MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1005–1018 doi: 10.1111/j.1365-294X.2011.05422.x

Numts help to reconstruct the demographic history of the ocellated lizard (*Lacerta lepida*) in a secondary contact zone

ANDREIA MIRALDO,*§ GODFREY M. HEWITT,* PAUL H. DEAR,† OCTAVIO S. PAULO‡ and BRENT C. EMERSON*¶

*School of Biological Sciences, University of East Anglia, Norwich NR4 7J, UK, †MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK, ‡Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciencias da Universidade de Lisboa, P-1749-016 Lisboa, Portugal

Abstract

In northwestern Iberia, two largely allopatric *Lacerta lepida* mitochondrial lineages occur, L5 occurring to the south of Douro River and L3 to the north, with a zone of putative secondary contact in the region of the Douro River valley. Cytochrome b sequence chromatograms with polymorphisms at nucleotide sites diagnostic for the two lineages were detected in individuals in the region of the Douro River and further north within the range of L3. We show that these polymorphisms are caused by the presence of four different numts (I–IV) co-occurring with the L3 genome, together with low levels of heteroplasmy. Two of the numts (I and II) are similar to the mitochondrial genome of L5 but are quite divergent from the mitochondrial genome of L3 where they occur. We show that these numts are derived from the mitochondrial genome of L5 and were incorporated in L3 through hybridization at the time of secondary contact between the lineages. The additional incidence of these numts to the north of the putative contact zone is consistent with an earlier postglacial northward range expansion of L5, preceding that of L3. We show that genetic exchange between the lineages responsible for the origin of these numts in L3 after secondary contact occurred prior to, or coincident with, the northward expansion of L3. This study shows that, in the context of phylogeographic analysis, numts can provide evidence for past demographic events and can be useful tools for the reconstruction of complex evolutionary histories.

Keywords: heteroplasmy, numts, refugia, secondary contact zone

Received 20 July 2010; revision received 16 November 2011; accepted 23 November 2011

Introduction

Mitochondrial DNA (mtDNA) is the most widely employed molecular marker in animal phylogenetic, phylogeographic and population genetic studies. This preferential use of mtDNA in evolutionary analysis is primarily because of its higher mutation rate compared to the nuclear genome and typically maternal inheritance that precludes sexual recombination. Under these conditions, mtDNA genes within an individual typically exhibit homoplasmy—sequences are identical copies of each other. However, several processes can result in exceptions to this general rule: the accumulation of somatic mutations within individuals (e.g. Khrapko *et al*. 1997), intramolecular recombination of mitochondrial genomes (e.g. Kajander *et al*. 2000; Lunt & Hyman 1997), the inheritance of gametic variation, and the inheritance of both maternal and paternal mtDNA genomes through paternal leakage (e.g. Kraytsberg *et al*. 2004; Kvist *et al*. 2003). These processes can all

Correspondence: Andreia Miraldo, Fax: +358 9 191 57694; E-mail: andreia.miraldo@helsinki.fi

§Present address: Metapopulation Research Group, University of Helsinki, Department of Biological and Environmental Sciences, P.O. Box 65, 00014 Helsinki, Finland.

¶Present address: Island Ecology and Evolution Research Group, IPNA-CSIC, C/Astrofísico Francisco Sánchez 3, 38206 La Laguna, Tenerife, Canary Islands, Spain.

lead to the detection of polymorphism upon sequencing mitochondrial genes, revealing a state of heteroplasmy. Nevertheless, the detection of a heteroplasmic signal in the absence of true heteroplasmy can also occur through the incorporation of mitochondrial genes, or fragments of genes, into the nuclear genome—(Numts, as first abbreviated by Lopez *et al.* 1994). Numts of recent origin can retain close homology to the original mitochondrial genes and therefore can be co-amplified together with mtDNA during PCR. Distinguishing between these very different processes that lead to a heteroplasmic signal upon sequencing mitochondrial genes is an important step in any study using mtDNA as a marker.

A recent phylogeographic study of the ocellated lizard, *Lacerta lepida*, in the Iberian Peninsula (Miraldo *et al.* 2011) revealed polymorphic sequence chromatograms for the mtDNA cytochrome b gene, both within and adjacent to a zone of secondary contact between two mtDNA lineages, L3 and L5, in the northwestern region of Iberia (Fig. 1). Chromatograms were polymorphic at nucleotide sites diagnostic for the two mtDNA lineages (Fig. 2), suggesting that genetic exchange has occurred among individuals possessing these mtDNA lineages and that this genetic exchange has involved heteroplasmy from the biparental inheritance (BI) of divergent mtDNA genomes and/or numts. The main objectives of this study are to (i) define more precisely the incidence of intra-individual mtDNA sequence polymorphism within the range of L3 and L5 individuals; (ii) determine the genomic origin of the polymorphism; and (iii) explain geographic patterns of mtDNA polymorphism, in the light of the mtDNA phylogeographic history of the two lineages.

In recent years, the processes controlling the active elimination of paternal mitochondria and their mtDNA genomes when inside the egg have been studied in some detail in several mammal species (Shitara *et al.* 1998; Sutovsky *et al.* 1999, 2000, 2004). After fertilization, sperm mitochondria are tagged within the egg by ubiquitin (a proteolytic marker) and are later selectively destroyed. Nevertheless, failure of the successful elimination of all paternal mitochondria in the egg has been shown in hybrid crosses, supporting the idea that the process of active elimination of sperm mitochondria is species specific (Kaneda *et al.* 1995; Sutovsky *et al.* 2000). When the mechanism of elimination breaks down, paternal leakage can occur with the generation of heteroplasmic individuals (biparental inheritance heteroplasmy, Fig. 2). With the exception of some bivalve species where distinct gender-associated mitochondrial DNA lineages coexist, a phenomenon known as doubly uniparental inheritance (see Skibinski *et al.* 1994; Zouros *et al.* 1994 for a detailed explanation of DUI systems),

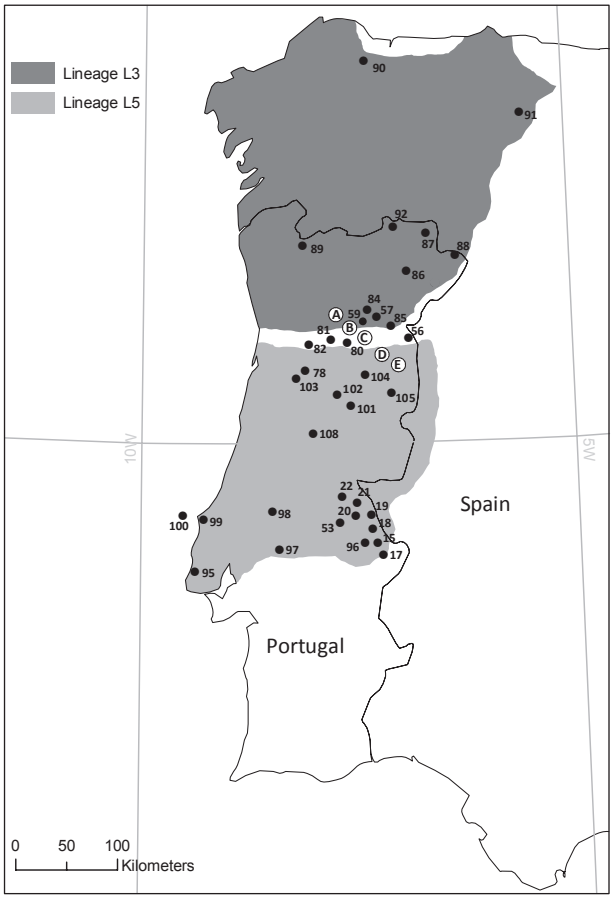

Fig. 1 Map of the western region of the Iberian Peninsula showing the distribution of *Lacerta lepida* mitochondrial lineages L3 and L5 as described in Miraldo *et al.* (2011). Numbers represent broad-scale sampling sites from Miraldo *et al.* study (2011), and letters A–E represent transect sampling sites regarding the fine-scale sampling conducted in this study. Sampling site numbers are the same as in Table 1.

most cases of BI heteroplasmy reported in the literature are incidental and typically associated with hybridization events. Such cases of BI mtDNA heteroplasmy have been reported in birds (Kvist *et al.* 2003), *Drosophila* (Kondo *et al.* 1990; Sherengul *et al.* 2006), mice (Gyllensten *et al.* 1991; Kaneda *et al.* 1995; Shitara *et al.* 1998), cows (Steinborn *et al.* 1998; Sutovsky *et al.* 2000), cicadas (Fontaine *et al.* 2007), mites (Van Leeuwen *et al.* 2008), fish (Ciborowski *et al.* 2007; Hoarau *et al.* 2002; Magoulas & Zouros 1993), bees (Meusel & Moritz 1993), sheep (Zhao *et al.* 2004), humans (Kraytsberg *et al.* 2004; Schwartz & Vissing 2002) and more recently in lizards (Ujvari *et al.* 2008). In the light of these facts, the association of sequence polymorphisms in *L. lepida* with a contact zone between divergent mtDNA lineages renders heteroplasmy as a possible explanation for the detected intra-individual polymorphism.

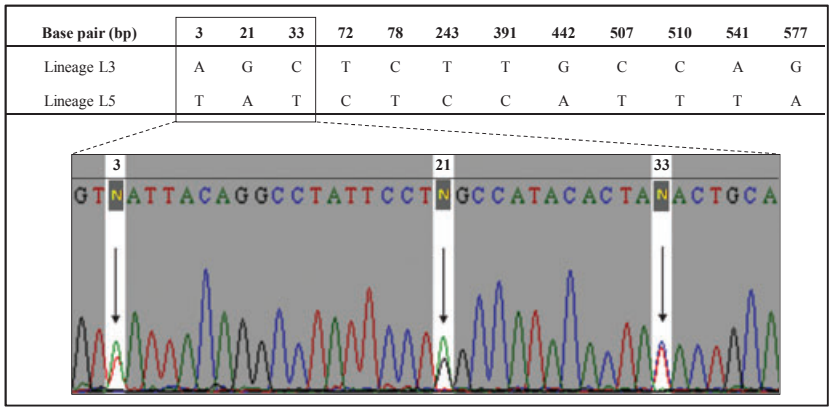

Fig. 2 Polymorphic sites at cytochrome b gene between *Lacerta lepida* mitochondrial lineages L3 and L5 and respective polymorphic sequence trace files.

However, individuals of hybrid ancestry are also a potential source of mtDNA sequence polymorphism arising from numts. Numts are common and have been recorded in many taxa (for reviews on Numts see Bensasson *et al.* 2001; Hazkani-Covo *et al.* 2010; Triant & DeWoody 2007; Zhang & Hewitt 1996). They can arise through single or several independent translocations to the nucleus and vary in number and size, from very small translocations comprising only fragments of mitochondrial genes to the incorporation of almost the entire mitochondrial genome (Hazkani-Covo *et al.* 2010; Leister 2005; Richly & Leister 2004). In vertebrates, it has been shown that the mutation rate of numts slows down relative to the mtDNA gene regions from which numts are derived (Arctander 1995; Collura & Stewart 1995; Fukuda *et al.* 1985; Lu *et al.* 2002; Smith *et al.* 1992; Zischler *et al.* 1995) in line with the slower rate of sequence evolution in this nuclear genome (Brown *et al.* 1982). Thus, the detection of numts that have recently arisen may be difficult, as the divergence between the numt and its functional mtDNA copy could be very small or even absent depending on the size of the fragment under study (Bertheau *et al.* 2011; Fukuda *et al.* 1985; Moulton *et al.* 2010; Zischler *et al.* 1995). Recently, Podnar *et al.* (2007) have demonstrated in *Podarcis* lizards the existence of a numt in *P. sicula* that is genetically more similar to the mtDNA genome of the related species *P. muralis*. Although their data did not allow for definitive conclusions regarding the origin of the numt in *P. sicula*, their study does present an interesting model whereby numts of recent origin may be more readily detectable within individuals of hybrid ancestry (Fig. 3). In essence, a recent numt that is genetically identical to the parental genome will remain undetected by PCR when it co-occurs with the parental genome in an individual. However, hybridization involving individuals with divergent mtDNA genomes can result in offspring where the recent numt and divergent mtDNA genome reside in the same cells and thus are detectable by PCR. Thus, to achieve our objectives, we test for the presence of both BI heteroplasmy and numts.

Materials and methods

Sampling

Lizards were captured under licence during 2006 from five sites (A–E) spanning the area of secondary contact between mitochondrial lineages L3 and L5 (Fig. 1). The five sites were approximately 20 km apart (Fig. 1). Lizards were captured using tomahawk traps or by hand, and tissue samples were taken by clipping 1 cm of the tail tip that was subsequently preserved in 100% ethanol. After tissue sampling, animals were immediately released back into the wild in the place of capture. Geographic coordinates of sampling sites were recorded with a GPS.

DNA extraction, amplification and sequencing

Total genomic DNA was extracted from ethanol-preserved muscle tissue, and a fragment of 627 base pairs (bp) of the mitochondrial DNA (mtDNA) cytochrome b (cytb) gene was amplified using primers CYTBF and CYTBR as described in Miraldo *et al.* (2011).

Identification of polymorphic individuals

All chromatograms of cytb sequences (627 bp) amplified with primers CYTBF and CYTBR were visually checked to assess sequence quality and the presence of double peaks. Sequences were classified as polymorphic if at least one double peak was detected. In order to control for contamination as the source of polymorphism, DNA

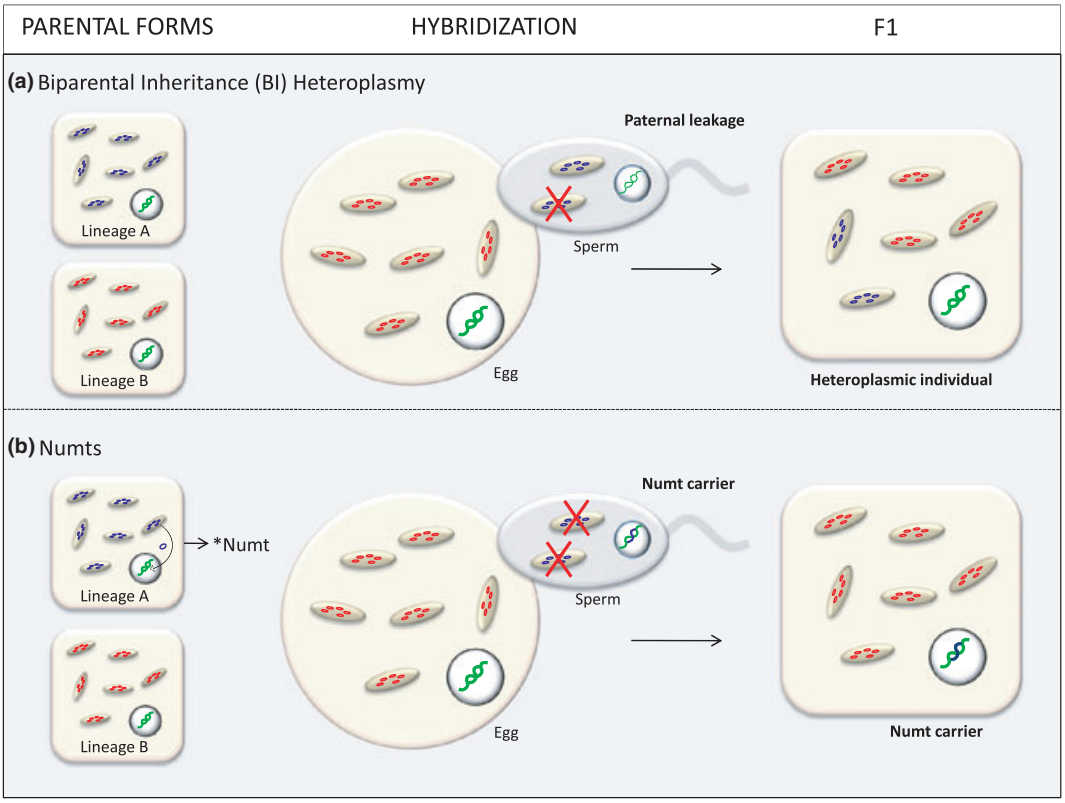

Fig. 3 (a) Biparental inheritance (BI) heteroplasmy—heteroplasmy is generated through paternal leakage at the time of fertilization. Hybridization of individuals from two different mitochondrial lineages results in F1 carrying two types of mitochondrial DNA. (b) Numts—incorporation of mitochondrial DNA in the nuclear genome of lineage A. When numt carrying males from lineage A hybridize with females from lineage B, F1 will be polymorphic for the transferred mitochondrial fragments, as they harbour the complete mitochondrial genome from lineage B (through maternal inheritance of mtDNA) and fragments of mitochondrial DNA from lineage A as numts. Somatic cells are represented as squares. Within each cell, mitochondria are shown as ellipses and mitochondrial DNA is represented as coloured ellipses inside mitochondria. Blue mtDNA represents mitochondrial lineage A, while red mtDNA represents mitochondrial lineage B. Nuclei are represented as white circles within each cell with nuclear genomes shown as green helices.

from eight polymorphic individuals and two individuals with no signs of polymorphism was re-extracted and the cytb fragment was amplified and sequenced using the same conditions as described before. Negative controls were used in every step of the experiment.

Quantification and characterization of intra-individual variation

Eighteen polymorphic individuals from the contact zone were used to quantify intra-individual variation. For those individuals, the cytb fragment was amplified as described above and the fragments were cloned using the StrataClone™ PCR Cloning Kit (Stratagene), following the manufacturer's recommendations. Three individuals with no signs of polymorphism were also cloned for control purposes. For each cloned individual, 6–10 positive clones were purified using the QIAprep® Spin Miniprep Kit (Qiagen) and directly sequenced in both directions using the conditions described above.

PCR cloning procedures are known to have some inherent and significant disadvantages. The disadvantages are mainly associated with the amplification step where PCR-derived mutations, template jumping and allelic preference are known to occur (Lin *et al.* 2002; Pääbo *et al.* 1990). These disadvantages become especially problematic when PCR cloning procedures are used to describe mutations that distinguish different gene copies. While PCR-induced errors (*in vitro* errors) will not be detectable upon sequencing, as at most they can only affect 25% of all molecules synthesized, upon cloning *in vitro* polymerase errors will become indistinguishable from *in vivo* mutations because each of the errors will affect all the molecules (100%) of a clone, just as a genuine *in vivo* mutation does. Thus, for mutations detected upon sequencing clones, we have to account for mutations that are caused by *Taq* errors. *Taq* error rates have been estimated to vary between 7.2×10^{-5} (Ling *et al.* 1991) and 2.0×10^{-4} (Saiki *et al.* 1988) errors per nucleotide per cycle, depending on PCR conditions.

From these error rates, we could expect to have up to 4.4 errors (or 1.6 using the lower error rate) in each amplified fragment, taking into account the size of the fragment (627 bp) and the number of cycles used for the PCR amplification (35 cycles). As a conservative approach, and taking into account the above-mentioned *Taq* error rates, all mutations present in sequences that differ by four mutations or fewer from any sequences that occur in more than one individual are considered as being potential *Taq* errors and are not considered for further analysis. Pairwise comparisons of uncorrected sequence divergences (p-distance) within each individual were estimated using PAUP* version 4.0 b10 (Swofford 2002).

Polymorphism origin: heteroplasmy vs. numts

Increasing amplicon size. To aid distinction between BI heteroplasmy and numts, a strategy of increasing amplicon size was adopted. In the case of BI heteroplasmy, a signature of heteroplasmy in sequence chromatograms irrespective of amplicon size should be expected. Nevertheless, unless the complete mtDNA genome has been incorporated as a numt, there should be an upper limit beyond which a numt will not be amplified, and therefore no polymorphism should be detected upon sequencing (Fig. 4). For all polymorphic individuals, the entire cytb gene was amplified (1143 bp), using modified versions of primers L14919 (TRNAGLU, 5′-AAC CAC CGT TGT ATT TCA ACT-3′) and L16064 (TRNATHR, 5′-CTT TGG TTT ACA AGA ACA ATG CTT TA-3′) (Burbrink *et al.* 2000) to increase specificity for *Lacerta lepida*, using information from published sequences in GenBank. The amplification of the complete cytb gene represents an increase of 516 bp from the amplified fragment that revealed the existence of polymorphisms. PCR conditions were the same as described for the 627-bp cytb fragment, and amplifications were conducted as follows: DNA was initially denatured at 94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 30 s and extension at 72 °C for 90 s, plus a final extension step at 72 °C for 3 min. Negative controls (no DNA) were included for all amplifications. PCR products were visualized on a 2% agarose gel and purified by filtration through QIAquick® columns (Qiagen) following the manufacturer's recommendations. Purified PCR products were sequenced with internal primers specifically designed for this study: CBF (5′-AAC CTC CTC TCA GCA ATA CC-3′) and CBR (5′-CCT GTG GGG TTG TTT GAA-3′).

Single-molecule PCR. Single-molecular PCR (smPCR) is essentially a normal PCR but where the template DNA is diluted to very low concentration. Performing a smPCR by limiting DNA dilution to one amplifiable mitochondrial genome should only result in the amplification of mitochondrial fragments. Therefore, by using this approach, it is possible to distinguish between heteroplasmy and numts as the source of mtDNA polymorphism. Four polymorphic individuals, previously subjected to cloning, were used for the smPCR. These samples were chosen based on the extent of polymorphism revealed from the cloning procedures, with the objective of using samples where several types of sequences were detected. Detailed explanation of the smPCR protocol is provided as supporting information (Appendix S1, Supporting information).

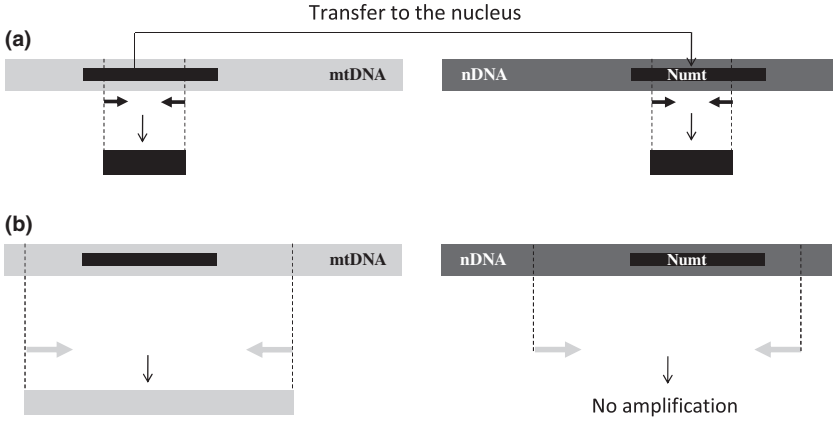

Fig. 4 Amplification of mtDNA and numts using different sets of primers. The numt is represented by a black bar within the nuclear genome. In (a), the primers used for PCR amplification are complementary to the mtDNA region that was transferred to the nuclear genome originating the numt, resulting in the amplification of both mtDNA and nDNA fragments. In (b), by increasing the amplicon size to be amplified designing primers that anneal outside the fragment that was transferred to the nuclear genome, the amplification will be specific for mtDNA and therefore the numt will not be amplified during the PCR.

Phylogeographic analysis

Intraspecific gene genealogies were inferred using two different network construction approaches: median-joining (MJ) (Bandelt *et al.* 1999) and statistical parsimony (SP) (Templeton *et al.* 1992). The MJ network was computed with the program NETWORK 4.5.0 (www.fluxus-engineering.com), and the SP network was inferred using the program TCS 1.21 (Clement *et al.* 2000). For detailed information about the parameter settings used, see Miraldo *et al.* (2011). For the polymorphic individuals, all sequences detected by cloning were added to the mitochondrial data set to determine the number of mutational steps separating these sequences from any given mitochondrial haplotype of the lineages under study.

Results

A total of 85 individuals were collected from five populations spanning the putative contact zone between the two mitochondrial lineages (Fig. 1, Table 1) and PCR-amplified and sequenced using primers CYTBF and CYTBR. Of these 85 individuals, 33 generated polymorphic sequences involving two from population A, 13 from population B and 18 from population C. In all cases, polymorphic sites included nucleotide positions and nucleotide states that are diagnostic between lineages L3 and L5 (Fig. 2).

Quantification and characterization of intra-individual variation

One hundred and 63 clones were sequenced from 21 individuals (one from population A, 10 from population B and 10 from population C). All clones sequenced in homoplasmic individuals (10 clones from individual B3, seven clones from B5 and nine clones from C1) represent the same haplotype (haplotype 46 in B3, haplotype 59 in B5 and haplotype 25 in C1) or represent sequences that differ from those by 1–4 unique point mutations (Table 2).

Among the 137 clones sequenced from the 18 polymorphic individuals, nine represent sequences of mitochondrial haplotypes sampled in Miraldo *et al.* study (2011) (two of haplotype 54 in individual A7, four of haplotype 59 in B18, two of haplotype 40 in C8 and one of haplotype 152 in C9) and seven represent sequences that differ from those haplotypes by four or fewer mutations. Among the remaining 121 clones, four sequences (25 clones) were found to occur in more than one individual (sequences I to IV): 14 clones represent sequence I and occur in seven individuals; four clones represent sequence II and occur in four individuals; two clones represent sequence III and occur in two individuals; and five clones represent sequence IV and occur in five individuals. The remaining 96 sequenced clones are either very similar to one of the previous sequences (I to IV), differing from those by four or fewer mutations (46 clones), or are consistent with recombinants (50 clones) between the different types of sequences present within each individual (Table 2). As *in vitro* recombination cannot be excluded as the origin of the recombinant-like sequences, and all but one of the recombinant sequences (sequence V, occurring in three clones from three different individuals) are found in single individuals, these were eliminated from further analysis. Sequence V was also removed for further analysis as it was easily generated *in vitro* by performing a PCR of mixed template from individuals that are homoplasmic for lineages 3 and 5 (data not shown). Thus, the sequences retained for further analysis are those that represent mitochondrial haplotypes that have been successfully sequenced in other studies and those whose origin cannot be attributable to PCR and cloning artefacts, and thus represent authentic sequences.

Polymorphism origin: heteroplasmy vs. numts

For the samples polymorphic for the 627-bp cytb gene fragment (33 from within the contact zone and identified in this study and seven from outside the contact zone and identified in Miraldo *et al.* (2011)), homoplasmic sequences were obtained for 38 samples with the amplification and sequencing of the 1143-bp fragment that represents the entire cytb gene. The complete cytb sequences obtained for all individuals represent haplotypes previously sampled (Miraldo *et al.* 2011). All haplotypes are from lineage L3 apart from haplotype 59, which belongs to lineage L1 described in Miraldo *et al.* (2011). This suggests that sequences I, II, III and IV are not mitochondrial sequences and are most likely of nuclear origin. Genetic distances (uncorrected *P*-values) between all sequences present in each polymorphic sample and the corresponding mtDNA sequence obtained by the amplification of the entire cytb gene range from 1.8% (Sequence III) to 17.0% (Sequence IV) (Table 2). Sequence I is the most common, occurring in 13 of 18 cloned individuals, and sequence II is the second most represented, occurring in seven individuals. Sequence III and IV occur in two and five individuals, respectively (Table 2). With the exception of sequence IV, all sequences detected in this study have open reading frames.

SmPCR was successfully performed for the four individuals analysed, and 55 sequences were obtained. All smPCR sequences from individuals C3 and C4 represent haplotype 40 and correspond to the sequence

Table 1 Sampled sites with name (Location), site number (Site) and country of origin (Country). For each site, the total number of samples collected (*n*), the respective sample code (Sample labels), the cytochrome b haplotypes detected (mtDNA haplotype) and their absolute frequency (frequency) are presented. Polymorphic individuals are denoted by bold underline font. Sampled sites numbers are the same as in Miraldo *et al.* (2011)

Site	Location	Country	*n*	Sample code	mtDNA haplotype (frequency)
15	Arronches	Portugal	1	BS83	25 (1)
17	Elvas	Portugal	1	BS82	16 (1)
18	Serra de Sao Mamede (South)	Portugal	3	BS84, BS85, BS86	25 (3)
19	Serra Sao Mamede (North)	Portugal	2	BS64, BS65	25 (2)
20	Crato	Portugal	1	BS87	25 (1)
21	Nisa (East)	Portugal	2	BS88, BS89	25 (2)
22	Nisa (West)	Portugal	3	BS61, BS62, BS63	17 (1), 25 (1), 109 (1)
53	Alter do Chao	Portugal	2	BS90, BS91	25 (1), 90 (1)
56	Fig. Castelo Rodrigo	Portugal	4	BS36, BS37, BS38, BS39	27 (2), 56 (1), 57 (1)
57	Foz Coa	Portugal	1	BS33	58 (1)
59	Penedono	Portugal	1	**BS34**	40 (1)
78	Serra do Caramulo (North)	Portugal	2	BS48, BS49	46 (1), 32 (1)
80	Satao	Portugal	1	**BS40**	148 (1)
81	Serra Liomil	Portugal	1	BS41	40 (1)
82	Castro Daire	Portugal	4	BS42, BS43, BS44, BS45	17 (1), 21 (1), 46 (1), 48 (1)
84	Foz Coa (North)	Portugal	1	**BS32**	49 (1)
85	Foz Coa (South)	Portugal	1	BS35	47 (1)
86	Macedo de Cavaleiros	Portugal	2	**BS30, BS31**	46 (1), 147 (1)
87	Parque Natural de Montesinho (East)	Portugal	3	BS26, BS27, BS28	42 (1), 52 (1), 53 (1)
88	Miranda do Douro	Portugal	1	**BS29**	46 (1)
89	Geres	Portugal	10	BS13, **BS14**, BS15, BS16, BS17, BS18, BS19, BS20, BS21, BS22	41 (1), 44 (2), 45 (1), 46 (2), 50 (1), 51 (1), 55 (1), 146 (1)
90	Galicia	Spain	11	BS1, BS2, BS3, BS4, BS5, BS6, BS7, BS8, BS9, BS10, BS11	41 (9), 43 (2)
91	Asturias	Spain	1	BS12	42
92	Parque Natural de Montesinho (West)	Portugal	3	BS23, BS24, BS25	41 (1), 46 (1), 48 (1)
95	Samarra	Portugal	3	BS69, BS70, BS71	14 (1), 17 (1), 20 (1)
96	Monforte	Portugal	1	BS66	13 (1)
97	Coruche	Portugal	1	BS68	15 (1)
98	Paul Boquilobo	Portugal	1	BS67	18 (1)
99	Peniche	Portugal	8	BS74, BS75, BS76, BS77, BS78, BS79, BS80, BS81	17 (3), 33 (3), 88 (1), 89 (1)
100	Berlengas	Portugal	2	BS72, BS73	30 (2)
101	Serra da Estrela (South)	Portugal	7	BS52, BS53, BS54, BS55, BS56, BS57, BS58	23 (1), 26 (2), 29 (2), 30 (2)
102	Serra da Estrela (North)	Portugal	1	BS59	17 (1)
103	Serra do Caramulo (South)	Portugal	2	BS46, BS47	19 (1), 32 (1)
104	Celorico da Beira	Portugal	1	BS50	35 (1)
105	Sabugal	Portugal	1	BS60	24 (1)
108	Lousa	Portugal	1	BS51	22 (1)
A	Peso da Regua	Portugal	7	A1, A2, A3, A4, A5, **A6, A7**	46 (4), 54 (2), 149 (1)
B	Nagosa	Portugal	19	B1 to B6, **B7 to B19**	46 (17), 59 (2)
C	Penedono (South)	Portugal	20	C1, C2, **C3 to C20**	1 (1), 25 (1), 40 (5), 46 (4), 51 (2), 150 (2), 151 (3), 152 (1), 153 (1)
D	Vila Franca das Naves	Portugal	14	D1 to D14	25 (4), 28 (3), 31 (5), 36 (1), 39 (1)
E	Lamegal	Portugal	25	E1 to E25	25 (7), 27 (1), 31 (10), 34 (1), 37 (1), 38 (1), 39 (1), 56 (3)

obtained by the amplification of the complete cytb gene. Although the majority of smPCR sequences in individuals C8 and C9 correspond to the sequence also obtained by the amplification of the complete cytb gene (haplotype 40 in C8 and 152 in C9), one sequence in each individual corresponds to a different haplotype, revealing

Table 2 Total number of clones sequenced per sample and identification of each type of sequence obtained

		mtDNA clones						Sequence I						Sequence II						Sequence III						Sequence IV								
Sample	Haplotype	0	1	2	3	4	Total	0	1	2	3	4	Total	0	1	2	3	4	Total	0	1	2	3	4	Total	0	1	2	3	4	Total	Rec.	Total no. clones	Divergence (%)
A7	54	2	1		1		**4**																			1					**1**	2	7	2.1–17.0
B7	46							1	3		1		**5**			1			**1**													2	8	2.2–2.4
B8	46							6					**6**																			3	9	2.2
B10	46								3	1			**4**																			3	7	2.2
B11	46								2	2			**4**																			3	7	2.2
B12	46								1	2		1	**4**													1					**1**	3	8	2.2–17.0
B13	46									2		1	**3**		1				**1**													3	7	2.2–2.4
B18	59	4	1				**5**																									5	10	n.a.
B19	46							3			1	1	**5**			1	1		**2**													1	8	2.2–2.4
C3	40								1				**1**	1					**1**							1					**1**	4	7	2.6–16.6
C4	40							1	2	2			**5**	1					**1**													2	8	2.6–2.7
C5	46							1		2	1		**4**																			2	6	2.2
C6	*							1	3	1			**5**	1					**1**													2	8	n.a.
C7	46							1	3		1		**5**	1		1			**2**													1	8	2.2–2.4
C8	40	2					**2**			1			**1**																			5	8	2.6
C9	152	1	1				**2**																			1					**1**	4	7	2.6–17.0
C10	46		3				**3**													1					**1**	1					**1**	3	8	1.8–17.0
C16	*																			1	2		1		**4**							2	6	n.a.
B3	**46**	2	4	2	2		**10**																										10	n.a
B5	**59**	3	2	1	1		**7**																										7	n.a
C1	**25**	4		3	1	1	**9**																										9	n.a

Sample codes are the same as in Table 1. Bold underlined samples represent control homoplasmic samples. For each sample, the mtDNA haplotype identified using the amplification of the entire cytb gene is shown (Haplotype). The total number of clones with 0, 1, 2, 3 and 4 mutations from the sequences representing the mtDNA, Sequence I, II, III and IV are indicated. The total number of cloned sequences representing recombinants (Rec.) and the genetic distance (pairwise uncorrected) between each sequence and the mtDNA detected in each sample are also represented [Divergence (%)]. Homoplasmic samples cloned are represented by bold underline font.
n.a., not applicable.
*Ran out of DNA.

the heteroplasmic state of these individuals. Individual C8 carries both haplotype 40 and 1 while individual C9 carries haplotype 152 and the new haplotype 153. SmPCR failed to recover sequences I to IV, supporting a nuclear origin for these sequences, compatible with the long amplicon results.

Geography and phylogenetic placement of numts

The geographic range of each sequence of inferred nuclear origin was assessed by examining sequence chromatograms for all seven polymorphic individuals found outside the contact zone (individuals BS14, BS29, BS30, BS31, BS32, BS34 and BS40). Nuclear sequences were inferred by subtracting the authentic mtDNA sequence, obtained by the amplification of the large amplicon, from the polymorphic sequence chromatogram for each individual. Diagnostic sites between sequence I–IV and mtDNA sequences from lineage L3 are shown in Table 3. The minimum number of mutations separating sequence IV from lineage L3 is 107; thus, its presence is easily recognizable from the polymorphic trace files. When sequence III is present together with sequence IV, the presence of sequence I cannot be detected with certainty as there are no independent diagnostic sites that distinguish it. Sequence II has a unique diagnostic mutation at site 385 (Table 3), and therefore detecting its presence is possible even in the presence of all other sequences. By adopting this approach, it was determined that sequence IV is present in all seven individuals and sequence III is present in five (BS29, BS30, BS31, BS32 and BS40). Sequence I is present in at least two individuals (BS14 and BS34), but it was not possible to determine its presence in the remaining individuals owing to the presence of both sequence III and IV.

Networks were constructed from 180 sequences: 83 from the contact zone, 91 from throughout the distribution of L3 and L5 characterized by Miraldo *et al.* (2011), two sequences representing the new mitochondrial haplotypes detected by smPCR for individuals C8 and C9 and four sequences representing the putative numts. Both methods used for network estimation (MS and SP) resulted in a single network, with the same topology yielding two loops (Fig. 5) that were easily resolved and 64 different haplotypes (GenBank accession numbers JN860936–JN860996). Sequences I and II are both derived from mtDNA lineage L5: sequence I corresponds to the sampled haplotype 37 and sequence II differs from I by a single mutation. Sequence III is connected to an unsampled or extinct haplotype near the root of the network, and sequence IV is the most divergent sequence and could not be connected to the network under a 95% parsimony connection limit.

Discussion

The distribution of *Lacerta lepida* is structured into two divergent mtDNA lineages in its northwestern range (lineage L3 and L5) that form a zone of secondary contact (Miraldo *et al.* 2011). Some individuals from lineage L3 exhibit several different copies of the cytochrome b gene, with a high incidence of this at the centre of the contact zone. Homoplasmic sequences obtained with the amplification of the entire cytb gene are concluded to represent sequences from the mitochondrial genome, suggesting that the polymorphic signal is originated by numts. Inter-genomic transfer of mtDNA fragments into the nuclear genome is a widespread phenomenon reported for a great number of taxa (Hazkani-Covo *et al.* 2010; Zhang & Hewitt 1996) and is the most plausible explanation for the polymorphisms associated with lineage L3. The phylogenetic relationships between the numts and the mtDNA sequences provide some information about where and when some of these translocations occurred. Numts I and II are both derived from mitochondrial lineage L5 and most probably originated within the geographic range of L5 with subsequent incorporation into the nuclear genome associated with L3 through hybridization and backcrossing when the ranges of these two mtDNA lineages came into secondary contact. Numt II differs from numt I by only a single mutation. The simplest explanation is that, rather than two independent translocations, numt II is an allele derived from numt I and therefore they represent the same numt. Numt III, on the other hand, is phylo-

Table 3 Diagnostic sites between Sequence I, II, III, IV and L3 mitochondrial lineage. Dots represent nucleotide matches to lineage L3. Light grey-shaded base pairs represent the diagnostic sites for Sequence II and III (see text for detailed explanation)

bp	3	21	33	72	78	171	243	264	348	385	391	442	443	507	510	541	567	577
L3	A	G	C	T	C	C	T	C	C	G	T	G	C	C	C	A	G	G
Sequence I	T	A	T	C	T	.	C	.	T	.	C	A	.	T	T	T	A	A
Sequence II	T	A	T	C	T	.	C	.	T	A	C	A	.	T	T	T	A	A
Sequence III	T	A	T	C	T	T	.	T	.	.	C	.	T	.	T	T	.	.
Sequence IV	T	.	.	C	.	T	C	T	T	.	C	A	.	T	.	C	A	A

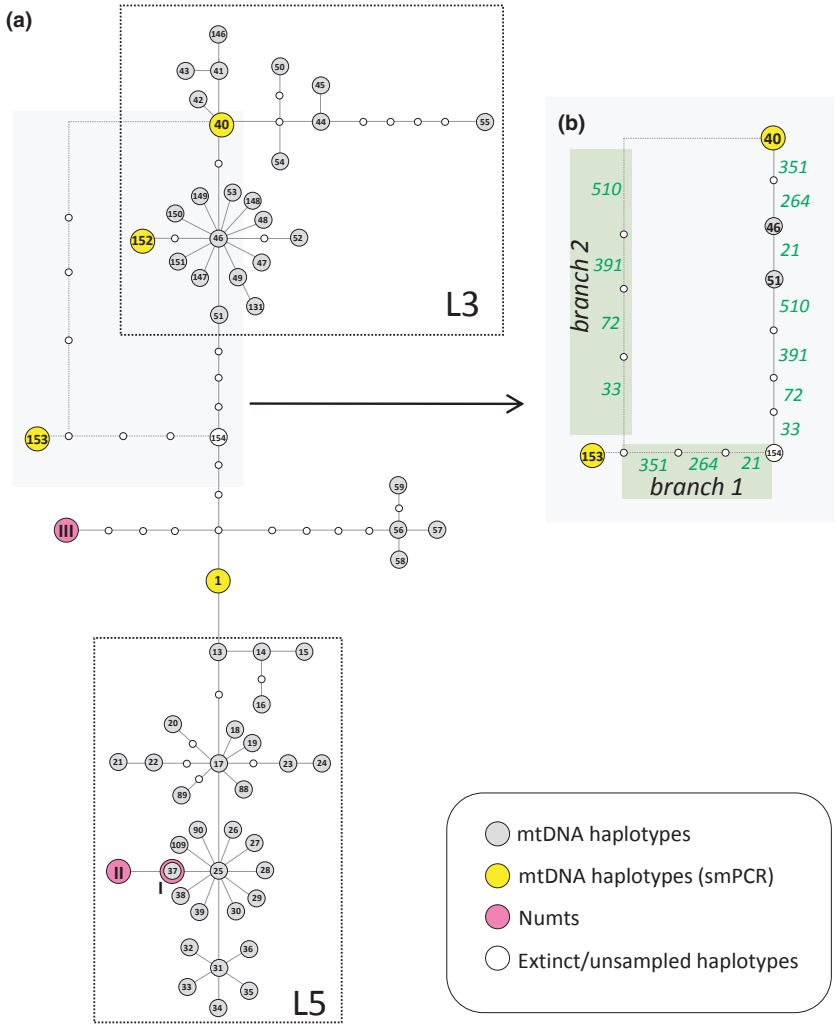

Fig. 5 (a) Statistical parsimony network of *Lacerta lepida* cytochrome b haplotypes. White circles represent unsampled or extinct mtDNA haplotypes; yellow circles represent mtDNA haplotypes detected through single-molecule PCR, and pink circles represent numts. Dashed lines represent ambiguities in the network. (b) Close-up of the loop connecting haplotypes 153, 40 and 154. Numbers in green refer to the nucleotide position of mutations.

genetically closer to the root of the network (Fig. 5) and thus is inferred to be the result of an older independent translocation. The very high divergence between numt IV and all mtDNA sequences indicates this translocation to be a much older event, pre-dating the divergence events that gave rise to mtDNA genetic diversity sampled within this study. Thus, the identified numts are descended from at least three independent transfers to the nucleus.

Phylogeographic utility of numts

Once identified, numts can be used as important tools in evolutionary biology, providing a unique window on past evolutionary events (reviewed in Bensasson *et al.* 2001; Zhang & Hewitt 1996), facilitating the estimation of phylogenetic relationships (Hazkani-Covo 2009), dating of divergence events (Schmitz *et al.* 2005) and allowing the quantification of past gene exchange between divergent lineages (Baldo *et al.* 2011). In this study, the geographic distribution of numt I within lineage L3 provides valuable information regarding the demographic history of the two mtDNA phylogeographic lineages under study, complementing information obtained through the analysis of mtDNA sequence data. Within lineage L5, mtDNA haplotype 13 is the most recent common ancestor (MRCA) and together with the closely related descendant haplotypes 14, 15 and 17 occur mainly in the southern limit of the lineage distribution, near the Tagus River valley. In contrast, the most derived haplotypes within lineage L5 (31, 32, 33, 34, 35 and 36) are nearly all (20 of 23) distributed in

the northern limit of the lineage distribution, just south of the Douro River. The same pattern is observed within lineage L3, with haplotype 51 (MRCA) and its immediate descendant haplotype 46 being more frequent in the southern limit of the lineage geographic range, south of Douro River. Eighty-three per cent of the individuals that possess either haplotype 46 or haplotype 51 are from southern sampling sites. In contrast, the derived haplotype 40 and its descendants (41, 42, 43, 44, 45, 50, 54, 55 and 146), all of which are derived from haplotype 46, are nearly all represented in more northerly sampling sites. This replicated pattern in both lineages of ancestral haplotypes being more frequent at southern latitudes and descendant haplotypes at northern ones is consistent with the hypothesis of range expansions from southern refugia as has been reported before (Miraldo *et al.* 2011) with the region around the Tagus River most probably functioning as a refugial area for lineage L5, and the area around the Douro river valley being a likely refugial area for L3. From these refugial areas, both lineages are inferred to have expanded their ranges when climatic conditions permitted, meeting at what is now a zone of secondary contact. Owing to earlier climate amelioration at southern latitudes, the northward range expansion of lineage L5 is expected to have been initiated earlier than the

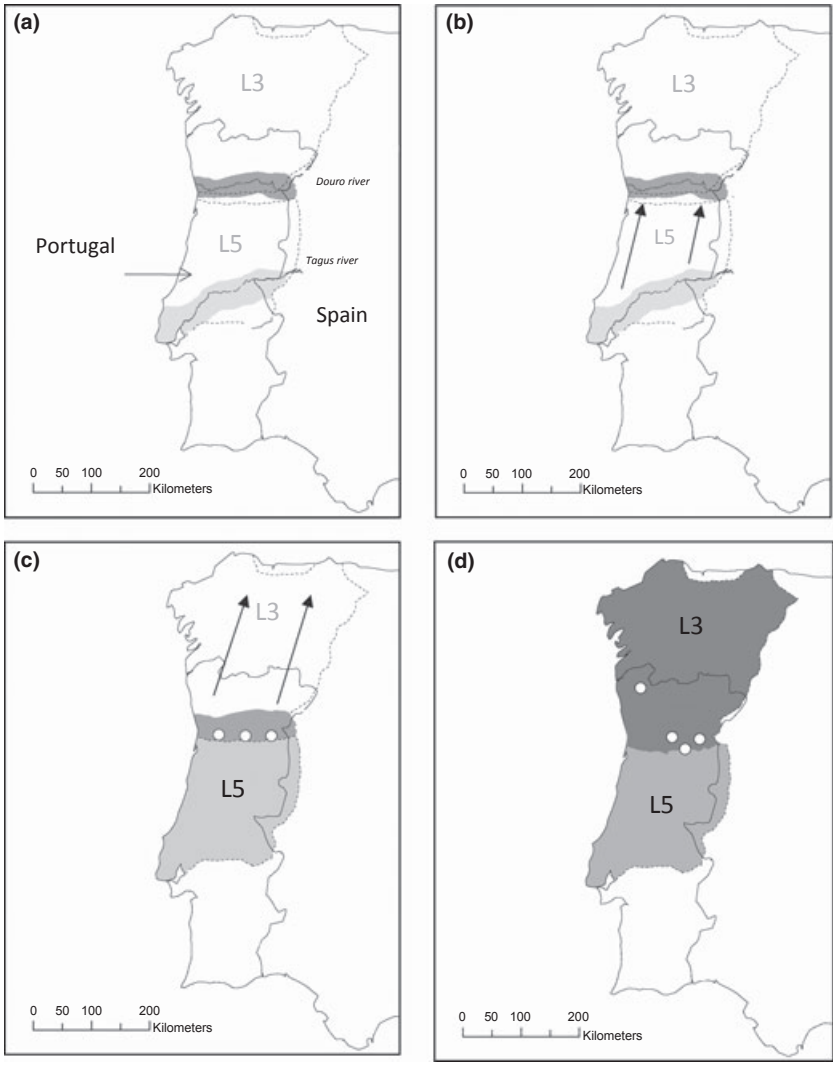

Fig. 6 Historical biogeographic scenario for *Lacerta lepida* mitochondrial lineages L3 and L5. (a) Separation of the lineages in two different Plio-Pleistocene glacial refugia as described in Miraldo *et al.* (2011); (b) earlier climatic mediated northwards range expansion of lineage L5 with lineage L3 remaining in its refugium; (c) establishment of secondary contact between lineages L3 and L5 with the consequent incorporation of mitochondrial DNA fragments from lineage L5 into the nuclear genome of L3 through hybridization; (d) range expansion of lineage L3 carrying numt I and II northwards. Dashed lines in (a), (b) and (c) represent the present distribution of *Lacerta lepida* mitochondrial lineages L5 and L3 as described in Miraldo *et al.* (2011); lineage L3 is represented in dark grey and lineage L5 in light grey. White circles in (d) represent the present distribution of numts I and II within lineage L3.

expansion of lineage L3. This is supported by the geographic distribution of the L5-derived numt I that provides information about the relative timing of the establishment of the secondary contact zone between both lineages, relative to the demographic and range expansions of each (Fig. 6). Numt I is derived from haplotype 37 that is a recently derived haplotype from lineage L5. Within the range of L5, haplotype 37 was only detected at the northern limit of the lineage distribution, in population E, adjacent to the contact zone of L3 and L5. The fact that numt I is found in L3 as far north as Geres, which is approximately 150 km north of the contact zone, suggests that admixture between both lineages was established before, or coincident with, the expansion of lineage L3 northwards, leading to the appearance of numt I in L3. The zone of secondary contact between lineages L3 and L5 is therefore consistent with the leading edge of lineage L5 expanding north and contacting lineage L3 in the vicinity of the L3 refugia, prior to or coincident with the northward range expansion of L3.

Heteroplasmy in Lacerta lepida

The co-occurrence of mitochondrial DNA haplotypes 1 and 40 in individual C8 and haplotypes 153 and 152 in C9 confirms the existence of low levels of heteroplasmy in these individuals. In animals, heteroplasmy can be achieved through the accumulation of somatic mutations (e.g. Khrapko *et al.* 1997), paternal leakage (Fontaine *et al.* 2007), the inheritance of gametic variation, or through intramolecular recombination (e.g. Kajander *et al.* 2000; Lunt & Hyman 1997). In the case of *Lacerta lepida*, the heteroplasmy documented in individuals C8 and C9 is most consistent with paternal leakage. In both individuals, the differences between the heteroplasmic haplotypes are too large to be explained by the accumulation of somatic mutations within an individual (11 mutational steps between haplotypes present in C8 and nine in C9). Additionally, it would require unprecedented levels of heteroplasmy in the case of haplotype 1 in individual C8. An increasing number of species have been shown to harbour some level of heteroplasmy associated with BI of mtDNA, and *Lacerta lepida* seems to be no exception.

The probable origin of haplotype 153 deserves further discussion. Haplotype 153 is connected in the network to haplotype 40 by four mutations (Fig. 5b, branch 2) and to the ancestral unsampled haplotype 154 by three mutations (Fig. 5b, branch 1). The mutations involved in branch 1 and 2 are the same seven mutations that occur from the ancestral unsampled haplotype to haplotype 40, resulting in the loop, suggesting haplotype 153 to be the product of a recombination event between haplotypes 40 and the ancestral unsampled haplotype. While homoplasy provides an alternative explanation, this would require a minimum of three homoplasious mutations, which is highly improbable. The origin of the haplotype 153 most likely derived from a recombination event following paternal leakage. Although not commonly reported, recombination following paternal leakage has been inferred as the most likely explanation for mtDNA recombinants detected in a contact zone between the conifers black spruce (*Picea mariana*) and red spruce (*Picea rubens*) (Jaramillo-Correa & Bousquet 2005), and in a contact zone between two mitochondrial lineages of the Australian frillneck lizard (*Chlamydosaurus kingii*) (Ujvari *et al.* 2007).

Conclusion

This study provides a detailed understanding of the recent demographic history of *Lacerta lepida* in the northwestern part of the Iberian Peninsula, where two divergent mitochondrial lineages, L3 and L5, occur. The fine-scale sampling of the contact zone allowed the quantification and characterization of the polymorphism detected in lineage L3. Our methodological approach has shown that the polymorphism in L3 is caused by the presence of at least four different numts together with low levels of heteroplasmy. The detailed phylogeographic analysis of L5-type numts within lineage L3 provided valuable information regarding the demographic history of the two mtDNA lineages under study, complementing information obtained through the analysis of mtDNA sequence data. In addition to the many evolutionary features of numts (see Bensasson *et al.* 2001), this study has shown that, in the context of phylogeographic analysis, numts can provide evidence for past demographic events, which is an exciting prospect for the field of phylogeography. We feel that in the era of second-generation parallel sequencing, where investigators are now seeking to exploit protein-coding mtDNA genes (e.g. Hajibabaei *et al.* 2011; Shokralla *et al.* 2011), numts and their utility for evolutionary analysis will receive renewed interest.

Acknowledgements

This work was supported by a grant from FCT, Portugal (POCI/BIA-BDE/59288/2004). A. M. was supported by a PhD grant from FCT (SFRH/BD/16996/2004). We thank Rui Osório, Pedro Figueiredo, Nuno Valente and Matthew Osborne for helping with fieldwork. We also thank Mário Miraldo and António Firmino for assembling the lizard traps. All samples were collected with appropriate licences from the Instituto de Conservação da Natureza in Portugal. We thank six anonymous reviewers for constructive criticism on earlier versions of this manuscript.

References

Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society of London B*, **262**, 13–19.

Baldo L, de Queiroz A, Hedin M, Hayashi CY, Gatesy J (2011) Nuclear–Mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping Bristletail *Mesomachilis*. *Molecular Biology and Evolution*, **28**, 195–210.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314–321.

Bertheau C, Schuler H, Krumböck S, Arthofer W, Stauffer C (2011) Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. *Molecular Ecology Resources*, **11**, 1056–1059.

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution*, **18**, 225–239.

Burbrink FT, Lawson R, Slowinski JB (2000) Mitochondrial DNA phylogeography of the polytypic north american rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution*, **54**, 2107–2118.

Ciborowski KL, Consuegra S, García de Leániz C *et al.* (2007) Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA. *Biology Letters*, **3**, 554–557.

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

Collura RV, Stewart C (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, **378**, 485–489.

Fontaine KM, Cooley JR, Simon C (2007) Evidence for paternal leakage in hybrid periodical cicadas (Hemiptera: Magicicada spp.). *PLoS ONE*, **2**, e892.

Fukuda M, Wakasugi S, Tsuzuki T (1985) Mitochondrial DNA-like sequences in the human nuclear genome. Characterization and implications in the evolution of mitochondrial DNA. *Journal of Molecular Biology*, **186**, 257–266.

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature*, **352**, 255–257.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, **6**, e17497.

Hazkani-Covo E (2009) Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. *Molecular Biology and Evolution*, **26**, 2175–2179.

Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular Poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. *PLoS Genetics*, **6**, e1000834.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, **19**, 2261–2264.

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics*, **171**, 1951–1962.

Kajander OA, Rovio AT, Majamaa K *et al.* (2000) Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. *Human Molecular Genetics*, **9**, 2821–2835.

Kaneda H, Hayashi J, Takahama S *et al.* (1995) Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proceedings of the National Academy of Sciences, USA*, **92**, 4542–4546.

Khrapko K, Coller HA, André PC *et al.* (1997) Mitochondrial mutational spectra in human cells and tissues. *Proceedings of the National Academy of Sciences, USA*, **94**, 13798–13803.

Kondo R, Satta Y, Matsuura ET *et al.* (1990) Incomplete maternal transmission of mitochondrial-DNA in *Drosophila*. *Genetics*, **126**, 657–663.

Kraytsberg Y, Schwartz M, Brown TA *et al.* (2004) Recombination of human mitochondrial DNA. *Science*, **304**, 981–981.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, **20**, 243–247.

Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends in Genetics*, **21**, 655–663.

Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. *Human Molecular Genetics*, **11**, 133–145.

Ling LL, Keohavong P, Dias C, Thilly WG (1991) Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. *Genome Research*, **1**, 63–69.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Lu X-M, Fu Y-X, Zhang Y-P (2002) Evolution of mitochondrial cytochrome b pseudogene in genus *Nycticebus*. *Molecular Biology and Evolution*, **19**, 2337–2341.

Lunt DH, Hyman BC (1997) Animal mitochondrial DNA recombination. *Nature Genetics*, **387**, 247.

Magoulas A, Zouros E (1993) Restriction-site heteroplasmy in Anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. *Molecular Biology and Evolution*, **10**, 319–325.

Meusel MS, Moritz RFA (1993) Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. *Current Genetics*, **24**, 539–543.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Moulton MJ, Song H, Whiting MF (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, **10**, 615–627.

Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry*, **265**, 4718–4721.

Podnar M, Haring E, Pinsker W, Mayer W (2007) Unusual origin of a nuclear pseudogene in the Italian wall lizard: intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus *Podarcis* (Lacertidae). *Journal of Molecular Evolution*, **64**, 308–320.

Richly E, Leister D (2004) NUMTs in sequenced Eukaryotic genomes. *Molecular Biology and Evolution*, **21**, 1081–1084.

Saiki RK, Gelfand DH, Stoffel S *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Schmitz J, Piskurek O, Zischler H (2005) Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in Primates. *Journal of Molecular Evolution*, **61**, 1–11.

Schwartz M, Vissing J (2002) Paternal inheritance of mtDNA in a patient with mitochondrial myopathy. *European Journal of Human Genetics*, **10**, 239–239.

Sherengul W, Kondo R, Matsuura ET (2006) Analysis of paternal transmission of mitochondrial DNA in *Drosophila*. *Genes and Genetic Systems*, **81**, 399–404.

Shitara H, Hayashi J, Takahama S, Kaneda H, Yonekawa H (1998) Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics*, **148**, 851–857.

Shokralla S, Zhou X, Janzen DH *et al.* (2011) Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE*, **6**, e21252.

Skibinski DOF, Gallagher C, Beynon CM (1994) Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics*, **138**, 801–809.

Smith MF, Thomas WK, Patton JL (1992) Mitochondrial DNA-like sequence in the nuclear genome of an akodontine rodent. *Molecular Biology and Evolution*, **9**, 204–215.

Steinborn R, Zakhartchenko V, Jelyazkov J *et al.* (1998) Composition of parental mitochondrial DNA in cloned bovine embryos. *FEBS Letters*, **426**, 352–356.

Sutovsky P, Moreno RD, Ramalho-Santos J *et al.* (1999) Development: ubiquitin tag for sperm mitochondria. *Nature Genetics*, **402**, 371–372.

Sutovsky P, Moreno RD, Ramalho-Santos J *et al.* (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biology of Reproduction*, **63**, 582–590.

Sutovsky P, Van Leyen K, McCauley T, Day BN, Sutovsky M (2004) Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. *Reproductive Biomedicine Online*, **8**, 24–33.

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

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.

Triant DA, DeWoody JA (2007) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy*, **88**, 908–920.

Ujvari B, Dowton M, Madsen T (2007) Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biology Letters*, **3**, 189–192.

Ujvari B, Dowton M, Madsen T (2008) Population genetic structure, gene flow and sex-biased dispersal in frillneck lizards (*Chlamydosaurus kingii*). *Molecular Ecology*, **17**, 3557–3564.

Van Leeuwen T, Vanholme B, Van Pottelberge S *et al.* (2008) Mitochondrial heteroplasmy and the evolution of insecticide resistance: non-Mendelian inheritance in action. *Proceedings of the National Academy of Sciences, USA*, **105**, 5980–5985.

Zhang D-X, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247–251.

Zhao X, Li N, Guo W *et al.* (2004) Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity*, **93**, 399–403.

Zischler H, Geiser H, Haeseler A, Pääbo S (1995) A nuclear ''fossil'' of the mitochondrial D-loop and the origin of modern humans. *Nature*, **378**, 489–492.

Zouros E, Ball A, Saavedra C, Freeman K (1994) An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences, USA*, **91**, 7463–7467.

A.M. is interested in the application of genetic data to infer species demographic histories and in the use of population genomic approaches to study adaptation, selection and speciation. G.M.H. is Professor Emeritus in evolutionary genetics with particular interests in Quaternary phylogeography, refugia and range changes, hybrid zones and speciation. P.H.D. works on single-molecule genomics and develops techniques for genomic analysis. O.S.P. is currently working on phylogeography and evolutionary ecological genomics. B.C.E. is an evolutionary biologist with interests in the application of molecular data to interpret phylogenetic history and population dynamics, particularly within arthropods and islands ecosystems.

Data accessibility

DNA sequences: GenBank accession numbers JN860936-JN860996. GenBank accession numbers of individual samples are available in Table S1 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Detailed information of single molecule PCR (smPCR) protocol.

Table S1 Complete dataset for each sample, including individual GenBank Accession nos.

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