Phylogeography and ecological niche modeling unravel the evolutionary history of the African green toad, Bufotes boulengeri boulengeri (Amphibia: Bufonidae), through the Quaternary

Violaine Nicolas, Abderrahmane Mataame, Pierre-André Crochet, Philippe Geniez, Soumia Fahd, Annemarie Ohler

Phylogeography and ecological niche modelling unravel the evolutionary history of the African green toad, *Bufotes boulengeri boulengeri* (Amphibia: Bufonidae), through the Quaternary

Violaine NICOLAS[1], Abderrahmane MATAAME[2], Pierre-André CROCHET[3], Philippe GENIEZ[4], Soumia FAHD[5], Annemarie OHLER[1]

1 Institut de Systématique, Evolution, Biodiversité, ISYEB - UMR 7205 - CNRS, MNHN, UPMC, EPHE, Muséum national d'Histoire naturelle, Sorbonne Universités, Paris, France. vnicolas@mnhn.fr, ohler@mnhn.fr

2 Institut Scientifique de Rabat, Rabat, Maroc. abdel_mat@hotmail.fr

3 CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, Montpellier, France. pierre-andre.crochet@cefe.cnrs.fr

4 PSL Research University, CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, Biogéographie et Ecologie des Vertébrés, Montpellier, France. Philippe.GENIEZ@cefe.cnrs.fr

5 Laboratoire "Ecologie, Biodiversité et Environnement", Département de Biologie, Faculté des Sciences de Tétouan, Université Abdelmalek Essaâdi, Tétouan, Maroc. soumiafahd@yahoo.fr

Corresponding author: Violaine NICOLAS

Tel: 00 33 1 40 79 48 52

vnicolas@mnhn.fr

Running title: Evolutionary history of the African green toad

Keywords

Amphibian - Ecological niche modelling - Genetic structure - Maghreb – Pleistocene

Abstract

Recent integration of ecological niche models in phylogeographic studies is improving our understanding of the processes structuring genetic variation across landscapes. Previous studies on the amphibian *B. b. boulengeri* uncovered a surprisingly weak intraspecific differentiation across the Maghreb region. We widely sampled this species from Morocco to Egypt, and sequenced one nuclear and three mitochondrial (mtDNA) genes to determine the level of genetic variability across its geographic range. We evaluated these data with ecological niche modelling to reveal its evolutionary history in response to climate change during the Quaternary. Our results highlight some mtDNA phylogeographic structure within this species, with one haplogroup endemic to coastal Morocco, and one haplogroup widely distributed throughout North Africa. No or little genetic differentiation is observed between isolated populations from the Hoggar Mountains, the Sabha district and the islands of Kerkennah and Lampedusa, compared to others populations. This can be explained by the expansion of the distribution range of *B. b. boulengeri* during glacial periods. This might have facilitated the species' dispersal and subsequent gene flow between most North African localities.

Introduction

Green toads of the *Bufotes viridis* (Laurenti, 1768) species complex are widespread in the Palearctic region where they have differentiated into several lineages (Batista et al. 2006; Stöck et al. 2006; Stöck et al. 2008). Using two mitochondrial genes, Stöck et al. (2006) suggested that at least five species can be recognised within this complex of species, but evidence of reproductive isolation or even distinct nuclear gene pools is lacking for several of these mitochondrial lineages. More detailed studies using mtDNA and two nuclear sequence markers, sex-chromosomes, and additional morphological, phenological, and bioacoustic data focused on the North African-Sicilian lineage (Stöck et al. 2008) demonstrated its evolutionary distinctiveness and documented strong reproductive isolation from the Italian mitochondrial lineage in Sicily. On the contrary, studies of contact zones between the younger lineages *viridis* and *balearicus* (Boettger, 1880) revealed extensive introgression of nuclear markers and lack of intrinsic reproductive isolation (Dufresnes et al. 2014). As a consequence Speybroeck et al. (2016) recognised only the North African-Sicilian lineage as a valid species, *Bufotes boulengeri* (Lataste, 1879), and proposed to postpone further species-level splits until the situation in other contact zones is documented. According to these authors *B. boulengeri*

is subdivided in two subspecies: *B. b. boulengeri* and *B. b. siculus* (Stöck et al. 2008) (availability of the nomen *siculus* is unclear, see Dubois et al. 2013; ICZN 2014).

The African green toad, *Bufotes b. boulengeri*, occurs as fragmented populations in North Africa from the Moroccan Atlantic coast eastwards to the Nile valley in Egypt, including several Mediterranean islands (Martínez-Solano et al. 2015). It is a characteristic inhabitant of relatively arid, open landscapes (Beukema et al. 2013) but it can inhabit a wide range of habitat: forested areas, scrubland, dry grassland, semi-deserts and deserts (Martínez-Solano et al. 2015). Several highly isolated populations are known from Algeria (e.g. Hoggar Mountains), Libya (Sabha district), as well as on the islands of Kerkennah and Lampedusa (Figure 1). In Morocco it is most abundant on the Atlantic coast (as far south as Dakhla) and in the High and Middle Atlas Mountains, with scattered populations in the north, north-east and east of the country and in the northern Saharan margins (Beukema et al. 2013). The species is mostly absent from the Rif Mountains, where it is limited to river corridors and is absent from the mountain cores characterized by a more humid climate (Donaire et al. 2011).

While strong phylogeographic structure is commonly found in amphibian species across the Maghreb region (Ben Hassine et al. 2016; Nicolas et al. 2015; Recuero et al. 2007; Vences et al. 2014), previous studies on *B. b. boulengeri* uncovered a surprisingly weak intraspecific differentiation over its geographic range (Batista et al. 2006; Stöck et al. 2006; Stöck et al. 2008). This is especially unexpected because fossil data indicates that the African green toad has been present in North Africa since the beginning of the Pleistocene (2.5 Ma in Morocco, (Bailon 2000). Climatic reconstructions and paleoenvironmental data show that North Africa regularly shifted from wetter to drier climatic conditions throughout the Quaternary (deMenocal 2004; Schuster et al. 2006; Tabel et al. 2016). This climatic instability led to periodic modifications of habitats, such as contraction/ expansion of the Sahara desert, emergence of more open habitats or deep modifications of hydrographic networks (Rognon 1989; Le Houérou 1997; Drake et al. 2011). Several geographical features, such as the Sahara desert, the Atlas Mountains or the Moulouya River valley are known to have acted as barrier to dispersal for many species and promoted diversification in several species or species complexes (Aulagnier and Thevenot 1986; Alvarez et al. 2000; Brown et al. 2002; Ben Faleh et al. 2012; Douady et al. 2003; Escoriza et al. 2006; Fritz et al. 2006; Barata et al. 2008; Nicolas et al. 2014).

In this study, we widely sampled African green toads from Morocco to Egypt (95 newly collected individuals in Morocco, Algeria and Egypt, and 39 specimens from Algeria, Egypt, Libya, Morocco, Tunisia and Lampedusa Island retrieved from GenBank) and sequenced

three mitochondrial (mtDNA) and one nuclear (nDNA) genes to determine the level of genetic variability within *B. b. boulengeri* across its geographic range. We specifically tested whether isolated populations from the Hoggar Mountains, the Sabha district and the islands of Kerkennah and Lampedusa are genetically differentiated from others populations. Moreover, we evaluated these genetic data with ecological niche modelling based on bioclimatic data to reveal the evolutionary history of *B. b. boulengeri* in response to climate change during the Quaternary.

Material and methods

DNA extraction, amplification and sequencing

Ninety-five newly collected individuals were included in this study (15 specimens from Algeria, two from Egypt and 78 from Morocco; Supporting information Table S1). When available, these specimens were deposited either at the Museum National d'Histoire Naturelle (MNHN, Paris, France) or at the collection of the UMR5175 CEFE-EPHE (Collection BEV, Montpellier, France); for several specimens only tissues samples were available. Genomic DNA was extracted from liver, spleen, muscle or toes preserved in 95% ethanol using the NucleoSpin Tissue Core kit (Macherey-Nagel, Hoerdt, France). The mitochondrial D-loop region (D-loop), 16S rRNA gene (16S; mtDNA), cytochrome c oxidase subunit 1 gene (CO1; mtDNA) and recombination activating protein 1 gene (RAG1; nDNA) were amplified using polymerase chain reaction (PCR) primers ControlB-H and CytbA-L (Goebel et al. 1999), 16Sar-L and 16Sbr-H (Palumbi et al. 1991), Amp-P3F and Amp-P3R (San Mauro et al. 2004) and MartFL1 and AmpRI (Chiari et al. 2004), respectively (see Supporting information Table S2 for primer sequences). The double-stranded PCR products were purified and sequenced at the Genoscope (Ivry-sur-Seine, France) or at EUROFINS (Ebersberg, France) using the Sanger dideoxy sequencing run on an ABI 3730XL sequencer (Applied Biosystems, Foster City, CA, USA). All newly determined haplotypes were submitted to GenBank (accession numbers KY498904 to KY499005).

In our analyses, we also included 39 specimens of *B. b. boulengeri* available in the GenBank database for which the D-loop, 16S and/or the CO1 genes were sequenced.

The combination of these two datasets (newly sequenced individuals and Genbank data) allowed us to include specimens from six countries and 40 localities, covering most of the geographical range of the species (Figure 1). A total of 122, 113, 88 and 35 individuals were included in our analyses for the D-loop, 16S, CO1 and RAG1 DNA regions, respectively (Table 1).

Median-joining network analyses

Relationships between haplotypes were inferred by constructing a network using the median-joining method available in NETWORK v4.500 (Bandelt et al. 1999). This method accounts for the coexistence of ancestral and descendent haplotypes, multifurcations and reticulate relationships (Posada and Crandall 2001b). We used the maximum parsimony post-processing option, which removes all superfluous median vectors and links that are not contained in the shortest trees of the network. Sequences of 545, 586, 806 and 397 bp were retained for the haplotype network analyses, for the D-loop, 16S, CO1 and RAG1 genes, respectively, to minimize the number of incomplete sequences; as adding ambiguous data in median-joining trees is problematic. For the same reason, we treated all genes separately in haplotype network analyses as sequences of one or two genes were missing for several specimens (Table 1 and Supporting information Table S1).

Alignment of 16S gene sequences required the insertion of three gaps of length one, and alignment of D-loop sequences required the insertion of 11 gaps: eight of length one and one of length three. Indels were taken into account in phylogenetic analyses, but were removed from genetic diversity and demographic history analyses.

For the nuclear RAG1 gene, the existence of heterozygous positions was investigated and an input file constructed from this information using SeqPHASE (Flot 2010). The phase of each haplotype and its reconstruction were carried out using PHASE v.2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005) by running the formerly built input file and by considering the default parameters of the software. Phased haplotypes were used in all subsequent analyses. The phasing of all specimens was straightforward (probability of 1.0), except for three of them (probability below 0.6). These three specimens were therefore removed from subsequent analyses (Supporting information Table S1).

Genetic diversity, population differentiation and demographic history

The average number of nucleotide differences, nucleotide diversity and haplotype diversity (Nei 1987) were calculated using DNASP 5.10 (Librado and Rozas 2009).

Inferences of population expansion were made using two different methods. First we calculated the Fs statistic of Fu (Fu 1997), which is based on the probability of having a number of haplotypes greater or equal to the observed number of samples drawn from a constant-sized population. Secondly we calculated the R2 statistic (Ramos-Onsins and Rozas 2002), which is based on the difference between the number of singleton mutations and the

average number of nucleotide differences. Ramos-Onsins and Rozas (2002) demonstrated that these statistics have the greatest power to detect population expansion for non-recombining regions of the genome under a variety of different circumstances, especially when population sample sizes are large or when sample sizes are small. The significance of Fu's Fs and R2 was obtained by examining the null distribution of 1,000 coalescent simulations of these statistics using DNASP 5.10. Significantly large negative Fu's Fs values and significantly positive R2 values were taken as evidence of a population expansion. The Fs statistic was considered significant when the p-value was below 0.02. We used Spatial analysis of Shared Alleles (SaShA) to test whether there is any genetic subdivision across the geographical space. This method can detect subtle genetic differentiation even when diversity is moderate, gene flow is moderate or high, and when sampling is poor –that is low sample size, or samples of uneven sizes and taken at uneven spatial intervals (Kelly et al. 2010). This analysis uses spatial and allele (haplotype) information to detect non-random allele distribution against an expectation of panmixia. The test statistic OM describes the observed mean distance between alleles. Where OM is less than the expected mean (EM), alleles are considered to be aggregated (underdistributed). Where OM is larger than EM, alleles are considered overdispersed (panmictic). A jackknifing procedure allows the identification of which alleles contribute to the distribution.

Phylogenetic analysis

To better understand phylogenetic relationships between the two haplogroups identified in the median-joining network analyses we also performed phylogenetic reconstruction on the 16S and D-loop sequences combined using Bayesian Inference (BI). Only specimens for which the two genes were available were included in this analysis. CO1 sequences were not included because we had no data for Tunisia and Egypt. We used MrBayes v. 3.2.5 (Ronquist & Huelsenbeck 2003). Analyses were conducted with two runs and four chains (three hot and one cold), each of 5,000,000 generations, with trees sampled every 1,000 generations. Stationarity was assessed by examining the average SD of split frequencies and the potential scale reduction factor. After a visual inspection, the first 25% of iterations was excluded as burn-in time and the resulting trees were combined in a majority rule consensus tree to obtain posterior probabilities. We used a two partition model (one for each gene). Each gene data set was tested for the most appropriate model of sequence evolution using the Akaike information criterion (Akaike 1973) test as implemented in MrModeltest v. 3.7 (Posada & Crandall 2001a) using MrMTgui (available from http://genedrift.org/software/mrmtgui.html).

The HKY + I + G and HKY + I models were chosen for the D-loop and 16S, respectively. To root our tree, we included four sequences of *B. b. siculus* (GenBank EU497543, EU497553, EU497538 and EU497548 for D-loop; EU497443, EU497453, EU497438 and EU497448 for 16S), the sister clade of *B. b. boulengeri*, as well as three most distantly related outgroups: *B. viridis* (EU497516 and EU497416), *B. balearicus* (EU497577 and EU497477) and *B. variabilis* (FJ882812 and DQ629623) (Dufresnes et al. 2014; Stöck et al. 2008).

Divergence time estimates

Obtaining robust divergence time estimates is not a trivial task (Ho et al. 2014). In the absence of a good dated fossil record for the genus *Bufotes* we used two approaches to estimate divergence times: we used the substitution rates derived from the paper of Stöck et al. (2008) and a secondary calibration point derived from the work of Garcia-Porta et al. (2012).

In these analyses, we only took into account the D-loop and 16S results as they are the only genes for which background data on substitution rates and DNA sequences of sister species were available. Divergence time estimates were inferred using BEAST version 1.4.7 (Drummond and Rambaut 2007). We used an uncorrelated lognormal Bayesian relaxed clock. To account for the fact that our trees mostly describe intraspecific relationships, we used a coalescent model tree prior with a Bayesian skyline model of population size. The computer program MRMODELTEST (Nylander 2004) was used to evaluate the fit of 24 nested models of nucleotide substitution to the data. The model chosen by MRMODELTEST according to the Akaike information criterion was then used in BEAST. Two distinct runs were carried out for each gene separately, each one with four independent chains of 60,000,000 generations, default priors, and trees sampled every 6,000 generations. Results were visually inspected using TRACER version 1.5 to ensure proper mixing of the Markov chain Monte Carlo. We applied a conservative burn-in of 25%. Samples from both runs were combined using the software LOGCOMBINER version 1.4.7, and a consensus chronogram was obtained using the program TREEANNOTATOR version 1.6.2 using the options maximum clade credibility tree and median heights. To estimate divergence times we used to approaches. First, we calibrated our tree using the same range of substitution rates as Stöck et al. (2008), which were derived from several general works on amphibians: we specified the prior for the mean substitution rate as a normal distribution, with a mean of 0.02 and a standard deviation of 0.007 substitution per site per My for the D-loop, and a mean of 0.008 and a standard deviation of 0.003 substitution per site per My for the 16S gene. To root our trees, we

included four sequences of *B. b. siculus* (GenBank EU497543, EU497553, EU497538 and EU497548 for D-loop; EU497443, EU497453, EU497438 and EU497448 for 16S), the sister clade of *B. b. boulengeri* (Dufresnes et al. 2014; Stöck et al. 2008). The model chosen by MRMODELTEST was HKY+I+G for the D-loop region, and HKY+I for the 16S gene. Second, we used a secondary calibration point derived from the work of Garcia-Porta et al. (2012). These authors estimated the separation between the Iberian and the African populations of *Bufo bufo* (Linnaeus, 1758) at 3.07 Ma (95% HPD = 1.91-4.36 Ma), based on four calibration points derived from geological and fossil evidences. We used a normal prior with a mean of 3.07 and a standard deviation of 0.70. The model chosen by MRMODELTEST for the 16S gene was GTR+G. We included in these analyses all our 16S sequences as well as sequences of *B. bufo* from Tunisia (Genbank JQ348746, JQ348749), Morocco (JQ348755, JQ348750) and Iberia (JQ348689, JQ348710, JQ348725).

Species distribution modeling

We modeled the climatic niche of *B. b. boulengeri* to approximate its current, Last Glacial Maximum (LGM) and mid-Holocene distributions. We applied ecological niche modeling methods, where environmental data are extracted from current occurrence records, and habitat suitability is evaluated across the landscape using program-specific algorithms (Elith et al. 2006). The present-day models were then projected on the climatic reconstructions of the LGM (a cold and dry period, about 22,000 years ago) and mid-Holocene (a wetter period, about 6,000 years ago) under the assumption that the climatic niche of the species remained conserved through time (Elith et al. 2010).

Distribution records of African green toads were assembled from literature (Bons and Geniez 1996; Frynta et al. 2000; Baha El Din 2006; Stöck et al. 2006; Fahd and Mediani 2007; Harris et al. 2008; Ibrahim 2008; Stöck et al. 2008; Brunet et al. 2009; Sicilia et al. 2009; García-Muñoz et al. 2010; Amor et al. 2011; Donaire et al. 2011; Filippi et al. 2011; Beukema et al. 2013;Ibrahim 2013; Damas-Moreira et al. 2014; Márquez-Rodríguez 2014; Essghaier et al. 2015; Rosado et al. 2016), museum collections (Global Biodiversity Information Facility, http://www.gbif.org, last access date: 26 August 2016) and our own unpublished databases (Supporting information Table S3). To correct sampling bias, a subsample of records regularly distributed in the geographical space was selected using ENMtools ver. 1.3 (Warren et al. 2010). This method is the most efficient at correcting sampling bias (Fourcade et al. 2014). The resolution of the reference grid was set to 2.5 arc min (nearly 5 × 5 km), that is the same as that used for the environmental layers. This coarse resolution was chosen to better

match the coordinate uncertainties associated with georeferenced, textual localities of museum specimens. After filtering, the final dataset used for modelling consisted of 702 distribution records (Supporting information Figure S1).

As environmental layers, we used available climatic data from the WorldClim database (Hijmans et al. 2005). We removed from our analyses variables BIO2, BIO3, BIO5, BIO7 and BIO15 because those variables show a high level of discrepancy between General Circulation Models (GCMs) for North Africa (Varela et al. 2015). Colinearity of the initial variables was measured by Pearson's correlation coefficient in ENMtools v1.3 (Warren et al. 2010). A total of five variables, all of which had a Pearson correlation coefficient lower than 0.75, were retained. The final set of environmental predictor variables used for the species distribution models (SDM) consisted of: Annual Temperature (BIO1), Precipitation of Wettest Quarter (BIO16), Precipitation of Driest Quarter (BIO17), Precipitation of Warmest Quarter (BIO18) and Precipitation of Coldest Quarter (BIO19).

Climatic variables were used for present conditions and for the LGM and mid-Holocene. Paleoclimate data were drawn from the general circulation model (GCM) simulations based on two climate models: the Community Climate System Model (CCSM, version 3) (Collins et al. 2006) and the Model for Interdisciplinary Research on Climate (MIROC, version 3.2) (Hasumi and Emori 2004). For a discussion of the uncertainties associated with the climatic data, see Schorr et al. (2012) and Varela et al. (2015). The use of these two different climate models enabled us to assess and account for modeling uncertainty due to LGM climate data.

To predict the potential distribution of the species in current conditions and in the LGM, we used MAXENT ver. 3.3.3 (Phillips et al. 2006), which returns a model with relative occurrence probability of a species within the grid cells of the study area. To ensure consistency of model predictions among repeated runs, we performed a 10-fold cross-validation with random seed. Several values of the regularization multiplier were tested, as suggested by Radosavljevic & Anderson (2014), and a final value of 2 was retained in our final analyses. To determine whether the predictions for current conditions generated by Maxent were better than random predictions, we used the area under the receiver operating characteristic curve (AUC), a commonly used measurement for comparison of model performance (Elith et al. 2006). The AUC ranges from 0 to 1, with greater scores indicating better discrimination ability. An AUC greater than 0.5 indicates that the model discriminates better than random.

Results

Molecular genetic analyses

The D-loop region is clearly the most variable of the four analyzed DNA regions, with a higher number of polymorphic sites, a higher number of haplotypes, and higher haplotype and nucleotide diversities (Table 2). The RAG1 nDNA fragment is the less variable one.

The D-loop network allows the identification of two main haplogroups (Figure 2). Haplogroup 1 groups 44 specimens (10 haplotypes) coming from coastal Morocco, from Aglou (locality 37) to Merja Zerga (locality 26; Figure 3). Haplogroup 2 is widely distributed throughout North Africa and groups 78 specimens (24 haplotypes) coming from Morocco to Egypt. Haplogroup 1 can be subdivided into two sub-haplogroups (1a and 1b) having sympatric geographic distributions in Morocco (Supporting information Table S1). Results of the 16S and CO1 networks are congruent with those obtained with the D-loop network, and confirm the presence of one haplogroup restricted to Morocco, and one haplogroup widely distributed throughout North Africa. No sub-structure is evident within the Moroccan haplogroup.

The RAG1 gene recovers only 13 haplotypes, differing from one another by one to five mutations. No phylogeographic structure is evident in the network: the three most common haplotypes differ from one another by one to three mutations and are geographically widely distributed. Haplotype h02 is present in the Algerian and Moroccan localities 17, 27, 28, 29, 34, 35, 38 and 39. Haplotype h01 is present in the Moroccan localities 27, 28, 29, 34 and 35, and haplotype h03 are present in the Moroccan localities 28, 33, 35 and 38.

For the D-loop region, the SAShA analysis of all specimens reveals that *B. boulengeri* haplotypes are significantly underdistributed (OM = 213 km, EM = 1285 km; p-value = p-value < 0.001). This is also true for haplogroups 1a (OM = 125 km, EM = 185 km; p-value = 0.002) and 2 (OM = 250 km, EM = 1495 km; p-value < 0.001), but not for haplogroup 1b which seems evenly distributed (OM = 208 km, EM = 201 km; p-value = 0.748). Our Jacknife SAShA analyses indicate the robustness of the overall results, which remain qualitatively the same when any one haplotype is removed. Within haplogroup 1a the most common haplotype (16 individuals) is restricted to localities situated in-between Sidi Mokhtar (locality 35 on Figure 1) and Berrechid (locality 28). Within haplogroup 2 the most common haplotype (22 individuals) is restricted to the Moroccan localities situated in-between the localities of Sidi Mokhtar (locality 35), Merja Zerga (locality 26) and Al Baten (locality 25). Only one haplotype is present in the Hoggar Mountains (10 individuals) and the Lampedusa Island (6

individuals), respectively, and each of these three haplotypes is endemic to each of these regions. One haplotype is endemic to Kerkennah Islands (3 individuals). One haplotype (5 individuals) in found on the Mediterranean coast from Cap Bon (Tunisia, locality12) to Sahahhat (Libya, locality 6). Finally one haplotype (3 individuals) is endemic to the three southernmost Moroccan localities of Tantan (locality 40), Abtith (locality 38) and Saqiyat al-Hamra (locality 39). In Sasha analyses only haplotypes found in at least three individuals are considered. We would like to underline that one haplotype represented by two individuals in our dataset was found in both Kekennah islands (locality 11) and Wadi El Natrum (locality 2). For both the 16S and CO1 genes, SASha analyses of all specimens reveals that *B. b. boulengeri* haplotypes are significantly underdistributed (OM = 266 km, EM = 996 km for 16S; OM = 264 km, EM = 756 km for CO1; p-values < 0.001). This is also true for haplotypes of haplogroup 2 (OM = 352 km, EM = 1254 km for 16S; OM = 345 km, EM = 1099 km for CO1; p-values < 0.001); while haplotypes of haplogroup 1 are evenly distributed (OM = 180 km, EM = 190 km for 16S; OM = 191 km, EM = 183 km for CO1; p-values = 0.221 and 0.433, respectively).

For the 16S gene, the most common haplotype within haplogroup 2 (23 individuals) is restricted to the Moroccan localities situated in-between the localities of Abtith (locality 38), Merja Zerga (locality 26) and Al Baten (locality 25). One haplotype (16 individuals) is endemic to the Hoggar Mountains (localities 17 to 22) and Al Baten (locality 25), and it is the only haplotype found in the Hoggar Mountains. Two haplotypes (8 and 5 individuals, respectively) are present in both coastal Tunisia (localities 9 or 12) and Kerkennah Island (locality 11). One haplotype (5 individuals) is endemic to Lampedusa Island (locality 15) and it is the only haplotype found in this island.

For the CO1 gene, the most common haplotype (20 individuals) within haplogroup 2 is restricted to the coastal Moroccan localities situated in-between the localities of Abtith (locality 38) and Merja Zerga (locality 26). One haplotype (15 individuals) is endemic to the Hoggar Mountains (localities 17 to 22) and Al Baten (locality 25), and it is the only haplotype found in the Hoggar Mountains. One haplotype (4 individuals) is evenly distributed (present in localities 25 and 26).

Negative values of Fu's Fs were obtained for all genes and all haplogroups but they were only significant for haplogroup 1 and the 16S gene, and for haplogroup 2 and the D-loop region (Table 2). It was almost significant P = 0.03) for haplogroup 1a and the D-loop region.

Positive values of R2 were obtained for all genes and all haplogroups, but they were only significant for haplogroup 1a and the D-loop region.

Visual examination of network characteristics (i.e. a star-like haplotype network where many new haplotypes are derived from one dominant haplotype via single of few mutation steps, indicating that they are the product of recent mutation events) indicate population expansion in haplogroup 1a for the D-loop region, and in haplogroup 1 for the 16S and CO1 genes. The central haplotype in the network, which could be interpreted as ancestral, is widely distributed and gives no information about the geographical origin of the population expansion.

To identify the center of a population expansion it can be interesting to compare haplotype and nucleotide diversity between populations. However this type of analysis cannot be done here due to low sample size per population.

Our phylogenetic tree, based on 16S and D-loop sequences (Figure 4), shows that haplogroup 1 is monophyletic (posterior probability = 1.00). The monophyly of haplogroup 2 is not confirmed (posterior probability = 0.69). All specimens from Algeria (in blue) cluster together in the tree, as well as all specimens from Lampedusa Island (in purple). Specimens from Tunisia (in red) cluster in two distinct groups, and the two specimens from Egypt (in green) are phylogenetically nested within one of these groups.

Based on our BEAST analyses and the substitution rates of Stöck et al. (2008), the TMRCA of all *B. boulengeri* is estimated at ca 0.773 Mya (range 0.210–1.787 Mya) and 0.601 Mya (0.128–1.716) according to our D-loop and 16S results, respectively. Based the separation between the Iberian and the African populations of *Bufo bufo* at ca 3.07 Ma, the TMRCA of all *B. boulengeri* is estimated at ca 1.140 Mya (range 0.242–2.455 Mya) with the 16S sequences.

Species distribution modelling

To study the relationships between genetic diversity and possible glacial refugia, we built a species distribution model (Figure 5) for *B. boulengeri* based on the known-presence localities of this species and bioclimatic layers. The AUC value is 0.980, which is considered to correspond to a useful predictive model. Under present-day climatic conditions the model reveals high climatic habitat suitability for this species across most coastal Morocco and Algeria (in-between the Atlantic Ocean or Mediterranean sea and the Atlas Mountains), from

the Souss valley to Alger, across coastal Tunisia, and across coastal Libya from the western border to Misourata, and from Adjedabia to Derna. High climatic habitat suitability also exists for this species south of the Atlas Mountains, at the limit with the Sahara desert, from Oued Guir in Morocco to Biskra in Algeria. Finally high climatic habitat suitability is recorded in the Hoggar Mountains.

According to the CCSM model, climatically suitable habitat for *B. boulengeri* was patchily distributed during the LGM on the Atlantic coast of Morocco, on the coast at the Moroccan-Algerian border, in coastal Tunisia, south of the Atlas Mountains, in the Hoggar Mountains, in south-east Libya and in most of Egypt, except the coast and the south-east. The MIROC model predicted larger areas of climatically suitable habitat than the CCSM model for *B. boulengeri* during the LGM in the south, with an almost continuous high habitat suitability range from the south of the Atlas Mountains to the Hoggar Mountains and in most Libya and Egypt.

For the Mid-Holocene the CCSM and MIROC models give similar results. The distribution of climatically suitable habitat for *B. boulengeri* at this period is predicted to be more or less similar to present-day conditions, except lower habitat suitability in the Anti-Atlas, Atlas and Middle Atlas mountains, in the Rif Mountains, and east of the Moulouya River in Morocco.

Discussion

Low phylogeographic structure across B. b. boulengeri *geographic range*

Our genetic results are mostly congruent with previous studies (Batista et al. 2006; Stöck et al. 2006; Stöck et al. 2008) in showing low genetic diversity in *B. b. boulengeri* across North Africa. However our more extensive sampling scheme, in terms of number of specimens and localities, allowed us to highlight some mtDNA phylogeographic structure within this species: one haplogroup (1) is endemic to coastal Morocco, from Aglou to Merja Zerga (Figure 3), and another haplogroup (2) is widely distributed throughout North Africa, from coastal Morocco to Egypt. Moreover, within haplogroup 2, several haplotypes are significantly underdistributed. Low genetic diversity values and no phylogeographic structure were obtained for the nDNA gene (RAG1). These results suggest that the mitochondrial phylogeographic structure of the species has a recent origin.

This low mitochondrial structure at continental scale contrasts with the east-west divergence that has been commonly found in previous phylogeographic studies of North African organisms (see Nicolas et al. 2015 for a series of examples). The African green toad is a relatively large amphibian that is supposed to travel considerable distances during its 4–5 year

lifespan, like its congener *B. viridis*, allowing considerable gene flow between populations (Langton 1990). Low genetic variation across North Africa was also observed for *Amietophrynus mauritanicus* (Schlegel, 1841), another toad species with high vagility and wide ecological niche preferences (Harris and Perera 2009). We thus hypothesise that a combination of high vagility and continuously suitable habitats across North Africa across glacial and interglacial periods (see Fig. 4) explains the unusual lack of mitochondrial divergence between eastern and western North Africa in *B. boulengeri*.

Given the current distribution of the species (Figure 1) one could however expect to find genetically divergent populations in the islands of Kerkennah and Lampedusa, the Hoggar Mountains and the Sabha district. In agreement with this hypothesis we found haplotypes endemic to Lampedusa Island (locality 15) and the Kerkennah Islands (locality 11) with the mtDNA markers. However these haplotypes differ from mainland Africa haplotypes by few mutations (1 mutation for the closest haplotype), and one haplotype was found in common in both Kerkennah Islands and Egypt (locality 2). With the fastest evolving mtDNA gene we found only one haplotype in the Hoggar Mountains, and it is endemic to these mountains. This haplotype differ by two mutations from the closest haplotype (present in locality 6). With the 16S and Co1 genes we found only one haplotype in the Hoggar Mountains, but this haplotype was also present in the locality of Al Baten (locality 25). No haplotype endemic to Sabha district was recorded in our analyses. Hence, no or little genetic differentiation was observed for populations from the islands of Kerkennah and Lampedusa, the Hoggar Mountains or the Sabha district. The islands of Kerkennah and Lampedusa are separated from mainland Tunisia by low-depth waters (maximum depth around -70m according to Google Earth) and were thus connected to mainland North Africa as recently as the LGM when sea levels were around 120m lower than today (Rohling et al. 1998, Ray and Adams 2001). Lack of divergence from mainland populations is thus not unexpected. Lampedusa and other islands around Sicily have been colonized by one amphibian (*B. boulengeri*) and several reptiles (see Corti et al. 1997), and phylogeographic studies suggest that a mix of natural range expansions and human-mediated colonization events explain the current distribution of amphibians and reptiles in these islands (Corti et al. 1997; Stöck et al. 2016).

Ecological niche modelling help understanding the observed phylogeographical pattern

Using different calibration priors resulted in slightly different divergence time estimates, but our mtDNA data indicate that all intraspecific divergent events within the African Green toad probably occurred during the last My. The Quaternary Period was dominated by Ice Ages and

involved repeated global cooling and increasing of the Arctic and Antarctic ice sheets (Hewitt 2004), and North Africa regularly shifted from wetter to drier climatic conditions (deMenocal 2004; Schuster et al. 2006; Tabel et al. 2016). In the absence of available climatic data for oldest glacial periods, we used the LGM conditions as a proxy of the paleo-climatic conditions of glacial periods. However it should be stressed that the intensity of glaciations was not always the same from one glacial cycle to another and that even for the LGM several uncertainties are associated with the climate data. One major point is the reliability of global circulation models for the LGM climate (Schorr et al. 2012). In our data, uncertainty resulting from climate model is evident from the LGM species distribution obtained with the CCSM versus MIROC models (Figure 5). However, despite these drawbacks, results of our climatic niche modelling analyses help to understand the observed phylogeographic pattern.

Our results for the LGM with the MIROC model suggest an expansion of the distribution range of *B. b. boulengeri* during glacial periods. This might have facilitated the species' dispersal and subsequent gene flow between most North African localities, explaining the quasi absence of genetic distinctiveness of Hoggar Mountains or Sabha district populations compared to other North African populations.

In Coastal Morocco no haplotype was found to be geographically underdistributed, which is congruent with the high climatic habitat suitability modeled for the African green toad in this area throughout the Quaternary.

We found one haplotype endemic to coastal Tunisia (localities 9, 11 and 12) and coastal Libya (locality 6). This pattern is easily explain by the actual and mid-Holocene modeled distributions of the species; highly climatically suitable habitat with no significant climatic barrier being present in this area at these periods.

Origin of the two mtDNA haplogroups

Our mtDNA sequencing results identify two haplogroups within *B. b. boulengeri*: one endemic to coastal Morocco, from Aglou to Merja Zerga (Figure 3), and another one widely distributed throughout North Africa, from coastal Morocco to Egypt. Thus the geographical range of the first haplogroup is embedded in the geographical range of the second haplogroup. Two hypotheses can be proposed to explain this pattern: i) vicariance of an eastern and western haplogroup followed by secondary expansion of the eastern haplogroup westward or ii) retention of deep mtDNA diversity in Morocco followed by loss of diversity during eastward expansion of the species out of Morocco.

i) Vicariance, probably around 1.1–0.7 Ma, could be linked to the late Early Pleistocene global climate transition, and particularly to MIS 22. MIS 22 (at ~0.87Ma) within the late Early Pleistocene global climate transition represents the first prominent cold stage of the Pleistocene. A burst of dust production and therefore aridity occurred at that time, and it corresponds to a period of important vegetation changes (forest withdrawal and onset of steppe or open vegetation), faunal turnover for mammals in North Africa and migration pulses of large herbivores and hominins from North Africa and Eastern Europe into Southern European refugia (Muttoni et al. 2010; Stoetzel 2013). At this period *B. b. boulengeri* may have survived in two distinct allopatric areas, leading to allopatric diversification. The first haplogroup would then have remained geographically restricted to the Atlantic Coast of Morocco (between Aglou and the Merja Zerga), its dispersion eastward and southward being limited by the Rif and Atlas Mountains, respectively. The second haplogroup would have spread throughout large parts of North Africa as the result of broad climatic suitability for the species South of the Atlas mountains and the Saharan Atlas during most of the Quaternary, and particularly during glacial maxima (according to the MIROC model).

ii) Alternatively, deep genetic diversity could have persisted in Morocco thanks to the continuous occurrence of climatically favorable conditions for the species and the complex topography of the country. A recent population expansion eastward throughout large parts of North Africa, favored by the broad climatic suitability for the species south of the Atlas mountains and the Saharan Atlas during most of the Quaternary, would have resulted in a loss of genetic diversity, with only the haplogroup 2 retained in the area of expansion.

Note that both scenarios imply a population expansion of haplogroup 2, for which we find no evidence in the tests we performed (see results), but the sparse sampling of haplogroup 2 outside Morocco might limit the power of the tests. The topology of our phylogenetic tree, based on 16S and D-loop sequences (Figure 4), supports the hypothesis of retention of deep mtDNA diversity in Morocco followed by loss of diversity during eastward expansion of the species out of Morocco. Haplogroup 2 does not form a strongly supported clade in our phylogenetic tree. All specimens from Algeria cluster together in the tree, as well as all specimens from Lampedusa Island. Specimens from Tunisia cluster in two distinct groups, and the two specimens from Egypt are phylogenetically nested within one of these groups. These results suggest that several waves of eastward expansion occurred.

To conclude, our integrative approach, combining phylogeography and climatic niche modelling, allowed us to propose a plausible hypothesis explaining the low phylogeographic structure observed within *B. b. boulengeri* across the Maghreb region.

Acknowledgments

This work has been funded by the ANR project PEX04-MOHMIE. Molecular analyses were supported by the 'Action Transversale du Museum: Taxonomie moléculaire, DNA Barcode & Gestion Durable des Collections', the Service de Systématique Moléculaire of the French National Museum of Natural History (UMS 2700, Paris, France) and the network 'Bibliothèque du Vivant' funded by the CNRS, the Muséum National d'Histoire Naturelle, the INRA, and the CEA (Genoscope). DNA sampling in Morocco was permitted thanks to the 'Haut Commissariat aux Eaux et Forêts et à la Lutte Contre la Désertification (authorizations 15 HCEFLCD/DLCDPN/DPRN/CFF)' in the frame of a joint collaboration between the Institut Scientifique of Rabat and the Muséum National d'Histoire Naturelle. We are grateful to the people who sent us tissue samples and/or field observations, particularly Abdelkhader Allali, Benjamin Allegrini, Menad Beddek, Wouter Beukema, Pierre Beaubrun, Jacques Bons, Fabrice Cuzin, Boualem Dellaoui, Rémi Destre, Bertrand Delprat, Guy Durand, Mohamed El-Ghazrani, Pascal Escudié, Thomas Gallix, Michel Geniez, Yves Hingrat, Paul Isenmann, Joan Martinez i Giner, Raphaël Leblois, José Antonio Mateo, Aïssa Moali, Olivier Peyre, Juan-Manuel Pleguezuelos, Philippe Roux, Paul Soto, Michel Thévenot, Gilles Trochard, Julien Viglione and Sohaib Liefried. Finally, we thank two anonymous reviewers for their constructive remarks that allowed to improve this manuscript.

References

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp 267–281.

Alvarez Y, Mateo JA, Andreu AC, Diaz-Paniagua C, Diez A, Bautista JM (2000) Mitochondrial DNA haplotyping of *Testudo graeca* on both continental sides of the Straits of Gibraltar. *J Hered* **91**:39-41.

Amor N, Farjallah S, Ben-Yacoub S, Merella P, Khaled S (2011) Morphological Variation of the African Green Toad, *Bufo boulengeri* (Amphibia: Anura) in Tunisia. *Pakistan J Zool* **43**:921-926.

Aulagnier S, Thevenot M (1986) Catalogue des mammifères sauvages du Maroc. *Trav Ins Sci Sér Zool* **41**:1-163.

Baha El Din S (2006) *A guide to the Reptiles and Amphibians of Egypt*. American University in Cairo Press, Cairo.

Bailon S (2000) Amphibiens et reptiles du Pliocène Terminal d'Ahl al Oughlam (Casablanca, Maroc). *Geodiversitas,* **22**:539-558.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**:37-48.

Barata M, Harris DJ, Castilho R (2008) Comparative phylogeography of northwest African *Natrix maura* (Serpentes: Colubridae) inferred from mtDNA sequences. *Afr Zool* **43**:1-7.

Batista V, Carranza S, Carretero MA, Harris DJ (2006a) Genetic variation within *Bufo viridis*: evidence from mitochondrial 12S and 16S RRNA DNA sequences. *Butll Soc Cat Herp* **17**:24-33.

Ben Faleh A, Granjon L, Tatard C, Ben Othmen A, Said K, Cosson JF (2012) Phylogeography of the Greater Egyptian Jerboa (*Jaculus orientalis*) (Rodentia: Dipodidae) in Mediterranean North Africa. *J Zool* **286**:208–220.

Ben Hassine J, Gutiérrez-Rodríguez J, Escoriza D, Martínez-Solano I (2016) Inferring the roles of vicariance, climate and topography in population differentiation in *Salamandra algira* (Caudata, Salamandridae). *J Zool Syst Evol Res* **54**:116-126.

Beukema W, De Pous P, Donaire-Barroso D, Bogaerts S, Garcia-Porta J, Escoriza D, Arribas OJ, El Mouden EH, Carranza S (2013) Review of the systematics, distribution, biogeography and natural history of Moroccan amphibians. *Zootaxa* **3661**:1-60.

Bons J, Geniez P (1996) *Amphibiens et Reptiles du Maroc (Sahara Occidental compris), Atlas biogéographique*. Asociación Herpetológica Española, Barcelona, Spain.

Brown RP, Suarez NM, Pestano J (2002) The Atlas mountains as a biogeographical divide in North-West Africa: evidence from mtDNA evolution in the Agamid lizard *Agama impalearis*. *Mol Phylogenet Evol* **24**:324-332.

Brunet P, Sanuy D, Oromi N, Ait Hammou M, Dahmani W (2009) Anuran studies from Tiaret region, north-west of Algeria. *Bol Asoc Herpetol Esp* **20**:68-72.

Chiari Y, Vences M, Vieites DR, Rabemananjara F, Bora P, Ramilijaona Ravoahangimalala O, Meyer A (2004) New evidence for parallel evolution of colour patterns in Malagasy poison frogs (*Mantella*). *Mol Ecol* **13**:3763-3774.

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, McKenna DS, Santer BD, Smith RD (2006) The Community Climate System Model Version 3 (CCSM3). *J Climate* **19**:2122-2143.

Corti, C, Lo Cascio, P, Vanni, S, Turrisi, GF, Vaccaro, A (1997). Amphibians and reptiles of the circumsicilian islands: new data and some considerations. *Boll Mus reg Sci nat Torino* **15**: 179-211.

Damas-Moreira I, Tome B, Harris DJ, Maia JP, Salvi D (2014) Moroccan herpetofauna: distribution updates. *Herpetozoa* **27**:96-102.

deMenocal P (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet Sc Lett* **220**:3-24.

Donaire D, Beukema W, de Pous P, Raul del Canto G (2011) A distributional review of *Bufo boulengeri* Lataste, 1879 in northern Morocco with emphasis on occurrence in the Rif Mountains. *Herpetology Notes* **3**:71-74.

Douady CJ, Catzeflis F, Raman J, Springer MS, Stanhope MJ (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proc Natl Acad Sci U S A* **100**:8325-8330.

Drake NA, Blench RM, Armitage SJ, Bristow CS, White KH (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proc Natl Acad Sci U S A* **108**:458-462.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**:214.

Dubois A, Crochet PA, Dickinson EC, Nemésio A, Aescht E, Bauer AM, Blagoderov V, Bour R, De Carvalho MR, Desutter-Grandcolas L, Frétey T, Jäger P, Koyamba V, Lavilla EO, Löbl I, Louchart A, Malécot V, Schatz H, A. O (2013) Nomenclatural and taxonomic problems related to the electronic publication of new nomina and nomenclatural acts in zoology, with brief comments on optical discs and on the situation in botany. *Zootaxa* **3735**:1-94.

Dufresnes C, Bonato L, Novarini N, Betto-Colliard C, Perrin N, Stöck M (2014) Inferring the degree of incipient speciation in secondary contact zones of closely related lineages of Palearctic green toads (*Bufo viridis* subgroup). *Heredity* **113**:9-20.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C,

Nakamura M, Nakazawa Y, Overton JM, Peterson T, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**:129–151.

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Meth Ecol Evol* **1**:330-342.

Escoriza D, Comas MM, Donaire D, Carranza S (2006) Rediscovery of *Salamandra algira* Bedriaga, 1883 from the Beni Snassen massif (Morocco) and phylogenetic relationships of North African *Salamandra*. *Amphibia-Reptilia* **27**:448-455.

Essghaier MFA, Taboni IM, Etayeb KS (2015) The diversity of wild animals at Fezzan Province (Libya). *Biodiversity Journal* **6**:245-252.

Fahd S, Mediani M (2007) *Herpetofaune du bassin versant de Oued Laou*. Report by Université de Tétouan.

Filippi E, Chetoui A, D'Alterio GL, Parker G, Woodfine T (2011) *Bufo boulengeri*: field observations on breeding phenology at a new Saharan site in Southern Tunisia. *Herpetology Notes* **4**:215-217.

Flot JF (2010) Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Mol Ecol Res* **10**:162-166.

Fourcade Y, Engler JO, Rodder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One* **9**:e97122.

Fritz U, Barata M, Busack SD, Fritzsch G, Castilho R (2006) Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, *Mauremys leprosa* (Reptilia, Testudines, Geoemydidae). *Zool Scripta* **35**:97–108.

Frynta D, Kratochvil L, Moravec J, Benda P, Dandova R, Kaftan M, Klosova K, Mikulova P, Nova P, Schwarzova L (2000) Amphibians and reptiles recently recorded in Libya. *Acta Soc Zool Bohem* **64**:17-26.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915-925.

García-Muñoz E, Jorge F, Rato C, Carretero MA (2010) Four types of malformations in a population of *Bufo boulengeri* (Amphibia, Anura, Bufonidae) from the Jbilet Mountains (Marrakech, Morocco). *Herpetology Notes* **3**:267-270.

Goebel AM, Donnelly JM, Atz ME (1999) PCR primers and amplification methods for 12S ribosomal DNA, the control region, cytochrome oxidase I, and cytochrome b in bufonids and other frogs, and an overview of PCR Primers which have amplified DNA in amphibians successfully. *Mol Phylogenet Evol* **11**:163-199.

Harris DJ, Perera A (2009) Phylogeography and genetic relationships of North African *Bufo mauritanicus* Schlegel, 1841 estimated from mitochondrial DNA sequences. *Biologia* **64**:356-360.

Harris DJ, Carretero MA, Brito JC, Kaliontzopoulou A, Pinho C, Perera A, Vasconcelos R, Barata M, Barbosa D, Carvalho S, Fonseca M, Perez-Lanuza G, Rato C (2008) Data on the distribution of the terrestrial herpetofauna of Morocco: records from 2001-2006. *Herpetological Bulletin* **103**:19-28.

Hasumi H, Emori S, eds. (2004) *K-1 Coupled GCM (MIROC) Description*. K-1 Model developers tech Rep 1, Center for Climate System Research, University of Tokyo, Tokyo, Japan.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc B* **359**:183-195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.

Ho SY, Duchene S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology* **23**:5947-5965.

Ibrahim A (2008) Contribution to the herpetology of southern Libya. *Acta Herpetologica* **3**:35-49.

Ibrahim A (2013) Amphibian in Libya: a status report. *Basic and Applied Herpetology* **27**:101-106.

ICZN (2014) Zoological Nomenclature and Electronic Publication—a reply to Dubois et al. (2013). *Zootaxa* **3779**:3-5.

Kelly RP, Oliver TA, Sivasundar A, Palumbi SR (2010) A method for detecting population genetic structure in diverse, high gene-flow species. *J Hered* **101**:423-436.

Langton TES (1990) Amphibians and roads. Proceedings of the Conference held in Proceedings of the Toad Tunnel Conference Rendsburg, Federal Republic of Germany.

Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *J Arid Environ* **37**:619-647.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451-1452.

Márquez-Rodríguez J (2014) *Bufotes boulengeri* (African Green Toad): New reproductive population in Tunisia. *Herpetological Bull* **127**:33-34.

Martínez-Solano I, Sindaco R, Romano A (2015) *Bufotes boulengeri*. *The IUCN Red List of Threatened Species 2015: eT153568A74497730* http://dxdoiorg/102305/IUCNUK2015-1RLTST153568A74497730en.

Muttoni G, Scardia G, Kent DV (2010) Human migration into Europe during the late Early Pleistocene climate transition. *Palaeogeogr, Palaeoclimat, Palaeoecol* **296**:79-93.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nicolas V, Mataame A, Crochet PA, Geniez P, Ohler A (2015) Phylogeographic patterns in North African water frog *Pelophylax saharicus* (Anura: Ranidae). *J Zool Syst Evol Res* **53**:239-248.

Nicolas V, Ndiaye A, Benazzou T, Souttou K, Delapre A, Denys C (2014) Phylogeography of the North African dipodil (Rodentia: Muridae) based on cytochrome-b sequences. *J Mammal* **95**:241-253.

Nylander JA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991) *The Simple Fool's Guide to PCR*. University of Hawaii Press, Honolulu.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. *Ecol Modelling* **190**:231-239.

Posada D, Crandall K (2001a) A comparison of different strategies for selecting models of DNA substitution. *Systematic Biology* **50**:580–601.

Posada D, Crandall KA (2001b) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* **16**:37-45.

Radosavljevic A, Anderson RP (2014) Making better MAXENT models of species distributions: complexity, overfitting and evaluation. *J Biogeogr* **41**:629-643.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**:2092-2100.

Ray N, Adams JM (2001) A GIS-based Vegetation Map of the World at the Last Glacial Maximum (25,000-15,000 BP). *Internet Archaeology* **11**.

Recuero E, Iraola A, Rubio X, Machordom A, Garcia-Paris M (2007) Mitochondrial differentiation and biogeography of *Hyla meridionalis* (Anura : Hylidae): an unusual phylogeographical pattern. *J Biogeogr* **34**:1207-1219.

Rognon PP (1989) *Biogéographie d'un désert*. Plon, Paris.

Rohling EJ, Fenton M, Jorissen FJ, Bertrand P, Ganssen G, Caulet JP (1998) Magnitudes of sea level lowstands of the past 500,000 years. *Nature* **394**: 162-165.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.

Rosado D, Harris DJ, Perera A, Jorge F, Tome B, Damas-Moreira I, Tavares I, Estrela H, Sousa A, Pereira A, Mantovani M, Salvi D (2016) Moroccan herpetofauna distribution updates including a DNA barcoding approach. *Herpetozoa* **28**:171–178.

San Mauro D, Gower DJ, Oommen OV, Wilkinson M, Zardoya R (2004) Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. *Mol Phylogenet Evol* **33**:413-427.

Schorr G, Holstein N, Pearman PB, Guisan A, Kadereit JW (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine Primula. *Ecol Evol* **2**:1260-1277.

Schuster M, Duringer P, Ghienne JF, Vignaud P, Mackaye HT, Likius A, Brunet M (2006) The age of the Sahara desert. *Science* **311**:821.

Sicilia A, Marrone F, Sindaco R, Turki S, Arculeo M (2009) Contribution to the knowledge of Tunisian amphibians: notes on distribution, habitat features and breeding phenology. *Herpetology Notes* **2**:107-132.

Speybroeck J, Beukema W, Bok B, van der Voort J, Velikov I (2016) *Field Guide to the Amphibians and Reptiles of Britain and Europe*. Bloomsbury, London & New York.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *Am J Hum Genet* **76**:449-462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**:978-989.

Stöck, M, Grifoni, G, Armor, N, Scheidt, U, Sicilia, A, Novarini, N (2016). On the origin of the recent herpetofauna of Sicily: Comparative phylogeography using homologous mitochondrial and nuclear genes. *Zoologischer Anzeiger* 261: 70-81.

Stöck M, Moritz C, Hickerson M, Frynta D, Dujsebayeva T, Eremchenko V, Macey JR, Papenfuss TJ, Wake DB (2006) Evolution of mitochondrial relationships and

biogeography of Palearctic green toads (*Bufo viridis* subgroup) with insights in their genomic plasticity. *Mol Phylogenet Evol* **41**:663-689.

Stöck M, Sicilia A, Belfiore NM, Buckley D, Lo Brutto S, Lo Valvo M, Arculeo M (2008) Post-Messinian evolutionary relationships across the Sicilian channel: mitochondrial and nuclear markers link a new green toad from Sicily to African relatives. *BMC Evol Biol* **8**:56-56.

Stoetzel E (2013) Late Cenozoic micromammal biochronology of northwestern Africa. *Palaeogeogr Palaeoclimatol Palaeoecol* **392**:359–381.

Tabel J, Khater C, Rhoujjati A, Dezileau L, Bouimetarhan I, Carre M, Vidal L, Benkaddour A, Nourelbait M, Cheddadi R (2016) Environmental changes over the past 25 000 years in the southern Middle Atlas, Morocco. *J Quater Sc* **31**:93-102.

Varela S, Lima-Ribeiro MS, Terribile LC (2015) A short guide to the climatic variables of the Last Glacial Maximum for biogeographers. *PLoS One* **10**:e0129037.

Vences M, de Pous P, Nicolas V, Díaz-Rodríguez J, Donaire D, Hugemann K, Hauswaldt JS, Amat F, Barnestein JAM, Bogaerts S, Bouazza A, Carranza S, Galán P, de la Vega JPG, Joger U, ansari A, El Mouden EH, Ohler A, Sanuy D, Slimani T, Tejedo M (2014) New insights on phylogeography and distribution of painted frogs (*Discoglossus*) in northern Africa and the Iberian Peninsula. *Amphibia Reptilia* **35**:305-320.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**:607-611.

Figure legends

Figure 1: Map showing the geographical range (shaded area) of *Bufotes boulengeri boulengeri* according to Martínez-Solano et al. (2015), sampling localities (black dots), and the main mountain ranges (light grey) and districts cited in the text.

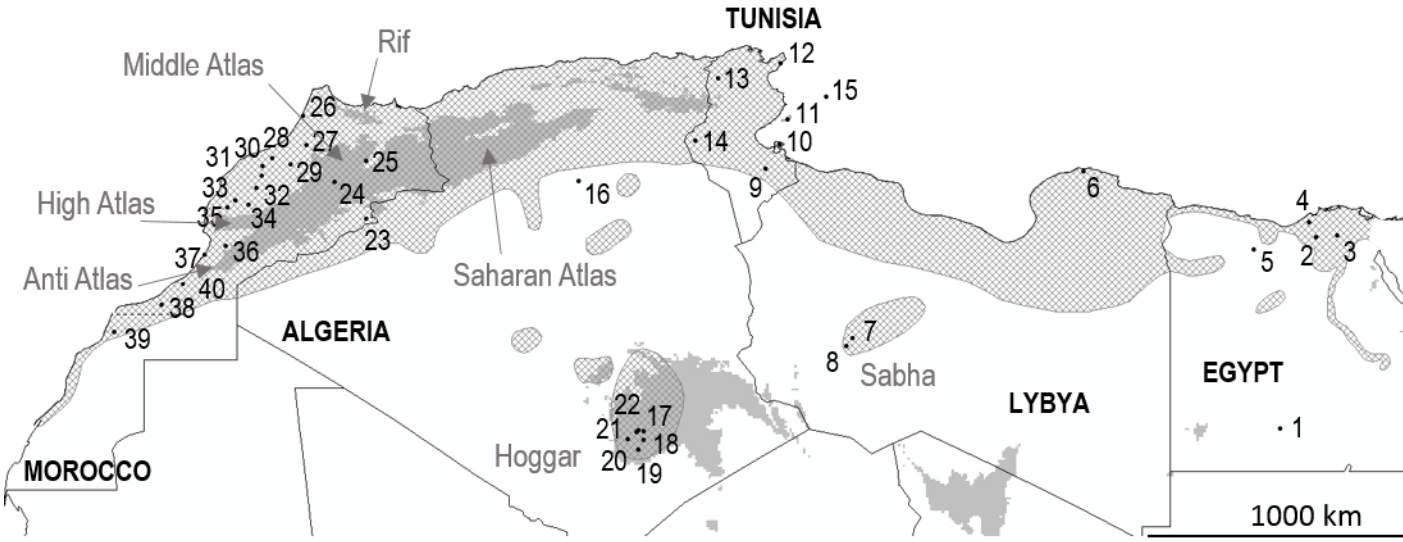


Figure 2. Median-joining network of D-loop (A), 16S (B), CO1 (C) and RAG1 (D) haplotypes of *Bufotes boulengeri boulengeri*. Circle sizes are proportional to the number of similar haplotypes observed in the data set. The number of mutations between haplotypes is proportional to the length of branches. For the D-loop, 16S and CO1 DNA fragments two haplogroups can be identified: haplogroup 1 (haplotypes in white), and haplogroup 2 (haplotypes in black). For the RAG1 fragment no genetic structure can be identified, but haplotypes are coloured in white and black according to the mitochondrial results.

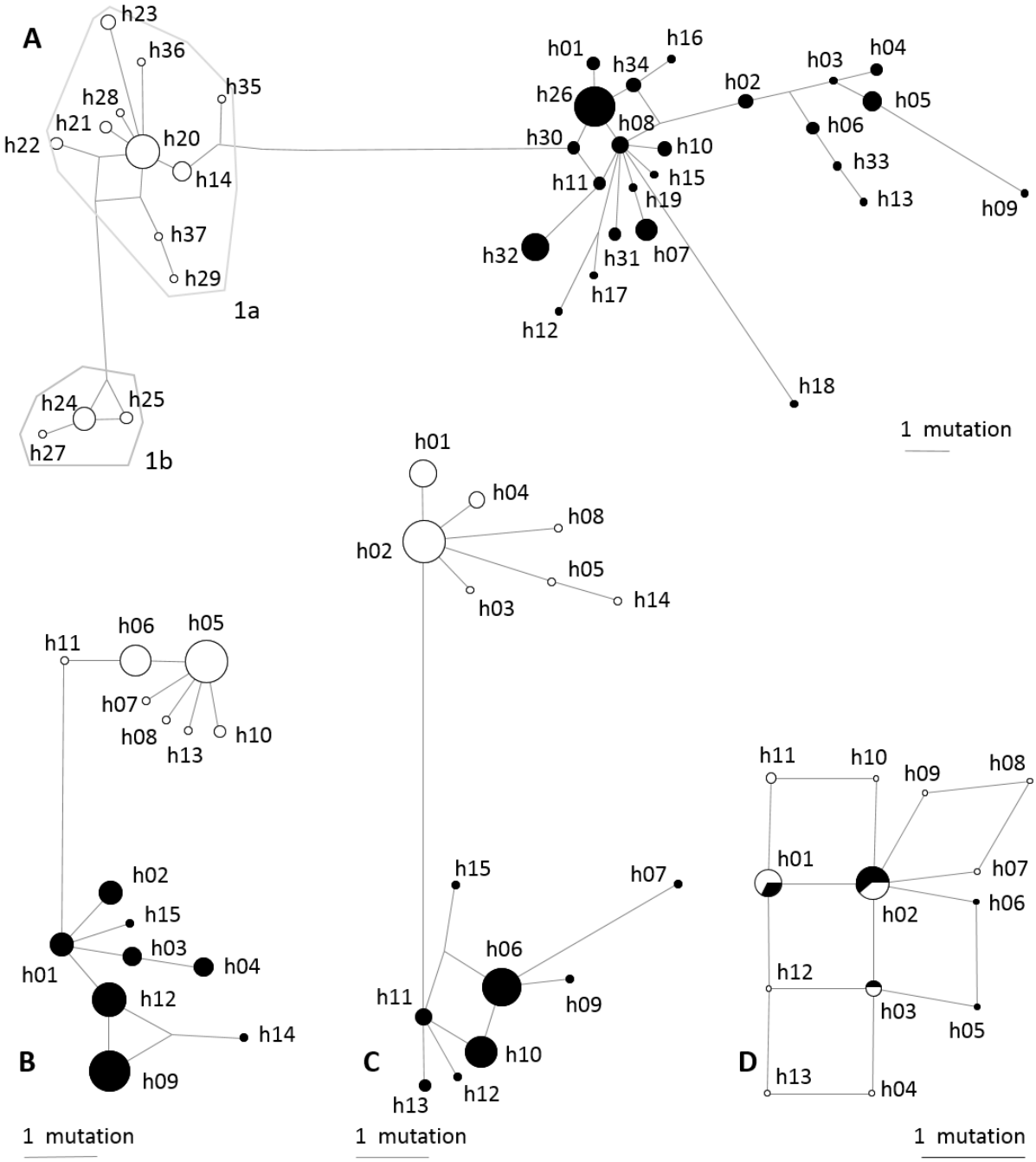
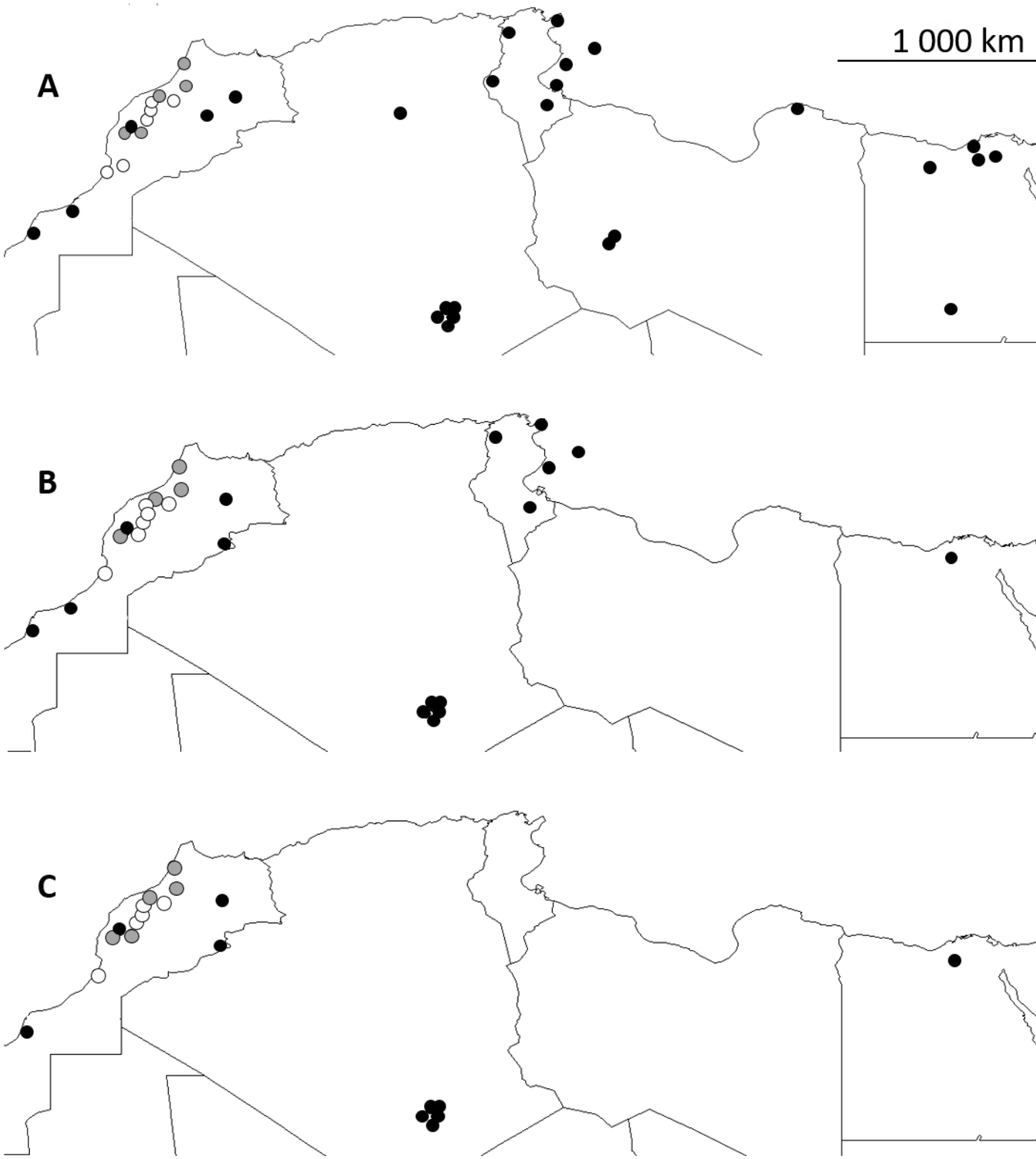


Figure 3. Geographical distribution of the main haplogroups obtained in the median-joining network analyses of the D-loop (A), 16S (B) and CO1 (C) haplotypes of *Bufotes boulengeri boulengeri*. White = only specimens belonging to haplogroup 1 are present in this locality; black = only specimens belonging to haplogroup 2 are present in this locality; grey = specimens of both haplogroups 1 and 2 co-occur in this locality.



1

Figure 4. Molecular phylogenetic tree inferred using Bayesian inference for the concatenated data set (D-loop and 16S). The average Bayesian posterior probabilities (left) of two runs (pp > 0.85) and bootstrap values (> 70%, right) calculated for 1,000 bootstrap replicates are indicated for each node. Colours indicate the geographical origin of the samples: black = Morocco, red = Tunisia, blue = Algeria, green = Egypt, purple = Lampedusa island.

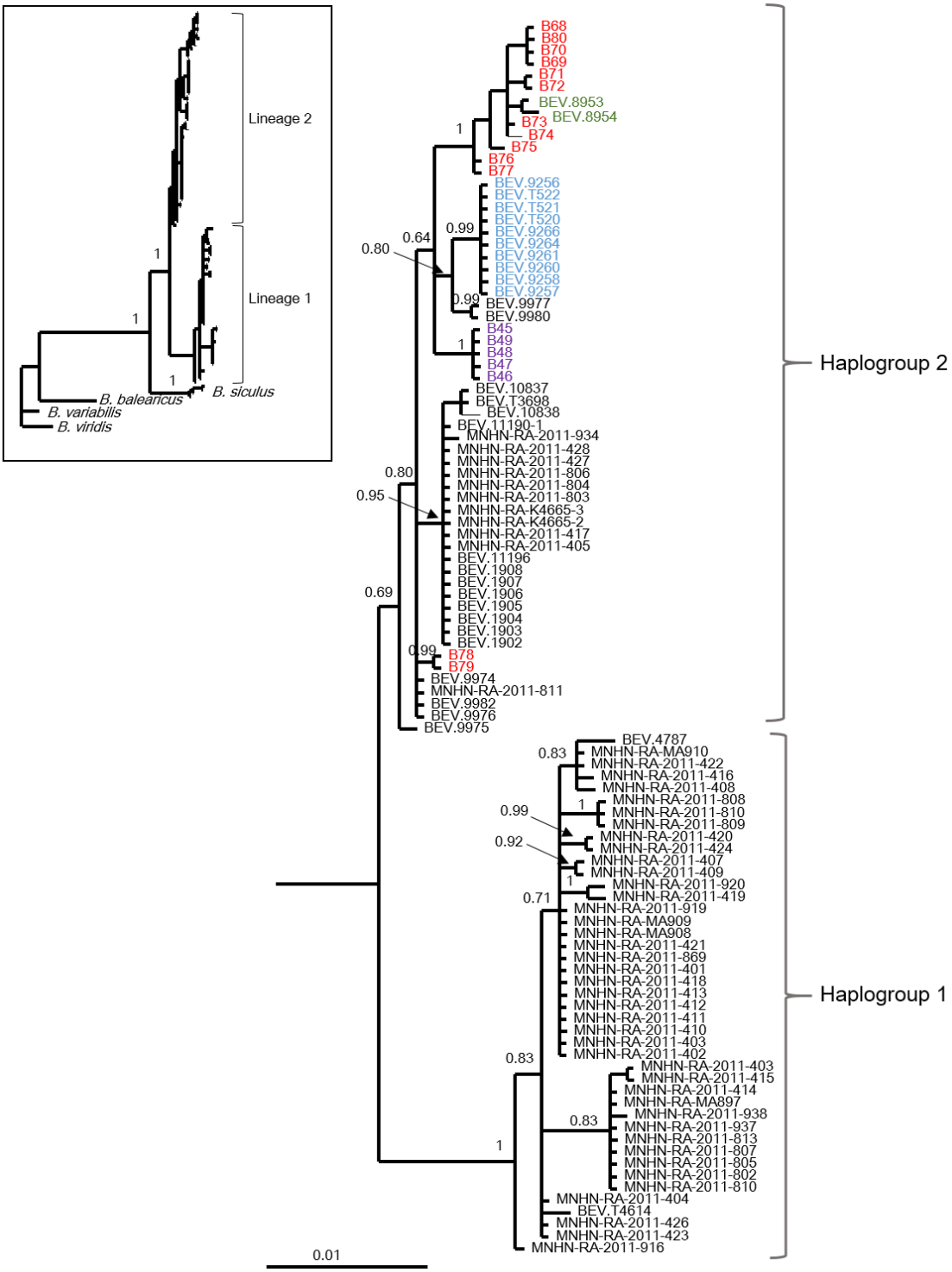
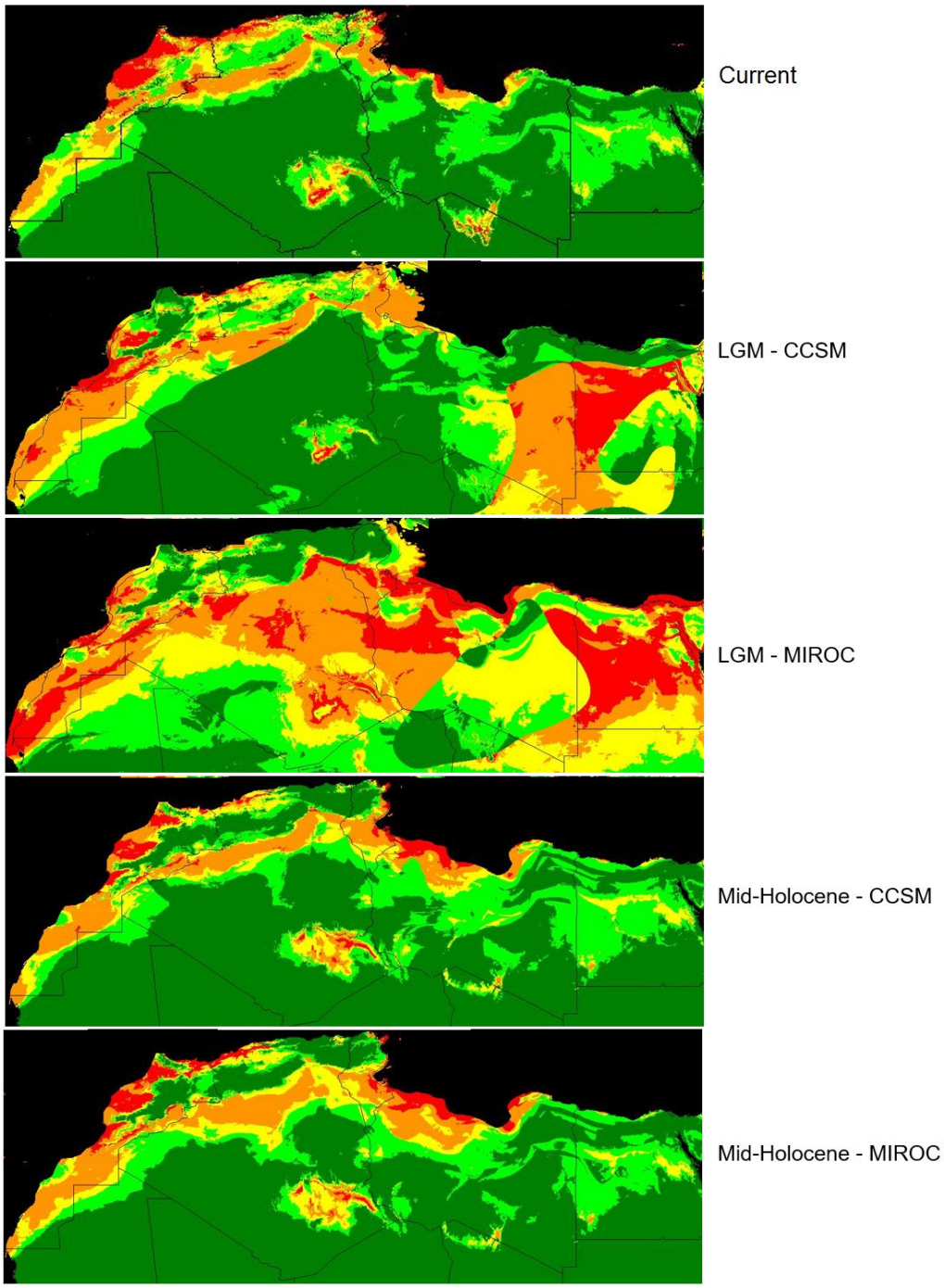


Figure 5. Species distribution modelling of *Bufotes boulengeri boulengeri* as estimated by Maxent for present-day conditions (current), for the last glacial maximum (LGM) and for the mid-holocene. For paleoclimatic data estimates are based on the Model for Interdisciplinary Research on Climate (MIROC) and on the Community Climate System Model (CCSM) paleoclimatic models. Warmer colours show areas with higher probability of presence.



Supporting Information

Table S1. List of specimens used in this study, with museum numbers, geographic origins (specimens are order by continent (Africa, Europe), country, latitude and longitude), GenBank accession numbers (GB)haplogroup (L) and haplotype numbers (H) for each gene. In bold: newly obtained sequence data.

Table S2. Primer sequences used in this study.

Table S3. Distribution records of African green toads (sorted by country name, latitude and longitude) assembled from literature, museum collections (Global Biodiversity Information Facility, http://www.gbif.org, last access date: 26 August 2016) and our own unpublished databases.

Figure S1. Maps of localities used to train the Maxent species distribution model (SDM) for *Bufotes boulengeri boulengeri*.

Table 1: Collection localities of *Bufotes boulengeri boulengeri* specimens included in this study, with the number of specimens sequenced per DNA region and per locality. The symbol "*" indicates that several collecting sites were sampled within a given locality (see Supporting information Table S1 for details on collecting sites), but they were grouped in our analyses for statistical reasons. Numbers in parentheses for the RAG1 gene indicate specimens that were sequenced but not included in the final analyses because phasing was doubtful. Localities are order by continent (Africa, Europe), country, latitude and longitude.

Country	Locality	Locality number on Figure 1	Latitude	Longitude	D-loop	16S	CO1	RAG1
Italy	Lampedusa island	15	35.508	12.600	6	5	-	-
Morocco	Saqiyat al-Hamra, 4 km WWWNW Edchera	39	27.030	-13.093	1	1	1	1
	15 km NNE Abtih	38	28.001	-11.367	1	1	1	1
	7 km W Tantan	40	28.438	-11.174	1	1	1	1 (1)
	Aglou-Plage, oued Oudoudou	37	29.807	-9.826	1	1	1	-
	Ait Baha	36	30.130	-9.080	1	-	-	-
	Sidi Mokhtar*	35	31.579	-9.019	15	16	15	7
	Sidi Chiker*	33	31.751	-8.736	9	9	9	4 (1)
	Oued N'Fis, 20 km W Marrakesh	34	31.630	-8.248	2	2	2	1
	Douar Iben Brahim	32	32.220	-7.970	1	1	1	-
	Doulad Bouziri	31	32.658	-7.779	2	2	2	-
	Settat, Douar Setoute	30	33.001	-7.741	1	1	1	-
	16 and 17 km EEENE Berrechid	28	33.296	-7.392	14	14	14	-
	Ouled Boughadi	29	33.076	-6.728	4	5	4	3
	Merja Zerga*	26	34.818	-6.291	9	11	9	4
	Dayet Er-Roumi*	27	33.752	-6.185	5	5	5	5 (1)
	High Atlas (ca 9.5 km SE Tounfite)	24	32.427	-5.156	2	-	-	-
	Dayet Termerzigat, near Merzouga	23	31.081	-4.042	-	1	1	-
	Al Baten	25	33.186	-3.990	6	7	6	5
Algeria	Ghardaia	16	32.483	3.667	1	-	-	-
	Hoggar, locality 1	17	23.454	5.995	2	2	2	2
	Hoggar, locality 2	18	23.126	5.989	3	4	3	-

	Hoggar, locality 3	20	23.157	5.423	1	2	1	-
	Hoggar, locality 4	21	23.421	5.762	1	2	3	-
	Hoggar, locality 5	22	23.478	5.810	1	2	2	-
	Hoggar, locality 6	19	22.786	5.814	2	2	2	-
Tunisia	El Kef	13	36.166	8.700	2	2	-	-
	Cap Bon, Lebna	12	36.728	10.931	3	3	-	-
	Kerkennah Islands, Chergui*	11	34.820	11.258	5	5	-	-
	Djerba island	10	33.800	10.900	1	-	-	-
	Nefta oasis	14	33.917	7.883	1	-	-	-
	Tataouine	9	32.903	10.416	3	3	-	-
Libya	Al' Fiayi	8	26.533	13.317	1	-	-	-
	Gabroon lake	7	26.803	13.535	3	-	-	-
	Sahahhat	6	32.817	21.867	3	-	-	-
Egypt	Matrouh	5	28.750	26.980	2	-	-	-
	Oasis Dakhla	1	25.553	28.948	2	-	-	-
	70 km S Alexandria	4	30.584	29.963	1	-	-	-
	Alexandria, El Menoufia	3	30.500	31.000	1	-	-	-
	Wadi El Natrun	2	30.413	30.285	2	2	2	-
Unknown	Unknown				-	1	-	-
Total					122	113	88	35 (3)

Table 2: Diversity and neutrality estimates for the main haplogroups identified in the median-joining analyses. Significant values are in bold. Number of sites (bp), number of sites excluding indels (bpi), number of sequences (N), number of polymorphic sites (Np), number of distinct haplotypes (Nh), haplotype diversity (Hd), nucleotide diversity (Pi) and average number of nucleotide differences (k) are given.

	bp	bpi	N	Np	Nh	Hd	Pi	k	Fu's Fs	R2
D-loop										
All	545	534	122	40	34	0.930 ± 0.012	0.01498 ± 0.00063	8.000		
Haplogroup 1a	545	542	33	15	10	0.744 ± 0.072	0.00326 ± 0.00061	1.769	-3.466	**0.058**
Haplogroup 1b	545	543	11	2	3	0.473 ± 0.162	0.00094 ± 0.00036	0.509	-0.659	0.165
Haplogroup 2	545	534	78	21	23	0.891 ± 0.025	0.00631 ± 0.00045	3.368	**-8.138**	0.078
16S										
All	586	583	113	14	12	0.791 ± 0.021	0.00617 ± 0.00019	3.597		
Haplogroup 1	586	585	47	6	7	0.601 ± 0.054	0.00122 ± 0.00017	0.712	**-3.316**	0.065
Haplogroup 2	586	583	66	4	5	0.585 ± 0.051	0.00138 ± 0.00015	0.807	-0.72	0.103
CO1										
All	806	806	88	23	15	0.833 ± 0.020	0.00629 ± 0.00018	5.071		
Haplogroup 1	806	806	43	8	7	0.627 ± 0.063	0.00115 ± 0.00022	0.930	-2.442	0.065
Haplogroup 2	806	806	45	9	8	0.695 ± 0.047	0.00144 ± 0.00023	1.164	-1.642	0.074
RAG										
All	937	937	64	7	13	0.728 ± 0.042	0.00117 ± 0.00013	1.101		