Impact of species delimitation and sampling on niche models and phylogeographical inference: A case study of the East African reed frog *Hyperolius substriatus* Ahl, 1931

Gabriela B. Bittencourt-Silva [a,*], Lucinda P. Lawson [b,c,d], Krystal A. Tolley [e,f], Daniel M. Portik [g], Christopher D. Barratt [a], Peter Nagel [a], Simon P. Loader [a,h,i]

[a] Department of Environmental Sciences, University of Basel, Basel 4056, Switzerland
[b] Committee on Evolutionary Biology, University of Chicago, 1025 E. 57th St., Culver Hall 402, Chicago, IL 60637, USA
[c] Field Museum of Natural History, 1400 S. Lake Shore Dr., Chicago, IL 60605, USA
[d] University of Cincinnati, 614 Rieveschl Hall, Cincinnati, OH 45221, USA
[e] South African National Biodiversity Institute, Rhodes Drive, Private Bag X7, Newlands, Claremont 7735, Cape Town, South Africa
[f] Department of Botany & Zoology, Stellenbosch University, Private Bag X1, Matieland 7602, Stellenbosch, South Africa
[g] Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, United States
[h] University of Roehampton, Department of Life Sciences, London SW15 4JD, United Kingdom
[i] Natural History Museum, Department of Life Sciences, London SW5 5BD, United Kingdom

ARTICLE INFO

ABSTRACT

Ecological niche models (ENMs) have been used in a wide range of ecological and evolutionary studies. In biogeographic studies these models have, among other things, helped in the discovery of new allopatric populations, and even new species. However, small sample sizes and questionable taxonomic delimitation can challenge models, often decreasing their accuracy. Herein we examine the sensitivity of ENMs to the addition of new, geographically isolated populations, and the impact of applying different taxonomic delimitations. The East African reed frog *Hyperolius substriatus* Ahl, 1931 was selected as a case study because it has been the subject of previous ENM predictions. Our results suggest that addition of new data and reanalysis of species lineages of *H. substriatus* improved our understanding of the evolutionary history of this group of frogs. ENMs provided robust predictions, even when some populations were deliberately excluded from the models. Splitting the lineages based on genetic relationships and analysing the ENMs separately provided insights about the biogeographical processes that led to the current distribution of *H. substriatus*.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Ecological Niche Models (ENMs; *sensu* Peterson and Soberón, 2012) have become an important part of phylogeographical studies (Chan et al., 2011). The approach provides a basis for estimating current, past, and future distributions of species. ENMs have provided major insights across many fields of biology, but their results can be difficult to interpret and quantifying their uncertainty difficult (e.g., Araújo and Guisan, 2006). Evaluating predictions of species distributions based on ENMs in African ecosystems, which are often characterised by incomplete information and uncertain evolutionary and taxonomic statuses, remains a major challenge. Identification of potential cryptic diversity is an important step toward

ameliorating these difficulties because it allows the correct identification of species lineages and thus, the appropriate selection of occurrence data to produce more accurate models.

Raxworthy et al. (2007) demonstrated how the accuracy of niche models can be improved by separating species lineages from species complexes. They generated ENMs for putative species of day geckos (*Phelsuma madagascariensis* species group) and concluded that ENMs based on the new species limits provided a better fit to the known distribution of the geckos than models treating all lineages as a single species. Likewise, de Aguiar et al. (2015) show that the addition of new samples from geographically distinct populations improved the precision of ENMs for leaf-nosed bats from South America.

The East African reed frog *Hyperolius substriatus* Ahl, 1931 occurs in moist and dry forests, from 300–2000 m above sea level (asl), across the eastern Afromontane region of East Africa. This

---

includes the Eastern Arc Mountains (EAM), Livingstone Mountains and Southern Highlands of Tanzania, and the highlands of Malawi and northern Mozambique (Fig. 1; IUCN SSC Amphibian Specialist Group, 2013). The degree of isolation or connectivity between forest-restricted communities throughout this region (such as those containing *H. substriatus*) is poorly known, but it is hypothesised that riverine corridors and watershed relationships may have maintained gene flow between some forest fragments in montane amphibians (Blackburn and Measey, 2009; Measey et al., 2007; Measey and Tolley, 2011).

In the past 10 million years, the landscape of East Africa has undergone substantial changes, including the uplift of the East African Rift System (Sepulchre et al., 2006), aridification (Maslin et al., 2014; Sepulchre et al., 2006), expansion of savannah ecosystems, and fragmentation of forests (deMenocal, 1995; Maslin et al., 2014). These changes were followed by large-scale shifts in faunal communities as species either adapted to newly emerging niches

of open savannah areas, or shifted distributions to maintain their ecological niches (moved into or out of the region). Communities of forest-restricted taxa became isolated in fragments scattered along the coast or at higher elevations (e.g., Maslin et al., 2014) as most lowland forested areas were converted to savannah.

Phylogenetic relationships and population structure within *Hyperolius substriatus* were previously investigated by Lawson (2013). She found that major and minor hydrological basins, which act as barriers to dispersal, best explained population-genetic differentiation, and population relatedness matched historical corridors of suitable habitat areas predicted from ENMs. Lawson (2013) identified three allopatric population lineages: northern/central Malawian Highlands, EAM, and southern Malawian Highlands. Populations from Mozambique (viz., Mt. Mabu, Mt. Namuli and Serra Jeci; Bayliss et al., 2014; Portik et al., 2013), found after the completion of Lawson's (2013) analyses, were predicted to cluster with populations from either the EAM or the southern
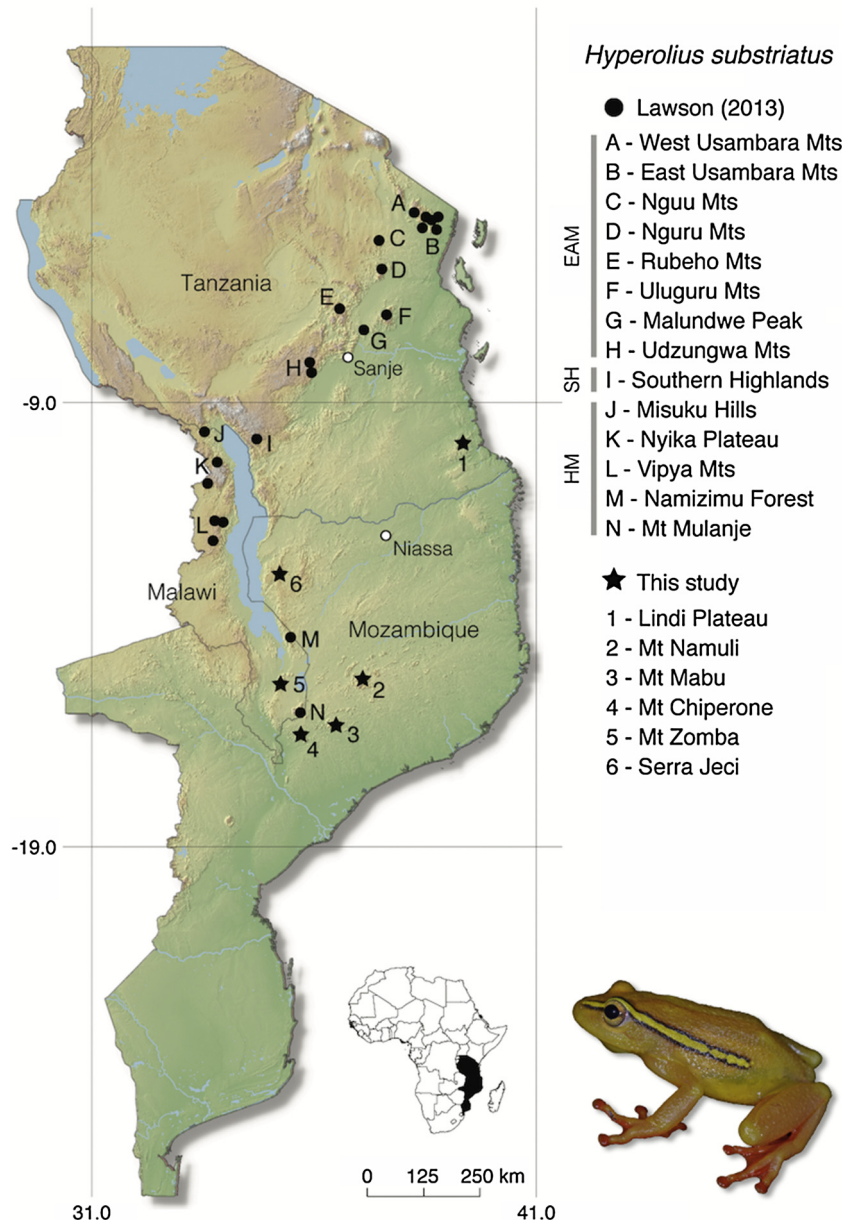

Fig. 1. Distribution map of *Hyperolius substriatus*. Black circles are populations included in Lawson's (2013) study. Stars are newly discovered populations. White circles are populations recorded but lacking molecular data. Eastern Arc Mountains (EAM); Southern Highlands and Livingstone Mountains (SH); Highlands of Malawi (HM).

Malawi highlands because they are found in the same major hydrological basin.

We investigated the impact of expanded phylogenetic sampling and improved phylogenetic resolution on ENM predictions. The recent discovery of populations of *H. substriatus* outside the previously described range offers a unique opportunity to assess the robustness of ENM predictions by examining the effect of inclusion of new locality data on the models. We also examined the impact of different species-level taxonomies on ENMs, given that the genetic variation uncovered in phylogeographic studies can be interpreted using different criteria. We examined ENMs generated from splitting the dataset into distinct species to investigate whether these allopatric lineages occupy similar niches (niche conservatism).

Our overall aim was to examine the sensitivity of ENMs to sample completeness and to species recognition, and thus their impact on phylogeographic studies. This may be particularly relevant in ecosystems in which there is great uncertainty in the genetic continuity of populations, such as in a tropical environment in which sampling may be incomplete and regionally biased by international borders that affect collecting efforts.

2. Material and methods

2.1. Samples and sequencing

Our sampling includes 28 geographic sites of *Hyperolius substriatus,* six of which are newly discovered. Four of the new sites are from northern Mozambique (Mts. Chiperone, Mabu, Namuli, and Serra Jeci), one from southern Malawi (Mt. Zomba), and one from coastal Tanzania (Lindi Plateau). Locality data, voucher numbers, and GenBank accession numbers are presented in Table A.1 in Supporting Information.

We extracted and sequenced DNA from liver and muscle of freshly collected specimens preserved in 95% ethanol following Bittencourt-Silva et al. (2016). To compare our newly collected samples with those presented in Lawson (2013), we sequenced the entire mitochondrial NADH dehydrogenase subunit 2 gene (*ND2*), and the nuclear genes pro-opiomelanocortin (*POMC*), and cellular myelocytomatosis proto-oncogene (*C-myc*). We also sequenced part of the mitochondrial 16S rRNA gene (*16S*) with the primers and PCR conditions described in Bittencourt-Silva et al. (2016). Summary statistics of selected molecular markers are shown in Table 1.

2.2. Data matrix and alignment

We combined sequences of *Hyperolius substriatus* from Lawson's (2010, 2013) studies, available on GenBank, with our newly sequenced samples. The total number of sequences included in the matrix per marker was 152 (*ND2*: 1152 bp), 155 (*POMC*: 628 bp), 152 (*C-myc*: 1303 bp), and 22 (*16S*: 527 bp). Sequences were aligned in Geneious v.7.1 (Kearse et al., 2012) using MAFFT v.7 (Katoh and Standley, 2013), applying the algorithm E-INS-i, and subsequently aligned by eye. The combined dataset represents 172 individuals (3083 bp) of *H. substriatus*, plus an outgroup (*H.*

mitchelli), based on the hypothetical relationships between these taxa and the availability of sequences on GenBank.

2.3. Phylogenetic analyses

Phylogenetic relationships were inferred for mitochondrial and nuclear genes with maximum likelihood (ML) and Bayesian inference (BI) approaches. We used PartitionFinder v.1.1.1 (Lanfear et al., 2012) to select the best-fit models of nucleotide substitution and partition schemes with the Bayesian information criterion (BIC; Table B.1). ML analyses were performed with RAxML v.8 (Stamatakis, 2014) and the GTR+GAMMA model, and support values were estimated using non-parametric bootstrap (auto-MRE). BI analyses were performed in BEAST v.2.4.1 (Bouckaert et al., 2014) in which time-calibrated trees were inferred based on estimated substitution rates: $16S = 0.00277$/lineage/my (Lemmon et al., 2007), $ND2 = 0.00957$/lineage/my (Crawford, 2003), *C-myc* $= 0.0006334$/lineage/my and $POMC = 0.000721$/lineage/my (Lawson, 2010). A strict clock model and a constant population size coalescent tree prior were used. The MCMC chains were run for 50 million generations and sampled at every 5000th generation; the first 10% of the sampled trees were discarded as burn-in. A species tree was estimated with a multilocus coalescent approach implemented in ∗BEAST v.2.4.1. (Bouckaert et al., 2014) to account for incomplete lineage sorting and to avoid overestimation of divergence times. ∗BEAST was run for 100 million generations, sampling every 1000th generation with the same substitution models, strict clocks, and rates as in the BEAST analyses. A Yule species tree with piecewise lineage and constant root prior was used with a random starting tree. Convergences for both the BEAST and ∗BEAST analyses were examined by checking the effective sample size (ESS) values in Tracer v.1.6 (Rambaut et al., 2014) using a threshold of >200. The program TreeAnnotator (Bouckaert et al., 2014) was used to create the maximum clade credibility (MCC) tree for each analysis.

2.4. Identification of operational taxonomic units (OTUs)

We used Lawson's (2013) three allopatric units, which were supported in our analyses, as a basis for lineage delimitation. We calculated the inter-lineage average and minimum uncorrected pairwise distance (p-distance) between 16S sequences with MEGA v.7 (Kumar et al., 2016). Lawson (2013) included populations from the Udzungwa Mountains in the EAM Lineage (named here Lineage 3), despite low support for association of the included samples. Given the high number of endemic species found on this mountain block (e.g., Menegon and Salvidio, 2005; Rovero et al., 2009), and the large genetic distances when compared to populations from other areas, we consider the populations from the Udzungwa Mountains to be a separate historic unit. Nevertheless, because of the reduced number of locality points compared to other groups, we grouped this unit with Lineage 3 for ENMs and niche analyses.

2.5. Ecological niche models

We generated ENMs for the three main lineages recovered by our species tree to determine whether there is a separation of ecological niches among them. For comparison, a niche model using all samples ("lumped") was also generated. ENMs were generated using present and past (Last Glacial Maximum; LGM and Mid-Holocene) climatic conditions to account for possible past connections between current forest fragments. Nine bioclimatic variables were downloaded from the WorldClim database (Global Climate Model: CCSM4; Hijmans et al., 2005) for current climate and Mid-Holocene (*ca.* 1-km resolution), and LGM (*ca.* 5-km resolution). The selection of variables follows Lawson (2013). ENMs were generated in MaxEnt v.3.3.3k (Phillips et al., 2006) with the logistic

Table 1
Summary of variable and constant sites of selected molecular markers.

Locus	N° sites	N° variable sites	N° parsimony-informative sites
16S	527	53	33
ND2	1152	142	242
C-myc	1303	48	48
POMC	628	20	43

outputs and the default settings from the program. The models were validated using a k-fold cross-validation algorithm to account for the small sample size of each lineage (Pearson et al., 2007). The area of analysis was limited to Tanzania, Malawi, and Mozambique. To facilitate comparison of models, we created binary presence/absence maps by applying a conservative threshold of the minimum training presence (Pearson et al., 2007).

Hyperolius substriatus populations have also been found at Sanje in the northeastern end of the Udzungwa Mountains (M. Menegon, pers. comm.) and in the Niassa Game Reserve of northern Mozambique (Branch, 2004). However, given the lack of molecular data, and hence the uncertainty regarding their phylogenetic position, these locality records were not included in the niche analyses.

2.6. Test for niche conservatism/divergence

The three lineages of *H. substriatus* have allopatric distributions, which could lead to misleading conclusions about niche divergence simply because of disjunct geographic ranges (i.e., spatial autocorrelation). To account for this issue, we conducted a background test to determine whether the ENMs of two disjunct populations are more or less similar than expected by chance, given the geographical region where they occur (Warren et al., 2008, 2010). For each lineage, a minimum convex polygon of the occurrence area was created and used as "background area" (i.e., environmental background where the species occurs). Given the proximity and small number of points (especially Lineages 1 and 3), a buffer zone of 10 km was created around each point before drawing the polygons. The population from Lindi in the coastal region of Tanzania occurs in an isolated patch of coastal forest considered to be a subcentre of endemism containing about 30 km² of forest (Burgess et al., 2007). This locality was excluded from the polygon because its inclusion would add a large area of habitat unsuitable (i.e., the surrounding *miombo* woodland) for *H. substriatus*.

To account for niche similarities among lineages, we calculated niche overlap between each pair of lineages with Schoener's *D* metric. If the observed value of the niche similarity among lineages (Schoener's *D*) is significantly higher (niche conservatism) or lower (niche divergence) than expected from the null distribution, the null hypothesis is rejected (i.e., niche conservatism/divergence between lineages is no more than expected based on the availability of habitat). For comparison, the test was conducted in two ways, (i) using continuous probabilities of habitat suitability generated by the ENMs, and (ii) applying a minimum training presence threshold (MTP). When this threshold is applied, the minimum value for habitat suitability (presence) corresponds to the lowest occurrence point, and every cell with a greater or equal value is treated as suitable (presence). Hence, it would be expected that the background area is less likely to include unsuitable areas for *Hyperolius substriatus*. Both niche-similarity analyses and background tests were conducted using ENMTools (Warren et al., 2010).

3. Results

3.1. Phylogenetic analyses

The topology of our estimated multilocus coalescent species tree agrees with the topology presented by Lawson (2013) with respect to the position of the three main lineages: (1) northern/central Malawian Highlands; (2) southern Malawian Highlands; and (3) Eastern Arc Mountain (i.e., Lineage 3 + Udzungwa; Fig. 1). However, the support for the sister-group relationship of Udzungwa Mountains and the rest of the EAM was weak (∗BEAST) or unresolved (RAxML) in both our study and that of Lawson

(2013). Herein, we consider the populations from the Udzungwa Mountains to be a separate lineage, and the remaining populations of the Eastern Arc Mountains to be part of Lineage 3. In the ∗BEAST tree, the six newly discovered populations of *Hyperolius substriatus* group within Lineage 2. The estimated divergence time between Lineage 1 and the other two lineages is about 6 mya (million years ago) in the late Miocene, and the split between Lineage 2 and 3 is estimated to have occurred in the late Pliocene about 3 mya. Tree topologies from multilocus, nuclear, and mitochondrial genes are provided in the supplementary material (Fig. A.1).

Both multilocus and mitochondrial trees have two sympatric populations in Serra Jeci (Fig. B.1)—one groups with populations from southern Malawi (Serra Jeci 1), and the other with populations from Lindi (Serra Jeci 2). The *16S* p-distance between Serra Jeci 1 (including the Southern Highlands and southern Malawi lineage) and Serra Jeci 2 (including Lindi) is 3.8%, whereas the difference between the latter and the lineage comprising Mts. Mabu, Namuli, and Chiperone (Mozambican lineage) is only 1.6%. Despite the short distance separating them, the difference between the lineage including Mt. Mulanje and the Mozambican lineage is more than twice (4.3%) the one observed between the latter and the Serra Jeci 2 + Lindi. Fig. B.1 shows the topology of the mitochondrial tree and the average *16S* p-distance between lineages.

3.2. Ecological niche models

Mean temperature of coldest quarter is the main constraint for Lineages 1, 2, and the "lumped" model, whereas for Lineage 3, this variable is the second most important. Precipitation of driest quarter is the main constraint for Lineage 3. The AUC values and the contributions of the bioclimatic variables that contributed the most to each model are summarised in Table B.2. Fig. 3 shows the ENMs for the "lumped" and separated lineages of *H. substriatus*.

If we consider *Hyperolius substriatus* to be one species (i.e., "lumped"), the ENM predicts the coastal region (Lindi) to be suitable; this area was not predicted in Lawson's (2013) ENM for current climate (Fig. 4 in Lawson 2013). If the Lindi point is removed, the area still is predicted as suitable by our new models, suggesting that some of the other newly added populations share some ecological niche characteristics with the coastal population and are contributing to the accuracy of the ENM.

As expected, the area of suitable habitat for *Hyperolius substriatus* has diminished substantially since the LGM (*ca.* 22 ka) in all models, especially for Lineages 1 and 2. When considering the ENMs of separate lineages, the connection between the isolated coastal population and other populations most likely occurred through the past distribution of the coastal forest in northern Mozambique (Coastal Forest belt; *sensu* Burgess et al., 2000), which extended inland toward the southern highlands of Malawi. Alternately, the "lumped" ENM suggests a historical dispersal corridor between the coastal region and the Udzungwa Mountains.

3.3. Test for niche conservatism/divergence

The background test showed sensitivity to the threshold applied (Fig. 4), especially when we compared niches of Lineages 1 and 3. When we used the continuous probability of habitat suitability, the results were non-significant (i.e., the similarity/difference between lineages is no more than expected based on the habitats in which they occur). However, when the minimum-presence threshold is applied, the results are significant, and the observed niche similarity is greater than the null distribution, indicating niche conservatism (Fig. 4d). In the case of Lineages 1 and 2, and Lineages 2 and 3, with and without the threshold (Fig. 4), niches are more similar (niche conservatism) than expected by chance.
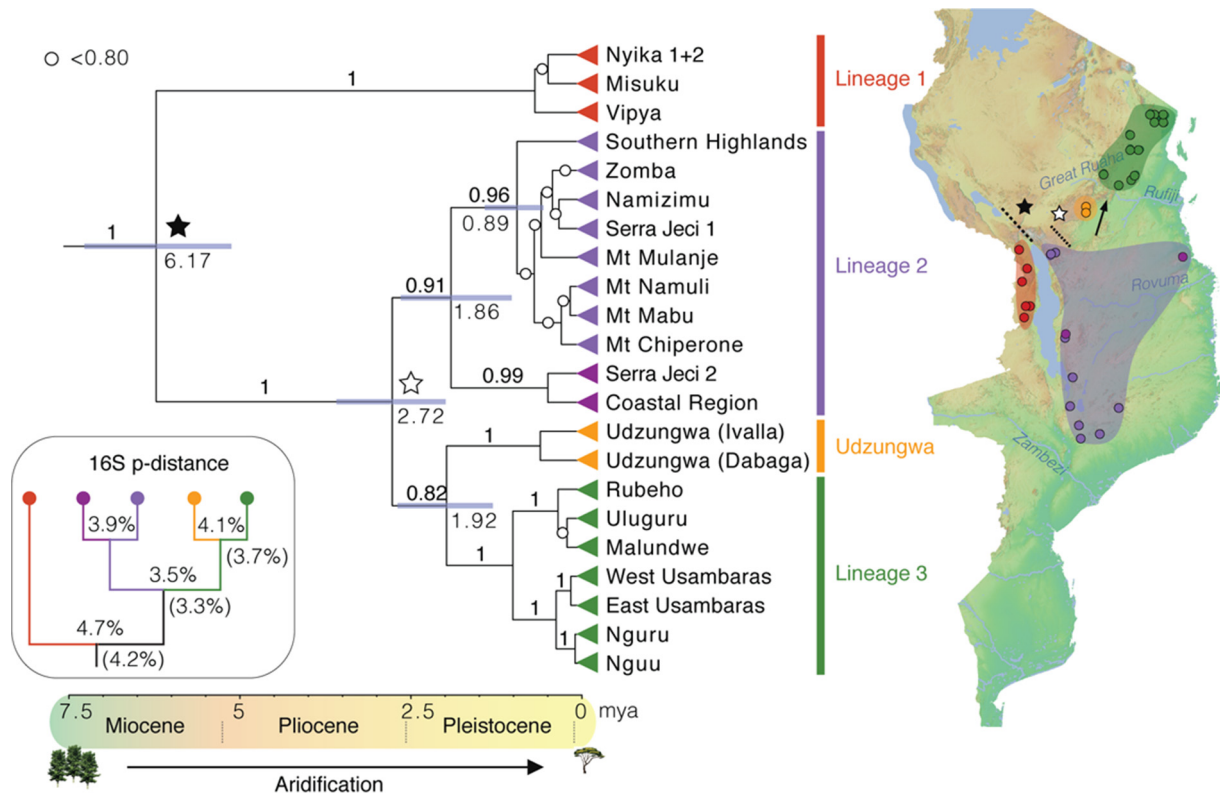
Fig. 2. StarBEAST (*BEAST) gene tree of populations of *Hyperolius substriatus* with branch lengths proportional to time. Values above the nodes are posterior probabilities; below are ages in millions years. Terminals are colour-coded according to each inferred population lineage. (Refer to Methods section for further information on lineage identification.) Bottom-left box shows average and minimum (in parenthesis) inter-lineage pairwise distances between 16S sequences. Map shows the distribution of each lineage (colour-coded). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Ecological niche analyses

The newly discovered populations from northern Mozambique were found in areas predicted as suitable by Lawson's (2013) ENM. Lindi, where the coastal population was found, was not predicted to be suitable for *Hyperolius substriatus* in Lawson's current climate ENM, but in her predictions for the LGM it appears as suitable. The addition of the new populations improved the accuracy of both the "lumped" and independent the ENMs because, even excluding the population from Lindi, the ENMs can now predict the coastal area to be suitable for *H. substriatus*.

The finding of the new population from Lindi strengthens the idea that this area represents a sub-centre of endemism, which is defined by Burgess et al. (1998) as an area possessing at least 10 species of endemic plants, or three endemic vertebrates or more than three species of butterflies. This population is a relict that was probably connected to populations from northern Mozambique via the Coastal Forest belt (LGM model in Fig. 4). A combination of climatic fluctuations, human activity, and sea-level changes provide a plausible explanation for the fragmentation of the coastal populations (Burgess et al., 1998).

The population from Niassa, in northern Mozambique (Fig. 1), was excluded from this study because it lacks molecular data and thus, could not be assigned to any lineage. When all locality points were used for the ENM ("lumped"), Niassa is not shown to be a suitable area for *Hyperolius substriatus*. However, the ENMs for Lineage 2 predict that the Niassa area is suitable, both in past and current predictions; for this reason the Niassa population is expected to group with the other populations from Lineage 2, especially Lindi and Serra Jeci.

The predictions for past climate revealed another advantage of splitting the lineages before generating the ENMs. In the case of Lineage 2, a connection between populations from northern Mozambique and Lindi via the Coastal Forest belt may have existed during the LGM and later disappeared. This may explain why the coastal population is more closely related to the populations from Mozambique than to the more close-lying Uluguru Mountains. Two other species of frogs have similar distributions through the Coastal Forest belt, occurring from the northern coast of Mozambique (Taratibu) to the west, toward Mt. Mulanje, and some coastal forest fragments in eastern Zimbabwe (Bittencourt-Silva et al., 2016; Farooq et al., 2015). Therefore, we agree with Raxworthy et al. (2007) that the accuracy of niche models is affected by the species concept applied, given that phylogeographic studies deal with cases above and below species level. Documentation of genetic variation that is associated with locality data will produce more reliable ecological niche models.

Wiens and Graham (2005) considered niche conservatism in allopatric sister lineages as an important signal of allopatric speciation. This scenario involves a formerly widespread ancestor having had its distribution bisected (via some vicariant process), with each daughter lineage remaining ecologically similar. The ENM predictions for the split lineages (Lineages, 2, and 3 + Udzungwa) allowed us to compare niches among the lineages and to test whether they show signals of niche conservatism or divergence. As we identified potential niche conservatism in these allopatrically distributed lineages, it is likely that physical barriers underlie these divergences. Within Lineage 1, the intervening areas between populations are predicted as suitable for dispersal, suggesting that dispersal and gene flow may occur. However, some areas separating populations within Lineages 2 and 3, respectively,
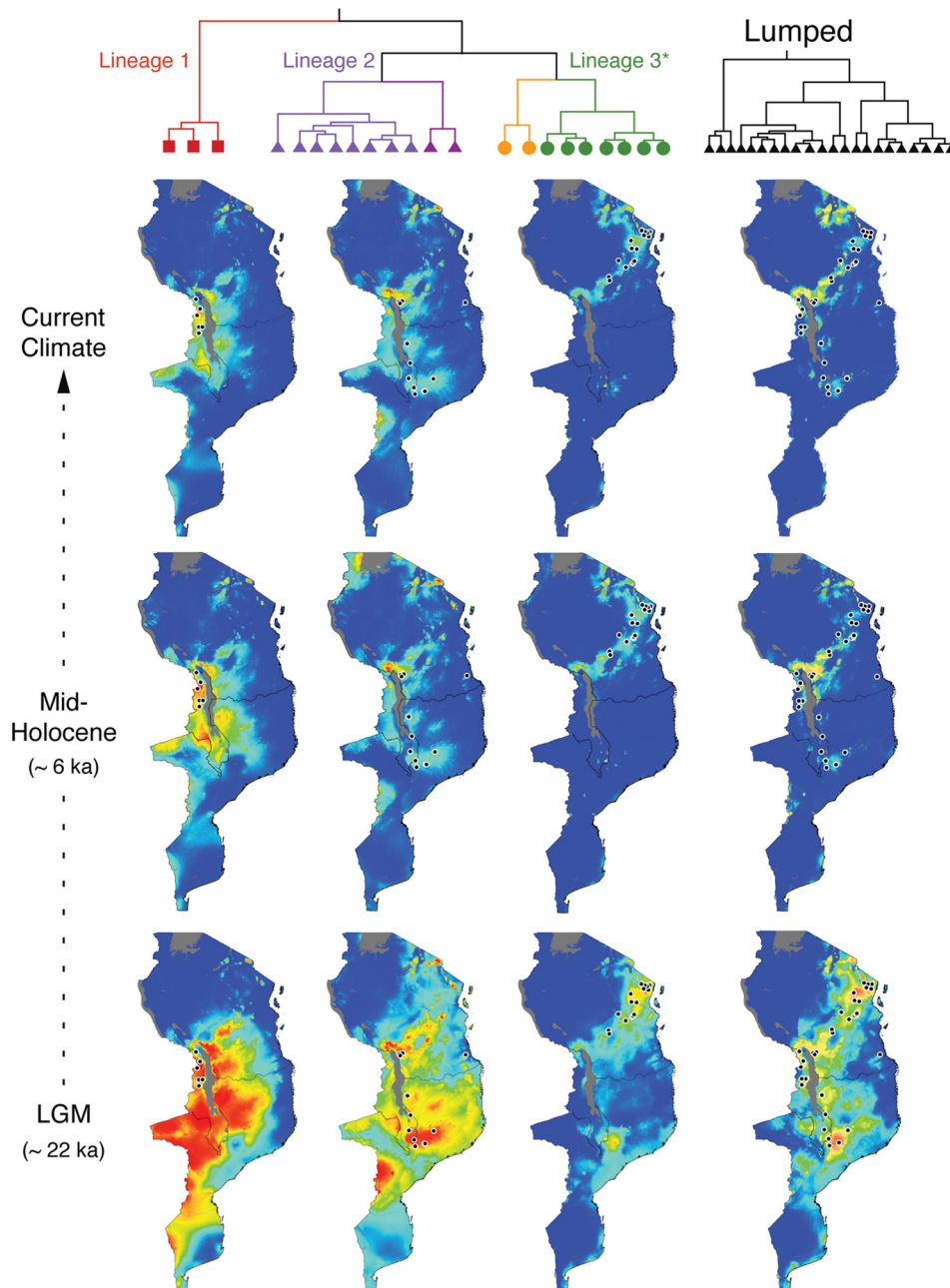
Fig. 3. Ecological niche models of *Hyperolius substriatus* using past and present bioclimatic data. Ecological niche models (ENMs) show the distribution of suitable habitats when treating each lineage from our phylogeny (Lineages 1, 2, and 3+ Udzungwa) as a different species lineage. The "lumped" model treats all populations as the same species. Warmer colours (toward red) indicate areas with suitable habitat for *H. substriatus*, and cooler colours (toward blue) indicate unsuitable habitats. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are predicted to be unsuitable (Fig. 3), indicating a climatic barrier for gene flow exists among some populations.

There are important caveats concerning ecological niche analyses. McCormack et al. (2010) suggested that using only ENMs to study niche divergence could be misleading because local-scale ecological differences might be overlooked. It also is possible that the environmental variables used for the ENMs did not capture the niche divergences between different lineages.

4.2. Genetic structure and biogeography of the three major population lineages

We identified three allopatrically distributed lineages of *Hyperolius substriatus* (Lineage 1, Lineage 2, and Lineage 3 + Udzungwa)

that occupy similar niches (niche conservatism). This observation supports the idea that barriers interrupted gene flow between them leading to an allopatric speciation process.

The genetic analyses show considerable structure within and among the three main lineages (Fig. 2). Inter-lineage differences observed for the 16S mitochondrial gene range from 3.5–4.5%, which is considered high among conspecific populations, though not substantial between sibling species of amphibians (Vences et al., 2005; Vieites et al., 2009). For *Hyperolius substriatus*, these distinct lineages may represent candidate species; however, additional lines of evidence (e.g., morphological, bioacoustical) are necessary to confirm their taxonomic status. This complex level of population structure, also noted by Lawson (2013), is similar to other amphibian species within the fragmented Eastern Arc
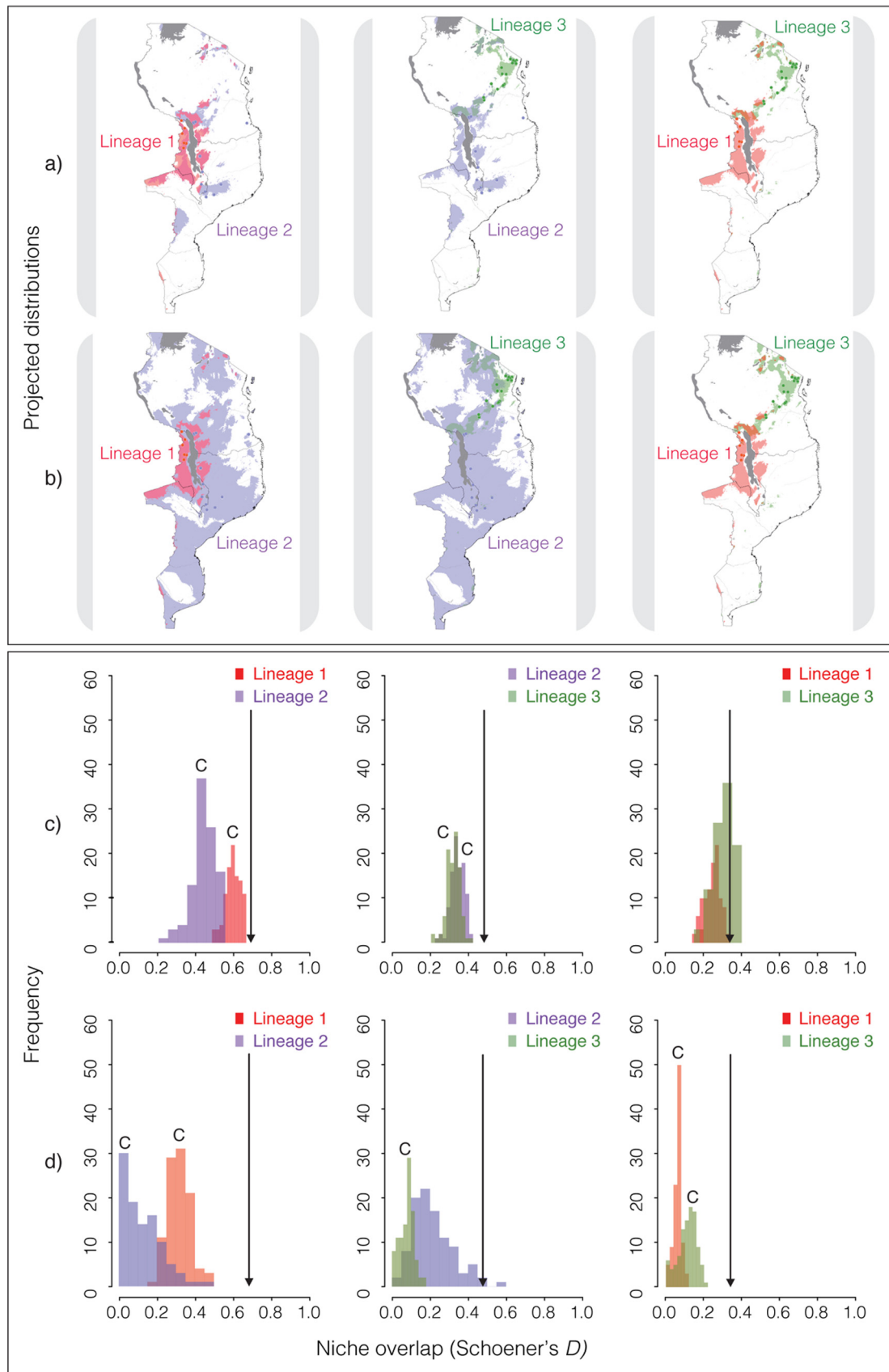
Fig. 4. Projected distribution of lineages and test of niche conservatism and divergence. Top box: shaded areas represent the projected distributions of lineage (a) without threshold, and (b) when applying the minimum training presence threshold (MTP). Lower box: Background test used to determine whether ENMs from any two lineages are more similar than expected by chance based on the geographical region in which they occur. (c) Using continuous measures of habitat suitability. (d) Applying MTP threshold. Niche conservatism ("C") is supported when the observed niche overlap (indicated by black arrow) is higher than expected under the null hypothesis.

Mountains (e.g. Burgess et al., 2007; Lawson et al., 2015; Loader et al., 2014).

The two sympatric populations from Serra Jeci, despite being morphologically indistinguishable, have a genetic signature of isolation (p-distance = 4%). In the mitochondrial tree, Serra Jeci 1 groups with populations from the Southern Highlands in Tanzania and southern Malawi, both of which are classified as montane forest-grassland mosaic (Fig. B.2; Burgess et al., 1998). In contrast, Serra Jeci 2 groups with Lindi, and with Mts. Mabu, Namuli, and Chiperone. These areas are encompassed by the Coastal Forest belt (Fig. B.2), which is composed of a different vegetation type from the EAM and the highlands of Malawi (Burgess et al., 1998). This difference in vegetation types was not fully explored here because modelling the niches of the two lineages separately would reduce the number of data points in each lineage and compromise the accuracy of the models. Targeted surveys could potentially solve this issue and should be considered for future studies with this group.

The estimations of divergence times among the three lineages correspond to two main vicariant events—viz., the southern rift formation and the aridification process in East Africa.

Southern rift formation – The divergence time between Lineage 1 (northern Malawi) and the other lineages coincides with the formation of the southern rift in northern Malawi; the rift is estimated to have started in this region about 8 Ma (Foster et al. 1997). This tectonic event affected the hydrological system of East Africa by reversing the flow direction of some rivers (Banister and Clarke, 1980; Stankiewicz and de Wit, 2006). Hence, rift formation is a plausible explanation for the cessation of gene flow between Lineage 1 and the other lineages, as it would have changed the boundaries of the major drainage basins and altered potential gene flow among populations. This geological event also is thought to have promoted cladogenesis in the Tanzanian mole rats *Fucomys* (Faulkes et al., 2010).

Aridification process – Climatic fluctuations also might have contributed to divergence of lineages in response to expansion of corridors of wetter/forested habitat between highland areas during warm/humid climatic cycles and contraction in cool/dry cycles (deMenocal, 1995). Variation in habitat connectivity may have compounded the effects of the rifting event, further isolating lineages. The estimated divergence time between the EAM lineage (Lineage 3 + Udzungwa) and Lineage 2, which includes the mountains of northern Mozambique, is about 3 Ma. This might have been driven by the gradual aridification of East Africa during the last 4.5 Myr (Sepulchre et al., 2006; Trauth et al., 2005). *Hyperolius substriatus* would have experienced a loss of mid- and low-elevation habitat as forests transitioned to open habitats (e.g., savannah and grasslands) during this period (Kissling et al., 2012). Similarly, Bryja et al. (2014) estimated the separation between the EAM lineage and the Mulanje + Mozambique lineage of the forest-dependent mountain rodent *Praomys* to have occurred in the early Pleistocene. The bat species complex *Rhinolophus hildebrandtii* also diverged at the same time (Taylor et al., 2012). Temporal congruence of divergence among these organisms may suggest a common biogeographic event driving these patterns. In contrast, some species such as the pygmy chameleons (*Rhampholeon* spp.) have been isolated on the Mozambican mountains since the mid-Miocene (Branch et al., 2014). Complex biogeographic patterns are to be expected given differences in vagility of organisms, and the many climatic fluctuations affecting East Africa's habitats (deMenocal, 1995; Maslin et al., 2014). Broader biogeographic studies focussed on multiple lineages will be required to clarify the biogeography of Mozambican mountains relative to other Eastern Afromontane areas.

4.3. Impact of waterways on gene flow

The role of rivers as dispersal barriers is controversial because studies of different faunal groups have contrasting results (e.g. Da Silva and Patton, 1998; Gascon et al., 2000). There are potentially contrasting patterns in this study—e.g., a river seems to impose a barrier to dispersal on the one hand, but on the other, a river seems to serve as a corridor connecting distant populations. In particular, the high genetic differentiation between the Udzungwa population (Dabaga and Ivalla) with respect to the EAM lineage (Lineage 3) suggests limited gene flow between them (p-distance between lineages = 4%; Fig. 2). The position of the Great Ruaha River may be a barrier hindering gene flow between these populations (Fig. 2). Stanley and Esselstyn (2010) suggested that the Great Ruaha separates two lineages of mouse shrew: *Myosorex geata* occurs in the EAM to the northeast of the river, and *M. kihaulei* in the Udzungwa and Southern Highlands to the southwest. If this river imposes a barrier between the Udzungwa to the south and the other mountains to the northeast, the population of *Hyperolius substriatus* from Sanje would be expected to group genetically with the other Udzungwa populations, despite the closer geographic proximity to Malundwe (Fig. 1).

Conversely, populations from Lineage 2, which are separated by approximately 500 km (Lindi-Serra Jeci), are distributed both north and south of the Rovuma River (Fig. 2). The genetically unsampled population from Niassa in northern Mozambique may represent a connection between the coastal region (Lindi) and Serra Jeci, near Lake Malawi; this awaits future testing. The riverine forest along the Rovuma River connecting Lindi to Serra Jeci might have facilitated dispersal, connecting these populations. Bryja et al. (2016) presents evidence of a connection between the East African coastal forest and the highlands in Malawi (Mt. Mulanje) possibly via riverine forest.

River width, flow speed, and seasonality are some of the factors that could determine whether the river is a barrier or a corridor for *Hyperolius substriatus*. Because these factors are affected by geological and climatic changes, reconstructing biogeographic history is a complex challenge. It is important to remember that tectonic events had considerable effects on the hydrological system of East Africa, when some rivers disappeared or had their flow direction reversed by geological processes (Banister and Clarke, 1980; Stankiewicz and de Wit, 2006). Our results show the differential impact of hydrological basins on phylogeographic patterns.

5. Conclusions

With the addition of locality data, we confirm predictions of previous niche models, thereby improving our understanding of population structure in *Hyperolius substriatus*. Modelling the niches for separated species lineages based on genetic relationships improved the accuracy of the models and provided new insights about the phylogeographical history of this frog.

Acknowledgments

121). The expedition to Mozambique in 2014 was funded by National Geographic Society (CRE9281-13) and the Competitive Program for Rated Researchers: National Research Foundation of South Africa (92776; both grants to KAT). Hanlie Engelbrecht, Werner Conradie, Michele Menegon and Cristóvão Nanvonamuquitxo provided field assistance and Bill Branch and Harith Farooq provided support for the 2014 expedition. We also thank Laura Nunes and Mark Wilkinson for their valuable comments on the manuscript.

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.2017.06.022.

References

de Aguiar, L.M.de.S., da Rosa, R.O.L., Jones, G., Machado, R.B., 2015. Effect of chronological addition of records to species distribution maps: the case of *Tonatia saurophila maresi* (Chiroptera, Phyllostomidae) in South America. Austral Ecol. 40, 836–844. http://dx.doi.org/10.1111/aec.12261.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688. http://dx.doi.org/10.1111/j.1365-2699.2006.01584.x.

Banister, K.E., Clarke, M.A., 1980. A revision of the large *Barbus* (Pisces, Cyprinidae) of Lake Malawi with a reconstruction of the history of the southern African Rift Valley lakes. J. Nat. Hist. 14, 483–542.

Bayliss, J., Timberlake, J., Branch, W., Bruessow, C., Collins, S., Congdon, C., Curran, M., de Sousa, C., Dowsett, R., Dowsett-Lemaire, F., Fishpool, L., Harris, T., Herrmann, E., Georgiadis, S., Kopp, M., Liggitt, B., Monadjem, A., Patel, H., Ribeiro, D., Spottiswoode, C., Taylor, P., Willcock, S., Smith, P., 2014. The discovery, biodiversity and conservation of Mabu forest – the largest medium-altitude rainforest in southern Africa. Oryx 48, 177–185. http://dx.doi.org/10.1017/S0030605313000720.

Bittencourt-Silva, G.B., Conradie, W., Siu-Ting, K., Tolley, K.A., Channing, A., Cunningham, M., Farooq, H.M., Menegon, M., Loader, S.P., 2016. The phylogenetic position and diversity of the enigmatic mongrel frog *Nothophryne* Poynton, 1963 (Amphibia, Anura). Mol. Phylogenet. Evol. 99, 89–102. http://dx.doi.org/10.1016/j.ympev.2016.03.021.

Blackburn, D.C., Measey, G.J., 2009. Dispersal to or from an African biodiversity hotspot? Mol. Ecol. 18, 1904–1915. http://dx.doi.org/10.1111/j.1365-294X.2009.04156.x.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. http://dx.doi.org/10.1371/journal.pcbi.1003537.

Branch, W.R., 2004. Herpetological Survey of the Niassa Game Reserve (Article). Northern Mozambique, Sociedade para a Gestão e Desenvolvimento da Reserva do Niassa.

Branch, W.R., Bayliss, J., Tolley, K.A., 2014. Pygmy chameleons of the *Rhampholeon platyceps* complex (Squamata: Chamaeleonidae): description of four new species from isolated "sky islands" of northern Mozambique. Zootaxa 3814, 1–36.

Bryja, J., Mikula, O., Patzenhauerová, H., Oguge, N.O., Šumbera, R., Verheyen, E., 2014. The role of dispersal and vicariance in the Pleistocene history of an East African mountain rodent, *Praomys delectorum*. J. Biogeogr. 41, 196–208. http://dx.doi.org/10.1111/jbi.12195.

Bryja, J., Šumbera, R., Kerbis Peterhans, J.C., Aghová, T., Bryjová, A., Mikula, O., Nicolas, V., Denys, C., Verheyen, E., 2016. Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the evolution of African forests since late Miocene. J. Biogeogr. http://dx.doi.org/10.1111/jbi.12890.

Burgess, N.D., Butynski, T.M., Cordeiro, N.J., Doggart, N.H., Fjeldså, J., Howell, K.M., Kilahama, F.B., Loader, S.P., Lovett, J.C., Mbilinyi, B., Menegon, M., Moyer, D.C., Nashanda, E., Perkin, A., Rovero, F., Stanley, W.T., Stuart, S.N., 2007. The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. Biol. Conserv. 134, 209–231. http://dx.doi.org/10.1016/j.biocon.2006.08.015.

Burgess, N.D., Clarke, G.P., et al., 2000. Coastal Forests of Eastern Africa. IUCN.

Burgess, N.D., Clarke, G.P., Rodgers, W.A., 1998. Coastal forests of eastern Africa: status, endemism patterns and their potential causes. Biol. J. Linn. Soc. 64, 337–367. http://dx.doi.org/10.1111/j.1095-8312.1998.tb00337.x.

Chan, L.M., Brown, J.L., Yoder, A.D., 2011. Integrating statistical genetic and geospatial methods brings new power to phylogeography. Mol. Phylogenet. Evol. 59, 523–537. http://dx.doi.org/10.1016/j.ympev.2011.01.020.

Crawford, A.J., 2003. Relative rates of nucleotide substitution in frogs. J. Mol. Evol. 57, 636–641. http://dx.doi.org/10.1007/s00239-003-2513-7.

Da Silva, M.N.F., Patton, J.L., 1998. Molecular phylogeography and the evolution and conservation of Amazonian mammals. Mol. Ecol. 7, 475–486. http://dx.doi.org/10.1046/j.1365-294x.1998.00276.x.

deMenocal, P.B., 1995. Plio-Pleistocene African climate. Science 270, 53–59. http://dx.doi.org/10.1126/science.270.5233.53.

Farooq, H., Liedtke, H.C., Bittencourt-Silva, G., Conradie, W., Loader, S.P., 2015. The distribution of *Mertensophryne anotis* with a new record in Northern Mozambique. Herpetol. Notes 8, 305–307.

Faulkes, C.G., Mgode, G.F., Le Comber, S.C., Bennett, N.C., 2010. Cladogenesis and endemism in Tanzanian mole-rats, genus *Fukomys*: (Rodentia Bathyergidae): a role for tectonics? Biol. J. Linn. Soc. 100, 337–352. http://dx.doi.org/10.1111/j.1095-8312.2010.01418.x.

Gascon, C., Malcolm, J.R., Patton, J.L., da Silva, M.N.F., Bogart, J.P., Lougheed, S.C., Peres, C.A., Neckel, S., Boag, P.T., 2000. Riverine barriers and the geographic distribution of Amazonian species. Proc. Natl. Acad. Sci. 97, 13672–13677. http://dx.doi.org/10.1073/pnas.230136397.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

IUCN SSC Amphibian Specialist Group, 2013. Hyperolius substriatus. The IUCN Red List of Threatened Species [WWW Document]. http://dx.doi.org/10.2305/IUCN.UK.2013-2.RLTS.T136152A18367261.en (accessed 10.10.16).

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. http://dx.doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.

Kissling, W.D., Eiserhardt, W.L., Baker, W.J., Borchsenius, F., Couvreur, T.L.P., Balslev, H., Svenning, J.-C., 2012. Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. Proc. Natl. Acad. Sci. 109, 7379–7384. http://dx.doi.org/10.1073/pnas.1120467109.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. http://dx.doi.org/10.1093/molbev/msw054.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. http://dx.doi.org/10.1093/molbev/mss020.

Lawson, L.P., 2013. Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the African spotted reed frog. Mol. Ecol. 22, 1947–1960. http://dx.doi.org/10.1111/mec.12229.

Lawson, L.P., 2010. The discordance of diversification: evolution in the tropical-montane frogs of the Eastern Arc Mountains of Tanzania. Mol. Ecol. 19, 4046–4060. http://dx.doi.org/10.1111/j.1365-294X.2010.04788.x.

Lawson, L.P., Bates, J.M., Menegon, M., Loader, S.P., 2015. Divergence at the edges: peripatric isolation in the montane spiny throated reed frog complex. BMC Evol. Biol. 15, 1–15. http://dx.doi.org/10.1186/s12862-015-0384-3.

Lemmon, E.M., Lemmon, A.R., Cannatella, D.C., 2007. Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). Evolution (N. Y). 61, 2086–2103. http://dx.doi.org/10.1111/j.1558-5646.2007.00181.x.

Loader, S.P., Ceccarelli, S.F., Menegon, M., Howell, K.M., Kassahun, R., Mengistu, A.A., Saber, S.A., Gebresenbet, F., de Sá, R., Davenport, T.R.B., Larson, J.G., Müller, H., Wilkinson, M., Gower, D.J., 2014. Persistence and stability of eastern Afromontane forests: evidence from brevicipitid frogs. J. Biogeogr. 41, 1781–1792. http://dx.doi.org/10.1111/jbi.12331.

Maslin, M.A., Brierley, C.M., Milner, A.M., Shultz, S., Trauth, M.H., Wilson, K.E., 2014. East African climate pulses and early human evolution. Quat. Sci. Rev. 101, 1–17. http://dx.doi.org/10.1016/j.quascirev.2014.06.012.

McCormack, J.E., Zellmer, A.J., Knowles, L.L., 2010. Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. Evolution (N. Y). 64, 1231–1244. http://dx.doi.org/10.1111/j.1558-5646.2009.00900.x.

Measey, G.J., Galbusera, P., Breyne, P., Matthysen, E., 2007. Gene flow in a direct-developing, leaf litter frog between isolated mountains in the Taita Hills. Kenya. Conserv. Genet. 8, 1177–1188. http://dx.doi.org/10.1007/s10592-006-9272-0.

Measey, G.J., Tolley, K.A., 2011. Sequential fragmentation of Pleistocene forests in an East Africa biodiversity hotspot: chameleons as a model to track forest history. PLoS ONE 6, 1–9. http://dx.doi.org/10.1371/journal.pone.0026606.

Menegon, M., Salvidio, S., 2005. Amphibian and reptile diversity in the southern Udzungwa scarp forest reserve, south-eastern Tanzania. In: African Biodiversity. Springer, US, Boston, MA, pp. 205–212. http://dx.doi.org/10.1007/0-387-24320-8_18.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117. http://dx.doi.org/10.1111/j.1365-2699.2006.01594.x.

Peterson, A.T., Soberón, J., 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. Nat. Conserv. 10, 102–107.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026.

Portik, D.M., Mulungu, E., Sequeira, D., McEntee, J.P., 2013. Herpetological surveys of the Serra Jeci and Namuli massifs, Mozambique, and an annotated checklist of the southern Afromontane archipelago. Herpetol. Rev. 44, 394–406.

Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2014. Tracer v1. 6 http://beast.bio.ed.ac.uk.

Raxworthy, C.J., Ingram, C.M., Rabibisoa, N., Pearson, R.G., 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (Phelsuma) from Madagascar. Syst. Biol. 56, 907–923. http://dx.doi.org/10.1080/10635150701775111.

Rovero, F., Marshall, A.R., Jones, T., Perkin, A., 2009. The primates of the Udzungwa Mountains: diversity, ecology and conservation. J. Anthropol. Sci. = Riv. di Antropol. JASS 87, 93–126.

Sepulchre, P., Ramstein, G., Fluteau, F., Schuster, M., Tiercelin, J.-J., Brunet, M., 2006. Tectonic uplift and Eastern Africa aridification. Science (80). 313, 1419–1423.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.

Stankiewicz, J., de Wit, M.J., 2006. A proposed drainage evolution model for Central Africa—Did the Congo flow east? J. African Earth Sci. 44, 75–84. http://dx.doi.org/10.1016/j.jafrearsci.2005.11.008.

Stanley, W.T., Esselstyn, J.A., 2010. Biogeography and diversity among montane populations of mouse shrew (Soricidae: Myosorex) in Tanzania. Biol. J. Linn. Soc. 100, 669–680. http://dx.doi.org/10.1111/j.1095-8312.2010.01448.x.

Taylor, P.J., Stoffberg, S., Monadjem, A., Schoeman, M.C., Bayliss, J., Cotterill, F.P.D., 2012. Four new bat species (Rhinolophus hildebrandtii complex) reflect Plio-Pleistocene divergence of dwarfs and giants across an Afromontane archipelago. PLoS ONE 7, 1–23. http://dx.doi.org/10.1371/journal.pone.0041744.

Trauth, M.H., Maslin, M.A., Deino, A., Strecker, M.R., 2005. Late Cenozoic moisture history of East Africa. Science (80). 309, 2051 LP–2053.

Vences, M., Thomas, M., van der Meijden, A., Chiari, Y., Vieites, D.R., 2005. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Front. Zool. 2, 1–12. http://dx.doi.org/10.1186/1742-9994-2-5.

Vieites, D.R., Wollenberg, K.C., Andreone, F., Köhler, J., Glaw, F., Vences, M., 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proc. Natl. Acad. Sci. 106, 8267–8272. http://dx.doi.org/10.1073/pnas.0810821106.

Warren, D.L., Glor, R.E., Turelli, M., 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography (Cop.) 33, 607–611. http://dx.doi.org/10.1111/j.1600-0587.2009.06142.x.

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62, 2868–2883. http://dx.doi.org/10.1111/j.1558-5646.2008.00482.x.

Wiens, J.J., Graham, C.H., 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. Annu. Rev. Ecol. Evol. Syst. 36, 519–539.