The role of climatic cycles and trans-Saharan migration corridors in species diversification: Biogeography of *Psammophis schokari* group in North Africa

D. Vasconcelos Gonçalves\(^a,b,c,\ast\), F. Martínez-Freiría\(^a\), P.-A. Crochet\(^d\), P. Geniez\(^e\), S. Carranza\(^c\), J.C. Brito\(^a,b\)

\(^a\) CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
\(^b\) Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua Campo Alegre, 4169-007 Porto, Portugal
\(^c\) Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, E-08003 Barcelona, Spain
\(^d\) CNRS-UMR 5175, Centre d’Ecologie Fonctionnelle et Evolutive, 1919 route de Mende, F-34293 Montpellier, France
\(^e\) EPHE, PSL Research University, CNRS, UM, SupAgro, IND, INRA, UMR 5175 CEFE, 34293 Montpellier, France

**A R T I C L E  I N F O**

**Keywords:**
- Climatic cycles
- Ecological niche-based modelling
- Paleoclimate
- Phylogeography
- Snakes
- Vicariance

**ABSTRACT**

Highlands, hydrographic systems and coastal areas have been hypothesised to form corridors across the hyperarid Sahara desert in North Africa, allowing dispersal and gene flow for non-xeric species. Here we aim to provide a genetic test for the trans-Saharan corridor model, and predict the location and stability of ecological corridors, by combining phylogeography and palaeoclimatic modelling. The model was the *Psammophis schokari* (Schokari sand racer) group, fast-moving and widely distributed generalist colubrids occurring mostly in arid and semiarid scrublands. We combined dated phylogenies of mitochondrial and nuclear markers with palaeoclimatic modelling. For the phylogeographic analysis, we used 75 samples of *P. schokari* and *P. aegyptius*, and Bayesian and Maximum-Likelihood methods. For the ecological models, we used Maxent over the distribution of *P. schokari* and West African lineages. Models were projected to past conditions (mid Holocene, Last Glacial Maximum and Last Inter-Glacial) to infer climatic stable areas. Climatic stability was predicted to be mostly restricted to coastal areas and not spatially continuous. A putative temporary trans-Saharan corridor was identified in Eastern Sahara, with a more stable one along the Atlantic coast. Six parapatric lineages were identified within *P. schokari*, four occurring in North Africa. These likely diverged during the Pliocene. The Tamanraset River might have been a vicariant agent. African lineages may have experienced further subsequent diversification during the late Pliocene. The main *P. schokari* refugia were probably located along the northern margins of the Sahara, allowing its North-to-South colonization. Trans-Saharan corridors seem to have played a role in *P. schokari* biogeography, allowing colonization of central Saharan mountains and Sahel. Some might have worked as refugia, and even the most stable corridors may have sections working as filters, depending on each climatic phase. We expect the use of trans-Saharan corridors to decrease for more mesic species or with less dispersal capabilities.

1. Introduction

Numerous geological and climatic events have affected the geographic and biological diversity of North Africa in the last few million years (Le Houérou, 1997; Fabre, 2005). Geological events include the opening of the Mediterranean to the Atlantic coast 7–9 million years ago (Ma), the subsequent closure 6 Ma and re-opening 5.3 Ma (MSC, Krijgsman et al., 1999), recurrent episodes of desiccation and refilling in the Red Sea area (Girdler, 1991; Bosworth et al., 2005), marine transgressions (Tawadros, 2011), or the Atlas mountains uplift (de Jong, 1998). These had climatic repercussions, but the most widespread climatic event was a shift from tropical to arid environments around mid-Miocene (Zachos et al., 2001) that eventually led to the appearance of the Sahara desert between 7 Ma and 2.5 Ma (Schuster et al., 2006; Swezey, 2009). Arid and humid conditions then alternated during the last few million years, causing a series of expansions and contractions of climatic zones (Le Houérou, 1992; Swezey, 2009) that largely determined current biodiversity patterns (Brito et al., 2014).

Diversification due to humid-arid cycles has been explained through genome rearrangements (Dobigny et al., 2005), adaptation to novel
habits (Boratynski et al., 2012; Carranza et al., 2002; Guillaumet et al., 2008) or, most commonly, divergence in allopatry (Brito et al., 2014). While during humid phases, the hyper-arid regions were reduced and probably isolated, in arid phases the mesic species were pushed towards coastal areas and mountains (Le Houérou, 1992; Messeri and Winiger, 1992). This resulted in disjoint distributions and allopatric diversification, currently best observed in mesic species, for example Mediterranean-Sahel separations (e.g. Gonçalves et al., 2012; Guillaumet et al., 2008), isolated populations in highlands (Geniez and Arnold, 2006; Metallinou et al., 2015), rock pools (Brito et al., 2011a; Vale et al., 2015) or desert-border refugia (Dobigny et al., 2013). Arid phases conversely allowed range expansions of xeric taxa (Arnold et al., 2008; Kissling et al., 2016; Leaché et al., 2017; Pook et al., 2009), later broken during humid phases (Metallinou et al., 2015; Pook et al., 2009).

Landscape features modulate gene flow (Brown and Lomolino, 1998). Depending on climatic-cycle phase and species’ ecological requirements, a geographic feature can constitute a barrier to gene flow, an ecological corridor (connecting two larger similar areas), a filter bridge (or barrier, a more selective connection), or a refugium (where a species survives during unfavourable periods). Lake Chad, for instance, has been identified as a refugium (Granjon and Dobigny, 2003) or a corridor connecting the Sahel to the Tibesti (Drake et al., 2011; Dumont, 1982) for mesic taxa and a vicariant barrier for xeric ones (Pook et al., 2009; Metallinou et al., 2015). Mountains, presently working as biodiversity hotspots and refugia for mesic species (Brito et al., 2014, 2011a; Vale et al., 2015), likely constitute barriers for lowland or xeric species. Lastly, mountains, coastal areas and hydrographic systems can be linked, forming ecological corridors for mesic species. Several areas likely to constitute North-South oriented ecological corridors have been proposed (Dumont, 1982); these geographic features are hereby referred to as trans-Saharan corridors (tS-corridors), to avoid confusion with the ecological corridor feature. They include the more transitory river drainages from central Sahara Mountains (Drake et al., 2011), or the more stable (thus possibly refugia for some species) Red Sea, Nile River or Atlantic Sahara (Brito et al., 2014). However, these have been proposed based on ecological/geological data and species’ distributions, and no genetic-level assessments have been conducted so far.

Our goal is to provide a genetic assessment of the validity of the tS-corridor model. For that, since most speciation events far predate the last humid/arid shift, data on the distribution of intraspecific genetic variability is required. While many mesic taxa may use the corridors during the humid phases, only those with high mobility and more adaptations to xeric environments are expected to do so during arid phases. Mesic-xeric species with broad ecological spectrum are thus expected to make more use of tS-corridors. We have selected the Psammophis schokari (Schokari sand racer) group in North Africa as a model since it is widely distributed in North Africa, occurring mostly in arid and semiarid habitats; it presents a continuous distribution along the Atlantic coast, thus apparently making use of this tS-corridor (Fig. 1). These snakes are large and mobile animals with likely good dispersal abilities. Thus, we do not expect limited dispersal or narrow distribution to restrict its use of ecologically suitable corridors. Previous work on P. schokari (Rato et al., 2007) identified several lineages but no conclusions were drawn regarding trans-Saharan dispersal.

Here we propose to model the species’ potential distribution in different climatic phases to assess the suitability of candidate tS-corridors as ecological corridors for Psammophis, thus allowing us to build clear hypothesis on the persistence of gene flow along tS-corridors for mesic species during the past climatic cycles. The general aim of this study is to assess the role of corridors in trans-Saharan dispersal, with particular focus on the Atlantic Sahara tS-corridor, in an integrative framework joining phylogeography and palaeoclimatic modelling. Using P. schokari, we aim to answer the questions: (1) where are the areas with higher climatic stability throughout the species range and particularly West Africa and where are the potential dispersal routes across the Sahara?; (2) how is the genetic variability spatially structured? By combining results from these two sections, we expect to find phylogeographic patterns coherent with refugia close to the Mediterranean coast and in the Saharan mountains, and tS-corridors connecting them.

2. Material and methods

2.1. Sampling and study areas

The snake genus Psammophis includes 34 diurnal fast-moving species occurring mostly throughout tropical Africa, with some species reaching the Middle East and South-Central Asia (Sindaco et al., 2013; Utz and Hošek, 2016). Psammophis schokari (FORSKål, 1775), commonly known as Schokari sand racer, is a common colubrid occurring from West Africa to India (Fig. 1), mostly in desert and xeric scrublands, marginally in sandy habitats in dry Mediterranean zones (Kelly et al., 2008; Schleich et al., 1996; Sindaco et al., 2013). Its sister species Psammophis aegyptius MARX, 1958, commonly known as Egyptian sand snake, is a typical Saharan species renowned for inhabiting even the driest areas of the eastern Sahara (Baha El Din, 2006).

We used 68 samples of P. schokari covering a representative part of the species distribution, particularly in West Africa (Table S1; Fig. 1). Seven samples of P. aegyptius were also included. For the phylogenetic analyses (see below), outgroups and other species of Psammophis were selected based mostly on Kelly et al. (2008). Additional P. schokari sequences were retrieved from GenBank. Two datasets were assembled: dataset 1, used to evaluate the phylogenetic relationships and placement within the genus of the P. schokari and P. aegyptius lineages; and dataset 2, used to date the divergence events among the lineages. Dataset 1 included all P. schokari and P. aegyptius, eleven Psammophis species representative of the diversity of the genus, and six outgroups (Table S1). Dataset 2 included eleven P. schokari and two P. aegyptius specimens representative of intraspecific lineages, plus representatives of other species spanning the Psammophis genus and the superfamily Colubroidae, plus Acrorodius granulatus as outgroup, in line with previous works (Table S1; see also Section 2.5).

For ecological models, a total of 748 observations (Figs. S3 and S4) were collected from fieldwork (n = 244), museum collections (n = 110) and bibliographic (n = 394). These observations were used to create two datasets: (i) Global, with 629 records at five arc-minute resolution (∼ 10 × 10 km); and (ii) Regional (Northwest Africa), with 379 records at 30 arc-second resolution (∼ 1 × 1 km). A 150 km buffer around minimum convex polygons including each dataset was used to delimit two corresponding study areas (see Section 2.6). All spatial analyses were conducted in ESRI ArcGIS 10. In order to reduce bias from uneven sampling, and to geographically and environmentally homogenize datasets (Merow et al., 2013), localities were randomly removed from clusters of species occurrence (e.g. Martinez-Frefía et al., 2015). The nearest-neighbour (NN) index (ArcGIS 10) was used as assessment. Low clustered distributions were obtained for both global (z-score = −19.67; NN-ratio = 0.478) and regional (z-score = −11.57; NN-ratio = 0.599) datasets, keeping 388 and 225 records, respectively (Table 2).

2.2. Climatic variables

Nineteen variables for current climatic conditions at 30 arc-second resolution (∼ 1 × 1 km) were downloaded from WorldClim (www.worldclim.org; Hijmans et al., 2005). Variables were clipped to each study area (Global and Regional) and, for the Global dataset, upscaled to five arc-minutes (∼ 10 × 10 km). After visual inspection, five variables were excluded due to the presence of spatial artefacts. The remaining 14 variables (Table S2) were considered for ecological models. Bivariate correlations among the 14 variables have been conducted so far.

Bivariate correlations among the 14 variables were tested within Global (z-score = −11.57; NN-ratio = 0.599) datasets, keeping 388 and 225 records, respectively (Table 2).
and Regional datasets. We retained the same five slightly correlated (R < 0.7) variables in both datasets (BIO 4, 10, 12, 14, 19; Table S2), which are commonly used in ecological niche-based modelling approaches developed for other snake species (e.g. Martínez-Freiría et al., 2015; Brito et al., 2011b).

For past conditions, the same five variables were downloaded from WorldClim for the Last Interglacial (LIG; ~120,000–140,000 years BP; Otto-Bliesner et al., 2006), Last Glacial Maximum (LGM; ~21,000 years BP; Paleoclimate Modelling Intercomparison Project Phase III – PMIP3-CMIP5) and middle Holocene (midHol; 6000 years BP; PMIP3-CMIP5) periods. These periods are generally accepted as qualitatively representative of previous cycles during the Pleistocene (e.g. Martínez-Freiría et al., 2017), in spite of variations in duration and intensity, since later cycles had the largest amplitude (Snyder, 2016). LIG and midHol variables were at 30 arc-second resolution. The corresponding Global Circulation Models (GCM) were the NCAR-CCSM and CCSM4, respectively. LGM variables were at 2.5 arc-second resolution (~5 × 5 km) and were retrieved for three GCMs: CCSM4, MIROC-ESM and MPI-ESM-P. Past variables were also upsampled to 5 arc-minute resolution for Global models and kept at original pixel size for Regional models.

2.3. DNA extraction and amplification

DNA was extracted from ethanol-preserved tissue using DNeasy Blood & Tissue Kit (Qiagen) as per manufacturer’s instructions. Amplifications were performed using MyTaq™ Mix. To benefit from sequence data available from GenBank (Kelly et al., 2008; Rato et al., 2007), specimens were bi-directionally sequenced for two mitochondrial (ND4, ND4L; CYTB, Cytb; oocyte maturation factor MOS, c-mos; Recombination activating gene 2, RAG2) gene fragments. ND4, CYTB, c-mos and RAG2 were amplified using primers from Arevalo et al. (1994), de Queiroz et al. (2002), Saint et al. (1998), and Vidal and Hedges (2005), respectively. PCR conditions were: pre-denaturation at 94 °C (15′), 40 cycles with 92 °C (30′) denaturing, 50°C (45′) annealing and 72 C (45′) extension; final extension at 60 °C (15′). Some samples required repeating the reactions with annealing temperatures in the 48–52 °C range.

2.4. Phylogenetic analyses

Sequences were aligned using mafft v7 (Katoh and Standley, 2013), with Auto option, then proofread by eye. No stop codons were found in coding genes. Each marker was individually analysed inferring independent ML trees in RAXML v8.1.21 (Stamatakis, 2014), in order to detect topological incongruences suggesting sample curation errors. For both sequence datasets, the best-fit partitioning scheme and models of molecular evolution were selected using PartitionFinder v.1.1.1 (Lanfear et al., 2012) with the following settings: linked branch lengths, beast models, BIC model selection criterion, and all partition schemes searched. Initial partitions schemes by gene and by codon were applied. Phylogenies were inferred using Bayesian and Maximum Likelihood (ML) methods. Gene and codon partitions were applied according to PartitionFinder. Bayesian analyses were performed in MrBayes version 3.2.6 (Ronquist et al., 2012) and beast v1.8.3 (Drummond et al., 2012). MrBayes was run for 10⁷ generations in two independent runs sampling every 1000 generations. Parameters of sequence evolution (statefreq, revmat, shape, pinvar) were unlinked for all partitions and the overall rate (ratepr) variable among them. beast was run in CIPRES gateway (Miller et al., 2010) in three independent runs of 5 × 10⁷ generations, sampling at every 5000, with unlinking substitution and clock models, under an uncorrelated lognormal relaxed clock (Drummond et al., 2006), and considering ambiguities in nuclear sequences (manually editing the xml file to UseAmbiguities = true). A constant population size coalescent tree prior (Kingman, 1982) was used for dataset 1, and a Yule speciation tree prior (Yule, 1925; Gernhard, 2008) for dataset 2. Burn-in was determined using Tracer v1.6 (Rambaut et al., 2014), upon stabilisation of log likelihood, average standard deviation of split frequencies, and ESS for all parameters. For beast analysis, runs were combined with LogCombiner and a maximum credibility tree was generated with TREEANNOTATOR (both in the BEAST package). Burn-in was determined using TRACER v1.6 (Rambaut et al., 2014). ML analyses were performed in RAXML v8.1.21 (Stamatakis, 2014) through RAxMLGUI 1.5b1 (Silvestro and Michalak, 2012), with partition schemes as above and GTR + G model of sequence evolution. The program was set to perform 10 ML searches and 1000 thorough bootstrapping replicates. To test the paraphyly against the monophyly of African populations, the best tree in which Algerian/Tunisian populations are sister taxa to the rest of African and Middle East populations was compared with the alternative constrained topology in which Africa was forced monophyletic. Per-site log likelihoods were obtained with RAxML, then used to run the Shimodaira-Hasegawa (SH) (Shimodaira and Hasegawa, 1999) and Approximately-Unbiased (AU) (Shimodaira, 2002) topology tests in consel (Shimodaira and Hasegawa, 2001).

MEGA6 (Tamura et al., 2013) was used to calculate uncorrected p-distances within and among groups of samples for each mitochondrial marker in dataset 1. DNAsp v.5.10.1 (Librado and Rozas, 2009) was used to calculate sequence and nucleotide diversity measures. The PHASE
2.1 algorithm (Stephens et al., 2001) implemented in DnaSP was used to infer haplotypes for nuclear sequences. The software was run for 10^4 generations with a thinning interval of five and a burn-in value of 1000, and repeated three times. Consistency was checked across runs by analysing haplotype frequency estimates and goodness-of-fit measures. Haplotypes were used to produce haplotype networks using NETWORK 5.0 (Fluxus-engineering.com) with median-joining algorithm and default parameters (Bandelt et al., 1999).

2.5. Time calibration

No calibration constraints are available within the genus Psammophis or the family Psammophiidae. Therefore, to obtain estimates of the divergence times among P. schokari lineages, the dataset was expanded to span the superfamily Colubroidea plus Achrochordus, and thus make use of the calibration scheme used by Wüster et al. (2008) and Pook et al. (2009). Other authors have used a slightly different prior set (Sanders and Lee, 2008; Kelly et al., 2009), so we also dated the phylogeny according to the revised version by Sanders et al. (2010). RAG2 was excluded given that it was not available for most of the dataset.

2.6. Palaeoclimate modelling

Palaeoclimate models were computed at a Global scale, which includes the whole species distribution and aims at a species-wide average niche, and at a Regional scale, including just the West African populations and targeting local environmental species preferences. Models were generated using the maximum entropy approach with MAXENT v3.3.1k (Phillips et al., 2006). This algorithm requires only presence data, performs well comparing to other methods (Elith et al., 2006), and has been used successfully in modelling snake species distributions (e.g. Brito et al., 2011b; Martínez-Fréiría et al., 2015). Thirty replicates were run with random seed and 70%/30% training/testing partition and using bootstrap with replacement. Models were run with auto-features, and the area-under-the-curve (AUC) of the receiver-operating characteristics (ROC) plots was taken as measure of individual model fit (Fielding and Bell, 1997). The importance of climatic variables in explaining the species’ distribution was determined by their mean percentage contribution to the models.

The individual model replicates were used to generate a mean forecast of probability of species occurrence under current conditions (Marmion et al., 2009). Standard deviation was used as indication of prediction uncertainty (e.g. Brito et al., 2011b; Martínez-Fréiría et al., 2015). Individual model replicates were projected to past climatic conditions (midHol, LGM and LIG) and subjected to the same procedure. Stable climatic areas, i.e. stable potential areas of occurrence that could serve as refugia in different time periods (Carnaval et al., 2009), were calculated by averaging the probability of occurrence across all time phases.

2.7. Niche overlap test

In order to assess the potential role of ecological processes in the evolutionary history of the lineages along the Atlantic corridor, we tested niche overlap, equivalency and similarity between regional lineages (see Warren et al., 2008). Tests were based on a 2D representation of climatic space of lineages (retrieved using 14 climatic variables; Table S2), represented by the first two components of a Principal Component Analysis (PCA). We used the PCA-env ordination approach developed by Broennimann et al. (2012) in R (R Core Team, 2016). This approach uses Gaussian functions to smooth densities of records and background climate and measures overlap through Schoener’s D metric. The significance of equivalency and similarity tests is approached via randomization; we used 100 repetitions.

3. Results

3.1. Phylogenetic relationships

Most samples were successfully sequenced for all markers (Table S1). Dataset 1 resulted in an alignment of 3111 positions (Table S1). Uncorrected p-distances and diversity measures for the mitochondrial markers of P. schokari and P. aegyptius lineages can be found in Table 1. Genetic identity of specimens was coherent with prior morphology-based species assignment: specimens identified as Psammophis schokari and P. aegyptius formed two reciprocally monophyletic groups in the concatenated phylogenetic ML and BI trees (Fig. 2). The nuclear DNA did not exhibit a clear separation in the network analysis, with a shared basal haplotype in c-mos, and some RAG2 schokari haplotypes being closer to aegyptius haplotypes than to the remaining schokari alleles (Fig. 2). Six monophyletic lineages of P. schokari were found, three occurring in Northwest Africa with apparently parapatric distributions, one in NE Africa and the Middle East, one in Iran, and one in SE Arabian Peninsula (Oman). ML and BI supported sub-structuring was found within all four lineages occurring in Africa. Except for the earliest-splitting Oman lineage, relationships among lineages were not resolved. Monophyly of the NW African lineages was recovered in some analysis, but the topology test was inconclusive (Table S4). A “Middle East” lineage, spreading from Egypt to the Persian Gulf, and a Mauritanian lineage occurring throughout the country were identified. The later apparently contacts with the Moroccan one. Most nuclear haplotypes of the Mauritanian lineage are shared by Moroccan populations (Fig. 2).

Table 1

<table>
<thead>
<tr>
<th>p-dist.</th>
<th>P. aegyptius</th>
<th>P. schokari</th>
<th>Morocco</th>
<th>Mauritania</th>
<th>Algeria</th>
<th>Middle East</th>
<th>Oman</th>
<th>N</th>
<th>h</th>
<th>Hdh</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7 ± 0.2</td>
<td>9.3 ± 0.8</td>
<td>9.4 ± 0.9</td>
<td>9 ± 0.9</td>
<td>9.4 ± 0.9</td>
<td>9.5 ± 0.9</td>
<td>9.3 ± 0.9</td>
<td>6</td>
<td>5</td>
<td>0.933 ± 0.122</td>
<td>0.00747 ± 0.00135</td>
<td></td>
</tr>
<tr>
<td>0.9 ± 0.2</td>
<td>4.0 ± 0.3</td>
<td>11.3 ± 1.1</td>
<td>4.1 ± 0.4</td>
<td>11.3 ± 1.2</td>
<td>0.6 ± 0.1</td>
<td>4.5 ± 0.6</td>
<td>4.8 ± 0.6</td>
<td>5.8 ± 0.6</td>
<td>6.6 ± 0.7</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>13</td>
<td>11</td>
<td>0.974 ± 0.039</td>
<td>0.00587 ± 0.00133</td>
<td></td>
</tr>
<tr>
<td>5.1 ± 0.7</td>
<td>1 ± 0.2</td>
<td>10.1 ± 1.1</td>
<td>5.6 ± 0.8</td>
<td>6.1 ± 0.8</td>
<td>6.1 ± 0.8</td>
<td>1.1 ± 0.3</td>
<td>10</td>
<td>6</td>
<td>0.844 ± 0.103</td>
<td>0.01172 ± 0.00195</td>
<td></td>
</tr>
<tr>
<td>2.4 ± 0.2</td>
<td>6.6 ± 0.6</td>
<td>11.6 ± 1.2</td>
<td>5.4 ± 0.7</td>
<td>5.8 ± 0.8</td>
<td>5.7 ± 0.8</td>
<td>1.9 ± 0.3</td>
<td>9</td>
<td>6</td>
<td>0.889 ± 0.091</td>
<td>0.01939 ± 0.00394</td>
<td></td>
</tr>
<tr>
<td>5.4 ± 0.9</td>
<td>7.1 ± 0.9</td>
<td>11.1 ± 1.2</td>
<td>5.8 ± 0.9</td>
<td>7.1 ± 0.9</td>
<td>6.1 ± 0.9</td>
<td>6 ± 0.8</td>
<td>11.3 ± 1.2</td>
<td>0.6 ± 0.1</td>
<td>5.6 ± 0.8</td>
<td>6.1 ± 0.8</td>
<td>1.1 ± 0.3</td>
</tr>
</tbody>
</table>

Note: Values above in regular font correspond to cytochrome b (CYTB), and in bold below correspond to NADH dehydrogenase subunit 4 (ND4). N = number of samples; h = number of haplotypes; Hdh = haplotype diversity; π = nucleotide diversity.
gence events during the late Pleistocene. Extremely reduced (D = 0.036; Fig. S7). Their niches are not equivalent (p > 0.05; Fig. S7).

3.4. Niche overlap

Niche overlap between Moroccan/Mauritanian lineages is extremely reduced (D = 0.036; Fig. S7). Their niches are not equivalent (Equivalency test, p < 0.05; Fig. S7) although a similar pattern of habitat selection between lineages cannot be rejected (Similarity test, p > 0.05; Fig. S7).

4. Discussion

The history of North Africa is rich in climatic and geological changes with significant impact on local biodiversity, making it hard to link timing of diversification with particular past events. Nevertheless, by combining geological/climatic knowledge and phylogenetic data, it is possible to approximate the processes that shaped Sahara colonization and the diversity of species in and around it. The models here presented include only climatic conditions, as no accurate geological data are available to past conditions for such large area. However, climatic variables should also be a reasonable surrogate for past habitat changes, as the distribution of desert species (particularly ectotherms) is strongly dependent on temperature and rainfall (Ward, 2009).

4.1. Migration corridors across the Sahara

Trans-Saharan corridors (tS-corridors) can be more or less permanent (i.e. active ecological corridors), with the coastal ones expected to be more permanent (Brito et al., 2014). The climatically stable areas mostly fit this prediction, with the Atlantic Sahara suggested even as potential refugium area. In a simple scenario, stronger gene flow along the Atlantic than from the coast to the Algerian Mountains would then be expected, but the distribution of genetic variability seems to indicate the opposite.

The same mitochondrial haplotype was found in E-Tunisia and Hoggar (Figs. 2 and 4). This contrasts with the genetic differentiation commonly observed in populations from Sahara mountains, illustrated by the high rate of endemism (more than half of Sahara-Sahel vertebrate endemics; Brito et al., 2014. But see Nicolas et al., 2015). The species’ broad niche can partly explain it, and this shows that the predicted level of connection of northern areas with south Algeria Mountains seems therefore sufficient to allow the species’ dispersal. According to the ecological models, the link is more likely to occur under current conditions, which suggests a very recent (re-) colonization or secondary contact. The most probable path to southern Algerian mountains seems to be through Fezzan region in Libya, but a tS-corridor from there to the Sahel seems unlikely. A putative tS-corridor seems more likely in Eastern Sahara, active under LGM-like conditions (CCSM and MIROC, Fig. S3), linking the Mediterranean to Gilf Kebir, Jebel Uweinat, Ennedi/Borkou, and possibly Tibesti and Marra mountains. However, identification of suitable areas in the Eastern Sahara is affected by the low availability of precise observational data (Fig. 1), resulting in less clear distribution patterns.

The Atlantic coast was the sole permanent tS-corridor for mesic species well supported and it seems to have been the most climatically stable dispersal path across the Sahara. The split of the Moroccan and Mauritanian lineages seems to contradict this pattern, but these lineages might have been separated by the Tamanrasset palaeoriver basin, which opened on the Atlantic coast in the north of present-day Mauritania (Fig. 4; Skonieczny et al., 2015). This river system was activated in the humid periods of the Sahara, which seem to have occurred cyclically since before the Pleocene (Rohling et al., 2015). With a drainage area comparable to the ones of Niger or Zambezi rivers (Vorosmarty et al., 2000), it could represent a barrier during humid
Fig. 2. Phylogeographic relationships among the lineages of *P. schokari* and *P. aegyptius*, and their position within the genus (RAxML and MRBAYES, dataset 1, all markers). The colours in the tree, map, and haplotype networks represent the lineages identified in the phylogenetic tree analysis. The RAxML tree is represented; posterior probability (pp) and bootstrap values (bss) are indicated as pp/bss; asterisks represent values of 100; values below 95%pp or 70bss are not shown, except for the node in light grey. Haplotype networks were calculated using the Median-Joining algorithm in NETWORK; small black dots represent mutation steps; red diamonds represent predicted missing haplotypes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
phases, as well as constitute an unsuitable sandy expanse during warm ones. However, it is still possible the observed pattern is a coincidence, thus phylogeographic studies with other taxa are needed to verify this hypothesis. Continued isolation may have subsequently led to some niche divergence among lineages, although the observed pattern could not be dissociated from the differences in local climatic conditions (Fig. S7). Predicted climatic connectivity may have decreased periodically (Figs. S3 and S4), but the fact that present conditions predict the narrowest suitable area and still allow species persistence along the coast seems to indicate that, at least climatically, some gene flow is always possible. Individual Moroccan-Mauritanian models predicted that lineage contact is most likely under LGM-like conditions (Figs. S5 and S6), similar to the Eastern Sahara corridor. The Red Sea coast was expected to be clearly represented in the paleomodels, given that the topography allows altitudinal range shifts according to climate fluctuations (Messerli and Winiger, 1992). Contrary to previously hypothesised, the Red Sea S-t-corridor seems not permanent for mesic species, especially during hot and dry phases similar to LIG (Fig. S3). Nevertheless, pixel size could be constraining model sensitivity.

All considered, it seems that the Atlantic coast works as a corridor when considering the climatic cycles altogether, but certain portions (e.g. Tamanrasset river basin) may work as filters depending on the climatic phase. For example, the predicted climatic suitability of northern costal Mauritania was not mirrored by availability of observations/samples, despite considerable local sampling efforts during the last decade (Sow et al., 2014), and is probably related to the unsuitable character of the wide sandy habitats present along the coast and Azeffâl dunes, in the Tamanrasset palaeoriver basin. However, Contact between Mauritanian and Moroccan populations could also be allowed through the Adrar Atar in Mauritania and Adrar Soutouf in Morocco. Assessing the detailed location of the corridors and the role of the landscape in gene-flow would however need further studies with fast-evolving markers (e.g. Velo-Antón et al., 2014) and more taxa. Still, the likelihood of a climatic corridor along the Atlantic coast is shown, and we thus expect similar genetic structure patterns to occur in other species with likely similar dispersal capabilities, like the Moila Snake (Malpolon monilis) or Saw-scaled Viper (Echis pyramidum; Pook et al., 2008). Stronger structuring is expected for more mesic species or with less dispersal capabilities.

4.2. Phylogenetic relationships

The status of sister species is supported for P. schokari and P. aegyptius, in agreement with Rato et al. (2007), although differentiation in the nuclear markers was not complete (Fig. 2). Although c-mos exhibits a typical pattern of shared ancestral polymorphism where the most common haplotype in the two species is still the ancestral haplotype, RAG2 exhibits clear separation between species, so the nuclear data do not contradict reproductive isolation between these two species. The distance found between the five P. schokari mtDNA lineages, compared in Table 1, was substantial (4.5-7.1%), and while not uncommon in reptiles, similarly deep evolutionary lineages were used to request taxonomy revisions in other species (e.g. Vipera latastei; Velo-Antón et al., 2012). Still, higher resolution nuclear markers or increased genomic coverage (e.g. Velo-Antón et al., 2014) would be needed to support it and to resolve the phylogenetic relations among lineages. Additional sampling would also be necessary to define lineage distributional limits and assess gene flow in contact zones.

4.3. Biogeography and diversification

Psammophis schokari and P. aegyptius seem to have diverged around 10 Ma (Fig. 3). The latter exists only in Africa, but P. schokari ranges from the Atlantic coast to India, and the earliest-splitting lineage identified here is found in Oman. If the African diversity is indeed embedded within Asian diversity, an Asian origin of P. schokari, for instance through vicariance from African populations (the ancestor of P. aegyptius), is plausible. However, clarification is dependent on covering sampling gaps, since an Oman lineage migration across the Red Sea cannot be ruled out, as exemplified by other faunal exchanges between Africa and Arabia during the progressive aridification of late Miocene (Metallinou et al., 2012; Smild et al., 2013; Tamar et al., 2016a). Ecological divergence into new more arid habitats could also have been involved, as suggested for other reptiles (Carranza et al., 2008, 2002; Metallinou et al., 2012; Pook et al., 2009; Tamar et al., 2016b).

The successive divergences among major lineages after the Messinian Salinity Crisis (Fig. 3) mirror the same diversification patterns in the Late Miocene and Pliocene found in many vertebrate and reptile groups (Carranza et al., 2008; Geniez and Arnold, 2006; Gonçalves et al., 2012; Wagner et al., 2011), and have been associated with the original onset of the Sahara desert. Lineage distributions are broadly coherent with the predicted climatically stable areas, which could suggest a vicariant effect of climate (Fig. 4). However, our models are based on Pleistocene conditions and do not cover the Pliocene. Still, dust flux records indicate that notwithstanding a slight increase in the last Ma, humid-arid cycles have been occurring similarly since 5 Ma (Trauth et al., 2009).

The roughly contemporaneous and independent divergences of North African sub-lineages after mid-Pleistocene (Fig. 3) may be linked to persistence and subsequent isolation in Quaternary climate refugia during humid-dry cycles. In Morocco for example, the sub-lineages are separated by mountain ranges (Fig. 2), which agrees with a break in climatically stable areas found in Global, Regional, and Moroccan models (Figs. 4 and S3-S5). This suggests a role for climate in the group diversification and matches the refugia patterns identified in the region (Husemann et al., 2014; Martinez-Freiría et al., 2017). Similarly, the decreased suitability predicted for north-eastern Algeria/Tunisia could be linked with the divergence of the two sub-lineages, a pattern also found in other taxa (Guiller and Madec, 2010; Husemann et al., 2014; Nicolas et al., 2015). In Mauritania, sporadic decreases in climatic connectivity or disruptive effects of rivers during humid phases (e.g. Dobigny et al., 2005) could help explain the distribution of the sub-lineages. However, given most of the phylogenetic signal is from mtDNA, effects of coalescence cannot be ruled out without employing multiple nuclear markers. The observed haplogroups in P. aegyptius are probably due to considerable geographic distance (isolation by distance) between sample clusters (Egypt and Niger).

4.4. Conclusions and future research

The work here presented contributes to a better understanding of how Saharan mesic corridors work according to an interplay of climatic fluctuations and geographic variability, and lays the ground for further hypothesis testing. Still, assessing how the shifts in the climatic conditions affect gene flow and evolution in other taxa is necessary to complete our understanding about the trans-Saharan corridor dynamics. Assessing the role of climate and landscape in gene-flow dynamics at local scales is also crucial. In addition to vicariance, ecological adaptation to different gradients of aridity (e.g. Tarentola along the Atlantic coast; Carranza et al., 2002) or habitat specialization (e.g. sympatric speciation in Jaculus; Boratsynski et al., 2014) can also lead to species diversification. We contributed to a better understanding of the history of P. schokari in North Africa, but still further sampling and additional nuclear markers are needed to understand the history of the species along its full distribution. Of particular importance are the role of the interplay of Arabia Peninsula and Africa, and the possibilities of secondary contacts between species and lineages, as well as their evolutionary and taxonomic status.
Fig. 3. Calibrated phylogenetic tree obtained from Dataset 2. Support values represent values from BEAST and RAXML (pp/bss); asterisks represent values of 100. Red dots mark the calibration constraints. Blue horizontal bars represent the 95% confidence intervals for age estimates. Detail of *P. schokari* and *P. aegyptius* on the left. Shadings represent (past to present): onset of aridification, oldest estimated Sahara, and youngest estimated Sahara. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Acknowledgements

Authors acknowledge BIODESERTS group members for their assistance during fieldwork. P. Lymberakis and A. Ohler provided access to samples; P. Campbell and R. Sindaco provided presence records for P. schokari. This work was funded by National Geographic Society (CRE 7629-04/8412-08), Mohamed bin Zayed Species Conservation Fund (11052709, 11052707, 11052499), Fundação para a Ciência e a Tecnologia (PTDC/BIA-BEC/099934/2008 and PTDC/BIA-BIC/2903/2012), and by ERDF funds through the Operational Programme for Competitiveness Factors - COMPETE (FCOMP-01-0124-FEDER-008917/028276). Individual support was given by FCT (contract IF/01425/2014, and grants SFRH/BD/78402/2011, SFRH/BPD/69857/2010) within QREN-POPH-T4.1 funded by ESF and Portuguese MEC. Logistic support for fieldwork was given by Pedro Santos Lda (Trimble GPS), Off Road Power Shop, P.N. Banc d’Arguin (Mauritania), and Ministère de l’Environnement et du Développement Durable of Mauritania.

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.2017.09.009.

References


