Gene 512 (2013) 166–173

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Evolutionary patterns of the mitochondrial genome in the Moorish gecko, *Tarentola mauritanica*

C. Rato [a,b,c,*], S. Carranza [c,1], A. Perera [a,2], D.J. Harris [a,b,2]

[a] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
[b] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal
[c] Instituto de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta, 37-49, E-08003 Barcelona, Spain

ARTICLE INFO

Article history:
Accepted 5 September 2012
Available online 18 September 2012

Keywords:
Tarentola mauritanica
Selective sweep
Positive selection
Background selection
Polymorphism
Bottleneck

ABSTRACT

A previous study on the evolutionary patterns of *Tarentola mauritanica* demonstrated that low levels of mitochondrial diversity observed in the European populations relative to nuclear markers were consistent with a selective sweep hypothesis. In order to unravel the mitochondrial evolutionary history in this European population and two other lineages of *T. mauritanica* (Iberian and North African clades), variation within 22 nearly complete mitogenomes was analyzed. Surprisingly, each clade seems to have a distinct evolutionary history; with both the European and Iberian clades presenting a decrease of polymorphism, which in the former is consistent with departure from neutrality of the mtDNA (positive or background selection), but in the latter seems to be the result of a bottleneck after a population expansion. The pattern exhibited by the North African clade seems to be a consequence of adaptation to certain mtDNA variants by positive selection.

1. Introduction

The vertebrate mitochondrial genome is around 16–18 kb in size, circular, with double stranded molecules, and typically composed of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, a non-coding control region, and some intergenic spacers. Over the last 30 years, mitochondrial DNA has been the most popular and commonly used molecular marker to estimate genetic diversity in a variety of organisms and its initial success was mainly because it was thought to be always maternally inherited, nearly neutral, and with an almost clock-like nature of its substitution rates. Because these organelles perform such important role in cellular respiration, and it is where most of the ATP is produced, any type of malfunction could be lethal or would seriously compromise an individual's fitness. Therefore, it was initially assumed that the mitochondrial genome would evolve primarily under constant purifying selection that would eliminate possible deleterious mutations. However, a number of exceptions have been reported showing evidences of positive selection acting on the mitochondria's evolution (Castoe et al., 2008, 2009; Grossman et al., 2004; Yu et al., 2011). Nevertheless, neutrality assumptions were previously challenged by Bazin et al. (2006), using a broad range of vertebrate and invertebrate animals, concluding that mtDNA genetic diversity is not affected by population size, which could be explained by recurrent selective sweeps in large populations. These recurrent sweeps are in agreement with the "genetic draft" model of Gillespie (2000). Considering the non-recombining nature of the mitochondria, this genetic hitch-hiking has the potential to cause strong selective sweeps (Hamilton, 2009). In fact, in the past few years several cases of possible selective sweeps have been recorded, namely in warblers (Bensch et al., 2006), birds from the genus *Emberiza* (Irwin et al., 2009), butterflies (Jiggins, 2003), and Hawaiian crickets (Shaw, 2002). More recently, it has been suggested that the mtDNA pattern presented by the European populations of two species of geckos, *Tarentola mauritanica* (Rato et al., 2010), and *Hemidactylus turcicus* (Rato et al., 2011) could have been shaped by a selective sweep leading to a decrease of the nucleotide diversity to levels below those presented by nuclear markers. However, at that time this hypothesis could not be verified at the genome level.

Abbreviations: Ala, Alanine; Asn, Asparagine; Asp, Aspartic acid; Arg, Arginine; ATP6, ATP synthase 6; ATP8, ATP synthase 8; bp, Base pair; COXI, Cytochrome c oxidase 1; COXII, Cytochrome c oxidase 2; COXIII, Cytochrome c oxidase 3; Cys, Cysteine; CYTB, Cytochrome b; dNTP, Deoxyribonucleoside triphosphate; Gln, Glutamine; Glu, Glutamic acid; Gly, Glycine; HKA test, Hudson, Kreitman and Aguadé test; His, Histidine; Ile, Isoleucine; Leu, Leucine; Lys, Lysine; Met, Methionine; MgCl2, Magnesium chloride; MLHKA, Maximum likelihood version of the HKA test; mtDNA, Mitochondrial DNA; ND1, NADH dehydrogenase 1; ND2, NADH dehydrogenase 2; ND3, NADH dehydrogenase 3; ND4, NADH dehydrogenase 4; ND4L, NADH-ubiquinone oxidoreductase chain 4L; ND5, NADH dehydrogenase 5; ND6, NADH dehydrogenase 6; Numt, Nuclear mitochondrial DNA; PCR, Polymerase chain reaction; Phe, Phenylalanine; Pro, Proline; rRNA, Ribosomal RNA; Ser, Serine; Thr, Threonine; tRNA, Transfer RNA genes; Trp, Tryptophan; Tyr, Tyrosine; Val, Valine.

* Corresponding author at: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. Tel./fax: +351 252660411, +351 252 661 780.
E-mail addresses: catarina.rato@cibio.up.pt (C. Rato), salvador.carranza@ibe.upf-csic.es (S. Carranza), perera@cibio.up.pt (A. Perera), james@cibio.up.pt (D.J. Harris).
[1] Tel./fax: +34 93 230 96 44, +34 93 230 95 55.
[2] Tel./fax: +351 252660411, +351 252 661 780.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.09.032

Nevertheless, genetic hitch-hiking can also take place when negative or purifying selection removes deleterious mutations from the population driving them to loss (Hamilton, 2009). This negative selection against deleterious mutations is also expected to reduce polymorphism through hitch-hiking in a process called background selection (Charlesworth et al., 1993).

In the present study, we have used 22 nearly complete mitochondrial genomes of *T. mauritanica* to test selective sweep hypothesis. This gecko species belonging to the Family Phyllodactylidae (Gamble et al., 2008), is apparently a complex of several highly divergent cryptic species (Harris et al., 2004a, 2004b, 2009; Perera and Harris, 2008), but evidences of a selective sweep were detected in only one lineage, geographically widespread across Europe and North Africa (Rato et al., 2010). Thus, used mitogenomes in this study correspond to three different lineages of *T. mauritanica*; one lineage previously identified as subject to a selective sweep, and two others where this process was not detected. Minimal genetic variation is expected in the European lineage, but also the detection of fixed mutations responsible for amino acid changes that could be the driving force behind the sweep. Populations from the Iberian Peninsula are also expected to present little mitochondrial genetic diversity, consistent with a post-bottleneck demographic expansion pattern, since they also present low values of variation for nuclear markers (Rato et al., 2010). We report on rates and patterns of mtDNA sequence variation, and perform specific tests to detect selection.

2. Material and methods

2.1. Sample collection and DNA extraction

The 22 specimens used in this study are representatives of three different lineages of *T. mauritanica* identified in previous studies (Harris et al., 2004a, 2004b, 2009; Perera and Harris, 2008; Rato et al., 2010); namely six specimens (DB11105, DB9107, DB9112, DB9113, DB9115, and DB9116) representing the widespread European lineage; five representatives (DB3832, DB3839, DB3843, DB3846, DB3853) from the Iberian Peninsula lineage; and 10 specimens (DB11003, DB11004, DB11007, DB11008, DB11009, DB11013, DB11022, DB11035, DB11042, DB2635) from the Central and Southern Morocco lineage. A complete mitochondrial genome from *T. mauritanica* already available in GenBank (EU443255.1 from Albert et al., 2009) was used in the analyses, and assigned to the European Clade according to the phylogenetic results from 12SrRNA and 16SrRNA fragments (data not shown).

Genomic DNA extraction from fresh tail muscle tissue was carried out using Qiagen's DNeasy Blood & Tissue Kit, following the manufacturer's protocol.

2.2. Primer design, amplification and sequencing

Since amplification of several mitochondrial regions was unsuccessful using the primers from Albert et al. (2009), new primers were developed for this study. The complete genome from these authors was used as a reference for primer design, as well as for the alignment of the obtained amplicons. Primers were designed by hand, and checked for quality using the software AmplifX v1.5.4 (http://ifrjr.nord.univ-mrs.fr/AmplifX-Home-page). The complete list of primers used in this study is represented in Table S1 from Additional file 1.

Mitochondrial fragments were amplified through a standard Polymerase Chain Reaction (PCR) protocol, using 25 μl volumes, containing 2.5 μl of 10× reaction buffer, 2.0 mM of $MgCl_2$, 0.5 μM of each dNTP, 0.2 mM of each primer, 1 U of Taq DNA polymerase (Invitrogen), and approximately 100 ng of template DNA, and under the following cycling conditions for all primer sets; 94 °C for 3 min of initial denaturation followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 48–54 °C for 30 s, and extension at 72 °C for 1 min and 30 s. Amplification ended with a 10 min at 72 °C final extension step. The amplicons were checked in a 2% agarose gel with a negative control. Sequencing of the PCR products was performed in Macrogen Inc. (Korea) using an ABI 3730XL DNA Analyzer (Applied Biosystems). All obtained genomes were deposited in GenBank with accession numbers from JQ425040 to JQ425060.

2.3. Genome annotation

The newly obtained sequences were assembled and aligned with the currently available mitochondrial genome of *T. mauritanica*, using the software package Geneious Pro v5.4.3 (Drummond et al., 2010). Alignment was carried out with MAFFT v6.814b (Katoh et al., 2002) considering the auto algorithm, gap open penalty = 1.53, offset value = 0.123, and the scoring matrix 200PAM/k = 2. The protein-coding gene sequences were identified and annotated in Geneious Pro v5.4.3 by the start and stop codons, and sequence similarity in the alignment. Inconsistencies between the obtained sequences in this study, and the mitogenome from Albert et al. (2009) were detected, regarding the orientation of two tRNAs (tRNA-Gln and tRNA-Glu). Using the program ARWEN for tRNA detection in metazoan mitochondrial sequences (Laslett and Canbäck, 2008), we identified an incorrect orientation of both tRNA fragments on the GenBank sequence. Since transfer or "transposition" of cytoplasmic mitochondrial DNA sequences into the separate nuclear genome of a eukaryotic organism (Numt) is a possible phenomenon (Bensasson et al., 2001), all amplicons were translated into amino acids in order to inspect for the presence of stop codons inside the protein-coding genes.

In order to identify the nucleotide fixed differences in the European clade compared to the remaining lineages, we used the software SITES (Hey and Wakeley, 1997), and classified these differences into synonymous and non-synonymous using the Geneious Pro v5.4.3 software package.

2.4. Population genetic analyses and tests for selection

Population genetic parameters, such as the nucleotide diversity (π), number of segregating sites, number of invariable sites and average number of nucleotide differences between lineages, were calculated using DnaSP v5.10 (Librado and Rozas, 2009) for each lineage and locus. In order to calculate the nucleotide diversity across the coding portion of the mitochondrial genome, we used a 500 steps sliding window with a step size of 100 bp. Calculations were performed for each lineage, and for the complete set of mtDNA genomes of *T. mauritanica*.

Within each lineage, all coding genes were tested for evidences of selection using Tajima's *D*, which compares the number of rare to intermediate-frequency variants (Tajima, 1989). However, a test sensitive to high-frequency variants such as Fay and Wu's H is particularly useful to detect hitch-hiking and therefore, positive selection (Fay and Wu, 2000). This test was implemented using as outgroups two mitogenome sequences of *Gekko gecko* (GenBank accession numbers: AY282753.1 and HM370130.1), which belongs to the Gekkonidae, the sister Family of Phyllodactylidae (Gamble et al., 2008). Both tests were implemented in the software package DnaSP v5.10 (Librado and Rozas, 2009), where also statistical significance was assessed through coalescence, considering 10,000 simulations.

Another neutrality test was implemented, namely a maximum likelihood version of the Hudson, Kreitman and Aguadé (HKA) test (Hudson et al., 1987) developed by Wright and Charlesworth (2004) and available from S.I. Wright (http://www.yorku.ca/stephenw/). Similar to the HKA test, the MLHKA compares sequence polymorphism within species to rates of divergence between species, in order to test for deviations from neutrality. Nevertheless, while the HKA test does not identify which loci deviate from expectations, MLHKA is able to test directly the specific loci for evidences of selection.

The MLHKA test was conducted for each locus and each lineage, using as the outgroup taxon the two mitogenomes of *Gekko gecko*

used in Fay and Wu's H. To conduct the tests, we ran 100,000 simulations in MLHKA to compare the fit of a neutral model to a model specifying each selected gene. The statistical significance was assessed by the likelihood ratio test, where twice the difference in log likelihood ($2^*\Delta\ln$) between the models is approximately chi-squared distributed, with the degrees of freedom (d.f.) equal to the difference in the number of parameters (d.f. = 1). The value of k (selection parameter) was also assessed, since $k<1$ means there is a reduction in polymorphism, while $k>1$ means there is an excess of nucleotide variation. All 13 coding genes from *Gekko gecko*'s mitogenome did not deviated from neutral expectations (sum of deviations = 27.6278; d.f. = 24; *P*-value = 0.27605) based on the standard multilocus HKA test, using the source code from J. Hey (http://genfaculty.rutgers.edu/hey/software#HKA). Correction for multiple testing was assessed using the false discovery rate test (Benjamini and Hochberg, 1995), implemented in R.

2.5. Protein modeling

Protein tertiary structures of each mitochondrial coding gene from each clade were predicted using Phyre^2 (Kelley and Sternberg, 2009), a webserver for modeling 3D atomic level structure of multi-subunit protein complexes, using the principles and techniques of homology modeling. Results from Phyre^2 were visualized and analyzed with the software Chimera (Pettersen et al., 2004). With this software we can localize and visualize the atomic effects when a certain amino acid is changed in the alignment. We specifically pinpointed the sites corresponding to the fixed non-synonymous differences in the European clade. We also identified interatomic clashes and contacts between protein structures (such as the COX complex) based on van der Waals (VDW) radii.

3. Results

The nearly complete mitochondrial genome was sequenced for all 21 individuals corresponding to a total of 15,354 bp for the European clade, and 15,373 bp and 15,347 bp for the Iberian and North African clades, respectively. The only regions that were unsuccessfully sequenced correspond to the control region, and portions of the tRNA-Phe and tRNA-Pro, as illustrated in Additional file 2, Fig. S1. Therefore, the obtained mitochondrial genome of *T. mauritanica*, is composed of 13 protein-coding genes, 22 transfer-RNAs (two of them partially sequenced), two ribosomal RNAs, and some intergenic spacers. The gene organization obtained conforms to the "vertebrate" consensus mitochondrial gene arrangement (Boore, 1999; Jameson et al., 2003), matching also the organization observed in some other reptiles (mole skink: Kumazawa and Nishida, 1999; green iguana: Janke et al., 2001). However, as already noted (Gissi et al., 2008), the majority of the reptile mtDNA genomes actually have rearrangements relative to this consensus pattern. Regarding the protein-coding genes, these were all of the same length between lineages, except NADH dehydrogenase 2 (ND2); this gene had a size of 1035 bp in the European clade, 1032 bp in the Iberian clade, and 1029 bp for the North African clade.

3.1. Patterns of DNA sequence variation in the mitochondrial genome

One of the most striking features of the genome-wide sequence variation analyses (Fig. 1) is the lower nucleotide diversity in the European and Iberian clades relative to the North African one, as already reported in a previous study (Rato et al., 2010). Within the European clade, the highest values occur in CYTB, and also in a region between COXII and COXIII, while the lowest nucleotide diversity occurs in the terminal end of ND4. In the Iberian clade, the peaks of sequence variation occur in COXI and ND5, and the lowest values are located in the terminal part of CYTB, ND1 and ND3, within ND4, and in the initial part of ND6. The North African clade is characterized by an elevated overall nucleotide diversity with several peaks across the genome, and lower variation in regions of COXI and COXII. When all lineages are considered, most mitochondrial regions describe elevated levels of diversity, with the peaks of maximum variability located in ATP8 and ND5 genes.

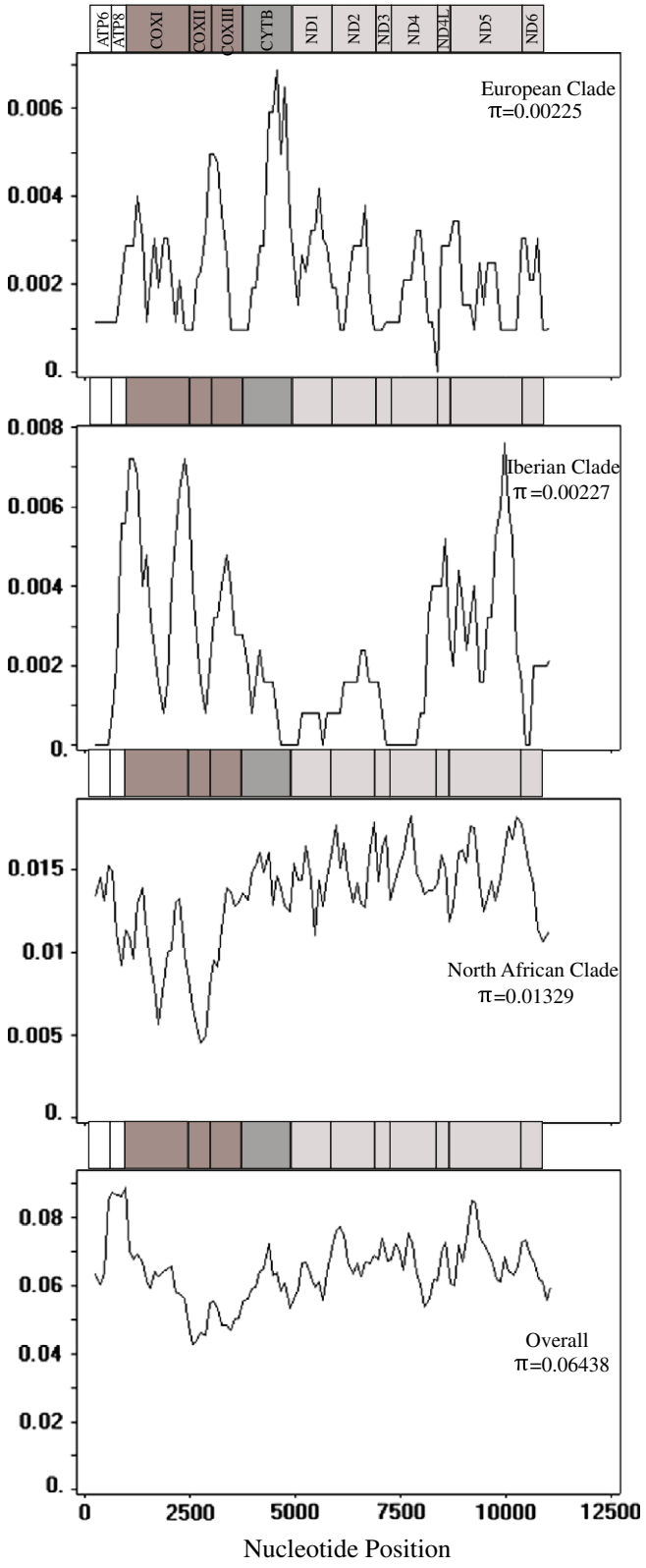

Fig. 1. Genome-wide nucleotide diversity (π) for all coding genes within the European, Iberian and North African clades and among all mitochondrial genomes of *T. mauritanica*. The values of π on each graphic correspond to the overall nucleotide diversity.

3.2. Selection constraint analyses

Regarding the results for the neutrality tests, for both European and Iberian clades no significant values were obtained for either Tajima's *D* or Fay and Wu's H. On the contrary, the North African clade presents significant values only for Fay and Wu's H for the majority of the genes (except ATP8, ND4L and ND6). However, since both these tests are determined by the frequency of mutations, when the polymorphism and number of segregating sites is reduced as in the cases of the European and Iberian clades, these tests might not have the power to detect deviations from neutrality as the MLHKA test, which is based on polymorphism and divergence. Values from the MLHKA test indicate that within the European clade the genes ATP6, ATP8, ND3, ND4, ND4L, ND5 and ND6 are under selection; within the Iberian clade the test was statistically significant for ATP6, CYTB, ND1, ND3 and ND4. For both these clades (European and Iberian), the genes identified as deviating from neutrality correspond to the zones in the genome with lower values of nucleotide diversity (Fig. 1). Regarding the North African clade, most genes (except COXI and ND5) deviate significantly from neutrality, and there is no clear correspondence to areas of low polymorphism of the genome, with exception of COXII. The remaining coding genes of this clade present high level of polymorphism compared to the remaining clades, which can be verified by the value of the selection parameter, k.

3.3. Protein modeling

Results for protein modeling were only possible for the COX subunits I, II and III as well as for the CYTB genes (Figs. 2 and 3), since these were the only models generated by Phyre[2] obtained with 100% confidence. However, the species chosen by this software as model templates were distinct for each of the two groups of genes (COX genes and CYTB). For the remaining genes, the confidence in the model was below 50%, and therefore these were not considered.

All models for the COX genes were generated using the tertiary protein structure of *Bos taurus* as a template. Both COXI and COXIII templates belong to the SCOP (Structural Classification of Proteins) database, with codes d1v54a, and d1v54c, respectively. For the construction of the COXII protein structure we used the template with the code ID c1v55B, from the Protein Data Bank. Compared to the template, COXI from all *Tarentola* lineages had 84% similarity; COXII from the European and North African clades had 69% similarity, and from the Iberian Clade 68%; and COXIII from the European and Iberian clades had 75% similarity, and from the North African clade, 76%.

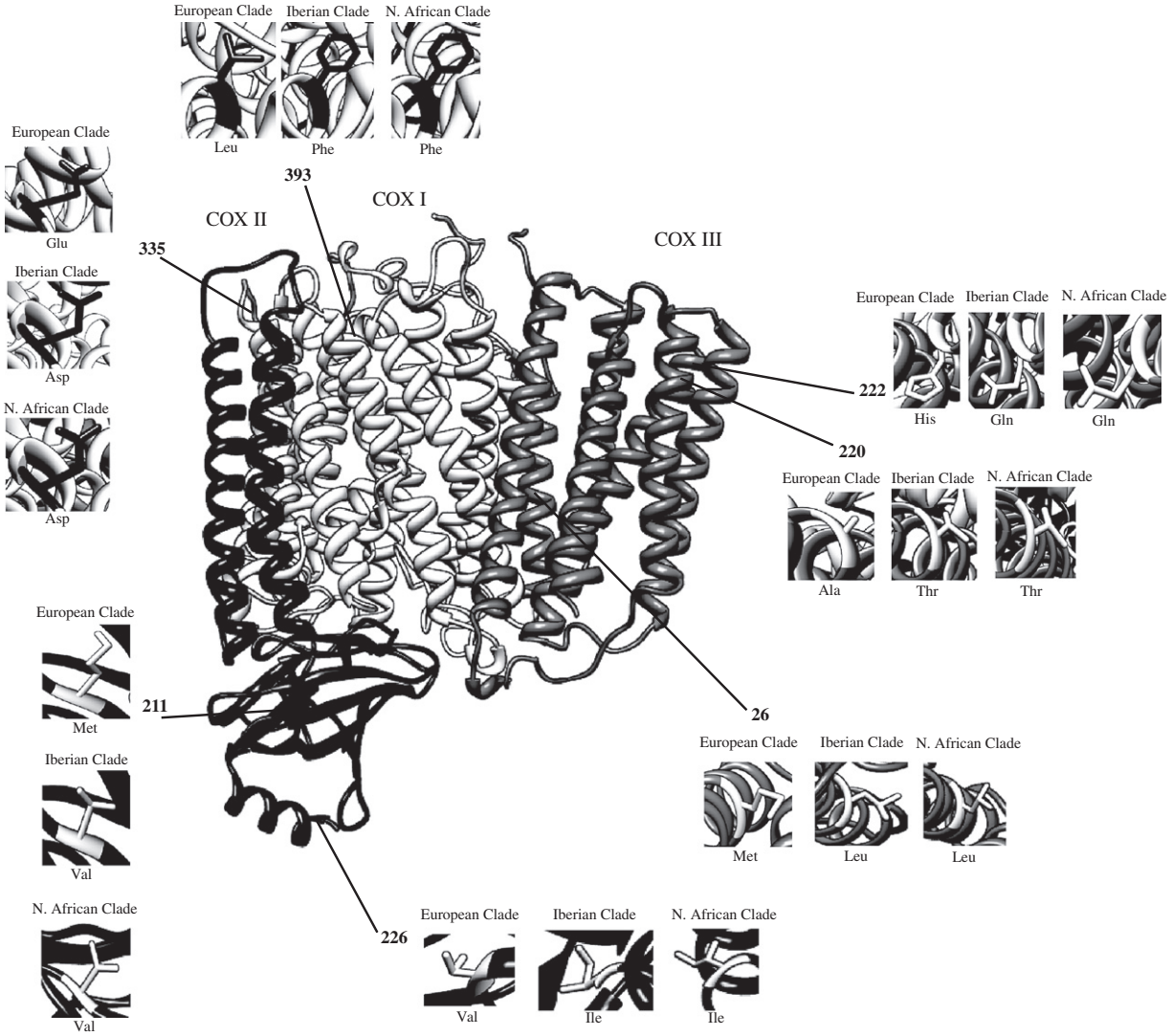

Fig. 2. Inferred model for the tertiary protein structure of cytochrome c oxidase (COX) complex, including subunits I, II, and III. The marked positions correspond to the non-synonymous fixed changes in the European clade and corresponding atomic structure in each lineage.

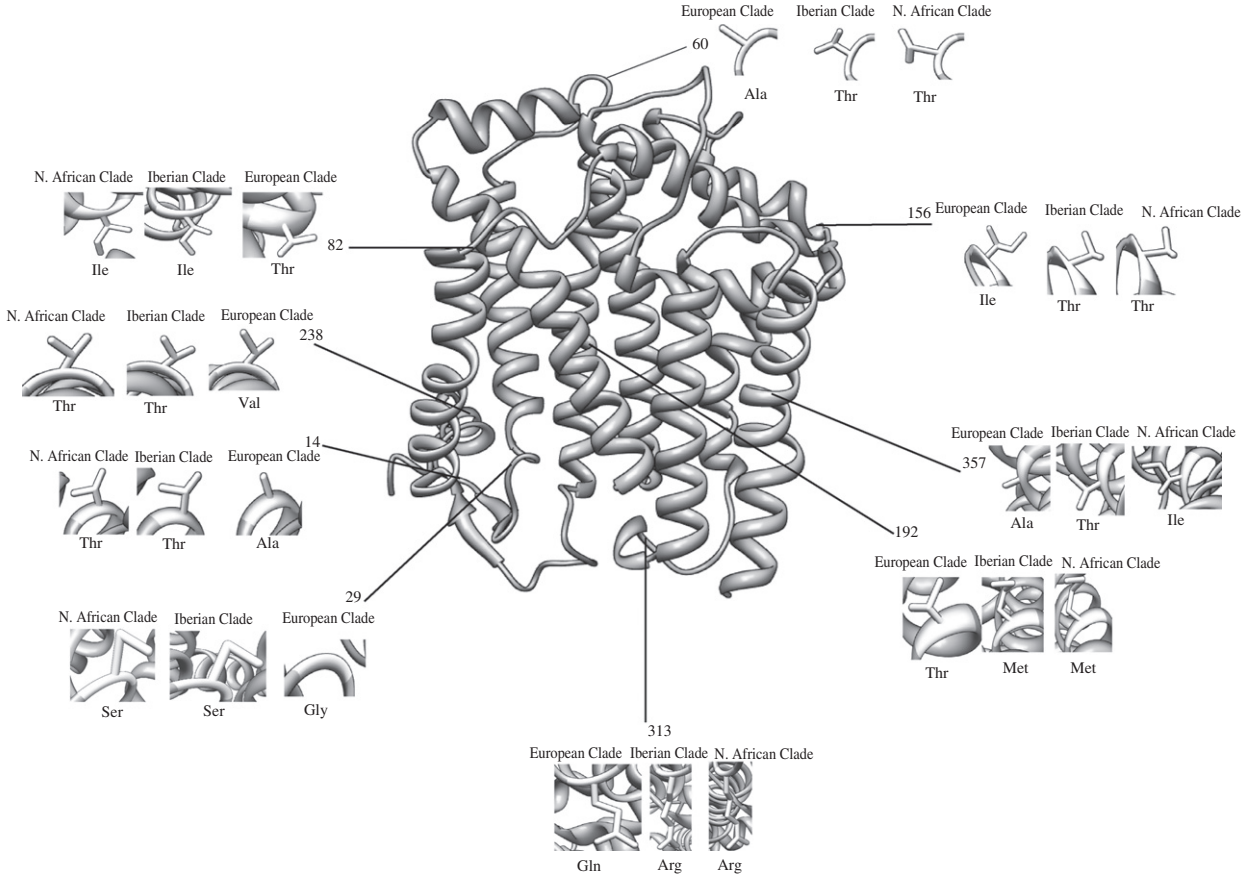

Fig. 3. Inferred model for the tertiary protein structure of cytochrome b (CYTB). The marked positions correspond to the non-synonymous fixed changes in the European clade and corresponding atomic structure in each lineage.

Regarding the CYTB gene, the corresponding 3D model was generated from the chicken cytochrome bc1 complex as template, with Protein Data Bank ID 3CWB. Amino acid similarity of the European, Iberian and North African clades towards the template was 72%, 73%, and 72%, respectively.

Using the results obtained from the software SITES (Table 2) we identified the fixed amino acid changes in the European clade, and located them in each protein model. The atomic changes within the tertiary structure of each protein are illustrated in Figs. 2 and 3. Since COX is a large transmembrane protein complex encoded by the three mitochondrial genes COXI, COXII and COXIII we have decided to represent them together (Fig. 2). Curiously within this protein complex, all non-synonymous fixed differences within the European clade were unique to this lineage, meaning that for the same sites the remaining clades were quite conserved presenting exactly the same amino acid residue. The same is demonstrated within CYTB, except for the position 357, where each clade presents a different residue. In fact, the overall patterns of sequence variation (Fig. 1) demonstrate that COXII, COXIII, and CYTB present the lowest values of nucleotide diversity, indicating that among these genes variation is quite conserved in *T. mauritanica*.

Table 1
Results obtained for Tajima's *D*, Fay and Wu's H, the selection parameter (k), and likelihood ratio test for each lineage and locus with the MLHKA test (Wright and Charlesworth, 2004). Significant values (*P*-value<0.05) are marked in bold and have been corrected with the false discovery rate test (Benjamini and Hochberg, 1995).

Locus	European clade					Iberian clade					North African clade				
	Tajima's *D*	Fay and Wu's H	MLHKA			Tajima's *D*	Fay and Wu's H	MLHKA			Tajima's *D*	Fay and Wu's H	MLHKA		
			k	2*Δln	*P*-value			k	2*Δln	*P*-value			k	2*Δln	*P*-value
ATP6	1.650 n.s.	−0.190 n.s.	0.139	7.324	**0.043**	n.a.	n.a.	0.000	14.298	**0.000**	−1.126 n.s.	**−14.133** (0.030)	0.310	18.034	**0.000**
ATP8	n.a.	n.a.	0.000	5.343	**0.043**	n.a.	n.a.	0.000	2.715	0.135	0.263 n.s.	−1.156 n.s.	0.421	21.755	**0.000**
COXI	1.208 n.s.	−1.333 n.s.	0.220	2.532	0.132	0.115 n.s.	−3.200 n.s.	0.659	3.251	0.119	−0.830 n.s.	**−23.467** (0.039)	0.707	2.322	0.128
COXII	1.459 n.s.	−1.095 n.s.	0.215	3.720	0.070	−0.175 n.s.	0.300 n.s.	0.164	2.542	0.135	−1.454 n.s.	**−4.800** (0.040)	0.169	11.204	**0.001**
COXIII	0.518 n.s.	0.810 n.s.	1.000	3.814	0.070	0.286 n.s.	−1.000 n.s.	0.365	1.694	0.209	−0.635 n.s.	**−9.422** (0.040)	0.479	18.397	**0.000**
CYTB	1.024 n.s.	0.286 n.s.	0.352	1.119	0.314	−1.048 n.s.	−1.200 n.s.	0.140	8.092	**0.022**	−1.370 n.s.	**−31.022** (0.030)	0.616	6.31	**0.014**
ND1	0.109 n.s.	−0.667 n.s.	0.416	0.118	0.731	−0.973 n.s.	−0.900 n.s.	0.099	7.868	**0.022**	−1.237 n.s.	**−18.044** (0.036)	0.453	20.524	**0.000**
ND2	0.797 n.s.	−0.143 n.s.	0.156	4.016	0.070	−1.094 n.s.	−1.800 n.s.	0.163	2.499	0.135	−1.518 n.s.	**−31.022** (0.030)	0.627	14.404	**0.000**
ND3	n.a.	n.a.	0.000	9.957	**0.026**	n.a.	n.a.	0.000	7.234	**0.023**	−1.214 n.s.	**−10.400** (0.036)	0.399	24.441	**0.000**
ND4	1.439 n.s.	−1.476 n.s.	0.204	5.613	**0.043**	−1.146 n.s.	−0.600 n.s.	0.206	4.896	**0.070**	−0.882 n.s.	**−19.200** (0.040)	0.508	11.212	**0.001**
ND4L	0.559 n.s.	−0.714 n.s.	0.165	5.147	**0.043**	1.225 n.s.	n.a.	0.110	3.215	0.119	−1.109 n.s.	−2.044 n.s.	0.379	19.841	**0.000**
ND5	0.416 n.s.	−1.333 n.s.	0.218	6.702	**0.043**	−0.512 n.s.	2.500 n.s.	0.326	1.361	0.243	−0.973 n.s.	**−30.844** (0.039)	0.571	4.684	0.033
ND6	0.559 n.s.	0.286 n.s.	0.076	5.840	**0.043**	0.243 n.s.	−1.200 n.s.	0.139	3.536	0.119	−1.200 n.s.	−4.444 n.s.	0.318	25.423	**0.000**

Table 2
Amino acid fixed differences in the European Clade. [n], nonpolar and neutral; [p], polar and neutral; [+], positively charged; [−], negatively charged; [h], hydrophobic; [H], hydrophilic.
Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B. 57, 289–300.

Amino acid residue	European clade	Iberian clade	North African clade
ND1			
27	Met^{n,h}	Leu^{n,h}	Leu^{n,h}
69	Ala^{n,h}	Ser^{p,H}	Thr^{p,H}
109	Leu^{n,h}	Ile^{n,h}	Ile^{n,h}
115	Gly^{n,H}	Ser^{p,H}	Ser^{p,H}
172	Thr^{p,H}	Ile^{n,h}	Val^{n,h}
304	Met^{n,h}	Leu^{n,h}	Leu^{n,h}
321	Val^{n,h}	Ile^{n,h}	Ile^{n,h}
322	Ser^{p,H}	Asn^{p,H}	Asn^{p,H}
ND2			
12	Thr^{p,H}	Ser^{p,H}	Ser^{p,H}
81	Ser^{p,H}	Thr^{p,H}	Thr^{p,H}
93	Val^{n,h}	Ile^{n,h}	Met^{n,h}
100	Met^{n,h}	Val^{n,h}	Val^{n,h}
124	Met^{n,h}	Leu^{n,h}	Leu^{n,h}
160	Ile^{n,h}	Thr^{p,H}	Ala^{n,h}
244	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
309	Thr^{p,H}	Ala^{n,h}	Ala^{n,h}
324	Thr^{p,H}	Ala^{n,h}	Ala^{n,h}
327	Met^{n,h}	Leu^{n,h}	Leu^{n,h}
COXI			
335	Glu$^{-,H}$	Asp$^{-,H}$	Asp$^{-,H}$
393	Leu^{n,h}	Phe^{n,h}	Phe^{n,h}
COXII			
211	Met^{n,h}	Val^{n,h}	Val^{n,h}
226	Val^{n,h}	Ile^{n,h}	Ile^{n,h}
ATP8			
7	Gly^{n,H}	Asn^{p,H}	Thr^{p,H}
33	Gln^{p,H}	Arg$^{+,H}$	Arg$^{+,H}$
34	Ser^{p,H}	Pro^{n,h}	Leu^{n,h}
45	Pro^{n,h}	His$^{+,H}$	Arg$^{+,H}$
ATP6			
11	Pro^{n,h}	Ser^{p,H}	Ser^{p,H}
14	Thr^{p,H}	Ile^{n,h}	Ile^{n,h}
50	Ser^{p,H}	Thr^{p,H}	Thr^{p,H}
135	Met^{n,h}	Val^{n,h}	Val^{n,h}
COXIII			
26	Met^{n,h}	Leu^{n,h}	Leu^{n,h}
220	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
222	His$^{+,H}$	Gln^{p,H}	Gln^{p,H}
ND3			
10	Thr^{p,H}	Ile^{n,h}	Ile^{n,h}
83	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
ND4L			
53	Asn^{p,H}	Thr^{p,H}	Thr^{p,H}
ND4			
22	Tyr^{p,H}	His$^{+,H}$	His$^{+,H}$
29	Ser^{p,H}	Ala^{n,h}	Ala^{n,h}
52	Val^{n,h}	Ile^{n,h}	Ile^{n,h}
172	Arg$^{+,H}$	Gln^{p,H}	Lys$^{+,H}$
185	Ser^{p,H}	Asn^{p,H}	Asn^{p,H}
254	Ser^{p,H}	Cys^{p,H}	Cys^{p,H}
310	Val^{n,h}	Met^{n,h}	Met^{n,h}
317	Ile^{n,h}	Leu^{n,h}	Leu^{n,h}
337	Met^{n,h}	Ile^{n,h}	Ile^{n,h}
369	Ile^{n,h}	Leu^{n,h}	Leu^{n,h}
395	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
419	Thr^{p,H}	Ala^{n,h}	Ala^{n,h}
440	Thr^{p,H}	Met^{n,h}	Met^{n,h}
443	Val^{n,h}	Ala^{n,h}	Ala^{n,h}
444	Tyr^{p,H}	His$^{+,H}$	His$^{+,H}$
446	Gly^{n,H}	Asp$^{-,H}$	Glu$^{-,H}$
ND4			
448	Met^{n,h}	Val^{n,h}	Val^{n,h}
449	Phe^{n,h}	Tyr^{p,H}	Tyr^{p,H}
ND5			
3	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
30	Ala^{n,h}	Ile^{n,h}	Ile^{n,h}
36	Thr^{p,H}	Ala^{n,h}	Ala^{n,h}
53	Ala^{n,h}	Ile^{n,h}	Ile^{n,h}, Met^{n,h}, Thr^{p,H}
54	Lys$^{+,H}$	Gln^{p,H}	Gln^{p,H}
70	Ser^{p,H}	Thr^{p,H}	Thr^{p,H}
156	Ala^{n,h}	Ser^{p,H}	Ser^{p,H}
185	Glu$^{-,H}$	His$^{+,H}$	His$^{+,H}$
195	Phe^{n,h}	Tyr^{p,H}	Tyr^{p,H}
306	Phe^{n,h}	Leu^{n,h}	Val^{n,h}
409	Ser^{p,H}	Thr^{p,H}	Thr^{p,H}
420	Leu^{n,h}	Phe^{n,h}	Phe^{n,h}
422	Ile^{n,h}	Leu^{n,h}	Leu^{n,h}
441	Ser^{p,H}	Asn^{p,H}	Asn^{p,H}
484	Phe^{n,h}	Leu^{n,h}	Met^{n,h}
488	Thr^{p,H}	Val^{n,h}	Ala^{n,h}
505	Arg$^{+,H}$	Pro^{n,h}	Pro^{n,h}
506	Phe^{n,h}	Tyr^{p,H}	Ser^{p,H}
509	Asn^{p,H}	Lys$^{+,H}$	Lys$^{+,H}$
523	Val^{n,h}	Ile^{n,h}	Ile^{n,h}
524	Val^{n,h}	Met^{n,h}	Met^{n,h}
529	Pro^{n,h}	Ser^{p,H}	Ser^{p,H}
585	Leu^{n,h}	Thr^{p,H}	Met^{n,h}
ND6			
126	Gln^{p,H}	His$^{+,H}$	His$^{+,H}$
111	Phe^{n,h}	Leu^{n,h}	Leu^{n,h}
104	Val^{n,h}	Ala^{n,h}	Ala^{n,h}
102	Trp^{n,h}	Gly^{n,H}	Gly^{n,H}
82	Asp$^{-,H}$	Ala^{n,h}	Ala^{n,h}
79	Tyr^{p,H}	His$^{+,H}$	His$^{+,H}$
46	Ser^{p,H}	Gly^{n,H}	Gly^{n,H}
cytb			
14	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
29	Gly^{n,H}	Ser^{p,H}	Ser^{p,H}
60	Ala^{n,h}	Thr^{p,H}	Thr^{p,H}
82	Thr^{p,H}	Ile^{n,h}	Ile^{n,h}
156	Ile^{n,h}	Thr^{p,H}	Thr^{p,H}
192	Thr^{p,H}	Met^{n,h}	Met^{n,h}
238	Val^{n,h}	Thr^{p,H}	Thr^{p,H}
313	Gln^{p,H}	Arg$^{+,H}$	Arg$^{+,H}$
357	Ala^{n,h}	Thr^{p,H}	Ile^{n,h}

Regarding the number of interactions (clashes/contacts) within the COX protein complex, we were able to estimate a total of 1527 between COXI and COXII (1237), and COXI and COXIII (290), and these were exactly the same for all lineages. None of the positions of these interactions correspond to the fixed non-synonymous changes in the European clade, meaning that each of these mutations will not, in principle, affect directly the adjacent molecule.

4. Discussion

A previous study from Rato et al. (2010) concerning the evolutionary history of *T. mauritanica* revealed that the low degree of mitochondrial diversity within the European populations was probably not only due to a recent colonization, but it must have been also the result of selective forces acting on the genome causing a genetic hitch-hiking, and a decrease of nucleotide diversity. These authors suggested that this hitch-hiking process might have been due to a mitochondrial selective sweep. While genetic hitch-hiking could in some cases involve positive selection on beneficial mutations (i.e. selective sweep), it is also possible that new mutations could be deleterious (Hamilton, 2009). In that case, negative or purifying selection removes these deleterious mutations from the population by driving them to loss, consequently

reducing polymorphism, in a process known as background selection (Charlesworth et al., 1993).

Reduction of nucleotide diversity in the European and Iberian clades is quite evident from the mtDNA genome-wide analysis of sequence variation (Fig. 1), especially when compared with the North African clade. Although sample size is more reduced in both the European and Iberian clades compared to the African clade, which might account for these variability differences, results from Rato et al. (2010) corroborate that the European and Iberian clades have low levels of mtDNA diversity. In fact, in this study the authors used a higher sample size, but these lineages exhibited lower levels of nucleotide diversity (European clade = 0.00056 and Iberian clade = 0.00203) compared to the values obtained in the present study. The highest peaks of nucleotide diversity in the European clade are located in the CYTB gene, and in a region between COXII and COXIII. This observed pattern of nucleotide diversity in the European clade is not, however, maintained within the remaining lineages, since the peaks of high sequence variation are located in different coding-genes, with the North African clade having exceptionally elevated values for the majority of them. Nevertheless, a common feature of all lineages is the highly heterogeneous pattern of nucleotide diversity across the mitochondrial genome. Such patterns of sequence diversity may influence the choice of regions in the mtDNA for phylogeographic studies; highly variable regions of the genome although more prone to homoplasy relative to slowly evolving genes (Galtier et al., 2006), will be excellent options for intraspecific genealogies. In the case of *T. mauritanica*, the genes ATP8 and ND5 would be a good choice as markers to be used in a within-lineage analysis where fast-evolving markers are necessary to resolve relationships.

According to the results for the selection constraint analyses, based on polymorphism and divergence (MLHKA) (Table 1), the European clade presents several coding genes under selection, which correspond to the areas of the genome with lower levels of polymorphism (Fig. 1). According to the neutral theory, regions of the genome evolving at high rates, as revealed by interspecific DNA sequence comparisons, will also exhibit high levels of polymorphism within species (Hudson et al., 1987). Therefore, selective forces acting on individuals from the European lineage seem to be causing a decrease of nucleotide diversity within species, leading to a departure from neutrality. The same is verified in the Iberian clade, where genes detected as under selection are also the ones exhibiting lower levels of nucleotide diversity. While MLHKA is quite conservative test to cases of population expansion, population bottlenecks might artificially result in significant *P* values (Wright and Charlesworth, 2004), and explain the polymorphism deficit displayed by the European and Iberian clades. However, according to the previous study on *T. mauritanica* (Rato et al., 2010), the nuclear loci from the European clade do not show this lack of variation and, in fact, they exhibit higher nucleotide diversity values, compared to the used mtDNA markers. Therefore, this particular polymorphism pattern does not seem consistent with a bottleneck event, but with a selection event that is shaping the mitochondrial genome of the European populations of the Moorish gecko. On the contrary, a bottleneck could explain the pattern observed in the Iberian clade since this clade presents the lowest nucleotide diversity for the nuclear loci used (Rato et al., 2010), which is expected after a bottleneck event where all loci are affected. In fact, there are several studies reporting bottleneck events in the European herpetofauna, commonly after the Pleistocene's glaciations, for example in the lizards *Podarcis muralis* (Giovannotti et al., 2010) and *Lacerta schreiberi* (Paulo et al., 2001), the tree frog, *Hyla intermedia* (Canestrelli et al., 2007) and the Alpine salamander, *Salamandra atra* (Riberon et al., 2001). After the Pleistocenic climate started to ameliorate, a post-glacial expansion of the individuals that were confined to their refugia occurred, and this founder effect may have led to the observed loss of diversity (Avise, 2000; Hewitt, 2000; Slatkin and Hudson, 1991). The latest phylogenetic studies on Mediterranean *Tarentola* (Rato et al., 2012) demonstrate that the Iberian clade was already in the Iberian Peninsula during the Pleistocene. Thus, a bottleneck after a post-glacial expansion could explain the nuclear and mitochondrial diversity patterns observed in these Iberian populations.

Finally, the North African clade shows significant results for the Fay and Wu's H test, known to measure an excess of high frequency variants compared to intermediate frequency variants, making this test extremely powerful in detecting signs of positive selection (Fay and Wu, 2000). According to the MLHKA results for the selection parameter (k) (Table 1) and nucleotide diversity sliding window (Fig. 1), this clade does not demonstrate a decrease of the intraspecific mitochondrial polymorphism (except for COXII) and on the contrary, compared to the other two clades presents an excess of diversity. High levels of polymorphism can result from relaxed constraint in the species, or if some local adaptation is causing populations to diverge (Charlesworth, 2006; Kreitman, 2000). In the case of the North African clade detection of positive selection in most mitochondrial coding genes indicates that there might be some adaptation of distinct mtDNA variants to different environmental conditions.

To conclude, each clade of *T. mauritanica* inspected in this study, presents different evolutionary patterns as revealed by the nucleotide diversity patterns and selection constraint analyses; both European and Iberian clades have been subject to a loss of mitochondrial polymorphism, which in the former was due to selective forces, while in the latter the pattern is consistent with a bottleneck effect; the pattern of the North African clade seems to be coherent with adaptation of certain mtDNA variants due to positive selection.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.09.032.

Acknowledgments

C.R. is supported by the PhD grant SFRH/BD/36387/2007 from the Fundação para a Ciência e Tecnologia (FCT), and A.P. is supported by the post-doc grant SFRH/BPD/26546/2006 (FCT). This research study was supported by the FCT projects PTDC/BIA-BEC/105327/2008 (to A.P.); and by the Ministério de Educación y Ciencia, Spain: CGL2009-11663/BOS (to S.C.). S.C. and C.R. are part of the Grup de Recerca Emergent of the Generalitat de Catalunya: 2009SGR1462. Molecular graphic images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081). The authors would also like to acknowledge the anonymous reviewers for the suggestions and corrections, leading to a considerable improvement of this manuscript.

References

Albert, E.M., Mauro, D.S., García-París, M., Rüber, L., Zardoya, R., 2009. Effect of taxon sampling on recovering the phylogeny of squamate reptiles based on complete mitochondrial genome and nuclear gene sequence data. Gene 441, 12–21.
Avise, J.C., 2000. Phylogeography. The History and Formation of Species. University Press, Harvard.
Bazin, E., Glémin, S., Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312, 570–572.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
Bensasson, D., Zhang, D.-X., Hartl, D.L., Hewitt, G., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314–321.
Bensch, S., Irwin, D.E., Irwin, J.H., Kvist, L., Åkesson, S., 2006. Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. Mol. Ecol. 15, 161–171.
Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.
Canestrelli, D., Cimmaruta, R., Nascetti, G., 2007. Phylogeography and historical demography of the Italian tree frog, *Hyla intermedia*, reveals multiple refugia, population expansions and secondary contacts within peninsular Italy. Mol. Ecol. 16, 4808–4821.
Castoe, T.A., Jiang, Z.J., Gu, W., Wang, Z.O., Pollock, D.D., 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS One 3, e2201.
Castoe, T.A., et al., 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc. Natl. Acad. Sci. U. S. A. 106, 8986–8991.
Charlesworth, D., 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2, e64. http://dx.doi.org/10.1371/journal.pgen.0020064.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on neutral genetic variation. Genetics 134, 1289–1303.
Drummond, A.J., et al., 2010. Geneious v5.4. Available from http://www.geneious.com/.
Fay, J.C., Wu, C.-I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.
Galtier, N., Enard, D., Radondy, Y., Bazin, E., Belkhir, K., 2006. Mutation hot spots in mammalian mitochondrial DNA. Genome Res. 16, 215–222.
Gamble, T., Bauer, A.M., Greenbaum, E., Jackman, T.R., 2008. Out of the blue: a novel, trans-Atlantic clade of geckos (Gekkota, Squamata). Zool. Scr. 37, 355–366.
Gillespie, J.H., 2000. Genetic drift in an infinite population: the pseudohitchhiking model. Genetics 155, 909–919.
Giovannotti, M., Nisi-Cerioni, P., Caputo, V., 2010. Mitochondrial DNA sequence analysis reveals multiple Pleistocene glacial refugia for *Podarcis muralis* (Laurenti, 1768) in the Italian Peninsula. Ital. J. Zool. 77, 277–288.
Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 191, 301–320.
Grossman, L.I., Wildman, D.E., Schmidt, T.R., Goodman, M., 2004. Accelerated evolution of the electron transport chain in anthropoid primates. Trends Genet. 20, 578–585.
Hamilton, M.B., 2009. Population Genetics. Blackwell Publishing, Chichester, West Sussex, UK.
Harris, D.J., Batista, V., Carretero, M.A., Ferrand, N., 2004a. Genetic variation in *Tarentola mauritanica* (Reptilia: Gekkonidae) across the Strait of Gibraltar derived from mitochondrial and nuclear DNA sequences. Amphibia-Reptilia 25, 451–459.
Harris, D.J., Batista, V., Lymberakis, P., Carretero, M.A., 2004b. Complex estimates of evolutionary relationships in *Tarentola mauritanica* (Reptilia: Gekkonidae) derived from mitochondrial DNA sequence. Mol. Phylogenet. Evol. 30, 855–859.
Harris, D.J., Carretero, M.A., Corti, C., Lo Cascio, P., 2009. Genetic affinities of *Tarentola mauritanica* (Reptilia: Gekkonidae) from Lampedusa and Conigli islet (SW Italy). North-West. J. Zool. 5, 197–205.
Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.
Hey, J., Wakeley, J., 1997. A coalescent estimator of the population recombination rate. Genetics 145, 833–846.
Hudson, R.R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.
Irwin, D.E., Rubtsov, A.S., Panov, E.N., 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). Biol. J. Linn. Soc. 98, 422–438.
Jameson, D., Gibson, A.P., Hudelot, C., Higgs, P.G., 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. 31, 202–206.
Janke, A., Erpenbeck, D., Nilsson, M., Arnason, U., 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B Biol. Sci. 268, 623–631.
Jiggins, F.M., 2003. Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. Genetics 164, 5–12.
Katoh, K., Misawa, K., Kumar, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kelley, L.A., Sternberg, M.J.E., 2009. Protein structure prediction on the web: a case study using the Phyre server. (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) Nat. Protoc. 4, 363–371.
Kreitman, M., 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. 1, 539–559.
Kumazawa, Y., Nishida, M., 1999. Complete mitochondrial DNA sequences of the green turtle and bluetailed mole skink: statistical evidence for archosaurian affinity of turtles. Mol. Biol. Evol. 16, 784–792.
Laslett, D., Canbäck, B., 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–175.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Paulo, O.S., Dias, C., Bruford, M.W., Jordan, W.C., Nichols, R.A., 2001. The persistence of Pliocene populations though the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. Proc. R. Soc. Lond. B Biol. Sci. 268, 1625–1630.
Perera, A., Harris, D.J., 2008. Genetic diversity in the gecko *Tarentola mauritanica* within the Iberian Peninsula. Amphibia-Reptilia 29, 583–588.
Pettersen, E.F., et al., 2004. UCSF Chimera: a visualization system for exploratory research and analysis. (http://www.cgl.ucsf.edu/chimera/docs/credits.html) J. Comput. Chem. 25, 1605–1612.
Rato, C., Carranza, S., Perera, A., Carretero, M.A., Harris, D.J., 2010. Conflicting patterns of nucleotide diversity between mtDNA and nDNA in the Moorish gecko, *Tarentola mauritanica*. Mol. Phylogenet. Evol. 56, 962–971.
Rato, C., Carranza, S., Harris, D.J., 2011. When selection deceives phylogeographic interpretation: the case of the Mediterranean house gecko, *Hemidactylus turcicus* (Linnaeus, 1758). Mol. Phylogenet. Evol. 58, 365–373.
Rato, C., Carranza, S., Harris, D.J., 2012. Evolutionary history of the genus *Tarentola* (Gekkota: Phyllodactylidae) from the Mediterranean Basin, estimated using multilocus sequence data. BMC Evol. Biol. 12. http://dx.doi.org/10.1186/1471-2148-12-14.
Riberon, A., Miaud, C., Grossenbacher, K., Taberlet, P., 2001. Phylogeography of the Alpine salamander, *Salamandra atra* (Salamandridae) and the influence of the Pleistocene climatic oscillations on population divergence. Mol. Ecol. 19, 2555–2560.
Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. 99, 16122–16127.
Slatkin, M., Hudson, R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129, 555–562.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Wright, S.I., Charlesworth, B., 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168, 1071–1076.
Yu, L., Wang, X., Ting, N., Zhang, Y., 2011. Mitogenomic analysis of Chinese snub-nosed monkeys: evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. Mitochondrion 11, 497–503.