Received: 3 June 2022 | Revised: 7 November 2022 | Accepted: 16 January 2023

DOI: 10.1111/mec.16867

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Evidence for ecotone speciation across an African rainforest-savanna gradient

Adam H. Freedman[1] | Ryan J. Harrigan[2] | Ying Zhen[2,3] | Alison M. Hamilton[4] | Thomas B. Smith[2,5]

[1]Faculty of Arts and Sciences Informatics Group, Harvard University, Cambridge, Massachusetts, USA

[2]Centre for Tropical Research and Institute of the Environment and Sustainability, University of California, Los Angeles, California, USA

[3]School of Life Sciences, Westlake University, Hangzhou, China

[4]Department of Biological Sciences, University of Massachusetts—Lowell, Lowell, Massachusetts, USA

[5]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California, USA

Correspondence
Adam H. Freedman, Faculty of Arts and Sciences Informatics Group, Harvard University, 38 Oxford Street, Cambridge, Massachusetts 01328, USA.
Email: adamfreedman@fas.harvard.edu

Funding information
Fulbright Association; National Geographic Society Education Foundation; National Science Foundation, Grant/Award Number: DEB-9726425, 1243524 and IRCEB9977072; QCB Collaboratory Postdoctoral Fellowship; United States Environmental Protection Agency STAR Fellowship Program

Handling Editor: Sean D Schoville

Abstract

Accelerating climate change and habitat loss make it imperative that plans to conserve biodiversity consider species' ability to adapt to changing environments. However, in biomes where biodiversity is highest, the evolutionary mechanisms responsible for generating adaptative variation and, ultimately, new species are frequently poorly understood. African rainforests represent one such biome, as decadal debates continue concerning the mechanisms generating African rainforest biodiversity. These debates hinge on the relative importance of geographic isolation versus divergent natural selection across environmental gradients. Hindering progress is a lack of robust tests of these competing hypotheses. Because African rainforests are severely at-risk due to climate change and other anthropogenic activities, addressing this long-standing debate is critical for making informed conservation decisions. We use demographic inference and allele frequency-environment relationships to investigate mechanisms of diversification in an African rainforest skink, *Trachylepis affinis*, a species inhabiting the gradient between rainforest and rainforest-savanna mosaic (ecotone). We provide compelling evidence of ecotone speciation, in which gene flow has all but ceased between rainforest and ecotone populations, at a level consistent with infrequent hybridization between sister species. Parallel patterns of genomic, morphological, and physiological divergence across this environmental gradient and pronounced allele frequency-environment correlation indicate speciation is mostly probably driven by ecological divergence, supporting a central role for divergent natural selection. Our results provide strong evidence for the importance of ecological gradients in African rainforest speciation and inform conservation strategies that preserve the processes that produce and maintain biodiversity.

KEYWORDS
adaptation, conservation, genetics, population genetics – empirical, speciation

1 | INTRODUCTION

Over half of the planet's biodiversity is centred in the tropics (Primack, 2005), making tropical rainforests crucial areas for conservation. Effective conservation efforts must consider the evolutionary processes that generate and maintain biodiversity. However, long-standing debates over the evolutionary processes that generate and maintain rainforest biodiversity and especially the lack of robust tests of competing hypotheses have impeded holistic approaches to conservation efforts. A central question in this debate hinges on the

relative importance of geographic isolation versus divergent natural selection across environmental gradients to speciation. Robust tests to disentangle their respective contributions are of immense practical importance, as these two mechanisms have different implications for the conservation of current and future diversity.

Genetic examinations of diverse African rainforest species, including birds (Endler, 1982; Fuchs & Bowie, 2015; Smith et al., 1997, 2011, 2021; Zhen et al., 2017), mammals (Anthony et al., 2007; Gonder et al., 2011; Mizerovská et al., 2019; Nicolas et al., 2011; Ntie et al., 2017; Quérouil et al., 2003), reptiles (Allen et al., 2021; Freedman, Thomassen, et al., 2010), amphibians (Barratt et al., 2018; Charles et al., 2018; Jaynes et al., 2021; Jongsma et al., 2018; Leaché et al., 2019; Portik et al., 2017), and plants (Debout et al., 2011; Ley et al., 2017; Piñeiro et al., 2017), have revealed varying degrees of population structure. Frequently, rough geographic concordance between patterns of genetic differentiation and hypothesized rainforest refugia are used to support arguments for allopatric isolation in ancient forest refugia as the primary generator of rainforest biodiversity; far fewer studies indicate an exclusive (Allen et al., 2021) or contributing (Anthony et al., 2007; Leaché et al., 2019; Portik et al., 2017) role for river barriers. Moreover, only a few recent studies have attempted to test the refugial model's fit to genome-scale data (Barratt et al., 2018; Charles et al., 2018; Jaynes et al., 2021; Leaché et al., 2019; Portik et al., 2017).

The cyclical dynamics of rainforest cover are not in question. Strong evidence for periodic fragmentation of rainforest into refugia is found in longitudinal palaeoecological studies (Anhuf et al., 2006; Maley, 1996; Maley & Brenac, 1998), and detailed plant genetic investigations (Helmstetter et al., 2020; Piñeiro et al., 2021). The latter jointly uncover population structure among putative refugial areas, reductions in genetic diversity with distance from them, and expected post-glacial population expansions out of them during warmer, wetter periods. However, claims for the importance of refugia to the diversification of taxa occupying rainforest environments are often made despite few efforts to examine possible roles of environmental gradients, contributions of environmental variation to genetic divergence, or analyses of adaptative phenotypic variation. In the absence of such data, an alternative hypothesis, namely divergence-with-gene-flow speciation (Rice & Hostert, 1993), cannot be adequately tested. Consequently, the refugia hypothesis can become an implicit null model without the evidence to justify this status. For speciation along the African rainforest-savanna gradient, evidence has been indirect and often many decades old and based on data such as the frequency of contact zones between forest and savanna habitats versus those between refugia (Endler, 1982).

In contrast, many more recent studies find divergent natural selection across environmental gradients may be an essential generator of African rainforest biodiversity (Freedman, Thomassen, et al., 2010; Kirschel et al., 2011; Smith et al., 1997, 2005, 2011; Zhen et al., 2017). These studies rely on either differentiation in phenotypic traits involved in local adaptation or evidence for isolation-by-environment (IBE; Wang & Bradburd, 2014), in which allele frequency–environment correlations cannot be explained solely by geographic distance—to provide evidence for selection along the rainforest-savanna gradient. While these studies indicate diversifying natural selection may be contributing to an ongoing process of genomic divergence, none of them have demonstrated that speciation across the gradient has actually occurred. Incomplete or partial inferences of refugial isolation may collectively highlight a gap in our knowledge regarding the mechanisms that generate and maintain biodiversity in African tropical forests. We address this gap, using genomic data to simultaneously test refugia and ecological gradient hypotheses of diversification in an African rainforest lizard, the skink *Trachylepis affinis*. *T. affinis* inhabits a wide range of bioclimates, spanning both rainforests where hypothesized refugia are located and the gallery forest-savanna transition zone (or ecotone). Rainforest and ecotone populations of the African rainforest skink *T. affinis* are morphologically differentiated in body size, limb length, and head shape (Freedman, Thomassen, et al., 2010), traits all well-known to be involved in local adaptation in reptiles (Calsbeek & Irschick, 2007; Donihue et al., 2018; Schneider et al., 1999) (Figure 1b,c). They are also physiologically divergent in a manner consistent with local adaptation to different temperature regimes (Yuan et al., 2018). Broad geographic distribution and environment-associated phenotypic variation make *T. affinis* an ideal system for testing alternative diversification hypotheses simultaneously. Our goals in this study were to use genomic data to: (1) quantify population structure and gene flow among *T. affinis* populations and relate these to potential geographic barriers to gene flow, (2) fit demographic models to genomic variation to test competing diversification hypotheses, (3) quantify signatures of local adaptation in the form of associations between allele-frequency and remotely sensed environmental variation, and (4) integrate results to understand the likely mechanism that promoted speciation in *T. affinis*.

2 | MATERIALS AND METHODS

2.1 | Specimen collection

Morphological measurements and tissue samples were collected from *T. affinis* captured at 12 sites in Cameroon and two sites in Gabon (Figure 1; sampling site geographic coordinates available as part of NCBI SRA project PRJNA763817), following protocols described in Freedman, Thomassen, et al. (2010). Within the geographic region where sampling took place—Cameroon and Gabon—sites were selected to maximize geographic coverage, capturing a range of interpopulation geographic distances, and substantial environmental variation. Despite intensive sampling effort totaling approximately 14 months over 4 years, we were unable to obtain samples from ecotone sites in Cameroon geographically intermediate between our forest and ecotone sites there. This is due to both very low population density of *T. affinis* in ecotone and logistical challenges working for extended periods in remote areas. Tissue samples and field-prepared whole-body skink specimens are housed at the Centre for Tropical Research at the University of California,

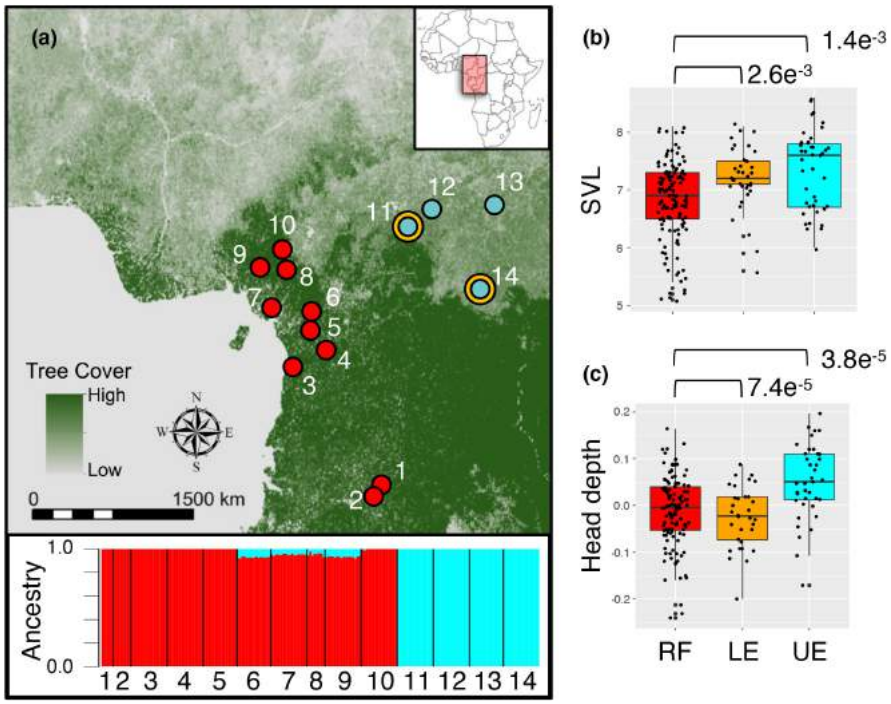

FIGURE 1 Sampling localities in Cameroon and Gabon, population structure in *Trachylepis affinis* and morphological divergence across habitats. (a) Sampling localities, with colours representing forest (red) and ecotone (cyan) populations. Background is tree cover derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Hansen et al., 2002), with the discontinuity in tree cover denoting the approximate boundary between rainforest and ecotone habitats; insert is ADMIXTURE plot for K = 2 indicating strong divergence, with minimal admixture, between forest and ecotone habitats. (b, c) For rainforest (RF), lower ecotone (LE) and upper ecotone (UE), size-corrected boxplots for snount-vent length (SVL) and head depth, respectively, with *p*-values for significant Wilcoxon rank sum tests indicated by bars. Physiological divergence between rainforest and ecotone *T. affinis* populations has been previously reported (Yuan et al., 2018).

Los Angeles and the Museum of Comparative Zoology at Harvard University.

2.2 | DNA sequencing and SNP discovery

Single digest RAD-seq libraries were prepared following the protocol described in Zhen et al. (2017), with minor modifications. A combination of single and paired-end libraries were sequenced on an Illumina HiSeq 2000 (Table S1). Following sequencing, we use FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to examine overall data quality. We processed RADseq reads using the STACKS pipeline version 1.32 (Catchen et al., 2011, 2013). First, we demultiplexed the data, removed low-quality reads and adapter sequences, and trimmed the reads to 94 bp (without barcode) altogether using process_radtags. Then, for paired-end reads, we removed PCR duplicates using clone_filter. We then combined data from multiple runs together, and de novo assembled RAD loci; for the STACKS denovo_map.pl script, parameter settings were: m = 3 M = 4 *n* = 4. The parameters for de novo assembly were determined empirically as described in Zhen et al. (2017) using methods adapted from previously published methods (Harvey et al., 2015; Ilut et al., 2014), such that values of m and M were selected to minimize the impact of duplicating loci. For genotype calling, STACKS implements a multinomial-based likelihood model by estimating the likelihood of two most frequently observed genotypes at each site and performing a standard likelihood ratio test using a χ^2 distribution (Catchen et al., 2011; Hohenlohe et al., 2010). We used the default χ^2 significance level of $\alpha = 0.05$. Paralogous loci that were stacked together were identified and removed by later quality control steps built into STACKS (e.g., max number of stacks per loci = 3) (Harvey et al., 2015; Ilut et al., 2014).

We then identified a set of high-quality RAD loci for downstream population genetic analysis using the following steps: We filtered out RAD loci that had more than 40 SNPs along the 94 bp RAD loci sequence, as these probably represented sequencing errors or over-clustering of paralogous loci. Because visual inspection of the RAD loci that have higher number of SNPs did not reveal any obvious SNP-causing misalignments, we did not apply any additional filters to avoid introducing additional biases. We used BLAT (Kent, 2002) to compare RAD loci sequences against each other, and removed ones that had a match. This step further removes duplicate RAD loci. Following these filters, we obtained our 59,699 final consensus set of RAD loci. We exported genotypes for these final consensus RAD loci in VCF format using stacks populations program (only biallelic SNPs). We additionally removed SNPs from the last 7 bp of each RAD locus as this part of the locus was enriched for erroneous SNPs due to the lower sequencing quality at the 3′ end of reads. This led to a set of RAD loci with median and maximum number of SNPs per locus of 3, and

20, respectively. Of 205,935 remaining SNPs, we randomly selected one per RAD locus, leaving a final data set of 56,055 SNPs. After performing exploratory analyses without further filtering on this data set, we discovered an inverse correlation between median depth of coverage in an individual, and the number of singletons in that individual, strongly suggestive that rare variants were enriched for erroneous genotype calls. Therefore, we filtered out sites with a minor allele frequency (MAF) < 0.05 and individuals with missing data at >25% of sites. This produced a final data set of 14,896 SNPs across 194 individuals, with populations.

2.3 | Population structure

To visualize genetic distances between sampled populations we first converted the SNP data to PLINK format, and then used PLINK version 1.90b3s (Purcell et al., 2007) to construct an identity-by-state matrix (IBS). We called the function nj in the R package ape to construct a neighbour-joining tree on 1-IBS, and visualized this phylogeny with the APE plotphylo function. We then performed principal components analysis with the R package SNPRelate (Zheng et al., 2012) version 3.1.5.

In a separate analysis, we inferred ancestry fractions and population structure for 2–10 estimated ancestral populations (K) using the maximum likelihood, model-based method implemented in the program ADMIXTURE (Alexander et al., 2009) version 1.3.0. For a set of genetic samples, ADMIXTURE uses maximum likelihood to estimate the proportions of ancestry contributed by a user-specified number of ancestral populations (K). We defined K that best fit the data as that which minimized the 10-fold cross-validation error rate. Finally, we calculated Weir and Cockerham's F_{ST} (Weir & Cockerham, 1984) between all pairs of populations with VCFTOOLS (Danecek et al., 2011) version 0.1.14.

2.4 | Migration patterns

To visualize patterns of migration in geographic space, we estimated migration surfaces with the software environmental evaluation modelling system (EEMS) (Petkova et al., 2016). EEMS takes SNP data and sample locality information to estimate dissimilarity between localities. Those dissimilarities are then interpolated across a defined study area in order to highlight regions where dissimilarity increases at a rate greater than expected under isolation-by-distance, that is, where gene flow is reduced relative to the expectation under isolation-by-distance; EEMS also constructs similar diversity surfaces that highlight regions where diversity is lower (or greater) than expected. We performed five separate EEMS runs with different starting seeds with which to initialize the MCMC, confirmed that chains were properly mixed (Figure S1), and combined runs when reporting the estimated migration (and diversity) surfaces as posterior mean migration rates (and diversity) as output by EEMS.

2.5 | Demographic inference

We used the diffusion approximation-based method implemented in the program *δaδi* (Gutenkunst et al., 2009) to evaluate the relative fit of site-frequency spectra derived from our polymorphism data to a large suite of demographic models. *δaδi* estimates joint site-frequency spectra (JSFS) from the observed data for up to three populations, and compares the estimated JSFS with that obtained through simulation under a specified demographic model. *δaδi* takes input parameters which describe the coarseness of the grid upon which to perform the diffusion approximation. Coarser grids provide a reasonable approximation of relative model fit, while finer grids, which are more computationally intensive, yield more accurate parameter estimates.

For model fitting and parameter estimation, we adapted the workflow implemented by Portik et al. (2017) and subsequent related research (Barratt et al., 2018; Charles et al., 2018), which can be found at https://github.com/dportik/dadi_pipeline. For the first round of optimization, we used threefold perturbed random starting parameters, performing 50 optimization replicates with 20 iterations per mode. We then fed the parameter estimates from the best fitting model (based upon Akaike's Information Criterion; AIC) to the second round of optimization, using twofold perturbed random starting parameters of 50 optimization replicates with 30 iterations per model. Finally, we fed the best fitting model parameters from that round to a final round of optimization using one-fold perturbed random starting parameters for 100 optimization replicates with 50 iterations per model.

Based upon the topology of the IBS neighbour-joining tree and ADMIXTURE results, for exploratory analyses we used the unfiltered data to fit 34 different models: 15 different two-population models employed by Portik et al. (2017) with forest and ecotone individuals comprising the two populations, and 19 three-population models with an initial split between forest and ecotone, and a subsequent split between southern forests (SF) and Southwest Province (SWP) (Appendix S1; Tables S3 and S4; Figures S2 and S3). AIC indicated that the 2D-JSFS models unequivocally fit the data better than the 3D-JSFS models (Table S4), so we did not consider three-population models further.

In order to obtain more accurate parameter estimates, we then repeated the model fitting procedure for the 2D-JSFS models with a finer grid, restricting our analyses to the 15 two-population models, and an additional seven two-population, 3-epoch models. The latter models were used to evaluate whether the two-epoch models might fail to detect temporal dynamics more indicative of refugial expansions and contractions. Specifically, inclusion of these seven additional models was motivated by the fact that, given the expected distribution of genealogies during severe bottlenecks, it may be difficult to distinguish our best-fit 2-epoch from 3-epoch models with histories that alternate between isolation and gene flow, and which might be equally consistent with refugial isolation or ongoing gene flow, depending upon the direction and timing of population size changes and periods of genetic exchange. All 22 models were fit to

the polymorphism data filtered on MAF and individual-level missingness. Because the data are filtered on MAF, there are, by definition, portions of the 2D-JSFS where genotypes are unobservable in the empirical data, that is, sites with low MAF in both forest and ecotone. In initial model-fitting of the filtered data, we determined that nearly all the unobservable fraction of the 2D-JSFS encompassed a 10 row by 10 column section (where row and column units are numbers of alleles), and so masked the 2D-JSFS in order to exclude this region from calculations of model fit.

In addition to evaluating AIC, we evaluated model fit with respect to the distribution of Anscombe residuals that quantify differences between the JSFS predicted by the model and that of the empirical data; Anscombe residuals were computed internally within *δaδi*. We visualize residuals in two ways using standard plotting functions available in *δaδi*: a one-dimension histogram and a two-dimensional representation that depicts residuals across the JSFS. A model that fits the data well should have (1) a 1D histogram that is centred on zero with a narrow distribution (residuals are relatively small), and (2) 2D residuals for which the spatial pattern of is essentially random such that large residuals do not cluster in sections of the JSFS. We further tested the goodness-of-fit of the top four models to the data using the *Simulate_and_Optimize.py* script updated pipeline that wraps goodness-of-fit functions in *δaδi*. Specifically, we compared the empirical χ^2 statistic to the distribution of χ^2 statistics obtained through model simulations, as well as the empirical and simulated log-likelihoods.

Transforming demographic parameters inferred with *δaδi* is challenging for three reasons: substantial uncertainty regarding mutation rate in reptiles, known genealogical biases in RAD-seq data (Arnold et al., 2013), and uncertainty regarding how to define the total length of sequences assayed. The total sequence length is required to extract the effective population size of the ancestral population, because in *δaδi*, $\theta = 4N_{ref}\mu$, but the mutation rate, μ, is not per site per generation, but per (total) sequence per generation. An estimate of N_{ref} is required for subsequent parameter conversions to biologically interpretable units. To do so, we assume a squamate mutation rate of 7.89×10^{-9} after Green et al. (2014). Based upon comments posted on the *δaδi* google site, and following work by others (Charles et al., 2018; Portik et al., 2017), we calculated total sequence length by multiplying the RAD locus length (94 bp) by the total number of loci output by STACKS, which includes invariant loci. Finally, we assume a generation time of 3 years (McCoy et al., 2010). While a cautionary note should be made concerning uncertainty in dating demographic events—particularly because of uncertainty surrounding the mutation rate used to transform demographic parameters into year scale—the relative length of intervals should be robust, indicating a long period of divergence-with-gene-flow followed by a relatively recent speciation event. Confidence intervals for both the untransformed parameter estimates of the best model were obtained by first generating 200 bootstrap samples without replacement of the empirical 2D-SFS using a custom python script, BoostrapDadiInfileWithReplacement.py. We then wrapped the dadi.Godambe.GIM_uncert function with GIMuncertaintyFullMatrix.py to produce a Godambe Information Matrix (GIM). The GIM was then converted to a variance–covariance matrix for the estimated demographic parameters (including θ), using the python numpy module via numpy.linalg.inv(GIM). For the untransformed parameters, confidence intervals were calculated as $\pm1.96\times$SE. Transformed parameters are extracted from composite parameters, for example, $N_{1a} = nu1a\times N_{ref}$, and thus variance estimates must incorporate covariance with θ. To do this, we follow guidance from developers of *δaδi* (https://dadi.readthedocs.io/en/latest/user-guide/uncertainty-analysis/), deriving variance estimators as described in Ku (1966).

2.6 | Demographic simulations and genome-wide F_{ST}

Because allele frequencies and summary statistics such as F_{ST} can be strongly influenced by population expansions (Excoffier & Ray, 2008; Lotterhos & Whitlock, 2014) such as those expected during glacial cycles (Anthony et al., 2007; Hewitt, 1996), we used simulations to evaluate the extent to which expansions estimated with *δaδi* contribute to overall patterns of genome-wide divergence between rainforest and ecotone. To better understand how the demographic history inferred by our best-fit model leads to the frequency distribution of F_{ST} for individuals SNPs in our empirical data, we first used VCFTOOLS with observed genotypes to produce Weir and Cockerham's F_{ST} between forest and ecotone. Next, to make sure that our simulation method did not produce any biases that would impact comparisons, we used *δaδi* to generate the empirical 2D-JSFS, then sampled allele frequencies from it, and from those frequencies generated a new "simulated from empirical data" vcf. Next, we applied VCFTOOLS to genotypes simulated under our best-fit model to produce a comparable set of F_{ST} values. Because the distributions from these three sets of F_{ST} –empirical data, empirical obtained from sampling the 2D-JSFS, and best-fit model—were similar, we then permuted parameters under the best-fit model to produce three additional simulated F_{ST} data sets. First, we entirely removed the initial demographic expansion occurring during divergence (T1a) by setting the effective population size scalars to be the same as the ancestral population, then simulated the F_{ST} distribution under this model; population size parameters in *δaδI* are scalars expressing the ratio of the population in question to the ancestral effective population size (N_{ref}). Second, we generated F_{ST} data by removing the population expansion at T1b relative to T1a, by setting the scalars for T1b to be equivalent to those for T1a. Third, we removed the bottleneck during T2 by setting the scalars to be equivalent to those for T1b.

2.7 | Genomic variation - environment relationships

In order to determine how genomic variation is distributed across a landscape, and which environmental variables were most closely associated with these changes, we ran gradient forest (Ellis

et al., 2012) models on the filtered SNPs. A gradient forest is a nonlinear, nonparametric machine learning technique that attempts to split a response variable (in our case, the allele frequencies of the suite of alleles mentioned above) into bifurcations in which variation in resulting bins is minimized. These bifurcations are repeated for all predictor variables across a subset of all records (i.e., our 14 *T. affinis* populations) in an iterative, random manner. Our environmental predictors consist of 16 environmental variables (Table S5; Fick & Hijmans, 2017; Hansen et al., 2002; Long et al., 2001; Zhang et al., 2005) that capture the environmental characteristics of habitats across the range of *T. affinis* in Central Africa, including aspects of temperature, precipitation, seasonality and vegetation. Each run represents a regression *tree*, and when they are run together, comprise a random forest. When multiple responses are tested simultaneously, a gradient forest is produced. In order to determine whether these gradient forests were providing true signal of the relationship between genomic and environmental variation, we also performed replicates of these models ($n = 100$ with $n = 200$ random trees in each), instead randomizing the association between allele frequencies and environmental characteristics. Should these models provide equivalent performance and/or explanatory power (based on number of SNPs with $R^2 > 0$ or with overall average R^2 of the full model), results from the actual model could be considered spurious and not valuable for assessing the relationship between genes and environment.

The result of these gradient forests models is (1) a ranking of environmental variables in terms of their importance in explaining genomic variation (as measured by the decrease in model performance when said variable is removed), and (2) a model object that describes the relationship between genomic variation and environmental variables at each location where data was collected. Using this model, we were then able to predict, based on the environmental variation found at 10,000 random pixels distributed across the range of *T. affinis*, the genomic variation that is expected at these sites, even though they were not sampled. The result of these predictions is a spatially explicit map of genomic variation, where changes in genomic variation are represented by changes in colour on a standard RGB scale, across the range of *T. affinis* in Central Africa.

3 | RESULTS

3.1 | Genomic divergence and restricted rainforest-ecotone gene flow

We use restriction site-associated DNA sequencing (RAD-Seq) to generate 56,055 single nucleotide variants (SNPs) for 208 individuals spanning 14 populations in Cameroon and Gabon (Figure 1a, Tables S1 and S2). Filtering out sites with MAF <0.05 and inidividuals with missing data at >25% of called sites led to a final data set of 14,896 SNPs in 194 individuals. Post-filtering number of individuals, number of SNPs, and nucleotide diversity per population are summarized in Table S2. Overall, the major bipartition of genomic variation is between forest and ecotone, with a second division between forest populations Cameroon's Southwest Province (SWP) and the other forest populations (Figures 1a and 2a, Figures S2–S4). Ancestry inference with ADMIXTURE reveals that, at K = 2, four forest populations north of the Sanaga River contained a minimal fraction of ecotone ancestry (Figure 1a). At higher levels of K assayed, there was no evidence of admixture between forest and ecotone populations. At K = 3, following forest versus ecotone, the next major partition of genomic variation was between forest populations from Southwest Province (SWP), Cameroon and the remaining southern forest populations (SF) (Figure S2b–f), a pattern supported by other methods for inferring population structure (Figure 2, Figures S2–S4). At K < 5 ecotone populations were recovered as a single cluster, but, at higher K, ecotone was partitioned into western and eastern subclusters (Figure S2d–f). Cross-validation error in ADMIXTURE analyses decreases with increasing K, but remains relatively stable from K = 4 onward (Figure S2g).

Admixture was detected across the Sanaga River, the major river in Cameroon that bisects the distribution of study site, thus river barriers do not appear to have played an important role in either rainforest-ecotone or intra-rainforest genetic structure in *T. affinis*. River barriers may have contributed to intra-ecotone population structure, as ADMIXTURE analyses at higher levels of K differentiated eastern and western sampling sites, which fall on either side of the Djerem River. Whether river barriers or local adaptation along an east–west climatic gradient drive this structure requires further study.

The hierarchical aspect of population structure is clearly illustrated by principal component analysis (PCA) (Figure 2a), with PC1 separating out ecotone from forest populations, and PC2 separating SF populations in Cameroon and Gabon, associated with the Congolian forest block, from SWP populations that probably have an origin in the Guinean forests. PC2 also captures finer latitudinal structure within SF, with larger values of PC2 found for populations closer to the equator. The magnitude of genetic differentiation between forest and ecotone populations cannot simply be attributed to geographic distance, as the slope of Slatkin's linearized F_{ST} over geographic distance is greater for forest versus ecotone comparisons than either forest versus forest or ecotone versus ecotone (Figure 2b). This is a classic signal of IBA, indicating an acceleration of genetic divergence due to local adaptation.

Inferring migration surfaces with EEMS, we find substantially less gene flow along the boundary between rainforest and ecotone than expected under an isolation-by-distance model (Figure 3). This pattern of reduced gene flow and population structure between rainforest and ecotone, is suggestive of recent or incipient speciation.

3.2 | Demographic inference of ecotone speciation

To determine whether gene flow between rainforest and ecotone skinks has ceased, we used the forward-time inference methods implemented in δaδi (Gutenkunst et al., 2009) to fit 41 different

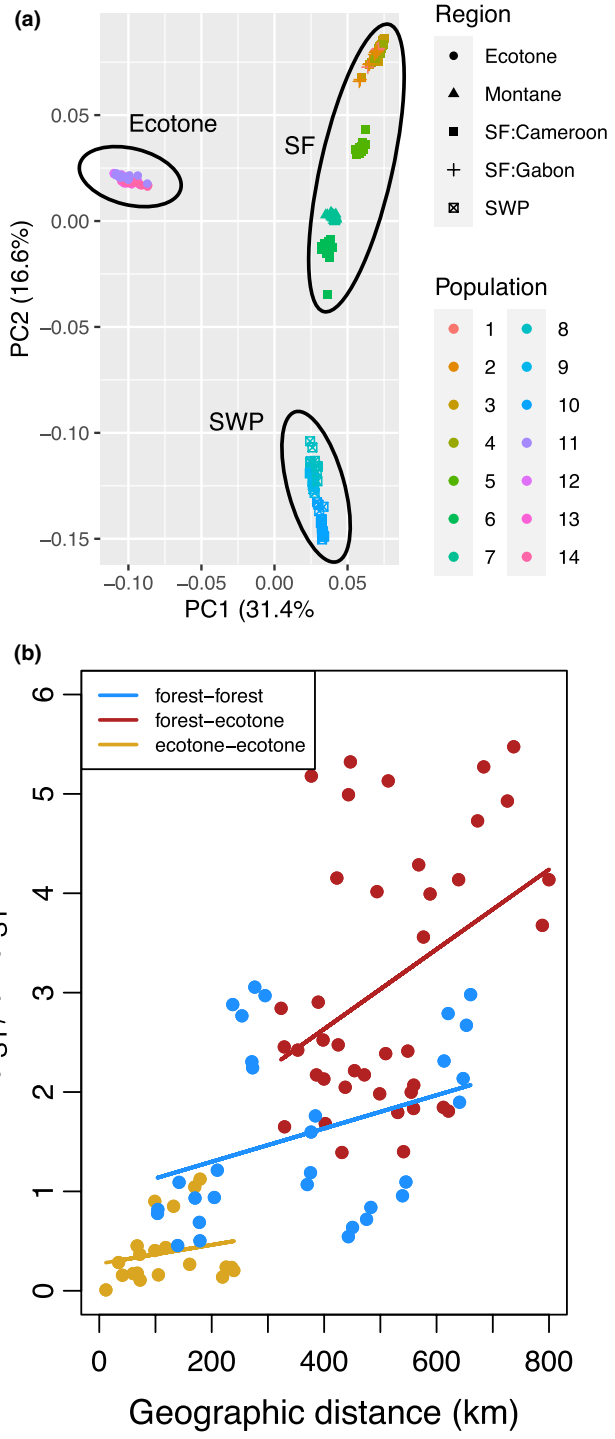

FIGURE 2 Population structure in *Trachylepis affinis*. (a) Principal component analysis (PCA) of genotypes for 14 populations of *Trachylepis affinis* in Cameroon and Gabon. PC1 uncovers genomic divergence between forest and ecotone populations, while PC2 primarily separates southern forest (SF) populations from Southwest Province (SWP) forest populations. Second, PC2 separates out SF populations latitudinally, with populations at lower latitudes characterized by larger values and vice versa. (b) Biplot of linearized F_{ST} over geographic distance, with colours and trendlines distinguishing within habitat versus between habitat comparisons. Trend lines fitted with the R lm function indicate greater genetic turnover with distance between habitats than within them, indicative of isolation-by-adaptation.

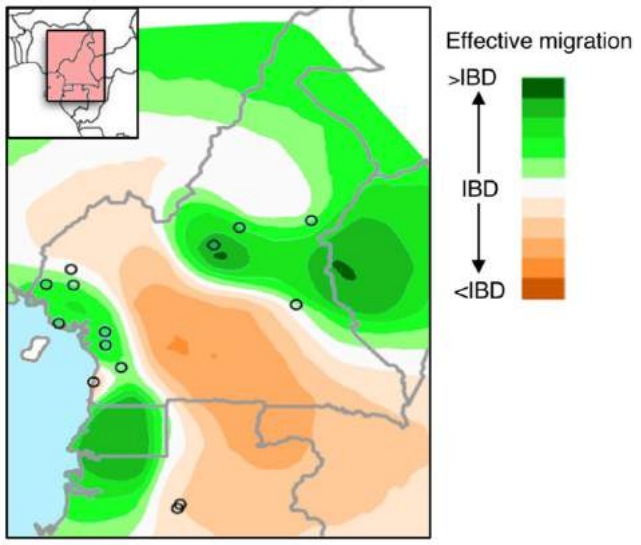

FIGURE 3 Effective migration surface estimated with environmental evaluation modelling system (EEMS) (Petkova et al., 2016), with brown and green colours indicating regions with less and more gene flow, respectively than expected under isolation-by-distance. EEMS infers reduced gene flow intervening between rainforest (left) and ecotone (right) populations. Insert indicates the location of the study area relative to the surrounding region.

demographic models to our polymorphism data (Figure S5). These comprised two and three-population models capturing the significant patterns of differentiation between forest and ecotone and between SF and SWP forest blocks, allowing us to test alternative diversification hypotheses for African rainforests formally.

We find strong support for a two-population forest versus ecotone model in which there are population expansions at the time of divergence between forest and ecotone populations, concurrent with asymmetric gene flow (T1a) for ~32 ky, followed by ~52 ky of more dramatic population expansions in both habitats, during which gene flow ceases (T1b), and a period of secondary contact (T2) of ~80 years that is accompanied by dramatic population bottlenecks (T1b) (Figure 4a, Tables S6 and S7); confidence intervals for all demographic parameter estimates are relatively narrow. Relative support for different demographic models is based upon model fit estimated with Akaike information criterion (AIC). The Akaike weight, which reflects the relative support among a set of models, indicates there is no support for alternative models relative to our top model. Fit of a particular model can be evaluated based upon correspondence between model and data in terms of the distribution of residuals (Figure S6) and goodness-of-fit tests (Figure S7). Residuals are narrowly distributed around zero and show weak spatial structure across the 2D-SFS indicative of good model fit (Figure S6). Similarly, χ^2 statistic evaluating the fit of the model to the empirical data falls within the distribution of that for data simulated under the model (Figure S7a). The log-likelihood of the data under the model falls just outside and to the left of the distribution for values simulated under the model (Figure S7b), an

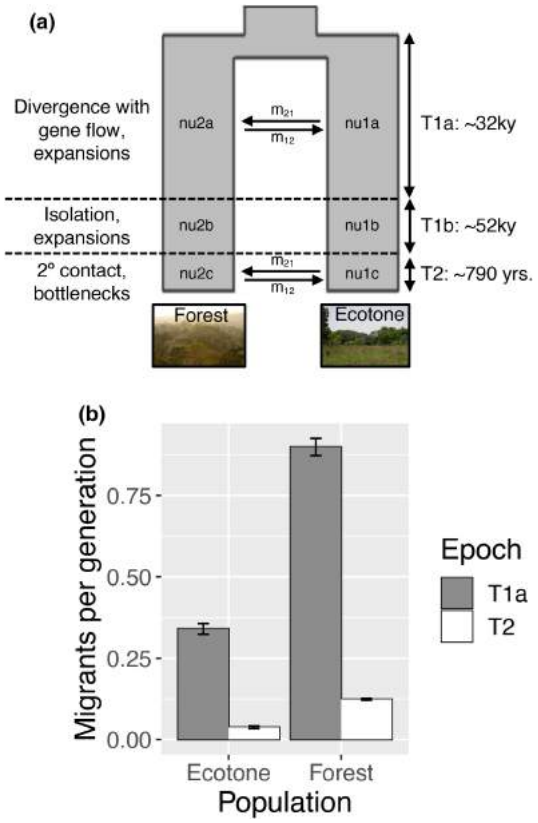

FIGURE 4 Demographic evidence for ecotone speciation. (a) Schematic of the best-fitting demographic model inferred with $\delta a \delta i$, consisting of three epochs, with population expansions in both forest and ecotone at divergence, during which there is asymmetric gene flow (during T1a), followed by bottlenecks and cessation of gene flow (T1b), followed by the most recent epoch leading up to the present (T2) in which there is the resumption of gene flow (i.e., secondary contact); levels of gene flow are extremely low. (b) Means ± 1.96 × SE for number of migrants per generation, indicating very low gene flow during T1a, and dramatically lower gene flow during the secondary contact phase (T2).

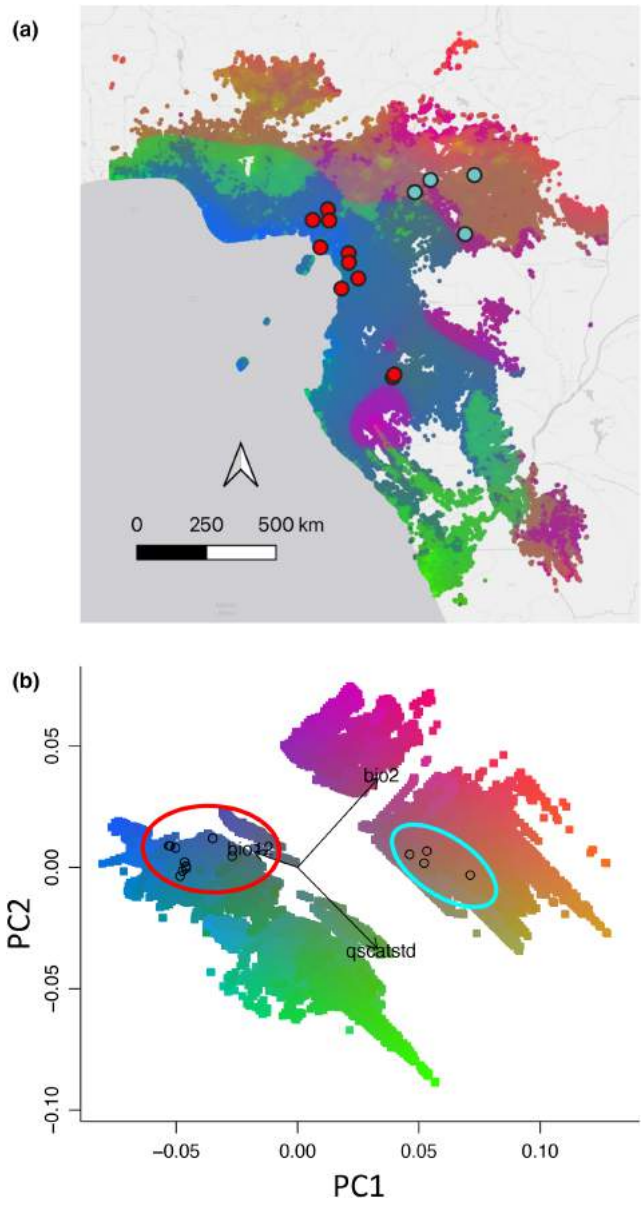

FIGURE 5 Environmental drivers of allele frequency divergence along the rainforest-savanna gradient. (a) Spatial projection of the gradient forest model fitting allele frequencies to environmental variables. (b) Principal component analysis (PCA) representation of the gradient forest model, with vectors of driving environmental variables superimposed in PCA space.

indication of a poorer fit than revealed by other measures. Overall, all but one of our performance metrics indicate there is good fit between the data and the inferred model.

Rates of gene flow between forest and ecotone inferred from the best model are very low. Proportions of populations comprising immigrants, while asymmetric, did not vary between T1a and T2, and are approximately 2.5 and 4.4 immigrants per 10,000 individuals, for ecotone and forest, respectively (Table S7). When we transform migration rates for T1a and T2 into numbers of migrants-per-generation (transformed migration rate × population-specific N_e), the number of migrants-per-generation during the divergence-with-gene flow phase are (mean ± 1.96 * SE) 0.340 ± 0.033 for ecotone, and 0.899 ± 0.052 for forest (Figure 4b). These rates decrease dramatically during the period of secondary contact: 95% CIs: 0.037 ± 0.007 and 0.124 ± 0.006 for forest and ecotone, respectively (Figure 3b). To better understand whether estimated rates of gene flow between forest and ecotone would leave a detectable signature of introgression, we conducted ancestry inference with ADMIXTURE at k = 2 for (1) genotypes sampled from the empirical SFS; (2) genotypes simulated under the best-fit model; (3) genotypes simulated under a modified version of the best-fit model, in which the period of isolation (T1b) is made extremely short, allocating all the time from T1b to T1a, that is, a model with an extended period of gene flow, and (4) the model described in 3, but increasing gene flow 10-fold. In all scenarios, almost no ecotone ancestry is assigned to forest individuals and no forest ancestry is inferred for forest individuals (Figure S8). This lack of an introgression signature is consistent with a scenario of sporadic, infrequent gene flow between distinct species over 10s of 1000s of years.

3.3 | Environment drives divergence across the rainforest-ecotone gradient

Fitting gradient forest models to allele-frequency environment associations, we find a clear environmental signature of genomic divergence between rainforest and ecotone populations, both in geographic space (Figure 5a) and principal component axes (Figure 5b). Driving this separation are differences in the diurnal temperature range (Bio2), temporal variation in vegetation roughness (standard deviation of Quickscat radar measurements, a proxy for seasonality), and mean annual precipitation (Figure S9a). In contrast, null models based upon randomized associations between genomic and environmental variation from the original data set exhibit weaker predictive performance (Figure S9b), suggesting that our model captures a genuine biological relationship between genomes and the environment.

If patterns of genome-wide F_{ST} cannot be attributed to population expansions—particularly the moderate spike in frequency of SNPs with high forest-ecotone F_{ST} –then it is more likely that diversifying selection plays an important role in generating associations between allele-frequency and environment. Specifically, the simulated genome-wide F_{ST} distribution obtained when removing population expansions during divergence-with-gene-flow in time T1a from our best model produce little if any deviation in the bimodal pattern of F_{ST} observed in our empirical data and removing the expansions during T1b dramatically increases the frequency of high F_{ST} SNPs (Figure S10). In contrast, removing the divergence bottlenecks in T2 eliminates the second peak of elevated F_{ST} (Figure S10). While demographic processes can contribute to allele frequency - environment correlations, the absence of a confounding effect of population expansions suggests that a non-trivial fraction of environment-associated allele frequency turnover along the rainforest-savanna gradient is due to local adaptation. Given the extremely low probability of detecting a directly selected site with RAD-seq, broad patterns of linkage between sequenced loci and genomic sites under divergent selection along the forest-ecotone gradient, not demography, shape the observed allele-frequency environment correlations.

4 | DISCUSSION

Parallel patterns among genomic, physiological, and morphological divergence in *T. affinis* across the rainforest-savanna gradient provides compelling evidence for recognizing forest and ecotone lineages *of T. affinis* as recently diverged species. Our findings support previous research on the possible role of ecotones in speciation in African rainforest birds (Smith et al., 1997; Smith et al., 2005; Smith et al., 2011; Zhen et al., 2017). More broadly, our results add further support to the role of ecological gradients in generating biodiversity (Schluter & Pennell, 2017). Below, we discuss population structure in *T. affinis*, evidence for ecotone speciation, and the conservation implications of rainforest-ecotone diversification.

4.1 | Population structure

Analyses of population structure, ancestry, and migration patterns consistently identify the rainforest-ecotone gradient as the major axis of genomic differentiation in *T. affinis*. This is consistent with the pattern previously recovered for this species with AFLPs (Freedman, Thomassen, et al., 2010). Forest-ecotone divergence does not coincide with any known geographic barrier, suggesting a role for divergence via habitat adaptation. We also recover weaker population structure within the rainforest zone between hypothesized rainforest refugia, with ancestry inference revealing substantial hybridization between them.

Rainforest-ecotone divergence in *T. affinis* is consistent with evidence found in other species, supporting the hypothesis that the rainforest-savanna gradient is an engine of diversification. Genetic, morphological, and song divergence across the gradient were observed in the little greenbul (*Andropadus virens*; Kirschel et al., 2011; Slabbekoorn & Smith, 2002; Smith et al., 1997, 2005; Zhen et al., 2017). The olive sunbird (*Cyanomitra olivacea*) exhibits morphological divergence along the gradient despite little population structure due to the homogenizing effects of gene flow (Smith et al., 2011). As rainforest and ecotone have distinct thermal regimes, we predict genetic divergence along the gradient will be prevalent in ectotherms such as reptiles and insects, resulting from local physiological adaptation reducing gene flow. Rainforest-ecotone physiological divergence has already been recorded for *T. affinis* (Yuan et al., 2018) and *Bicyclus dorothea*, a Central African butterfly (Dongmo et al., 2021).

4.2 | Diversification hypotheses: predictions versus data

The demographic history we inferred for *T. affinis* involves sequential periods of divergence-with-gene-flow, no gene flow, and a resumption of gene flow, spanning an estimated time of approximately 85 ky. A greater understanding of the evolutionary process underlying this model requires identifying the correct drivers of genomic divergence and evaluating whether the magnitudes of divergence and gene flow are indicative of speciation or merely ongoing intraspecific divergence.

Regarding the first of these considerations, rainforest-ecotone divergence is inconsistent with a model of refugial isolation for several reasons. First, rainforest-ecotone divergence initially takes place in the face of gene flow, while the refugial model predicts that divergence occurs in allopatry. Second, population expansions inferred under our model are temporally inconsistent with predictions of a refugial model. Glacial cycles occur approximately every 100 ky, with the LGM 12.5 kya. Thus, at 100 kya, the African rainforest would have been near its maximal extent, contracting into refugia with minimal extent at the LGM (Maley, 1996; Maley & Brenac, 1998). The contraction of species ranges into refugia predicts a genetic signal of population bottlenecks. Our inferred model

indicates the opposite for *T. affinis*, with both rainforest and ecotone populations undergoing expansion over much of the last glacial cycle. Furthermore, the most dramatic expansions take place during a period in which gene flow ceases. In contrast, if cycles of refugial isolation drive divergence, any gene flow should occur upon secondary contact of lineages previously isolated in refugia. Finally, a vanishing refuge scenario (Vanzolini & Williams, 1981) in which ecotone populations get stranded in a refugial ecotone isolate and undergo local adaptation without the evolutionary constraint of gene flow from the forest is highly unlikely. During the cool, dry conditions occurring in glacial periods, as rainforest habitats would have contracted towards the equator, the ecotone would have followed, occupying regions previously covered by rainforest and remaining adjacent to the shrinking rainforest zone. Ecological niche modelling (ENM) of tree cover at the LGM replicates this phenomenon, and the predicted distribution of *T. affinis* also contracts, tracking their suitable habitat along shifting rainforest-ecotone gradient (Freedman, Thomassen, et al., 2010).

Rather than a model of divergence between stable refugia, we suggest persistent parapatry between rainforest and ecotone-adapted populations even as the location and extent of the zone of parapatry shifted through time. Consistent with ENMs for both rainforest cover and the distribution of *T. affinis*, this would allow divergence-with-gene-flow along the rainforest-savanna gradient to steadily augment genomic divergence. Under this scenario, the cessation of gene flow during the second model epoch could result from evolving reproductive isolation, a reduction in habitat connectivity around the LGM impeding already weak gene flow, or synergy between these two processes.

4.3 | IBA and sampling bias

One concern regarding the interpretation of forest-ecotone divergence and demographic models indicating minimal gene flow is that ecotone populations intermediate between our forest and ecotone sampling sites in Cameroon were not sampled. Should those unsampled populations fall within a broad hybrid zone, we may have underestimated the magnitude of gene flow along the gradient and, conversely, overestimated the extent to which forest and ecotone gene pools have become isolated from one another. The strong signal of IBA (Figure 2b) lessens this concern for two reasons. First, the higher rate of genetic turnover between habitats compared to turnover within them implies that reproductive isolation has evolved in response to local adaptation. Second, the relationships between linearized F_{ST} and geographic distance predict that, for any particular geographic distance, forest-ecotone divergence will remain higher than that within either forest or ecotone. Thus, forest-ecotone population pairs separated by geographic distances equivalent to those between unsampled ecotone populations and our sampled forest sites are still predicted to experience less gene flow than what would be expected due to (1) clinal variation simply resulting from limited dispersal or (2) across a broad hybrid zone with little selection against hybrids. Even if gene flow was underestimated due to sampling bias, simulations with 10× more gene flow across both the divergence-with-gene-flow and isolation periods of our original model predict minimal amounts of ecotone ancestry in rainforest and no rainforest ancestry in ecotone. This lack of admixture suggests that our inferences are robust and not biased. In summary, the pattern of IBA, and the meagre rates of gene flow inferred suggest that, if a hybrid zone exists, it is quite narrow with strong selection against hybrids. If this were to be the case it would be more typical of a hybrid zone between two well-defined species.

4.4 | Evidence for ecotone speciation

Levels of gene flow estimated by our demographic model are so low that simulations under the model fail to recover individuals of admixed ancestry. Gene flow between rainforest and ecotone *T. affinis* is lower than commonly observed in hybridizing taxa (Harrison & Larson, 2014; Pavón-Vásquez et al., 2021; Taylor & Larson, 2019; Yang et al., 2021). Even when we modified our model so that gene flow was 10-fold higher than originally inferred, and occurred across the entire history of divergence, analysis simulated genotypes indicated little admixture and only in the forest. Our results and the growing body of studies documenting interspecific hybridization are consistent with a "genic view of speciation" (Barton & Hewitt, 1985; Barton & Hewitt, 1989; Wu, 2001). Support for this view has grown with the availability of genome-scale data for diverse nonmodel organisms (Taylor & Larson, 2019). The assertion here is that genomes are semipermeable, with gene flow between species varying among genomic regions: less taking place in genomic regions containing genes related to local adaptation, more in regions less constrained by selective forces (Barton & Hewitt, 1985; Barton & Hewitt, 1989; Harrison & Larson, 2014; Wu, 2001). We assume a generation time in skinks of 3 years (McCoy et al., 2010), but a generation time of 2 years has also been reported (Vitt & Cooper, 1986). If the generation time of *T. affinis* were between two and 4 years, speciation would have occurred sometime between 57 ky and 113 ky ago, during the Pleistocene. It remains an open question whether climate cycles, resulting changes in habitat structure and distribution, and diversifying selection may have acted synergistically to accelerate diversification in *T. affinis* along the rainforest-savanna gradient over such a short time scale. The concordance between allele frequency with environment variables, morphological and physiological divergence along the rainforest-savanna gradient are all consistent with a role for diversifying selection in the speciation process.

4.5 | Conservation implications in a warming world

The ecotone encompassing Africa's Congo Basin's rainforest edge and the joining forest-savanna mosaic can be as much as 1000 km wide, and historically covered over 8 million square kilometres, an area immense enough to support the evolutionary dynamics and the

progression to new species. The slope of an ecological gradient is a critical determinant of how natural selection operates (Bachmann et al., 2020; Doebeli & Dieckmann, 2003) and, ultimately, how biodiversity is generated. Anthropogenic flattening of the rainforest-ecotone gradient can substantially weaken differential selection (Freedman, Buermann, et al., 2010) along the gradient. Ongoing land cover changes caused by land clearing and urbanization in tropical Africa may homogenize ecological gradients that are essential drivers of adaptive diversification. Viewed in the context of climate change, the ongoing destruction of the rainforest-ecotone gradient has far-reaching implications for biodiversity conservation efforts (Smith et al., 2021). While the potential for species to persist in a warming world will be determined by many factors (Catullo et al., 2019), maximizing adaptive genetic variation across natural populations serves a useful bet-hedging measure to increase the probability that at least some populations will survive. Our evidence for ecotone speciation strengthens the view of the rainforest-ecotone gradient as an essential engine of adaptive genetic variation that is probably crucial for species persistence. The continued erosion of rainforest-ecotone gradients may well hinder evolutionary responses to climate change (Smith et al., 2021), increase extinction rates, and accelerate the speed at which tropical Africa crosses a possible "biodiversity climate horizon" (Trisos et al., 2020), beyond which collapse of ecological and evolutionary processes become more likely.

AUTHOR CONTRIBUTIONS

Adam H. Freedman, Alison M. Hamilton, and Thomas B. Smith collected field samples. Adam H. Freedman performed all data analyses, except for gradient forest analyses performed by Ryan J. Harrigan, and genotyping by Ying Zhen. Adam H. Freedman wrote the manuscript with Thomas B. Smith, with input from all other authors.

ACKNOWLEDGEMENTS

The computations in this study were run on UCLA's Hoffman2 cluster, and the cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. We thank the University of California, Berkeley's Vincent J. Coates Genomic Sequencing Laboratory (GSL) for sequencing services. We also thank the Agence Nationale des Parcs Nationaux (ANPN, Gabon), the Centre National de la Recherche Scientifique et Technologique (CENAREST, Gabon), the Ministère des Forêts et de la Faune (MINFOF, Cameroon) and the Ministère de la Recherche Scientifique et de l'innovation (MINRESI, Cameroon) for research and sampling permits. In addition, we thank Roger Fotso and the Wildlife Conservation Society in Cameron for logistical support. We thank Chris Schneider for assisting T. B. S. collect samples in Cameroon. We thank Emmanuel Benda for indispensable assistance in the field in Cameroon as driver, village negotiator, mechanic, organizer of field teams, and coconspirator in navigating the endless array of roadway checkpoints and hinterland security concerns. We also thank Dolph Schluter for comments on an earlier version of the manuscript, along with members of Harvard's Faculty of Arts and Sciences Informatics Group for feedback on data analysis. This work was supported by U. S. Environmental Protection Agency STAR and Fulbright fellowships to A. H. F, a QCB Collaboratory Postdoctoral Fellowship to YZ, and National Geographic grants and National Science Foundation grants IIA PIRE no. 1243524, DEB-9726425 and IRCEB9977072 to T. B. S.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

All demultiplexed reads have been stored in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA763817 (Freedman, 2021a). Variant call format (vcf) genotype files derived with the STACKS pipeline, and study site geographic coordinates and environmental data are available on Dryad (Freedman, 2021b) at https://doi.org/10.5061/dryad.xwdbrv1f4. Code used in analysis and generating figures are available at https://github.com/adamfreedman/TrachylepisAffinisSpeciation.

ORCID

Adam H. Freedman https://orcid.org/0000-0003-4714-3925
Ying Zhen https://orcid.org/0000-0001-8382-2533

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allen, K. E., Greenbaum, E., Hime, P. M., Tapondjou, N. W. P., Sterkhova, V. V., Kusamba, C., Rödel, M.-O., Penner, J., Peterson, A. T., & Brown, R. M. (2021). Rivers, not refugia, drove diversification in arboreal sub-Saharan African snakes. *Ecology & Evolution*, *11*, 6133–6152.

Anhuf, D., Ledru, M.-P., Behling, H., Da Cruz, F. W., Jr., Cordeiro, R. C., Van der Hammen, T., Karmann, I., Marengo, J. A., De Oliveira, P. E., Pessenda, L., Siffedine, A., Albuquerque, A. L., & Da Silva Dias., P. L. (2006). Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *239*(2006), 510–527.

Anthony, N. M., Johnson-Bawe, M., Jeffrey, K., Clifford, S. L., Abernethy, K. A., Tutin, C. A., Lahm, S. A., White, L. J. T., Utley, J. F., Wickings, E. J., & Bruford, M. W. (2007). The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in Central Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20432–20436.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, *22*, 3179–3190.

Bachmann, J. C., Jansen van Rensburg, A., Cortazar-Chinarro, M., Laurila, A., & Van Buskirk, J. (2020). Gene flow limits adaptation along steep environmental gradients. *American Naturalist*, *195*(3), E67–E86.

Barratt, C. D., Bwong, B. A., Jehle, R., Liedtke, H. C., Nagel, P., Onstein, R. E., Portik, D. M., Streicher, J. W., & Loader, S. P. (2018). Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Molecular Ecology*, *27*(21), 4289–4308.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation, and hybrid zones. *Nature*, *341*, 497–503.

Calsbeek, R., & Irschick, D. J. (2007). The quick and the dead: Correlational selection on morphology, performance, and habitat use in Island lizards. *Evolution*, *61*(11), 2493–2503.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Catullo, R. A., Llewelyn, J., Phillips, B. L., & Moritz, C. C. (2019). The potential for rapid evolution under anthropogenic climate change. *Current Biology*, *29*(19), R996–R1007.

Charles, K. L., Bell, R. C., Blackburn, D. C., Burger, M., Fujita, M. K., Gvoždík, V., Jongsma, G. F. M., Kouete, M. T., Leaché, A. D., & Portik, D. M. (2018). Sky, sea, and forest islands: Diversification in the African leaf-folding frog *Afrixalus paradorsalis* (Anura: Hyperoliidae) of the lower Guineo-Congolian rain forest. *Journal of Biogeography*, *45*(8), 1781–1794.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Luner, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Debout, G. D. G., Doucet, J.-L., & Hardy, O. J. (2011). Population history and gene dispersal inferred from spatial genetic structure of a central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity*, *106*(1), 88–99.

Doebeli, M., & Dieckmann, U. (2003). Speciation along environmental gradients. *Nature*, *421*(6920), 259–264.

Dongmo, M. A. K., Hanna, R., Smith, T. B., Fiaboe, K. K. M., Fomena, A., & Bonebrake, T. C. (2021). Local adaptation in thermal tolerance for a tropical butterfly across ecotone and rainforest habitats. *Biology Open*, *10*(4), 58619. https://doi.org/10.1242/bio.058619

Donihue, C. M., Herrel, A., Fabre, A.-C., Kamath, A., Geneva, A. J., Schoener, T. W., Kolbe, J. J., & Losos, J. B. (2018). Hurricane-induced selection on the morphology of an Island lizard. *Nature*, *560*(7716), 88–91.

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168.

Endler, J. A. (1982). Pleistocene forest refuges: Fact or fancy? In G. T. Prance (Ed.), *Biological diversification in the tropics* (pp. 641–657). Columbia University Press.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecology & Evolution*, *23*, 347–351.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315.

Freedman, A. H. (2021a). *Population genomics of the African rainforest skink Trachylepis affinis*. NCBI SRA; [dataset]. https://www.ncbi.nlm.nih.gov/sra/PRJNA763817

Freedman, A. H. (2021b). *Ecotone speciation across an African rainforest-savanna gradient*. Dryad. https://doi.org/10.5061/dryad.xwdbrv1f4 [dataset].

Freedman, A. H., Buermann, W., Mitchard, E. T. A., DeFries, R. S., & Smith, T. B. (2010). Human impacts flatten rainforest-savanna gradient and reduce adaptive diversity in a rainforest bird. *PLoS One*, *5*, e13088. https://doi.org/10.1371/journal.pone.0013088

Freedman, A. H., Thomassen, H. A., Buermann, W., & Smith, T. B. (2010). Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, *19*(17), 3773–3788.

Fuchs, J., & Bowie, R. C. K. (2015). Concordant genetic structure in two species of woodpecker distributed across the primary west African biogeographic barriers. *Molecular Phylogenetics & Evolution*, *88*, 64–74.

Gonder, M. K., Locatelli, S., Ghobrial, L., Mitchell, M. W., Kujawski, J. T., Lankester, F. J., Stewart, C.-B., & Tishkoff, S. A. (2011). Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 4766–4771.

Green, R. E., Braun, E. L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M. W., St. John, J. A., Capella-Gutiérrez, S., Castoe, T. A., Kern, C., Fujita, M. K., Opazo, J. C., Jurka, J., Kojima, K. K., Caballero, J., Hubley, R. M., Smit, A. F., Platt, R. N., … Ray, D. A. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*, *346*(6215), 1254449. https://doi.org/10.1126/science.1254449

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. https://doi.org/10.1371/journal.pgen.1000695

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimicelli, C., & Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sensing of Environment*, *83*(1–2), 303–319.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Heredity*, *105*, 795–809.

Harvey, M. G., Judy, C. D., Seeholzer, G. F., Maley, J. M., Graves, G. R., & Brumfeld, R. T. (2015). Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. *PeerJ*, *3*, e895. https://doi.org/10.7717/peerj.895

Helmstetter, A. J., Béthune, K., Kamdem, N. G., Sonké, B., & Couvreur, T. L. P. (2020). Individualistic evolutionary responses of central African rain forest plants to Pleistocene climatic fluctuations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(51), 32509–32518.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their rold in divergence and speciation. *Biological Journal of the Linnean Society*, *58*, 247–276.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Ilut, D. C., Nydam, M. L., & Hare, M. P. (2014). Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, *2014*, 675158. https://doi.org/10.1155/2014/675158

Jaynes, K. E., Myers, E. A., Gvoždík, V., Blackburn, D. C., Portik, D. M., Greenbaum, E., Jongsma, G. F. M., Rödel, M. O., Badjedjea, G., Bamba-Kaya, A., Baptista, N. L., Akuboy, J. B., Ernst, R., Kouete, M. T., Kusamba, C., Masudi, F. M., McLaughlin, P., Nneji, L. M., Onadeko, A. B., … Bell, R. C. (2021). Giant tree frog diversification in west and Central Africa: Isolation by physical barriers, climate, and reproductive traits. *Molecular Ecology*, *31*, 3979–3998. https://doi.org/10.1111/mec.16169

Jongsma, G. F. M., Barej, M. F., Barratt, C. D., Burger, M., Conradie, W., Ernst, R., Greenbaum, E., Hirschfeld, M., Leaché, A. D., Penner, J., Portik, D. M., Zassi-Boulou, A.-G., Rödel, M. O., & Blackburn, D. C. (2018). Diversity and biogeography of frogs in the genus Amnirana (Anura: Ranidae) across sub-Saharan Africa. *Molecular Phylogenetics & Evolution*, *120*, 274–285.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664.

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*(11), 3162–3174.

Ku, H. H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of the National Board of Standards – C Engineering and Instrumentation*, *70*, 263–273.

Leaché, A. D., Portik, D. M., Rivera, D., Rödel, M. O., Penner, J., Gvoždík, V., Greenbaum, E., Jongsma, G. F. M., Ofori-Boateng, C., Burger, M., Eniang, E. A., Bell, R. C., & Fujita, M. K. (2019). Exploring rain forest diversification using demographic model testing in the African foam-nest treefrog *Chiromantis rufescens*. *Journal of Biogeography*, *46*(12), 2706–2721.

Ley, A. C., Heuertz, M., & Hardy, O. J. (2017). The evolutionary history of central African rain forest plants: Phylogeographical insights from sister species in the climber genus *Haumania* (Marantaceae). *Journal of Biogeography*, *44*(2), 308–321.

Long, D. G., Drinkwater, M. R., Holt, B., Saatchi, S., & Bertoia, C. (2001). Global ice and land climate studies using scatterometer image data. *EOS Transactions American Geophysical Union*, *82*(43), 503.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*, 2178–2192.

Maley, J. (1996). The African rainforest-main characteristics of changes in vegetation and climate from the upper cretaceous to the quaternary. *Proceedings of the Royal Society of Edinburgh*, *104B*, 31–73.

Maley, J., & Brenac, P. (1998). Vegetation dynamics, palaeoenvironments and climate changes in the forests of western CameroonCameroon during the last 28,000 B.P. *Review of Palaeobotany and Palynology*, *99*, 157–187.

McCoy, E. D., Richmond, J. Q., Mushinsky, H. R., Britt, E. J., & Godley, J. S. (2010). Long generation time delays the genetic response to habitat fragmentation in the Florida sand skink. *Journal of Herpetology*, *44*(4), 641–644.

Mizerovská, D., Nicolas, V., Demos, T. C., Akaibe, D., Colyn, M., Denys, C., Kaleme, P. K., Katuala, P., Kennis, J., Kerbis Peterhans, J. C., Laudisoit, A., Missoup, A. D., Šumbera, R., Verheyen, E., & Bryja, J. (2019). Genetic variation of the most abundant forest-dwelling rodents in Central Africa (*Praomys jacksoni* complex): Evidence for Pleistocene refugia in both montane and lowland forests. *Journal of Biogeography*, *46*(7), 1466–1478.

Nicolas, V., Missoup, A. D., Denys, C., Kerbis Peterhans, J., Katuala, P., Couloux, A., & Colyn, M. (2011). The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *Journal of Biogeography*, *38*(1), 191–207.

Ntie, S., Davis, A. R., Hils, K., Mickala, P., Thomassen, H. A., Morgan, K., Vanthomme, H., Gonder, M. K., & Anthony, N. (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

Pavón-Vásquez, C. J., Brennan, I. G., & Keogh, S. (2021). A comprehensive approach to detect hybridization sheds light on the evolution of Earth's largest lizards. *Systematic Biology*, *70*(5), 877–890.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.

Piñeiro, R., Dauby, G., Kaymak, E., & Hardy, O. J. (2017). Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the ice ages. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1866), 20171800.

Piñeiro, R., Hardy, O. J., Tovar, C., Gopalakrishnan, S., Vieira, F. G., & Gilbert, M. T. P. (2021). Contrasting genetic signal of recolonization after rainforest fragmentation in African trees with different dispersal abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(27), e2013979118. https://doi.org/10.1073/pnas.2013979118

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M. O., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo- Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263.

Primack, R. B. (2005). *Tropical rain forests: An ecological and biogeographical comparison*. Blackwell Publishers.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Quérouil, S., Verheyen, E., Dillen, M., & Colyn, M. (2003). Patterns of diversification in two African forest shrews: *Sylvisorex johnstoni* and *Sylvisorex ollula* (Soricidae, Insectivora) in relation to paleoenvironmental changes. *Molecular Phylogenetics & Evolution*, *28*(1), 24–37.

Rice, W. R., & Hostert, E. E. (1993). Laboratory experiments on speciation: What have we learned in 40 years? *Evolution*, *47*(6), 1637–1653.

Schluter, D., & Pennell, M. W. (2017). Speciation gradients and the distribution of biodiversity. *Nature*, *546*(7656), 48–55.

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*(24), 13869–13873.

Slabbekoorn, H., & Smith, T. B. (2002). Habitat-dependent song divergence in the little greenbul: An analysis of environmental selection pressures on acoustic signals. *Evolution*, *56*(9), 1849–1858.

Smith, T. B., Calsbeek, R., Wayne, R. K., Holder, K. H., Pires, D., & Bardeleben, C. (2005). Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. *Journal of Evolutionary Biology*, *18*(2), 257–268.

Smith, T. B., Fuller, T. L., Zhen, Y., Zaunbrecher, V., Thomassen, H. A., Njabo, K., Anthony, N. A., Gonder, M. K., Buermann, W., Larison, B., Ruegg, K., & Harrigan, R. J. (2021). Genomic vulnerability and socio-economic threats under climate change in an African rainforest bird. *Evolutionary Applications*, *14*(5), 1239–1247.

Smith, T. B., Thomassen, H. A., Freedman, A. H., Sehgal, R. N. M., Buermann, W., Saatchi, S., Pollinger, J., Milà, B., Pires, D., Valkiūnas, G., & Wayne, R. K. (2011). Patterns of divergence in the olive sunbird *Cyanomitra olivacea* (Aves: Nectariniidae) across the African rainforest–savanna ecotone. *Biological Journal of the Linnean Society*, *103*(4), 821–835.

Smith, T. B., Wayne, R. K., Girman, D. J., & Bruford, M. W. (1997). A role for ecotones in generating rainforest biodiversity. *Science*, *276*(5320), 1855–1857.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*, 170–177.

Trisos, C. H., Merow, C., & Pigot, A. L. (2020). The projected timing of abrupt ecological disruption from climate change. *Nature*, *580*(7804), 496–501.

Vanzolini, P. E., & Williams, E. F. (1981). The vanishing refuge: A mechanism for ecogeographic speciation. *Papéis Avulsos de Zoologica*, *34*, 251–255.

Vitt, L. J., & Cooper, W. E. (1986). Skink reproduction and sexual dimorphism: *Eumeces fasciatus* in the southeastern United States, with notes on *Eumeces inexpectatus*. *Journal of Herpetology*, *20*(1), 65–76.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*, 851–865.

Yang, W., Feiner, N., Pinho, C., While, G. M., Kaliontzopoulou, A., Harris, J. D., Salvi, D., & Uller, T. (2021). Extensive introgression and mosaic genomes of Mediterranean endemic lizards. *Nature Communications*, *12*(1), 2762. https://doi.org/10.1038/s41467-021-22949-9

Yuan, F. L., Freedman, A. H., Chirio, L., LeBreton, M., & Bonebrake, T. C. (2018). Ecophysiological variation across a forest-ecotone gradient produces divergent climate change vulnerability within species. *Ecography*, *41*(10), 1627–1637.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., & Liu, Z. (2005). Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. *Journal of Geophysical Research: Atmospheres*, *110*, D12103. https://doi.org/10.1029/2004JD005263

Zhen, Y., Harrigan, R. J., Ruegg, K. C., Anderson, E. C., Ng, T. C., Lao, S., Lohmueller, K. E., & Smith, T. B. (2017). Genomic divergence across ecological gradients in the central African rainforest songbird (*Andropadus virens*). *Molecular Ecology*, *26*(19), 4966–4977.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Lauri, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

SUPPORTING INFORMATION

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

How to cite this article: Freedman, A H., Harrigan, R J., Zhen, Y., Hamilton, A M., & Smith, T B. (2023). Evidence for ecotone speciation across an African rainforest-savanna gradient. *Molecular Ecology*, *00*, 1–14. https://doi.org/10.1111/mec.16867