Biological Journal of the Linnean Society, 2017, **122**, 801–813. With 3 figures.

Diversification in a biodiversity hotspot: genomic variation in the river frog *Amietia nutti* across the Ethiopian Highlands

JOSEPH D. MANTHEY[1], JACOBO REYES-VELASCO[1], XENIA FREILICH[2] and STÉPHANE BOISSINOT[1]*

[1]*New York University Abu Dhabi, Saadiyat Island Campus, PO Box 129188, Abu Dhabi, United Arab Emirates*
[2]*Department of Biology, Queens College, The City University of New York, 65-30 Kissena Boulevard, Flushing, New York 11367, USA*

Received 25 May 2017; revised 8 August 2017; accepted for publication 8 August 2017

Isolation is a major driver of population differentiation, due to long-term biogeographic barriers as well as habitat fragmentation across smaller geographic scales. East Africa is a highly heterogeneous and fragmented landscape and contains high numbers of endemic plants and animals, particularly in the Afromontane regions of the Eastern Arc Mountains, Albertine Rift Mountains and Ethiopian Highlands. Despite high levels of endemism, few studies have looked at intraspecific genetic variation in these regions, and fewer still within a genome-wide context. Here, we use two mitochondrial genes (*ND2* and 16S) and large single nucleotide polymorphism panels to investigate patterns and processes of diversification in Ethiopian Highland populations of the river frog *Amietia nutti*. We identify a potential out-of-Ethiopia pattern of diversification, although further investigation is required to confirm this scenario. Within Ethiopia, regional populations diverged in isolation with little or no gene flow. Biogeographic barriers—including the Great Rift, Blue Nile and Omo River Valleys—and geographic distance, not environmental differences or watershed barriers, explain a large proportion of the genomic differentiation between populations. The timing of diversification in the group suggests genomic differentiation may have been facilitated by Pleistocene shifts of suitable habitat between glacial and interglacial periods. Strong regional genetic differences across *A. nutti* populations suggest that highland-adapted species harbour unique genetic variation across several regions of Ethiopia. This highlights the necessity for sampling many populations in phylogeographic analyses across this region to fully understand patterns of genetic variation and diversification across the complex topographical landscape of the Ethiopian Highlands.

ADDITIONAL KEYWORDS: *Amietia* – Ethiopian Highlands – frogs – phylogeography – RAD-seq.

INTRODUCTION

Habitat fragmentation is a major driver of genetic differentiation across spatiotemporal scales, from biogeographic barriers at large time scales to microgeographic habitat fragmentation over few generations (MacArthur & Wilson, 1967; Saunders, Hobbs & Margules, 1991; Jackson & Fahrig, 2016). During the current Anthropocene period, we are experiencing extreme habitat fragmentation (Haddad *et al.*, 2015) and associated loss of biological diversity (Dirzo *et al.*, 2014). As such, understanding biodiversity in threatened global ecosystems is central to our knowledge of baseline levels of genetic and species diversity in order to implement appropriate conservation strategies.

Despite the importance of understanding global patterns of biodiversity, large gaps exist in our current knowledge of genetic diversity across the worldwide landscape (Miraldo *et al.*, 2016). This is particularly problematic in biodiversity hotspots, regions of high biodiversity threatened by loss of native habitat (Myers *et al.*, 2000). One such biodiversity hotspot, East Africa, has a high diversity of species and ecosystems, especially in the Afromontane regions of the Albertine Rift, Eastern Arc Mountains and Ethiopian

*Corresponding author. E-mail: stephane.boissinot@nyu.edu

Highlands (Zachos & Habel, 2011). While two levels of biodiversity—species and ecosystems—are generally well characterized in these Afromontane regions, relatively little is known of the genetic diversity in the region, reflecting global deficiencies in large portions of Africa and Asia (Miraldo *et al.*, 2016). The lack of knowledge regarding genetic diversity in the Horn of Africa's montane regions requires a more comprehensive survey, because much of the diversity within species is likely underappreciated. For example, birds in the Eastern Arc Mountains often have distinct populations across several sets of isolated mountains, resulting in the risk of undervaluing important regions in conservation plans (Kahindo, Bowie & Bates, 2007).

The Ethiopian Highlands are part of the Horn of Africa biodiversity hotspot (Zachos & Habel, 2011) and possess high levels of endemic species. In addition to charismatic species such as the gelada baboon (*Theropithecus gelada*) and the Ethiopian wolf (*Canis simensis*), the Ethiopian Highlands are rich in endemic species of plants (Friis *et al.*, 2001), mammals (Yalden & Largen, 1992), frogs (Largen & Spawls, 2010) and birds (Klerk *et al.*, 2002). However, despite the high levels of endemism, few studies have focused on intraspecific diversity across the Ethiopian Highlands. Research has shown that several montane species are genetically differentiated across the Great Rift Valley, including the gelada baboon (Belay & Mori, 2006), Ethiopian wolf (Gottelli *et al.*, 2004), groove-toothed rats (Taylor *et al.*, 2011), a giant lobelia (Kebede *et al.*, 2007), coffee (Silvestrini *et al.*, 2007), as well as several species of frogs (Evans *et al.*, 2011; Freilich, Tollis & Boissinot, 2014; Freilich *et al.*, 2016). However, few studies have examined intraspecific genetic variation across multiple populations, and none with the use of genomic-scale data.

In this study, we used the frog *Amietia nutti* (Boulenger, 1896) as a model to study the role that habitat fragmentation played across the Ethiopian Highlands in reducing gene flow between populations, as well as creating and maintaining genomic variation. *Amietia nutti* is a widespread species in East Africa (Fig. 1A), including localities in Uganda, Democratic Republic of Congo, Rwanda, Burundi, Tanzania, Kenya and Ethiopia (Channing *et al.*, 2016; Larson *et al.*, 2016). Across its distribution, *A. nutti* is found in slow moving streams, agricultural water channels and ponds at a variety of elevations (Channing *et al.*, 2016). *Amietia nutti* was originally described as *Rana nutti* (Boulenger, 1896) and later synonymized with *R. angolensis* (de Witte, 1930), although this was not universally accepted. Several different naming schemes for the species were used throughout the past century, including generic labels of *Rana*, *Afrana* and *Amietia* (Visser & Channing, 1997; Channing, 2001; Scott, 2005; Frost, 2011). Most recently, the species has been referred to as either *A. angolensis* or *A. lubrica* (Pickersgill, 2007; Largen & Spawls, 2010; Larson *et al.*, 2016) and was reclassified as *A. nutti* by Channing *et al.* (2016) based on a phylogenetic analysis with genetic and morphological data.

Large amounts of genetic variation have been observed across the range of members of the genus *Amietia*, and specifically in *A. nutti* in Eastern Africa

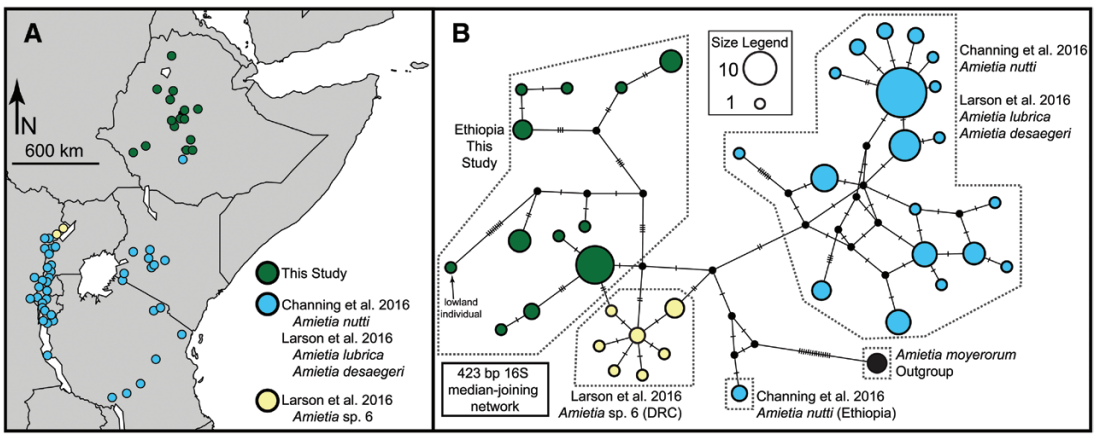

Figure 1. Distribution of *Amietia nutti* in Eastern Africa and genetic structure based on 16S mtDNA. A, known localities of *A. nutti* described in recent papers on the taxonomy and phylogeny of the genus *Amietia* (light blue and light yellow circles) and new samples used in this study (dark green circles). Note that Channing *et al.* (2016) synonymized all naming designations presented in this figure to *A. nutti*. B, haplotype network of 16S mtDNA gene for all *A. nutti* individuals and a closely related outgroup (*A. moyerorum*). Dashes on lines indicate number of mutational steps between haplotypes (circles). Haplotypes are sized based on the number of individuals with a particular haplotype. Dotted lines are not meant to convey support for genetic structure, but rather to match samples with specific studies and the map on the left.

(Channing *et al.*, 2016; Freilich *et al.*, 2016; Larson *et al.*, 2016). A large proportion of the genetic variation in *A. nutti* appears to be influenced by topographical variation in the landscape (Channing *et al.*, 2016; Larson *et al.*, 2016). In these studies, only a single locality of *A. nutti* was identified genetically from Ethiopia (Fig. 1A). More recently, a phylogeographic analysis of *A. nutti* in Ethiopia used one mitochondrial and three nuclear loci to identify phylogeographic breaks across the Blue Nile River Valley and Great Rift Valley (Freilich *et al.*, 2016). These studies suggest that the species is susceptible to gene flow barriers resulting in subsequent isolation and diversification across the landscape of the Ethiopian Highlands.

In order to study how environmental characteristics and landscape topography influence population genomic patterns of variation in *A. nutti* across the Ethiopian Highlands, we sequenced two mitochondrial genes and obtained thousands of single nucleotide polymorphisms (SNPs) using restriction site-associated DNA sequencing (RAD-seq) to investigate the following questions: (1) How are Ethiopian populations of *A. nutti* related to other *A. nutti* populations in Eastern Africa? (2) What are the mechanisms promoting creation, maintenance or degradation of genetic differentiation across this unique montane landscape?

METHODS

Sampling

We collected 36 samples of *A. nutti* between 2011 and 2016 (Table S1; Figs 1A, 2A), with permits obtained from the Ethiopian Wildlife Conservation Authority. We obtained three additional individuals of *A. nutti* from Uganda and Burundi to use as outgroup samples in our analyses (Table S1). These outgroup samples were loaned from the University of Texas at El Paso Biodiversity Collections. For all samples, we extracted genomic DNA from muscle tissue using a magnetic bead DNA extraction protocol (Rohland & Reich, 2012) and quantified and standardized DNA concentrations using Qubit Fluorometric Quantitation (Life Technologies).

Mitochondrial DNA Sequencing and Analysis

We sequenced the mitochondrial genes 16S ribosomal RNA (16S; n = 35) and NADH dehydrogenase 2 (*ND2*; n = 18) for most individuals. For some individuals, *ND2* was already available from a previous study (Freilich *et al.*, 2016) and we retrieved these sequences from GenBank (Table S2). Additionally, we obtained all published 16S sequences of *A. nutti* from previous taxonomic assessments (Channing *et al.*, 2016; Larson *et al.*, 2016). We used these samples to determine if Ethiopian Highland *A. nutti* form a clade, and to understand their evolutionary relationships to other populations across the range of the species. We used polymerase chain reaction (PCR) with primers developed for 16S sequences in frogs (Bossuyt & Milinkovitch, 2000) or specifically for *ND2* in *Amietia* (Freilich *et al.*, 2016). Sanger sequencing of all PCR products was performed by the Beijing Genomics Institute. All raw sequence chromatograms were manually inspected in Geneious v9.1.7 (BioMatters Ltd), and then aligned using the Geneious-implemented MUSCLE (Edgar, 2004). In total, we obtained 422 bp of 16S and 731 bp of *ND2* for our final alignments. We then used the R package 'pegas' (Paradis, 2010) to infer nucleotide diversity for each clade identified from our double digest restriction site-associated DNA sequencing (ddRAD-seq) analyses.

We chose the best model of sequence evolution for each gene using the Bayesian information criterion (K80 + I for 16S and HKY + G for *ND2*) in PAUP v4.0a151 (Swofford, 2003). We used these models of sequence evolution and the resulting alignments of each gene for phylogenetic analyses using maximum likelihood (ML) and Bayesian frameworks. First, we used PhyML (Guindon & Gascuel, 2003), implemented in Geneious, to produce an ML phylogeny, and used 1000 bootstrap replicates to assess support of relationships. Second, in MrBayes v3.2.6 (Ronquist & Huelsenbeck, 2003), we ran four Markov chains for 100 million generations, sampling every 100 000th. We discarded the first 500 trees as burn-in and obtained a consensus topology from the remaining 500 trees.

RAD-seq and Bioinformatics

We used a ddRAD-seq protocol (Peterson *et al.*, 2012) to obtain many homologous loci and associated SNPs from across the genome. We initially digested samples with the restriction enzymes SbfI and MspI, cleaned the samples with AMPure magnetic beads (Agencourt) and ligated adapters with attached barcodes (Table S1). We pooled samples (eight individuals per pool), performed a subsequent bead cleanup, and size selected DNA fragments in the range between 400 and 550 bp using a Pippin Prep electrophoresis cassette (Sage Science). We used a brief PCR reaction (15 cycles) in quadruplicate reactions to attach additional indices (Peterson *et al.*, 2012), and purified with beads a final time. Lastly, we used a BioAnalyzer high sensitivity analysis chip (Agilent) and quantitative PCR to check library quality and quantity, respectively. All libraries were pooled and sequenced (along with samples from unrelated projects) on the Illumina MiSeq or Illumina HiSeq2500 at the New York University Abu Dhabi Center for Genomics and Systems Biology.

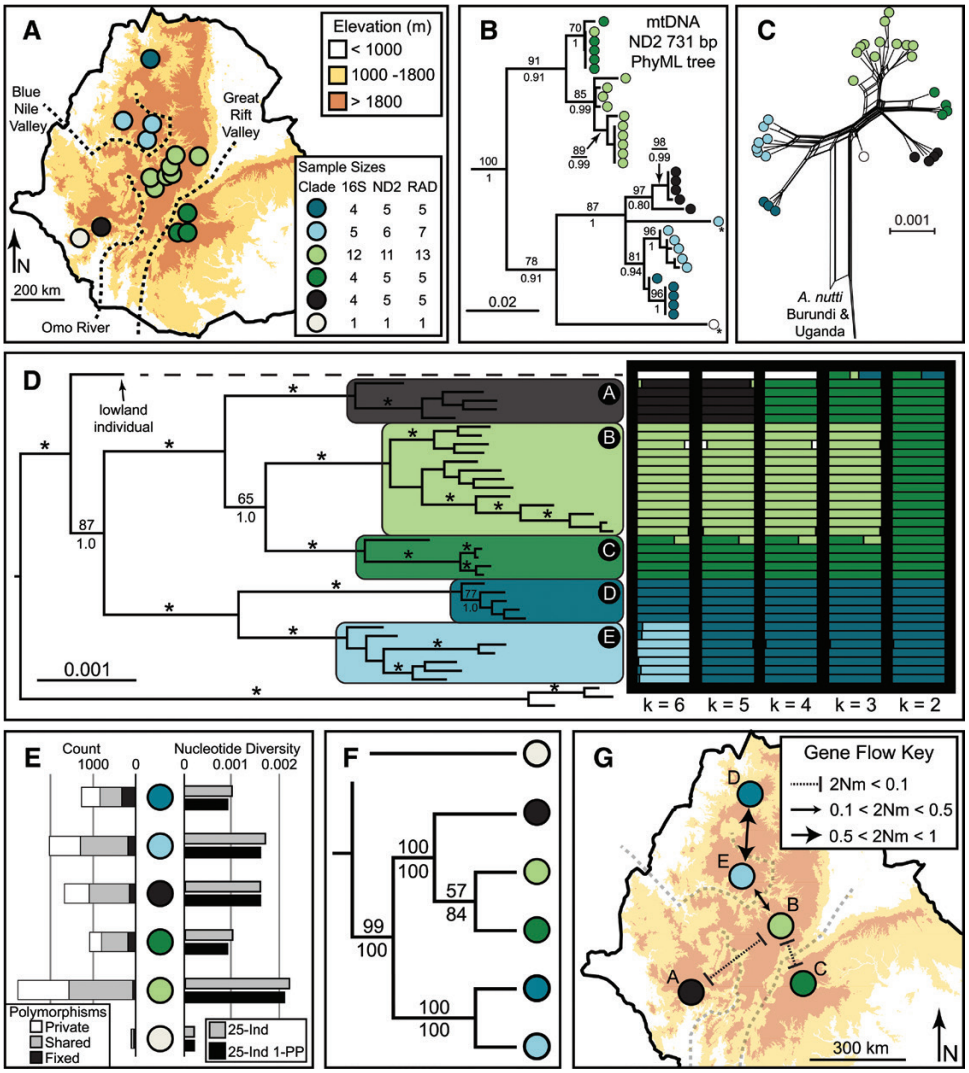

Figure 2. Sampling and phylogeographic structure in *Amietia nutti*. A, sampling map with points coloured according to geographic structure identified in the ddRAD-seq data set. In all other panels, colours correspond to these localities. B, maximum likelihood gene tree inferred using RAxML for the mitochondrial gene *ND2*. Asterisks (*) at tips indicate individuals on long branches referenced in text. C, network of ddRAD-seq SNPs based on uncorrected *p*-distances. D, maximum likelihood phylogeny estimated in RAxML for the concatenated loci obtained from ddRAD-seq. The tree is rooted with *A. nutti* individuals from Burundi and Uganda. Support from RAxML and MrBayes are shown above and below the branches, respectively. An asterisk (*) indicates support > 95% in both analyses. At the tips of the phylogeny are STRUCTURE results for 2–6 genetic clusters. The colours in each bar indicate the probability of genetic cluster assignment for each individual. Labels for each clade (in black circles) identify populations used for ∂a∂i analyses (Table 1; Fig. 2G). E, polymorphism counts for the 25-Individual data set and estimates of genetic diversity for both SNP data sets (see Table 1). F, species tree estimated in SVDquartets, rooted with samples of *A. nutti* from Burundi and Uganda. Support values above and below the branches are for the 25-Individual and 25-Individual 1-PP SNP data sets, respectively. G, estimates of gene flow between geographically neighbouring populations as estimated in ∂a∂i (full results in Table S3). ddRAD-seq, double digest restriction site-associated DNA sequencing; SNP, single nucleotide polymorphism.

We used the STACKS pipeline to assemble loci de novo from the raw sequencing reads (Catchen *et al.*, 2011). We initially used the *process_RADtags* python script included in the STACKS pipeline to demultiplex inline barcodes as well as remove reads of poor quality, lacking either of the restriction sites, or containing ambiguous inline barcodes. For inclusion, we required sequences to have an average *phred* score of ten or

Table 1. Summary of the SNP data sets used in analyses

Data set	# Loci	# SNPs	# SNPs ingroup	% Missing	SNPs per locus*
25-Ind†	2296	8553	7092	22.50	3 (2.08)
25-Ind 1-PP‡	1104	4265	3344	22.39	3.5 (2.17)
∂a∂i A-B§	938	2035	–	8.53	2 (1.31)
∂a∂i B-C§	779	1592	–	8.73	2 (1.19)
∂a∂i B-E§	835	2033	–	7.69	2 (1.39)
∂a∂i D-E§	1405	2906	–	8.55	2 (1.24)

SNP, single nucleotide polymorphism.
*Median (SD) SNPs per polymorphic locus.
†Data set requiring a SNP to be in a minimum of 25 individuals.
‡Data set requiring a SNP to be in a minimum of 25 individuals and represented by at least one individual in each clade (Fig. 2D).
§Data sets requiring a minimum of 80% of individuals from each group (see Fig. 2) for use in ∂a∂i analyses.

higher in sliding windows of 15 bp. Next, we used the FASTX Toolkit (Gordon & Hannon, 2010) to trim restriction sites, trim all sequences to the same length (because of sequencing with different Illumina machines) and convert the sequences to FASTA format. All paired-end reads were concatenated for a total length of 188 bp per combined read.

We used the *ustacks* module in STACKS using the default settings to cluster sequences within each individual. The output of *ustacks* was then used in the STACKS module *cstacks* to match loci between individuals, using a range (N = 1–11) of allowed mismatches between individuals to be assigned to each locus. Here, we assessed how changing the number of mismatches across individuals affected downstream estimates of genetic diversity and differentiation using a subset of 20 individuals (Table S3). Based on this preliminary analysis, we continued with five mismatches allowed between individuals to be assigned to the same locus. Lastly, we used the *sstacks* module in STACKS with default settings to compile results from the *ustacks* and *cstacks* modules.

With the STACKS catalogues, we used the *populations* module to filter the data sets and create final SNP matrices. Here, we required a minimum of five reads coverage for a SNP to be included per individual, a minimum minor allele frequency of 0.05 and maximum heterozygosity less than 50% across all individuals to reduce inclusion of paralogs. From this, we created two SNP data sets based on different coverages: (1) present in a minimum of 25 individuals and (2) present in a minimum of 25 individuals and at least one individual per identified clade. We initially tested the effects of changing the minimum required read depth for SNPs to be included (m = 5, 10, 15). Here, genetic differentiation between sampling localities was consistent across the data sets (all $R^2 > 0.99$). Based on this test, we continued downstream analyses with a minimum read depth of five to maximize the number of loci and SNPs maintained in the final data sets.

RAD-seq Analyses

To investigate the phylogenetic relationships of all individuals, we created a network as well as inferred phylogenies in ML and Bayesian frameworks using the concatenated data set from a minimum of 25 individuals. First, in SplitsTree v4.14.4 (Huson & Bryant, 2006), we used uncorrected p-distances and averaged ambiguous sites to create a network of all individuals. For phylogenetic analyses, we initially used PAUP* v4.0.a151 (Swofford, 2003) to compare models of sequence evolution and the Bayesian information criterion to choose the best model (GTR + I + G). We then used RAxML v8 (Stamatakis, 2014) to estimate an ML tree with support assessed using 1000 quick bootstrap replicates. Next, we used MrBayes v3.2.6 (Ronquist & Huelsenbeck, 2003) with four MCMC chains for 5 million generations sampled every 5000 iterations. We discarded the first 50% of trees as burn-in and used the remaining trees to identify a consensus topology and to provide posterior support values.

To infer genetic structure across individuals, we used the software STRUCTURE (Pritchard, Stephens & Donnelly, 2000). Here, we used the two SNP data sets, and subsampled one random SNP from each locus to reduce effects of linkage on genetic structure inference. We performed two replicates of subsampling to ensure our random subsets did not identify different patterns. In STRUCTURE, we initially inferred lambda with the number of populations set to one (k = 1). We subsequently set a constant lambda and ran a range of possible values for the number of genetic groupings (k = 1–9). For each value of k, we ran five replicates. In STRUCTURE, we used the admixture model with correlated allele frequencies, using a burn-in of 50 000 MCMC generations followed by an additional 50 000. In the results, we manually looked for consistency across replicates and across data sets. All random SNP data sets identified similar patterns of genetic structure; therefore, we used the most inclusive SNP

data set (minimum of 25 individuals for a SNP to be included) for reporting results across a range of likely numbers of genetic clusters up to a maximum that appeared to be biologically relevant (Meirmans, 2015).

We used SVDquartets (Chifman & Kubatko, 2014), as implemented in PAUP* v4.0.a151 (Swofford, 2003), to infer relationships between populations (i.e. species tree analysis). We assigned individuals to populations based on the genetic clusters recovered from STRUCTURE and the phylogenetic analyses. SVDquartets infers a phylogeny for quartets of individuals, and then uses all sampled quartets to identify a species tree phylogeny. Here, we sampled all possible quartets of individuals, inferred the species tree from those quartets and assessed support with 100 bootstrap replicates.

We used the program ∂a∂i (Gutenkunst *et al.*, 2009) to test different scenarios of divergence between geographically neighbouring genetic clusters. We omitted the single lowland individual, because it was a lone sample that appeared to be genetically distinct. For input in ∂a∂i, we reran STACKS for each population pair to obtain a large number of SNPs represented by at least 80% of individuals in each of two genetic clusters in the comparison and with no enforced minor allele frequency (Table 1). We tested four divergence scenarios: (1) no population split, (2) isolation with migration following divergence, (3) strict isolation following divergence and (4) equal bidirectional migration following divergence (Table S4). In ∂a∂i, we tested these models using a composite log-likelihood-based multinomial approach using site frequency spectra (Gutenkunst *et al.*, 2009) derived from one biallelic SNP per locus. All frequency spectra were based on minor allele frequencies (folded terminology in ∂a∂i) and each divergence scenario was performed in triplicate to ensure reproducible results (Table S4).

Correlates of population divergence

To assess possible restrictions to gene flow, we used a multivariate regression of distance matrices (Zapala & Schork, 2006) implemented in the R package 'ecodist' (Goslee & Urban, 2007). Here, we aimed to explain genetic differentiation (as measured by F_{ST}) between localities using a variety of distance measures: (1) geographic distance, (2) environmental distance, (3) number of a priori biogeographic barriers separating populations and (4) number of watersheds separating populations. The number of biogeographic barriers was manually scored for population comparisons; these included the Blue Nile Valley, the Great Rift Valley and the Omo River Valley (Fig. 2A). The number of watersheds separating populations was calculated using ArcMap v10. Here, watersheds were based on a flow direction grid calculated from an elevation raster at 1-km resolution. Geographic distance (km) between localities was calculated using the R package 'fossil' (Vavrek, 2011).

We calculated environmental differences between populations using the methods of Lira-Noriega & Manthey (2014). Briefly, we created a niche model using all known occurrence points of *A. nutti* in the Ethiopian Highlands. One-third of occurrence points (n = 10) were not used to train the model, and were used as a testing dataset to assess the quality of the prediction. We used a subset of the bioclimatic variables from the WORLDCLIM database at a resolution of 2.5', reflecting averages and extremes of temperature and precipitation across the landscape (Hijmans *et al.*, 2005). The subset of variables (Table S5) was chosen to reduce dimensionality as well as remove some of the autocorrelation among layers (all r < 0.7 across Ethiopia). These environmental variables were used with occurrence points in Maxent (Phillips, Anderson & Schapire, 2006) to predict the geographic distribution of *A. nutti* in the Ethiopian Highlands. The model training region was set to Ethiopia. We converted the Maxent output to a binary presence/absence matrix by setting a threshold on the output model inclusive of ≥ 95% of training points, with 5% omission to allow for some error (Peterson, Papeş & Soberón, 2008). We assessed the model performance using a cumulative binomial test; here, the null probability of successfully predicting occurrence of the test points was the proportion of the model region predicted as suitable (~16%). Based on the binomial test, the model predicted the test occurrences better than by chance (P < 0.0001). To assess the environmental space for the species, we extracted the bioclimatic variables' values for our genetically sampled points and 1000 random points from the binary niche model, and subsequently transformed the environmental data using principal components analysis (PCA) in R. From these points, we measured the Euclidean distance between sampled localities in the PCA-transformed environmental space.

Using all of the aforementioned distances (Table S6), we used multivariate regression of distance matrices (Zapala & Schork, 2006) to assess the effects of various distances on the F_{ST} estimates of genetic differentiation. Significance of the model was assessed using 10 000 matrix permutations. Because correlations between explanatory variables can impact model inferences, we used redundancy analyses implemented in the variance partitioning method (varpart) of the R package 'vegan' (Oksanen, 2011) to assess the amount of variance explained in

the response data (genetic differentiation) by each of the explanatory variables.

RESULTS

Mitochondrial DNA sequencing

We obtained alignments of 423 and 731 bp for the 16S and *ND2* genes, respectively. When all samples were included (our samples plus sequences from GenBank), we obtained 47 parsimony informative sites (PIS) in the 16S gene and 101 in the *ND2* gene. When we included only our newly sequenced Ethiopia samples (and no outgroups), we obtained 22 PIS in 16S and 58 PIS in *ND2*.

Bayesian and ML phylogenies of the 16S gene generally showed a lack of support for any branches (Fig. S1), but did indicate that all our *Amietia* samples were closely related to other eastern African *A. nutti* samples. There was limited support that Ethiopian *A. nutti* are paraphyletic, with all other *A. nutti* derived from within the Ethiopian lineages. Because of the lack of phylogenetic support in the 16S phylogeny, we plotted a median-joining network (Bandelt, Forster & Röhl, 1999) using PopART (Leigh & Bryant, 2015) for visualization (Fig. 1B). Based on the haplotype network, our Ethiopian *A. nutti* samples appeared to be most closely related to samples from the Democratic Republic of Congo sequenced by Larson *et al.* (2016). Additionally, the Ethiopian *A. nutti* samples from Channing *et al.* (2016) appeared to be sister to all other *A. nutti* (most closely related to the outgroup; Fig. 1B). Although there is limited support, the haplotype network and phylogeny of the 16S data suggest that diversification of *A. nutti* originated in the Ethiopian Highlands with expansions to other eastern African regions.

Our *ND2* phylogeny contained more PIS and as such the resulting topology was more resolved with strong support for several branches (Fig. 2B). We identified five main clades corresponding to geographic regions of Ethiopia (Fig. 2A, B). The clades largely appear to be separated by hypothesized biogeographic barriers including the Great Rift Valley, Omo River and Blue Nile Valley. Two individuals, one north-west of the Blue Nile Valley as well as the lone lowland individual in the south-west, are on their own long braches (Fig. 2B). With the exception of one potentially introgressed mtDNA haplotype from neighbouring clades across the Great Rift Valley (between dark and light green populations in Fig. 2), all other geographic regions showed strong genetic structure and a lack of mtDNA introgression. Genetic diversity within major phylogeographic clades was not consistent between the 16S and *ND2* data sets (Table 2), likely due to a lack of variable sites within some clades in the 16S data.

RAD-seq analyses

We obtained a total of ~105 million reads across 39 *A. nutti* individuals (Table S1), with a highly variable number of reads per individual (mean ~ 2.7 million reads, SD ~ 1.5 million reads). This resulted in between 5628 and 24 962 RAD-tags per individual (Table S1; mean = 13 600 RAD-tags, SD = 4558). The final SNP data sets used for analyses of genetic variation included between 1100 and 2300 loci, inclusive of ~4300 to 8500 SNPs, respectively (Table 1). This resulted in a median of three SNPs per locus (SD = 2.08) or about one SNP per 60 bp on average.

A network based on uncorrected *p*-distances (Fig. 2C) and phylogenetic analyses (Fig. 2D) identified five geographically structured highland clades. The single lowland individual was sister to all highland Ethiopian *A. nutti*, with strong support for all clades (Fig. 2D). We consider the lowland individual separate from all highland clades because of its elevational dissimilarity to all other sampled individuals. The lowland individual was found at 1010 m, while the lowest elevation of all other sampled individuals was 1861 m (Table S1). The phylogeographic clades generally correspond with regions of Ethiopia and are separated by hypothesized biogeographic barriers: (1) east-southeast of the Great Rift Valley (dark green in Fig. 2); (2) south-west of the

Table 2. Nucleotide diversity in the five highland clades of *Amietia nutti* in this study

Clade	Data set*			
	16S	*ND2*	RAD (25-Ind)	RAD (25-Ind 1-PP)
(A) South-west	0	0.0055	0.0016	0.0016
(B) North-west of Rift	0	0.0076	0.0022	0.0021
(C) South-east of Rift	0.0059	0.0016	0.0010	0.0009
(D) Far North	0.0019	0.0016	0.0010	0.0009
(E) North-west of Blue Nile	0.0092	0.0128	0.0017	0.0016

*The data sets include two mitochondrial genes (16S and *ND2*), and two restriction site-associated DNA sequencing data sets with different filtering restrictions: minimum of 25 individuals (25-Ind) and minimum of 25 individuals and including a minimum of one individual per clade (25-Ind 1-PP).

Omo River (black in Fig. 2); (3) central highlands surrounded by the Great Rift Valley, Omo River and Blue Nile Valley (light green in Fig. 2); (4) north-west of the Blue Nile Valley (light blue in Fig. 2); and (5) Far North near the Simien Mountains (dark blue in Fig. 2). While most breaks correspond with hypothesized inhospitable low elevation biogeographic barriers for highland species, there is also a split of the northernmost population from all others (dark blue in Fig. 2), potentially due to an isolated population, geographic distance or an unidentified biogeographic barrier. The highland clades correspond directly with clades identified with the mtDNA *ND2* gene (Fig. 2B). While most of the phylogenetic relationships among these groups are consistent between data sets, the south-western clade (coloured black and labelled 'A' in Fig. 2) appears in different lineages in the mtDNA and RAD data sets.

STRUCTURE analyses showed genetic clusters consistent with the RAD phylogeny (Fig. 2D) and indicated little evidence for admixture between the five highland clades and single lowland individual. Nucleotide diversity within each of these clades was consistent across data sets (Fig. 2E), but was not correlated with mtDNA nucleotide diversity (all *P*-values > 0.15; Table 2). None of the clades were obviously enriched for fixed or private polymorphisms (Fig. 2E); each clade had a proportion of fixed differences, but a much larger proportion of SNPs shared across clades (Fig. 2E). The species tree inferred in SVDquartets was consistent with the genetic clusters identified in the individual-based phylogeny of genome-wide SNPs (Fig. 2F). The SVDquartets species tree was rooted with individuals of *A. nutti* from Uganda and Burundi.

Using a multivariate regression of distance matrices, biogeographic barriers and geographic distance alone can explain about 19% of the variance in genetic differentiation between localities (as estimated by F_{ST}; Table 3). The entire model, including all variables and possible interactions, explained ~36% of the variation in the data (Table 3). Variance partitioning showed that the strongest explanatory variables for genetic differentiation were geographic distance between localities and biogeographic barriers, suggestive of a large role for landscape topography in shaping phylogeographic structure.

Table 3. Explanatory model of genetic differentiation between populations

	Coefficient	*P*-value	Var-Part
Full model (R^2 = 0.364, P < 0.001)			
Intercept	0.5042	0.494	
Environmental distance	−0.0291	0.140	0.033
Geographic distance	0.0007	0.012	0.123
Watershed crossings	−0.0435	0.281	0.011
# Biogeographic barriers	0.0774	0.049	0.066

Results of the multiple regression of distance matrices, including all explanatory variables of genetic differentiation. Var-Part indicates the amount of variation explained by each variable alone based on variance partitioning.

All divergence scenarios estimated in ∂a∂i based on site frequency spectra suggested very low or complete lack of gene flow between clades following divergence (Fig. 2G; Table S4). Across several of the comparisons, multiple scenarios might equally well explain the data (based on ΔAIC values; Table S4). For example, a scenario of strict isolation with no gene flow may be equally likely to an isolation with migration model with very low levels of gene flow (2Nm < 0.2) between clades. Nevertheless, across all comparisons, no scenarios existed with 2Nm > 1 (Table S4), which would suggest low enough levels of gene flow so as not to counteract the effects of genetic drift in isolation.

DISCUSSION

Phylogeographic patterns

We identified strong phylogeographic structure across the range of *A. nutti* in the Ethiopian Highlands, based on two mitochondrial genes and thousands of SNPs (Figs 1, 2). We identified five highland phylogeographic clades, with little evidence for gene flow between lineages. This is consistent with previous studies that showed pronounced genetic variation across the range of *A. nutti* (Channing *et al.*, 2016; Larson *et al.*, 2016). Although we found high levels of genetic structure across the species' range in Ethiopia, we did not observe markedly different patterns of genetic diversity within each of the clades (Fig. 2E; Table 2).

The phylogeographic clades are split by three hypothesized biogeographic barriers: Great Rift Valley, Blue Nile Valley and Omo River. A putative biogeographic barrier separates the Far North population from all others, possibly due to large tracts of dry and inhospitable areas; this region requires further investigation. These phylogeographic splits observed in *A. nutti*'s nuclear genome are consistent with previous studies that identified phylogeographic breaks across Ethiopia using mtDNA (Fig. 3). The Great Rift Valley has acted as a barrier to dispersal for several taxa, including species of mammals (Gottelli *et al.*, 2004; Belay & Mori, 2006; Taylor *et al.*, 2011), plants (Kebede *et al.*, 2007; Silvestrini *et al.*, 2007) and other frogs (Evans *et al.*, 2011; Freilich *et al.*, 2014, 2016). Fewer studies have investigated taxa across the other barriers identified here, although there are at least two other taxa sharing each of the breaks identified in highland populations of *A. nutti* (Fig. 3). Several frog species have distributional limits bordering these

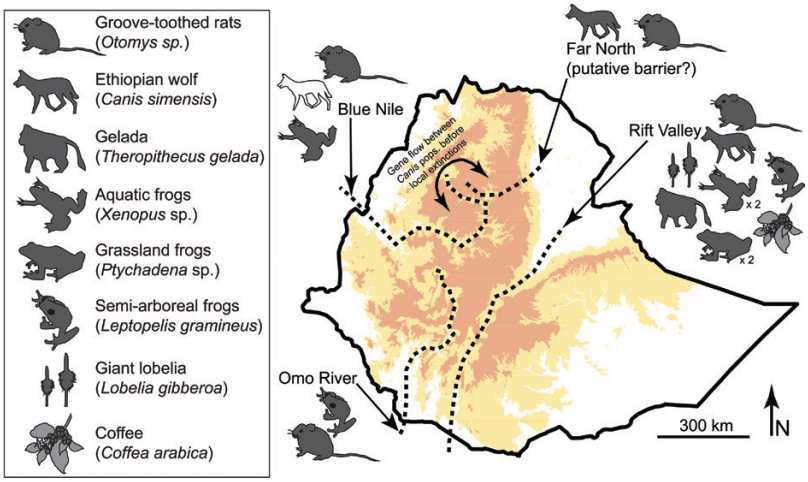

Figure 3. Phylogeographic barriers shared across Ethiopian montane taxa. Four phylogeographic barriers are observed between highland populations of *Amietia nutti* (dashed lines). Taxa that share phylogeographic breaks on either side of the barrier are denoted at each label. In the Ethiopian wolf, one extinct population north-west of the Blue Nile was distinct from populations to the south-east (white shading), but formed a genetic cluster with individuals north of the Far North barrier (denoted by arrows indicating gene flow on map). Two species of *Ptychadena* show no phylogeographic structure across the Omo River barrier. References: *Otomys*: Taylor *et al.* (2011); *Canis*: Gottelli *et al.* (2004); *Theropithecus*: Belay & Mori (2006); *Xenopus*: Evans *et al.* (2011); *Ptychadena*: Freilich *et al.* (2014); *Leptopelis*: Freilich *et al.* (2016); *Lobelia*: Kebede *et al.* (2007); *Coffea*: Silvestrini *et al.* (2007).

barriers, supporting the hypothesis that the barriers limit gene flow. *Leptopelis yaldeni* and *Ptychadena* cf. *neumanni* 5 are range-restricted to the region north-west of the Blue Nile Valley (Largen & Spawls, 2010; Freilich *et al.*, 2014) and *L. vannutellii* is only found west of the Great Rift Valley and south of the Blue Nile Valley (Largen & Spawls, 2010).

While the Omo River has not been hypothesized as a barrier in previous phylogeographic studies of Ethiopian taxa, it separates tropical forest in the south-west from the plateaus of central Ethiopia. Some amphibian species, such as the Ethiopian caecilian (*Sylvaecilia grandisonae*) and Clarke's banana frog (*Afrixalus clarkeorum*), are endemic to this region south-west of the Omo River (Largen & Spawls, 2010). In addition, several species of Central African affinity, such as De Brazza's monkey (*Cercopithecus neglectus*), are found in this region. In two species, *A. nutti* (Fig. 2) and *L. gramineus* (Freilich *et al.*, 2016), individuals collected south-west of the Omo River appear genetically distinct from other populations. In contrast, we may not expect the Omo River as a barrier to species not completely restricted to highland or forest habitats. This idea is supported by a lack of phylogeographic structure across the Omo River in two mid-elevation grassland frogs (*P.* cf. *neumanni* 1 & *P. erlangeri*) (Freilich *et al.*, 2014, 2016). Overall, based on our extensive sampling and identification of phylogeographic structure across many highland regions of Ethiopia, several regions (i.e. far northern and south-western areas of Ethiopia) harbour genetically distinct populations that have been undetected previously. This warrants more widespread sampling of genomic variation in Ethiopian taxa in general.

Identification of previously under or unappreciated genetic variation within species is a common trend in highland species of the Horn of Africa region, with complex and fragmented montane systems in the Ethiopian Highlands, the Eastern Arc Mountains and mountain ranges of the Albertine Rift. Several frog species show complex patterns of diversification and contain high amounts of genetic variation across several mountain chains in the Albertine Rift (Portillo *et al.*, 2015; Larson *et al.*, 2016) and Eastern Arc Mountains (Lawson, 2010; Lawson, 2013). These general trends show the prevalence of distinct evolutionary units across taxa in isolated mountain ranges of the Eastern Africa Rift System and the Horn of Africa biodiversity hotspot. It is necessary to continue identifying unique genetic variation across many regions in the Horn of Africa to minimize the risk of devaluing certain areas for conservation strategies, an issue that has already been emphasized with a focus on birds by Kahindo *et al.* (2007).

Evolutionary mechanisms

The diversification of *A. nutti* appears to have originated in the Ethiopian Highlands with expansions to

other eastern African regions. Non-Ethiopian lineages of *A. nutti* appear derived from within the lineages of Ethiopian *A. nutti* based on 16S data (Fig. 1B; Fig. S1). While the support for an out-of-Ethiopia hypothesis is weak because of the limited amount of phylogenetic information in the 16S data, this pattern has also been described in other east African taxa, including rodents in the genera *Mus* and *Lophuromys* (Lavrenchenko *et al.*, 2007; Bryja *et al.*, 2014). Alternatively, a widespread ancestor of all *A. nutti* may have colonized Ethiopia in one or two different events. These alternative possibilities will require further investigation.

In Ethiopia, the history of divergence between populations may be best characterized as isolation with a lack of gene flow or small levels of gene flow (Fig. 2; Table S4). The general pattern of isolation and little effective gene flow is evident by a general lack of haplotype sharing between mtDNA clades (Fig. 2B), a lack of apparent mixing in the nuclear genome based on genetic structure results (Fig. 2D) and demographic modelling using SNP site frequency spectra (Fig. 2G).

While we did not perform divergence dating with the RAD SNP data, a previous study (Freilich *et al.*, 2016) used mtDNA and three nuclear genes to identify and date the most divergent lineages (clades B, C and E in Fig. 2D); they found that all of the diversification within *A. nutti* in Ethiopia likely took place within the last 2 million years. While the orogeny of the region is old (pre-dating the Miocene), the formation of the Ethiopian Rift Valley began around 20 Mya and continues to this day (Sepulchre *et al.*, 2006; Frisch, Meschede & Blakey, 2010). The rift formation includes the main portion of the Great Rift Valley as shown in the figures in this paper, as well as the area around Lake Tana, i.e. the mouth of the Blue Nile River (Frisch *et al.*, 2010). Even with some error allowed in divergence dating, the formation of major geological structures does not coincide with diversification in *A. nutti*. Rather, the diversification within Ethiopian *A. nutti* appears to be explained best by the colonization of new areas followed by little or no gene flow between neighbouring regions. Habitat elevational shifts across the Ethiopian Highlands during Pleistocene climatic cycles may have facilitated colonization of new areas as well as allowed a relatively increased amount of connectivity between currently separated populations, especially during glacial maxima. This is supported by increased potential suitable habitat of several highland Ethiopian frog species during glacial maxima as inferred via ecological niche modelling (Freilich *et al.*, 2016). However, despite potential increased connectivity between populations during glacial maxima, we recovered little evidence that significant gene flow occurred between the major regions of the Ethiopian Highlands. Overall, we suggest that genetic variation across populations of the Ethiopian Highlands is due to a combination of topographic complexity limiting gene flow and colonization of new areas during distributional shifts facilitated by Pleistocene climate cycles. These hypotheses are supported by the combined patterns of genetic structure (Fig. 2), lack of gene flow between major regions (Fig. 2G) and explanatory variables best able to describe genetic differentiation: geographic distance and biogeographic barriers (Table 3).

In one clade (south-western Ethiopia), we found conflicting relationships between mtDNA and nuDNA (Fig. 2). In both cases, all samples formed a clade, with strong support for phylogenetic relationships. As has been shown across many vertebrates, discordant mito-nuclear phylogeographic patterns are common (Toews & Brelsford, 2012). There are several possible explanations for this pattern depending on the context in particular species, including incomplete lineage sorting and introgression due to hybridization. Here, mito-nuclear discordant patterns are likely due to stochasticity associated with lineage sorting, as we find a complete lack of evidence for gene flow between the south-western Ethiopia clade (clade A, black colour in Fig. 2) and other clades.

CONCLUSIONS

We investigated the patterns and processes of diversification of the frog species *A. nutti* in the Ethiopian Highlands, part of the Horn of Africa biodiversity hotspot. We found strong phylogeographic structure across the region, with diversification largely due to colonization followed by isolation and a general lack of gene flow. Several of the phylogeographic breaks observed here in *A. nutti* are concordant with other taxa from the Ethiopian Highlands. For one clade, we identified discordant mito-nuclear patterns, likely due to stochasticity associated with lineage sorting. In general, our results suggest the necessity for assessing patterns of genetic variation across species' entire distributions in the fragmented and topographically complex montane areas of the East African Rift System to minimize devaluation of peripheral montane areas in conservation strategies.

ACKNOWLEDGEMENTS

We thank the Ethiopian Wildlife Conservation Authority for providing us with permits to work in Ethiopia. Collection of the samples would not have been possible without the help of many field hands over the years, including Sudikchya Shrestha, Jigar

Govind, Mona Nehme, Victor Okoth and Jayson Slovak. Sami Woldeyes, Megersa and Solomon Sebsebie of Rock Hewn Tours were indispensable for planning, transportations, translation and fieldwork across many years. We thank Eli Greenbaum (University of Texas El Paso) for providing three outgroup tissue samples. We thank Marc Arnoux from the Genome Core Facility at NYUAD for sequencing RAD libraries. This research was supported by New York University Abu Dhabi Research Funds AD180 (to S.B.). NYUAD Sequencing and Bioinformatics Cores are supported by NYUAD Research Institute grant G1205-1205A to the Center for Genomics and Systems Biology at NYUAD.

REFERENCES

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Belay G, Mori A. 2006. Intraspecific phylogeographic mitochondrial DNA (D-loop) variation of gelada baboon, *Theropithecus gelada*, in Ethiopia. *Biochemical Systematics and Ecology* **34:** 554–561.

Bossuyt F, Milinkovitch MC. 2000. Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. *Proceedings of the National Academy of Sciences* **97:** 6585–6590.

Boulenger G. 1896. Descriptions of two new frogs from Lake Tanganyika, presented to the British Museum by Mr. WH Nutt. *Journal of Natural History* **18:** 467–468.

Bryja J, Mikula O, Šumbera R, Meheretu Y, Aghová T, Lavrenchenko LA, Mazoch V, Oguge N, Mbau JS, Welegerima K, Amundala N, Colyn M, Leirs H, Verheyen E. 2014. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology* **14:** 256.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1:** 171–182.

Channing A. 2001. *Amphibians of Central and Southern Africa*. Ithaca, NY: Cornell University Press.

Channing A, Dehling JM, Lötters S, Ernst R. 2016. Species boundaries and taxonomy of the African river frogs (Amphibia: Pyxicephalidae: Amietia). *Zootaxa* **4155:** 1–76.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30:** 3317–3324.

de Witte G. 1930. Liste des batraciens du Congo Belge (collection du Musée du Congo Belge à Tervuren). Première partie. *Revue Zoologie et de Botanique Africaines* **19:** 232–274.

Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJ, Collen B. 2014. Defaunation in the Anthropocene. *Science* **345:** 401–406.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Evans BJ, Bliss SM, Mendel SA, Tinsley RC. 2011. The Rift Valley is a major barrier to dispersal of African clawed frogs (*Xenopus*) in Ethiopia. *Molecular Ecology* **20:** 4216–4230.

Freilich X, Anadón JD, Bukala J, Calderon O, Chakraborty R, Boissinot S. 2016. Comparative phylogeography of Ethiopian anurans: impact of the Great Rift Valley and Pleistocene climate change. *BMC Evolutionary Biology* **16:** 206.

Freilich X, Tollis M, Boissinot S. 2014. Hiding in the highlands: evolution of a frog species complex of the genus *Ptychadena* in the Ethiopian highlands. *Molecular Phylogenetics and Evolution* **71:** 157–169.

Friis I, Edwards S, Ensermu K, Sebsebe D. 2001. Diversity and endemism in the flora of Ethiopia and Eritrea–what do the published flora volumes tell us. *Biologiske Skrifter* **54:** 173–193.

Frisch W, Meschede M, Blakey RC. 2010. *Plate tectonics: continental drift and mountain building*. Berlin, Germany: Springer Science & Business Media.

Frost DR. 2011. Amphibian species of the world: an online reference. Version 5: 01.

Gordon A, Hannon G. 2010. *Fastx-toolkit: FASTQ/A short-reads preprocessing tools*. Available at:http://hannonlab.cshl.edu/fastx_toolkit

Goslee SC, Urban DL. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22:** 1–19.

Gottelli D, Marino J, Sillero-Zubiri C, Funk SM. 2004. The effect of the last glacial age on speciation and population genetic structure of the endangered Ethiopian wolf (*Canis simensis*). *Molecular Ecology* **13:** 2275–2286.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696–704.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5:** e1000695.

Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A, Holt RD, Lovejoy TE, Sexton JO, Austin MP, Collins CD, Cook WM, Damschen EI, Ewers RM, Foster BL, Jenkins CN, King AJ, Laurance WF, Levey DJ, Margules CR, Melbourne BA, Nicholls AO, Orrock JL, Song DX, Townshend JR. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances* **1:** e1500052.

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

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Jackson ND, Fahrig L. 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. *Landscape Ecology* **31:** 951–968.

Kahindo C, Bowie RC, Bates JM. 2007. The relevance of data on genetic diversity for the conservation of Afromontane regions. *Biological Conservation* **134:** 262–270.

Kebede M, Ehrich D, Taberlet P, Nemomissa S, Brochmann C. 2007. Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology* **16:** 1233–1243.

Klerk HM, Crowe TM, Fjeldså J, Burgess ND. 2002. Patterns of species richness and narrow endemism of terrestrial bird species in the Afrotropical region. *Journal of Zoology* **256:** 327–342.

Largen M, Spawls S. 2010. *The amphibians and reptiles of Ethiopia and Eritrea*. Frankfurt am Main: Edition Chimaira.

Larson TR, Castro D, Behangana M, Greenbaum E. 2016. Evolutionary history of the river frog genus *Amietia* (Anura: Pyxicephalidae) reveals extensive diversification in Central African highlands. *Molecular Phylogenetics and Evolution* **99:** 168–181.

Lavrenchenko LA, Verheyen WN, Verheyen E, Hulselmans J, Leirs H. 2007. Morphometric and genetic study of Ethiopian *Lophuromys flavopunctatus* Thomas, 1888 species complex with description of three new 70-chromosomal species (Muridae, Rodentia). *Bulletin de L'Institut Royal des Sciences Naturelles de Belgique* **77:** 77–117.

Lawson LP. 2010. The discordance of diversification: evolution in the tropical-montane frogs of the Eastern Arc Mountains of Tanzania. *Molecular Ecology* **19:** 4046–4060.

Lawson LP. 2013. Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the African spotted reed frog. *Molecular Ecology* **22:** 1947–1960.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6:** 1110–1116.

Lira-Noriega A, Manthey JD. 2014. Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution* **68:** 1082–1093.

MacArthur RH, Wilson EO. 1967. *Theory of island biogeography*. Princeton, NJ: Princeton University Press.

Meirmans PG. 2015. Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology* **24:** 3223–3231.

Miraldo A, Li S, Borregaard MK, Flórez-Rodríguez A, Gopalakrishnan S, Rizvanovic M, Wang Z, Rahbek C, Marske KA, Nogués-Bravo D. 2016. An Anthropocene map of genetic diversity. *Science* **353:** 1532–1535.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403:** 853–858.

Oksanen J. 2011. Multivariate analysis of ecological communities in R: vegan tutorial. *R Package Version* **1:** 11–12.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26:** 419–420.

Peterson AT, Papeş M, Soberón J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213:** 63–72.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7:** e37135.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Pickersgill M. 2007. *Frog search, results of expeditions to Southern and Eastern Africa from 1993–1999*. Frankfurt am Main: Edition Chimaira.

Portillo F, Greenbaum E, Menegon M, Kusamba C, Maximilian Dehling J. 2015. Phylogeography and species boundaries of *Leptopelis* (Anura: Arthroleptidae) from the Albertine Rift. *Molecular Phylogenetics and Evolution* **82:** 75–86.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22:** 939–946.

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

Saunders DA, Hobbs RJ, Margules CR. 1991. Biological consequences of ecosystem fragmentation: a review. *Conservation Biology* **5:** 18–32.

Scott E. 2005. A phylogeny of ranid frogs (Anura: Ranoidea: Ranidae), based on a simultaneous analysis of morphological and molecular data. *Cladistics* **21:** 507–574.

Sepulchre P, Ramstein G, Fluteau F, Schuster M, Tiercelin JJ, Brunet M. 2006. Tectonic uplift and Eastern Africa aridification. *Science* **313:** 1419–1423.

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB, Colombo CA. 2007. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genetic Resources and Crop Evolution* **54:** 1367–1379.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Swofford DL. 2003. *PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4*. Sunderland, MA: Sinauer.

Taylor PJ, Lavrenchenko LA, Carleton MD, Verheyen E, Oosthuizen CJ. 2011. Specific limits and emerging diversity patterns in east African populations of laminate-toothed rats, genus *Otomys* (Muridae: Murinae: Otomyini): revision of the *Otomys typus* complex. *Zootaxa* **3024:** 1–66.

Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21:** 3907–3930.

Vavrek MJ. 2011. Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* **14:** 1T.

Visser J, Channing A. 1997. A new species of river frog from the Swartberg, South Africa (Ranidae: Afrana). *Journal of African Zoology* **111:** 191–198.

Yalden D, Largen M. 1992. The endemic mammals of Ethiopia. *Mammal Review* **22:** 115–150.

Zachos FE, Habel JC. 2011. *Biodiversity hotspots: distribution and protection of conservation priority areas*. Berlin, Germany: Springer Science & Business Media.

Zapala MA, Schork NJ. 2006. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences* **103:** 19430–19435.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. ML tree of 16S with labelled groups.
Table S1. Sampling information for this study. This table summarizes data for each individual, including sampling locality, sample number, sequencing coverage, number of RAD-tags recovered and coverage of all loci used for downstream analyses.
Table S2. List of NCBI GenBank accession numbers for mitochondrial DNA sequences from previous studies that were included in this study.
Table S3. Testing the *cstacks* module *N* parameter effects. This table summarizes how changing the *cstacks N* parameter changes inference of genetic diversity within and genetic differentiation between a subset of individuals.
Table S4. Full results of demographic analyses performed in ∂a∂i. Highlighted results indicate models that explained the site frequency spectra best or close to the best model ($\Delta AIC < 10$).
Table S5. Bioclimatic variables used for ecological niche modelling and calculation of environmental distances between populations.
Table S6. Distances used for multivariate regression of distance matrices. Population labels correspond to those identified in Table S1 (Locality).

SHARED DATA

All mtDNA Sanger sequencing data is accessioned in NCBI's Genbank (MF780625-MF780677). All Illumina raw sequence data (ddRAD-seq) are available at the NCBI Sequence Read Archive accessioned under BioProject ID PRJNA387691.