The taxonomy of the *Tarentola mauritanica* species complex (Gekkota: Phyllodactylidae): Bayesian species delimitation supports six candidate species

Catarina Rato a,⇑, David James Harris a, Salvador Carranza b, Luís Machado a,b, Ana Perera a

a CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, N.º 7, 4485-661 Vairão, Vila do Conde, Portugal

b Institute of Evolutionary Biology (CSIC-UPF) – Passeig Marítim de la Barceloneta, 37-49, E-08003 Barcelona, Spain

A R T I C L E   I N F O

Article history:
Received 26 March 2015
Revised 29 July 2015
Accepted 8 September 2015
Available online 21 September 2015

Keywords:
*Tarentola mauritanica*
Species tree
Bayesian species delimitation

A B S T R A C T

The lack of morphological diagnosable characters typical of cryptic species, poses a particular problem to taxonomists. This is especially true when taxa are closely related, sharing considerable amounts of ancestral polymorphism. Phylogenetic studies on the Moorish gecko species-complex, *Tarentola mauritanica*, uncovered extremely high levels of mtDNA diversity with six identified clades, including one from the Canary Islands identified as *T. angustimentalis*. Because of the conserved morphology of this species and its paraphyletic status with respect to *T. angustimentalis*, it was suggested that *T. mauritanica* is a cryptic species complex. Nevertheless, none of the nuclear loci used were reciprocally monophyletic regarding the mitochondrial lineages due to retention of ancestral polymorphism. In this study, we added three new intron markers to the already available dataset and used additional tools, namely phylogenetic gene trees, species tree and species limits within a Bayesian coalescent framework to confirm the support of the main lineages. Bayesian clustering analysis supports all six mtDNA lineages as independent groups, despite showing signs of ancestral polymorphism or possibly gene flow between the Maghreb/South Iberia and Central Morocco clades. The species tree recovered two major groups; one clustering taxa from Europe and Northern Maghreb and another one encompassing the lineages from Central/Southern Morocco, Central Morocco and Canary Islands, indicating that the ancestor of *T. angustimentalis* came from the Central/Southern Morocco region. Finally, Bayesian coalescent species delimitation analysis supports all six mitochondrial clades as “unconfirmed candidate species”, pending morphological data to define them.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Species are considered as fundamental units of analysis in many disciplines, including biogeography, ecology, macroevolution and conservation biology and a better understanding of the processes and mechanisms which lead to their establishment, entails that systematists employ methods to delimit species in an accurate and objective way (Sites and Marshall, 2003). However, over the past half century, species delimitation has often been confused with that of species conceptualization itself (de Queiroz, 2007). Recent advances in this field have proposed a unified species theory known as the “General Lineage Concept” (de Queiroz, 1998), rejecting the interpretation of traits, such as reproductive isolation, reciprocal monophyly, ecological distinctiveness and phenotypic diagnosability, as necessary characteristics but suggesting they be treated as lines of evidence relevant for species limits inference. This understanding of species as independent population-level lineages (de Queiroz, 2007; Mayden, 1997) and the “renaissance” of the interest in species delimitation (Sites and Marshall, 2003), has led to the development of a more integrative perspective of taxonomy (Cardoso et al., 2009; Dayrat, 2005; Padial et al., 2010).

Traditionally, species have been identified and described using morphological traits, either typological or quantitative, and these are still considered by some authors as essential lines of evidence (Bauer et al., 2011). Nevertheless, similar selective pressures or severe environmental extremes may mean that morphological characters follow a convergent evolutionary pathway (Nevo, 2001). Furthermore, morphological variation may be a result of short-term adaptation to local conditions (Kaliontzopoulou et al., 2010) or an output of phenotypic plasticity (Zhdanova and...
Zakhariv, 2006). Hence, the use of morphological data alone may underestimate the number of species and, in particular, could fail to identify “cryptic” species (e.g. Barley et al., 2013; Smith et al., 2011).

The routine acquisition of molecular markers has led to a shift from a purely phenotypic approach to the inclusion of molecular characters in the investigation of the evolutionary history of taxa. This change in methodology has led to the discovery of higher levels of differentiation than previously expected using a morphological approach and to a significant increase of cases describing cryptic species complexes (e.g. Camargo et al., 2006; Leavitt et al., 2007; McLeod, 2010; Pinho et al., 2007). This accelerated rate of detection of cryptic species when using molecular markers, suggests that genetic data should be included in alpha taxonomy, whenever possible (Ahmadzadeh et al., 2013). Typically cryptic species are considered “sibling species” (Saez and Lozano, 2005), highlighting the fact that, usually, cryptic species are part of a complex of closely related forms (Bickford et al., 2007). In such cases, biologists are sometimes challenged with poorly resolved phylogenetic relationships and a lack of monophyly inferred from individual genealogies. This scenario can occur due to incomplete lineage sorting or because of past or ongoing gene flow between closely related (but not necessarily sister) taxa. Currently, these gene tree conflicts can be overcome by applying a Bayesian framework, which allows the integration of gene tree uncertainties as well as the incorporation of an explicit model of lineage sorting by the coalescent process model (e.g. BTPP in Yang and Rannala, 2010).

The Moorish gecko, Tarentola mauritanica (Linnaeus, 1758) has a widespread geographic distribution across Europe, some Macaronesian islands and the Maghreb region of North Africa, namely Morocco, Algeria and Tunisia (Bons and Geniez, 1996; Schleich et al., 1996). It is characterized by an extremely elevated mitochondrial DNA genetic variability, with currently six lineages identified (Carranza et al., 2000; Harris et al., 2004a,b; 2009; Perera and Harris, 2008; Rato et al., 2012, 2010). This elevated genetic diversity is particularly evident in North Africa, with Morocco harboring two endemic lineages and sharing most of the remaining ones with the Iberian Peninsula (Harris et al., 2004a,b; Rato et al., 2012). This high mtDNA divergence in T. mauritanica, and the apparent morphological homogeneity of the group, has led to the hypothesis that this corresponds to a cryptic species complex (Harris et al., 2004a,b). Two distinct subspecies, T. m. juliae Joger, 1984 and T. m. pallasii Geniez et al., 1999 have been described from Morocco based on morphology, but these do not correspond to independent genetic lineages (Harris et al., 2004a). Moreover, the paralogy of this species with respect to T. angustimentalis present in the Canary Islands (Carranza et al., 2000; Harris et al., 2004a,b; 2009; Perera and Harris, 2008; Rato et al., 2012, 2010), reinforces the hypothesis that T. mauritanica is, in fact, a complex of several species. However, when nuclear markers have been analyzed, including C-mos, RAG1, MC1R, ACM4 and PDC, these have shown limited variability and consequently, little capacity in resolving some of the phylogenetic relationships (Harris et al., 2004a; Rato et al., 2012, 2010). Most probably, this is due to the fact that all of these taxa are closely related, so that they share a substantial amount of polymorphism (Rato et al., 2012). With the goal of assessing the specific status of each of the six mitochondrial lineages of T. mauritanica, we here add two new molecular markers to the previous dataset from Rato et al. (2012). We estimated the species tree and tested the nuclear markers for the six lineages using a Bayesian clustering analysis and within a coalescent species delimitation framework. The results obtained here will provide the foundation for a future taxonomic revision of T. mauritanica.

2. Material and methods

2.1. Sampling and molecular methods

Genomic DNA was extracted from a total of 60 individuals from the T. mauritanica species-complex, belonging to each of the six previously identified mitochondrial DNA (mtDNA) lineages (Rato et al., 2012) (10 individuals per lineage). For all individuals, we already had available from previous studies the sequences of two mitochondrial genes, 12S rRNA and 16S rRNA (Harris et al., 2004a,b; Perera and Harris, 2008; Rato et al., 2012, 2010) and for some of them, four nuclear protein-coding gene fragments: the acetylcholinergic receptor M4 (ACM4), the melanocortin 1 receptor (MC1R), the phosducin (PDC), and the recombination activating gene 2 (RAG2) (Rato et al., 2012, 2010). We carried out the amplification and sequencing of these four exons for the missing samples, and also obtained amplicons for three introns (bzwl, lztfl1 and frag), which have shown to be quite variable at the intraspecific level in other Gekkota (Fujita et al., 2010; Tonione et al., 2011). For amplification and sequencing of ACM4 the primers tg-F and tg-R published by Gamble et al. (2008) were used. Regarding the MC1R and PDC fragments the primers MC1RF and MC1RR from Pinho et al. (2010) and PHOF2 and PHOR1 (Bauer et al., 2007) respectively, were used for amplification and sequencing. Amplification of ACM4, MC1R and PDC fragments were carried out in 25 μl volumes, containing 5.0 μl of 10 x reaction Buffer, 2.0 mM of MgCl₂, 0.5 mM each dNTP, 0.2 μM each primer, 1 U of Taq DNA polymerase (Invitrogen), and approximately 100 ng of template DNA. Finally, amplification and sequencing of the RAG2 gene fragment was performed using two sets of primers; 31FN.Venk/Lung.460R (amplification) and Lung.35F/Lung.320R (amplification and sequencing) published by Hoegg et al. (2004), PCR conditions were those described in Chiari et al. (2004).

Because the primers described in Fujita et al. (2010) for the amplification of both bwzl and lztfl1 introns were not effective in T. mauritanica, new ones were designed. We used the Ensembl database (http://www.ensembl.org/index.html) transcript information to determine the exon–intron structure in both the chicken and Anolis carolinensis lizard orthologs, and using the sequences of the primers published by Fujita et al. (2010), we designed new primers on exons to amplify the intervening introns. Details are provided in Table 1. Quality of the primers, such as GC percentage, and self-dimerisation was assessed using the software AMPLIFIX v1.6.2 (by Nicolas Jullien; CNRS, Aix-Marseille Université – http://crn2m.univ-mrs.fr/pub/amplifx-dist).

We used the same general PCR protocol for all intron markers. Each PCR had a total volume of 25 μl, using the MyTaq™ Red DNA Polymerase (BIOLINE), with 5 μl of Buffer, 5 mM of each dNTP, 15 mM MgCl₂, 1 U of Taq DNA Polymerase, 0.4 mM of each primer, and 10–25 ng of DNA. The amplification protocol for all PCR reactions of the three introns was the same as described in Fujita et al. (2010). All PCR products were visualized on 2% agarose gels stained with GelRed nucleic acid stain (BIOTIUM) and successful fragment PCR amplification was confirmed by GelRed nucleic acid stain (BIOTIUM) and successful fragment PCR amplification was confirmed by Bioanalyzer 2100 (Agilent). PCR products were purified using the PCR Purification Kit from GeneClean III (Bio101) and sequenced using an ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA) with the same primers as the PCR amplification. Sequence chromatograms were checked and assembled using SeqMan (DNASTAR).}

Phase reconstruction of the nuclear DNA (nDNA) dataset, was performed using the program PHASE v.2.1.1 (Stephens and Donnelly, 2003; Stephens et al., 2001), considering a threshold of 60% (p = q 0.6), since this has been demonstrated to reduce the number of genotype uncertainties with little or no increase in false positives (Garrick et al., 2010). The input files to use in PHASE were created in SeqPHASE (Flot, 2010). All polymorphic sites with a
probability of <0.6 were coded in both alleles with the appropriate IUPAC ambiguity code.

Mega v. 5.0 (Tamura et al., 2011) was used to estimate uncorrected p-distances between lineages. Specimens used in this study, corresponding GenBank accession numbers and localities are provided as Supplementary Material in Table S1.

2.2. Gene trees

In order to determine the best fitting nucleotide model for each gene (mtDNA and nDNA), we used the software jModelTest v2.1.4 (Darriba et al., 2012), under the Akaike Information Criterion (Akaike, 1974). Maximum Likelihood (ML), and Bayesian Inference (BI) phylogenetic analyses were performed for both the concatenated nDNA fragments and for all the loci combined (mtDNA + nDNA). A phylogenetic analysis using only mitochondrial markers for these same specimens was already available (Rato et al., 2012). ML analyses were conducted with the software RAXML v7.2.8 alpha (Stamatakis, 2006) under the GTR+G+Γ model, partitioning the dataset per locus. For all analyses 20 thorough ML searches were performed in order to obtain the ML tree with 1000 bootstrap inferences. A majority rule consensus tree was generated using the software Phyutility (Smith and Dunn, 2008). BI was implemented with the program MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) under a partitioned model (dataset divided into genes), and considering the model of nucleotide substitution estimated with jModelTest. The models selected for the different partitions were: TPM1uf+Γ+Γ for 12S rRNA; GTR+G for 16S rRNA; TIM1+G+Γ for bzw1; TPM1uf+Γ+Γ for Ifabp; TrN+Γ for Iztfl1; TPM1uf for MC1R; K80 for PDC and HKY +Γ for Rag2. The Bayesian posterior probabilities were estimated using a Metropolis-coupled Markov chain Monte Carlo sampling approach, with both runs starting with random trees running for 10 × 10^6 generations, saving one tree every 100 generations to produce 100,000 trees. Both convergence and appropriate sampling were confirmed by examining the standard deviation of the split frequencies between the two simultaneous runs and the Potential Scale Reduction Factor (PSRF) diagnostic. The first 25,000 trees of each run were included in the burn-in period and discarded, and a majority-rule consensus tree was generated from the remaining trees. In all phylogenetic analyses Tarentola chazulaiæ was used as outgroup.

2.3. Population structure

A Bayesian clustering analysis was applied to all nuclear DNA haplotypes using the program STRUCTURE 2.3.4 (Falush et al., 2003; Pritchard et al., 2000). To do this, the multilocus genetic data were converted to haplotype frequencies. The analysis was run considering all possible combinations between the Ancestry and Allele Frequency models’ parameters to compare the results. No priors of individual identification were used. STRUCTURE was run for 10 iterations with a burn-in length of 50,000 steps, followed by 550,000 Markov chain Monte Carlo repetitions, forced to vary between K = 2 and K = 8, the latter corresponding to the number of mtDNA clades identified in Rato et al. (2012) plus two. Determination of the number of clusters (K) within each clade was assessed using the method described by Evanno et al. (2005), implemented in the online software STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

2.4. Species tree inference

A species tree was calculated using a coalescent-based approach as implemented in +BEAST (Heled and Drummond, 2010). We used the groups obtained from the Bayesian clustering analysis as candidate species, and therefore a six species model was applied. This Bayesian assignment test has been shown to provide a species delimitation congruent with integrative taxonomic practices (Miralles and Vences, 2013). Both mitochondrial fragments and the phased nuclear alleles were used in the +BEAST analyses, resulting in nine independent partitions. We performed two independent runs of 20 × 10^7 generations, sampling every 10,000 generations, from which 10% were discarded as burn-in. Models and prior specification applied were as follows [otherwise by default]: both 12S rRNA and bzw1, GTR+G+Γ; both 16S rRNA and Ifabp, GTR+G; ACM4 and Rag2, HK+Γ; Iztfl1, GTR+Γ; and both MC1R and PDC the HKI model; Relaxed Uncorrelated Lognormal Clock (estimate); Yule process of speciation; random starting tree; alpha Uniform (0, 10). Convergence for all model parameters was assessed by examining trace plots and histograms in tracer v1.4 (Rambaut and Drummond, 2007) after obtaining an effective sample size (ESS) >200. Runs were combined using LogCombiner, and maximum credibility trees with divergence time means and 95% highest probability densities (HPDs) were produced using Tree Annotator (both part of the BEAST package). Trees were visualized using FigTree v1.3.1 (available at http://tree.bio.ed.ac.uk/software/figtree).

2.5. Bayesian species delimitation

Bayesian species delimitation was conducted using BP&P v2.2 (Rannala and Yang, 2003; Yang and Rannala, 2010), with the full phased dataset for the seven nuclear loci. The model assumes no admixture following speciation, which is an assumption motivated by the biological species concept and a reversible-jump Markov Chain Monte Carlo (rjMCMC) to estimate the posterior distribution for species delimitation models. Ensuring adequate rjMCMC mixing involves specifying a reversible jump algorithm to achieve dimension matching between species delimitation models with different numbers of parameters (Yang and Rannala, 2010). We

---

Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Ensembl a</th>
<th>Intron</th>
<th>Size (bp)</th>
<th>TA (°C)</th>
<th>Primers (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>bzw1</td>
<td>ENSGALG000000088220</td>
<td>3</td>
<td>490–773</td>
<td>63</td>
<td>bzw1_Tar1: GATGCTTCTGGKGCAAAAACGTGAATTCGC a</td>
</tr>
<tr>
<td></td>
<td>ENSGAC000000016934</td>
<td>4</td>
<td></td>
<td></td>
<td>bzw1_Tar2: AGCAAATGCTCGATGTTGCTGTC b</td>
</tr>
<tr>
<td>Iztfl1</td>
<td>ENSGALG00000011810</td>
<td>2</td>
<td>488</td>
<td>61.6</td>
<td>Iztfl1_Tar1: CAACGCTGCTGCTGTCAAAATAGCAGC a</td>
</tr>
<tr>
<td></td>
<td>ENSGAC000000014164</td>
<td>2</td>
<td></td>
<td></td>
<td>Iztfl1_Tar2: CTGGCCGACAGAAGAAWACATTGGTC c</td>
</tr>
<tr>
<td>LfABP</td>
<td></td>
<td></td>
<td>841</td>
<td>61</td>
<td>n32: AGACCTATGCTAGAGGACATATGACAG a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n33: GGATTTCAATGCTGCCTATATAC a</td>
</tr>
</tbody>
</table>

a Accession number for the gene in chicken (ENSGALG) and Anolis carolinensis (ENSACA).

---

C. Rato et al. / Molecular Phylogenetics and Evolution 94 (2016) 271–278
used algorithm0 with a large fine-tuning parameter $e = 15$, in order to guarantee a good mixing (Yang and Rannala, 2010). Each species delimitation model was assigned equal prior probability (species-modelprior $= 0$). The rjMCMC analysis was run for 200,000 generations with a sampling interval of 1 and a burn-in of 20,000 (10%). This method accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism and a fixed tree topology. The rjMCMC algorithm estimates the posterior distribution for species delimitation models that differ in the number of species, all of which are compatible with the starting guide tree. As guide tree we used the species tree obtained with $+$BEAST to assess the candidate species within the T. mauritanica species-complex. Sequence data included only the nuclear loci, in order to test species limits within the nuclear data on the topology derived from the complete data set. The estimation of the marginal posterior probability of speciation associated with each node in the guide tree (i.e. speciation probabilities as in Leaché and Fujita, 2010) is performed by summarizing the probabilities for all models that support a particular speciation event. Speciation probability values of $\geq 95\%$ are considered as a strong support for a speciation event (Leaché and Fujita, 2010).

The posterior probabilities for models can be mainly affected by the prior distributions on the ancestral population size ($\theta$) and root age ($\tau$), with large values for $\theta$ and small values for $\tau$ favoring conservative models containing fewer species (Yang and Rannala, 2010). Therefore, we conducted the analyses by considering three different combinations of prior settings for these parameters (as in Leaché and Fujita, 2010), which are given as gamma $G(\alpha, ~\beta)$ distributions: relatively large ancestral population size and shallow divergences ($\theta = G[2, 10]$, $\tau = G[2, 2000]$), relatively large ancestral population size and deep divergences ($\theta = G[2, 10]$, $\tau = G[2, 10]$), and relatively small ancestral population size and shallow divergences ($\theta = G[2, 2000]$, $\tau = G[2, 2000]$). The first combination is a conservative one, favoring models containing fewer species. Only speciation events supported by all models were accepted. This cumulative approach, although quite conservative is, in our opinion, the most objective one to interpret the results. All analyses were run twice to confirm consistency between runs.

3. Results

For all 60 individuals of T. mauritanica and T. angustimentalis we obtained 857 bp for the mitochondrial DNA (368 bp of 12S; 489 bp of 16S) and 4193 bp for the concatenated nuclear dataset (405 bp for ACM4; 773 bp for bzw1; 760 bp for lfabp; 419 bp for ltf1f1; 668 bp for MC1R; 392 bp for PDC and 776 bp for Rag2). For the bzw1 intron fragment, some specimens possess a 276 bp indel.

Genetic distances between clades vary between 2.7% and 7.7% for the 12S and 3.4% and 6.5% for the 16S (Table 2). The lower values occur between the Iberian and the European/North African clades for both genes and between Iberian and Central/Southern Morocco for the 16S, whilst the highest genetic divergence for the 12S is observed between Central Morocco and the Maghreb/South Iberia and for the 16S between the Canary Islands and Central Morocco.

3.1. Gene trees

The phylogeny based on the combined nuclear DNA fragments (Fig. 1A), supports as monophyletic the Iberian, Maghreb/South Iberia, the Canary Islands and the Central Morocco groups. The European/North African group is paraphyletic and the Central/Southern Morocco group is also not monophyletic, with two haplotypes more closely related to the remaining members of the phylogeny. On the contrary, the genealogy obtained when all markers are concatenated (Fig. 1B) supports all six groups as monophyletic. Nevertheless, some of the phylogenetic relationships between both analyses are similar, although not highly supported. For both analyses, ML and BI inferred the same topology of the preferred tree.

3.2. Population structure

The obtained K differed according to the combination between ancestry and allele frequency models. For both combinations “Admixture/Frequencies Correlated” and “No Admixture/Frequencies Independent” the best resulting K was K = 6 (Fig. 2). The combination “Admixture/Frequencies Independent” resulted in K = 4 and for “No Admixture/Frequencies Correlated”, the best result was obtained with K = 5. Since two of the models are in agreement regarding the number of clusters (six), we consider K = 6 as the most appropriate. Therefore, the results obtained from the Bayesian clustering analyses support the six clades as independent groups. However, it is noteworthy that two individuals from the Maghreb/South Iberian clade (DB329 and DB5113) share some degree of ancestry with the Central Morocco Clade (52% and 72%, respectively), as well as with the European/North African Clade.

3.3. Species tree and Bayesian species delimitation

The results from STRUCTURE with K = 6 were used to define the “species” for the species tree analysis in +BEAST. The tree inferred with information from both mitochondrial and phased nuclear markers (Fig. 3) recovered a similar topology as observed in Fig. 1B, with higher posterior probabilities. We identified two major groups; one clustering all clades present in Europe and Northern Maghreb and another comprising the clades endemic to Morocco and the Canary Islands. The cladogenesis between the Canary Islands and Central/Southern Morocco was unsupported (53%) and also the split between these two groups and the Central Morocco showed low posterior probability (74%).

Bayesian species delimitation analysis supports the topology of the guide tree when applying both the 95% threshold and the cumulative approach; regardless of the prior distribution for both

| Table 2 | MIDI-norrected p-distances (%) between all clades of T. mauritanica and T. angustimentalis (Canary Islands). Above diagonal are represented the values for the 16S rRNA and below for the 12S rRNA. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Central/Southern Morocco | Iberian Peninsula | Europe/North Africa | Central Morocco | Canary Islands | Maghreb/South Iberia |
| Central/Southern Morocco | 3.4 | 4.6 | 4.6 | 4.9 | 5.1 |
| Iberian Peninsula | 3.5 | 3.4 | 4.4 | 4.9 | 4.0 |
| Europe/North Africa | 3.7 | 2.7 | 5.3 | 4.3 | 5.3 |
| Central Morocco | 7.6 | 7.0 | 5.5 | 6.3 | 6.1 |
| Canary Islands | 5.9 | 5.5 | 4.6 | 7.6 | 5.0 |
| Maghreb/South Iberia | 5.8 | 6.0 | 5.5 | 7.7 | 5.0 |
4. Discussion

The mitochondrial genetic divergence observed between clades is quite variable, ranging from 2.7% to 7.7% for the 12S and 3.4% to 6.5% for the 16S. Some of these values are higher than the ones observed within other geckos that have been suggested to represent species complexes (Stenodactylus arabicus with 12S: 7.7% and 16S: 5.0% in Metallinou et al., 2012; Hemidactylus turcicus with maximum of 7.2% for cytb in Moravec et al., 2011; Ptyodactylus oudrii with 7.4% for the 12S in Perera and Harris, 2010) and lower genetic distances than these, are found between recently described new species of Tarentola from Cape Verde (e.g. divergence between T. gigas and T. rudis of 2.4% for cytb in Vasconcelos et al., 2012). Nevertheless, these yardsticks for setting species boundaries remain highly questionable (Harris and Froufe, 2005) and should be merely descriptive.

The phylogenetic analysis based solely on the concatenated nuclear loci failed to identify some of the mtDNA clades as monophyletic, although they are monophyletic when all data are combined. The lack of monophyly for the nuclear markers was already recorded in another study on the evolutionary history of the genus Tarentola (Rato et al., 2012). Nevertheless, in our study the topology has considerably better resolution but still with two mtDNA lineages being paraphyletic: Europe/North Africa and Central/Southern Morocco. The paraphyly of the Central/Southern Morocco clade is actually due to two individuals from this group that are more closely related to the remaining clades.

Although reciprocal monophyly is one of the main criteria to delimit species (de Queiroz, 2007), recently derived species can be accurately identified long before reciprocal monophyly is achieved after speciation (Knowles and Carstens, 2007). Additionally, analytical methods for delimiting species that typically rely upon genetic distances or gene tree monophyly (Sites and Marshall, 2003, 2004), require subjective decisions concerning the thresholds that demark the species boundary (Hey, 2009). To resolve this, species discovery should be amenable to statistical exploration and Bayesian methods based on coalescence allow this (Fujita et al., 2012; Leaché and Fujita, 2010). Coalescent-based species delimitation methods use probabilistic approaches that do not require reciprocal monophyly of all alleles or fixed differences. This is an important distinction from other non-coalescent methods, since not all molecular markers are expected to be reciprocally monophyletic across lineages, especially if we are testing recent speciation events (Fujita et al., 2012; Hudson and Coyne, 2002). Instead, coalescent-based species delimitation uses multilocus data to test alternative hypotheses of lineage divergence that allow for gene tree discordance under genetic drift (Fujita et al., 2012 and...
references herein). In the particular case of T. mauritanica, although the protein-coding genes alone are not able to differentiate some of the clades (Rato et al., 2012), when analyzed in combination with the introns, BP&P is able to distinguish them by avoiding the idiosyncratic histories of the gene trees, such as incomplete lineage sorting.

We acknowledge that ideally a combination of genetic, morphological and ecological criteria should be taken into account when delimiting species (e.g. Ahmadzadeh et al., 2013; Vasconcelos et al., 2012). Although as far as we know, no morphological or ecological traits have been studied with this objective, ecophysiological results demonstrate that patterns of water loss in T. mauritanica are distinct between the Iberian and the Europe/North African clades (Rato and Carretero, 2015). Therefore, all available data point to significant differences between the identified genetic lineages.

The phylogenetic relationships obtained here for the species tree, are distinct from Rato et al. (2012) but geographically coherent. The clades distributed across Europe and Northern Maghreb are closely related forming one major group, while the lineages endemic to Morocco and the Canary Islands form another. Although the current distribution of both Europe/North Africa and Maghreb/South Iberian clades is the result of a recent dispersal to Europe (Perera and Harris, 2008; Rato et al., 2012), they occur in sympatry in Northern Morocco the same way the Iberian and the Europe/North African clades are in the Iberian Peninsula. A recent study on niche evolution of T. mauritanica, concludes that genetic divergence patterns of this species complex resulted from a mix of both niche divergence and niche conservatism (Rato et al., 2015). Specifically, the Maghreb/South Iberia, the Europe/North Africa and the clade from the Canary Islands show signs of niche divergence with respect to the remaining groups. Thus, results obtained in our study and the ones from Rato et al. (2015) indicate that the clades occurring in Europe and Northern Maghreb regions most probably derived from ecological divergence. According to the species tree genealogy illustrated in Fig. 3, T. angustimentalis from the Canary Islands (clade in pink) is closely related to the Central/Southern Morocco clade (which is also the one geographically closer), indicating that this insular species probably originated from colonizers coming from this region of Morocco.

Results from the Bayesian clustering analysis, estimated a $K = 6$ with a correct assignment of the individuals to their mitochondrial clade. The existence of two individuals from the Maghreb/South Iberian clade sharing ancestry with the Central Morocco clade, indicates that in the past these two taxa have been in contact. Currently they are not known to occur in sympatry and so the observed pattern is presumed to be a case of ancestral polymorphism. On the other hand, the genotype sharing between the Maghreb/South Iberian clade and the Iberian clade, could be due to a recent contact and ongoing gene flow since their geographic ranges overlap in the south of the Iberian Peninsula. Clearly additional sampling across contact zones, and prospecting for potential areas of sympatry in North Africa, will be needed to better distinguish between present and past gene flow.

Bayesian clustering analyses, such as STRUCTURE have been commonly used for detecting species groups and assigning individuals in a variety of studies of species delimitation (Edwards and Knowles, 2014; Leaché and Fujita, 2010; Welton et al., 2013). Although estimating the true value of $K$ may sometimes be difficult, the *ad hoc* statistic $\text{AK}$ implemented in STRUCTURE, has been shown to provide an accurate evaluation on the uppermost hierarchical level of structure (Jakobsson and Rosenberg, 2007) and to properly infer the number of subpopulations, even when genetic differentiation is low (Latch et al., 2006).

Supporting the number of groups detected by the Bayesian cluster analysis are the results derived from the Bayesian species delimitation analyses with maximum node support, regardless of the models for the ancestral population size and root age. All analyses therefore indicate that six unconfirmed candidate species could be recognized within the group, pending a morphological diversity assessment to define them.

5. Conclusions

For several years, the Moorish gecko Tarentola mauritanica has been considered a classic case of cryptic speciation, harboring extremely high levels of mitochondrial diversity and a relatively homogenous morphological variation (Carranza et al., 2000; Harris et al., 2004a,b; 2009; Perera and Harris, 2008; Rato et al., 2012, 2010). However, because the nuclear markers used so far have repeatedly failed to show any variation (Harris et al., 2004a; Rato et al., 2012, 2010), it was never possible to determine the number and distribution of candidate species.

The results obtained from the multilocus phylogenetic gene trees, species tree and analysis of the nuclear markers by Bayesian clustering and species tree delimitation approaches, support six recognized mitochondrial lineages as "unconfirmed candidate species". These results represent a definitive molecular evidence for the taxonomic revision of Tarentola mauritanica. Nevertheless, thorough quantitative multivariate analysis of diagnostic morphological characters and ideally other ecological evidences are still needed in order to approach the revision of this group from an integrative perspective.
Acknowledgments

CR is supported by a FCT (Fundação para a Ciência e Tecnologia) postdoctoral fellowship [SFRH/BDP/92343/2013], LM by a FCT doctoral fellowship [SFRH/BD/89820/2012] and AP by the IF FCT postdoctoral fellowship [SFRH/BPD/92343/2013]. SC is supported by the Systems Research Fund from the Linnean Society of London (to CR), and by the FEDER funds through the Operational Programme for Competitiveness Factors–COMPETE and by National Funds through FCT-Foundation for Science and Technology under the “PTDC/BIA-BEC/105327/2008” and “FCOMP-01-0124-FEDER-008970”.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.09.008.

References


