AFRICAN JOURNAL OF HERPETOLOGY
https://doi.org/10.1080/21564574.2018.1509139

Geographic variation in West African *Agama picticauda*: insights from genetics, morphology and ecology

Sneha Krishnan[a], Caleb Ofori-Boateng[b], Matthew K Fujita[c] and Adam D Leaché[d]

[a]Colorado State University, College of Veterinary Medicine and Biomedical Sciences 1350 Center Ave, Fort Collins, CO 80521, Fort Collins, CO, USA 80526, 4257858934; [b]Kwame Nkrumah University of Science and Technology, Wildlife and Range Management Department, Kumasi, Ghana; [c]University of Texas at Arlington, Biology, Arlington, TX, USA 76019; [d]University of Washington Department of Biology, Seattle, WA, USA 98195-1800.

ABSTRACT

West Africa is a region characterised by high species diversity and endemism, making it an ideal region to study the evolution of genetic and phenotypic differentiation. Species in this region often vary in morphology and genetics; however, the specific drivers of these differences have not been clearly identified. Focusing on populations in Ghana, we tested for correlations between geography, genetic diversity, morphology and ecology in the West African rainbow lizard (*Agama picticauda*). Genetic data were collected from 102 specimens using double-digest restriction site-associated DNA sequencing (ddRADseq) and the final data matrix included 5 976 loci and 20 624 single nucleotide polymorphisms (SNPs). Morphological data were collected from 42 specimens and included 6 meristic and 4 mensural characters. Ecological data were obtained from the WorldClim database to a resolution of 30 arcseconds. Population structure analyses supported up to five distinct populations of *A. picticauda* in Ghana. Discriminant function analyses were used to classify samples using the morphological data and ecological data. Ecology showed the strongest correlation with population genetic structure, whereas morphological data were only able to weakly differentiate three populations. We discuss the factors that might be responsible for correlations between phenotypes and genotypes in the context of *A. picticauda* natural history.

ARTICLE HISTORY
Received 26 July 2017
Revised 31 July 2018
Accepted 3 August 2018

KEYWORDS
Agamidae; *Agama agama*; ddRADseq; Ghana; morphometrics; niche; population structure; single nucleotide polymorphism

Introduction

Understanding the processes that drive patterns of geographic variation is an integral part of evolutionary biology. In the face of global change and ensuing modifications of biodiversity patterns, research on species distribution is a prime focus in ecology and conservation (Brncic et al. 2015). A wide range of factors, such as environment and geography, contribute to evolutionary processes that could alter the adaptive potential of a population, such as gene flow, genetic drift and selection (Manel et al. 2003). Landscape genetics aims to identify the processes responsible for generating patterns of population

CONTACT Sneha Krishnan snehakrishnan4@gmail.com

African Journal of Herpetology is co-published by NISC Pty (Ltd) and Informa Limited (trading as Taylor & Francis Group)

Published online 07 Mar 2019

genetic differentiation and diversity and how landscape features shape gene flow and genetic connectivity (Storfer et al. 2010).

Historical, geographic and environmental effects can contribute to patterns of geographic variation among populations (Manthey and Moyle 2015). Therefore, it is necessary to thoroughly sample across the genomic, environmental, morphological and spatial landscapes. A variety of statistical methods are used to evaluate landscape variables and the roles they play in shaping population structure (Storfer et al. 2007). An integrative approach is necessary to understand spatial patterns of variation.

The West African region is characterised by high biodiversity and endemicity. The savannah and tropical forest biomes of Africa are highly dynamic, creating a perfect setting to study the evolution and diversification of species (Kissling et al. 2016). The biological and ecological diversity of Ghana is a good representation of the biodiversity seen across West Africa. Ghana exhibits a distinct ecological gradient, with a large savannah region in the north and a deciduous forest in the south. With these differences in ecology, geography and climate between the northern and southern regions, Ghana is an ideal region to study the drivers of geographic variation within a species.

In this study, we couple phylogeographic data with niche characterization and morphometrics to evaluate geographic variation in the West African rainbow lizard, *Agama picticauda*. These diurnal insectivores have a broad distribution across West and Central Africa (Leaché et al. 2017) and are distributed across a wide range of habitats in Ghana, including the savannah regions, open areas in semi-deciduous forests, margins of forests and even in urban settings (Mediannikov et al. 2012). There is some debate concerning the validity of the name *Agama picticauda* (Leaché et al. 2014) versus the more familiar name *Agama agama* (Linnaeus, 1758). The lack of type specimens and invalid syntypes is the major source of confusion. Wagner et al. (2009) showed that the syntypes of *Agama agama* examined by Linnaeus (1758), which were illustrated by Seba (1734), are not attributable to *A. agama* nor to any of the specimens remaining in the Seba and Uppsala collections and therefore designated a neotype for *Lacerta agama* (Linnaeus, 1758) from northern Cameroon to preserve the stability of *Agama agama* (Linnaeus, 1758) and all subsequent names connected to this taxon. Leaché et al. (2014) proposed to use the available name *Agama picticauda* (Peters, 1877) for the remaining populations found across West Africa. Mediannikov et al. (2012) argued for the continued use of *A. agama* for all populations across West and Central Africa, believing that Linnaeus (1758) had access to additional specimens other than those illustrated by Seba (1734). In this study, we follow the definition of *A. agama* according to Wagner et al. (2009) and continue to use *A. picticauda* for populations in West Africa.

We tested the relationships between genetic distance, geographic distance, ecological variability and morphological variability in *Agama picticauda*. Genetic data was collected and processed to assess population relationships and genetic distances. We sample single nucleotide polymorphisms (SNPs) from throughout the genome to quantify genetic diversity within and among populations. Morphological data included measurements that are relevant to the biology and physiology of the species. Ecological data included 19 bioclimatic variables, which we use as a proxy for ecological niche. Given the distinct differences in climate between the north and south of Ghana, we hypothesised that there would be significant genetic, phenotypic and niche differentiation between populations in these regions.

Materials and methods

Genetic data

We collected ddRADseq (Peterson et al. 2012) for 102 samples from 19 localities (Table 1; Figure 1). The high abundance and genome-wide distribution of SNPs make them a valuable source of genetic variation for studies of population structure and phylogeny (Brumfield et al. 2003; Leaché and Oaks 2017). Genomic DNA was isolated from fresh tissue samples (liver) using QIAGEN DNeasy extraction kits (QIAGEN Inc.). We double-digested 500 ng of genomic DNA for each sample with 20 units each of a rare cutter SbfI (restriction site 5′-CCTGCAGG-3′) and a common cutter MspI (restriction site 5′-CCGG-3′) in a single reaction with the manufacturer recommended buffer (New England Biolabs) for 8 hours at 37 °C. Fragments were purified with Sera-Mag SpeedBeads beads before ligation of barcoded Illumina adaptors onto the fragments. The oligonucleotide sequences used for barcoding and adding Illumina indexes during library preparation are provided in Peterson et al. (2012). The libraries were size-selected (between 415 and 515 bp after accounting for adapter length) on a Blue Pippin Prep size fractionator (Sage Science). The final library amplification used proofreading *Taq* and Illumina's indexed primers. The fragment size distribution and concentration of each pool was determined on an Agilent 2200 TapeStation and qPCR was performed to determine sequenceable library concentrations before multiplexing equimolar amounts of each pool for sequencing on one Illumina HiSeq 2500 lane (50 bp, single-end reads) at the QB3 facility at UC Berkeley.

We processed raw Illumina reads using the program iPyRADv0.6.10 (Eaton 2014; http://github.com/dereneaton/ipyrad). We demultiplexed samples using their unique barcode sequences with no mismatches allowed. Sites with Phred quality scores under 99% (Phred score = 20) were changed into N characters and reads with ≥5 Ns were discarded. Each locus was reduced from 50 to 39 bp after the removal of the 6 bp restriction site overhang and the 5 bp barcode. The filtered reads for each sample were clustered

Table 1. Sample localities of Ghanaian *Agama picticauda* included in the analysis.

Region	Locality	Latitude (decimal degrees)	Longitude (decimal degrees)	SNP samples	Morphology
Ashanti	Barekese	6.829	−1.721	2	1
	Kumasi	6.690	−1.569	7	3
Brong Ahafo	Asumura	6.701	−2.761	4	4
	Bui	8.293	−2.287	1	2
Eastern	Mamang	6.253	−1.032	1	0
	New Tafo	6.222	−0.364	4	1
Greater Accra	Accra	5.605	−0.224	4	4
	Dangme	5.841	0.115	8	3
	Shai Hills	5.884	0.038	7	2
Northern	Bimbila	8.872	0.043	6	0
	Buipe	8.768	−1.477	4	4
	Mole	9.295	−1.845	8	0
Upper West	Fian	10.388	−2.465	5	10
	Gbele	10.420	−2.074	1	1
	Wa	10.058	−2.497	2	1
Volta	Kiri	8.265	0.517	12	2
	Kyabobo	8.348	0.601	12	0
	Pawa	8.456	0.567	6	2
Western	Ankasa	5.281	−2.641	8	1

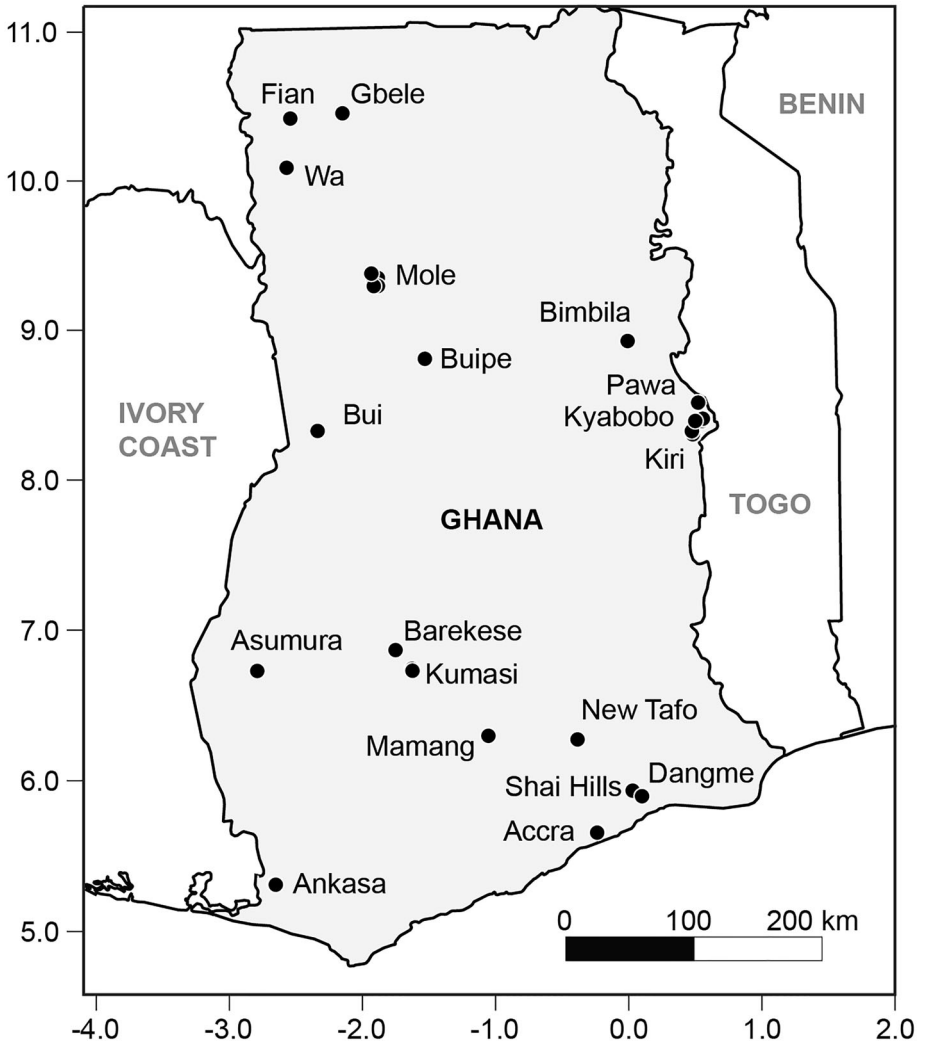

Figure 1. Map of Ghana and *Agama picticauda* populations included in the analysis.

using the program VSEARCH v1.1.0 (Edgar 2010) and then aligned with MUSCLE (Edgar 2004). We used a clustering threshold of 92% to assemble reads. Consensus sequences that had low coverage (<6 reads), excessive undetermined or heterozygous sites (>8), or too many haplotypes (>2 for diploids) were discarded. The consensus sequences were clustered across samples using the same threshold used to cluster data within species (92%). Each locus was aligned with MUSCLE and a paralog filter that removes loci with excessive shared heterozygosity among samples was applied (paralog filter = 0.5). Loci present for >50% of the samples were included in the final matrix.

We explored two methods to estimate population structure from the SNP data. First, we used a clustering approach with the program Adegenet (Jombart 2008) in R. We analysed all samples together (102 total) using one random SNP per RAD locus (5 976 SNPs). We used discriminant analysis of principal components (DAPC) to describe the genetic

diversity among the 19 sample locations, with the expectation that samples collected from the same location would cluster together, reflecting their genetic similarity. We conducted a preliminary analysis using the maximum number of principal components (PCs = 34) and then used the opt.pca function to select the optimal number of PCs to retain in the final analysis (PCs = 20). Second, we estimated population structure using the likelihood-based method STRUCTURE v2.3 (Pritchard et al. 2000). We used the admixture model with correlated allele frequencies without treating sampling locations as *a priori* information (Pritchard et al. 2000, 2002). The program was run with a burn-in of 50 000 iterations, followed by 500 000 MCMC steps. Each value of *K* (number of populations) between 1 and 10 was run 5 times. Replicate runs of each *K* value were combined using the program CLUMPP (Jakobsson and Rosenberg 2007) and visualised with DISTRUCT v1.1 (Rosenberg 2004). To select the optimal value of *K,* we inspected the LnP(*K*) curve for an asymptote and then picked the lowest value of *K* with minimal standard deviation (Pritchard et al. 2000).

Species trees were estimated for *Agama picticauda* using two methods. We used SVDquartets (Chifman and Kubatko 2014), implemented in PAUP v4.0a152 (Swofford 2002), to estimate a species tree using one random SNP from each locus (5 976 loci). We evaluated 100 000 quartets and used a taxon partition that assigned samples to the 19 sample locations. Ambiguous data were treated as missing. We evaluated support for the species tree using bootstrapping with 100 pseudoreplicates. The multispecies coalescent method SNAPP v.1.3.0 (Bryant et al. 2012), implemented in BEAST v2.4.5 (Bouckaert et al. 2014), was used to estimate a species tree for the biallelic SNPs that were shared among all 19 sample locations (= 456 SNPs). As a result of computational constraints, only two random samples per sample locality were included in the data matrix. The mutation rates u and v were set to 1.0 and not sampled. The coalescent rate prior was sampled and set to 10 and the remaining priors were left at default settings. We ran four replicate SNAPP analyses, each for 100 000 generations (sampling interval = 50) and removed the initial 20% of the samples as burn-in (leaving 6,400 trees). The maximum clade credibility (MCC) tree was calculated using TreeAnnotator.

Morphological data

Forty two (42) male and female adult specimens were examined for the morphological analysis. 10 traits were examined for each specimen (Table 2). Measurements were taken with a digital calliper to the nearest 0.01 mm and some measurements utilised a

Table 2. Morphological characters measured for *Agama picticauda.*

	Mensural characters
SVL	Snout-vent length from tip of snout to cloaca
HH	Head height – head height at angle of jaw
HW	Head width – head width at angle of jaw
HL	Head length from tip of snout to angle of jaw
Meristic characters	
LL4	Subdigital lamellae of left 4th toe
RL4	Subdigital lamellae of right 4th toe
SL	Number of supralabial scales
IL	Number of infralabial scales
PR	Number of postrostral scales
MB	Number of scales around midbody

stereomicroscope. Specimens from the following institutions were included in the study: The Museum of Vertebrate Zoology (MVZ), Berkeley, USA; The Burke Museum of Natural History (UWBM), Seattle, USA; The Louisiana Museum of Natural History (LSUMZ), Baton Rouge, USA. For morphological analysis, the snout-vent length (SVL) character was size corrected and log10 transformed by performing a linear regression with the log scaled SVL and using residual values for each mensural character. Using the prcomp function in the biostats package in R, we carried out a multivariate statistical analysis of the data. We used principal component analysis (PCA) to evaluate the 10 morphological characters and plotted the results using the ggbiplot function. We also explored analyses that separated male and females. Euclidean distances among PCA eigenvalues for each specimen were calculated using the data.dist function in R and converted into a matrix. Five of the 19 localities in the study were excluded, because of a lack of specimens, therefore distances were calculated among 14 locations. We also performed classification tests using discriminant function analysis (DFA).

Ecological data

For the ecological analysis, 19 bioclimatic variables were extracted from http://www.worldclim.org with a 30 arcsecond resolution (Hijmans et al. 2005). We used QGIS software to extract the data and generate 19 values for each of the 19 localities. Euclidean distances between the eigenvalues for each bioclimatic variable were calculated using the data.dist function in R and converted into a matrix using the as.matrix function. We tested whether the ecological data could be used to correctly classify sample localities into genetic populations using discriminant function analysis.

Isolation by distance

We tested for isolation by distance (IBD) using Mantel tests in the program Adegenet. We first calculated a matrix of genetic distances (Nei's genetic distances; Nei 1978) using the SNP data and a matrix of geographic distances. Correlation between genetic and geographic distances can result from continuous clines of genetic differentiation, as well as from distant and differentiated populations. Consequently, we tested for IBD using individuals, as well as populations. We visualised IBD using scatterplots of genetic and geographic distances. To visualise the density of points in the IBD graph, we measured local density using a two-dimensional kernel density estimator. A scatterplot with one single consistent cloud of points fits a continuous model of IBD, whereas patchy or discontinuous scatterplots might reflect population differentiation.

Results

Genetics

The ddRADseq protocol provided an average of over 1.8 million sequences for each of the 102 *Agama picticauda* samples (Table 3). The final number of loci produced by the iPyRAD assembly with 50% missing data was 5 976 loci and 20 624 SNPs (Table 3). The

Table 3. Summary of meristic and mensural data.

Species ID	SVL	HH	HL	HW	LL4	RL4	SL	IL	PR	MB
UWBM 5566	105	12.02333	22.78	18.23	16.33333	17.66667	10	10	4	62.66667
UWBM 5564	91	9.69	20.24667	15.38333	19	18.66667	9	9	4	70.66667
UWBM 5556	134	14.8	25.91667	22.14	20	20	10	9	4	72.33333
UWBM 5558	112	12.31667	23.41667	20.26667	18	18	10.33333	9	6	70.66667
UWBM 5555	108	12.18	22.74333	20.01667	16	15.33333	11	9	5	69.33333
UWBM 5552	99	12.13	20.85667	18.05	16.66667	16.66667	9	11	5	71.33333
UWBM 5567	148	16.89667	29.48	26.52667	18.66667	17	9	9.666667	5	78
UWBM 5554	157	29.71333	29.71333	26.09667	19	18	8	10	5.666667	73.33333
UWBM 5551	90	9.976667	19.24333	16.31333	16	19.33333	9	8	5	66.66667
UWBM 5557	109	12.42667	22.23	17.87	15.66667	17	10	9	5	70.66667
LSUMZ 86952	104.85	12.69333	24.06333	20.10333	19	19.66667	10	7.333333	5	61.33333
LSUMZ 86952	127.46	14.61667	27.76667	22.32	22.66667	23	7.666667	8	5	64.66667
LSUMZ 87130	100.97	12.46333	24.75667	19.98	20.66667	20.33333	10	9	6	68.66667
LSUMZ 87120	33.48	5.92	11.37	7.633333	20	16	10	11	5	62.66667
LSUMZ 87213	118.13	15.27333	25.37	22.57667	20.33333	21	8	9	6	70.66667
LSUMZ 86950	102.21	11.16333	21.81333	18.55667	22	21.33333	9	8	5	70.66667
LSUMZ 86842	97.04	12.10333	21.51333	18.52667	13	18	10	9	6	56.66667
LSUMZ 86814	85.03	11.13667	18.95667	17.43333	19	18	8	9	4	62.66667
LSUMZ 87122	83.39	10.53333	18.9	16.49667	18.66667	17.33333	10	8	5	69.33333
LSUMZ 86951	92.92	11.87	21.40333	17.89667	20.66667	19	9	8	5	67.33333
LSUMZ 86872	76.03	10.36667	18.77333	15.47667	16.66667	17	8	7	6	63.33333
LSUMZ 86843	63.92	8.083333	15.73	12.8	17	15.66667	9	8	4	63.33333
LSUMZ 87121	37.61	5.336667	10.5	8.366667	16.66667	16	10	9	5	60.66667
UWBM 5550	89	10.47	20.06	15.98	23	20.33333	10	9	5	72.66667
UWBM 5560	91	10.21333	19.67333	15.8	23	22.33333	9	9	4	80.33333
UWBM 5548	123	13.64333	25.31333	21.88333	19.66667	21	8	8	6	70
UWBM 5551	92	10.40667	19.68333	16.39	16.66667	16.66667	9	9	6	64
UWBM 5559	93	11.07	19.86	17.68333	13.66667	14	10	8	5	80.66667
UWBM 5573	121	13.77	23.80333	20.32	22.33333	22.33333	9	8	5	68
UWBM 5562	120	13.63667	25.20333	20.54333	19	21.33333	9	7	6	64.66667
UWBM 5571	90	9.873333	20.45333	15.7	22.66667	21	10	9	5	72.66667
MVZ 249640	135.2167	15.89333	27.38667	21.01667	20	18.33333	7.333333	6	5	70.666
MVZ 245236	130.2067	15.86333	28.65	20.61667	22	21.33333	7	8	6	56.66667
MVZ 245234	100.2633	18.85	22.16	18.85	21	20	10	9	5	72.66667
MVZ 245259	82.55	11.34	19.24	16.22667	18	17.33333	8	8.33333	5	56.66667
MVZ 249608	122.2067	14.26667	23.80667	19.13667	18.66667	19.33333	8	7	6	64

(*Continued*)

Table 3. Continued.

Species ID	SVL	HH	HL	HW	LL4	RL4	SL	IL	PR	MB
MVZ 245246	135.6733	15.72667	26.20333	20.91333	19.66667	20.33333	7.333333	8	5.666667	66
MVZ 245237	133.4267	15.74667	25.76	21.93333	20.33333	19.66667	10	7	5	65.33333
MVZ 245235	128.49	15.77	25.91333	21.45333	23.66667	21.66667	10	8	5	66
MVZ 245238	109.7433	22.71667	21.66667	17.44333	21.66667	21.66667	9	9	5	65.33333
MVZ 249621	126.5967	16.40333	28.05667	21.68667	22.66667	21.66667	9	7	5	43.33333

demultiplexed sequence reads are available for download on the NCBI Sequence Read Archive (SRA study accession #SRP108438).

The genetic diversity among populations is shown in the DAPC plot (Fig. 2). The samples from Accra, New Tafo, Dangme and Kiri cluster together and also form a group in the SVDquartets tree (Fig. 2). The Pawa and Shai Hills populations are genetically distinct from other populations in the DAPC plot (Fig. 2). The remaining populations overlap and are not well differentiated (Fig. 2). The STRUCTURE results suggest that the LnP(K) curve reaches an asymptote at $K = 5$ and that higher values of K only increase the standard deviation (Fig. 3). The major groups detected at $K = 5$ (depicted in Figs 2 and 4) were are follows: (1) northern, seven localities from Northern Ghana; (2) southwest: four localities from Ankasa National Park to Kumasi; (3) coastal, four localities, including one from near the Togo Hills (Kiri) near the Togo Hills; (4) central, two localities, including Kyabobo National Park and Mamang Forest Resource Reserve (5); Pawa, a single locality near the Togo Hills.

The phylogenetic relationships estimated among sample localities using the coalescent method SNAPP are shown in Fig. 4A. The samples from Mamang, Kumasi and Kyabobo form a weakly supported clade (posterior probability = 0.54) that is sister to the remaining samples. The population in northern Ghana forms a clade (i.e. Bui, Bimbila, Gbele, Fian, Mole, Buipe, Wa), but the interrelationships within this clade are highly uncertain (posterior

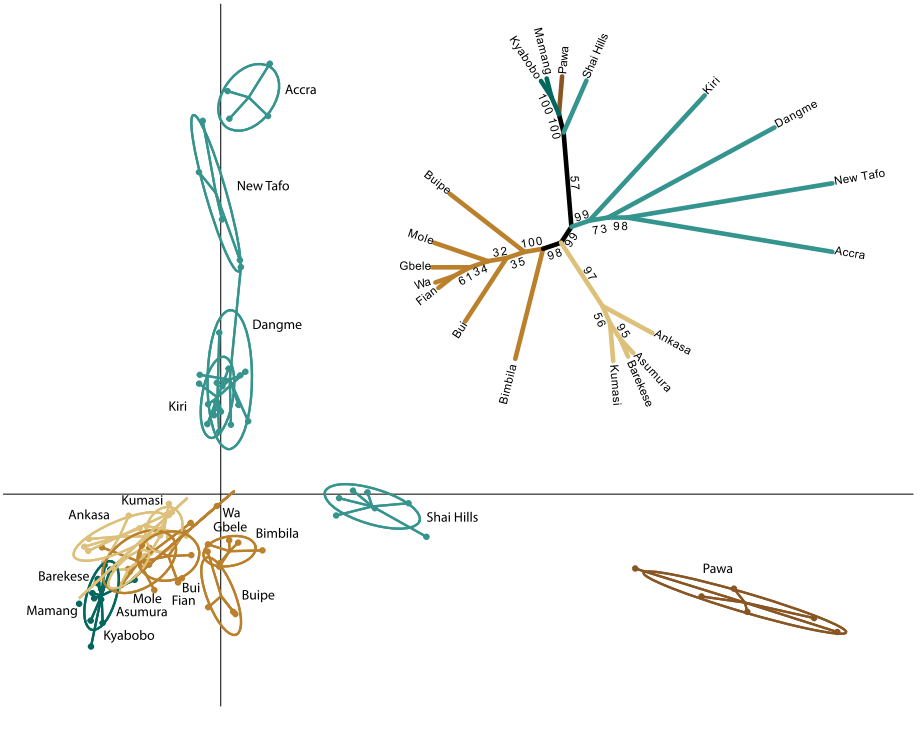

Figure 2. Scatterplot of genetic variation using 5,976 SNPs among *Agama picticauda* in Ghana. The unrooted species tree (estimated using SVDquartets) shows the phylogenetic relationships among the 19 sample locations. Numbers on internal branches are bootstrap support values. Samples are coloured to reflect the population assignment with the highest proportion of membership inferred from STRUCTURE assuming $K = 5$.

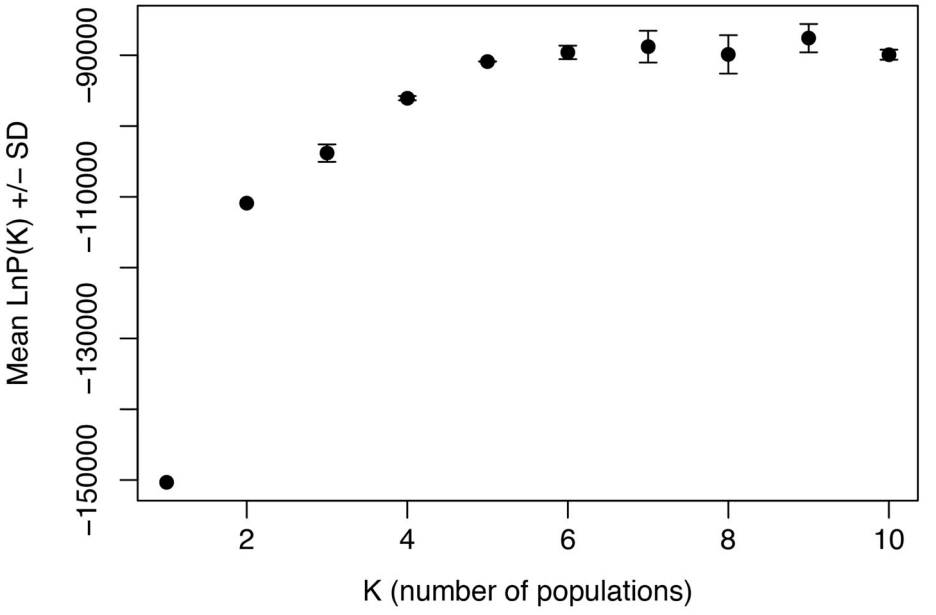

Figure 3. Population structure inference using the program STRUCTURE supports a $K = 5$ model for *Agama picticaudA.* Results are averages (and standard deviations) of five replicate runs for K values ranging from 1 to 10.

probabilities <0.5). There is strong support (posterior probability = 1) for a clade containing Anakasa, Barekese and Asumura and weak support for a southern coastal clade containing Shai Hills, Accra, Dangme, Kiri and New Tafo.

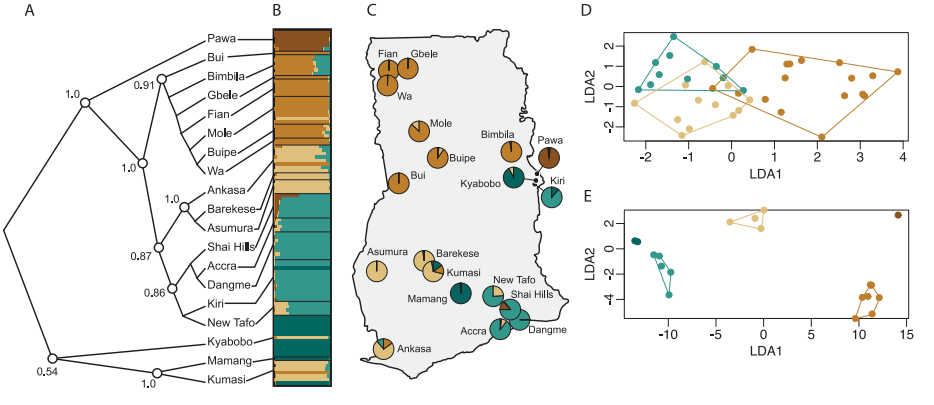

Figure 4. (A) Species tree for *Agama picticauda* estimated using SNAPP with the 456 SNPs shared among the 19 sample locations. Numbers on nodes are Bayesian posterior probability values. (B) Barplot showing the population assignments from STRUCTURE ($K = 5$). (C) Population structure of *A. picticauda* in Ghana. Pie charts show population membership proportions for each of the sampled localities. (D) Discriminant function analysis (DFA) results for the morphological data for the northern, southwestern and coastal populations. The scores for the first and second discriminant functions are shown (LDA1 and LDA2). (E) DFA results for the ecological data, using the first and second discriminant functions (LDA1 and LDA2).

Morphology

The principal component analysis of the morphological data is shown in Fig. 5. The first principal component (PC) explains 28.2% of the variation in the data and the second PC explains 17.3% of the variation. Males and females did not differ significantly in trait values and were therefore not separated for the analysis. There were no apparent groupings among populations. The DFA analysis included samples from the northern, coastal and southwestern populations. Classification success of samples into these three populations was $\leq 65\%$ (Table 4) and samples from the coastal and southwestern populations are largely overlapping (Fig. 4D).

Ecology

The principal component analysis of the ecological data is shown in Fig. 6. The first principal component (PC1) explains 69.8% of the variation and PC2 explains 25.7% of the variation. The PC loading values are provided in Table 5 for the BioClim variables that had the largest values. Temperature seasonality and precipitation of coldest quarter varied together on PC1. Variables related to precipitation also varied together and annual precipitation had the largest effect (Table 6). The DFA analysis of the climate variables correctly classified all sample localities into their proper genetic population (Table 4; Fig. 4E).

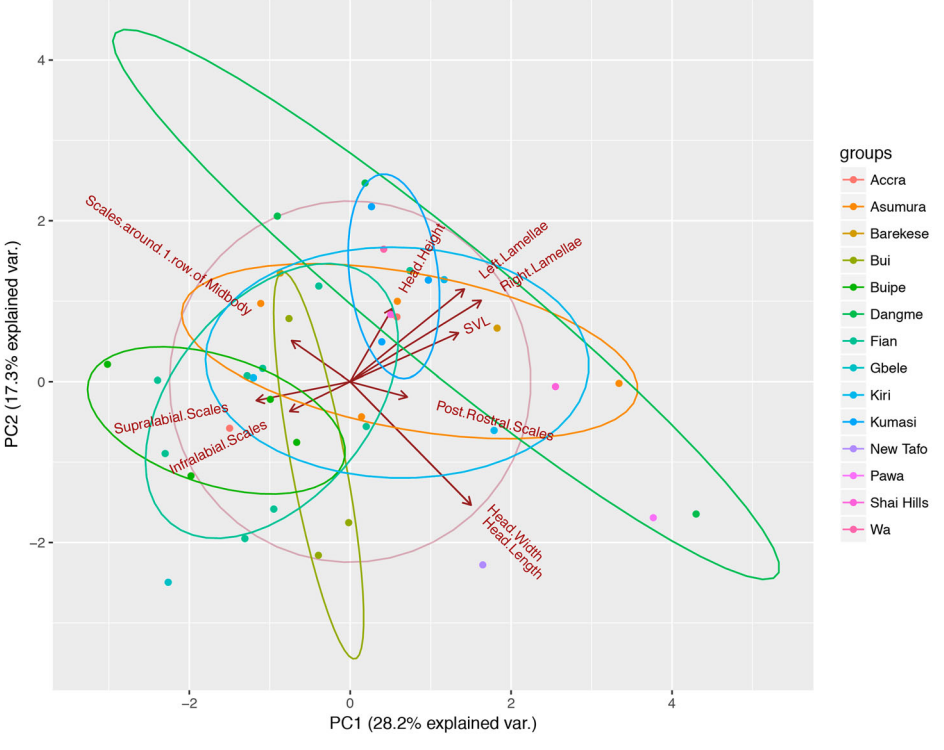

Figure 5. PCA plot of morphological data. Points represent specimens and coloured ellipses represent the variation within populations. The horizontal axis shows the direction of maximum variation (i.e. the first principal component) and the vertical axis represents the second principal component.

Table 4. Summary of the ddRADseq data.

Raw reads[1]	1 802 370
Reads passing filters[1]	1 802 201
Total clusters[1]	98 836
Heterozygosity[1]	0.00837
Error[1]	0.00215
Loci in assembly[1]	4 548
Total loci	5 976
Total SNPs	20 624

[1]Averaged across 102 samples.

Isolation by distance

The SNP data did not support strong correlations between geographic and genetic distance (Fig. 7). A test for IBD that included all samples was insignificant ($p = 0.447$; Fig. 7A). Tests for IBD within populations were also insignificant (southwestern population, $p = 0.449$; coastal population, $p = 0.151$; northern population, $p = 0.069$). The population-level IBD plots shows discontinuous and patchy densities that could reflect additional population structure.

Discussion

The goal of this study was to investigate correlations between genetics, morphology, ecology and geographic distance to identify patterns of variation in *Agama picticauda* in Ghana. The SNP data support fine-scale population structure in Ghana, with as many as $K = 5$ populations. Isolation by distance was not detected for all samples or within any of the more thoroughly sampled populations. There was a strong correlation between population structure and climate variables. The correlation between morphology and population structure was weak.

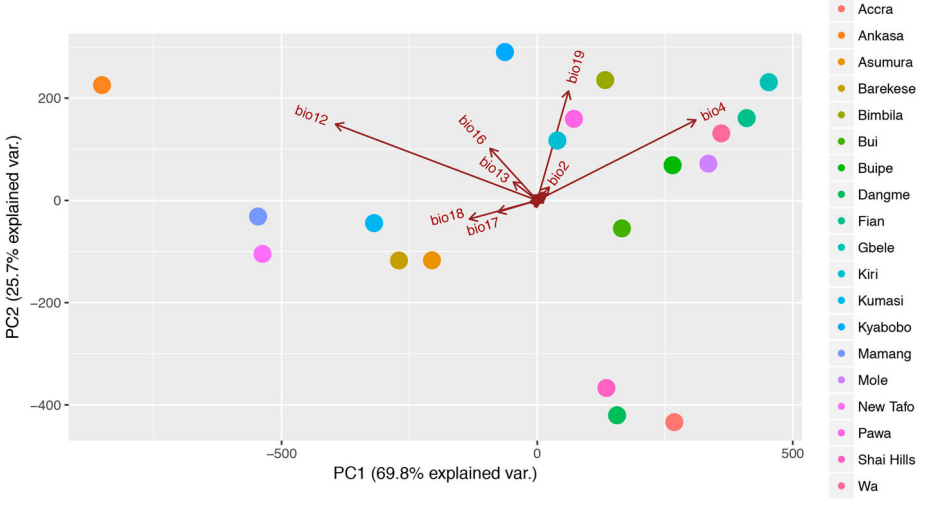

Figure 6. PCA plot of ecological data. Each point represents one of the 19 sample locations. Horizontal axis shows the direction of maximum variation (i.e. the first principal component) and the vertical axis represents the second principal component.

Table 5. Discriminant function analysis classification results.

	Morphology		Climate	
Population	*n*	% correct	*n*	% correct
Coastal	12	50	5	100
Northern	17	65	7	100
Southwest	10	60	4	100
Central	–	–	2	100
Pawa	–	–	1	100

Table 6. Ecological variables with the largest loading values.

Bioclimatic variable	PC1	PC2
BIO4 = Temperature seasonality	0.574	0.479
BIO19 = Precipitation during coldest quarter	0.113	0.652
BIO13 = Precipitation during wettest month	−0.086	0.110
BIO17 = Precipitation during driest quarter	−0.142	−0.069
BIO16 = Precipitation during wettest quarter	−0.171	0.309
bio18 = Precipitation during warmest quarter	−0.246	−0.111
bio12 = Annual precipitation	−0.730	0.454

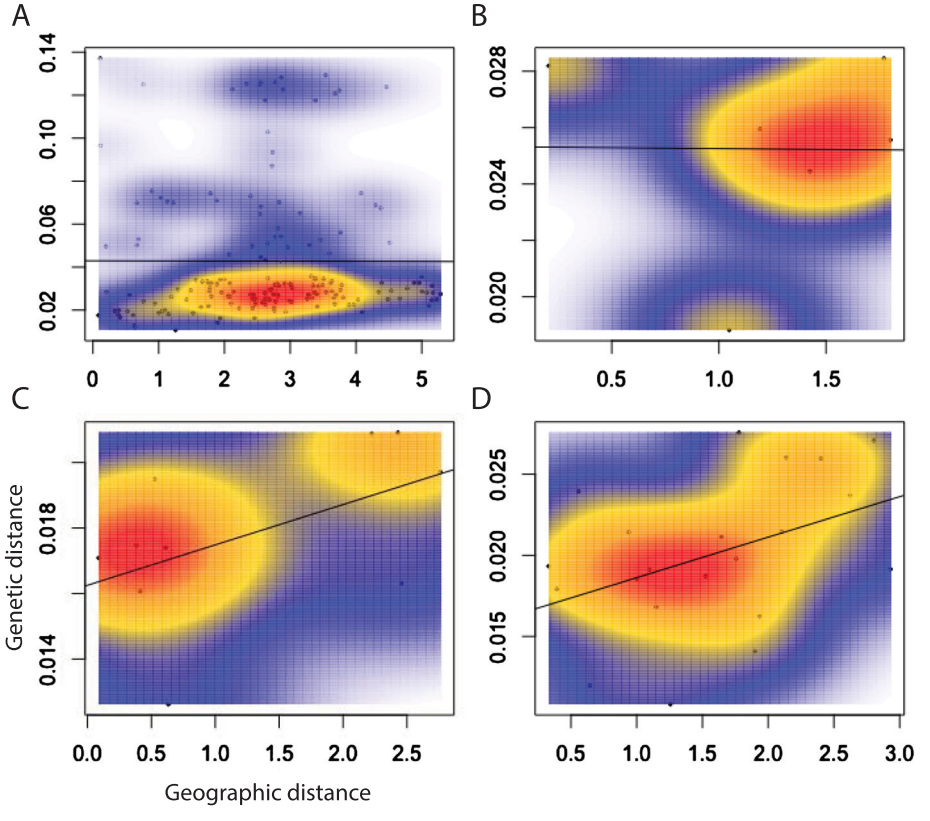

Figure 7. Isolation by distance (IBD) plots showing the relationship between geographic and genetic distance for (A) all 102 *Agama picticauda* samples, (B) the southwestern population, (C) the coastal population and (D) the northern population. Colours represent local sampling density (white, lowest; red, highest).

Genetics

In genetic studies of natural populations, various factors can influence spatial patterns of variation. Genetic diversity is often correlated with geographic distance and populations that are far apart are typically more genetically differentiated compared with geographically proximate populations, resulting in a pattern of isolation by distance. We did not find a significant correlation between genetic distance and geographic distance in *A. picticauda*. One potential factor that can obscure the IBD pattern is gene flow across large distances. *Agama picticauda* are human commensals and highly mobile, with introduced populations as far as Madagascar and North America (Enge et al. 2004). We found some evidence for gene flow across large distances, based on the genetic clustering results (Figs 2 and 4) and the uncertainty in the species tree (Fig. 4). The movement of lizards along with vehicles and goods could result in a pattern of close genetic relatedness across distant localities. Another factor that can obscure the relationship between genetic distance and geography is population structure. We found strong evidence of population structure in the SNP data (Figs 2, 3 and 4). IBD tests within populations failed to detect IBD, further suggesting that gene flow within *A. picticauda* might not be limited to small spatial scales. Distinct population boundaries obscure patterns of IBD, because population boundaries produce large genetic distances across short geographic distances. Large population sizes could also maintain high levels of population similarity and there is genetic evidence that *A. picticauda* experienced a recent and rapid population expansion (Leaché et al. 2017). These population dynamics, namely gene flow, population structure and large population sizes, will influence genetic diversity measures and potentially obscure any correlation between genetic distance and geographic distances.

Morphology

The morphological data analysed here did not provide a clear correlation with population structure. The first two axes of the PCA represent only a small part of the total variation (45.5%), indicating that variation is complex and represented over many dimensions. Accordingly, phenotypic variation in *Agama picticauda* is mostly attributed to individual or within-population variation. It is possible that the 10 morphological characters examined in this study are not relevant to genetic structure and that more characters are necessary to discriminate populations. A morphological analysis of *Agama* species occurring in the Horn of Africa using 67 characters was able to distinguish species, including variation among populations of *A. spinosa* (Wagner et al. 2013). One of the most prominent features of *Agama* lizards is adult male head coloration, which is known to vary across regions (Mediannikov et al. 2012). The head colour of adult males changes gradually from reddish or dark orange in coastal areas with high levels of precipitation to yellow or white in the arid northern savannah regions. We would expect this trait to vary with ecology and geography in the male specimens in the study. However, adult coloration is not preserved in museum specimens and this important trait was not quantified for our morphological analysis.

Another trait important to lizards is scale size. Dorsal scale size has been proposed to vary along climatic gradients, with species in warmer climates exhibiting larger scales to reduce heat load (Oufiero et. al. 2011). Inclusion of this trait in the morphological analysis

might reveal patterns between scale size and bioclimatic variables. Future morphological studies would benefit from sampling adult male lizards and quantifying colour in live specimens, as well as collecting measurements of scale size and shape.

Ecology

We used climate variables as a measure of ecological niche in *Agama picticauda*. We found a strong correlation between population genetic structure and climate variables and the DFA analysis was able to classify populations with 100% accuracy (Table 4). The bioclimatic variables included were related to temperature and precipitation, which vary in substantially in Ghana from the moist semi-deciduous forests in the south to the dry savannah regions in the north. *Agama picticauda* is a generalist species that is commonly found in human-modified landscapes across most habitat types in Ghana (Mediannikov et al. 2012). The only habitat type that *A. picticauda* is not found are closed forests, but they can be found on the forest edge habitats. Our results suggest that *A. picticauda* has a wide climatic niche (Fig. 6) and this agrees with the widespread geographic distribution of the species in both natural and man-made landscapes.

Conclusions

The results of this study provide evidence for population structure in *Agama picticauda* in Ghana using SNP data. The phylogenetic analysis provides strong evidence for relationships among populations in some regions, whereas other population relationships remain obscure, possibly as a result of gene flow. The morphological characters were only able to weakly differentiate several populations of *A. picticauda*. Factors related to temperature and precipitation are known to be related to certain morphological characters, such as adult male head coloration and scale size and should be included in future studies.

Acknowledgements

We thank Jared Grummer and Kevin Epperly for their help with the collection of genetic data using the ddRADseq protocol. For help with analysis of morphological and ecological data, we thank Itzue W Caviedes Solis and Sima Bouzid.

Funding

This project was supported by a Mary Gates Research Scholarship to SK and funds from the National Science Foundation to ADL and MKF (NSF DEB-1456098 and DEB-1457232). This work used the Vincent J Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

References

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLOS Comput Biol. Apr 10;10(4):e1003537.

Brncic T, Amarasekaran B, McKenna A, Mundry R, Kühl HS. 2015. Large mammal diversity and their conservation in the human-dominated land-use mosaic of Sierra Leone. Biodivers Conserv. 24 (10):2417–2438.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol. 18(5):249–256.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol. (8):1917–1932.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30(23):3317–3324.

Eaton DAR. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 30(13):1844–1849.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26 (19):2460–2461

Enge KM, Krysko KL, Talley BL. 2004. Distribution and ecology of the introduced African rainbow lizard, *Agama agama africana* (Saura: Agamidae), in Florida. Fla Sci. 67:303–310.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 25(15):1965–1978.

Jakobsson M, Rosenberg NA.2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23 (14):1801–1806.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24(11):1403–1405.

Kissling WD, Blach-Overgaard A, Zwaan RE, Wagner P. 2016. Historical colonization and dispersal limitation supplement climate and topography in shaping species richness of African lizards (Reptilia: agaminae). Sci Rep. 6(1):34014

Leaché AD, Oaks JR. 2017. The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. Annu Rev Ecol Evol Syst. 48(1):69–84.

Leaché AD, Wagner P, Linkem CW, Böhme W, Papenfuss TJ, Chong RA, Lavin BR, Bauer AM, Nielsen SV, Greenbaum E, *et al*. 2014. A hybrid phylogenetic-phylogenomic approach for species tree estimation in African Agama lizards with applications to biogeography, character evolution, and diversification. Mol Phylogenet Evol. 79:215–230.

Leaché AD, Grummer JA, Miller M, Krishnan S, Fujita MK, Böhme W, Schmitz A, Lebreton M, Ineich I, Chirio L, et al. 2017. Bayesian inference of species diffusion in the West African *Agama agama* species group (Reptilia, Agamidae). Syst Biodivers. 15(3):192–203.

Manel S, Schwartz M, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 18(4):18.

Manthey JD, Moyle RG. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis)* of the Madrean Archipelago sky islands: a landscape genomics approach. Mol Ecol. 24 (14):3628–3638.

Mediannikov O, Trape S, Trape J-F. 2012. A molecular study of the genus *Agama* (Squamata: Agamidae) in West Africa, with description of two new species and a review of the taxonomy, geographic distribution, and ecology of currently recognized species. Russ J Herpetol. 19:115–142.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89(3):583–590.

Oufiero CE, Gartner GE, Adolph SC, Garland T Jr. 2011. Latitudinal and climatic variation in body size and dorsal scale counts in *Sceloporus* lizards:a phylogenetic perspective. Evolution. 65(12):3590–3607.

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 (5):e37135.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2):945–959.
Pritchard JK, Falush D, Stephens M. 2002. Inference of population structure in recently admixed populations. Am J Hum Genet. 71:177.
Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. Mol Ecol Notes. 4(1):137–138.
Seba A. 1734. Locupletissimi Rerum Naturalium Thesauri Accurata Descriptio, et Iconibus Artificiosissimis Expressio, per Universam Physices Historiam. Opus, cui, in hoc Rerum Genere, Nullum par Exstitit. Ex Toto Terrarum Orbe Collegit, Digessit, Descripsit, et Depingendum Curavit Albertus Seba, Etzela Oostfrisius, Academiæ Caesareæ Leopoldino Carolinæ Naturæ Curiosorum Collega Xenocrates dictus; Societatis Regiæ Anglicanæ, et Instituti Bononiensis, sodalis. Tomus I, Amstelaedami: Apud J. Wetstenium, and Gul. Smith, and Janssonio-Waesbergios. 33, 178 p., 111 pls.
Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP. 2007. Putting the "landscape" in landscape genetics. Heredity. 98(3):128–142.
Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. 2010. Landscape genetics: where are we now? Mol Ecol. 19(17):3496–3514.
Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland (Massachusetts): Sinauer Associates.
Wagner P, Wilms TM, Bauer A, Böhme W. 2009. Studies on African Agama V. On the origin of *Lacerta agama* Linnaeus, 1758 (Squamata: Agamidae). Bonn Zool Beitr. 56(4):215–223.
Wagner P, Mazuch T, Leaché AD, Böhme W. 2013. Additions to the lizard diversity of the Horn of Africa: two new species in the *Agama spinosa* group. Amphib-Reptil. 34(3):363–387.

Appendices

Materials examined

Voucher specimens for genetic samples: LSUMZ-H 20027, 20055-20056, 20085, 20162-20165, 20333-20336, 20342-20345; MVZ 245231-245238, 245240-245242, 245244-245250, 245253-245257, 245260, 245262, 245264, 245431-245434, 249600-249602, 249604-249605, 249607-249613, 249616-249617, 249618-249619, 249621-249627, 249629, 249631, 249633-249634, 249636-249640, 249644-249646, 249649-249654, 252405; UWBM 5548, 5550-5551, 5553-5554, 5557-5560, 5562-5563, 5566, 5570-5571.

Voucher specimens for morphological samples: LSUMZ-H 20055, 20056, 20027, 20085, 20162-20165, 20333-20336, 20342; MVZ 245234-245238, 245246, 245259, 249608, 249621, 249640; UWBM 5548, 5550-5552, 5554-5559, 5560, 5562, 5564, 5566, 5567, 5571, 5573.