϕ_{ST}), among populations within groups (ϕ_{SC}), and among groups (ϕ_{CT}). With sequence data, the number of mutations between haplotypes is taken into account for the AMOVA calculations. Populations were initially grouped according to the clades identified by phylogenetic analysis or according to geography, but after initial analysis several other hierarchical arrangements were tested. The arrangement that maximized ϕ_{CT}, the among-group variation, and was statistically significant, was assumed to be the most plausible geographical subdivision. Random permutation procedures were used to test the significance of the variance components and ϕ-statistics, which avoids the parametric assumptions of normality and independence that are not met by molecular distance measures. For this analysis, due to the small sample size, samples from the site of Peña de Francia was analysed together with Bejar, and those from Gredos and Pegerinos combined with Guadarrama. AMOVAs were performed using the software ARLEQUIN v2 (Schneider *et al*. 2000).

Population history and structure

In addition to AMOVA, we used NCA (Templeton *et al*. 1995; Templeton 1998; Avise 2000) which comprises several sequential procedures.

Haplotype network estimation

Templeton *et al*. (1992) developed a statistical parsimony approach to construction of haplotype trees or networks at the intraspecific level, based on the most parsimonious connection of pairs of haplotypes. Cladogram estimation was carried out using TCS software version 1.06 (Clement *et al*. 2000).

Nested statistical design

Standard nesting rules (Templeton *et al*. 1987; Templeton & Sing 1993) are then used to convert the resulting cladogram into a nested design. Clades are formed and nested in accordance to the number of mutational changes between them, until the final level encloses the entire cladogram.

Nested contingency analysis

The nested design can be used to test for geographical association in two different ways. The simplest procedure is a nested contingency test, described by Templeton & Sing (1993). This test does not include the geographical distances or positions among the sampled locations, which reduces its power when compared with nested geographical distance analysis (see below). The test assumes that each location is a category variable and makes combinations between locations and the clades at each chosen clade level. An exact permutational contingency test is performed to test the null hypothesis of no association of clades with geographical location. The observed χ^2 value is compared to a distribution of χ^2 values generated from permutations of the original data.

Nested geographical distance analysis

This second method to detect geographical association uses the nested design and a matrix of geographical distances between the sampled locations. Clade distance (D_c), the average distance between the location of the members of the clade and the geographical centre of the clade, is calculated for each clade. The nested clade distance (D_n), the average spatial distance between the members of each

clade and the geographical centre of the entire nesting clade is also calculated (Templeton *et al*. 1995). The method also distinguishes two topological classes of clades: tip clades, with only one connection to the remaining cladogram, and interior clades with two or more connections to other clades.

Based on these definitions two other measures can be calculated, the difference between the clade distance of tips and interior clades of a nesting clade (D_cI-D_cT), and the difference between nested clade distance of interior and tip clades (D_nI-D_nT). The hypothesis of spatial panmixia is tested in the following way. The observed distribution of clade distances is compared to the distribution generated by random permutation of clades against sampled locations. The randomization is constrained to keep the clade frequency and sample size constant (1000 or more permutations). The rejection of the null hypothesis for a certain clade only indicates that there is a significant association between haplotype and geography. To infer the population structure and history at each clade level, Templeton *et al*. (1995) have built an inference key. We made use of the upgraded version of the key for the nested haplotype tree analysis of geographical distances originally presented in the appendix of Templeton *et al*. (1995). Both nested contingency analysis and nested geographical distance analysis were carried out with GEODIS version 2.0 computer package (Posada *et al*. 2000).

Results

A total of 24 unique haplotypes were identified out of the 83 specimens assayed. Across the 663 base pairs analysed for each sample no indels or premature stop codons were found. Alignments were straightforward and all sequences were translated into amino acids suggesting that they were functional. Nucleotide sequences are deposited in GenBank under accession numbers AF372103-AF372126.

Population variability

The distribution of 83 sequences among the 18 sampled populations is presented in Table 1. There was a remarkably low number of haplotypes observed, even in populations with a moderate number of samples analysed. The overall species nucleotide diversity ($\Pi = 0.026 \pm 0.005$) was relatively high for a species with such a low number of haplotypes, due to the deep divergence between them. There was a general trend for populations belonging to the Coastal Northern Region (CNR, Fig. 2) to show lower nucleotide diversity than Coastal Southern Region (CSR). Populations from CSR tended to have the highest nucleotide diversity, except for the very small populations of Sintra ($N < 100$) and Cercal ($N < 1000$). The remaining populations of Guadarrama, Bejar, Toledo and Guadalupe tended to have lowest haplotype and nucleotide diversity (Table 2).

Nested Clade Analysis

NCA (Tables 3 and 4, Figs 3 and 4) detected two main clades, 3–1 (the coastal clade) and 3–2 (the inland clade). The haplotype with the highest outgroup probability was used to assign interior status to Clade 3–1. A nested contingency test (Table 3) clearly showed an association between each

Table 1 Distribution of haplotypes at each sampled site. The number of samples, different haplotypes and unique haplotypes are indicated for each row

Haplotypes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	Different	Unique
Cantabria	1	1																							2	2	1
Galicia	2		1	1																					4	3	2
Geres	2				1	1																			4	3	2
Estrela	5																								5	1	0
Malcata							3	1																	4	2	2
Mamede	3								1																4	2	1
Caldas										1	2			1											4	3	2
Montejunto												2	1	1											4	3	2
Sintra														4											4	1	0
Monchique														2	2										4	2	0
Cercal															4										4	1	0
Guadarrama																7	1								8	2	1
Pegerinos																1									1	1	0
Gredos																		1							1	1	1
Bejar																			8	1					9	2	1
Peña de Francia																			1						1	1	0
Toledo																					9	1			10	2	1
Guadalupe																					5		4	1	10	3	2
	13	1	1	1	1	1	3	1	1	1	2	2	1	8	6	8	1	1	9	1	14	1	4	1	83		

Table 2 Nucleotide (π) and haplotype (h) diversity and corresponding standard deviation (SD), for populations with four or more individuals sampled. CNR-Coastal Northern Region, CSR-Coastal Southern Region, ISR-Inland Southern Region, INRw-Inland Northern Region west, INRe-Inland Northern Region east

Region	Population	h ± SD	Π ± SD
CNR	Galicia	0.833 ± 0.222	0.00151 ± 0.00051
CNR	Geres	0.833 ± 0.222	0.00151 ± 0.00051
CNR	Estrela	0	0
CNR	Malcata	0.500 ± 0.265	0.00151 ± 0.00080
CNR	Mamede	0.500 ± 0.265	0.00075 ± 0.00040
CSR	Caldas	0.833 ± 0.222	0.00427 ± 0.00145
CSR	Montejunto	0.833 ± 0.222	0.00251 ± 0.00091
CSR	Sintra	0	0
CSR	Monchique	0.667 ± 0.204	0.00201 ± 0.00062
CSR	Cercal	0	0
INRe	Guadarrama	0.250 ± 0.180	0.00038 ± 0.00027
INRw	Bejar	0.222 ± 0.166	0.00034 ± 0.00025
ISR	Toledo	0.200 ± 0.154	0.00030 ± 0.00023
ISR	Guadalupe	0.644 ± 0.101	0.00114 ± 0.00026

Table 3 Results of the tests of nested contingency analysis of geographical associations. Clade column refers to the clades of Fig. 4. Only clades with some geographical or genetic variation were tested

Permutational		
Clade	chi-square statistic	Probability
1–1	57.5	0.004
1–2	5.8	0.194
1–7	0.5	1
1–9	0.1	1
1–10	10.1	0.386
2–2	31.6	< 0.001
2–3	20.0	< 0.001
2–4	6.7	0.034
3–1	43.0	< 0.001
3–2	40.0	< 0.001
Entire Cladogram	83.0	< 0.001

of these two main clades and geographical location (see also Fig. 2).

The only uncertainty found in the cladogram was the connection between clades 3–1 and 3–2, where several equidistant alternatives can be found. However these alternative connections were always between clade 1–1 and clade 1–10.

Within each of the main clades two secondary clades were detected. The upgraded version of Templeton's (1998) inference key suggested past fragmentation as the process that produced the two 2-step clades [2–1 (CNR) and 2–2 (CSR)] inside the 3-step (coastal) clade. It also suggested that allopatric fragmentation created the two 2-step clades 2–3 (Inland Northern Region – INR) and 2–4 (Inland Southern Region – ISR) inside the other 3-step (inland) clade.

Clade 2–3 had a geographical pattern compatible with a past fragmentation event between nested clades 1–9 (the western part of the Inland Northern Region – INRw) and 1–10 (the eastern part of the Inland Northern Region – INRe). The three other 2-step clades showed a genetic structure that was compatible with restricted gene flow with isolation by distance (2–1 = 1–1) (CNR) and restricted gene flow/dispersal

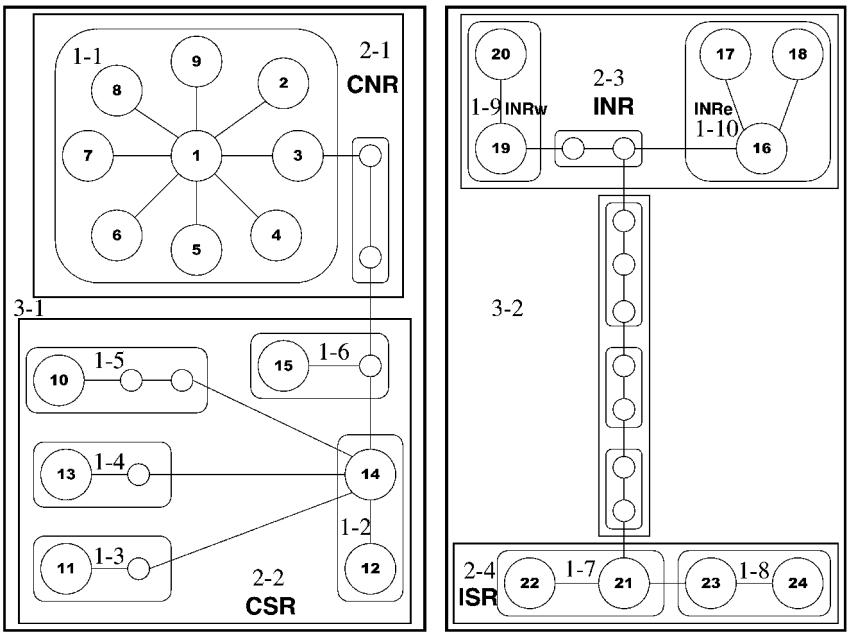

Fig. 3 The nesting design inferred from the cladogram estimation of the 24 haplotypes detected for *Lacerta schreiberi*. Each line in the network represents one mutational change. Small empty circles represent the inferred, nondetected interior haplotypes. The number inside each circle identifies the detected haplotypes, as in Table 1. Thin-lined rounded rectangles indicate the haplotypes grouped together into 1-step clades and the thin-lined rectangles indicate the nesting 2-step clades, while medium-lined rectangles indicate the 3-step clades. The connection between the two main clades, 3–1 and 3–2 is between clade 1–1 and 1–10, see text for details.

Table 4 Analysis of variance among populations of *Lacerta schreiberi*. Fixation indices are indicated as well as the percentage of the total variance that is explained by the grouping and its significance

Groups	ϕ_{ST}	ϕ_{SC}	ϕ_{CT}	% among groups	P
[Coastal Region] [Inland Region]	0.978	0.882	0.817	82.00	< 0.01
[Coastal Region] [INR] [ISR]	0.975	0.800	0.877	87.69	< 0.01
[CNR] [CSR] [INR] [ISR]	0.971	0.603	0.928	92.79	< 0.01
[CNR] [CSR] [INRw] [INRe] [ISR]	0.970	0.448	0.946	94.64	< 0.01
[CNR] [INRw] [INRe] [ISR]					
[Caldas, Montejunto, Sintra] [Cercal, Monchique]	0.969	0.313	0.956	95.56	< 0.01
[CNR] [INRw] [INRe] [Guadalupe] [Toledo]					
[Caldas, Montejunto, Sintra] [Cercal, Monchique]	0.968	0.332	0.953	95.27	< 0.01
[Cantabrian, Galicia, Geres] [Estrela, Malcata,					
Mamede] [Caldas, Montejunto, Sintra]					
[Cercal, Monchique] [INRw] [INRe] [ISR]	0.968	0.322	0.953	95.30	< 0.01

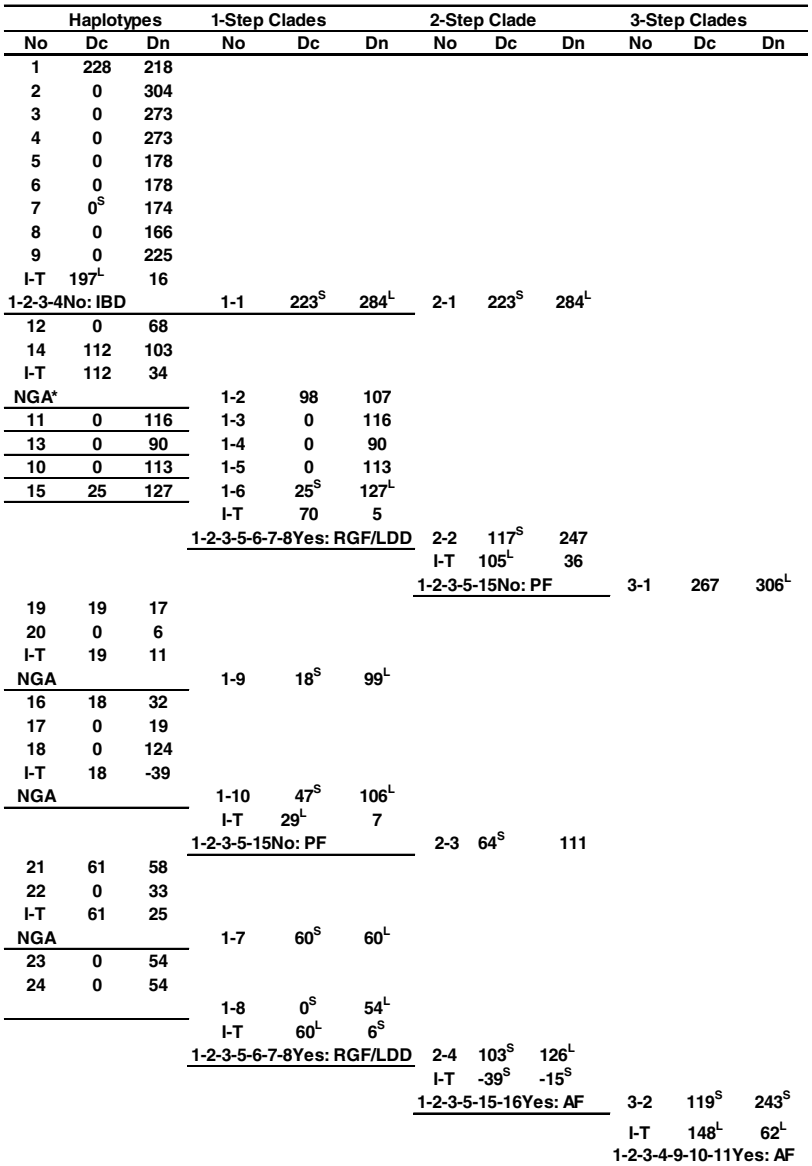

Fig. 4 Results of the nested geographical analysis of the *Lacerta schreiberi* mtDNA haplotypes. The nested design, haplotypes and clade designation are given in Fig. 3. The name or number of each clade are in the first column of each step clades, followed by the clade and nested clade distance. When there are interior and tip clades inside a nesting clade, the average differences between them are given for both distance measures. A superscript S means that the distance measure was significantly small at the 5% level, and a superscript L means that the distance was significantly large. Each horizontal line in each step clade column defines a clade. The sequences of numbers, immediately above the horizontal lines, are the results of the inference key and are shown with the respective biological inference and refer to the adjacent nesting clade. IBD is recurrent gene flow restricted by isolation by distance, RGF/LDD is restricted gene flow but with long distance dispersal, PF is past fragmentation and AF is allopatric fragmentation.

but with some long distance dispersal over intermediate areas not occupied by the species (2–2 and 2–4).

Clade 1–2 was an example of a non–geographical association that was close to statistical significance ($P = 0.054$). However if a significant geographical association of the haplotypes was accepted, then isolation by distance with restricted gene flow was the process inferred through use of Templeton's key. Nested contingency tests, although less powerful, corroborated all the geographical associations found by the nested distance tests (Table 3).

Hierarchical analyse of molecular variance

The NCA subdivision results were supported by the results of analysis of molecular variance. AMOVA identified significant genetic structure at all levels in the several hierarchical designs examined (Table 4). Of these, a six-group combination maximized the allocation of variance among groups (95.56%). These six groups were the five clades and subclades of the phylogenetic analysis with a further subdivison of CSR into two sets of populations [the southern isolates (Monchique and Cercal) and the central semi-isolated populations (Sintra, Montejunto e Caldas)]. Any other combination or further subdivison including the subdivison of ISR, marginally indicated by the phylogenetic analysis (Paulo *et al*. 2001), did not increase the ϕ_{CT} value. In the four-group combination, the among-group variance was considerably smaller. The ϕ_{ST} values were very high for all combinations, suggesting limited gene flow between populations. As expected, intragroup variation among populations (ϕ_{SC}) was markedly reduced in combinations that maximized regional variance.

The two analytical approaches used here emphasize different information contained in the data set, but here were generally in agreement and produced congruent results. The clades recovered by both methods were essentially the same. Five clades were common to the two techniques, while the sixth clade detected by AMOVA corresponded to a subdivison of the NCA clade 2–2 (CSR). For this clade NCA detected a process of restricted gene flow but with some long distance migration across unsuitable habitat.

Discussion

The results confirm and extend the analyses of a previous paper (Paulo *et al*. 2001), where we argued that the major subdivision of the phylogeny probably dates from Pliocene, whilst the secondary divisions of each major clade probably occurred in the Pleistocene, around 1 and 0.6 Ma. This set of Pleistocene genetic events coincides with climatic cycles. The timing of more recent events cannot be estimated accurately due to the inherent variability of the phylogenetic branching pattern and variation introduced by demographic events (Nichols 2001).

Assessment of the predictions

We expected multiple refugia to have been associated with mountains and coastal areas (Prediction i) and that they were generated by allopatric fragmentation (Prediction ii). In all, we could identify a total of four refugia, two clearly allopatric and two probably allopatric. These results were in agreement with the first and second predictions.

The two allopatric fragmentation events detected by NCA, and clearly corroborated by AMOVA, fitted the predicted pattern of persistence. The inland and coastal refugia, were defined by the first allopatric event, an explanation that was supported by the 27 substitution differences found between the two clades (Table 4). Additionally, the inland refuge seems to have been the focus of allopatric isolation between the INR and ISR. It is likely that these were interglacial mountain refugia separated by the Tagus basin. Past fragmentation was inferred by NCA for two situations, that is, the split between CNR and CSR and between INRw and INRe.

The third prediction was not supported by the results on genetic diversity. We did not find that genetic diversity was higher in inland populations than coastal populations. These results do not necessarily refute the 'Phalanx' dispersal for short-distance range shifts since, after the range-shift, these populations could have been affected by long bottlenecks induced by later climatic extremes.

The final two predictions (iv and v) were not endorsed by NCA. The first discrepancy is the interpretation obtained from the inference key of NCA for the clades 2–2 (CSR) and 2–4 (ISR). The key suggested restricted gene flow/dispersal but with long distance migration across intermediate areas not occupied by the species. For clade 2–2 the widespread occurrence of the interior haplotype 14 clearly contrasted with the restricted distribution of all other haplotypes, either tip or interior. For Clade 2–4 it was the distribution of the interior haplotype 21, shared between the two allopatric populations that contrasted with the restricted distribution of the remaining haplotypes.

It seems improbable that this lizard could have migrated over intermediate areas of such currently unsuitable habitat. It is much more plausible that the current genetic pattern reflects a contraction of the range of these populations produced during the postglacial period, maybe even during the Holocene, or even after the Younger Dryas (8 kya) climatic event. A widespread and genetically variable population could have occupied southern areas during a glacial age and then been fragmented by extinction of intermediate populations as the climate warmed. This sequence of events would have reduced population size, range and genetic variability. The current pattern can therefore be explained by different outcomes of drift in each of the remaining isolated or semi-isolated populations.

The second discrepancy concerned the inference of restricted gene flow with isolation by distance assigned to the Clade 2–1 (or 1–1), where we predicted a range expansion and loss of diversity due to founder events. Indeed other lines of evidence corroborate this hypothesis. The star-like phylogeny combined with low nucleotide diversity and high haplotype diversity strongly suggests a population expansion.

The NCA key suggested that the genetic pattern was produced because of isolation by distance. In this case, either the prediction was wrong or, alternatively, that NCA is relatively insensitive to previous range expansions.

Comparison of analytical methods

AMOVA is a widely used and conceptually simple technique that describes current population structure and gene flow based on Wright's island model. It detects geographical subdivison within a species, incorporating the historical relationships between haplotypes with the design of analysis of variance. Its relative importance as a tool to elucidate population subdivision clearly increases with the decreasing influence of older or allopatric phylogenetic events in the population history. However, the technique also has two limitations. It is not designed to indicate the temporal order of events that generated the detected pattern, and it is unable to detect if differentiation between populations is due to restriction caused by current isolation by distance, or other events like past fragmentation or range expansion.

NCA attempts to deal with some of these limitations. However, its inference key lacks some alternative scenarios that can generate the same genetic pattern. But perhaps more important is the central assumption about the distribution of the tip clades (derived) and interior (ancestral) clades. Under the null hypothesis (of non-geographical association), ancestral haplotypes are more widespread than derived ones, because of their age. Consequently, range expansions can be detected by the inversion of this pattern, when tip clades become more widespread and ancestral haplotypes are restricted to the ancestral range (Templeton 1998). However, in the case of clade 2–1, it seems likely that the population went through a population bottleneck before expanding rapidly in size and then range. It shows the characteristic of an expanded population predicted by coalescent theory, that ancestral (interior) haplotypes are more frequent than recent haplotypes (Donnelly & Tavaré 1986; Crandall & Templeton 1993). As the founder events that accompany range expansion are most likely to fix the more frequent haplotypes, they would be expected to have fixed and spread these ancestral haplotypes. Under these particular circumstances NCA seems unable to detect range expansions. This possibility is partially recognized by Templeton (1998) and might merit further investigation, perhaps through the creation of simulated datasets.

Acknowledgements

This work was supported by a NATO Fellowship Portugal Invotan, 03/A/97/PO and by the Institute of Zoology, Zoological Society of London. We thank C. Dias, J. Pinheiro, J. Brito, R. Godinho, J. Gonzalez de la Vega and J. Barnestein for their assistance on the field work, to M. Kadwell and J. Groombridge for their advice on laboratory work. All captures were performed under permit from national government of Portugal, and regional governments of Spain.

References

Allen JM, Huntley B, Watts WA (1996) The vegetation and climate of northwest Iberia over the last 14 000 yr. *Journal of Quaternary Science*, **11**, 125–147.

Anderson S, Bankier AT, Barrell BG *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.

Avise JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, MA.

Ballard OWJ, Kreitman M (1995) Is mitochondrial-DNA a strictly neutral marker? *Trends in Ecology and Evolution*, **10**, 485–488.

Bennett KD, Tzedakis PC, Willis KJ (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103–115.

Brito JC, Abreu FE, Paulo OS, Rosa HD, Crespo EG (1996) Distribution of Schreiber green lizard (*Lacerta-schreiberi*) in Portugal: a predictive model. *Herpetological Journal*, **6**, 43–47.

Brito JC, Godinho R, Luis C, Paulo OS, Crespo EG (1999) Management strategies for conservation of the lizard *Lacerta schreiberi*. Portugal. *Biological Conservation*, **89**, 311–319.

Brito JC, Paulo OS, Crespo EG (1998) Distribution and habitats of Schreiber's green lizard (*Lacerta schreiberi*) in Portugal. *Herpetological Journal*, **8**, 187–194.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Crowley TJ, North GR (1991) *Paleoclimatology*. Oxford University Press, New York.

Donnelly P, Tavaré S (1986) The ages of alleles and coalescent. *Advances in Applied Probability*, **18**, 1–19.

Excoffier L (1990) Evolution of human mitochondrial-DNA — evidence for departure from a pure neutral model of populations at equilibrium. *Journal of Molecular Evolution*, **30**, 125–139.

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

Galan P (1984) Estudio sobre la biometria, folidosis, diseno y coloration de *Lacerta schreiberi*, Bedriaga,1978 (Sauria, Lacertidae). *Alytes*, **2**, 25–54.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13 000 Years Ago*. Cambridge University Press, Cambridge.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.

Kocher TD, Thomas WK, Meyer A *et al*. (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

Marco A (1994) *Autoecología y Biología Reproductora del Lagarto Verdinegro (Lacerta Schreiberi. Bedriaga 1878) en una Poblacíon de Medio Montaña (Siera de Béjar-Salamanca)*. PhD Thesis, Universidad de Salamanca, Spain.

Merilä J, Björklund M, Baker AJ (1997) Historical demography and present day population structure of the Greenfinch, *Carduelis chloris* — An analysis of mtDNA control-region sequences. *Evolution*, **51**, 946–956.

Moritz C, Schneider CJ, Wake DB (1992) Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. *Systematic Biology*, **41**, 273–291.

Nichols RA (2001) Gene trees and species trees are not the same. *Trends in Ecology and Evolution*, **16**, 358–364.

Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. *Heredity*, **72**, 312–317.

Paulo OS (2001) *The phylogeography of reptiles of the Iberian Peninsula*. PhD Thesis, University of London.

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. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **268**, 1625–1630.

Peñalba MC (1994) The history of the Holocene vegetation in northern Spain from pollen analysis. *Journal of Ecology*, **82**, 815–832.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

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

Ramil-Rego P, Muñoz-Sobrino C, Rodríguez-Guitián M, Gómez-Orellana L (1998a) Differences in the vegetation of the North Iberian Peninsula during the last 16 000 years. *Plant Ecology*, **138**, 41–62.

Ramil-Rego P, Rodríguez-Guitián M, Muñoz-Sobrino C (1998b) Sclerophyllous vegetation dynamics in the north of the Iberian Peninsula during the last 16 000 years. *Global Ecology and Biogeography*, **7**, 335–351.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology*, **2**, 321–428.

Riva I (1987) Zoogeografia de *Lacerta schreiberi* Bedriaga, 1878. *Revista Española de Herpetología*, **2**, 49–69.

Rozas J, Rozas R (1999) DNASP, Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15, 174–175.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring. Harbor Laboratory Press, New York.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN, *Version 2.000: a Software for Population Genetics Data Analysis. Geneva, Switzerland*: Genetics and Biometry Laboratory. University of Geneva.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

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

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

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

Zagwijn WH (1992) Migration of vegetation during the Quaternary in Europe. *Courier Forshunginstitut Senckenberg*, **153**, 9–20.