








ORIGINAL ARTICLE

Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians

Christopher D. Barratt^{1,2}  | Beryl A. Bwong^{1,3} | Robert Jehle⁴  |
 H. Christoph Liedtke^{1,5}  | Peter Nagel¹ | Renske E. Onstein²  | Daniel M. Portik^{6,7}  |
 Jeffrey W. Streicher⁸  | Simon P. Loader^{1,8} 

¹Department of Environmental Sciences, University of Basel, Basel, Switzerland

²German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

³Herpetology Section, National Museums of Kenya, Nairobi, Kenya

⁴School of Environment and Life Sciences, University of Salford, Salford, UK

⁵Ecology, Evolution and Developmental Group, Department of Wetland Ecology, Estación Biológica de Doñana (CSIC), Sevilla, Spain

⁶Department of Biology, The University of Texas at Arlington, Arlington, Texas

⁷Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona

⁸Department of Life Sciences, Natural History Museum, London, UK

Correspondence

Christopher D. Barratt and Simon P. Loader, Department of Environmental Sciences, University of Basel, Basel, Switzerland. Emails: c.d.barratt@gmail.com (CDB); simon.loader@nhm.ac.uk (SPL)

Funding information

Humer Foundation for Academic Talent; Freiwilige Akademische Gesellschaft Basel

Abstract

High-throughput sequencing data have greatly improved our ability to understand the processes that contribute to current biodiversity patterns. The “vanishing refuge” diversification model is speculated for the coastal forests of eastern Africa, whereby some taxa have persisted and diversified between forest refugia, while others have switched to becoming generalists also present in non-forest habitats. Complex arrangements of geographical barriers (hydrology and topography) and ecological gradients between forest and non-forest habitats may have further influenced the region's biodiversity, but elucidation of general diversification processes has been limited by lack of suitable data. Here, we explicitly test alternative diversification modes in the coastal forests using genome-wide single nucleotide polymorphisms, mtDNA, spatial and environmental data for three forest (*Arthroleptis xenodactyloides*, *Leptopelis flavomaculatus* and *Afrivalus sylvaticus*) and four generalist (*Afrivalus fornasini*, *A. delicatus*, *Leptopelis concolor* and *Leptopelis argenteus*) amphibians. Multiple analyses provide insight about divergence times, spatial population structure, dispersal barriers, environmental stability and demographic history. We reveal highly congruent intra-specific diversity and population structure across taxa, with most divergences occurring during the late Pliocene and Pleistocene. Although stability models support the existence of some forest refugia, dispersal barriers and demographic models point towards idiosyncratic diversification modes across taxa. We identify a consistent role for riverine barriers in the diversification of generalist taxa, but mechanisms of diversification are more complex for forest taxa and potentially include topographical barriers, forest refugia and ecological gradients. Our work demonstrates the complexity of diversification processes in this region, which vary between forest and generalist taxa, but also for ecologically similar species with shared population boundaries.

KEYWORDS

demographic inference, diversification, niche modelling, phylogeography, population dynamics, RAD-seq

1 | INTRODUCTION

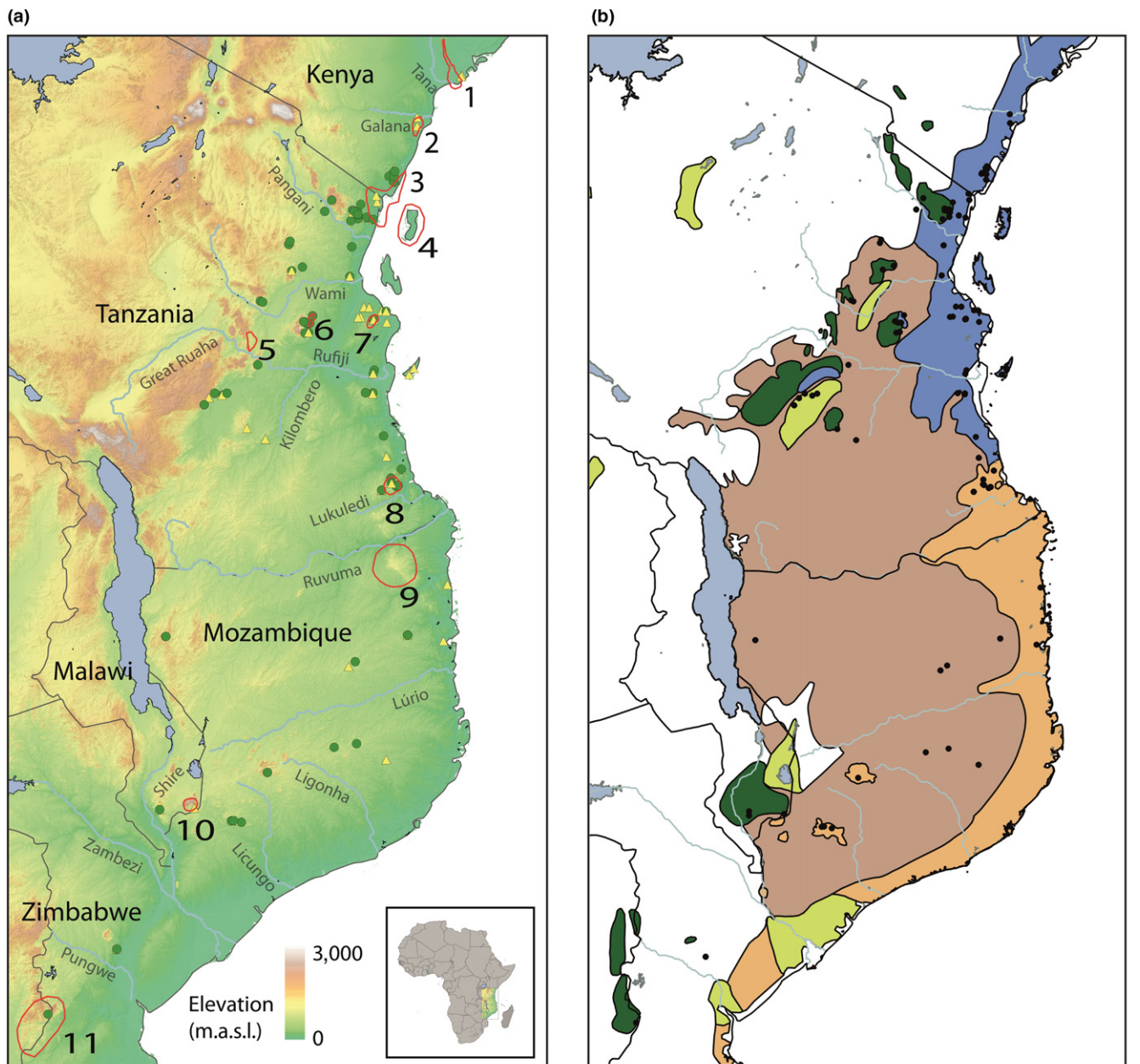
Biodiversity is unequally distributed across the earth, with the greatest concentration occurring in tropical regions (Gaston, 2000). Understanding the processes that have generated and maintained this pattern has been a major question in biology for the best part of a century (Brown, 2013). Landscape changes are hypothesized to have driven vicariant evolution by fragmenting species distributions that were formerly continuous, generating congruent spatial and temporal patterns of genetic differentiation across co-distributed taxa (Rosenzweig, 1995; Smith et al., 2014). Vicariance may occur due to a number of different processes, including the formation of dispersal barriers by rivers (the riverine barrier hypothesis; Gascon et al., 2000; Haffer, 1997; Moritz, Patton, Schneider, & Smith, 2000; Plana, 2004; Voelker et al., 2013; Wallace, 1852), by mountains (Fjelds  & Lovett, 1997) or by areas of unsuitable habitat (Kirschel et al., 2011; Schneider, Smith, Larison, & Moritz, 1999). In some regions, landscape changes over time may have been so severe that particular areas would have functioned as refugia, while diversity in surrounding areas was entirely lost (e.g., the forest refuge hypothesis: Endler, 1982; Haffer, 1969, 1997; Mayr & O'Hara, 1986; Moreau, 1954; Moritz et al., 2000; Plana, 2004). In some cases, diversification may have been triggered by climate change, with taxa making an ecological switch from forest to non-forest habitats in order to persist (the vanishing refuge hypothesis, Vanzolini & Williams, 1981). Finally, ecotones can facilitate disruptive selection on phenotypes despite the presence of gene flow, which may lead to dispersal with incomplete genetic barriers (e.g., range expansion), ultimately driving parapatric divergence (Moritz et al., 2000; Smith, Wayne, Girman, & Bruford, 1997). To understand how and why biodiversity accumulates in tropical regions, the underlying diversification processes need to be tested (e.g., Charles et al., 2018; Ntje et al., 2017; Portik et al., 2017). However, this has remained difficult for most tropical biodiversity hotspots due to a lack of thorough taxonomic, geographic and molecular (genome-wide) sampling.

The Coastal Forests of Eastern Africa (CFEA) are one of Africa's foremost centres of diversity and a designated global biodiversity hotspot (Burgess, D'Amico Hales, Underwood, Dinerstein, & Ecoregion, 2004). Along with the adjacent Eastern Arc Mountains, the CFEA form an important area of endemism highly threatened by anthropogenic impacts (Barratt et al., 2014; Burgess & Clarke, 2000; Burgess, Clarke, & Rodgers, 1998; Burgess, Mwasumbi, Hawthorne, Dickinson, & Doggett, 1992). Distributional data from plants, vertebrates and invertebrates demonstrate a high number of narrow-ranged endemics and a pronounced north–south biogeographic divide caused by the rain shadow of Madagascar (Burgess & Clarke, 2000; Burgess et al., 1998). Diverse communities of taxa, many endemic, are found in the remaining forest fragments of Kenya (Tana River, Kwale, Arabuko-Sokoke), Tanzania (East Usambara, Pemba island, Uluguru, Udzungwa, Pugu hills and Lindi), Mozambique (Bazaruto archipelago), Malawi (Mulanje massif) and Chimanimani and Haroni–Rusitu in Zimbabwe (Burgess et al., 1998; Figure 1a). Historical

environmental change across tropical East Africa has been frequent since the Miocene, and the current CFEA are considered to be the remnants of a once continuous forest that has expanded and contracted for the past 40 million years (Axelrod & Raven, 1978; Demenocal, 1995; Maslin et al., 2014; Mumbi, Marchant, Hooghiemstra, & Wooller, 2008). Combined knowledge of endemism patterns and environmental change have led to the assumption that current CFEA biodiversity mainly originated from the isolation and persistence of ancient lineages in forest refugia, with local extinctions and in some cases adaptation to non-forest habitats across the rest of the region (Azeria, Sanmart n,  s, Carlson, & Burgess, 2007; Barratt et al., 2017a; Burgess et al., 1998). The coastal forests have thus been described as a “vanishing refuge” (Burgess et al., 1998), although to date forest refugial processes have not been thoroughly tested against alternative modes of diversification (Damasceno, Strangas, Carnaval, Rodrigues, & Moritz, 2014; Kirschel et al., 2011; Schneider et al., 1999; Zhen et al., 2017).

To test competing hypotheses of ecological association over long time periods (i.e., millions of years), it is helpful to use taxa that retain ancestral variation at small spatial scales. Amphibians are ideal candidates due to their poor dispersal abilities and highly deme-structured species in East Africa (Bittencourt-Silva et al., 2017; Blackburn & Measey, 2009; Lawson, 2013; Zimkus et al., 2017). Here, we use nuclear DNA (nucDNA) single nucleotide polymorphism (SNP) data acquired from restriction site-associated DNA sequencing (RAD-seq), mitochondrial DNA (mtDNA) and georeferenced occurrences for seven widespread amphibians found in forest and non-forest habitats across the CFEA. Although these seven taxa represent only a small proportion of the 55 known amphibians in this region (Barratt et al., 2017a), they cover a variety of life histories and ecological associations (from generalists to forest specialists), which should reflect the evolutionary processes that have occurred in this region. Our sampling represents almost the complete range of each of the study taxa, allowing us to test the forest refuge hypothesis against alternative modes of diversification, including allopatric divergence across landscape barriers (rivers and topographic) and parapatric divergence across ecological gradients. Using our analytical framework, we make an important first step towards understanding fine-scale diversification patterns and processes in this highly threatened biodiversity hotspot which is a suitable model for similar studies in other tropical regions. For each focal taxon, we assess (a) How many distinct populations are there, and how are they related? (b) Do population boundaries coincide with geographic features and dispersal barriers, and are these boundaries shared across taxa? (c) Which demographic mechanisms have played a role in population diversity and divergence?

For forest taxa which may be older in terms of their evolutionary history, we predict concordant population structure across taxa (Burgess et al., 1998), with distributions that have remained locally stable over multiple time periods (i.e., refugia), and demographic signals of allopatric divergence and size expansion/secondary contact which would represent forest contraction and expansions. Generalist taxa present in non-forest habitats such as Miombo woodland and



FOREST REFUGIA:

- 1 - Tana river, 2 - Arabuko-Sokoke, 3 - Usambara-Kwale,
- 4 - Pemba island, 5 - lowland Udzungwa, 6 - lowland Uluguru, 7 - Pugu hills, 8 - Lindi, 9 - Maçondes plateau,
- 10 - Mulanje, 11 - Chimanimani (including Haroni-Rusitu)

SAMPLING: ● Forest taxa ▲ Generalist taxa

TERRESTRIAL ECOREGIONS:

- East African montane forests
- Northern Zanzibar-Inhambane coastal forest mosaic
- Southern Zanzibar-Inhambane coastal forest mosaic
- Eastern Miombo woodlands
- Zambezi coastal flooded savanna/flooded grasslands
- RAD-seq and mtDNA sampling locality

FIGURE 1 Study region in East Africa, encompassing the Coastal Forests of Eastern Africa and surrounding areas, with sampling locations marked (forest and generalist taxa). (a) Major hydrological and topographical features, including forest refugia demarcated by red polygons based on data from Burgess et al. (1998). (b) Terrestrial ecoregions indicating the location of known biogeographic realms, associated with climate gradients, and sampling localities used in this study (black dots) [Colour figure can be viewed at wileyonlinelibrary.com]

savannah, on the other hand, should demonstrate signals of migration between populations and incomplete dispersal barriers, indicating the potential role of ecological gradients in line with the

vanishing refuge hypothesis. Given the large number of major rivers, mountains and raised plateaus intersecting our CFEA sampling, we additionally investigate the role of landscape barriers, which would

result in clear population structure without subsequent migration, size changes or secondary contact. We make a distinction between forest and non-forest (Miombo woodland and savannah) to aid interpretation of results with regard to the long-term (pre-Pleistocene) environmental stability of these areas and to the ecological preferences of the study taxa.

2 | MATERIALS AND METHODS

2.1 | Study system and conceptual framework

The leaf-folding frogs, *Afrixalus fornasini* (Bianconi, 1849), *Afrixalus delicatus* (Pickersgill, 1984), and *Afrixalus sylvaticus* (Schlötter, 1974), the tree frogs, *Leptopelis argenteus* (Pfeffer, 1893), *Leptopelis concolor* (Ahl, 1929) and *Leptopelis flavomaculatus* (Günther, 1864), and the forest leaf-litter frog, *Arthroleptis xenodactyloides* (Hewitt, 1933), have large geographic ranges distributed throughout the CFEA and surrounding areas. *Afrixalus fornasini*, *A. delicatus*, *L. argenteus* and *L. concolor* are generalist taxa, inhabiting forest edges, woodlands and savannah, whereas *A. xenodactyloides*, *L. flavomaculatus* and *A. sylvaticus* are considered forest-restricted (IUCN, 2017, see Table 1). For *A. delicatus/A. sylvaticus* and *L. argenteus/L. concolor*, geographic boundaries are poorly defined, and some taxonomic confusion exists (IUCN, 2017; Poynton, 2006), which we attempt to elucidate in this study. Due to their wide distributional ranges, these taxa should harbour diversity that reflects population expansion, contraction and persistence, events caused by past environmental changes in the region.

We firstly investigate population structure within taxa to define putative geographic boundaries of populations. We then use explicit demographic model selection to reach a conclusion on whether refugial models of diversification are applicable for forest and generalist taxa. We support demographic results by modelling evolutionary relationships, geographic distributions, effective migration and diversity, and long-term environmental stability based on ecological niche models (ENMs). This approach allows us to assess the congruence of

patterns across taxa and evaluate their consistency with putative forest refugia, geographic features (river and topographic barriers) and known biogeographic regions caused by ecological (rainfall) gradients.

2.2 | Sample collection

Tissue samples (leg muscle, liver or toe clips, $n = 191$) stored in 100% ethanol or RNase Later to preserve DNA were collected across the study region in 2013–2015. Geographic coordinates for all samples were recorded with a handheld GPS device. Additional samples ($n = 40$) held in collections at the University of Basel, University of Jena, Natural History Museum, London, Science Museum of Trento, Museum of Comparative Zoology, Harvard and Museum of Vertebrate Zoology, Berkeley (collected between 2001 and 2012), were used to complement recently collected field data (Supporting Information Table S1).

2.3 | DNA sequencing and data filtering

Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's instructions. A partial fragment of the mitochondrial 16S gene was amplified via polymerase chain reaction to verify species identity using the NCBI BLAST tool against our own barcoding database of amphibians across the region (full details can be found in Barratt et al., 2017a; GenBank Accession nos included in Supporting Information Table S1). DNA was quantified prior to RAD-seq library preparation using a Qubit fluorometer (Invitrogen). We followed the Etter, Bassham, Hohenlohe, Johnson, and Cresko (2011) laboratory protocol to prepare RAD-seq libraries using the *SbfI* restriction enzyme (Supporting Information Appendix S1). Final RAD-seq libraries included 43 *A. fornasini* from 30 sites, 49 *A. delicatus/A. sylvaticus* from 35 sites (comprising of 22 *A. delicatus* from 15 sites and 27 *A. sylvaticus* from 20 sites), 59 *L. flavomaculatus* from 24 sites, 27 *L. argenteus/L. concolor* from 18 sites (comprising of 12 *L. argenteus* from 8 sites and 15 *L. concolor* from 10 sites) and 53

TABLE 1 Summary of taxa studied including information on taxonomy and currently recognized species according to the IUCN red list

Species group	Recognized species (including recently synonymised)	Type locality	Habitat
<i>Afrixalus fornasini</i>	<i>Afrixalus fornasini</i> (Bianconi, 1849)	Mozambique	Generalist
	<i>Afrixalus unicolor</i> (Boettger, 1913) ^a (<i>A. fornasini</i>)	Tanzania	Generalist
<i>Afrixalus stuhlmanni</i>	<i>Afrixalus stuhlmanni</i> (Pfeffer, 1894)	Zanzibar, Tanzania	Generalist
	<i>Afrixalus brachycnemis</i> (Boulenger, 1896)	Malawi	Generalist
	<i>Afrixalus sylvaticus</i> (Schlötter, 1974)	Kwale, Kenya	Forest
	<i>Afrixalus delicatus</i> (Pickersgill, 2005)	St. Lucia, South Africa	Generalist
<i>Arthroleptis xenodactyloides</i>	<i>Arthroleptis xenodactyloides</i> (Hewitt, 1933)	Chimanimani, Zimbabwe	Forest
	<i>Arthroleptis stridens</i> (Pickersgill, 2007) ^b	East Usambara, Tanzania	Forest
<i>Leptopelis argenteus</i>	<i>Leptopelis argenteus</i> (Pfeffer, 1893)	Bagamoyo, Tanzania	Generalist
	<i>Leptopelis concolor</i> (Ahl, 1929)	Witu, Kenya	Generalist
	<i>Leptopelis broadleyi</i> (Poynton, 1985) ^a (<i>L. argenteus</i>)	Mozambique	Generalist
<i>Leptopelis flavomaculatus</i>	<i>Leptopelis flavomaculatus</i> (Günther, 1864)	Ruvuma bay, Tanzania	Forest

Notes. Habitat: Forest = species is primarily forest associated; Generalist = species is a generalist, not confined to forest.

^aSpecies synonymized with conspecific (in brackets). ^bRecognized by IUCN red list but taxonomic status is uncertain.

A. xenodactyloides from 35 sites (Supporting Information Table S1). We sequenced individuals across five RAD-seq libraries (45–51 samples each), with a unique barcode adapter per individual in each library to demultiplex sequences bioinformatically. The final eluted products were sequenced in a single run on an Illumina Hi-seq 2500 (100-bp single-end reads) at the D-BSSE sequencing facility in Basel, Switzerland (Supporting Information Table S2).

We used STACKS 1.41 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to process RAD-seq data and produce single nucleotide polymorphism (SNP) data sets. We used the *process_radtags* module to filter out low-quality reads, demultiplexing individuals into their own fastq file. The standard workflow of *ustacks*, *cstacks* and *sstacks* modules was used to align reads into stacks, to build a catalogue of consensus loci by merging alleles across individuals and to match individuals to the catalogue, respectively (data sets summarized in Supporting Information Table S3). Catalogues of loci were built separately for *A. fornasini*, *A. xenodactyloides* and *L. flavomaculatus*. For *A. delicatus/A. sylvaticus* and *L. argenteus/L. concolor*, the catalogues included the combined individuals in each species pair due to possible admixture between them given their overlapping distributions. We wished to capture all possible loci in the catalogue and then filter data, so for all downstream analyses, we subsequently separated each of these taxa based on population structure results, resulting in seven separate data sets. As sampling bias is inherent in RAD-seq data sets, we acknowledge that our results could potentially be affected by allelic dropout and null alleles, PCR duplicates and genotyping errors, and variance in depth of coverage amongst loci (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), which we mitigated against as described in each relevant section. Our final catalogues of loci used a minimum depth of sequencing coverage of 5 \times , and a maximum of 2 bp mismatches between the fragments, with only loci present in at least half of the individuals in each catalogue retained. Data matrices were then generated using the *populations* module, retaining only a single random SNP per RAD locus to avoid linkage disequilibrium (Andrews et al., 2016).

2.4 | Population structure

We filtered STACKS “haplotype” files to remove loci that were invariant between samples, loci with at least one individual with more than two alleles (i.e., potentially paralogous loci), and loci that were not bi-allelic. We investigated population structure per taxon using discriminant analysis of principal components (DAPC) in the ADEGENET R package (Jombart & Ahmed, 2011), after converting STACKS output files into Fstat format using PGDSPIDER 2.1.0.3 (Lischer & Excoffier, 2012). Unlike model-based clustering methods, the DAPC method is free of assumptions regarding Hardy–Weinberg equilibrium (Jeffries et al., 2015; Jombart & Ahmed, 2011) and less sensitive to minor allele frequency thresholds (Linck & Battey, 2017). We defined values of k between 1 (i.e., a single panmictic population) and 8, using Bayesian information criterion (BIC, Schwarz, 1978) scores across tested k values to infer the number of populations. Due to the taxonomic uncertainty of some samples of *A. delicatus/A. sylvaticus* and

L. argenteus/L. concolor, we conducted an initial analysis of the full data set to confirm species memberships, followed by analyses of each taxon separately. To complement our DAPC analyses, we also ran ADMIXTURE (Alexander, Novembre, & Lange, 2009) with formatted bed files converted using PLINK 1.07 (Purcell et al., 2007). As with DAPC analyses, k ranged between 1 and 8, and we used the 10-fold cross-validation procedure to estimate the number of population clusters. Population membership of each individual for chosen k values was verified by inspecting the clustering analysis plots in DAPC and ancestry coefficients in ADMIXTURE barplots (Supporting Information Tables S4 and S5). DAPC clustering for multiple values of k is shown in Supporting Information Figure S1.

2.5 | Demographic model selection

To evaluate the likelihoods of alternative demographic models within each taxon based on RAD-seq data, we used the diffusion approximation method implemented in $\delta a \delta i$ (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) to analyse two- and three-dimensional Joint Site Frequency Spectra (JSFS). The number of dimensions used in models (2D or 3D) refers to the number of populations being compared (based on population structure results), using folded JSFS because we lacked out-group information. For *L. flavomaculatus* we excluded the Mozambique populations, which only contained two individuals after data filtering. Following Portik et al. (2017), included parameters in models allowed for the broad categorization of models into three competing diversification modes. Landscape (i.e., riverine or topographic) barriers fall under a general allopatric model of population splitting with no gene flow, and no assumptions of migration or secondary contact. Forest refugia models follow a similar model of allopatric divergence, but expect initial population isolation followed by size change and/or secondary contact. Parapatric models expect gene flow accompanying divergence and subsequent isolation and represent divergence due to ecological gradients. Although these broad categorizations are simplistic, they enable comparisons across taxa to be more readily made, although it should be noted that other processes which are not explicitly captured by our model parameters may also contribute to the demographic patterns observed (e.g., range expansion, local adaptation, recent anthropogenic impacts). We ran a total of 15 alternative 2D models that differed in parameters for migration rates, periods of isolation and population size changes (visually represented in Supporting Information Figure S2), including a null model of no divergence between populations. A set of 15 models was run for 3D population comparisons (visually represented in Supporting Information Figure S3), including several models that account for the simultaneous divergence of populations based on potential polytomies in dating analyses for each of *A. delicatus*, *A. sylvaticus* and *L. flavomaculatus*. Because the SFS cannot be constructed using an incomplete data matrix, it is necessary to first down-project the data to smaller sample sizes of alleles. We did this by exploring a range of values per population (between 2 and 30), choosing the configuration per data set with the largest number of segregating sites. We ran three sets of increasingly focused

optimizations for each model before performing the final model selection. We did not transform obtained parameters into absolute migration rates and divergence times because our primary aim was to perform model selection, and these parameter values should ideally be estimated using an accurate mutation rate which we lack for our study taxa (Gutenkunst et al., 2009). We therefore compare the relative time intervals of population divergences obtained from $\delta\text{a}\delta\text{i}$ with those obtained from a Bayesian coalescent-based approach using bi-allelic SNPs as described in the next section. To determine the best fitting models, the AIC and log likelihoods were inspected, and ΔAIC scores were used to calculate relative likelihoods and Akaike weights (w_i). The model with the highest Akaike weight was selected as the most likely for each divergence event (Burnham & Anderson, 2002). To explore the possible influence of recent anthropogenic impacts, we ensured that a variation in the top-ranked model per taxon was also tested that included a size change step. Two-dimensional models already had these models within the original fifteen tested, and we built an additional three 3D models to test for each of *A. fornasini*, *L. flavomaculatus* and *A. xenodactyloides*. To verify that our models were reasonable explanations of the JSFS, we performed goodness of fit tests. For each taxon, we fit the top-ranked model using our optimized parameters, scaled the resulting model spectrum by the inferred theta value and used the model spectrum to generate 100 Poisson-sampled frequency spectra. We then optimized each simulated frequency spectrum to obtain a distribution of log-likelihood scores and Pearson's chi-squared test statistic and subsequently determined whether our empirical values were contained within these distributions. A more detailed description of demographic model selection and goodness of fit tests is shown in Supporting Information Appendix S1. To support our demographic model selection, we also estimated evolutionary relationships, effective migration and diversity, and the long-term stability of taxa using ENMS, as described below.

2.6 | Evolutionary relationships

We explored evolutionary relationships with mtDNA and RAD-seq data separately. Sequences of 16S mtDNA for all samples included in RAD-seq libraries were edited in GENEIOUS 6 (Kearse et al., 2012) and aligned with the RAXML tree estimator using a GTRCAT model in SATÉ-II (Liu et al., 2012) before analyses in BEAST 2.4.8 (Bouckaert et al., 2014). We used BMODELTEST (Bouckaert & Drummond, 2017) to average over all possible substitution models instead of selecting a single model. We used a strict clock with a log-normal prior distribution to estimate divergence times in millions of years. Estimated mitochondrial substitution rates for 16S in amphibians range from 0.16% to 1.98% pairwise divergence per million years (Jongsma et al., 2017), and the prior mean was set to the mid-point of this range (1.07%, set as 0.00535 substitutions/site/MYR) with a standard deviation of 0.3 for the interquartile range to reach the approximate lower and upper range limits. A birth–death tree prior was used, running the MCMC for 20,000,000 generations, sampling every 1,000 trees.

We reconstructed phylogenomic relationships using SNAPP 1.3 (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012) implemented in BEAST 2.4.8 (Bouckaert et al., 2014). SNAPP is a package that infers species or population trees from unlinked markers such as bi-allelic SNPs, implementing a coalescent model with an algorithm to integrate all possible gene trees rather than explicitly sampling them. To reduce computational requirements and run times, we selected 2–6 representative individuals per population (based on population structure results) with at least 50% complete data matrices. Backwards (u) and forwards (v) mutation rates (expected mutations per site per generation) were estimated from the data in SNAPP, with the birth rate (λ) of the Yule prior (indicating the rate at which populations diverge from one another) based on the number of samples used (Supporting Information Table S6). The run for each data set used a chain length of 1,000,000 generations, sampling every 1,000 trees. We inspected final log files and created maximum clade credibility trees (median node heights) by combining two independent runs per taxon in TREEANNOTATOR 2.4.6 after discarding 10% as burn-in. To verify that the selected individuals did not severely bias SNAPP results, we repeated each analysis using a different random selection of individuals per population (non-overlapping where possible, Supporting Information Table S6).

Given that an accurate mutation rate for amphibians is unavailable, we refrained from inferring the absolute timing of divergence events based on bi-allelic SNP data. Furthermore, as the SNAPP model is coalescent-based, it can account for incomplete lineage sorting, but the presence of high gene flow can cause underestimates of node ages (Leaché, Harris, Rannala, & Yang, 2014a). Additionally, the assembly and filtering of our RAD-seq data sets may also adversely affect dating estimates due to high numbers of retained loci being under selection or linked, which would potentially reduce calculated genetic diversity (Huang & Knowles, 2016). Despite the uncertainties of absolute dating, our SNAPP and $\delta\text{a}\delta\text{i}$ analyses enabled a relative comparison of the divergence time intervals between populations of each species. We investigated divergence estimates for each of these analyses alongside mtDNA estimates and conducted Pearson's correlation tests between the relative divergence time intervals to aid our discussion of forest and generalist taxa.

2.7 | Effective migration and diversity surfaces

We visualized effective migration and diversity surfaces (i.e., gene flow and barriers) using the ESTIMATED EFFECTIVE MIGRATION SURFACES (EEMS) program 0.0.0.9000 (Petkova, Novembre, & Stephens, 2016). This program identifies geographic areas where genetic similarity is greater than expected under isolation by distance using spatial and SNP data, without the need for environmental and topographic information. The effective migration and effective diversity estimates are interpolated across geographic space and provide a visual representation of observed genetic dissimilarities, including regions with higher or lower gene flow (i.e., barriers) and effective (i.e., genetic) diversity than expected. We converted filtered STACKS “haplotype” files into PLINK format (.bed files, Purcell et al., 2007) and used the

BED2DIFFS program to calculate dissimilarity matrices for each taxon. Based on the size of the habitat and the number of demes required to fill that habitat, we defined a deme size (the density of populations) of 700 in EEMS, but also explored results using smaller deme sizes of 250 and 500 using the SNP version of EEMS (runeems_snps). We used a MCMC length of 25,000,000 with a burn-in of 1,000,000, each for three replicates. We combined results using the EEMS R plotting (rEEMSplot) package and plotted surfaces of effective migration (m) and effective diversity rates (q).

2.8 | Ecological niche and stability models

To evaluate the geographic distributions over time per taxon (i.e., stability), we first used ENMS to define the realized macroclimatic niche of each taxon based on current environmental conditions. We then used this model to determine whether similar environmental conditions were found at a specific past epoch. Finally, we summed models over multiple time periods to visualize the potential spatio-temporal distributions of each taxon. We collected all available locality information per taxon using the Global Biodiversity Information Facility (GBIF), published data (Barratt et al., 2017a; Burgess & Clarke, 2000; Ohler & Frétey, 2015) and our own and collaborator's fieldwork. Data were filtered to remove any imprecise or ambiguous localities, or points that could not be accurately matched to each taxon with certainty. This resulted in 59–144 unique records per taxon (Supporting Information Table S7). To ensure that georeferenced data were not spatially autocorrelated, we rarefied all points to be a minimum of 10 km apart, resulting in 20–112 unique localities per taxon ($n = 112, 86, 36, 62, 38, 20, 32$, for *A. xenodactyloides*, *L. flavomaculatus*, *A. sylvaticus*, *A. fornasini*, *A. delicatus*, *L. argenteus* and *L. concolor*, respectively). We built ENMS in MAXENT per taxon using elevation data derived from the USGS (<http://csgtm.isgcm.org/dataset/gtopo30>) and six bioclim variables (mean diurnal temperature range, temperature seasonality, maximal temperature of warmest month, annual precipitation, precipitation of driest month and precipitation of warmest quarter), at 30 arc-second resolution (approximately 1 km² grid cells). These variables were obtained from the WordClim Database, based on the Community Climate System Model (CCSM) global circulation model (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), and selected based on low between-variable correlations (Pearson's $r < 0.7$), to minimize model overfitting. Background data from 10,000 random points were sampled from a minimum convex polygon defined by a 150 km buffer around each occurrence record. To select optimal parameters for models, we tested a range of feature classes (hinge, quadratic, linear, product and threshold), each with a regularization multiplier between 1 and 3 in increments of 1. To project models into the past, this procedure was repeated for three different palaeoclimate epochs (the mid-Holocene ca. 6 kybp, the Last Glacial Maximum ca. 21 kybp and the Last Interglacial ca. 120 kybp). We selected the best models for each taxon (Supporting Information Table S8) based on the lowest test omission and highest AUC scores (Brown, 2014). To create stability models, we calculated the mean of the negative log suitability per

epoch, using the exponent of this value to create continuous stability models per taxon ranging between 0 (absent) and 1 (present, Rosauer, Catullo, Vanderwal, & Moussalli, 2015). As the Last Glacial Maximum data were only available at 2.5 arc-seconds (approximately 5 km²), continuous stability models are resampled to this resolution. The palaeoclimatic data we used for the ENM projection only cover the period until the Last Interglacial (120 kybp). We therefore use the stability models to reflect recent major climatic events, and as a proxy for deeper time in the absence of accurate palaeoclimate data further back to the Miocene.

3 | RESULTS

3.1 | Illumina reads and filtered loci

We obtained single-end Illumina reads for 43 *A. fornasini* (182,663,928 reads in total), 49 *Afrixalus sylvaticus/Afrixalus delicatus* (243,690,376 reads), 27 *Leptopelis argenteus/Leptopelis concolor* (154,933,766 reads), 59 *Leptopelis flavomaculatus* (299,581,783 reads) and 53 *Arthroleptis xenodactyloides* individuals (199,514,898 reads, Supporting Information Table S2). STACKS output haplotype files contained between 1,930 (*A. delicatus/A. sylvaticus*) and 9,867 (*L. flavomaculatus*) loci. After excluding invariant, paralogous and non-bi-allelic loci, and individuals with high amounts of missing data (>90%), final numbers of SNPs per taxon used for subsequent analyses (Supporting Information Table S3) were as follows; 3,753 (*A. fornasini*), 1,646 (*A. delicatus* and *A. sylvaticus*), 3,371 (*L. argenteus* and *L. concolor*), 1,505 (*A. xenodactyloides*) and 8,598 (*L. flavomaculatus*).

3.2 | Population structure

DAPC and ADMIXTURE analyses produced congruent results for the number of inferred populations (*L. flavomaculatus*: $k = 4$; *A. delicatus*: $k = 3$; *A. sylvaticus*: $k = 3$, *L. argenteus*: $k = 2$; *L. concolor*: $k = 2$; Figure 2, Supporting Information Tables S4 and S5). In two taxa (*A. fornasini* and *A. xenodactyloides*), there was discordance across analyses, and for *A. fornasini*, the DAPC analysis suggested a tripartition ($k = 3$), while ADMIXTURE suggested a single panmictic population ($k = 1$). For *A. xenodactyloides*, DAPC and ADMIXTURE results were also incongruent ($k = 4$ and $k = 8$, respectively). As DAPC is known to be less sensitive to minor allele frequency thresholds than other model-based clustering methods (Linck & Battey, 2017), we based subsequent *A. fornasini* analyses on the DAPC results given their congruence with phylogenetic and phylogenomic inferences. Similarly, for *A. xenodactyloides*, subsequent analyses were based on $k = 3$ following consistently lowest BIC and CV scores in DAPC and ADMIXTURE analyses (Supporting Information Table S4), and consistency with phylogenetic clustering. Results for other values of k are shown in Supporting Information Figure S1. Population structure analyses across all taxa revealed that thirteen of the twenty populations were restricted to six allopatric areas which are non-overlapping. The remaining seven populations exhibited wider ranges associated with either the northern or southern parts of the CFEA region (Figure 3).

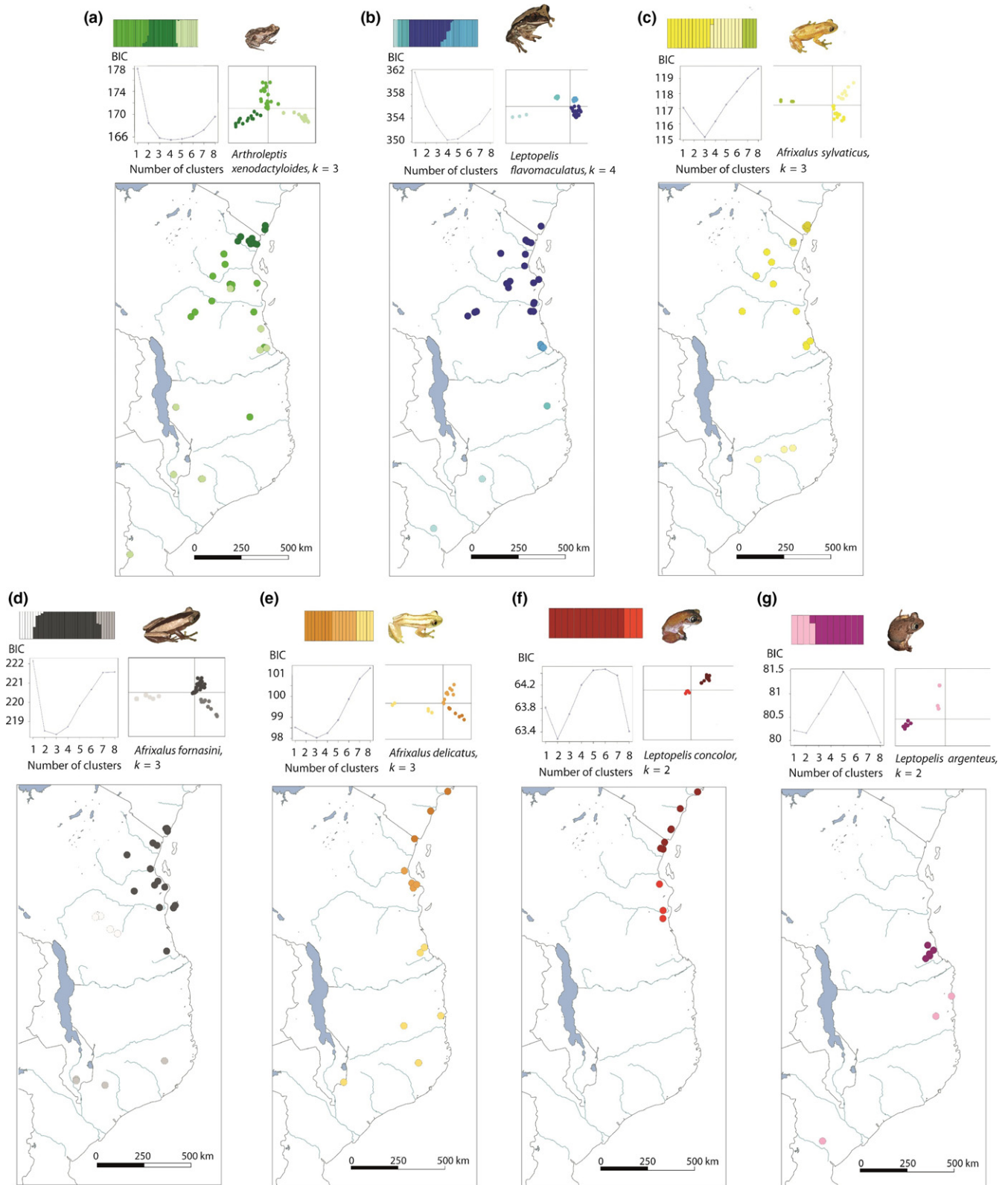


FIGURE 2 Population structure for each of the seven amphibian taxa including ADMIXTURE plots detailing ancestry coefficients and discriminant analyses of principal components (DAPC) showing the most likely numbers of population clusters based on BIC scores (line graph). (a–c) Forest taxa (*Arthroleptis xenodactyloides*, *Leptopelis flavomaculatus*, *Arthroleptis sylvaticus*). (d–g). Generalist taxa (*Afrixalus fornasini*, *Afrixalus delicatus*, *Leptopelis concolor*, *Leptopelis argenteus*). Populations are coloured corresponding to their spatial distributions on accompanying maps [Colour figure can be viewed at wileyonlinelibrary.com]

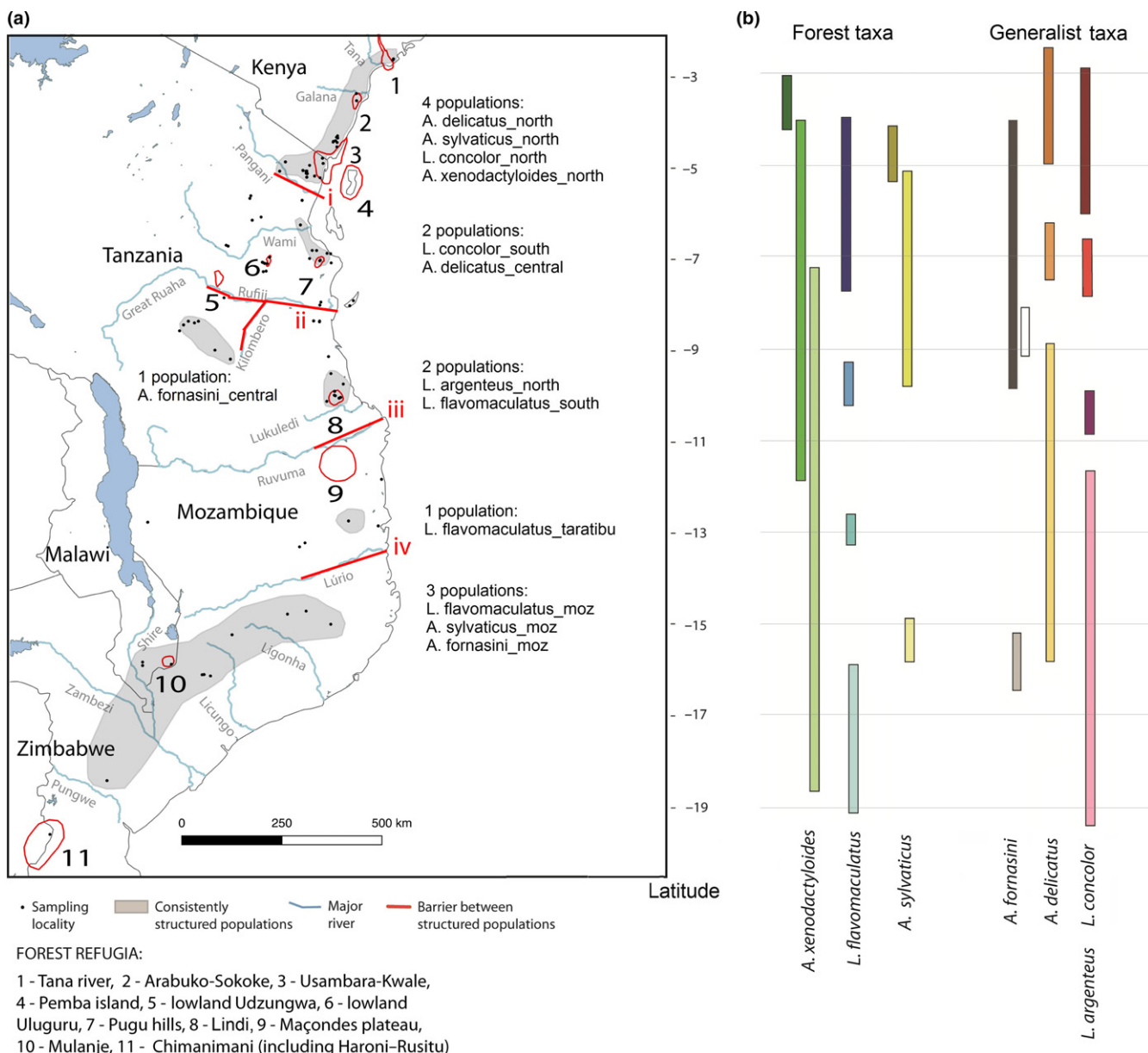


FIGURE 3 (a) Cumulative summary of major population breaks shown in 13 of the 20 discovered populations, with nearby river systems (i–iv) and forest refugia (1–10) labelled. (b) Latitudinal range of each population grouped into forest and generalist taxa. Colours correspond to populations identified in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Demographic model selection

The consistency of log likelihoods and parameters across replicates of each model increased after the first, second and third optimization rounds and indicated convergence for the best-ranked models across replicates (Supporting Information Table S9, Figure S4). Null models (no divergences) were consistently lowest ranked for 2D models, supporting the genetic distinctiveness of populations. For forest taxa, demographic model selection was inconsistent across taxa (Table 2; Figure 4a), with the best-ranked model selected as historical gene flow for *A. xenodactyloides*, and allopatric divergence models selected for *L. flavomaculatus* (historical isolation followed by secondary contact) and *A. sylvaticus* (divergence and isolation). For all generalist

taxa, allopatric models were consistently found as the best-ranked (Table 2; Figure 4b), with diversification without subsequent migration and/or size change. Based on Akaike weights (Supporting Information Table S10), best-ranked models were significantly better than the next best alternative models for all taxa ($\omega_i > 0.99$) with the exception of *L. argenteus* ($\omega_i = 0.74$) and *L. concolor* ($\omega_i = 0.65$), for which the second-best model was only characterized by the addition (*L. argenteus*) or the absence of a size change (*L. concolor*). The amended top-ranked models that included a size change step over multiple epochs implied that human impacts could potentially have played a role for *L. concolor*, and, to a lesser extent, *L. argenteus*, which were characterized by high Akaike weights for size change models and a recent time interval since the change occurred.

TABLE 2 Demographic model selection using $\delta a\delta i$ for each species group

Habitat	Species	Population comparison	JSFS model type	Proj. sample size	Best model	General model	In-l	AIC	Akaike weight (ω_i)
Forest	<i>Arthroleptis xenodactyloides</i>	Montane—south—north	3D	9,9,10	Ancient migration 2 (shorter isolation)	Ecological gradient	-266.22	546.44	0.999
Forest	<i>Leptopelis flavomaculatus</i>	Taratibu—south—north	3D	15,29,36	Simultaneous split in refugia, symmetrical migration (adjacent)	Forest refugia	-641.51	1297.02	0.999
Forest	<i>Afrivalus sylvaticus</i>	North—central—montane_south	3D	4,10,13	Simultaneous split, no migration	Landscape barrier	-203.2	414.4	0.999
Generalist	<i>Afrivalus fornasini</i>	Central—south—moz	3D	6,8,10	Split, no migration	Landscape barrier	-605.56	1223.12	1
Generalist	<i>Afrivalus delicatus</i>	South—central—north	3D	9,8,6	Simultaneous split, no migration	Landscape barrier	-96.02	200.04	0.999
Generalist	<i>Leptopelis argenteus</i>	South—north	2D	5,6	Split, no migration	Landscape barrier	-50.13	106.26	0.739
Generalist	<i>Leptopelis concolor</i>	South—north	2D	6,6	Split, no migration, size change	Landscape barrier/ Anthropogenic	-69.45	150.9	0.65

Note. Model type is shown (2D or 3D-JSFS) along with sample projection size. Best model from all tested models is listed along with its final log likelihood, AIC score and Akaike weight (ω_i).

Goodness of fit tests showed that most models were within reasonable expectations of the simulated data, with the exception of *L. concolor* for which the empirical result lay outside of the simulated data distributions of log likelihoods and log-transformed Pearson's chi-squared test statistics (Figure 4). Visualizations of the data, models and residuals are shown in Supporting Information Figure S4, along with likelihood plots of the best model to demonstrate convergence across each of the three rounds of replicates. Parameters and full results of best scoring replicates for all models across taxa are shown in Supporting Information Table S10, and full goodness of fit results are shown in Supporting Information Table S11.

3.4 | Evolutionary relationships

Mitochondrial DNA-inferred relationships (Figure 5a) were mostly congruent with those inferred using bi-allelic SNPs from RAD-seq data, despite some discordance in population membership and the evolutionary relationships between populations, which we suspect is due to incomplete lineage sorting of the 16S mtDNA locus. For *A. fornasini*, *L. flavomaculatus* and *A. xenodactyloides*, mtDNA haplotype clades were consistent with the RAD-seq inferred population clusters, but in *A. delicatus*, and *L. argenteus/L. concolor*, population clusters did not always form monophyletic haplotype groups. According to mtDNA data, the earliest divergences within taxa are during the Pliocene (3.67 mya, 95% HPD 1.59–6.89 mya for *A. sylvaticus* and 3.44 mya, 95% HPD 1.55–6.30 mya for *A. xenodactyloides*). The large confidence intervals indicate that the earliest divergences in *A. fornasini* and *A. delicatus* could have also occurred around the late Pliocene (as early as 2.37 and 3.68 mya, respectively). The remainder of later population divergences in these taxa and all divergences within *L. flavomaculatus*, *L. argenteus* and *L. concolor* occurred at different stages throughout the Pleistocene.

Evolutionary relationships reconstructed from bi-allelic SNPs in SNAPP (Figure 5b) showed that divergences between taxa occurred at broadly comparable timescales. Contrary to our expectations that forest taxa diverged earlier than generalists, we found similar timescales for most generalists and forest taxa except for the much more recently diverged generalist populations of *L. argenteus* and *L. concolor*. Analyses based on different random representatives of individuals per population showed similar results in terms of relationships and divergence rates, with one exception (the south Taratibu population divergence for *L. flavomaculatus*) indicating that results may potentially be sensitive to the population sets used in some cases (Supporting Information Table S6). Our comparison of time interval parameter estimates between SNAPP and $\delta a\delta i$ (Supporting Information Table S6) yielded remarkably similar results (Pearson's $r > 0.88$ for forest taxa and $r > 0.93$ for generalist taxa across replicates), indicating that the divergence times obtained are congruent in relative terms in spite of the aforementioned problems of absolute dating. Mitochondrial DNA divergence dates were more closely correlated with $\delta a\delta i$ and SNAPP estimates across replicates for generalist taxa ($r > 0.87$) than for forest taxa due to the handful of incongruent relationships between mtDNA and SNP data ($r = 0.43$).

Cumulatively, population structure, phylogenetic and phylogenomic results indicate a strong north-south pattern of differentiation between populations, in addition to at least four major breaks that segregate populations across multiple taxa (Figure 3). Divergences between southern populations in Mozambique, Zimbabwe and Malawi, and the remaining CFEA regions in Tanzania and Kenya are approximately located around major river barriers (Figure 3, labels i–iv). In Tanzania and Kenya, populations are regularly separated from each other in the Lindi region (evident in *A. xenodactyloides*, *L. flavomaculatus*, *L. argenteus* and *A. delicatus*). The Usambara-Kwale region

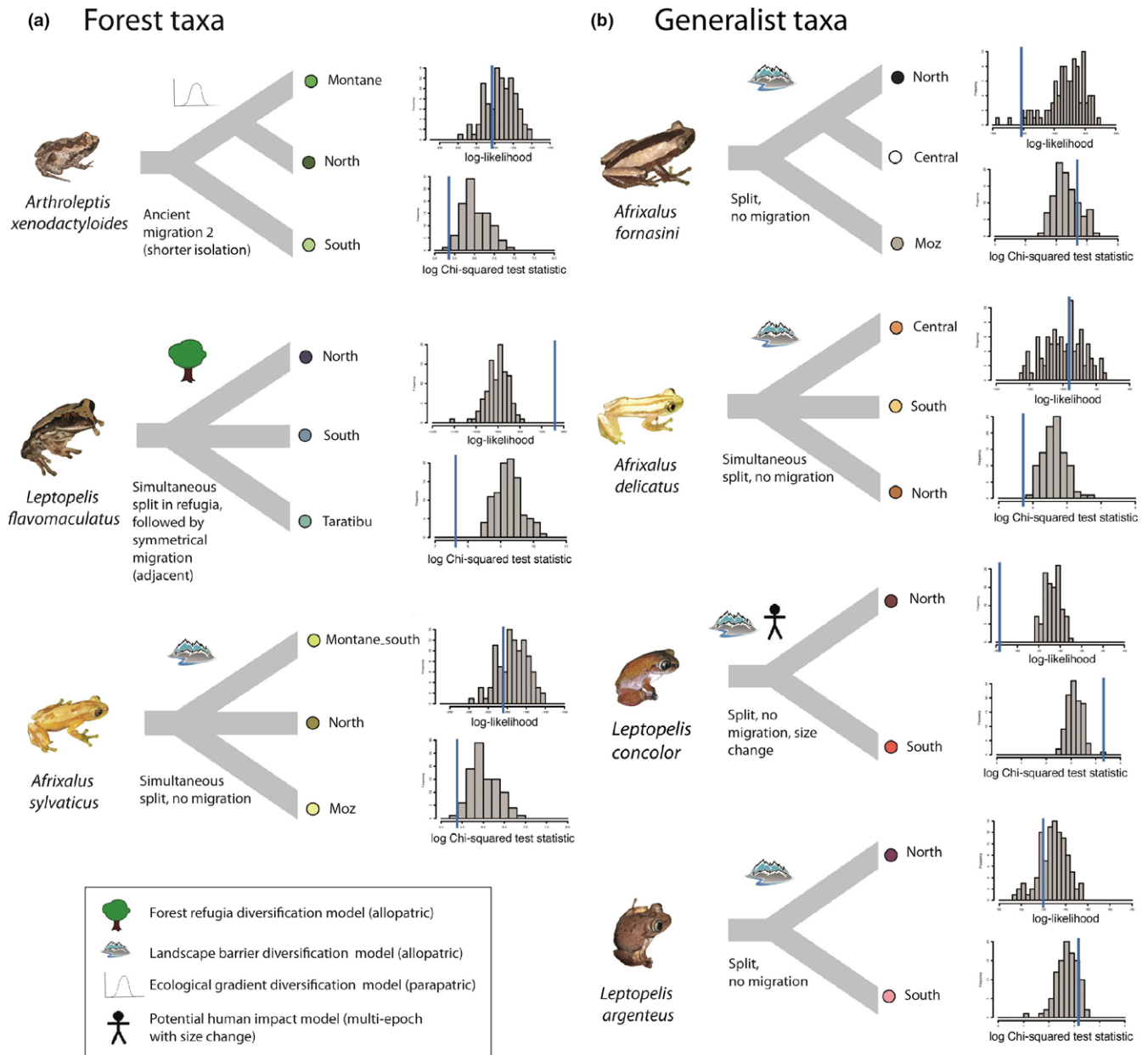


FIGURE 4 Visual representation of the best-ranked demographic models selected by $\delta a \delta i$ from a choice of 15 alternative 2D JSFS demographic models and 15 alternative 3D JSFS models. Next to each model are goodness of fit test results, showing the location of the empirical (blue lines) value within the distribution of simulation values (grey bars) for log likelihood and Pearson's log-transformed chi-squared test statistic. Empirical values occurring within distributions indicate good fits, with poorer fits indicated by a lower log likelihood (left of distribution) or a higher log-transformed chi-squared test statistic (right of distribution). (a) Forest taxa 3D best models (population divergences in *Arthroleptis xenodactyloides*, *Leptopelis flavomaculatus* and *Afrixalus sylvaticus*), (b) generalist taxa best 2D and 3D models (*A. fornasini*, *Afrixalus delicatus*, *Leptopelis concolor*, *Leptopelis argenteus*). Populations are colour coded matching Figures 2 and 5. Comparisons of the pairwise joint site frequency spectra for the data, the model and resulting residuals can be found in Supporting Information Figure S4, along with likelihood estimates across optimization rounds to indicate convergence. Parameter values and results for all models are shown in Supporting Information Table S9 [Colour figure can be viewed at wileyonlinelibrary.com]

spanning across the Tanzania–Kenya border also divides unique populations within *A. xenodactyloides*, *L. concolor* and *A. delicatus*.

3.5 | Effective migration and diversity surfaces

The EEMS analyses based on smaller deme sizes (250, 500, not shown) merged numerous localities that we consider important to

keep separate, particularly around the border of Tanzania and Kenya, and within southern Tanzania. Therefore, we only present results from deme sizes of 700 (Figure 6). The analyses revealed several barriers to migration mostly matching with the population breaks shown in Figure 3. For the forest taxa *A. xenodactyloides* and *L. flavomaculatus*, at least three major barriers in Tanzania and one in Mozambique were found, also approximately coinciding with the

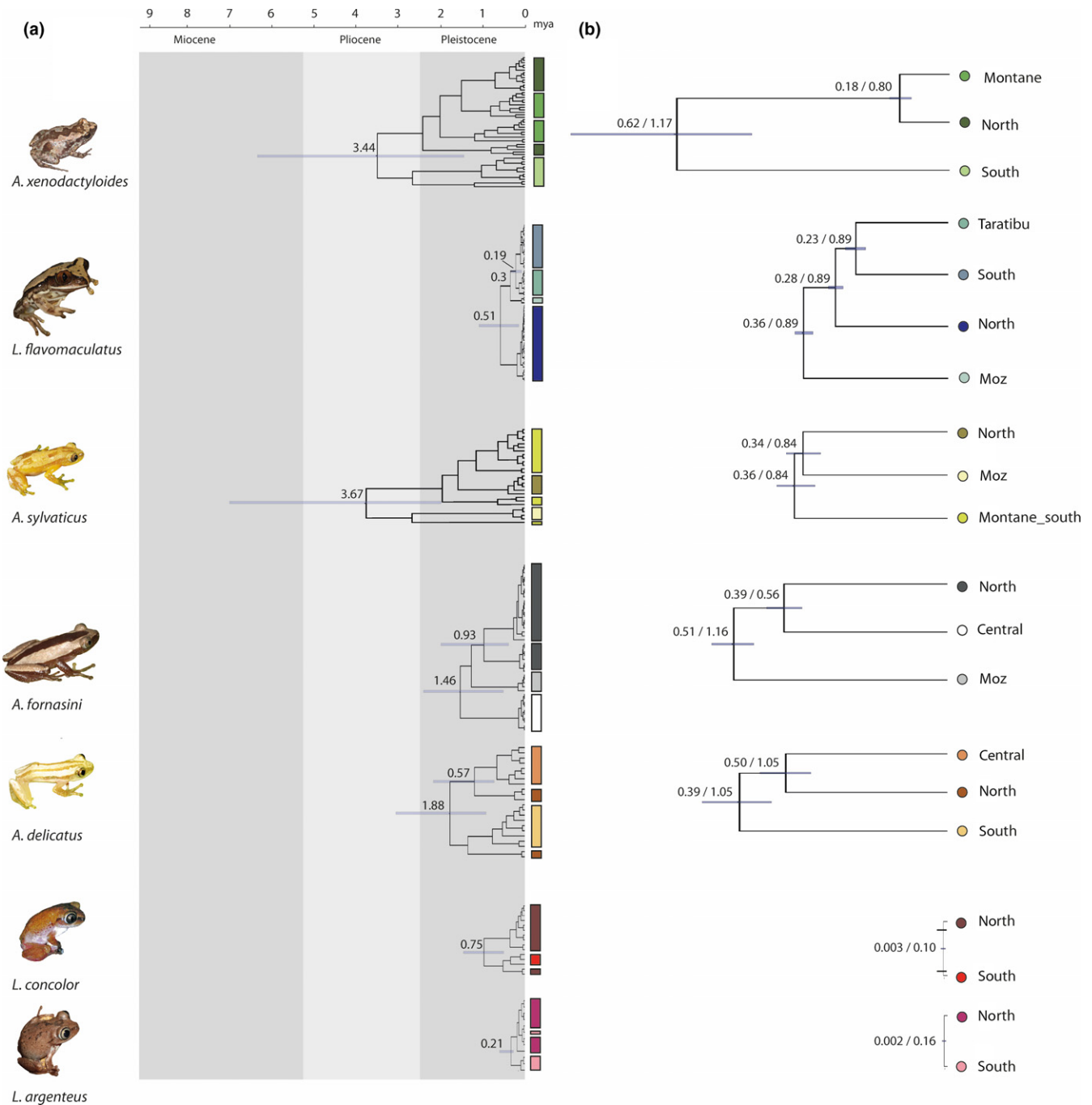
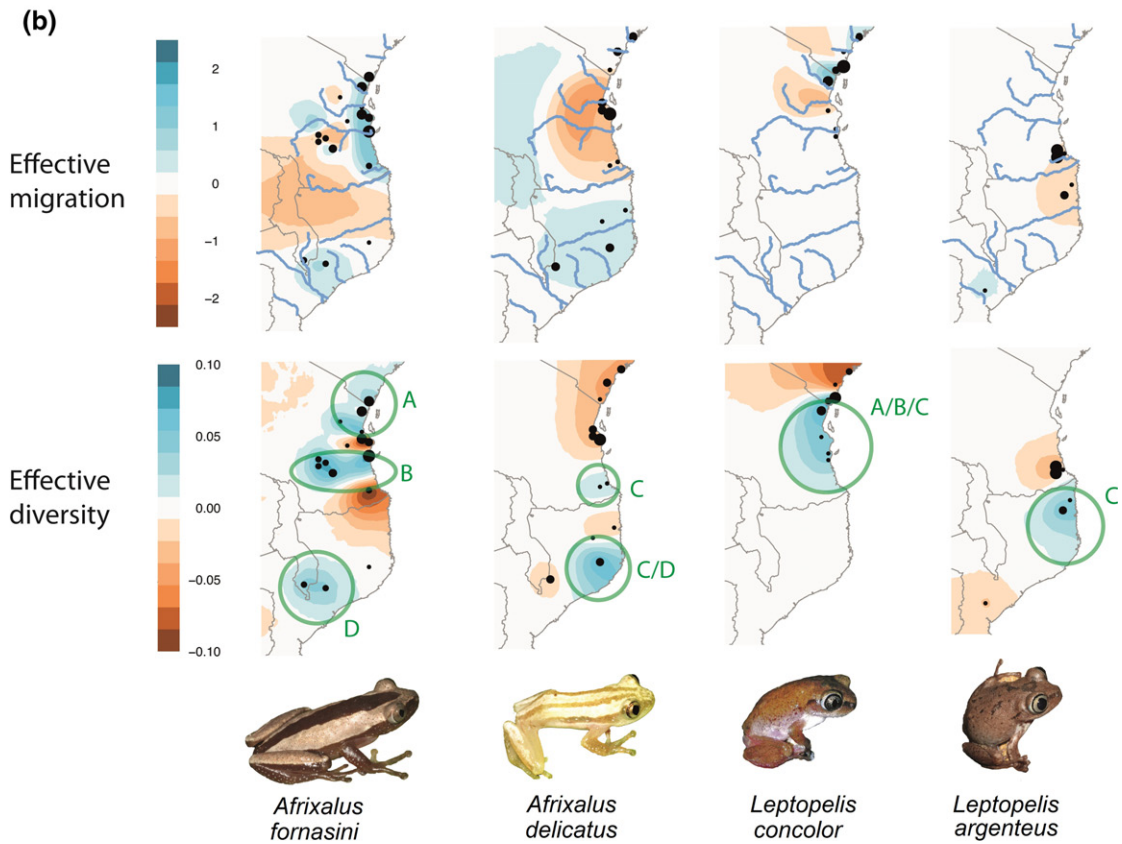
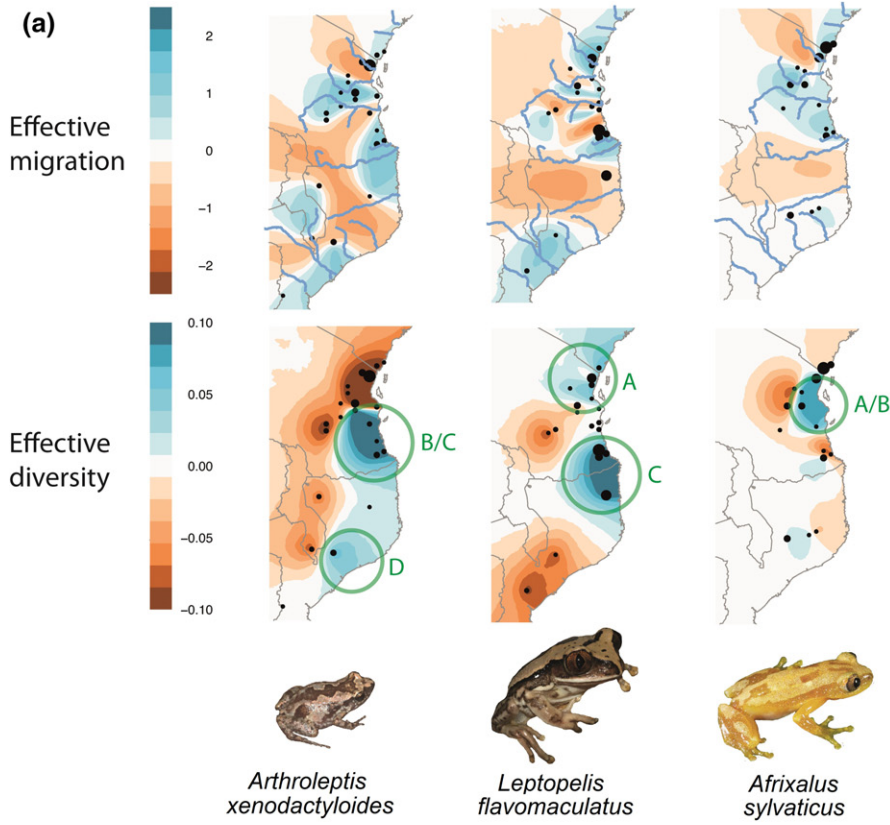


FIGURE 5 (a) Dated phylogenies per species in BEAST based on 16S mtDNA sequence data. Node labels between populations indicate divergence time in millions of years. (b) Population trees per species from SNAPP based on bi-allelic SNPs from RAD-seq data. Node labels indicate node ages in SNAPP (numbers after the slash are time intervals in $\delta a\delta i$). Error bars represent SNAPP 95% HPD (Highest Posterior Density) intervals [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 6 Maps representing posterior means of effective migration and diversity surfaces for all seven taxa. Size of sampling dots represents the number of samples merged into a locality. (a) Forest taxa. (b) Generalist taxa. Upper panel: Effective migration surfaces (m); blue colours represent areas of gene flow, orange colours represent genetic barriers, and rivers are shown in blue. Lower panel: Effective diversity surfaces (q); blue colours represent areas of higher than expected diversity (green circles correspond to approximate location of refugia shown in Figure 1; A—Usambara-Kwale, B—Pugu/inland to Udzungwa, C—Lindi/Maçondes plateau, D—Mt. Mulanje, Mt. Mabu and surrounding lowlands), and orange colours represent areas of lower diversity



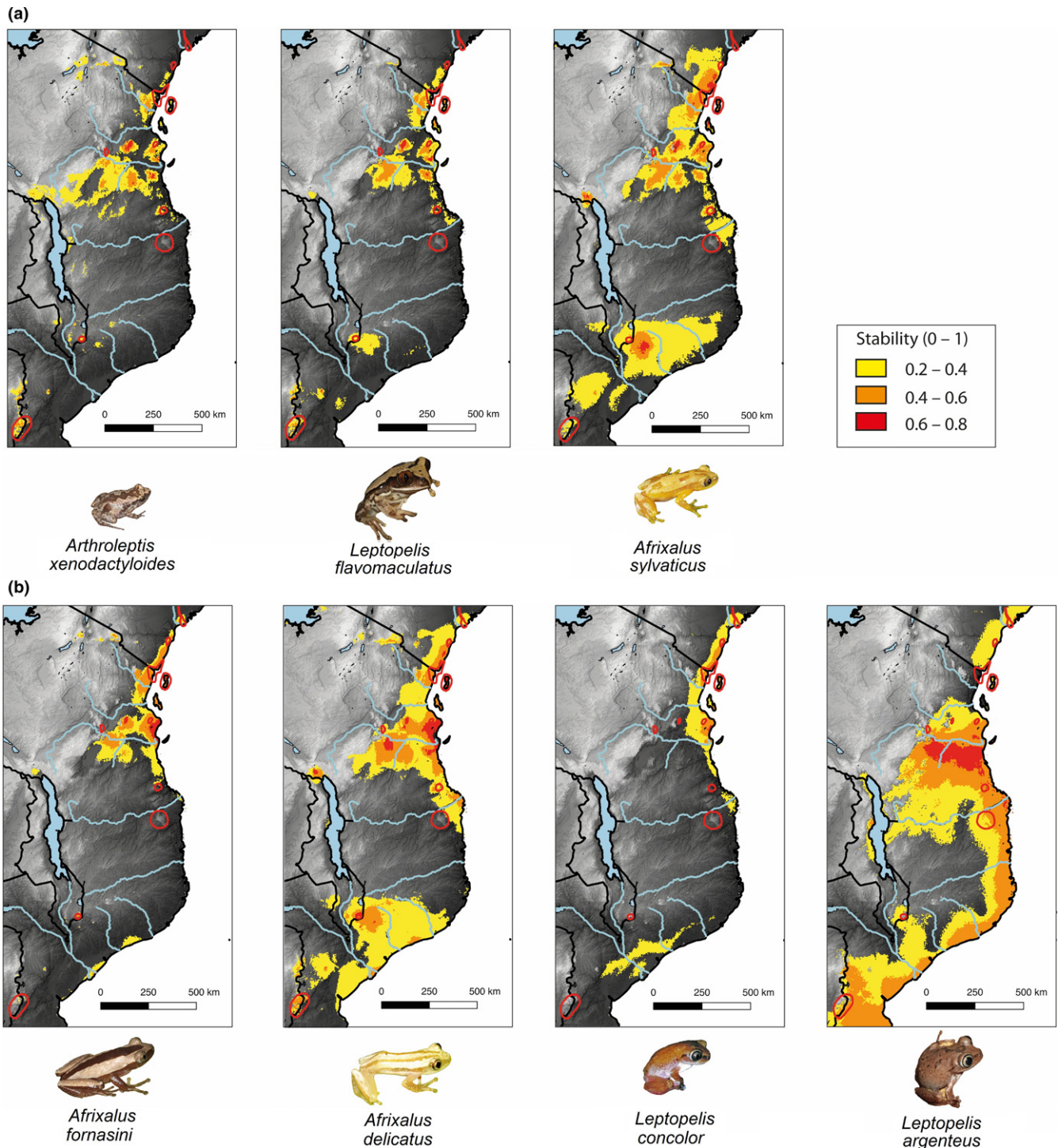


FIGURE 7 Stability models representing areas of persistent suitable habitat across from the Last Interglacial period (120 kya) until the present. (a) Forest taxa. (b) Generalist taxa. Black and white colours represent topography (white = higher elevation, black = lower elevation), and yellow-red colours represent highest stability (ranging between 0 and 1). Rivers are also shown, along with refugia identified by Burgess et al. (1998) shown as red polygons, matching the labelled refugia in Figure 1a [Colour figure can be viewed at wileyonlinelibrary.com]

location of major rivers (Figure 3; i: Pangani, ii: Rufiji-Great Ruaha-Kilombero, iii: Lukuledi/Ruvuma. iv: Lúrio). For *A. sylvaticus*, only the Pangani and Lúrio rivers represented barriers. For the generalist taxa *A. fornasini* and *L. argenteus*, we identified a single barrier around the Ruvuma/Lúrio rivers in Mozambique; *A. delicatus* and *L. concolor* showed low migration around the Pangani River,

however, extending across a wider area of coastal Tanzania for *A. delicatus*. Despite population barriers across taxa coinciding closely with rivers, we caution that they do not consistently structure all populations, and other features such as topography, ecological gradients and forest refugia may have played a role in shaping the observed patterns.

We identified at least three main areas of high effective diversity (Figure 6), which correspond to the refugial areas outlined in Figure 1a as well as contact areas where genetically distinct populations meet. High effective diversity is found in a large area surrounding the Usambara-Kwale refuge for *L. flavomaculatus*, *A. sylvaticus*, *A. fornasini* and *L. concolor*, extending southwards to southern Tanzania for the latter two, generalist taxa. A second large area of high effective diversity is found in southern Tanzania and northern Mozambique, including the Lindi and Maçondes plateau refugia (for the forest-restricted *A. xenodactyloides* and *L. flavomaculatus* and the generalists *A. delicatus* and *L. argenteus*). The Afromontane regions of Mozambique close to the Mt. Mulanje massif, Mt. Mabu and surrounding lowlands further support high effective diversity for *A. xenodactyloides*, *A. fornasini*, *A. sylvaticus* and *A. delicatus*. However, patterns are idiosyncratic across taxa, and some of these same areas demonstrate lower effective diversity for *A. xenodactyloides* and *A. delicatus* (Usambara-Kwale), *A. fornasini*, *A. sylvaticus* and *L. argenteus* (southern Tanzania/northern Mozambique), and *L. flavomaculatus* and *L. argenteus* (Mt. Mulanje massif and surrounding areas).

3.6 | Ecological niche and stability models

The best ENMS per taxon had AUC scores between 0.89 and 0.99 (Supporting Information Table S8), with Bio14 (precipitation of driest month) consistently having the highest contribution to the models for six taxa along with input from Bio4 (temperature seasonality) and Bio2 (mean diurnal range). Together, these three variables accounted for 70.9%–94.9% of the variation in the models. For *L. argenteus* and *L. concolor*, altitude had the highest model contribution, along with Bio2 (mean diurnal range) and Bio18 (precipitation of warmest quarter), together contributing to 63.3%–87.8% of the variation in the models. Stability models per taxon (Figure 7) showed that several areas in the Northern Zanzibar-Inhambane coastal forest ecoregion have remained more climatically stable over time. A difference between stability models for forest taxa (*A. xenodactyloides*, *L. flavomaculatus*, *A. sylvaticus*) and generalist taxa is apparent, with areas of stability being smaller and more fragmented for forest taxa. These areas of stability are situated at proposed refugia (in particular Usambara-Kwale, Pugu hills, Uluguru, Udzungwa, Lindi) and in southern CFEA regions such as Mt. Mulanje (Malawi), Mt. Mabu (Mozambique) and Haroni-Rusitu (Zimbabwe). With the exception of *L. argenteus*, which is restricted to the Southern Zanzibar-Inhambane ecoregion, the generalist taxa display mostly congruent patterns, with larger areas of stability that extend into non-forest (savannah and Miombo woodland) habitats.

4 | DISCUSSION

At present, our understanding of diversification in the CFEA is far from complete. We used analyses of population structure, demographic model selection, evolutionary relationships and stability models to evaluate diversification patterns and processes with

genome-wide data for the first time within this highly threatened biodiversity hotspot. Utilizing such data for multiple co-distributed taxa provides insight into the spatial distribution of biodiversity and the underlying diversification processes (e.g., Bell et al., 2017), which smaller spatial and genetic data sets have previously been unable to address in detail (see Davey & Blaxter, 2010; Lexer et al., 2013). Using a multi-faceted analytical strategy, our findings suggest that, although biodiversity patterns across the CFEA appear consistent with a forest refugia-driven model of diversification, this hypothesis alone is insufficient to fully explain the region's biological diversity even for forest taxa. We show that multi-taxon approaches can help to develop a more comprehensive understanding of the biotic history of the region.

4.1 | Broad-scale phylogeographic patterns

Previous species richness and endemism studies in the CFEA have recognized the existence of a biogeographic division situated between the northern (Zanzibar) and southern (Inhambane) Zanzibar-Inhambane ecoregions (Azeria et al., 2007; Burgess et al., 1992, 1998, 2004). Together, our analyses across taxa lend support to this division, with congruent divergences between southern populations in Mozambique, Zimbabwe and Malawi, and the remaining CFEA regions in Tanzania and Kenya. The patterns shown by our data closely match numerous other phylogeographic studies in vertebrates, notably between Tanzania and Mozambique (Bertola et al., 2016; Bryja et al., 2017; Levinsky et al., 2013; Lorenzen, Heller, & Siegismond, 2012). Analyses of effective migration and diversity surfaces confirm these patterns, but show that the location of the major Tanzania-Mozambique barrier across different sampled taxa varies, being present in a similar location for some (*L. flavomaculatus*, *A. sylvaticus*, *A. fornasini*), geographically shifted (*Arthroleptis xenodactyloides*), smaller (*Leptopelis argenteus*) or absent (*Leptopelis concolor*, *Afrivalus delicatus*) for others. These differences may reflect true variation across taxa but are most likely explained by variations in geographic sampling, which may have influenced results. Further targeted sampling in underrepresented areas would help to address issues concerning robustness of our spatial estimations, especially in northern and central Mozambique and around the other major phylogeographic barriers revealed by our analyses (e.g., through Tanzania and Mozambique in the vicinity of the Pangani, Rufiji-Great Ruaha-Kilombero, Lukuledi-Ruvuma and Lúrio rivers, which also occur between refugial areas). The presence of range-restricted diversity in these areas has been documented for several taxonomic groups (Burgess & Clarke, 2000; Burgess et al., 1998), including amphibian populations (Barratt, 2017; Barratt et al., 2017a, 2017b; Bwong et al., 2017). In East Africa, such patterns are often associated with vicariant diversification through a forest refuge model of speciation (Endler, 1982; Haffer, 1969, 1997; Mayr & O'Hara, 1986; Moreau, 1954; Moritz et al., 2000; Plana, 2004) or attributed to ecological change (Lorenzen et al., 2012).

Most divergence events observed in our focal taxa occurred during the Pliocene–Pleistocene based on mtDNA analysis. Differences

in topologies with SNP data and difficulties in inferring absolute dates without an accurate mutation rate for amphibians make direct comparisons of node ages across these data sets difficult. Despite this, we note that almost all divergences are temporally consistent between *SNAPP* and *δαδi*, with time interval parameters (in genetic units) being highly correlated. Although gene flow can have adverse effects on divergence date estimates using Bayesian coalescent models (Leaché, Fujita, Minin, & Bouckaert, 2014b), we found evidence for a lack of gene flow in six of our seven population comparisons, indicating age estimates are not likely to be underestimated from this cause. However, we do note that divergence times may be overestimated when dealing with closely related lineages due to inappropriate models of nucleotide evolution using genome-wide data (Lischer, Excoffier, & Heckel, 2014). The Plio-Pleistocene divergences observed in our focal taxa for both forest and generalist taxa are particularly interesting as they provide further evidence of older diversification events across the CFEA, in addition to the more recent divergences likely due to Pleistocene forest refugia. Together, evolutionary relationships and demographic analyses lend support to the high complexity of temporal and spatial diversification processes that have occurred across the CFEA, which have resulted in idiosyncratic responses across forest and generalist taxa.

4.2 | Diversification processes in forest vs. generalist taxa

For the forest taxa (*A. xenodactyloides*, *L. flavomaculatus*, and *A. sylvaticus*), geographic distributions of populations were well defined, divergences between populations were in general early, and stability models showed more fragmented distributions of stable habitats since at least the Last Interglacial than for generalist taxa. Effective migration and diversity surfaces highlighted more numerous but geographically smaller dispersal barriers than detected for the generalist taxa, which encompass areas of low elevation and arid habitat (e.g., north of the Rufiji and Lukuledi rivers and south of the Ruvuma). Although these lines of evidence are in general accordance with the forest refuge hypothesis, more detailed analyses through demographic models revealed mixed results. For *L. flavomaculatus*, simultaneous divergence followed by secondary contact is consistent with forest refugial processes. Conversely, we found evidence of parapatric divergence (ancient migration followed by size change) in *A. xenodactyloides* and a simple model of allopatric divergence (no gene flow or size changes) for *A. sylvaticus*, indicating other processes have driven divergence in forest taxa. Taken together, our molecular results and the spatial location of forest taxa populations only partially support a previous assertion that some CFEA taxa may have evolved from isolated forest refugial taxa since the Miocene, with predicted population divergence and periods of re-connectivity during the cyclical expansion and contraction of forests during the Pliocene and Pleistocene (Blackburn & Measey, 2009; Pickersgill, 2005). However, landscape barriers (*A. sylvaticus*) and ecological gradients (*A. xenodactyloides*) may have also contributed to the divergence of forest taxa, with no strong evidence for recent size

changes caused by human influence. The overall patterns of generalist taxa (*A. forasini*, *L. argenteus*, *L. concolor* and *A. delicatus*) demonstrated earlier than expected divergences for all taxa except *L. argenteus* and *L. concolor*, and limited gene flow between populations despite fewer dispersal barriers shown by *EEMS* analyses. Demographic model selection results are highly congruent across generalist taxa, with all four patterns consistent with allopatric divergence without subsequent migration. As for forest taxa, the geographic distribution of populations is highly structured and appears to correspond more closely to river barriers (Pangani, Rufiji, Great Ruaha, Kilombero, Lukuledi, Ruvuma and Lúrio) than topographic barriers, in line with previous results for amphibian populations in sub-Saharan Africa (Lawson, 2013; Measey, Galbusera, Breyne, & Matthysen, 2007; Portik et al., 2017). Stability models for these taxa highlight several areas that could be considered as potential refugia (e.g., Usambara-Kwale, Pugu hills, Lindi), although the large sizes of these are likely due to the broader niche of generalists compared to those of forest taxa. However, generalists may facultatively occupy forest habitats, being sympatric with forest taxa in some cases (e.g., see *A. delicatus* and *A. sylvaticus* distributions in Figure 2c,e). In summary, the diversity patterns shown by generalist taxa are best interpreted as the result of vicariance due to landscape barriers (most likely rivers) rather than ecological gradients or forest refugia, and our goodness of fit tests indicate that our best-ranked models are good explanations of the JSFS in all taxa except *L. concolor*, which should be interpreted with caution, although human impacts could potentially have played an additional role.

4.3 | Understanding tropical diversification using multiple taxa

Taxa found across the CFEA, as in other heterogeneous tropical biodiversity hotspots, are a rich mixture of old and young taxa, each with unique ecological characteristics. Inferring diversification processes in tropical biodiversity hotspots is therefore challenging, as even congruent biodiversity patterns are likely to be generated by highly complex processes that vary both temporally and spatially. With this study, we managed to capture some of these complex processes for forest and generalist taxa and were able to evaluate potential diversification processes at work across taxa using several high-throughput data sets. In doing so, we were able to explicitly test existing hypotheses against alternative diversification modes for the first time in this region and make a broad evaluation of the forest refuge model. Forest refugial processes appear to be only partially responsible for the current diversity in the CFEA, and a range of other diversification mechanisms support the idiosyncrasy of these processes across taxa. Although we found some clear differences between forest specialists and generalist taxa, counter to our predictions the forest specialists were less consistent in diversification mechanisms than generalists.

While the conceptual framework we employ is by no means the only available option to address such hypotheses, approaches such as ours can help to reveal the nuances of diversification which lie

behind the correct but rather coarsely defined forest refuge hypothesis. Building upon this work, new statistical tools and techniques for handling high-throughput data, such as minimizing allelic dropout while maximizing amounts of informative data and building more accurate demographic models will be key to leveraging optimal information from RAD-seq data sets in particular. Moreover, newly published model-based methods for genome-wide SNP data (Smith et al. 2017; Xue & Hickerson, 2017) as well as those already existing for mtDNA (Huang, Takebayashi, Qi, & Hickerson, 2011; Oaks, 2014) will enable further work for assessing concordant patterns of co-diversification amongst co-distributed taxa in tropical hotspots. These methods allow empirical investigation of the synchronicity of divergences within and between species and populations, detecting if co-distributed taxa exhibit shared responses to major historical events. We strongly encourage future work in tropical regions to take advantage of these emerging tools as well as those we utilized here, and as burgeoning quantities of genome-wide spatial and molecular data become available, we advocate that research collaborations incorporating multiple taxonomic groups will further improve our understanding of the evolutionary processes occurring in biodiversity hotspots.

ACKNOWLEDGEMENTS

CDB was funded by a PhD scholarship from the Humer Foundation via the Centre for African Studies Basel. Fieldwork in Tanzania was carried out by CDB under COSTECH permit 2013-341-NA-2013-121, and Kenyan fieldwork was conducted by BAB under Kenya Wildlife Service (KWS/BRM/5001) and Kenya Forest service (MUS/1/KFS/VOL.II/4) permits, funded by Freiwillige Akademische Gesellschaft Basel. Sequencing costs were partially contributed to by Universität Basel Fonds zur Förderung des akademischen Nachwuchses DUW2100 awarded to HCL, and we acknowledge a grant from the European Science Foundation ConGenOmics program (6720 to CDB). Thanks to Kim Howell, Wilirk Ngalason and Chacha Werema at the University of Dar es Salaam and Tanzania Forest Conservation Group for field advice and to Patrick K. Malonza, Victor Wasonga, Joash Nyamache and Vincent Muchai at the National Museums of Kenya. We are grateful to Lucinda Lawson, Mark Wilkinson, David Gower, Krystal Tolley, Harith Farooq, W. Daniel Kissling, Gabriela Bittencourt-Silva, Michele Menegon, Simon Maddock and Marco Crotti for discussion and advice. The Natural History Museum, London, Frontier, Field Museum of Natural History, Chicago, Museum of Vertebrate Zoology, California, Museum of Comparative Zoology, Harvard (Breda Zimkus and Joanna Larson), are thanked for additional specimens. CDB and REO further acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—FZT 118. This project has been partially conducted in the framework of the iDiv-Flexpool—the internal funding mechanism of iDiv. We are indebted to Walter Salzburger and Marius Roesti at the University of Basel for their advice on RAD-seq sampling design and to Zuzana Musilova and Bernd Egger

for guidance on library preparation. Sequencing was conducted at the D-BSSE facility in Basel, Switzerland, and calculations were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing centre at University of Basel. Finally, we thank Veit Arlt (Centre for African Studies Basel), Tim Littlewood (NHM London) and the iDiv Open Science Publication Fund for supporting Open Access publication of this article.








DATA ACCESSIBILITY

All raw, unprocessed sequences are deposited in NCBI GenBank (Accession nos MG871749–MG871982 for 16S mtDNA) and the Sequence Read Archive (Accession no. SRP150605 for RAD-seq). We provide a large package on DRYAD (<https://doi.org/10.5061/dryad.315km76>) that includes our final RAD-seq filtered “haplotypes” files, input files for analyses, including mtDNA (also submitted to GenBank). We include data and results for our analyses in this package (ADMIXTURE, DAPC, BEAST, SNAPP, EEMS, $\delta a\delta i$, STACKS and ENMS). All newly created 3D demographic models, along with scripts for performing goodness of fit tests, are freely available in an updated version of the $\delta a\delta i$ analysis pipeline found at https://github.com/dportik/dadi_pipeline.

AUTHOR CONTRIBUTIONS

The research project was designed by C.D.B. and S.P.L., with guidance on analyses from D.M.P., R.J., R.E.O., H.C.L. and J.W.S. Fieldwork was conducted and molecular data were collected by C.D.B., B.A.B., H.C.L., S.P.L. All data processing and analyses were performed by C.D.B., with assistance from D.M.P. The manuscript was written by C.D.B. and S.P.L., with contributions from all authors.

ORCID

Christopher D. Barratt  <http://orcid.org/0000-0003-3267-8855>
 Robert Jehle  <http://orcid.org/0000-0003-0545-5664>
 H. Christoph Liedtke  <http://orcid.org/0000-0002-6221-8043>
 Renske E. Onstein  <http://orcid.org/0000-0002-2295-3510>
 Daniel M. Portik  <http://orcid.org/0000-0003-3518-7277>
 Jeffrey W. Streicher  <http://orcid.org/0000-0002-3738-4162>
 Simon P. Loader  <https://orcid.org/0000-0003-4162-0575>

REFERENCES

- Ahl, E. (1929). Zur Kenntnis der afrikanischen Baumfrosch-Gattung *Lep-topelis*. *Sitzungsberichte der Gesellschaft Naturforschender Freunde zu Berlin*, 1929, 185–222.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17, 81–92. <https://doi.org/10.1038/nrg.2015.28>

- Axelrod, D., & Raven, P. H. (1978). Late Cretaceous and Tertiary vegetation history of Africa. In *Biogeography and Ecology of Southern Africa*, pp. 77–130. <https://doi.org/10.1007/978-94-009-9951-0>
- Azeria, E., Sanmartín, I., Ås, S., Carlson, A., & Burgess, N. (2007). Biogeographic patterns of the East African coastal forest vertebrate fauna. *Biodiversity and Conservation*, 16, 883–912. <https://doi.org/10.1007/s10531-006-9022-0>
- Barratt, C. D. (2017). Biodiversity patterns and conservation of the coastal forests of Eastern Africa. PhD thesis, University of Basel. <https://edoc.unibas.ch/55679/>
- Barratt, C. D., Bwong, B. A., Onstein, R. E., Rosauer, D. F., Menegon, M., Doggart, N., ... Loader, S. P. (2017a). Environmental correlates of phylogenetic endemism in amphibians and the conservation of refugia in the coastal forests of Eastern Africa. *Diversity and Distributions*, 23, 875–887. <https://doi.org/10.1111/ddi.12582>
- Barratt, C. D., Lawson, L. P., Bittencourt-Silva, G. B., Doggart, N., Morgan-Brown, T., Nagel, P., & Loader, S. P. (2017b). A new, narrowly distributed, and critically endangered species of spiny-throated reed frog (Anura: Hyperoliidae) from a highly threatened coastal forest reserve in Tanzania. *Herpetological Journal*, 27, 13–24. <https://www.thebhs.org/publications/the-herpetological-journal/volume-27-number-1-january-2017/944-02-a-new-narrowly-distributed-and-critically-endangered-species-of-spiny-throated-reed-frog-anura-hyperoliidae-from-a-highly-threatened-coastal-forest-reserve-in-tanzania/file>
- Barratt, C. D., Tonelli, E., Menegon, M., Doggart, N., Bowkett, A., Harris, W. E., Howell, K., Ngason, W., & Loader, S. P. (2014). Fragmented habitats and species: The challenges of amphibian conservation in Tanzania today. *Froglog*, 111, 63–64.
- Bell, R. C., Parra, J. L., Badjedje, G., Barej, M. F., Blackburn, D. C., Burger, M., ... Zamudio, K. (2017). Idiosyncratic responses to climate-driven forest fragmentation and marine incursions in reed frogs from Central Africa and the Gulf of Guinea Islands. *Molecular Ecology*, 26, 5223–5244. <https://doi.org/10.1111/mec.14260>
- Bertola, L. D., Jongbloed, H., van der Gaag, K. J., de Knijff, P., Yamaguchi, N., Hooghiemstra, H., ... de longh, H. H. (2016). Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the lion (*Panthera leo*). *Scientific Reports*, 6, 30807. <https://doi.org/10.1038/srep30807>
- Bianconi, G. G. (1849). “1848”. Lettera al Dottore Filippo De Filippi, Professore di Zoologia a Torino, sopra alcune nuove specie di Rettili del Mozambico. *Nuovi Annali delle Scienze Naturali. Serie 2, Bologna*, 10, 106–109.
- Bittencourt-Silva, G. B., Lawson, L. P., Tolley, K., Portik, D. M., Barratt, C. D., Nagel, P., & Loader, S. P. (2017). 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. *Molecular Phylogenetics and Evolution*, 114, 261–270. <https://doi.org/10.1016/j.ympev.2017.06.022>
- Blackburn, D. C., & Measey, G. J. (2009). Dispersal to or from an African biodiversity hotspot? *Molecular Ecology*, 18, 1904–1915. <https://doi.org/10.1111/j.1365-294X.2009.04156.x>
- Bouckaert, R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17, 42. <https://doi.org/10.1186/s12862-017-0890-6>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Brown, J. H. (2013). Why are there so many species in the tropics? *Journal of Biogeography*, 41, 8–22.
- Brown, J. L. (2014). SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, 5, 694–700. <https://doi.org/10.1111/2041-210X.12200>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917–1932. <https://doi.org/10.1093/molbev/mss086>
- Bryja, J., Šumbera, R., Kerbis Peterhans, J. C., Aghová, T., Bryjová, A., Mikula, O., ... Verheyen, E. (2017). Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the evolution of African forests since late Miocene. *Journal of Biogeography*, 44, 182–194. <https://doi.org/10.1111/jbi.12890>
- Burgess, N., & Clarke, G. P. (2000). Coastal Forests of Eastern Africa. In *IUCN Forest Conservation Programme*, 2010, 443 pp.
- Burgess, N. D., Clarke, G. P., & Rodgers, W. A. (1998). Coastal forests of eastern Africa: Status, endemism patterns and their potential causes. *Biological Journal of the Linnean Society*, 64, 337–367. <https://doi.org/10.1111/j.1095-8312.1998.tb00337.x>
- Burgess, N., D'Amico Hales, J., Underwood, E., Dinerstein, E., & Ecoregion, W. W. F. (Eds.) (2004). Terrestrial Ecoregions of Africa and Madagascar: A Conservation Assessment. In *World Wildlife Fund Ecoregion Assessments*, 2nd ed. Washington DC: Island Press.
- Burgess, N. D., Mwasumbi, L. B., Hawthorne, W. J., Dickinson, A., & Doggett, R. A. (1992). Preliminary assessment of the distribution, status and biological importance of coastal forests in Tanzania. *Biological Conservation*, 62, 205–218. [https://doi.org/10.1016/0006-3207\(92\)91048-W](https://doi.org/10.1016/0006-3207(92)91048-W)
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed. New York, NY: Springer-Verlag.
- Bwong, B. A., Nyamache, J. O., Malonza, P. K., Wasonga, D. V., Ngwava, J. M., Barratt, C. D., ... Loader, S. P. (2017). Amphibian diversity in Shimba Hills National Reserve, Kenya: A comprehensive list of specimens and species. *Journal of East African Natural History*, 106, 19–46. <https://doi.org/10.2982/028.106.0104>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Charles, K. L., Bell, R. C., Blackburn, D. C., Murger, M., Fujita, M. K., Gvoždík, V., ... Portik, D. M. (2018). Sky, sea, and forest islands: Diversification in the African leaf-folding frog *Afraxalus paradorsalis* (Anura: Hyperoliidae) of the Lower Guineo-Congolian rain forest. *Journal of Biogeography*, 45, 1781–1794. <https://doi.org/10.1111/jbi.13365>
- Damaseno, R., Strangas, M. L., Carnaval, A. C., Rodrigues, M. T., & Moritz, C. (2014). Revisiting the vanishing refuge model of diversification. *Frontiers in Genetics*, 5, 353.
- Davey, J. L., & Blaxter, M. W. (2010). RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, 9, 416–423. <https://doi.org/10.1093/bfgp/elq031>
- Demenocal, P. B. (1995). Plio-Pleistocene African climate. *Science*, 270, 53–59. <https://doi.org/10.1126/science.270.5233.53>
- Endler, J. A. (1982). Pleistocene forest refuges: Fact or fancy. In G. T. Prance (Ed.), *Biological diversification in the tropics* (pp. 641–657). New York: Columbia University Press.
- Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology*, 772, 157–178.
- Fjeldså, J., & Lovett, J. C. (1997). Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation*, 6, 325–346. <https://doi.org/10.1023/A:1018356506390>
- Gascon, C., Malcom, J. R., Patton, J. L., da Silva, M. N. F., Bogart, J. P., Lougheed, S. C., ... Boag, P. T. (2000). Riverine barriers and the geographic distribution of Amazonian species. *Proceedings of the National Academy of Sciences USA*, 97, 13672–13677. <https://doi.org/10.1073/pnas.230136397>

- Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, 405, 8.
- Günther, A. C. L. G. (1864). Report on a collection of reptiles and fishes made by Dr. Kirk in the Zambesi and Nyassa regions. *Proceedings of the Zoological Society of London*, 1864, 303–314.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, e1000695.
- Haffer, J. (1969). Speciation in Amazonian forest birds. *Science*, 165, 131–137. <https://doi.org/10.1126/science.165.3889.131>
- Haffer, J. (1997). Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation*, 6, 451–477. <https://doi.org/10.1023/A:1018320925954>
- Hewitt, J. (1933). Descriptions of some new reptiles and a frog from Rhodesia. *Occasional Papers of the National Museum of Southern Rhodesia*, 2, 45–50.
- 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. *International Journal of Climatology*, 25, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Huang, H., & Knowles, L. (2016). Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology*, 65, 357–365. <https://doi.org/10.1093/sysbio/syu046>
- Huang, W., Takebayashi, N., Qi, Y., & Hickerson, M. J. (2011). MTML-Bayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics*, 12, 1.
- IUCN. (2017). *The IUCN Red List of Threatened Species*. Version 2017-2. Available from: <http://www.iucnredlist.org>. Accessed: 6th September 2017.
- Jeffries, D. L., Copp, G. H., Lawson Handley, L. J., Olsén, H., Sayer, C. D., & Hänfling, B. (2015). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, a real data informed perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, 25, 2997–3018.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jongsma, G. F. M., Barej, M. F., Barratt, C. D., Burger, M., Conradie, W., Ernst, R., ... Blackburn, D. C. (2017). Diversity and biogeography of frogs in the genus *Amnirana* (Anura: Ranidae) across sub-saharan Africa. *Molecular Phylogenetics and Evolution*, 120, 274–285. <https://doi.org/10.1016/j.ympev.2017.12.006>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kirschel, A. N. G., Slabbekoorn, H., Blumstein, D. T., Cohen, R. E., de Kort, S. R., Buermann, W., & Smith, T. B. (2011). Testing alternative hypotheses for evolutionary diversification in an African songbird: Rainforest refugia versus ecological gradients. *Evolution*, 65, 3162–3174. <https://doi.org/10.1111/j.1558-5646.2011.01386.x>
- Lawson, L. P. (2013). Diversification in a biodiversity hot spot: Landscape correlates of phylogeographic patterns in the African spotted reed frog. *Molecular Ecology*, 22, 1947–1960. <https://doi.org/10.1111/mec.12229>
- Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014b). Species delimitation using genome-wide SNP data. *Systematic Biology*, 63, 534–542. <https://doi.org/10.1093/sysbio/syu018>
- Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014a). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, 63, 17–30. <https://doi.org/10.1093/sysbio/syt049>
- Levinsky, I., Araújo, M. B., Nogués-Bravo, D., Haywood, A. M., Valdes, P. J., & Rahbek, C. (2013). Climate envelope models suggest spatio-temporal co-occurrence of refugia of African birds and mammals. *Global Ecology and Biogeography*, 22, 351–363. <https://doi.org/10.1111/geb.12045>
- Lexer, C., Mangili, S., Bossolini, E., Forest, F., Stölting, K. N., Pearman, P. B., ... Salamin, N. (2013). “Next generation” biogeography: Towards understanding the drivers of species diversification and persistence. *Journal of Biogeography*, 40, 1013–1022. <https://doi.org/10.1111/jbi.12076>
- Linck, E. B., & Battey, C. J. (2017). Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *Biorxiv*. <https://doi.org/10.1101/188623>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- Lischer, H. E. L., Excoffier, L., & Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of *Microtus* voles. *Molecular Biology and Evolution*, 31, 817–831. <https://doi.org/10.1093/molbev/mst271>
- Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P., & Linder, C. R. (2012). SATé-II: Very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology*, 61, 90. <https://doi.org/10.1093/sysbio/syr095>
- Lorenzen, E. D., Heller, R., & Siegmund, H. R. (2012). Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21, 3656–3670. <https://doi.org/10.1111/j.1365-294X.2012.05650.x>
- 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. *Quaternary Science Reviews*, 101, 1–17. <https://doi.org/10.1016/j.quascirev.2014.06.012>
- Mayr, E., & O'Hara, R. J. (1986). The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution*, 40, 55–67. <https://doi.org/10.1111/j.1558-5646.1986.tb05717.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. *Conservation Genetics*, 8, 1177–1188. <https://doi.org/10.1007/s10592-006-9272-0>
- Moreau, R. E. (1954). The distribution of African evergreen-forest birds. *Proceedings of the Linnean Society of London*, 165, 35–46. <https://doi.org/10.1111/j.1095-8312.1954.tb00707.x>
- Moritz, C., Patton, J. L., Schneider, C. J., & Smith, T. B. (2000). Diversification of rainforest faunas: An integrated molecular approach. *Annual Review of Ecology and Systematics*, 31, 533–563. <https://doi.org/10.1146/annurev.ecolsys.31.1.533>
- Mumbi, C. T., Marchant, R., Hooghiemstra, H., & Wooller, M. J. (2008). Late Quaternary vegetation reconstruction from the Eastern Arc Mountains, Tanzania. *Quaternary Research*, 69, 326–341. <https://doi.org/10.1016/j.yqres.2007.10.012>
- Ntie, S., Davis, A., Hils, K., Mickala, P., Thomassen, H., Vanthomme, M. K., ... Anthony, M. (2017). Evaluating the role of Pleistocene refugia, rivers and environmental variation in the diversification of central African duikers (genera *Cephalophus* and *Philantomba*). *BMC Evolutionary Biology*, 17, 212. <https://doi.org/10.1186/s12862-017-1054-4>
- Oaks, J. R. (2014). An improved approximate-Bayesian model-choice method for estimating shared evolutionary history. *BMC Evolutionary Biology*, 14, 150.
- Ohler, A., & Frétey, T. (2015). Going Back to Rovuma: The frog fauna of a coastal dry forest, and a checklist of the amphibians of Mozambique. *Journal of East African Natural History*, 103, 73–124. <https://doi.org/10.2982/028.103.0203>

- Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48, 94–100. <https://doi.org/10.1038/ng.3464>
- Pfeffer, G. (1893). Ostafrikanische Reptilien und Amphibien, gesammelt von Herrn Dr. F. Stuhlmann im Jahre 1888 und 1889. *Jahrbuch der Hamburgischen Wissenschaftlichen Anstalten*, 10, 69–105.
- Pickersgill, M. (1984). Three new *Afrixalus* (Anura: Hyperoliidae) from South-Eastern Africa". *Durban Museum Novitates*, 13, 203–220.
- Pickersgill, M. (2005). The taxonomy and ethology of the *Afrixalus stuhlmanni* complex (Anura: Hyperoliidae). *Steenstrupia*, 29, 1–38.
- Plana, V. (2004). Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359, 1585–1594. <https://doi.org/10.1098/rstb.2004.1535>
- Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., ... Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26, 5245–5263. <https://doi.org/10.1111/mec.14266>
- Poynton, J. C. (2006). On dwarf spiny reedfrogs in Tanzanian Eastern lowlands (Anura: *Afrixalus*). *African Journal of Herpetology*, 55, 167–169. <https://doi.org/10.1080/21564574.2006.9635548>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- Rosauer, D. F., Catullo, R. A., Vanderwal, J., & Moussalli, A. (2015). Lineage range estimation method reveals fine-scale endemism linked to Pleistocene stability in Australian rainforest herpetofauna. *PLoS ONE*, 10, e0126274. <https://doi.org/10.1371/journal.pone.0126274>
- Rosenzweig, M. (1995). *Species diversity in space and time*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511623387>
- Schiøtz, A. (1974). Revision of the genus *Afrixalus* (Anura) in eastern Africa. *Videnskabelige Meddelelser fra Dansk Naturhistorisk Forening i Kjøbenhavn*, 137, 9–18.
- Schneider, C. J., Smith, T. B., Larison, B., & Moritz, C. (1999). A test of alternative models of diversification in tropical rainforests: Ecological gradients vs. rainforest refugia. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 13869–13873. <https://doi.org/10.1073/pnas.96.24.13869>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., ... Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, 515, 406–409. <https://doi.org/10.1038/nature13687>
- Smith, T. B., Wayne, R. K., Girman, D. J., & Bruford, M. W. (1997). A role for ecotones in generating rainforest biodiversity. *Science*, 276, 1855–1857. <https://doi.org/10.1126/science.276.5320.1855>
- Smith, M. L., Ruffley, M., Espíndola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2017). Demographic model selection using random forests and the site frequency spectrum. *Molecular Ecology*, 26, 4562–4573.
- Vanzolini, P. E., & Williams, E. F. (1981). The vanishing refuge: A mechanism for ecogeographic speciation. *Papéis Avulsos de Zoologia*, 34, 251–255.
- Voelker, G., Marks, B. D., Kahindo, C., Agenonga, U., Bapeamoni, F., Duffie, L. E., ... Light, J. E. (2013). River barriers and cryptic biodiversity in an evolutionary museum. *Ecology and Evolution*, 3, 536–545. <https://doi.org/10.1002/ece3.482>
- Wallace, A. R. (1852). On the monkeys of the Amazon. *Proceedings of the Zoological Society London*, 20, 107–110.
- Xue, A. T., & Hickerson, M. J. (2017). Multi-DICE: R package for comparative population genomic inference under hierarchical co-demographic models of independent single-population size changes. *Molecular Ecology Resources*, 17, e212–e224.
- Zhen, Y., Harrigan, R. J., Rugg, K. C., Anderson, E. C., Ng, T. C., Lao, S., ... Smith, T. B. (2017). Genomic divergence across ecological gradients in the Central African rainforest songbird (*Andropadus virens*). *Molecular Ecology*, 26, 4966–4977. <https://doi.org/10.1111/mec.14270>
- Zimkus, B., Lawson, L., Barej, M., Barratt, C. D., Channing, A., Dehling, J. M., ... Lötters, S. (2017). Leapfrogging into new territory: How Mascarene ridged frogs have diversified across Africa. *Molecular Phylogenetics and Evolution*, 106, 254–269. <https://doi.org/10.1016/j.ympev.2016.09.018>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Barratt CD, Bwong BA, Jehle R, et al. Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Mol Ecol*. 2018;27:4289–4308. <https://doi.org/10.1111/mec.14862>