Phylogeography of the Rufous Vanga and the role of bioclimatic transition zones in promoting speciation within Madagascar

Running title: Phylogeography of the Rufous Vanga

Jane L. Younger[1], Phoenix Dempster[1], Árpád S. Nyári[2], T. Olivia Helms[1], Marie J. Raherilalao[3,4], Steven M. Goodman[3,5] and Sushma Reddy[1,6]

Author affiliations:

1. Department of Biology, Loyola University Chicago, 1032 W. Sheridan Road, Chicago, IL, 60660, USA.
2. The University of Tennessee, Department of Ecology and Evolutionary Biology, 569 Dabney Hall, Knoxville, TN, 37996, USA.
3. Association Vahatra, BP 3972, Antananarivo 101, Madagascar.
4. Mention Zoology and Animal Biodiversity, University of Antananarivo, BP 906, Antananarivo 101, Madagascar.
5. Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL, 60605, USA.
6. Bell Museum of Natural History and Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul, MN, 55108, USA.

*Corresponding author: jyounger@luc.edu

Keywords: *Schetba*, Vangidae, phylogenetics, Passeriformes, ecological niche modeling, diversification

Abstract

Madagascar is known as a biodiversity hotspot, providing an ideal natural laboratory for investigating the processes of avian diversification. Yet, the phylogeography of Madagascar's avifauna is still largely unexamined. In this study, we evaluated phylogeographic patterns and species limits within the Rufous Vanga, *Schetba rufa*, a monotypic genus of forest-dwelling birds endemic to the island. Using an integrative taxonomic approach, we synthesized data from over 4,000 ultra-conserved element (UCE) loci, mitochondrial DNA, multivariate morphometrics, and ecological niche modeling to uncover two reciprocally monophyletic, geographically circumscribed, and morphologically distinct clades of *Schetba*. The two lineages are restricted to eastern and western Madagascar, respectively, with distributions broadly consistent with previously described subspecies. Based on their genetic and morphological distinctiveness, the two subspecies merit recognition as separate species. The bioclimatic transition between the humid east and dry west of Madagascar likely promoted population subdivision and drove speciation in *Schetba* during the Pleistocene. Our study is the first evidence that an East-West bioclimatic transition zone played a role in the speciation of birds within Madagascar.

Introduction

Madagascar is a biodiversity hotspot that has been described as a model region for plant and animal diversification studies (Wilmé *et al.*, 2006; Yoder & Heckman, 2006; Vences *et al.*, 2009; Brown *et al.*, 2014). The island's long history of geographic isolation (approx. 88 million years; Storey *et al.*, 1995), coupled with its complex landscape heterogeneity, provided ample opportunities for in-situ lineage diversification resulting in a unique and largely endemic biota (de Wit, 2003; Wilmé *et al.*, 2006). However, phylogeographic structure within the Malagasy avifauna is still largely unexamined, with no published genetic data for almost half of the island's endemic species (Reddy, 2014). Recent discoveries of cryptic species-level diversity within Malagasy birds (Younger *et al.*, 2018), small mammals (Everson *et al.*, 2016; Hotaling *et al.*, 2016; Everson *et al.*, 2018), reptiles (Florio *et al.*, 2012), and amphibians (Brown *et al.*, 2014), coupled with alarming rates of deforestation (Vieilledent *et al.*, 2018), highlight the need for further efforts to comprehend the full

breadth of the biodiversity endemic to Madagascar.

Bioclimatically, the forests of Madagascar can be coarsely subdivided into the humid east and the dry west (Gautier & Goodman, 2003). The eastern edge of the island is characterized by evergreen forest with high precipitation levels, and elevation increases sharply from the coast to the montane forests of the Central Highlands. The western portion of the island receives considerably less precipitation and experiences a pronounced dry season. The biome consists of dry deciduous forest throughout most of the west, spiny bush formations in the subarid southwest, and some smaller areas of subhumid forest. The Central Highlands is situated between the east and west biomes, and is thought to have historically consisted of a matrix of forest and wooded grasslands (Yoder *et al.*, 2016), but now has little remaining native forest habit.

The abrupt bioclimatic transition between eastern and western Madagascar has been hypothesized to act as a facilitator for speciation via ecogeographic isolation (Yoder & Heckman, 2006; Vences *et al.*, 2009). According to this hypothesis, populations of a widespread, generalist ancestral species differentially adapted to conditions in the humid east versus the dry west, producing sister taxa with minimal overlap in their environmental niches. Following this initial divergence, secondary subdivision of populations may occur within the eastern and western bioclimatic zones. Vences *et al.*, (2009) also proposed an alternative mechanism for east-west divergences without adaptation termed the 'Western rainforest refugia mechanism'. In this scenario, an ancestral species adapted to humid forest habitat may have been widespread during warmer (wet) periods of the Pleistocene, but then became isolated in forest refugia during glacial (dry) periods, eventually speciating in allopatry without ecological divergence (Vences *et al.*, 2009). Under this model, sister lineages in eastern and western Madagascar would be expected to occupy similar environmental niches. The east-west bioclimatic transition appears to have facilitated speciation in a range of taxa, including reptiles (Nussbaum & Raxworthy, 1994; Nussbaum & Raxworthy, 1998; Raxworthy *et al.*, 2007; Orozco-Terwengel *et al.*, 2008; Florio *et al.*, 2012), amphibians (Vences *et al.*, 2000; Andreone *et al.*, 2002; Vences & Glaw, 2002; Köhler *et al.*, 2007), insects (Lees *et al.*, 2003), and mammals (Everson *et al.*, 2016; Yoder *et al.*, 2016). So far, little is known about an east-west speciation pattern in birds. A study of Malagasy Scops-owl (*Otus rutilus*) found a

pattern of subtle genetic differentiation between east and west Madagascar, suggesting the bioclimatic transition could contribute to population divergence in birds (Fuchs *et al.*, 2007).

Schetba, or the Rufous Vanga, is a monotypic genus within an endemic Malagasy radiation of songbirds, the Vangidae (Yamagishi *et al.*, 2001; Reddy *et al.*, 2012; Jønsson *et al.*, 2012). The current taxonomy of the genus comprises a single species, *S. rufa*, with two subspecies: *S. r. rufa* (Linnaeus, 1766) and *S. r. occidentalis* (Delacour, 1931). *S. r. rufa* occupies the humid and littoral forests of eastern Madagascar, whereas *S. r. occidentalis* is found in the deciduous forests and subhumid forests of western Madagascar (Goodman & Raherilalao, 2013; Schulenberg, 2013). Given the geographic division of these subspecies in the east and west, it is possible that the bioclimatic transition between these regions played a role in the divergence of *Schetba.* However, the degree of divergence between these subspecies, which are based on slight differences in bill dimensions and plumage, has not been corroborated with genetic or ecological data. Previous genetic studies each included only a single representative of *S. rufa* (Yamagishi *et al.*, 2001; Jønsson *et al.*, 2012; Reddy *et al.*, 2012). Furthermore, the distributional limits of *S. r. occidentalis* and *S. r. rufa* are somewhat ill-defined and might be attributed to clinal variation (Delacour, 1932; Schulenberg, 2013). Given the dearth of conclusive information, some have suggested that *S. rufa* may be better regarded as monotypic (Schulenberg, 2013).

Here, we aim to (1) clarify the taxonomy of *Schetba* in light of recent findings of cryptic species-level diversity within Madagascar (e.g. Younger *et al.*, 2018), and (2) investigate whether the bioclimatic transition between the humid east and dry west of Madagascar may have facilitated speciation within birds. We synthesized data from over 4,000 ultra-conserved element (UCE) loci, mitochondrial DNA, morphometrics, and ecological niche modeling in an integrative systematics approach to assess species limits and explore phylogeographic patterns within the genus.

Materials and Methods

Taxon sampling

We sampled *Schetba* from across its geographic range in order to assess phylogeographic patterns, subspecies definitions, and subspecies distributional limits (Figure 1). Tissue samples used for genotyping are associated with vouchered specimens held at the Field Museum of Natural History (FMNH; Chicago) and the Mention Zoologie et Biologie Animale, Université d'Antananarivo (UADBA; Antananarivo, formerly Département de Biologie Animale). We genotyped 27 individuals of *S. rufa*, plus two outgroup species of Vangidae (*Euryceros prevostii* and *Newtonia amphichroa*). Morphometric data was collected from 20 adult *S. rufa* study skins (five individuals of each sex for each subspecies) in the FMNH and American Museum of Natural History (AMNH; New York) collections. For detailed location information (locality, latitude, and longitude), accession numbers, and data collected from each specimen, please refer to Supplementary Table 1.

Sequencing

DNA was extracted using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's protocol. UCE libraries for 28 taxa (27 *Schetba* plus *Euryceros* outgroup) were prepared following described methods (Faircloth *et al.*, 2012; McCormack *et al.*, 2013) with minor modifications. Briefly, purified DNA was normalized to 10 ng/µL and fragmented via sonication (Covaris, Model #M220) to approximately 550 base pairs (bp). Samples were end-repaired, A-tailed and Illumina TruSeqHT adapters were ligated using either a TruSeq DNA HT Sample Prep Kit (Illumina) or a KAPA Hyper Prep Kit (Kapa Biosystems), following the manufacturer's instructions. Libraries were then amplified by limited-cycle (16—18) PCR using Kapa HiFi DNA polymerase (Kapa Biosystems), normalized, and pooled into sets consisting of eight libraries each (along with taxa for other studies) with a total of 500 ng of sample. We enriched these pooled libraries for 5,060 UCE loci using MYbaits capture kits (Terapods 5K v1, MYcroarray) following the manufacturer's instructions. Enriched libraries were quantified using qPCR (Kapa Library Quantification Kit) and a Qubit Flourometer (Invitrogen), normalized, and pair-end sequenced (2 x 250 bp) on the Illumina HiSeq2500 platform. DNA sequence reads are archived on NCBI SRA (XXXXX).

We amplified and sequenced the mitochondrial gene NADH dehydrogenase 3 (ND3) for 26 taxa (including outgroups *Euryceros* and *Newtonia*) using standard PCR and Sanger sequencing methods with primers ND3-L10751 (5'-GACTTCCAATCTTTAAAATCTGG-3') and ND3-H11151 (5'-GATTTGTTGAGCCGAAATCAAC-3'). We used Geneious 9.0.5 for alignment and sequences were deposited in GenBank (TBA — TBA). We also extracted mitochondrial cytochrome b (CYTB) sequences from off-target contigs of the UCE protocol using the Megablast function within Geneious 9.0.5, and successfully recovered CYTB for 24 of the *S. rufa* individuals.

Bioinformatics

We used the PHYLUCE 1.5 package (Faircloth, 2015) to prepare alignments of UCE loci for phylogenetic analysis. The demultiplexed reads were trimmed to remove adapters and low-quality bases using Illumiprocessor (Faircloth, 2013), then assembled into contigs using Trinity 2.0.4 (Grabherr *et al.*, 2011). UCE loci were extracted from among the contigs using PHYLUCE and then aligned with MAFFT 7 (Katoh *et al.*, 2002; Katoh & Standley, 2013). The alignments were trimmed using the edge-trimming algorithm available in PHYLUCE, and then a data matrix of 75% completeness was generated, where 'completeness' refers to the minimum number of taxa sequenced for a locus to be included in the matrix.

We prepared a dataset of single nucleotide polymorphisms (SNPs) for the 27 *Schetba* individuals, following the methods of the seqcap_pop pipeline (Harvey *et al.*, 2016), with some modifications. In brief, following cleaning of the reads with Illumiprocessor, we used Trinity 2.0.4 to assemble reads across all specimens into contigs *de novo*. Contigs matching UCE probes were then extracted using PHYLUCE and were used as a reference for SNP calling. The reads for each individual were mapped to the reference contigs using BWA (Li & Durbin, 2009), with a maximum of four mismatches allowed per read. We used SAMtools (Li *et al.*, 2009) and Picard (http://broadinstitute.github.io/picard/) to convert sam files to bam format, soft-clip reads beyond the reference, add read groups for each sample, and then merge bam files across all samples in the dataset. We used the Genome Analysis Toolkit (GATK; McKenna *et al.*, 2010) to realign reads and indels, call SNPs, annotate SNPs and indels, mask indels, remove SNPs with a quality score < Q30, and to conduct read-backed

phasing. At this point we output a dataset of phased SNPs in vcf format for further filtering. We filtered the SNP dataset using VCFtools 0.1.15 (Danecek *et al.*, 2011): we specified a minimum read depth of three for a genotype call; removed any SNPs with a minor allele count < 2 (these are potential sequencing errors and generally uninformative loci); restricted to biallelic SNPs; and removed any variants not genotyped in 100% of individuals. We then used a custom python script to select one SNP at random per contig to reduce linkage in the final dataset. VCFtools 0.1.15 was used to calculate mean sequencing coverage of each SNP. Because our analysis found two highly distinct groups within *Schetba* (East and West groups), we also prepared separate SNP datasets for each of these group to allow for separate clustering analyses within the East and West to detect fine-scale genetic structure. After the final filtering with VCFtools on the entire dataset as described, we divided the dataset into East and West datasets, then applied a minor allele count filter to remove positions that are invariant within these groups, and finally selected one SNP at random per contig. PGDSpider 2.1.0.0 (Lischer & Excoffier, 2012) was used to convert vcf files into other formats required for analysis.

Phylogenetic analysis

We inferred maximum likelihood (ML) phylogenies for the UCE dataset using RAxML 8.2.7 (Stamatakis, 2014). We performed both unpartitioned and partitioned concatenated analyses. To find the most appropriate partitioning scheme for the UCE dataset we used the Sliding-Window Site Characteristics (SWSC) entropy based method (Tagliacollo & Lanfear, 2018) to generate partitions that account for within-locus heterogeneity (e.g., the flanking regions of UCE loci are typically more variable than the ultraconserved core). These partitions were then input to PartitionFinder 2 (Lanfear *et al.*, 2014; Lanfear *et al.*, 2016), to estimate the optimal partitioning scheme for phylogenetic analysis by grouping together similar subsets from the SWSC output. For each RAxML analysis, we conducted rapid bootstrapping analysis and a search for the best-scoring ML tree in a single program run, using the MRE-based bootstopping criterion (Pattengale *et al.*, 2010) to ascertain when sufficient bootstrap replicates had been generated. All searches were conducted under the GTR GAMMA site-rate substitution model.

We also inferred a phylogeny under the multispecies coalescent method. Gene-tree based coalescent methods may have reduced accuracy when inadequately resolved gene trees are included, which can result from using loci with low phylogenetic signal (Gatesy & Springer, 2014; Xi *et al.*, 2015; Hosner *et al.*, 2016; Meiklejohn *et al.*, 2016). We therefore selected the 25% of UCE loci with the greatest number of parsimony informative sites for analysis. This subset contained 1,062 loci with between five and 26 parsimony informative sites each. A gene tree was estimated for each locus with 100 ML searches under GTR GAMMA using RAxML, and these were then reconciled into a gene tree-species tree using ASTRAL 4.10.12 with default settings (Mirarab & Warnow, 2015).

Divergence time estimation

We performed time-calibrated Bayesian phylogenetic analyses on mtDNA sequences (ND3 and CYTB) using BEAST 2.4.4 (Bouckaert *et al.*, 2014) to estimate divergence times among *Schetba* lineages. The mtDNA genes were used because estimates of divergence rates in birds are available for these loci (Lerner, Meyer, James, Hofreiter, & Fleischer, 2011; Weir & Schluter, 2008). Furthermore, the mtDNA gene trees resolved the same well-supported clades as the UCE dataset. The data was partitioned into ND3 and CYTB, with nucleotide substitution models specified as HKY for both genes to reflect the optimal models selected by PartitionFinder 2 (Lanfear *et al*., 2016). We used the Yule tree prior with a strict molecular clock. The molecular clock was calibrated using two different reference rates; (1) the divergence rate of CYTB for Passeriformes of 2.07% (± 0.20) per million years (Weir & Schluter, 2008; lognormal, mean = 0.01035, SD = 0.05); and (2) the substitution rates estimated for ND3 and CYTB for Hawaiian honeycreepers (Lerner *et al*., 2011; ND3: lognormal, mean = 0.024, SD = 0.09; CYTB: lognormal, mean = 0.014, SD = 0.05). Two independent analyses were performed for each to ensure reproducibility of the posterior distributions. The MCMCs were run until convergence of the posteriors, as confirmed using Tracer v1.6 (Rambaut & Drummond, 2007). We estimated maximum clade credibility trees with mean node heights from each posterior after removing the first 10% of samples as burn-in.

Genetic clustering analyses and summary statistics

To estimate the number of genetic clusters in the *Schetba* SNP dataset, we performed Discriminant Analysis of Principal Components method (DAPC; Jombart *et al.*, 2010), and Bayesian clustering within Structure 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). The DAPC method, implemented in *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011), creates discriminant functions to maximize variance among, whilst minimizing variance within, genetic clusters. The most likely number of clusters in the dataset, and the assignment of individuals to those clusters, was estimated using successive *K*-means clustering, with the number of clusters selected based on minimum BIC. Then DAPC was performed, using the cross-validation method (1000 replicates) to determine the optimal number of PCs to retain. Finally, we plotted the posterior membership probability of all *Schetba* taxa to the genetic clusters.

For a given number of clusters (*K*), Structure identifies genetic clusters within the dataset and estimates the corresponding membership coefficients for each. We performed Structure analyses for the entire *Schetba* dataset, as well as for *S. r. rufa* and *S. r. occidentalis* separately in order to detect any fine-scale genetic differentiation within the eastern and western sectors of the island. For all analyses, we used the admixture model with correlated allele frequencies and ran the model without sampling locations as priors. For each dataset, we performed an initial run of 100,000 generations, discarding the first 50,000 as burn-in, with $K = 1$ and lambda allowed to vary in order to estimate a value for lambda (the allele frequencies prior) for the dataset. For subsequent runs, the value of lambda was set to the estimated value, and the number of clusters was allowed to vary from $K = 1$ to $K = 10$ (for the full dataset), and from $K = 1$ to $K = 5$ for the analyses on the East and West groups. Each analysis was run for 500,000 generations, discarding the first 100,000 as burn-in, and repeated ten times. We used Structure Harvester Web 0.6.94 (Earl, 2012) to assess convergence across replicates, to determine the most optimal value of K for the three datasets (based on the log likelihood of each value of *K*, and the Evanno method (Evanno *et al.*, 2005)), and to prepare input files for CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007). CLUMPP was then used to calculate average membership coefficients from across the replicates. Distruct 1.1 (Rosenberg, 2004) was used to visualize the final results for several values of *K*, in order to better understand the levels of genetic structure within *Schetba*. Previous work suggests that a "true" value of *K* does not usually exist (Gilbert *et al.*,

2012; Benestan *et al.*, 2016; Janes *et al.*, 2017), and that in order to gain insight into different levels of genetic structure it is best practice to view multiple *K*-values.

We used Genodive 2.0b27 (Meirmans & Van Tienderen, 2004) to calculate the Weir and Cockerham unbiased weighted F_{ST} estimator (Weir & Cockerham, 1984) between the East and West clades, with significance calculated using 10,000 permutations of the data. We also used Genodive to calculate observed (H_O) and expected (H_S) heterozygosity.

Ecological niche modeling

Our occurrence dataset comprised a total of 16 spatially unique latitude/longitude combinations for *S. r. occidentalis* and 18 for *S. r. rufa*. Bioclimatic variables for Madagascar were used to summarize aspects of temperature and precipitation from the latter half of the 20th century (Hijmans *et al.*, 2005), as well as for the Last Glacial Maximum (LGM; ~21,000 years BP; under both Community Climate System Model (CCSM) and Model for Interdisciplinary Research on Climate (MIROC) scenarios). We used bioclimatic GIS layers (http://www.worldclim.org) at a spatial resolution of 2.5 arc-minutes. To account for dimensionality across environmental spaces and time scales, we used a subset of six of the 19 layers that showed lowest correlation ($p < 0.7$): annual mean temperature (bio1), mean diurnal range (bio2), maximum temperature of warmest month (bio5), annual precipitation (bio12), precipitation of wettest month (bio13), and precipitation of driest month (bio14). We used MaxEnt v.3.4.1 (Phillips *et al.*, 2006) to construct ecological niche models of each subspecies. Owing to the low number of unique occurrences for each subspecies, we set the algorithm to perform cross validation with five replicates and a 10% training presence threshold. The spatial extent of our model training was kept at the level of the entire island of Madagascar, while our models were run using climatic variables for the present time frame and then projected onto LGM past conditions (CCSM and MIROC scenarios). We performed an additional MaxEnt run with all occurrence points of the two taxa combined (*S. rufa* sensu lato; total of 34 unique points), in order to examine potential ecological and biogeographical divergences and breaks within this taxon. For this run we once again used cross validation with five replicates and a 10% training presence threshold. Niche similarity between the two taxa was assessed by calculating Schoener's *D* metric using the Maxent

estimates of present-day potential distributions. To evaluate statistical significance of the niche similarity measure, we generated a null distribution of *D* values for each of the two taxa through 100 simulated models based on the same environmental layers and background extent, and random samples of background in place of occurrence records (Warren *et al*., 2008).

Morphological variation

We measured 20 *Schetba* skin specimens (10 per subspecies) to examine morphological variation. One of us (TOH) took standard linear measurements of bill length from the crown to tip (BL), bill width at the anterior edge of nares (BW), bill depth at nares (BD), tarsus length (TL), hallux length (HL), tail length (Tail), and wing chord length (WL). These measurements followed the descriptions in (Baldwin *et al.*, 1931). Wing and tail lengths were measured with a wing rule to an accuracy of 1 mm, all other measurements were taken with Mitutoyo Digital Calipers to an accuracy of 0.01 mm. All measurements were repeated three times, checked for outliers (by confirming that all measurements for an individual were within one standard deviation), and then averaged. The summary statistics of these measurements for the two clades are given in Supplementary Table 2. We first tested whether males and females exhibit significant variation by conducting an ANOVA for each variable between sexes within each clade. Next, we log-transformed and standardized all measurements and conducted principal components analysis (PCA) on all specimens to examine the morphological variation between the two genetic clades. We conducted a multivariate analysis of variance (MANOVA) to determine whether the centroids of the two clades were statistically different. There were five specimens for which wing measurements could not be made and since missing data is problematic in multivariate analyses, we removed wing length and used only the remaining six variables for these analyses. We also conducted ANOVA tests for each measured trait with clade as a factor to determine which traits differed significantly between clades. We used the R statistical package for all statistical analyses.

Results

Sequence capture of UCE loci

After removal of adapters, low quality bases and unpaired reads, an average of 350 million bp of sequence per individual remained (46 million – 589 million bp). These reads were assembled into an average of 15,448 contigs per individual, with a mean contig length of 508 bp. An average of 4,235 UCE loci were recovered per individual (3,139–4,421), with 4,951 UCE loci recovered across all taxa. The 75% complete data matrix used for analysis consisted of 4,243 loci with a mean locus length of 784 bp. The concatenated alignment was 3,328,172 bp in length, and contained 15,392 parsimony informative sites.

The recovered UCE loci contained a total of 56,701 SNPs. Our filtering protocols reduced this to 12,045 SNPs, and after thinning to one SNP per contig our final dataset contained 3,609 SNPs for use in subsequent analyses. The mean sequencing coverage of these SNPs was 68X. The SNP datasets we prepared for *S. r. rufa* and *S. r. occidentalis* contained 2,873 and 3,044 SNPs, respectively.

Phylogenetic relationships

Our phylogenetic analyses converged on a strongly supported topology showing a clear division of *S. rufa* into two reciprocally monophyletic clades (Figure 2), corresponding to eastern and western Madagascar. The ML phylogenies also indicated several well-supported clades within each of the eastern and western clades, corresponding with latitudinal subdivision (details in section on fine-scale genetic structure, below). The topology recovered from ML analysis of the 4,243 UCE loci dataset was robust to partitioning scheme (Figure 2, Supplementary Figure 1). The ASTRAL species tree constructed from the 1,062 most informative UCE loci had 100% support for the eastern and western clades, and had a normalized quartet score of 0.42 (Supplementary Figure 2). The sub-clades within the eastern and western clades were less well supported in the ASTRAL tree, indicating a degree of either incomplete lineage sorting or gene flow, as expected for intraspecific comparisons.

The eastern and western clades were also reciprocally monophyletic and 100% supported in the mitochondrial tree (phylogeny not shown). We estimated that the divergence of eastern and western clades of *S. rufa* occurred approximately 854,000 years ago (median estimate,

95% HPD: 0.582 – 1.16 MYA), based on the Weir & Schluter (2008) calibration for all Passeriformes. Our estimates of divergence times based on the two calibration strategies had overlapping 95% HPDs, with a slightly younger estimate of lineage divergence based on substitution rates in Hawaiian honeycreepers (Lerner *et al*., 2011) of 0.536 MYA (median estimate, 95% HPD: 0.380 – 0.736 MYA).

Clustering analyses and differentiation measures (eastern vs. western *Schetba rufa*)

The optimal number of genetic clusters in our Structure analysis of the 27 *S. rufa* individuals was $K = 2$, based on both the maximum posterior log likelihood and the rate of change in log probability (deltaK, Evanno method). Assignments of individuals to these clusters was consistent with the results of our phylogenetic analyses, dividing *S. rufa* into two genetic groups originating in eastern and western Madagascar (Figure 3a, Figure 2). Successive *K*-means clustering also clearly indicated $K = 2$ as the most likely number of clusters, and DAPC was able to differentiate between these with 100% support (root mean squared error = 0), even when only a single PC was retained for analysis (Supplementary Figure 3a). The posterior membership probabilities for all taxa were 100% to their respective clusters in both Structure and DAPC, with no evidence of admixture between the east and west groups (Figure 3a, Supplementary Figure 3b).

Our estimate of F_{ST} between the east and west groups was 0.256 (95%CI: 0.235 – 0.277, *p*-value < 0.0001), suggesting strong, statistically significant genetic differentiation between them. There were 103 fixed SNPs between the two clades (across the full SNP dataset). The expected (H_S) heterozygosity for the western clade was greater than that of the eastern clade (0.143, 95%CI: 0.138 – 0.147; compared to 0.133, 95%CI: 0.128 – 0.138).

Fine-scale genetic structure

To investigate finer-scale divergences within the eastern and western groups of *S. rufa*, we conducted further Structure analyses on these two groups separately. For the *S. r. rufa* (the eastern clade), the posterior log likelihood was maximized at $K = 3$, whereas deltaK was maximized at $K = 2$. In the two-cluster scenario, individuals from the northeast humid forest (Masoala National Park) are clearly differentiated from those in the southeastern humid

forest with minimal admixture (Figure 3b). In the three-cluster scenario, this division between northeast and southeast is still apparent, and three individuals from the northwestern sector of Masoala National Park (near Hiaraka village) are largely assigned to a third cluster, distinct from the other Masoala National Park individuals (Figure 3c). The four-cluster scenario is consistent with this finding, showing no further genetic structure (Figure 3d). In our phylogenetic analysis, the individuals from the southeastern forest were monophyletic with 100% bootstrap support, but the individuals from Masoala National Park were paraphyletic, with those individuals from the eastern sector (Sarahandrano Forest) appearing the most divergent (Figure 2). Based on this inconsistency regarding genetic subdivision in Masoala, we conclude that there are most likely two genetic populations of *S. rufa* in the eastern humid forests; in Masoala National Park and in the southeastern region.

For the individuals from western Madagascar, the optimal number of clusters in our Structure analyses was four, based on both the posterior log likelihood and deltaK. In a two-cluster scenario (Figure 3e), the individuals from the northwest forest (Namoroka and Ankarafantsika) were differentiated from the rest of the western clade, a split which was also supported in our phylogeny (Figure 2). When $K = 3$ further subdivision is apparent, with Namoroka and Ankarafantsika individuals largely assigned to distinct clusters (Figure 3f). These groups are located south and north of the Betsiboka River, respectively, and this split has 100% support in our phylogenetic analysis (Figure 2). In the four-cluster scenario, there is further divergence between individuals from the southwest and central-west regions (Figure 3g); this split has 100% support in our phylogeny (Figure 2). Therefore, it appears that there are four genetically differentiated populations in western Madagascar, separated latitudinally. Overall, our genetic data provide evidence for an initial divergence in the *S. rufa* complex between the east and west of Madagascar during the mid-Pleistocene, followed by more recent divergences within these two regions, which perhaps reflect the fragmented nature of Madagascar's forest habitat and/or low levels of dispersal of *Schetba*.

Ecological niche modeling

Our ecological niche models for *Schetba* provided a good fit to their contemporary distribution (Goodman & Raherilalao, 2013; Schulenberg, 2013), with the caveat that the

actual inhabited area is smaller than predicted in the model owing to recent deforestation (Vieilledent *et al.*, 2018). Of the five model replicates for the separate subspecies (*S. r. occidentalis* and *S. r. rufa*) and the single taxon (*S. rufa s.l.*), we selected the run with the best performance (highest AUC values and lowest testing data omission error) for further interpretation.

Our combined single taxon ecological niche model (pooled dataset of 34 unique points) recovered two distinct areas of suitability (Figure 4), corresponding to the eastern and western clades evident in our phylogenetic analyses. Separate MaxEnt models of each subspecies (*S. r. occidentalis* and *S. r. rufa*) produced similar geographic signatures, with the individual models showing suitable habitat in western and eastern Madagascar, respectively. Slight differences in the individual models compared to the pooled dataset (*S. rufa s.l.*) were observed in an apparent connection between the two subdivided habitats in western Madagascar, which correspond to the genetic break between the northwest forest (Namoroka/Ankarafantsika) and the remainder of the western clade of *S. r. occidentalis* (Figures 2, 3e). While this subdivision was not recovered in the present-day model of *S. r. occidentalis*, this separation was visible in the LGM model projections for this taxon. Models of the combined dataset also differed from the individual models for *S. r. occidentalis* and *S. r. rufa*, by rendering areas of eastern Madagascar as largely habitable by *Schetba* during the LGM scenarios, but omitting suitable habitats in the northwest. The individual model for *S. r. occidentalis* produced models with suitable areas in western Madagascar during both LGM scenarios (Figure 4). Notable for the *S. r. occidentalis* models is the difference between the CCSM and MIROC LGM scenarios, where under the former scenario only the southwestern region of the island is predicted as having large extents of suitable areas, while the northwestern suitable area is reduced to a smaller, isolated patch (Figure 4). The observed niche similarity between the two taxa based on Schoener's *D* was 0.182. This value was outside the lower bound of the 95% confidence interval of the simulated null distributions of *D* values, indicating that the niches of the two taxa are significantly dissimilar.

Morphological variation

There was no significant difference between sexes within each clade based on our ANOVA, so we used all individuals together for subsequent analyses. Univariate ANOVA of each measurement separately showed that *S. r. rufa* and *S. r. occidentalis* were significantly different in terms of bill length, bill depth, tarsus, and tail length (Supplementary Table 2). We used all 20 individuals and six variables (removing wing length due to missing data) for the PCA, which resulted in six PCs, with the first four explaining more than 90% of the variance (see Supplementary Table 3). The two *Schetba* clades formed distinct clusters in morphospace (Figure 5; Supplementary Figure 4). Our MANOVA test determined that the clade centroids were significantly different ($p < 0.001$).

Discussion

Previously unrecognized species diversity within *Schetba*

We found that the two *Schetba rufa* subspecies are geographically, genetically, ecologically, and morphologically distinct. The *S. r. rufa* and *S. r. occidentalis* lineages are restricted to eastern and western Madagascar, respectively, occupying distinct ecological niches separated by a large expanse of unfavorable habitat (the Central Highlands). The subspecies formed reciprocally monophyletic clades in all of our analyses. We estimate that these lineages diverged 854,000 years ago (95% HPD: 0.582 – 1.16 MYA), and have since accumulated fixed SNP differences in their nuclear genomes and diverged in their genetic diversity levels. They have also diverged in morphology, such that *S. r. occidentalis* has a significantly longer tail, longer tarsus, and longer and heavier bill than *S. r. rufa*. This result is consistent with other morphological studies (Schulenberg, 2013).

The genetic and morphological differences described here suggest that the two *S. rufa* subspecies merit recognition as separate species. We therefore propose that within the currently defined *S. rufa*, the western subspecies, *occidentalis*, should be elevated to species level, *S. occidentalis.* We suggest the common name ‘Western Rufous Vanga’ for this new species, to reflect its geographic distribution. The eastern subspecies, *rufa*, would remain *S. rufa*. A full description for the *S. r. occidentalis* subspecies already exists (Delacour, 1931), therefore we do not include a species description for *S. occidentalis* here. These two species

are on separate evolutionary trajectories, and their distinctiveness should be taken into consideration in future conservation plans and biodiversity studies. Only by recognizing and conserving the full spectrum of genetic and morphological variation can the adaptive potential of *Schetba* be maximized (Funk *et al.*, 2012; D'Amen *et al.*, 2013).

Our discovery of unrecognized species-level diversity within *Schetba*, coupled with the recent discovery of other cryptic species diversity within the endemic Vangidae family (Younger *et al.*, 2018), suggests that the avian species richness of Madagascar may still be underestimated. This is concerning given the high rates of deforestation and forest fragmentation (Vieilledent *et al.*, 2018) that are currently threatening the island's avifauna. Recent efforts in avian taxonomy suggest that unrecognized species may be a widespread problem, leading to substantial underestimates of avian biodiversity levels and fine-scale endemism (Barrowclough *et al.*, 2016; Hosner *et al.*, 2018). Given that most conservation plans rely on species-level designations (Barrowclough *et al.*, 2016), it is crucial to continue efforts to comprehend the full breadth of avian species diversity.

Phylogeography of *Schetba*

Our genetic data indicate an initial divergence in *Schetba* between the east and west of Madagascar during the mid-Pleistocene. Although other studies have proposed that the bioclimatic transition between the humid east and dry west of Madagascar may promote population subdivision and speciation (Yoder & Heckman, 2006; Vences *et al.*, 2009), this is the first evidence for this speciation mechanism in birds. Sister species pairs restricted to east and west Madagascar could form via predominantly adaptive processes (i.e. ecogeographic isolation), or via non-adaptive processes (i.e. biogeographic isolation) (Vences *et al.*, 2009). In the case of *Schetba*, the two species differ in ecological niche based on our models (Figure 4), therefore adaptive processes most likely played a role in their divergence. These results fit the hypothesis for ecogeographic isolation, with sister taxa in east and west Madagascar that differ in ecological niche. The divergence of *Schetba* does not appear to be consistent with the 'Western rainforest refugia' speciation mechanism put forward by Vences *et al*. (2009), given that the two taxa have significantly dissimilar environmental niches. Interestingly, based on its distribution, *S. occidentalis* does not appear to be a strictly dry-adapted species. It

occupies subhumid and deciduous forests, and is not found in the arid spiny bush habitat. For example, there is a population of *S. occidentalis* in the high elevation areas of the subhumid forest of Analavelona (Figure 1), and the species is not found in the non-forested area surrounding the massif. The flora of the Analavelona region shares characteristics of the mid-altitude forests of the east, and has been considered a Pleistocene relict when portions of southwestern Madagascar was wetter than today (Goodman *et al.*, 2018).

Biogeographic isolation may have also played a role in the divergence of *S. occidentalis* and *S. rufa*. Our ecological niche models for *Schetba* recovered two distinct areas of suitable habitat in east and west Madagascar, separated by a large expanse of unfavorable habitat in central Madagascar. The natural forest habitats of the Central Highlands have been degraded over hundreds of years (Green & Sussman, 1990; Gade, 1996), but during the Pleistocene this region may have consisted of mosaic habitat of wooded savannah and closed canopy forests (Yoder *et al.*, 2016). Whether this region has been a biogeographic barrier to *Schetba* dispersal over the past 854,000 years is unclear. *Schetba* has a broad elevational range (0 – 1829 m, (Goodman & Raherilalao, 2013)) and, hence, in principal could disperse across these highlands given the necessary ecological conditions. On the other hand, both species are strictly closed canopy forest dependent and found in large tracts of relatively undisturbed forest habitat (Schulenberg, 2013), therefore wooded savannah habitat may have acted as a biogeographic barrier to dispersal.

Overall, especially given our low number of occurrence records for ecological niche modeling, we cannot say conclusively whether ecogeographic or biogeographic isolation was the predominant cause of speciation, and it may be the case that both ecological and biogeographic mechanisms played a significant role in generating and maintain these species.

Concluding remarks

Madagascar has been considered a model region for species diversification studies, yet the phylogeography and diversification processes of the island's avifauna are still largely unexamined. Here we provide the first evidence that the bioclimatic transition between the humid east and dry west of Madagascar has facilitated speciation within birds. More

importantly, our findings of unrecognized diversity within *Schetba*, and cryptic diversity within *Newtonia* (Younger *et al.*, 2018), suggest there may be other species awaiting recognition in this biodiversity hotspot. Appreciating the full spectrum of diversity is likely to alter conservation priorities for Madagascar, and we urge that further studies are needed to quantify the island's biodiversity before it is lost to deforestation.

Acknowledgements

This study was funded by NSF grant DEB-1457624 awarded to SR. Funding was also provided by the Pritzker Laboratory for Molecular Systematics and Evolution, operated with support from the Pritzker Foundation. We gratefully acknowledge the Field Museum of Natural History, the American Museum of Natural History, and the Mention Zoologie et Biologie Animale at the Université d'Antananarivo for access to specimens and tissue samples. We are thankful to Robert Lauer for his assistance preparing sampling maps, and to Chris Kyriazis and Dylan Maddox for their molecular laboratory work.

Data Accessibility

The Illumina short reads are available from the NCBI sequence read archive, link_TBA and Sanger sequences are available from GenBank link_TBA.

Author Contributions

JY collected, analyzed, and interpreted the data, wrote the manuscript, and participated in conceiving and designing the study. PD carried out phylogenetic analyses, AN conducted ecological niche modeling, TOH collected the morphometric data. MJR collected genetic samples. SMG collected genetic samples, and participated in interpreting the data and conceiving the study. SR conceived and designed the study, and carried out morphometric analyses.

References

Andreone, F., Vences, M., Guarino, F.M., Glaw, F. & Randrianirina, J.E. (2002) Natural history and larval morphology of Boophis occidentalis (Anura: Mantellidae: Boophinae) provide new insights into the phylogeny and adaptive radiation of endemic Malagasy frogs. *Journal of Zoology*, **257**, 425-438.

Baldwin, S.P., Oberholser, H.C. & Worley, L.G. (1931) Measurements of birds. *Scientific Publications of the Cleveland Museum of Natural History*, **2**, 1-165.

Barrowclough, G.F., Cracraft, J., Klicka, J. & Zink, R.M. (2016) How many kinds of birds are there and why does it matter? *PLoS One*, **11**, e0166307.

Benestan, L.M., Ferchaud, A.L., Hohenlohe, P.A., Garner, B.A., Naylor, G.J., Baums, I.B., Schwartz, M.K., Kelley, J.L. & Luikart, G. (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, **25**, 2967–2977.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brown, J.L., Cameron, A., Yoder, A.D. & Vences, M. (2014) A necessarily complex model to explain the biogeography of the amphibians and reptiles of Madagascar. *Nature Communications*, **5,** 5046.

D'Amen, M., Zimmermann, N.E. & Pearman, P.B. (2013) Conservation of phylogeographic lineages under climate change. *Global Ecology and Biogeography*, **22**, 93-104.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T. & Sherry, S.T. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

de Wit, M.J. (2003) Madagascar: heads it's a continent, tails it's an island. *Annual Review of Earth and Planetary Sciences*, **31**, 213-248.

Delacour, J. (1931) Description de neuf oiseaux nouveax de Madagascar. *Oiseau et Revue Française d'Ornithologie*, pp. 473-486. Société ornithologique de France, Paris, France.

Delacour, J. (1932) Les Oiseaux de la mission zoologique franco-anglo-americaine a Madagascar. *L'Diseauet la Revue Francaise D'Ornithologie*, **2**, 1-96.

Earl, D.A. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Everson, K.M., Soarimalala, V., Goodman, S.M. & Olson, L.E. (2016) Multiple loci and complete taxonomic sampling resolve the phylogeny and biogeographic history of tenrecs (Mammalia: Tenrecidae) and reveal higher speciation rates in Madagascar's humid forests. *Systematic Biology*, **65**, 890-909.

Everson, K.M., Hildebrandt, K.B., Goodman, S.M. & Olson, L.E. (2018) Caught in the act: Incipient speciation across a latitudinal gradient in a semifossorial mammal from Madagascar, the mole tenrec Oryzorictes hova (Tenrecidae). *Molecular Phylogenetics and Evolution*, **126**, 74-84.

Faircloth, B.C. (2013) illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming.
Faircloth, B.C. (2015) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786-788.
Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717-726.
Florio, A., Ingram, C., Rakotondravony, H., Louis, E. & Raxworthy, C. (2012) Detecting cryptic speciation in the widespread and morphologically conservative carpet chameleon (*Furcifer lateralis*) of Madagascar. *Journal of Evolutionary Biology*, **25**, 1399-1414.
Fuchs, J., Pons, J.-M., Pasquet, E., Raherilalao, M. J. & Goodman, S. M. (2007) Geographical structure of the genetic variation in the Malagasy Scops-owl (*Otus rutilus*) inferred from mitochondrial sequence data. *The Condor*, **109**, 409-18.
Funk, W.C., McKay, J.K., Hohenlohe, P.A. & Allendorf, F.W. (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**, 489-96.
Gade, D.W. (1996) Deforestation and its effects in highland Madagascar. *Mountain Research and Development*, 101-116.
Gatesy, J. & Springer, M.S. (2014) Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, **80**, 231-266.
Gautier, L. & Goodman, S. (2003) Introduction to the flora of Madagascar. *The Natural History of Madagascar* (ed. by S.M. Goodman and J.P. Benstead). University of Chicago Press, Chicago, United States.
Gilbert, K.J., Andrew, R.L., Bock, D.G., Franklin, M.T., Kane, N.C., Moore, J.S., Moyers, B.T., Renaut, S., Rennison, D.J. & Veen, T. (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Molecular Ecology*, **21**, 4925-4930.
Goodman, S.M. & Raherilalao, M.J. (2013) Birds of the class Aves. *Atlas of Selected Land Vertebrates of Madagascar* (ed. by S.M. Goodman and M.J. Raherilalao). Association Vahatra, Antananarivo, Madagascar.
Goodman, S.M., Raherilalao, M.J. & Wohlhauser, S. (eds) (2018) *The terrestrial protected areas of Madagascar: Their history, description, and biota.* Association Vahatra, Antananarivo, Madagascar.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. & Zeng, Q. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-652.
Green, G.M. & Sussman, R.W. (1990) Deforestation history of the eastern rain forests of Madagascar from satellite images. *Science*, **248**, 212-215.
Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C. & Brumfield, R.T. (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, **65**, 910-924.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.
Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. & Kimball, R.T. (2016) Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, **33**, 1110-1125.

Hosner, P.A., Campillo, L.C., Andersen, M.J., Sánchez-González, L.A., Oliveros, C.H., Urriza, R.C. & Moyle, R.G. (2018) An integrative species delimitation approach reveals fine-scale endemism and substantial unrecognized avian diversity in the Philippine Archipelago. *Conservation Genetics*, **19**, 1153-1168.

Hotaling, S., Foley, M.E., Lawrence, N.M., Bocanegra, J., Blanco, M.B., Rasoloarison, R., Kappeler, P.M., Barrett, M.A., Yoder, A.D. & Weisrock, D.W. (2016) Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology*, **25**, 2029-2045.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I. & Andrew, R.L. (2017) The K= 2 conundrum. *Molecular Ecology*, **26**, 3594-3602.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070-3071.

Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jønsson, K.A., Fabre, P.-H., Fritz, S.A., Etienne, R.S., Ricklefs, R.E., Jørgensen, T.B., Fjeldså, J., Rahbek, C., Ericson, P.G. & Woog, F. (2012) Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences*, **109**, 6620-6625.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772-780.

Katoh, K., Misawa, K., Kuma, K.i. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059-3066.

Köhler, J., Glaw, F. & Vences, M. (2007) A new green treefrog, genus Boophis Tschudi 1838 (Anura Mantellidae), from arid western Madagascar: phylogenetic relationships and biogeographic implications. *Tropical Zoology*, **20**, 215-227.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, **14**, 82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772-773.

Lees, D., Kremen, C., Raharitsimba, T., Goodman, S. & Benstead, J. (2003) Classification, diversity, and endemism of the butterflies (Papilionoidea and Hesperioidea): a revised species checklist. *The Natural History of Madagascar* (ed. by S.M. Goodman and J.P. Benstead). University of Chicago Press, Chicago, United States.

Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology, 21*(21), 1838-1844.

727 Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler
728 transform. *Bioinformatics*, **25**, 1754-1760.
729 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. &
730 Durbin, R. (2009) The sequence alignment/map format and SAMtools.
731 *Bioinformatics*, **25**, 2078-2079.
732 Linnaeus, C. (1766) *Systema naturæ per regna tria naturæ, secundum classes, ordines,*
733 *genera, species, cum characteribus, differentiis, synonymis, locis. 12th Edition,*
734 *Volume 1.* Salvius, Stockholm, Sweden.
735 Lischer, H. & Excoffier, L. (2012) PGDSpider: an automated data conversion tool for
736 connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298-
737 299.
738 McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C. & Brumfield, R.T.
739 (2013) A phylogeny of birds based on over 1,500 loci collected by target enrichment
740 and high-throughput sequencing. *PLoS One*, **8**, e54848.
741 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
742 Altshuler, D., Gabriel, S. & Daly, M. (2010) The Genome Analysis Toolkit: a
743 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
744 *Research*, **20**, 1297-1303.
745 Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T. & Braun, E.L. (2016) Analysis of a
746 rapid evolutionary radiation using ultraconserved elements: evidence for a bias in
747 some multispecies coalescent methods. *Systematic Biology*, **65**, 612-627.
748 Mirarab, S. & Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with
749 many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, 44-52.
750 Nussbaum, R. & Raxworthy, C. (1994) The genus Paragehyra (Reptilia: Sauria: Gekkonidae) in
751 southern Madagascar. *Journal of Zoology*, **232**, 37-59.
752 Nussbaum, R.A. & Raxworthy, C.J. (1998) Revision of the genus Ebenavia Boettger (Reptilia:
753 Squamata: Gekkonidae). *Herpetologica*, 18-34.
754 Orozco-Terwengel, P., Nagy, Z.T., Vieites, D.R., Vences, M. & Louis Jr, E. (2008)
755 Phylogeography and phylogenetic relationships of Malagasy tree and ground boas.
756 *Biological journal of the Linnean Society*, **95**, 640-652.
757 Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M. & Stamatakis, A. (2010)
758 How many bootstrap replicates are necessary? *Journal of Computational Biology*, **17**,
759 337-354.
760 Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species
761 geographic distributions. *Ecological Modelling*, **190**, 231-259.
762 Rambaut, A. & Drummond, A.J. (2007) Tracer v1. 4.
763 Raxworthy, C.J., Ingram, C.M., Rabibisoa, N. & Pearson, R.G. (2007) Applications of
764 ecological niche modeling for species delimitation: a review and empirical evaluation
765 using day geckos (Phelsuma) from Madagascar. *Systematic Biology*, **56**, 907-923.
766 Reddy, S. (2014) What's missing from avian global diversification analyses? *Molecular*
767 *Phylogenetics and Evolution*, **77**, 159-165.
768 Reddy, S., Driskell, A., Rabosky, D., Hackett, S. & Schulenberg, T. (2012) Diversification and
769 the adaptive radiation of the vangas of Madagascar. *Proceedings of the Royal Society*
770 *B: Biological Sciences*, **279**, 2062-2071.
771 Rosenberg, N.A. (2004) DISTRUCT: a program for the graphical display of population
772 structure. *Molecular Ecology Notes*, **4**, 137-138.

Schulenberg, T.S. (2013) Schetba. *The Birds of Africa. Volume VIII: The Malagasy Region.* (ed. by R. Safford and F. Hawkins), pp. 828-830. Christopher Helm, London, UK.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312-1313.
Storey, M., Mahoney, J.J., Saunders, A.D. & Duncan, R.A. (1995) Timing of hot spot-related volcanism and the breakup of Madagascar and India. *Science*, **267**, 852.
Tagliacollo, V.A. & Lanfear, R. (2018) Estimating improved partitioning schemes for UltraConserved Elements (UCEs). *Molecular Biology and Evolution*, msy069-msy069.
Vences, M. & Glaw, F. (2002) Molecular phylogeography of Boophis tephraeomystax: a test case for east-west vicariance in Malagasy anurans. *Spixiana*, **25**, 79-84.
Vences, M., Glaw, F., Jesu, R. & Schimmenti, G. (2000) A new species of Heterixalus (Amphibia: Hyperoliidae) from western Madagascar. *African Zoology*, **35**, 269-276.
Vences, M., Wollenberg, K.C., Vieites, D.R. & Lees, D.C. (2009) Madagascar as a model region of species diversification. *Trends in Ecology and Evolution*, **24**, 456-465.
Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T.F. & Achard, F. (2018) Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation*, **222**, 189-197.
Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution,* **62**, 2868-2883.
Weir, J.T. & Schluter, D. (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321-2328.
Wilmé, L., Goodman, S.M. & Ganzhorn, J.U. (2006) Biogeographic evolution of Madagascar's microendemic biota. *Science*, **312**, 1063-1065.
Xi, Z., Liu, L. & Davis, C.C. (2015) Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution*, **92**, 63-71.
Yamagishi, S., Honda, M., Eguchi, K. & Thorstrom, R. (2001) Extreme endemic radiation of the Malagasy Vangas (Aves: Passeriformes). *Journal of Molecular Evolution*, **53**, 39-46.
Yoder, A.D. & Heckman, K.L. (2006) Mouse lemur phylogeography revises a model of ecogeographic constraint in Madagascar. *Primate Biogeography: Progress and Prospects* (ed. by J. Fleagle and S.M. Lehman), pp. 255-268. Kluwer Press, Dordrecht, Netherlands.
Yoder, A.D., Campbell, C.R., Blanco, M.B., dos Reis, M., Ganzhorn, J.U., Goodman, S.M., Hunnicutt, K.E., Larsen, P.A., Kappeler, P.M. & Rasoloarison, R.M. (2016) Geogenetic patterns in mouse lemurs (genus Microcebus) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences*, **113**, 8049-8056.
Younger, J.L., Strozier, L., Maddox, J.D., Nyári, Á.S., Bonfitto, M.T., Raherilalao, M.J., Goodman, S.M. & Reddy, S. (2018) Hidden diversity of forest birds in Madagascar revealed using integrative taxonomy. *Molecular phylogenetics and evolution*, **124**, 16-26.

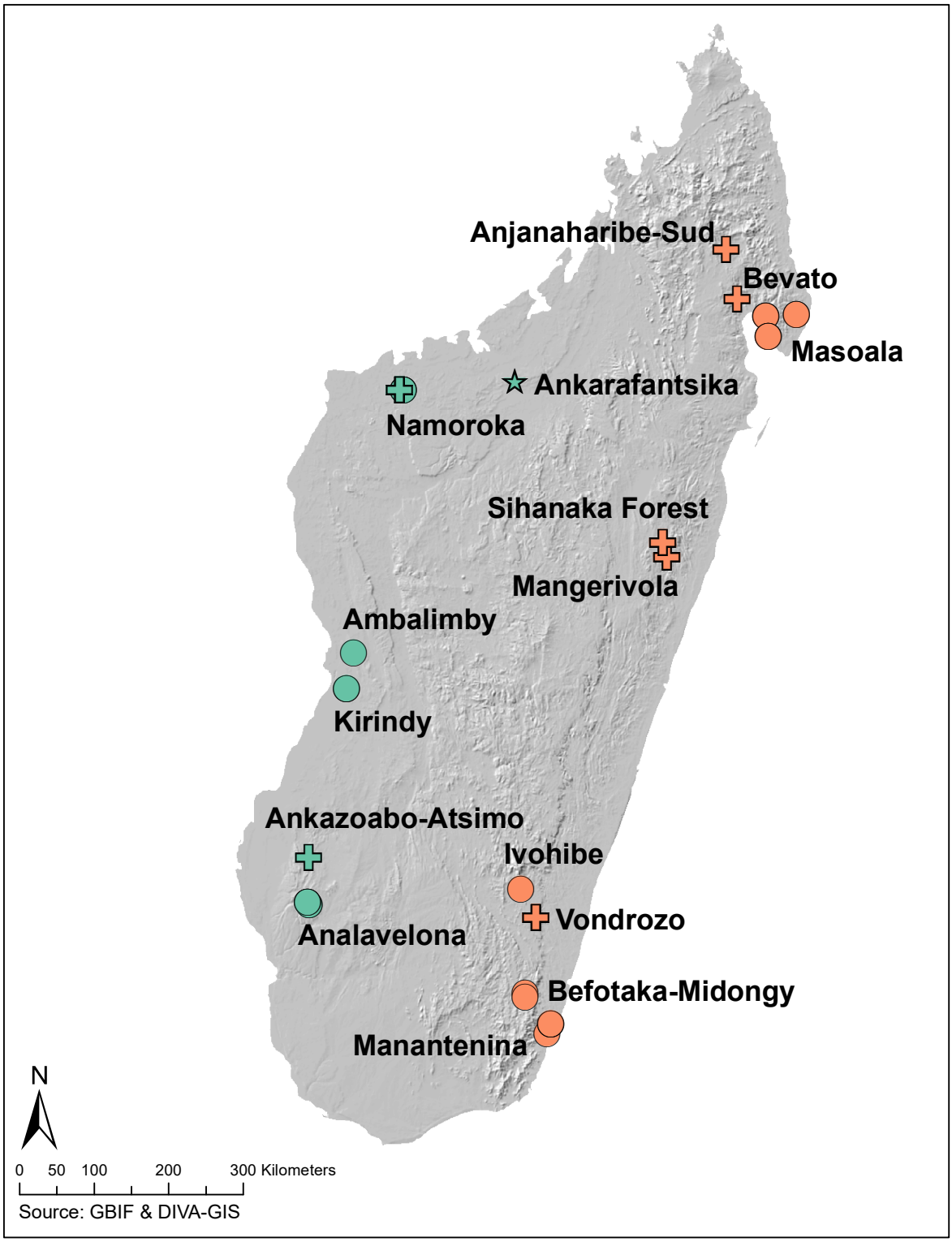


Figure 1. Map of study sampling sites, with *S. r. rufa* indicated by orange icons, and *S. r. occidentalis* by green icons. Green star indicates western population outside the documented range of *S. r. occidentalis*, but confirmed as *occidentalis* in this study. Circles indicate genetic sampling, crosses indicate morphological sampling only. See Table S1 for latitude/longitude and accession numbers.

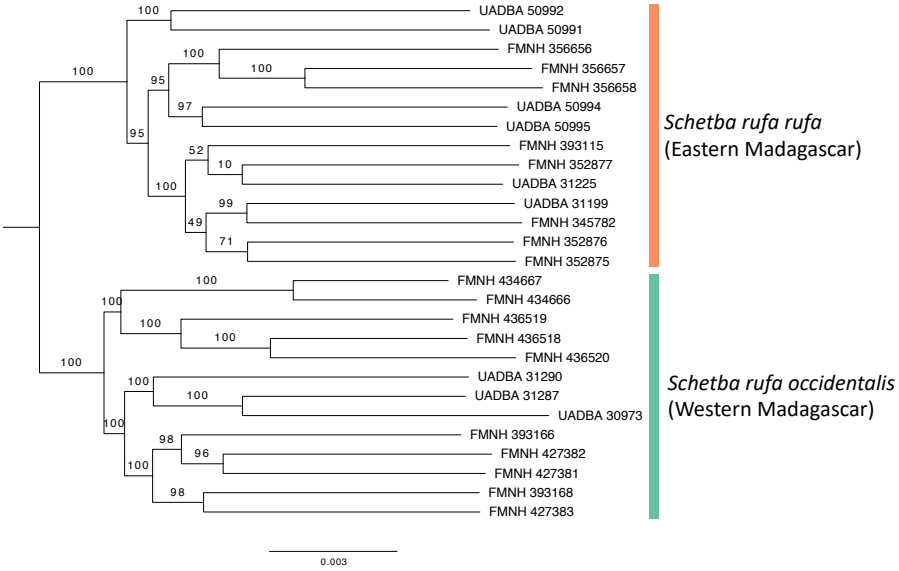
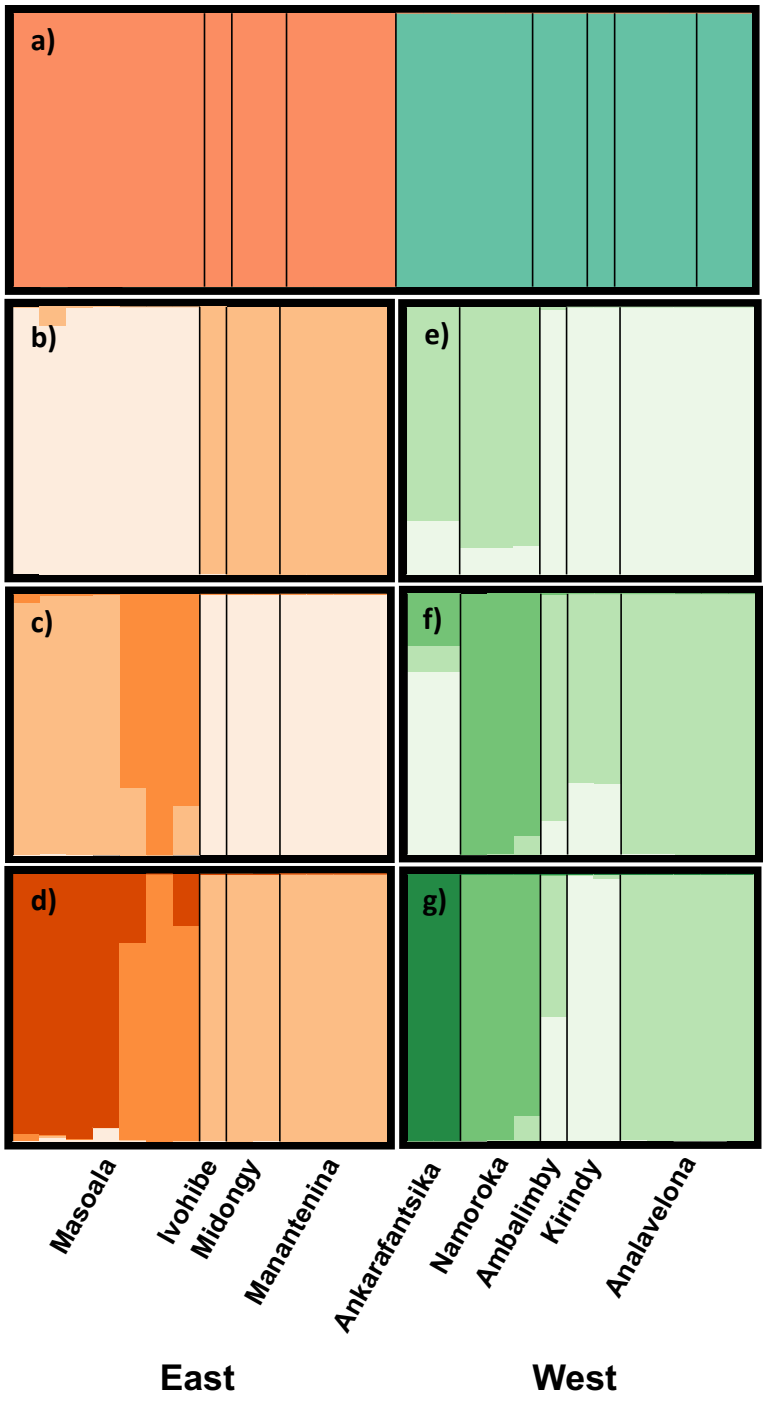


Figure 2. Phylogenetic relationships within *Schetba*. Partitioned maximum-likelihood phylogeny of 4,243 concatenated UCE loci (3,328,172 bp). Support values are shown for nodes that received >70% bootstrap support.



828
829 **Figure 3. Structure plot showing the membership coefficients for *Schetba* individuals to**
830 **genetic clusters.** (a) All *Schetba*, assigned to two genetic clusters ($K = 2$). All individuals from
831 eastern Madagascar have 100% assignment to the orange cluster, whereas all western
832 Madagascar individuals have 100% assignment to the green cluster. Panels (b) – (d);
833 assignment of *S. r. rufa* and *S. r. occidentalis* individuals for $K = 2$ – $K = 4$. Labels refer to the
834 area of collection of individuals.

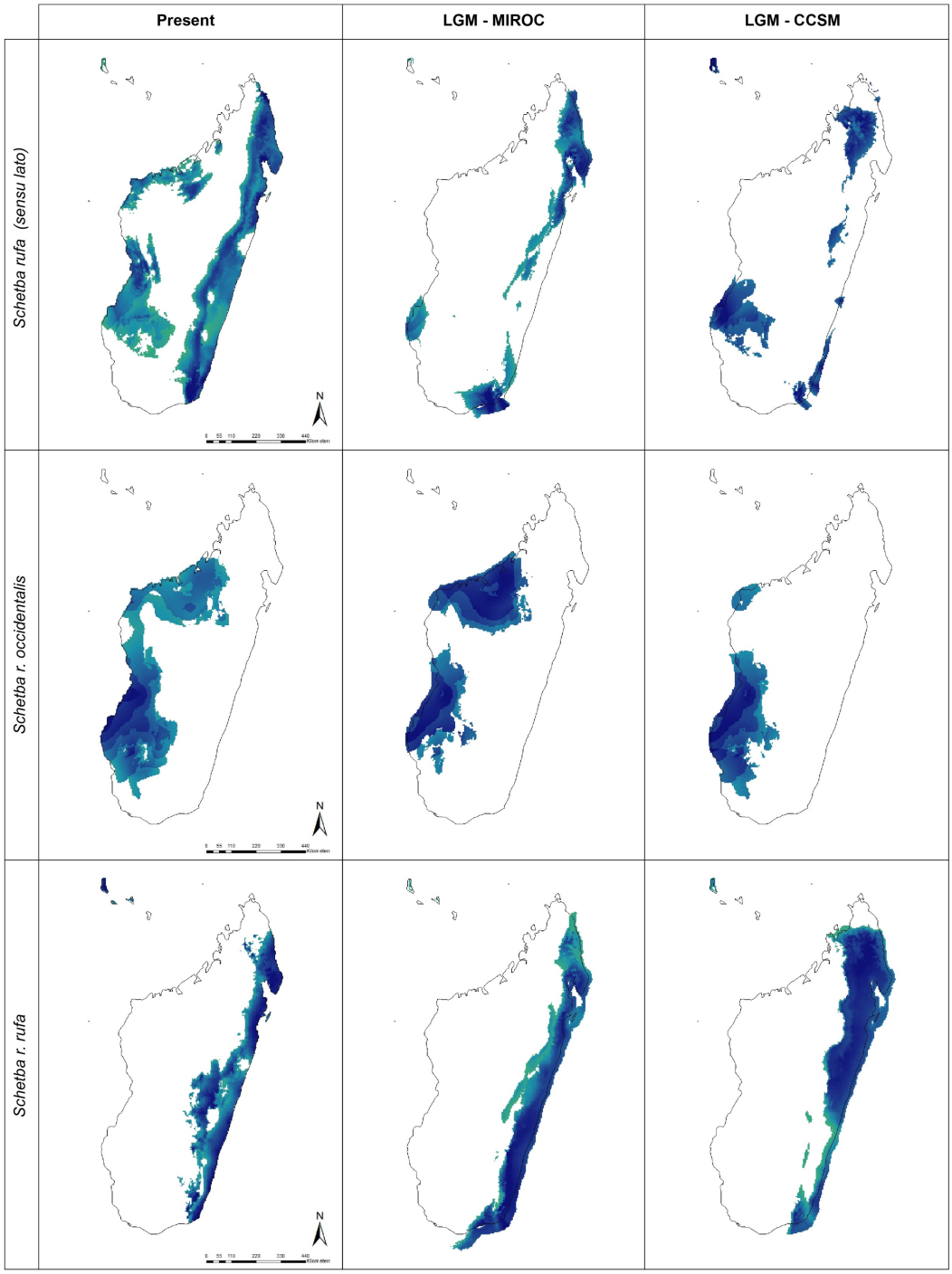


Figure 4. Ecological niche models for *Schetba*, demonstrating suitable habitat in the present and at the Last Glacial Maximum based on two alternate climate scenarios (MIROC and CCSM). Top row represents the best model (AUC = 0.712; omission error = 0.286) for *Schetba rufa* (*sensu lato*) obtained from the pooled set of occurrences (N = 34). The middle row corresponds to the best model (AUC = 0.891; omission error = 0.000) for *S. r. occidentalis* (N = 16), while the bottom row shows the best model output (AUC = 0.925; omission error = 0.000) for *S. r. rufa* (N = 14). Dark blue areas represent higher occurrence probability, while light blue and turquoise indicates lower presence probability.

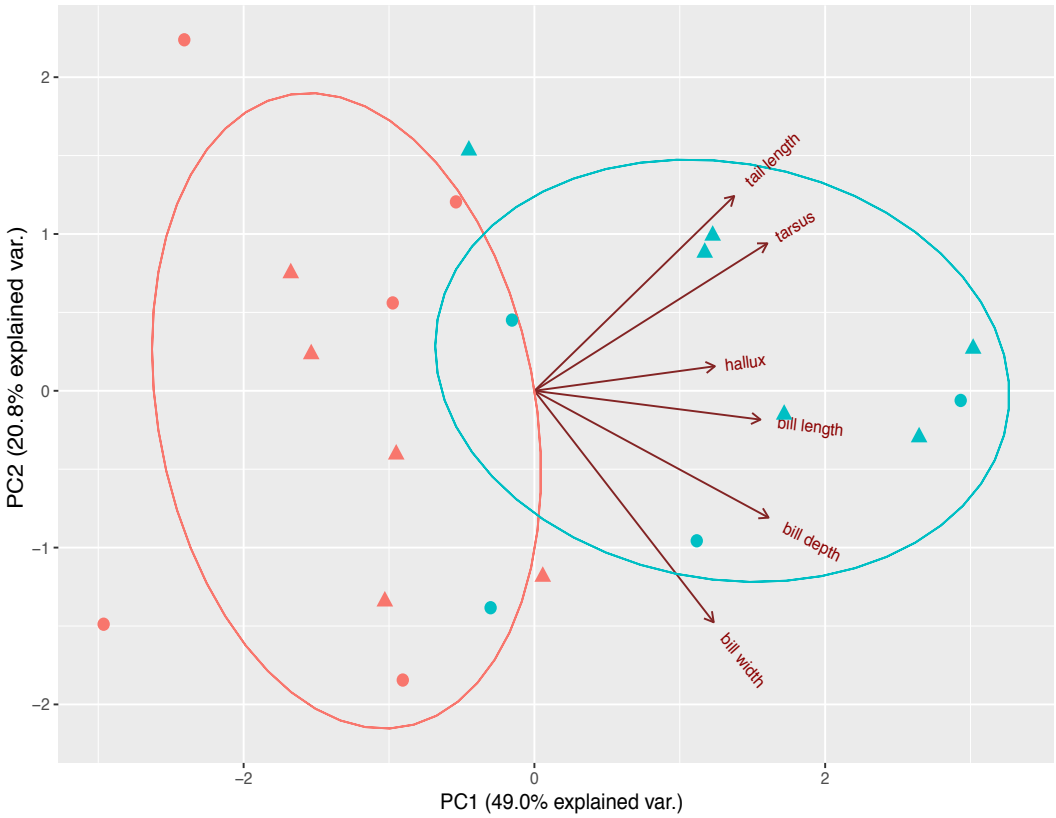


Figure 5. Principal components analysis of morphometric comparisons across *Schetba*. Biplot of PC1 versus PC2, which together explain ~70% of the variation. Centroids of each clade (orange = *S. r. rufa*; green = *S. r. occidentalis*) were significantly different ($p < 0.001$) according to a MANOVA. Circles indicate 95% confidence ellipses around the centroid of each clade; symbols indicate sex (dots = females; triangles = males). Since there was no significant difference between sex, all individuals of each clade were analyzed together. Arrowed lines show direction and magnitude of the coefficients of each variable (abbreviations in text).