Heredity (2006) 96, 454–463

www.nature.com/hdy

Genealogy of the nuclear β-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone

R Godinho[1], B Mendonça[1], EG Crespo[2] and N Ferrand[1,3]

[1]*CIBIO – Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal;* [2]*Centro de Biologia Ambiental e Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal;* [3]*Departamento de Zoologia e Antropologia, Faculdade de Ciências, Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal*

The study of nuclear genealogies in natural populations of nonmodel organisms is expected to provide novel insights into the evolutionary history of populations, especially when developed in the framework of well-established mtDNA phylogeographical scenarios. In the Iberian Peninsula, the endemic Schreiber's green lizard *Lacerta schreiberi* exhibits two highly divergent and allopatric mtDNA lineages that started to split during the late Pliocene. In this work, we performed a fine-scale analysis of the putative mtDNA contact zone together with a global analysis of the patterns of variation observed at the nuclear β-fibrinogen intron 7 (*β-fibint7*). Using a combination of DNA sequencing with single-strand conformational polymorphism (SSCP) analysis, we show that the observed genealogy at the *β-fibint7* locus reveals extensive admixture between two formerly isolated lizard populations while the two mtDNA lineages remain essentially allopatric. In addition, a private *β-fibint7* haplotype detected in the single population where both mtDNA lineages were found in sympatry is probably the result of intragenic recombination between the two more common and divergent *β-fibint7* haplotypes. Our results suggest that the progressive incorporation of nuclear genealogies in investigating the ancient demography and admixture dynamics of divergent genomes will be necessary to obtain a more comprehensive picture of the evolutionary history of organisms.
Heredity (2006) **96,** 454–463. doi:10.1038/sj.hdy.6800823; published online 5 April 2006

Keywords: phylogeography; nuclear genealogy; *β-fibrinogen* gene; hybrid zone; intragenic recombination; *Lacerta schreiberi*

Introduction

The study of nuclear gene genealogies in natural populations of nonmodel organisms is still in its infancy and only a few reports are available to date (see, for example, Antunes *et al*, 2002; Broughton and Harrison, 2003). Nevertheless, we expect that the generalized application of those genealogies will provide novel insights into the evolutionary history of populations, especially when developed in the framework of well-established mtDNA phylogeographic scenarios. This is because nuclear and mtDNA markers offer contrasting molecular and population properties that include bi-parental and uni-parental inheritance, recombining and nonrecombining histories and a 4:1 ratio of expected coalescence times. These contrasting properties will be most valuable when employed to investigate complex patterns of genetic diversity like those exhibited by hybrid zones where differential rates of sex-related dispersal, intricate dynamics of nuclear admixture and fluctuations in effective population sizes still constitute enormous challenges.

In the last few decades, the accumulation of empirical data documenting the patterns of variation in hybrid zones was paralleled by the development of a strong theoretical background (Barton and Hewitt, 1985, 1989; Harrison, 1993). This is clearly reflected in our current understanding of a number of fundamental issues about hybrid zones that include their origin, maintenance and fate (Barton and Hewitt, 1985), their role as promoters of evolutionary novelties (Arnold, 1997) and the evaluation of the balance between exogenous and endogenous selection processes and implications for speciation (Jiggins and Mallet, 2000). A remarkable pattern associated with most well-studied hybrid zones is the occurrence of novel alleles that are generally absent from the parental populations outside of the hybrid zone (Barton and Hewitt, 1985). When reviewing the subject, Woodruff (1989) considered the previously used expression 'rare allele phenomenon' as inappropriate and introduced the term *hybrizyme* to describe unexpected allelic electromorphs observed in hybrid zones. Among the genetic mechanisms suggested as an explanation for this observation were point mutations (Woodruff, 1989), intragenic recombination (Sage and Selander, 1979; Golding and Strobeck, 1983; Woodruff, 1989), gene

Correspondence: R Godinho, CIBIO/UP Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. E-mail: rgodinho@mail.icav.up.pt
Received 9 August 2005; accepted 6 March 2006; published online 5 April 2006

conversion (Hillis *et al*, 1991) or transposition (Kidwell, 1990), but only detailed DNA sequencing of such new variants could clarify this phenomenon. With the advent of polymerase chain reaction (PCR) and other corresponding technical advances, most of the difficulties associated with the sequencing of nuclear genes and the definition of haplotypes have been removed. It is thus surprising that only a few reports investigating the origin of *hybrizymes* are available (Bradley *et al*, 1993; Hoffmann and Brown, 1995), eventually encouraging Schilthuizen *et al* (2001) to state that 'sequencing studies have shown that hybrizyme alleles are caused by simple point mutations, and cannot be explained by either recombination or transposition'.

The age of most hybrid zones in central and northern Europe is probably very recent and corresponds to the postglacial colonization of biota (Hewitt, 1999, 2000). Various sources of evidence suggest that this process was rapid, concordant with the typically lower genetic diversity of northern populations (Hewitt, 1996). In contrast, populations and species have persisted in southern European refugia over several ice ages, leading to the accumulation of much higher levels of genetic diversity. In the Iberian Peninsula, the molecular signature of deep genetic subdivisions, elevated haplotype richness, and older contact zones have recently been described in a variety of organisms including salamanders (Alexandrino *et al*, 2000), lizards (Paulo *et al*, 2001) and rabbits (Branco *et al*, 2002). Results described for the Schreiber's Green lizard (*Lacerta schreiberi*), an Iberian endemic confined to the stream and river margins of typical Atlantic habitats of the peninsula, are particularly interesting. At the mitochondrial DNA (mtDNA) level, Paulo *et al* (2001) described the occurrence of two highly divergent and allopatric lineages located in the Portuguese and Spanish sides of the Iberian Central System, whereas the analysis of protein polymorphism suggested that those same population groups have been in contact for a long time (Godinho *et al*, 2001, 2003).

In this work, we combine a fine-scale analysis of the mtDNA contact zone with a global analysis of the patterns of variation observed at the nuclear β-fibrinogen intron 7 (*β-fibint7*). We have been able to (i) contrast the phylogeographical patterns observed for mtDNA and the nuclear *β-fibint7*, (ii) more precisely define the geographical location of the hybrid zone between the two divergent groups of lizards, (iii) show evidence of extensive admixture, (iv) describe the occurrence of unique haplotypes restricted to hybrid populations at the *β-fibint7* locus, and (v) suggest that genetic mechanisms other than simple point mutations are likely involved in the generation of those haplotypes.

Materials and methods

Sample collection

We analysed a total of 406 individual tissue samples collected in 19 populations covering the whole distribution area of *L. schreiberi*. Five out of these 19 populations were sampled along an east–west transect that crosses the putative mtDNA contact zone (Figure 1).

Amplification and RFLP analysis of mtDNA

An 850 bp sized fragment of mtDNA, including most of the cytochrome *b* (*cyt b*) gene, was PCR amplified with *Taq* polymerase according to manufacturer's instructions (Promega). The reaction mixture included 0.1 U of *Taq*, 3 mM $MgCl_2$ and approximately 50 ng of genomic DNA. PCR was carried out in a Biometra T3 thermocycler over 35 cycles with annealing temperature set at 50°C. The primers used were gluDG and cb3, described in Palumbi (1996). The Restriction Map subroutine of BioEdit 5.0.9 software (Hall, 1999) was used to locate the site gain or loss for each endonuclease profile and the diagnostic profiles of *Nla*III and *Nla*IV endonucleases were selected to distinguish between the two main mtDNA lineages described in Paulo *et al* (2001) as the coastal and inland clades. In our work, these two main clades were named A and B, respectively. Additionally, the two sublineages described for each main clade by Paulo *et al* (2001) as coastal northern/southern and inland northern/southern and referred to in this study as A_1/A_2 and B_1/B_2, respectively, were distinguished using the *Bse*GI and the *Nla*IV endonucleases. RFLP patterns were visualized under UV light after electrophoretic separation on 2% agarose gels.

Amplification, SSCP analysis and sequencing of β-fibrinogen

The entire *β-fibint7* was PCR-amplified using FIB-B17U and FIB-B17L primers as described by Prychitko and

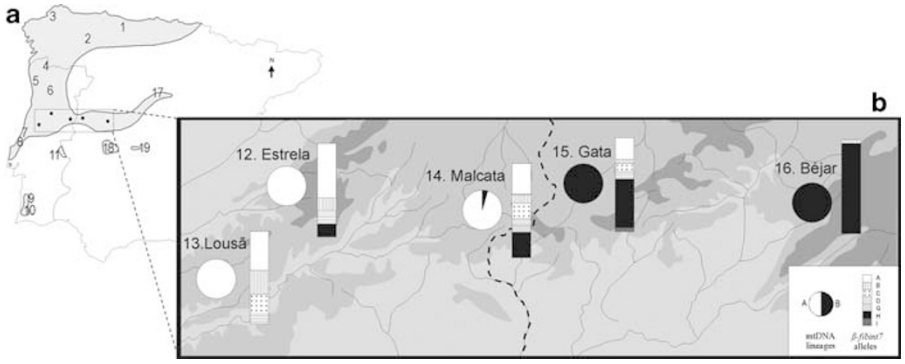

Figure 1 (**a**) Distribution area of *L. schreiberi* in the Iberian Peninsula (shaded) and sampling locations: 1 – Asturias, 2 – Ancares, 3 – Ferrol, 4 – Gerês, 5 – Gião, 6 – Montemuro, 7 – C. Rainha, 8 – Montejunto, 9 – Cercal, 10 – Monchique, 11 – S. Mamede, 17 – Guadarrama, 18 – Guadalupe, 19 – Toledo. (**b**) Detailed map with indication of relief and rivers along a west–east transect crossing the putative mtDNA contact zone. The approximate frequencies of the two mtDNA lineages and of the *β-fibint7* alleles are represented by pie charts and bars, respectively.

Moore (1997), but increasing the annealing temperature to 57°C. A preliminary survey of nucleotide polymorphism was obtained by sequencing *β-fibint7* (788 bp) in a total of 12 individuals sampled in different localities. This strategy allowed the identification of a cluster of polymorphic positions close to the 5′ end of *β-fibint7* covering less than 130 bp and the subsequent design of PCR-SSCP primers (Fib7LsF 5′-CTA GTC ATA CCC AAA TGT G-3′ (forward) and Fib7LsR 5′-CTA ATT CAG GGG GAG CTA-3′ (reverse)) for an extended population analysis that included a total of 406 samples. The PCR mixture was made for a total volume of 10 μl using 0.1 U of *Taq* polymerase (Promega), 3 mM $MgCl_2$ and approximately 25 ng of DNA. Amplifications were done in a Biometra T3 thermocycler over 32 cycles with annealing temperature set at 54°C and a fragment of 180 bp was obtained. Discrimination of the SSCP conformers was made in a 12% polyacrylamide gel (39:1 acrylamide: methylbisacrylamide) with 0.5 × TBE buffer on a vertical electrophoresis system (BIORAD Protean II). The electrophoresis was performed at a constant voltage of 250 V and constant temperature of 12°C for 15 h. Results were visualized by silver staining. In addition, a sample of three to six homozygous individuals for each SSCP allele collected in different populations was also sequenced for the entire *β-fibint7* region in order to evaluate the consistency of SSCP results and explore further variation. In the case of low frequency alleles, the identification of double-peaks in electrophoregrams from sequenced heterozygous individuals allowed the use of ARMS-primers (Amplification Refractory Mutation System, Newton *et al*, 1989) to unambiguously identify haplotypes. Sequencing followed the ABI Prism BigDye Terminator Cycle sequencing protocol in an ABI PRISM 310 Genetic Analyser. The sequences of *β-fibint7* alleles in *L. schreiberi* were deposited in GenBank (Accession numbers DQ097101, DQ097102 and DQ299919 to DQ299932). The separation of haplotypes B and C was difficult in routine SSCP gels and confirmation of their identity was additionally obtained by examining RFLP patterns with the use of the endonuclease *Stu*I. Four closely related *Lacerta* species (*L. viridis*, *L. agilis*, *L. strigata* and *L. trilineata*) were used as outgroups (Godinho *et al*, 2005).

Data analysis

Allele frequencies and measures of genetic variation such as heterozygosity and number of alleles were calculated using the Genetix software, version 4.01 (Belkhir *et al*, 2000). Estimates of nucleotide diversity (π) were obtained with the DnaSP 3.51 software (Rozas and Rozas, 1999) while *β-fibint7* nucleotide contents and the transition/tranversion ratio were calculated using the MEGA 2.1 software (Kumar *et al*, 2000). The Network 3.1 software (Röhl, 2000) was used to construct reduced median-joining networks for both the *β-fibint7* SSCP alleles and the entire intron 7 fragments. We further explored the information content of the *β-fibint7* geneology by splitting the total gene tree according to the four sublineages A_1, A_2, B_1 and B_2. These networks were used to (i) investigate phylogenetic relationships between haplotypes; (ii) evaluate the concordance between mtDNA and *β-fibint7* phylogeographic patterns; and (iii) evaluate the effectiveness of the SSCP fragment in capturing the evolutionary trajectory of the entire *β-fibint7*. The minimum number of recombination events in the history of the sampled sequences was estimated using the four-gamete test (Hudson and Kaplan, 1985) as implemented in the DnaSP 3.51 software (Rozas and Rozas, 1999).

Results

The distribution of mtDNA lineages in the hybrid zone

A total of 107 lizard samples collected in five locations along an east–west transect across the putative hybrid zone of the two major divergent population groups were screened with *Nla*III and *Nla*IV for the generation of distinct cleavage patterns distinguishing the mtDNA lineages A and B. On the western side, the Portuguese populations of Lousã ($n = 10$) and Estrela ($n = 28$) were found to be fixed for mtDNA lineage A, while on the eastern side the Spanish Central System populations of Béjar ($n = 22$) and Gata ($n = 21$) showed only lineage B. In contrast, the intermediate location of Malcata exhibited both lineages, although only one individual out of 25 was typed as mtDNA B (Figure 1). These results are essentially in accordance with those reported by Paulo *et al* (2001), but increased sample sizes allowed the identification of an admixed population that reveals the approximate location of the mtDNA secondary contact zone. In addition, two to three samples from the remaining 14 populations, together with those in the above-mentioned west–east transect, were digested with *Bse*GI and *Nla*IV for the identification of sublineages (Table 1).

Polymorphism of *β-fibint7*

A 180 bp-sized fragment of *β-fibint7* was resolved generating a total of 10 haplotypes using SSCP (Figure 2). With the exception of Béjar, the other populations included in the west–east transect revealed high levels of expected heterozygosity and nucleotide diversity when compared to the remaining 14 lizard populations (Table 1), which is in agreement with our hypothesis of a well-defined hybrid zone in this geographical location. Haplotype frequencies showed considerable variation across populations, resulting in a high amount of population structure ($F_{st} = 0.39$). However, the F_{st} value was considerably lower ($F_{st} = 0.25$) when only the five populations included in the west–east hybrid zone were analysed, suggesting the occurrence of restricted gene flow between the two highly divergent population groups. Haplotypes A and H were typical of north-western Iberia and the Spanish Central System, respectively, and showed opposite clinal gradients across the hybrid zone, while haplotype B was less frequent but widespread. Western isolated populations exhibited high frequencies of haplotypes E and F, while more central isolates showed the occurrence of haplotype J. Central populations of the main distribution area were characterized by haplotypes C and D, whereas two rare and private haplotypes, G and I, were found in the populations of Malcata and Gata, respectively.

The sequence of the whole *β-fibint7* locus in *L. schreiberi* provided a 788 bp fragment, including 71 bp of exon 7 and 103 bp of exon 8. A total of 20 segregating sites (eight in the SSCP fragment) were found, from which nine

Table 1 Number of individuals sampled (*N*), mtDNA lineage, haplotypic frequencies, number of haplotypes per population, expected heterozygosity (He) and nucleotide diversity (π) at the nuclear *β-fibint7* locus. Boxes highlight the highest values observed for the genetic diversity parameters

	mtDNA lineage	N	β-fibint7 *haplotypes*										*No. of haplotypes*	*HE*	$\pi \pm SD$
			A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*			
1. Asturias	A_1	13	0.92	0.08	—	—	—	—	—	—	—	—	2	0.136	0.00304±0.00288
2. Ancares	A_1	21	0.93	0.07	—	—	—	—	—	—	—	—	2	0.148	0.00153±0.00187
3. Ferrol	A_1	19	0.76	0.21	0.03	—	—	—	—	—	—	—	3	0.383	0.00400±0.00340
4. Gerês	A_1	26	0.71	0.29	—	—	—	—	—	—	—	—	2	0.419	0.00470±0.00376
5. Gião	A_1	26	0.71	0.29	—	—	—	—	—	—	—	—	2	0.419	0.00470±0.00376
6. Montemuro	A_1	17	0.74	0.20	0.03	0.03	—	—	—	—	—	—	4	0.428	0.00430±0.00358
7. C.Rainha	A_2	26	—	0.04	—	—	0.88	0.08	—	—	—	—	3	0.214	0.00245±0.00247
8. Montejunto	A_2	22	—	0.02	—	—	0.71	0.27	—	—	—	—	3	0.439	0.00315±0.00290
9. Cercal	A_2	17	0.65	—	—	—	—	0.35	—	—	—	—	2	0.471	0.00262±0.00260
10. Monchique	A_2	28	0.39	—	—	—	0.39	0.22	—	—	—	—	3	0.657	0.00543±0.00415
11. S.Mamede	A_1	35	0.41	0.17	—	—	—	—	—	0.16	—	0.26	4	0.705	0.00657±0.00473
12. Estrela[a]	A_1	27	0.59	0.13	—	0.13	—	—	—	0.15	—	—	4	0.604	0.00633±0.00463
13. Lousã[a]	A_1	10	0.45	0.25	0.20	0.10	—	—	—	—	—	—	4	0.732	0.00620±0.00471
14. Malcata[a]	A_1/B_1	25	0.38	0.10	0.18	0.04	—	—	0.06	0.24	—	—	6	0.766	0.00626±0.00460
15. Gata[a]	B_1	22	0.26	0.03	0.14	0.05	—	—	—	0.50	0.02	—	6	0.669	0.00637±0.00467
16. Béjar[a]	B_1	19	—	0.03	—	—	—	—	—	0.97	—	—	2	0.053	0.00059±0.00111
17. Guadarrama	B_1	22	—	0.32	—	—	—	—	—	0.68	—	—	2	0.444	0.00499±0.00393
18. Guadalupe	B_2	13	0.73	—	—	—	—	—	—	—	—	0.27	2	0.409	0.00229±0.00241
19. Toledo	B_2	18	1.00	—	—	—	—	—	—	—	—	—	1	—	—

[a]Indicates the five populations included in the east-west transect that crosses the putative mtDNA contact zone.

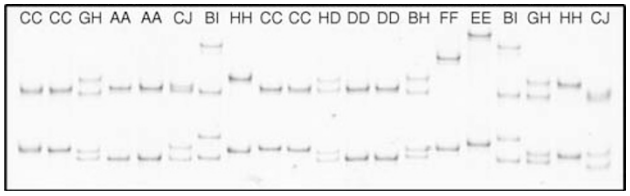

Figure 2 Separation of allelic variants of the *β-fibint7* locus (180 bp) by SSCP analysis in 12% polyacrylamide gels. Alleles B and C were further confirmed with the use of the endonuclease *Stu*I.

(five in the SSCP fragment) were parsimony informative (Table 2). Two 1 bp deletions and one insertion of a trinucleotide repeat were found. A transition/transversion substitution ratio of 1.7 and an A + T biased content of 63% were observed in *β-fibint7* sequences, which were in agreement with values reported for this intron in other species (Prychitko and Moore, 1997, 2000; Johnson and Clayton, 2000).

Phylogenetic relationships of SSCP *β-fibint7* haplotypes

A sample of 3–12 chromosomes for each SSCP allele was sequenced and confirmed that all mutations were detected with the technique described. This result thus highlights the high-resolving power of SSCP in the detection of DNA sequence variation (Orti *et al*, 1997; Sunnucks *et al*, 2000). Haplotype genealogy was displayed with a reduced median network and inference of the ancestral haplotype was made from *β-fibint7* sequences obtained in four other closely related *Lacerta* species. Haplotype C, present in only a few populations, contains the ancestral state at each of the eight polymorphic sites in the *β-fibint7* locus and was, accordingly, identified as the root of the network. The two major haplotypes A and H are one and two mutations distant from the root, respectively, and most of the remnant haplotypes are easily connected in the network (Figure 3a). The remarkable exception is the position of both low frequency haplotypes G and J that are at the origin of the two loops in the network thus suggesting the occurrence of recurrent mutation or recombination. We will argue below that these rare haplotypes are the product of intragenic recombination and not the result of recurrent mutation.

When the haplotype network is split according to the essentially allopatric mtDNA lineages and sublineages, we gain additional information on the history of the *β-fibint7* polymorphism. The populations described by Paulo *et al* (2001) as exhibiting the ancestral mtDNA sub-lineages A_1 and B_1 are characterized by the presence in high frequency of *β-fibint7* haplotypes A and H, respectively, and also show a diverse array of other less frequent haplotypes (Figure 3b). In contrast, populations possessing the more derived mtDNA sub-lineages A_2 and B_2 show much lower levels of diversity, a result that is consistent with a more recent colonization of the central and southern areas of the Iberian Peninsula, as described by Paulo *et al* (2001). In the western areas of Iberia, the mtDNA sublineage A_2 is concordant with the occurrence of the derived *β-fibint7* haplotypes E and F, which mark the southwestern range expansion of *L. schreiberi*. A similar phenomenon is not apparent in the Spanish central isolates that are characterized by the mtDNA sublineage B_2 but that are almost fixed for the *β-fibint7* haplotype A.

Phylogenetic relationships of extended *β-fibint7* haplotypes and evidence for recombination

When the same sample of chromosomes was sequenced for the entire *β-fibint7*, a similar network was obtained (Figure 4), clearly confirming the usefulness of our SSCP approach in determining the major haplotypes and capturing their respective phylogenetic relationships. Interestingly, the basic difference between the two networks is the increase in the number of variants

Table 2 (a) Variable positions found at the *β-fibint7* locus in *L. schreiberi* defining a total of 16 haplotypes. The fragment (180 bp) corresponding to the SSCP analysis is shaded. The ancestral haplotype was derived from the analysis of *β-fibint7* sequences in four closely related *Lacerta* species. Position 1 corresponds to position 8636 of human *β*-fibrinogen. (b) Reconstruction of the hypothetical recombination events originating haplotypes G and J

Haplotypes	1 2	8 3	1 2 6	1 5 8	1 6 7	1 9 3	1 9 5	1 9 8	2 3 2	2 4 8	2 9 9	3 0 6	3 8 7	3 9 7	3 9 8	—	4 9 2	5 3 2	6 0 7	6 1 2
(a)																				
Ancestral	C	C	G	T	G	A	C	C	G	T	G	T	A	G	G	.	T	C	T	T
A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A_{v1}	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	—	C
A_{v2}	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	AAT	.	.	.	.
D	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
E	.	.	A	G	.	.	.	T	.	.	.	.	.	.	A	.	A	.	.	.
F	.	.	A	G	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.
B	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.
B_{v1}	.	.	.	.	.	.	.	.	A	.	.	G	G	.	.	.	.	.	.	.
C	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
H	.	.	.	.	.	G	.	.	.	—	.	.	.	.	.	.	.	A	.	.
H_{v1}	.	.	.	.	.	G	.	.	.	—	C	.	.	.	.	.	.	A	.	.
H_{v2}	.	.	.	.	.	G	.	.	.	—	.	.	.	T	.	.	.	A	.	.
H_{v3}	T	T	.	.	.	G	.	.	.	—	.	.	.	.	.	.	.	A	.	.
I	.	.	.	.	A	G	.	.	.	—	.	.	.	.	.	.	.	A	.	.
G	.	.	A	.	.	.	.	.	.	—	.	.	.	.	.	.	.	A	.	.
J	.	.	A	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.
(b)																				
A	.	.	A	.	.	A	.	.	.	T	.	.	.	.	.	.	.	C	.	.
G	.	.	A	.	.	A	.	.	.	—	.	.	.	.	.	.	.	A	.	.
H	.	.	G	.	.	G	.	.	.	—	.	.	.	.	.	.	.	A	.	.
A	.	.	A	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.
J	.	.	A	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.
B	.	.	G	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.

derived from the two major *β-fibint7* haplotypes, A and H. The complete *β-fibint7* network additionally reveals that (i) at this level of resolution, haplotype C is no longer the root of the network due to a T→A transversion at position 306, (ii) the ancestral haplotype was not found in any of the studied populations, and (iii) the derivation of haplotypes E and F is further evidenced by the G→A transition and T→A transversion at positions 398 and 492, respectively.

The additional information provided by the extended *β-fibint7* haplotypes supports the interpretation that recombination is at the origin of the rare haplotypes G and J. Accordingly, two different criteria indicate that the possibility of recurrent mutation is highly unlikely. First, both haplotypes occur in populations that show evidence of extensive admixture between the two major population groups of *L. schreiberi*. Haplotype G is private from the Malcata population, which is precisely the only location where both mtDNA lineages A and B were found in sympatry (Table 1 and Figure 2). A close examination of the characteristics of haplotype G (Table 2b) clearly suggests that this haplotype is a mosaic of the more common parental haplotypes A and H, indicating that recombination is the probable mechanism involved in its origin. Haplotype J was found in the central isolate of S. Mamede, a peculiar population that exhibits mtDNA lineage A but clusters with populations from the Spanish Central System when polymorphic proteins are analysed (Godinho *et al*, 2003), thus suggesting another situation of extensive admixture between two divergent groups. Although our evidence is not as clear as for haplotype G, we suggest that haplotype J is the result of a recombination event between the two more common haplotypes found in S. Mamede, A and B (Table 2b). Second, none of the mutations implied in the origin of these haplotypes can be attributed to the hypermutability of CpG dinucleotides. In vertebrate genomes, CpG dinucleotides are known to be hotspots for nucleotide substitutions because of their high content of 5-methylcytosine (5mC). Methylation of cytosine results in a high level of mutation due to the propensity of 5mC to undergo deamination to form thymine, which explains the high frequency of C-to-T and G-to-A transitions (reviewed in Cooper and Krawezak, 1993). Finally, the four-gamete test identified four pairs of sites with the four gametic types (Table 2), of which two (positions 126, 248 and 126, 532) are involved in the recombination between haplotypes A and H, and the other two (positions 126, 232 and 126, 306) in the recombination between haplotypes A and B.

Discussion

Comparing mtDNA and *β*-fibint7 phylogeographies in Lacerta schreiberi

The previous phylogeographic analysis of the mtDNA molecule in *Lacerta schreiberi* suggested that the major differentiation process in two main lineages was initiated in the late Pliocene. Populations of this species would have persisted through the Pleistocene in allopatric

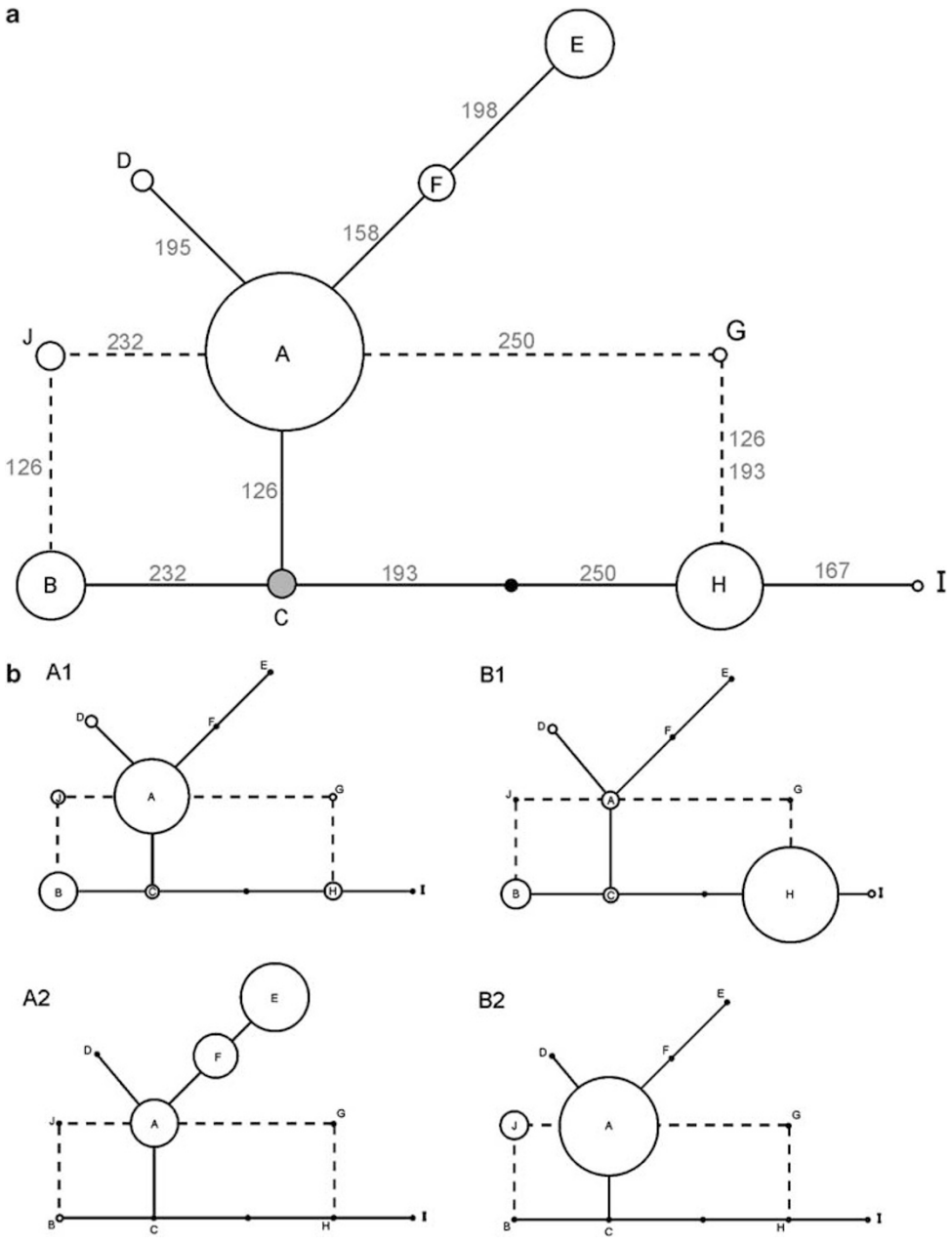

Figure 3 (**a**) Reduced median-joining network representing the phylogenetic relationships between the 10 SSCP *β-fibint7* haplotypes. Circle size is proportional to the frequency of each haplotype in the total sample. The shaded circle indicates the root of the network, black points represent potential intermediates and dashed lines indicate loops that may result from recombination. (**b**) Parsimony networks for four groups of amalgamated populations according to the mtDNA sublineages (A_1, A_2, B_1 and B_2). The population of Malcata was included in group A_1. Circle size is proportional to the frequency of each haplotype in the respective amalgamated sample.

refugia, but the well known climatic fluctuations of this period additionally caused the emergence of allopatric mtDNA sublineages (Paulo *et al*, 2001). Although Hudson and Turelli (2003) convincingly argued that the use of mtDNA divergence to predict monophyly of nuclear loci is not reliable, we believe that the remarkable phylogeographic structure of this lizard offers an excellent opportunity to evaluate the utility of nuclear gene genealogies in obtaining a more complete picture of the evolutionary history of the species. The results of our study confirmed this prediction and showed that the global analysis of variation at the *β-fibint7* locus provides additional and important information on the processes that shaped the present-day genetic architecture of *L. schreiberi*.

The Pliocene divergence between the two major population groups located in coastal and inland Iberia left a clear molecular signature in the *β-fibint7* network by separating haplotypes A and H with four mutations (Figure 4). These haplotypes represent the two main clades of that network and while haplotype A exhibited high frequencies in all populations showing mtDNA sublineage A_1, haplotype H characterized populations possessing sublineage B_1. Moreover, both haplotypes

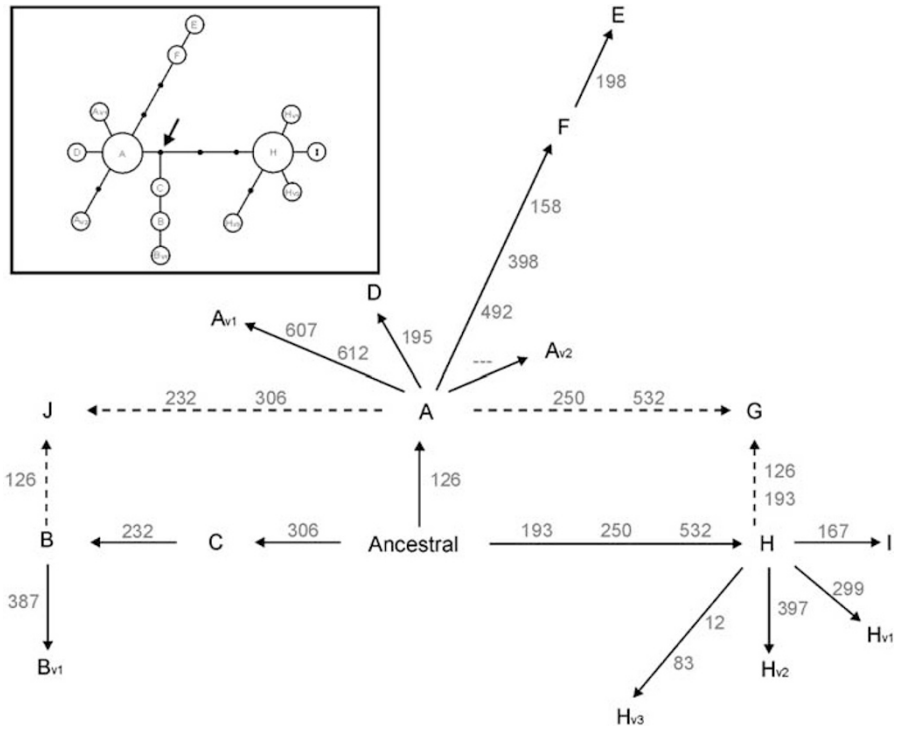

Figure 4 Reduced median-joining network representing the phylogenetic relationships between the 16 extended *β-fibint7* haplotypes. The ancestral allele is indicated. Numbers correspond to single point mutation events. Dashed lines indicate loops that may result from recombination. The inset represents the estimated network after the removal of recombinant haplotypes. The root is indicated by an arrow, and black points show potential intermediates.

showed opposite clinal variation along the west–east transect studied in the Iberian Central System, indicating the occurrence of gene flow between two formerly isolated populations. Paulo *et al* (2002) further suggested that central and southern isolates were the result of old expansions to the south during a glacial age followed by a recent contraction and the extinction of intermediate populations. In coastal regions, the populations were characterized by the derived mtDNA sublineage A_2 (Table 1) and concordantly showed the *β-fibint7* haplotypes E and F, which are clearly derived from the more common and ancestral haplotype A. Surprisingly, however, the inland isolated populations of Guadalupe and Toledo that exhibited the mtDNA sublineage B_2 were characterized by the *β-fibint7* haplotypes A and J. Given that the divergence between mtDNA sublineages B_1 and B_2 is almost twice the divergence between A_1 and A_2, we could expect the occurrence of concordantly derived *β-fibint7* haplotypes in those populations. We suggest that their absence may be explained by the occurrence of recent gene flow from western populations, eventually driving the extinction of those putative haplotypes. This hypothesis is supported by (i) the presence of haplotype J, which is likely to have originated in the more western mountain of S. Mamede (see below); (ii) additional nuclear data indicating that recent gene flow is stronger from the west than from the geographically closer populations located in the mountains of Gredos and Guadarrama in the east (Godinho, 2004); and (iii) environmental and climatic modelling of the species distribution showing a possible corridor linking S. Mamede with Guadalupe and Toledo (Godinho, 2004).

Evidence for recombination at the β-fibint7 in the hybrid zone

In recent years, sequencing of a couple of alleles that are exclusively found in hybrid zones suggested that they may have originated by simple point mutations instead of the more likely process of recombination (Bradley *et al*, 1993; Hoffmann and Brown, 1995). Our results, however, suggest that the two haplotypes G and J, detected in admixed populations, are likely the product of recombination between the more common and widespread parental haplotypes A/H and A/B, respectively.

The population of Malcata is notable because it corresponds to the single geographical location where the two deeply divergent mtDNA lineages were found in sympatry (Figure 1). The highest values of genetic diversity at the *β-fibint7* locus (number of haplotypes, expected heterozygosity and nucleotide diversity; see Table 1) and also from other nuclear data (Godinho, 2004; Godinho *et al*, 2006) give additional evidence that this population is close to the centre of a hybrid zone. This scenario of extensive admixture between two groups of populations that show evidence of long-term historical isolation was ultimately responsible for the contact of the distinct *β-fibint7* haplotypes A and H, creating an opportunity for the generation of recombinant haplotypes. In the Malcata population, the private haplotype G is easily viewed as a mosaic of the more common haplotypes A and H (Table 2). Additionally, the absence of hotspot mutation motifs together with the equitable base composition and nucleotide substitution probabilities (see Prychitko and Moore, 2000) decrease the likelihood of recurrent mutations and strengthens the confidence in our interpretation. Essentially, the same

criteria support the hypothesis that haplotype J originated in the population of S. Mamede by a recombination event between the more common haplotypes A and B. However, more data are necessary to draw a firm conclusion in this respect.

The utility of nuclear genealogies in well-defined phylogeographic contexts

The extraordinary explosion of phylogeographical studies since the seminal paper of Avise *et al* (1979) has made an enormous contribution to our understanding of evolution and created a new research discipline (Avise, 2000). However, the fact that most phylogeographical studies are based solely on mtDNA, which possesses well-recognized limitations (Zhang and Hewitt, 2003), has hampered further developments in the field. Nevertheless, it is now clear that recent progress in molecular genetic techniques combined with the development of new statistical tools will lead to an increased use of nuclear DNA polymorphisms. Our study of the nuclear *β-fibint7* locus in a lizard species for which a well-established phylogeographical scenario was previously described highlights the advantages of those types of polymorphisms and offers some prospects for future work.

To our knowledge, this is the first report using the *β-fibint7* nuclear marker for an intraspecific study in a vertebrate species. Although this intron has been previously characterized as having a substitution rate approximately four to six times slower than mtDNA (Johnson and Clayton, 2000; Weibel and Moore, 2002), we found a total of 20 segregating sites (Table 2) that allowed the simple construction of highly informative haplotype networks (Figures 3 and 4). Together with the four-gamete test, these networks additionally revealed that two out of ten different *β-fibint7* haplotypes are probably recombinants, shedding new light on the population history and dynamics of the nuclear genome of *L. schreiberi*.

Recently, Broughton and Harrison (2003) examined four nuclear gene genealogies in three closely related but morphologically, behaviourally and ecologically distinct *Gryllus* species, and concluded that those genealogies were of limited phylogeographical use but provided important insights into the historical, demographic and selective forces that shaped North American field crickets. In this study, we offer a somewhat different perspective in which the genealogy of the nuclear *β-fibint7* combined with mtDNA information considerably improves our understanding of the evolutionary history of a species exhibiting a pronounced genetic structure and, simultaneously, a remarkable uniformity with respect to morphology, behaviour and ecology.

Implications for the study of hybrid zones

Most well-known hybrid zones are probably very recent and resulted from the post-glacial colonization of previously unsuitable territories by expanding genomes that were formerly constrained to low latitude refugia (Hewitt, 2000). Consequently, we may consider that these hybrid zones are very limited windows into the evolutionary processes for which relevant timescales are orders of magnitude higher. However, it is becoming clear that much older hybrid zones may exist in low latitude refugia like the Iberian Peninsula. Apart from the lizard species that is the focus of this study, recent examples of deep lineage divergence followed by secondary contacts have been described in a variety of organisms (Alexandrino *et al*, 2000; Branco *et al*, 2002; García-París *et al*, 2003; Sequeira *et al*, 2005). This is certainly a consequence of the extreme topographical heterogeneity and associated habitat diversity of the Iberian Peninsula, resulting in multiple sub-refugia within the refugium.

Our data suggest that divergent lineages of *L. schreiberi* have been contracting and expanding repeatedly during the glacial and interglacial periods of the Pleistocene, thus creating ample opportunities for ancient admixture and the establishment of an old hybrid zone (Godinho, 2004). We additionally provide compelling evidence in favour of an important role of recombination in the generation of new alleles in hybrid zones, thus contradicting the recent statement by Schilthuizen *et al* (2001) but clearly confirming expectations based on both empirical and theoretical evidence (Golding and Strobeck, 1983; Woodruff, 1989). We note, however, that the generation of hybrizymes through intronic recombination depends on the accumulation of nonsynonymous divergence in the flanking exons. While future sequencing of *β*-fibrinogen exons are necessary to effectively demonstrate the occurrence of hybrizyme products at this locus, we anticipate that the high degree of population subdivision observed in a variety of organisms in low latitude refugia has likely resulted in divergent protein products that can be at the origin of novel variants in admixed populations. Finally, our results support a high frequency of occurrence of recombination even in very short nuclear DNA sequences, as recently described by Ibrahim *et al* (2002), and suggest that the multiple pulses of expansion and admixture that characterized the Pleistocene era in Southern Europe will be most fruitfully investigated through careful studies incorporating the detection of recombinant haplotypes, eventually leading to an accurate perception of the dates of contacts between hybridizing populations (Baird, 1995). Our demonstration of using the properties of recombination to have more insights into past processes of population contacts and admixture gain relevance in light of the recently described lack of phylogeographical structure in European mammals before the last ice age (Hofreiter *et al*, 2004). These authors proposed that the deep mtDNA divergence times observed in many different species do not reflect long separations of populations but simply the beginning of the differentiation process that happened frequently before the Pleistocene. Subsequent processes of expansion, recolonization and admixture are erased at each new ice age because mtDNA clades remain fixed at the different refugia (Figure 5 of Hofreiter *et al*, 2004) and, at the same time, are unable to retain the memory, through recombination, of periods of extensive mixing.

Taken together with present and previously published mtDNA data, our *β-fibint7* study in the Schreiber's green lizard anticipates the invaluable utility of autosomal DNA markers in future characterizations of the admixture dynamics of divergent genomes, ultimately helping in the elucidation of biological phenomena occurring in hybrid zones that are not revealed by haploid markers like mtDNA or the mammalian Y-chromosome.

Acknowledgements

We thank C Moritz, J Rocha and S Baird for helpful discussions and DJ Harris, JC Brito, P Alexandrino, PC Alves and S Weiss for manuscript comments. We thank the Subject Editor and two anonymous referees for their thorough and helpful comments on a previous version of the manuscript. We are grateful to all colleagues that assisted in fieldwork, specially C Luís. We also thank the collecting permits given by Portuguese and Spanish authorities. This work was partially supported by POCTI/BSE/40987/2001 and POCTI/BSE/41395/2001 projects. R Godinho worked under a PhD grant (PRAXIS XXI/BD/16087/98) and is presently supported by a post-doctoral grant (SFRH/BPD/12723/2003). N Ferrand was supported by a sabbatical grant (SFRH/BSAB/318/2002).

References

Alexandrino J, Froufe E, Arntzen JW, Ferrand N (2000). Genetic subdivision, glacial refugia and postglacial recolonization in the golden-striped salamander, *Chioglossa lusitanica* (Amphibia: Urodela). *Mol Ecol* **9**: 771–781.

Antunes A, Templeton AR, Guyomard R, Alexandrino P (2002). The role of nuclear genes in intraspecific evolutionary inference: genealogy of the transferring gene in the brown trout. *Mol Biol Evol* **19**: 1272–1287.

Arnold M (1997). *Natural Hybridization and Evolution*. Oxford University Press: New York.

Avise JC (2000). *Phylogeography, the history and formation of species*. Harvard University Press: Cambridge, Mass.

Avise JC, Lansman RA, Shade RO (1979). The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution of the genus *Peromyscus*. *Genetics* **92**: 279–295.

Baird SJE (1995). A simulation study of multilocus clines. *Evolution* **49**: 1038–1045.

Barton NH, Hewitt GM (1985). Analysis of hybrid zones. *Annu Rev Ecol Syst* **16**: 113–148.

Barton NH, Hewitt GM (1989). Adaptation, speciation and hybrid zones. *Nature* **341**: 497–503.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) GENETIX 4.01, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.

Bradley RD, Bull JJ, Johnson AD, Hillis DM (1993). Origin of a novel allele in a mammalian hybrid zone. *Proc Natl Acad Sci USA* **90**: 8939–8941.

Branco M, Monnerot M, Ferrand N, Templeton AR (2002). Postglacial dispersal of the european rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* **56**: 792–803.

Broughton RE, Harrison RG (2003). Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* **163**: 1389–1401.

Cooper DN, Krawezak M (1993). *Human gene mutation*. BIOS Scientific: Oxford.

García-París M, Alcobendas M, Buckley D, Wake DB (2003). Dispersal of viviparity across contact zones in Iberian populations of fire salamanders (*Salamandra*) inferred from discordance of genetic and morphological traits. *Evolution* **57**: 129–143.

Godinho R (2004) Evolutionary history of the Schreiber's green lizard, *Lacerta schreiberi*: Phylogenetic analysis, reconstruction of processes of population expansion, contraction and miscigenation, and integration in the biogeographic context of Northwestern Iberian Peninsula. PhD thesis: University of Lisbon.

Godinho R, Crespo EG, Ferrand N, Harris DJ (2005). Phylogeny and evolution of the green lizards, *Lacerta* spp. (Squamata: Lacertidae) based on mitochondrial and nuclear DNA sequences. *Amphibia-Reptilia* **26**: 271–285.

Godinho R, Domingues V, Crespo RG, Ferrand N (2006). Extensive intraspecific polymorphism detected by SSCP at the nuclear C-*mos* gene in the endemic Iberian lizard *Lacerta schreiberi*. *Mol Ecol* **15**: 731–738.

Godinho R, Ferrand N, Crespo EG (2001). Phylogeography of the Iberian Schreiber's Green Lizard (*Lacerta schreiberi*): preliminary data on mitochondrial and nuclear markers reveal discrepant patterns. *Mertensiella* **13**: 33–39.

Godinho R, Paulo OS, Ferrand N, Luis C, Rosa HD, Crespo EG (2003). Major patterns of population differentiation in the Iberian Schreiber's Green Lizard (*Lacerta schreiberi*) inferred from protein polymorphism. *Herpet J* **13**: 35–42.

Golding GB, Strobeck C (1983). Increased number of alleles found in hybrid populations due to intragenic recombination. *Evolution* **37**: 17–29.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.

Harrison RG (1993). *Hybrid Zones and The Evolutionary Process*. Oxford University Press: New York.

Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc* **58**: 247–276.

Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biol J Linn Soc* **68**: 87–112.

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

Hillis DM, Moritz C, Porter CA, Baker RJ (1991). Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308–310.

Hoffmann SMG, Brown WM (1995). The molecular mechanism underlying the 'rare allele phenomenon' in a subspecific hybrid zone of the California Field Mouse, *Peromyscus californicus*. *J Mol Evol* **41**: 1165–1169.

Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conard N *et al* (2004). Lack of phylogeography in European mammals before the last glaciation. *Proc Natl Acad Sci USA* **101**: 12963–12968.

Hudson RR, Kaplan NL (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.

Hudson RR, Turelli M (2003). Stochasticity overrules the 'Three-Times Rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**: 182–190.

Ibrahim KM, Cooper SJB, Hewitt GM (2002). Testing for recombination in a short nuclear DNA sequence of the European meadow grasshopper, *Chorthippus parallelus*. *Mol Ecol* **11**: 583–590.

Jiggins CD, Mallet J (2000). Bimodal hybrid zones and speciation. *Trends Ecol Evol* **15**: 250–255.

Johnson KP, Clayton DH (2000). Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). *Mol Phylogenet Evol* **14**: 141–151.

Kidwell MG (1990). Evolutionary aspects of hybrid dysgenesis in *Drosophila*. *Can J Zool* **68**: 1716–1726.

Kumar S, Tamura K, Jakobsen I, Nei M (2000). *MEGA: Molecular Evolutionary Genetics Analysis*, v. 2.1. The Pennsylvania State University: University Park.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N *et al* (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* **17**: 2503–2516.

Orti G, Hare MP, Avise JC (1997). Detection and isolation of nuclear haplotypes by PCR-SSCP. *Mol Ecol* **6**: 575–580.

Paulo OS, Dias C, Bruford MW, Jordan WC, Nichols RA (2001). The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proc Biol Sci* **268**: 1625–1630.

Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002). Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard. *Mol Ecol* **11**: 809–819.

Palumbi SR (1996). Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Moritz C, Mable BK (eds) *Molecular Systematic*. Sinauer Associates: Sutherland, Mass. pp 205–247.

Prychitko TM, Moore WS (1997). The utility of DNA sequences of an intron from the β-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves:Picidae). *Mol Phylogenet Evol* **8**: 193–204.

Prychitko TM, Moore WS (2000). Comparative evolution of the mitochondrial cytochrome *b* gene and nuclear β-fibrinogen intron 7 in woodpeckers. *Mol Biol Evol* **17**: 1101–1111.

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

Röhl A (2000). *Network: Phylogenetic Network Analysis*, v. 3.1.1.1. Fluxus Technology, Ltd: Germany.

Sage RD, Selander RK (1979). Hybridization between species of the *Rana pipiens* complex in Central Texas. *Evolution* **33**: 1069–1088.

Schilthuizen M, Hoekstra RF, Gittenberger E (2001). The 'rare allele phenomenon' in a ribosomal spacer. *Mol Ecol* **10**: 1341–1345.

Sequeira F, Alexandrino J, Rocha S, Arntzen JW, Ferrand N (2005). Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Mol Ecol* **14**: 245–254.

Sunnucks P, Wilson ACC, Beheregaray LB, Zenger K, French J, Taylor AC (2000). SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol* **9**: 1699–1710.

Weibel AC, Moore WS (2002). A test of a mitochondrial gene-based phylogeny of woodpeckers (Genus Picoides) using an independent nuclear gene, β-fibrinogen intron 7. *Mol Phylogenet Evol* **22**: 247–257.

Woodruff DS (1989). Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of new electromorphic variants called hybrizymes. *Biol J Linn Soc* **36**: 281–294.

Zhang D-X, Hewitt GM (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* **12**: 563–584.