DR. EDWARD MYERS (Orcid ID : 0000-0001-5369-6329)
MR. ALEXANDER T XUE (Orcid ID : 0000-0003-1521-6448)
DR. MARCELO GEHARA (Orcid ID : 0000-0001-9899-1970)
DR. CHRISTIAN L COX (Orcid ID : 0000-0003-1247-7247)

Article type : From the Cover

Environmental Heterogeneity and Not Vicariant Biogeographic Barriers Generate Community Wide Population Structure in Desert Adapted Snakes

Short running title: Climate drives population structure

Edward A. Myers[1,2,*], Alexander T. Xue[3], Marcelo Gehara[2], Christian Cox[4], Alison R. Davis Rabosky[5], Julio Lemos-Espinal[6], Juan E. Martínez-Gómez[7], & Frank T. Burbrink[2]

[1]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

[2]Department of Herpetology, The American Museum of Natural History, Central Park West and 79th Street, New York, NY 10024 USA

[3]Department of Genetics, Rutgers University, 604 Allison Road, Piscataway, NJ 08854

[4]Department of Biology, Georgia Southern University, Statesboro, GA, 30460

[5]Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, 1109 Geddes Ave., Ann Arbor, MI 48103 USA

[6]Laboratorio de Ecología. UBIPRO, FES Iztacala, Universidad Nacional Autónoma de México, Av. De los Barrios # 1, Col. Los Reyes Iztacala, Tlalnepantla, Estado de México. 54090, Mexico

[7]Instituto de Ecología, A.C., Red de Interacciones Multitróficas, Carretera Antigua a Coatepec 351, Xalapa, Veracruz 91070, México

*Corresponding author: eddie.a.myers@gmail.com

Abstract

Genetic structure can be influenced by local adaptation to environmental heterogeneity and biogeographic barriers, resulting in discrete population clusters. Geographic distance among populations, however, can result in continuous clines of genetic divergence that appear as structured populations. Here we evaluate the relevant importance of these three factors over a landscape characterized by environmental heterogeneity and the presence of a hypothesized biogeographic barrier in producing population genetic structure within 13 codistributed snake species using a genomic dataset. We demonstrate that geographic distance and environmental heterogeneity across western North America contribute to population genomic divergence. Surprisingly, landscape features long thought to contribute to biogeographic barriers play little role in divergence community wide. Our results suggest that isolation by environment is the most important contributor to genomic divergence. Furthermore, we show that models of population clustering that incorporate spatial information consistently outperform nonspatial models, demonstrating the importance of considering geographic distances in population clustering. We argue that environmental and geographic distances as drivers of community-wide divergence should be explored before assuming the role of biogeographic barriers.

Key words: comparative phylogeography, biogeographic barriers, generalized dissimilarity modeling, community ecology, population structure, gene flow

Introduction

Population structure across a species' range is typically produced by isolation by distance (IBD), isolation by environment (IBE), or isolation by resistance (IBR). Isolation by distance, which is commonly reported in empirical data sets (Pelletier & Carstens, 2018; Wang, Glor, & Losos, 2013; Wright, 1943), is defined as spatial auto-correlation in the distribution of genetic variation and is the outcome of limited dispersal abilities which reduces opportunity for gene flow across the extent of a species' geographic distribution. Limited dispersal therefore results in negative associations with genetic relatedness and geographic distance (Vekemans & Hardy, 2004). Because IBD simply correlates Euclidian distance in geography and genetic distance, this metric ignores heterogeneity in the environment and landscape. By contrast, IBE predicts spatial genetic divergence based on environmental differences between sampled demes, regardless of geographic distance (Wang & Bradburd, 2014). Isolation by environment can result from several

unique processes, such as natural selection against immigrants, reduced hybrid fitness or biased rates of dispersal (Wang & Bradburd, 2014). Lastly, resistance distances across a heterogeneous landscape can structure spatial genetic divergence (McRae, 2006). Such resistance distances are often used to capture features of the landscape that may be acting as physically isolating barriers to dispersal rather than an adaptive barrier as is the case with IBE. Therefore, IBR may be considered the main force driving population structure at biogeographic barriers. Isolation by resistance is calculated as the probability that an individual will migrate from one population to the other, weighted by a friction to dispersal across unsuitable habitats and/or physical barriers (McRae, 2006; Wang & Bradburd, 2014). A pattern of IBR arises when characteristics of the landscape modify gene flow between demes such that resistance across these landscapes (e.g., across rivers or over mountains) provide a more appropriate predictor of genetic differentiation than do Euclidean distances or (non-spatial) environmental distances (McRae, 2006).

One or more of these three patterns may explain patterns of divergence in population genomic data and differentiating them may be difficult. In addition, if patterns of IBD dominate population genetic structure, inferences of discrete population clusters may be spurious (Bradburd, Coop, & Ralph, 2018; Meirmans, 2012), and these spurious inferences may also extend to local adaptation to clinal variation in environment. By contrast, sharp environmental transitions or migration resistance across biogeographic barriers will likely produce discrete population structure. Because distance, environment, and landscape are often spatially auto-correlated with one another, failure to examine the effects of all of these variables may potentially result in incorrect estimates of the drivers of population divergence (Reid, Mladenoff, & Peery, 2017). Taking into account geographic distances, environmental variation, and heterogeneity in the landscape will help to understand the factors that facilitate adaptation and species diversification.

Comparative studies of multiple codistributed species can advance our understanding of organism-landscape interactions, reveal factors that generate population genetic structure, and address whether multiple species are affected in similar ways to shared environments (Wang & Bradburd, 2014). Responses to shared landscapes can vary from concordant (Jackson et al., 2018), to entirely discordant population genetic structure (Phillipsen et al., 2015). The degree to which spatial genetic structure is shared across codistributed species may be affected by

organismal traits (Phillipsen et al., 2015; Reid et al., 2017). For example, genomic divergence in taxa with greater dispersal abilities may have little to no signature of IBD compared to taxa with lower dispersal abilities (Phillipsen et al., 2015).

Within arid, southwestern North America, several studies have demonstrated that codistributed species have a signature of population divergence between the Sonoran and Chihuahuan Deserts across the Cochise Filter Barrier (CFB; Zink et al., 2001; Pyron and Burbrink, 2010; Myers et al., 2017b), potentially making this an important regional biogeographic barrier (Figure 1). This region is both geographically and topographically complex and provides opportunities for allopatric divergence. The river networks of southwestern North America may have also driven allopatric divergence and population structure in numerous taxa (e.g., the Pecos River, the Rio Grande, and the Colorado River, Figure 1; Wood et al., 2013; Graham et al., 2015; O'Connell et al., 2017; Myers et al., 2019). The two deserts are also environmentally heterogeneous, with differences in temperature and precipitation (Figure 1). Divergence due to environmental variation across many species within an assemblage could potentially lead to codiversifcation at the community level (Johnson & Stinchcombe, 2007; Wang & Bradburd, 2014).

The snake fauna codistributed across southwestern North America is an assemblage of ecologically, behaviorally and physiologically diverse taxa that presents an opportunity to examine how genomic variation is distributed across the landscape. For example, this community is composed of both oviparous and ovoviviparous species (e.g., *Lampropeltis getula* and *Trimorphodon biscutatus* versus *Crotalus* spp. and *Thamnophis marcianus*), strictly nocturnal and strictly diurnal taxa (e.g., *Hypsiglena torquata* versus *Masticophis flagellum*), and taxa that specialize on an invertebrate diet as well as those that feed primarily on small rodents (e.g., *Sonora semiannulata* versus *Pituophis catenifer*). These differences might be reflected in the determinants of population structure (Phillipsen et al., 2015; Reid et al., 2017). Previously, it has been shown that many of these snake taxa are reciprocally monophyletic in mtDNA gene trees across the CFB (Myers et al., 2017) and that geographic distance is an important variable in explaining genetic variation across these taxa. The authors concluded that divergence times were asynchronous among east-west population pairs in 12 snake taxon groups, indicating non-shared histories (Myers et al., 2017). Furthermore, numerous species delimitation studies have elevated

species east and west of the CFB (Anderson & Greenbaum, 2012; Cox et al., 2018; Devitt, LaDuc, & McGuire, 2008; Mulcahy, 2008; O'Connell & Smith, 2018; Pyron & Burbrink, 2009) while additional studies have suggested widespread cryptic diversity within these snake species (Dahn et al., 2018; Myers et al., 2017a), therefore distinct population structure is likely present across this biogeographic barrier.

Given the previous research conducted within this region, we hypothesize that the CFB drives population divergence across an entire assemblage of species, all of which are widely distributed across arid North America. We predict that IBR will be a key determinant of genomic divergence and that the location of the CFB will be concordant with the lowest effective migration rates in nearly all species. To test these predictions, we generate a reduced-representation genomic data set, analyzing these data with both nonspatial and spatial population clustering methods. We then explicitly test for the impacts of IBD, IBE, and IBR on genetic structure, as well as quantify which environmental variables and geographic features (e.g., climate, riverine barriers or elevation) are most important in producing patterns of population genetic structure.

Methods

Sample Collection

A total of 383 tissue samples were obtained throughout the range of each of the 13 snake species groups studied here (*Arizona elegans*, *Crotalus atrox*, *Crotalus molossus*, *Crotalus scutulatus*, *Hypsiglena torquata*, *Lampropeltis getula*, *Masticophis flagellum*, *Pituophis catenifer*, *Rhinocheilus lecontei*, *Salvadora hexalepis*, *Sonora semiannulata*, *Thamnophis marcianus*, and *Trimorphodon biscutatus*), with collecting efforts focused on sampling from within the Chihuahuan and Sonoran Deserts. The number of individuals per taxon ranged from 15 – 44 and averaged 29.5 (Table 1). Snakes are difficult to collect in large numbers and therefore while sampling efforts were focused on collecting these thirteen species, samples were often collected opportunistically yet with the goal of broadly sampling each species within the Sonoran and Chihuahuan Deserts across the Cochise Filter Barrier.

Generation of Sequence Data

Genomic DNA was extracted from muscle or liver tissues using DNeasy kits (Qiagen, Valencia, CA, USA) following manufacturer's protocols. Double stranded DNA concentrations were quantified using a Qubit (Thermo Fisher Scientific, Waltham, MA USA). We sent up to 30,000 ng of DNA from each sample to Cornell Institute of Genomic Diversity for genotyping-by-sequencing (GBS; Elshire et al., 2011). GBS is a technique for building reduced representation libraries, similar to other restriction-site associated DNA sequencing methods where a restriction enzyme is used to reduce the complexity of the genome before sequencing (Elshire et al., 2011). Specifically, the method implemented uses methylation-sensitive restriction enzymes which targets low copy regions of the genome avoiding repetitive regions (Elshire et al., 2011). Genomic DNA was digested with the Pst1 enzyme and sample-specific barcode adapters as well as a common adapter were ligated to the sticky end of the fragments. Libraries were sequenced on a 100 bp single-end Illumina HiSeq 2000 at the Cornell Core Lab Center.

Bioinformatics/SNP calling

We processed raw Illumina reads using the bioinformatics pipeline pyRAD v.3.0 (Eaton, 2014) to assemble de novo GBS loci. Each species group was analyzed independently and samples were demultiplexed using their unique barcode sequence. The maximum number of sites allowed with a Phred score <20 was set to 4 (these sites were changed to N's), minimum sequence depth was set to 10 reads per locus, and we used a clustering threshold of 90%. All fragments >50 bps were retained. Additionally, we filtered sequences where loci with excessive heterozygous sites (>3) were removed to reduce the chances of keeping paralogous sequences. Lastly, minimum-taxon coverage was set at 75% of all individuals, allowing for 25% missing data per locus in the final sequence alignments. Filtered reads for each sample were clustered using vsearch (https://github.com/torognes/vsearch) and aligned with MUSCLE (Edgar, 2004). Only one SNP per locus was retained for downstream analyses, in order to reduce the possibility of linked SNPs.

Isolation by Distance

As an initial exploration of IBD within these data, we fit a linear model between genetic distance and Euclidian geographic distance for all sampled individuals, and calculated an r^2 value and p-value. Genetic distances were calculated as absolute genetic distances, without making any assumptions regarding mutation or genetic drift (Prevosti's genetic distances; Kamvar, Tabima, & Grünwald, 2014; Prevosti, Ocaña, & Alonso, 1975), in adegenet using a matrix of one SNP per locus for each taxon and geographic distances between sampling localities were calculated using the R package fossil (Vavrek, 2011).

We also implemented the spatial method *Estimated Effective Migration Surface* (EEMS; (Petkova, Novembre, & Stephens, 2015), that is used to find patterns of genetic diversity across a landscape that deviate from a null expectation of IBD. We applied this method as an exploratory tool to find regions of the landscape that may act as biogeographic barriers in this system (e.g., the Cochise Filter Barrier or major river systems) and to explore if there are common patterns shared across taxa. This method is based on a stepping-stone model where individuals migrate locally between demes and migration rates are allowed to vary by location (Petkova et al., 2015). To capture continuous genetic diversity, the landscape is divided into demes and each deme can only exchange migrants with its neighbors. Under this model, expected genetic dissimilarities depend on sample location and migration rates (Petkova et al., 2015). EEMS explicitly represents genetic differentiation as a function of migration rates and correlates genetic variation with geography, producing visualizations that highlight portions of a species range where population divergence deviates from patterns expected under IBD. These regions are indicative of areas of the landscape that act as barriers to gene flow, or conversely promote gene flow acting as species corridors (Richmond et al., 2017). For example, regions where EEMS identifies spatial genetic patterns that have lower than expected effective migration under pure IBD is suggestive of population clustering (i.e., a region of lower than expected migration under IBD is potentially a contact zone between genetically distinct populations). Using the above genetic distance matrices, we ran EEMS using a deme size of 1200 (i.e., the density of populations), with three independent starting chains for 5×10^6 MCMC iterations following a burn-in of 1×10^6, with a thinning of 5000 and different starting seeds, for each taxon. Posterior plots were compared across independent runs for each taxon to ensure convergence. These three runs per taxon were

combined and visualized using the R package reemsplots2 (https://github.com/dipetkov/reemsplots2).

Spatial Population Clustering: *Spatial vs. Nonspatial*

Whether genetic divergence should be represented as discrete clusters or continuous clines of variation is a well-known problem in population genetics (Bradburd et al., 2018). Here we implemented conStruct to avoid this potential issue. conStruct is a model-based method that simultaneously infers continuous and discrete patterns of population structure by estimating ancestry proportions for each sampled individual from two-dimensional population layers, where within each layer a rate at which relatedness decays with distance is estimated (Bradburd et al., 2018). This method also allows for a cross validation procedure for model selection, between both spatial and nonspatial models as well as the number of underlying layers (Bradburd et al., 2018). This analysis allows us to specifically test whether population structure con be attributed to IBD versus IBE or IBR. For example, under s scenario of pure IBD we would expect conStruct to find a strong support for spatial model a single population ($K = 1$), alternatively if the CFB has structured populations we expect to find support for two populations, with geographic distributions that meet approximately at the Western Continental Divide. Based on preliminary runs, large amounts of missing data may bias results. Therefore, with the unlinked SNP dataset, individual samples missing more than 75% of genotypes were removed and after these individuals were removed we again removed loci to ensure that there was only 25% missing data within a locus (see Table 1). Pruning of these data sets was conducted in vcftools (Danecek et al., 2011) and vcf files were converted to Structure input files (Pritchard, Stephens, & Donnelly, 2000) using plink (Purcell et al., 2007). The cross-validation procedure to test between discrete clusters versus continuous variation within conStruct was then run for each taxon with $K = 1 – 6$, or until the predictive accuracy reached a value of 0, with 10 repetitions per each K value, 100,000 iterations per repetition, and a training proportion of 0.9. When choosing a best fit value of K we required that all layers contribute >2% to the total covariance of the model.

Determinants of Population Genomic Structure: IBD, IBE, IBR

To determine what variables best predict genomic divergence, we implemented generalized dissimilarity modeling (GDM; Ferrier *et al.* 2007). This method is a matrix regression technique that models variation in distance matrices by relating dissimilarity in genetic distances between sites to differences in environmental distances and the degree to which these sites are isolated from one another (e.g., geographical or resistance distances; Fitzpatrick & Keller 2015; Thomassen et al., 2010). GDM can fit nonlinear relationships of environmental/distance variables to genetic variation through the use of I-spline basis functions (Ferrier et al., 2007). This method uses the percent deviance explained as a measure of model fit (Fitzpatrick & Keller, 2015). We used this method to simultaneously examine the effects that geographic distance, environmental variables, and several potential resistance surfaces have on generating genomic divergence.

For environmental variables we downloaded the 19 BioClim variables (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) at 30 second resolution. We then reduced this to a set of variables such that correlation among variables was <0.7 using the raster.cor.matrix function in the ENMTools R package (Warren, Glor, & Turelli, 2010). This resulted in retaining nine Bioclim varaibles for use in GDM models (Annual Mean Temperature, Mean Diurnal Range, Isothermality, Temperature Seasonality, Mean Temperature of Wettest Quarter, Mean Temperature of Driest Quarter, Annual Precipitation, Precipitation of Driest Month, and Precipitation Seasonality). Environmental variation for each collecting locality for all species was extracted from this set of uncorrelated variables. We used three different resistance surfaces that may better reflect patterns of genomic divergence than pure geographic distance, these are resistance around major rivers of southwestern North America, elevation, and potential geographic distributions based on ecological niche models (ENMs). Major rivers were selected given that numerous studies have suggested that the Pecos River, the Rio Grande, and the Colorado River (e.g., Graham et al., 2015; Myers et al., 2019; O'Connell et al., 2017; Wood et al., 2013) are barriers to gene flow. Elevation was selected as a resistance surface because the Cochise Filter Barrier is often associated with the Western Continental Divide (Castoe, Spencer, & Parkinson, 2007), a high elevation region between major watersheds in North America and because the Central Mexican Plateau has been associated with lineage divergence in previous studies of the same taxa (Schield et al., 2018). A shape file of rivers was obtained from

https://www.naturalearthdata.com/downloads/50m-physical-vectors/ and elevation was obtained from https://research.cip.cgiar.org/gis. Both of these were converted to an ascii file using the raster library in R (Hijmans & van Etten, 2012). Lastly, we chose to use ENMs as a resistance surface because potential routes of dispersal and gene flow among populations are likely restricted by suitable habitat. ENMs were created for each taxon independently by first retrieving 500 locality records from the Global Biodiversity Information Facility (GBIF.org) using the R package spocc (Chamberlain, Ram, & Hart, 2016). Any records outside the known geographic distributions of these species were then removed. Furthermore, occurrences outside our study region were then dropped (-126, -90, 18, 50) and thinned so that sampled localities within 50 km were removed, using spThin (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015). Using Biomod2 (Thuiller, Georges, & Engler, 2013) we sampled 10,000 pseudoabsence points within the study region and Maxent v3.4.1 (Phillips, Anderson, & Schapire, 2006) was used to construct ENMs using all 19 bioclim variables. We used all available bioclim variables because the regularization method implemented in MaxEnt is stable even if variables are correlated, therefore removing potentially correlated variables or preprocessing covariates through the use of PCA and selecting only the dominant axes for using analysis is unnecessary (Elith et al., 2011). Each analysis was replicated for 5000 iterations, reserving 25% of samples as a training dataset for model evaluation, and we created response curves and jackknifed our data to measure variable importance. The average of these ENMs were then projected and saved as ascii files. The ascii files were normalized to values of 0 – 1. In the case of rivers and elevation greater values represent increased resistance rates across the landscape (e.g., in the case of potential riverine barriers, rivers = 1 and non-river = 0) and were used as resistances in Circuitscape 4.0 (McRae, 2006; McRae et al., 2016) implemented in Julia. In the case of ENMs, the ascii files were also normalized to values of 0 – 1, but these were used as conductance surfaces in Circuitscape analyses.

Using GDM, we tested how these geographic distance (IBD), environmental variation (IBE; all uncorrelated Bioclim variables), and three models of distance matrices (IBR) contribute to genomic divergence. Our previously generated absolute genetic distance matrices (from all potentially unlinked SNPs) were used as the response variable and the gdm R package (Manion, Lisk, Ferrier, Nieto-Lugilde, & Fitzpatrick, 2016) was used to fit generalized dissimilarity

297 models. We also calculated Nei's D genetic distances from our unlinked SNPs and repeated all
298 GDM analyses using this measure of genetic distance as the response variable (Nei, 1972). We
299 ran seven independent tests for each taxon with different sets of predictor variables: (1) a full
300 model with geographic distance, environmental variables, and the resistance surfaces, (2) a
301 model with geographic distance and environmental variables, (3) a model with geographic
302 distance and resistance distances, (4) a model with environmental variables and resistance
303 distances, (5) a model with environmental variables only, (6) distance only, and (7) resistance
304 distances only. We used the gdm.varImp function in the gdm R package on on all seven models,
305 which uses a matrix permutation to perform model and variable significance testing and
306 estimates variable importance in a GDM.
307 Because a large percent deviance can be explained in our GDM models, we tested
308 whether nucleotide diversity or sample size was correlated with percent deviance explained.
309 Nucleotide diversity was calculated for each species in PopGenome (Pfeifer, Wittelsbürger,
310 Ramos-Onsins, & Lercher, 2014) package of R. We then fit linear models between nucleotide
311 diversity and percent deviance as well as between the total number of samples collected per
312 species and percent deviance explained, an r^2 and p-value were calculated for these two linear
313 models. We also tested if environmental variation in the Bioclim variables can be explained by
314 geographic distance alone. To do this we used GDM, for each set of collecting localities for each
315 taxon. In GDMs we used the 19 BioClim variables as a response variable and latitude and
316 longitude as the predictor variables.
317
318 **Results**
319 *Sequencing and Bioinformatics*
320 We generated GBS data for 383 specimens resulting in 1,009,845,311 reads and 72.12
321 GB of raw data with an average of 2,120,912.5 ± 1,446,417.4 reads per individual (see
322 Supporting Information). After excluding loci with more than 25% missing data, 11,681 –
323 46,444 total SNPs and 5,496 – 21,259 SNPs when restricted to one SNP per locus, depending on
324 the species group were retained (Table 1). Raw sequence data are available on the NCBI
325 Sequence Read Archive (Accession: PRJNA554495) and the assembled GBS data used in this
326 study are available on Dryad (doi:10.5061/dryad.2172qg4).

Patterns of IBD

The r^2 values from linear models of correlations between genetic distances and geographic distances range from 0.13 – 0.73 (in *P. catenifer* and *C. molossus*, respectively) and in all cases p-values < 0.05 (Supporting Information). The EEMS analyses highlight regions of lower than expected migration across the geographic distributions of all 13 taxa. Many taxa show regions of reduced gene flow that run north to south separating populations into the Sonoran and Chihuahuan Deserts (Figure 2; e.g., *C. atrox* and *H. torquata*). However, within some taxa, the geographic features that might be creating these regions of reduced gene flow are less clear and are not strictly associated with the Cochise Filter Barrier, nor are there shared, community wide patterns of reduced gene flow (Figure 2). For example, rates of migration are reduced across much of the geographic distribution of *C. scutualtus* and not associated with any biogeographic barriers (e.g., these regions are not tightly associated with the CFB or major rivers). Likewise, within *A. elegans* reduced rates of migration are inferred across many of the sampling localities within the western portion of this species' range and nearly all of the Sonoran Desert has reduced rates of migration within *Sal. hexalepis* (Figure 2).

Spatial Population Clustering

In cross-validation analyses of spatial versus nonspatial population clustering across all 13 taxa, a model that includes spatial information outperforms nonspatial models using conStruct (Supporting Information). These analyses suggest that incorporating geographic information, which may be a reflection of a pattern of IBD, are important for determining the number of genetic clusters in all species across this assemblage. These cross-validation analyses coupled with a required threshold of 0.02 minimum contribution of each layer to total covariance, suggest that between $K = 1 – 4$ layers sufficiently describe the genomic data within each species (Figure 3; Supporting Information). Within *C. scutulatus* and *P. catenifer* the best support is for a spatial model with $K = 1$ (e.g., adding an additional layer at $K = 2$ for *C. scutulatus* only contributed to explaining an additional 0.5% of the model covariance; for *P. catenifer* this additional layer only explained an additional 0.08%; see Supporting Information), suggesting that genomic variation within these two taxa are indicative of a continuous cline of ancestry, a pattern of IBD. With the

exception of these two groups, conStruct results provide strong support for discrete population structure across arid North America. Seven species show a strong signal of population divergence across the CFB (Figure 3) suggesting that IBR may have influenced population genetic structure in these groups. However, the cause of population structure in some species is less clear, for example the cause of population structure in *R. lecontei*, *Sal. hexalepis*, *Son. semiannulata*, and *Th. marcianus* are unidentifiable (Figure 3). Maps of all tested levels of *K* layers are included in the Supporting Information for both spatial and nonspatial models.

Determinants of Population Genomic Structure: IBD, IBE, IBR

Ecological niche models for all taxa had reasonable performance with AUC values ranging from 0.9 (*P. catenifer*) to 0.97 (*C. molossus*; Table 2 and Supporting Information for projected ENMs). BioClim variables related to temperature, specifically mean temperature of the coldest quarter, contributed the most to ENMs in the majority of species (Table 2). Only in two taxa, *T. biscutatus* and *Sal. hexalepis*, did variables related to precipitation contribute more to ENMs than did variables related to temperature (Table 2). Output ascii files for each ENM are available from Drayd (doi:10.5061/dryad.2172qg4).

When using absolute genetic distances as a response variable, the GDM models that account for all possible predictor variables (geographic distance, environmental variation, and resistances distances) potentially generating genomic variation across these 13 species explained between 35.9 and 95.4% (average deviance of 65.6% ±20%) of the total observed genomic variation and were significant in all of the 13 species (Table 2). The variables that contributed the most to models that included all potential predictor variables varied by taxa but most often included geographic distance (9/13 species), and rarely included resistance surfaces generated from Circuitscape (3/13 species; Table 2). In each of the three cases where resistance surfaces were important predictor variables, the resistance variable differed (e.g., in *L. getula* elevation was important, where as in *P. catenifer* resistance distances around ENMs were important, and in *S. hexalepis* rivers as barriers were important). Furthermore, the climatic variables of most importance in explaining genomic variation from the full GDM models were never the same variables as contributing the most to the generated ENMs (Table 2). This suggests that the variables that predict the geographic distribution of a species are not the same as those promoting

population divergence. GDM models which only included climatic variables or climatic variables and geographic distance performed nearly as well as the full model (i.e., all predictor variables), while the GDM that included only geographic or resistances distances predicted much less variation alone (Table 2). Variable importance values resulting from model permutations and statistical significance are presented in the Supporting Information. The exact predictor variables differed slightly when using Nei's D genetic distances as the response variable instead of absolute genetic distances (Supplemental Information). However, models that incorporated environmental variation or environmental variation and geographic distances were consistently the top models in explaining Nei's D genetic distances within species, whereas models that only consisted of IBR distances explained less genetic differentiation (Supplemental Information).

GDM models were also able to explain between 19 – 76.2% of the variation in correlations between geographic distance and the climate variables used above, however GDM models were inconclusive in several cases suggesting that geographic distance is not always correlated with environmental variation across the geographic distribution of these thirteen species (Table 2). There was no correlation between observed nucleotide diversity and percent deviance explained in GDM models (Supplemental Materials), however there was a correlation between the number of samples per species and the deviance in genetic differentiation explained in our GDM models ($r^2 = 0.65$, p-value < 0.05). This suggests that smaller samples sizes result in a larger percent deviance explained when using GDM models (see Supporting Information).

Discussion

Using comparative population genomic data across 13 codistributed snake species, we demonstrate that isolation by environment and isolation by distance are common patterns in population genomic divergence across an entire assemblage. Surprisingly, features of the landscape thought to contribute to biogeographic barriers (e.g., differences in elevation, for example the CFB, and rivers) play little role in population differentiation. Genetic clustering methods that explicitly account for spatial information consistently outperformed nonspatial clustering methods, which regularly over-split the number of populations within a species (Supporting Information). These spatial clustering analyses demonstrate that some species have population structure across the CFB, however, this pattern is inconsistent across the entire

species assemblage (Figures 2 & 3). Together these results suggest that local environmental conditions, not shared biogeographic barriers, are likely driving lineage divergence, and importantly the determinants of population divergence are taxon specific.

IBE Plays a Dominant Role in Population Structure

For 13 codistributed species, we find that both IBD and IBE contribute to spatial genomic divergence and that on average IBE contributes to approximately 2.5 times more genomic divergence than does IBD alone (mean IBE 62.0% vs. mean IBD 33.9%; Table 2). These two combined variables contributed to a large portion of genomic divergence in all taxa (e.g., up to 95.4% in *Sal. hexalepis*; Table 2), suggesting that our analyses are capable of detecting the underlying processes of diversification. Results are consistent across taxa where environmental divergence was always highly predictive of genomic divergence. However, we also found that the most important environmental variable in driving genomic divergence varied among taxa and it was evenly divided whether temperature or precipitation was the most important variable in predicting divergence (Table 2). Therefore, while climatic differences are broadly important for driving divergence, the key components of diversification are species specific. Because much genomic divergence can be explained by environmental heterogeneity, future studies should focus on differential selection and functional adaptive differences between populations to separate ecological from historical processes in driving speciation within this region (Sobel, Chen, Watt, & Schemske, 2010). However, it is important to point out that the amount of genomic divergence explained by GDM models is sensitive to the total number of samples included in analyses, where GDM models explain more deviance with smaller sample sizes (Supporting Information). However, these models are statistically significant (Table 2) as are most of the variables of importance using permutation tests (Supporting Information).

The predominant role of environmental heterogeneity in shaping genomic divergence in this system suggests that local adaptation is an important process in structuring populations and potentially responsible for species level diversification (Nosil, 2012; Sexton, Hangartner, & Hoffmann, 2014; Shafer & Wolf, 2013). However, a dominant role of IBE in promoting genomic divergence is not the outcome of other similar studies. For example, the majority of mtDNA variation within Caribbean *Anolis* lizards can be attributed to patterns of IBD (Wang et al.,

2013). Similarly, genomic variation within Australian skinks is best explained by a pattern of IBD (Singhal et al., 2018). Because of the contrasts between these previous studies and our results, it is important to highlight that the drivers of genomic divergence may vary greatly across taxa under investigation or study region (e.g., differentiation on islands compared to continental radiations).

Although comparative population genomics studies can identify correlations between landscape and environmental characteristics and population genetic structure, the underlying relationship between species traits and genetic variation can be difficult to determine (Reid et al., 2017). It is likely that species traits are important in structuring population genetic patterns (Zamudio, Bell, & Mason, 2016) and therefore even closely related, codistributed species, while subjected to similar landscapes and environmental variation, can have very different population structure. For example, all taxa within the tribe Lampropeltinii (*A. elegans*, *L. getula*, *P. catenifer*, and *R. lecontei*) examined here, though closely related (divergence time ~12.2 mya; Chen, Lemmon, Lemmon, Pyron, & Burbrink, 2017), have unique determinants of population structure (Table 2). This may be an expected outcome of such comparative analyses given that previous studies have found landscape genetics patterns to be influenced by species-specific dispersal abilities, life history traits or habitat preferences (Reid et al., 2017; Robertson et al., 2018). Therefore, understanding differences in species specific traits may ultimately help elucidate what landscape features promote connectivity and gene flow among populations (Zamudio et al., 2016). However, determining which traits are useful for predicting patterns of population genetic structure and gene flow may prove to be difficult. For example, codistributed species with very different physiologies and life histories can become locally adapted in response to similar environmental variation. Within our study taxa, two groups of distantly related taxa have similar determinates of population structure. For example, with both *A. elegans* and *Tha. marcianus*, genomic distance between populations is best explained by both geographic distance and Precipitation of Driest Month (Table 2). While these two taxa have broadly overlapping geographic distributions, they are very distantly related (diverged approximately ~42 mya; Pyron & Burbrink, 2012) with unique physiologies and ecologies; *A. elegans* is a medium sized, nocturnal, oviparous colubrine that preys largely on lizards (Rodríguez-Robles, Bell, & Greene, 1999) and *T. marcianus* is a semi-aquatic, viviparous species that feeds on fish, anurans, and

477 invertebrates (Ernst & Ernst, 2003). Why these two species would have similar responses in
478 population genetic structure to environmental heterogeneity is unclear. Additionally, GDM
479 analyses demonstrate that geographic distance and Mean Temperature of Driest Quarter are the
480 best predictors of population genetic structure in *C. atrox*, *C. molossus*, and *M. flagellum*. It is
481 also unclear why these three species have similar determinates of population genetic structure,
482 for example, while *C. atrox* and *C. molossus* are closely related, they occupied distinct habitats
483 across arid North America (*C. atrox* is found throughout creosote bush/desert flats while *C.*
484 *molossus* is a higher elevation taxon, rarely found in the desert flats).
485
486 *Spatial Phylogeography & Co-Vicariance*
487 It is often assumed that cyclical climatic changes during the Quaternary coupled with
488 biogeographic barriers were responsible for lineage formation (Hewitt, 2000). Within arid North
489 America numerous studies have cited the CFB as a soft ecological barrier promoting
490 diversification across entire communities that are now in secondary contact (Myers et al., 2017b;
491 Pyron & Burbrink, 2010; Riddle & Hafner, 2006). The CFB has also been described as an
492 ecotonal region dividing the Chihuahuan and Sonoran Deserts (Laport & Minckley, 2013) where
493 there are also climatic gradients from east to west (Figure 1; Schmidt Jr, 1979). Additional
494 geographic features throughout the southwest have also been proposed as important barriers
495 including major river systems (Graham et al., 2015; Myers et al., 2019; O'Connell et al., 2017;
496 Wood et al., 2013) and increases in elevation at the Central Mexican Plateau (Bryson, García-
497 Vázquez, & Riddle, 2011; Schield et al., 2018). Our analyses that incorporate spatial information
498 to account for continuous genetic variation best fit the observed genomic data for 13
499 codistributed species (Supporting Information), with less than half of these taxa showing clear
500 population structure across the CFB (Figure 3), while GDM models suggest little genetic
501 divergence is explained by resistance distances that may be indicative of biogeographic barriers
502 (Table 2). This implies that determinants of population divergence are dissimilar across many
503 members of a biological community and that the CFB as a vicariant biogeographic barrier is not
504 the direct cause of assemblage wide species diversification (Figures 2 & 3). The emphasis on
505 identifying and supporting vicariant barriers within the field of phylogeography may have
506 hampered our understanding of the direct causes of lineage divergence (e.g., Irwin, 2002). The

roles of neutral divergence resulting in clinal variation (e.g., IBD) and that of ecological differentiation due to climatic variation (e.g., IBE) have not been fully appreciated in driving diversification when compared to biogeographic barriers promoting allopatric divergence.

The patterns observed here might be expected to be general to other taxonomic groups regionally and likely at other potential biogeographic barriers globally. Because the geographic locations of population boundaries appear to be concordant with a physical barrier (e.g., a river, ecotone, or elevation) that does not imply that this geologic feature is the root cause of population divergence. Therefore, careful interpretation of phylogeographic results are necessary, specifically across regions proposed as model systems to understand comparative phylogeographic patterns and processes. This is especially important as additional genomic datasets are generated to reinvestigate previous studies based on single locus analyses. At the CFB, numerous single locus phylogeographic studies suggest this region is responsible for lineage divergence (e.g., Pyron and Burbrink, 2010; Myers et al., 2017b). However, our analyses here suggest that spatial patterns in genomic divergence do not match those found in mtDNA analyses, and therefore our understanding of phylogeographic barriers and locations of Pleistocene refugia, particularly in regions that are currently continuously distributed, may need to be reinterpreted. To fully understand the process of speciation and lineage divergence, additional comparative studies from disparate regions of the globe, with sampling across taxonomic diversity, and increased genome scale data, are necessary to explore what is really driving lineage divergence and speciation across communities. However, we suggest that environmental and geographic distances be explored as potential drivers of community-wide divergence before it is assumed that regional biogeographic barriers have promoted diversification.

Empirical data may also be prone to over interpretation. For example, forcing discrete population clusters on continuous data may result in a confirmation bias regarding regional biogeographic barriers. This can occur because new data may be interpreted in a manner that is consistent with preconceived ideas of where phylogeographic barriers are thought to occur (Carstens, Stoute, & Reid, 2009; Nickerson, 1998). This may incorrectly suggest the presence of common biogeographic barriers in comparative studies and ultimately influence all downstream phylogeographic analyses, such as isolation with migration models, species delimitation, and

comparative phylogeography. Phylogeographic studies should routinely analyze population genomic data with both discrete and continuous spatial analyses to avoid these issues. Notably, the taxa here that do not exhibit strong patterns of IBD have qualitatively similar population structure when comparing discrete and continuous population clustering results (e.g., *C. atrox*, *H. torquata*, and *L. getula*; Supporting Information). Furthermore, while IBE is common in nearly all species, climatic variables are also associated with geographic distance (Table 2). Because of auto-correlation between climate and distance, the use of spatially explicit models of population clustering should perform well given information on geographic distance alone.

Conclusions

Here, using a genomic data set generated across 13 codistributed species, we have demonstrated that population divergence across an entire assemblage of snakes has not been produced by vicariant biogeographic barriers (e.g., the Cochise Filter Barrier or major rivers). This is in contrast to our predictions based on what was previously thought about this region. Instead population genetic structure is largely influenced by variation in climate and geographic distance between sampled individuals across arid North America, resulting in patterns of isolation by environment and isolation by distance, that can explain a large proportion of genomic divergence. Given these results, we suggest that future phylogeographic studies explore multiple determinates of population structure before pointing to proposed biogeographic barriers.

Acknowledgements

We thank the following for donating tissue samples used in these analyses, Museum of Vertebrate Zoology (J. McGuire, C. Spencer), Royal Ontario Museum (R. Murphy, A. Lathrop), Texas Natural History Collection, UT-Austin (D. Cannatella, T. LaDuc, and D. Hall), the Sternberg Museum, FHSU (T. Taggart, C. Schmidt, J. Collins), the UTEP Biodiversity Collection, University of Texas El Paso, the Monte L. Bean Life Science Museum at BYU, D. Shepard, S. Ruane, A. Pyron, R. Bryson, P. Lindsey and W. Wüster. EAM would also like to acknowledge A. C. Carnaval, M. J. Hickerson, C. J. Raxworthy, S. E. Alter, E. R. Hekkala, and FTB who served as his dissertation committee, this manuscript is a heavily revised chapter of

that dissertation. The authors would also like to thank A. McKelvy, J. Servis, A. Kuhn, B. Reid, I. Overcast, K. Provost, and the Bell Lab at NMNH for assistance with analyses and/or providing comments on previous drafts of this paper. M. Fitzpatrick, G. Bradburd, and D. Petkova provided comments and assistance with analyses. A number of researchers helped in field collection of samples and we owe a great deal of gratitude to those individuals; M. Soley, M. Gomez, M. A. Dominguez, M. Sosa, A. L. González, V. São Pedro, E. Faria, Smoochy, N. E. Lara-Díaz and C. A. López González. Additional support for fieldwork was generously provided by the Estación del Desierto de Mapimí – INECOL and Rancho El UNO of Nature Conservancy México, as well as logistical support provided by the Laboratorio de Zoología, Facultad de Ciencias Naturales, Universidad Autónoma de Querétaro. Collection of samples in the US were collected under NM permit no. 3559, AZ permit no. SP626898 CLS, and TX permit no. SPR-0413-054 issued to EAM.

Funding

This work was supported by the National Science Foundation (DEB 1500448 awarded to EAM and FTB; and CNS-0855217 and CNS-0958379 that facilitated computational analyses at the CUNY HPCC); a National Geographic Society Young Explorer Grant; City University of New York-Graduate Center Doctoral Student Research Grant and Dissertation Year Fellowship; The Explorer's Club Exploration Fund Grant; American Museum of Natural History Theodore Roosevelt Memorial Fund; Society of Systematic Biologists Graduate Student Research Award; and The Systematics Association/Linnean Society Systematics Research Fund awarded to EAM.

References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5). doi:10.1111/ecog.01132

Anderson, C. G., & Greenbaum, E. (2012). Phylogeography of northern populations of the black-tailed rattlesnake (Crotalus molossus Baird and Girard, 1853), with the revalidation of C. ornatus Hallowell, 1854. *Herpetological Monographs*, *26*(26), 19–57.

doi:10.1655/HERPMONOGRAPHS-D-11-00012.1

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*(1), 33–52.

Bryson, R. W., García-Vázquez, U. O., & Riddle, B. R. (2011). Phylogeography of Middle American gophersnakes: Mixed responses to biogeographical barriers across the Mexican Transition Zone. *Journal of Biogeography*, *38*(8), 1570–1584. doi:10.1111/j.1365-2699.2011.02508.x

Carstens, B. C., Stoute, H. N., & Reid, N. M. (2009). An information-theoretical approach to phylogeography. *Molecular Ecology*. doi:10.1111/j.1365-294X.2009.04327.x

Castoe, T. A., Spencer, C. L., & Parkinson, C. L. (2007). Phylogeographic structure and historical demography of the western diamondback rattlesnake (Crotalus atrox): A perspective on North American desert biogeography. *Molecular Phylogenetics and Evolution*, *42*(1), 193–212. doi:10.1016/j.ympev.2006.07.002

Chamberlain, S., Ram, K., & Hart, T. (2016). spocc: Interface to Species Occurrence Data Sources. R package version 0.7. 0.

Chen, X., Lemmon, A. R., Lemmon, E. M., Pyron, R. A., & Burbrink, F. T. (2017). Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Molecular Phylogenetics and Evolution*, *111*, 206–218.

Cox, C. L., Davis Rabosky, A. R., Holmes, I. A., Reyes-Velasco, J., Roelke, C. E., Smith, E. N., … Campbell, J. A. (2018). Synopsis and taxonomic revision of three genera in the snake tribe Sonorini. *Journal of Natural History*, *52*(13–16), 945–988.

Dahn, H. A., Strickland, J. L., Osorio, A., Colston, T. J., & Parkinson, C. L. (2018). Hidden diversity within the depauperate genera of the snake tribe Lampropeltini (Serpentes, Colubridae). *Molecular Phylogenetics and Evolution*, *129*, 214–225.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330

Devitt, T. J., LaDuc, T. J., & McGuire, J. a. (2008). The Trimorphodon biscutatus (Squamata: Colubridae) Species Complex Revisited: A Multivariate Statistical Analysis of Geographic Variation. *Copeia*, *2008*(2), 370–387. doi:10.1643/CH-07-045

626 Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses.
627 *Bioinformatics*, *30*(13), 1844–1849. doi:10.1093/bioinformatics/btu121
628 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
629 throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. doi:10.1093/nar/gkh340
630 Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical
631 explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.
632 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell,
633 S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity
634 species. *PLoS ONE*, *6*(5). doi:10.1371/journal.pone.0019379
635 Ernst, C. H., & Ernst, E. M. (2003). *Snakes of the United States and Canada*. Washington DC:
636 Smithsonian Books.
637 Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity
638 modelling to analyse and predict patterns of beta diversity in regional biodiversity
639 assessment. *Diversity and Distributions*, *13*(3), 252–264. doi:10.1111/j.1472-
640 4642.2007.00341.x
641 Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level
642 modelling of biodiversity: mapping the genomic landscape of current and future
643 environmental adaptation. *Ecology Letters*, *18*(1), 1–16.
644 Graham, M. R., Hendrixson, B. E., Hamilton, C. A., & Bond, J. E. (2015). Miocene extensional
645 tectonics explain ancient patterns of diversification among turret-building tarantulas
646 (Aphonopelma mojave group) in the Mojave and Sonoran deserts. *Journal of*
647 *Biogeography*, *42*(6), 1052–1065. doi:10.1111/jbi.12494
648 Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913.
649 doi:10.1038/35016000
650 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high
651 resolution interpolated climate surfaces for global land areas. *International Journal of*
652 *Climatology*, *25*(15), 1965–1978. doi:10.1002/joc.1276
653 Hijmans, R. J., & van Etten, J. (2012). raster: Geographic analysis and modeling with raster data.
654 *R Package Version*, *1*, 9–92.
655 Irwin, D. E. (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution*,

56(12), 2383–2394. doi:10.1111/j.0014-3820.2002.tb00164.x

Jackson, J. M., Pimsler, M. L., Oyen, K. J., Koch-Uhuad, J. B., Herndon, J. D., Strange, J. P., … Lozier, J. D. (2018). Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. *Molecular Ecology*. doi:10.1111/mec.14735

Johnson, M. T. J., & Stinchcombe, J. R. (2007). An emerging synthesis between community ecology and evolutionary biology. *Trends in Ecology and Evolution*, *22*(5), 250–257. doi:10.1016/j.tree.2007.01.014

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. doi:10.7717/peerj.281

Laport, R. G., & Minckley, R. L. (2013). Cytogeography of Larrea tridentata at the Chihuahuan-Sonoran Desert ecotone. In *In: Gottfried, Gerald J.; Ffolliott, Peter F.; Gebow, Brooke S.; Eskew, Lane G.; Collins, Loa C. Merging science and management in a rapidly changing world: Biodiversity and management of the Madrean Archipelago III and 7th Conference on Research and Resource Management in the Southwestern Deserts; 2012 May 1-5; Tucson, AZ. Proceedings. RMRS-P-67. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 218-224.* (Vol. 67, pp. 218–224).

Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D., & Fitzpatrick, M. C. (2016). gdm: Functions for generalized dissimilarity modeling. *R Package*.

McRae, B. H. (2006). Isolation By Resistance. *Evolution*, *60*(8), 1551–1561. doi:10.1554/05-321.1

Mcrae, B., Shah, V., & Edelman, A. (2016). Circuitscape: modeling landscape connectivity to promote conservation and human health. *The Nature Conservancy*, *14*. doi:10.13140/RG.2.1.4265.1126

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846.

Mulcahy, D. G. (2008). Phylogeography and species boundaries of the western North American Nightsnake (Hypsiglena torquata): Revisiting the subspecies concept. *Molecular Phylogenetics and Evolution*, *46*(3), 1095–1115. doi:10.1016/j.ympev.2007.12.012

686 Myers, E. A., Bryson Jr, R. W., Hansen, R. W., Aardema, M. L., Lazcano, D., & Burbrink, F. T.
687 (2019). Exploring Chihuahuan Desert diversification in the gray-banded kingsnake,
688 Lampropeltis alterna (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, *131*,
689 211–218.
690 Myers, E. A., Burgoon, J. L., Ray, J. M., Martínez-Gómez, J. E., Matías-Ferrer, N., Mulcahy, D.
691 G., & Burbrink, F. T. (2017). Coalescent Species Tree Inference of *Coluber* and
692 *Masticophis*. *Copeia*, *105*(4), 642–650. doi:10.1643/CH-16-552
693 Myers, E. A., Hickerson, M. J., & Burbrink, F. T. (2017). Asynchronous diversification of
694 snakes in the North American warm deserts. *Journal of Biogeography*, *44*(2), 461–474.
695 doi:10.1111/jbi.12873
696 Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283–
697 292.
698 Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review*
699 *of General Psychology*, *2*(2), 175.
700 Nosil, P. (2012). *Ecoloigcal Speciation*. Oxford, UK: Oxford University Press.
701 O'Connell, K. A., & Smith, E. N. (2018). The effect of missing data on coalescent species
702 delimitation and a taxonomic revision of whipsnakes (Colubridae: Masticophis). *Molecular*
703 *Phylogenetics and Evolution*.
704 O'Connell, K. A., Streicher, J. W., Smith, E. N., & Fujita, M. K. (2017). Geographical features
705 are the predominant driver of molecular diversification in widely distributed North
706 American whipsnakes. *Molecular Ecology*, *26*(20), 5729–5751. doi:10.1111/mec.14295
707 Pelletier, T. A., & Carstens, B. C. (2018). Geographical range size and latitude predict
708 population genetic structure in a global survey. *Biology Letters*, *14*(1), 20170566.
709 Petkova, D., Novembre, J., & Stephens, M. (2015). Visualizing spatial population structure with
710 estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.
711 doi:10.1038/ng.3464
712 Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an
713 efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and*
714 *Evolution*, *31*(7), 1929–1936.
715 Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species

716 geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.
717 doi:http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026
718 Phillipsen, I. C., Kirk, E. H., Bogan, M. T., Mims, M. C., Olden, J. D., & Lytle, D. A. (2015).
719 Dispersal ability and habitat requirements determine landscape-level genetic patterns in
720 desert aquatic insects. *Molecular Ecology*, *24*(1), 54–69. doi:10.1111/mec.13003
721 Prevosti, A., Ocaña, J., & Alonso, G. (1975). Distances between populations of Drosophila
722 subobscura, based on chromosome arrangement frequencies. *Theoretical and Applied*
723 *Genetics*, *45*(6), 231–241. doi:10.1007/BF00831894
724 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
725 Multilocus Genotype Data. *Genetics*, *155*(2), 945–959.
726 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P.
727 C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based
728 Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.
729 doi:10.1086/519795
730 Pyron, R. A., & Burbrink, F. T. (2009). Systematics of the Common Kingsnake (Lampropeltis
731 getula; Serpentes: Colubridae) and the burden of heritage in taxonomy. *Zootaxa*, *2241*(27),
732 22–32.
733 Pyron, R. A., & Burbrink, F. T. (2010). Hard and soft allopatry: Physically and ecologically
734 mediated modes of geographic speciation. *Journal of Biogeography*, *37*(10), 2005–2015.
735 doi:10.1111/j.1365-2699.2010.02336.x
736 Pyron, R. A., & Burbrink, F. T. (2012). Extinction, ecological opportunity, and the origins of
737 global snake diversity. *Evolution*, *66*(1), 163–178. doi:10.1111/j.1558-5646.2011.01437.x
738 Reid, B. N., Mladenoff, D. J., & Peery, M. Z. (2017). Genetic effects of landscape, habitat
739 preference and demography on three co-occurring turtle species. *Molecular Ecology*, *26*(3),
740 781–798. doi:10.1111/mec.13962
741 Richmond, J. Q., Wood, D. A., Westphal, M. F., Vandergast, A. G., Leaché, A. D., Saslaw, L.
742 R., … Fisher, R. N. (2017). Persistence of historical population structure in an endangered
743 species despite near-complete biome conversion in California's San Joaquin Desert.
744 *Molecular Ecology*, *26*(14), 3618–3635. doi:10.1111/mec.14125
745 Riddle, B. R., & Hafner, D. J. (2006). A step-wise approach to integrating phylogeographic and

746 phylogenetic biogeographic perspectives on the history of a core North American warm
747 deserts biota. *Journal of Arid Environments*, *66*(3), 435–461.
748 doi:http://dx.doi.org/10.1016/j.jaridenv.2006.01.014
749 Robertson, J. M., Murphy, M. A., Pearl, C. A., Adams, M. J., Páez-Vacas, M. I., Haig, S. M., …
750 Funk, W. C. (2018). Regional variation in drivers of connectivity for two frog species (Rana
751 pretiosa and R. luteiventris) from the US Pacific Northwest. *Molecular Ecology*, *27*(16),
752 3242–3256.
753 Rodríguez-Robles, J. A., Bell, C. J., & Greene, H. W. (1999). Food habits of the glossy snake,
754 Arizona elegans, with comparisons to the diet of sympatric long-nosed snakes, Rhinocheilus
755 lecontei. *Journal of Herpetology*, 87–92.
756 Ruiz-Gonzalez, A., Cushman, S. A., Madeira, M. J., Randi, E., & Gómez-Moliner, B. J. (2015).
757 Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling
758 carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, *24*(20), 5110–5129.
759 Schield, D. R., Adams, R. H., Card, D. C., Corbin, A. B., Jezkova, T., Hales, N. R., … Smith, L.
760 L. (2018). Cryptic genetic diversity, population structure, and gene flow in the Mojave
761 Rattlesnake (Crotalus scutulatus). *Molecular Phylogenetics and Evolution*.
762 Schmidt Jr, R. H. (1979). A climatic delineation of the 'real'Chihuahuan Desert. *Journal of Arid*
763 *Environments*, *2*(3), 243–250.
764 Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or
765 distance: Which pattern of gene flow is most common? *Evolution*. doi:10.1111/evo.12258
766 Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological
767 speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*. doi:10.1111/ele.12120
768 Singhal, S., Huang, H., Grundler, M. R., Marchán-Rivadeneira, M. R., Holmes, I., Title, P. O.,
769 … Rabosky, D. L. (2018). Does Population Structure Predict the Rate of Speciation? A
770 Comparative Test across Australia's Most Diverse Vertebrate Radiation. *The American*
771 *Naturalist*, *192*(4), 432–447.
772 Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation.
773 *Evolution*. doi:10.1111/j.1558-5646.2009.00877.x
774 Thomassen, H. A., Buermann, W., Milá, B., Graham, C. H., Cameron, S. E., Schneider, C. J., …
775 Smith, T. B. (2010). Modeling environmentally associated morphological and genetic

variation in a rainforest bird, and its application to conservation prioritization. *Evolutionary Applications*, *3*(1), 1–16. doi:10.1111/j.1752-4571.2009.00093.x

Thuiller, W., Georges, D., & Engler, R. (2013). biomod2: Ensemble platform for species distribution modeling. *R Package Version*. doi:10.1017/CBO9781107415324.004

Vavrek, M. J. (2011). Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, *14*(1), 1T.

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935. doi:10.1046/j.1365-294X.2004.02076.x

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*. doi:10.1111/mec.12938

Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, *16*(2), 175–182. doi:10.1111/ele.12025

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, *33*(3), 607–611.

Wood, D. A., Vandergast, A. G., Barr, K. R., Inman, R. D., Esque, T. C., Nussear, K. E., & Fisher, R. N. (2013). Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Diversity and Distributions*, *19*(7), 722–737.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8041–8. doi:10.1073/pnas.1602237113

Zink, R. M., Kessen, A. E., Line, T. V, & Blackwell-Rago, R. C. (2001). Comparative phylogeography of some aridland bird species. *The Condor*, *103*(1), 1–10.

Data accessibility statement: The data generated during the current study are available from the NCBI Sequence Read Archive (Accession: PRJNA554495; see Supplemental Appendix 1 for individual specimens accession IDs) and the assembled GBS data used in this study are available

on Dryad (doi:10.5061/dryad.2172qg4). Outputs from ENMs are also available from Dryad (doi:10.5061/dryad.2172qg4).

Author Contributions: EAM and FTB conceived the study. EAM collected data, performed analyses, and wrote the manuscript. ATX and MG performed analyses. CC, ARDR, JLE, and JEMG collected data. All authors approved of the final manuscript.

Figure 1: Study system. A) The geographic distribution of the Sonoran and Chihuahuan Deserts in western North America. B) The major river systems of western North America. C) Elevation and the western continental divide. D) Climatic variation averaged across 32.0 and 32.01 latitude. This latitude corresponds to a transect from the Sonoran Desert through the Cochise Filter Barrier into the Chihuahuan Desert, the horizontal solid line represents the location of the Western Continental Divide. Data are from WorldClim (http://www.worldclim.org/). The *x*-axis is longitude and *y*-axes are environmental variables.

Figure 2: EEMS plots for all thirteen species. White areas indicate regions where migration rates are consistent with a pattern of IBD, highlighted blue regions have higher than expected rates of migration, and orange shaded regions have lower than expected rates of migration. Circles on each plot represent sampled localities. A. *Arizona elegans*; B. *Crotalus atrox*; C. *Crotalus molossus*; D. *Crotalus scutulatus*; E. *Hypsiglena torquata*; F. *Lampropeltis getula*; G. *Masticophis flagellum*; H. *Pituophis catenifer*; I. *Rhinochelius lecontei*; J. *Salvadora hexalepis*; K. *Sonora semiannulata*; L. *Thamnophis marcianus*; M. *Trimorphodon biscutatus*.

Figure 3: Sampling localities and populations inferred from clustering analyses in conStruct plotted over the distributions of each species (in gray) and the Western Continental Divide (in black, is often used to delineated the Cochise Filter Barrier). Also shown are representatives of some of the major lineages of snakes from this study. Each circle represents an individual sample, the color of the circle is representative of clustering results where the proportion of the color corresponds to the population assignment of that individual. A. *Arizona elegans*; B., O. *Crotalus atrox*; C. *Crotalus molossus*; D. *Crotalus scutulatus*; E. *Hypsiglena torquata*; F. *Lampropeltis getula*; G. *Masticophis flagellum*; H. *Pituophis catenifer*; I., P. *Rhinochelius lecontei*; J. *Salvadora hexalepis*; K. *Sonora semiannulata*; L., N. *Thamnophis marcianus*; M.

836 *Trimorphodon biscutatus*. Geographic distribution data was obtained from the IUCN website
837 (iucnredlist.org) for species A. – K, distributions for L. – M. were generated from locality
838 information downloaded from VertNet.

839 Table 1: Total number of samples and number of SNPs per taxon used in analyses.

Taxon	**Number of Samples**	**Total Number of SNPs**	**Number of Unlinked SNPs**	**Number of Samples Used For conStruct Analysis**	**Number of SNPs Used For conStruct Analysis**
A) *Arizona elegans*	43	18,993	7,438	37	599
B) *Crotalus atrox*	44	11,710	7,929	40	3,955
C) *Crotalus molossus*	20	15,245	7,784	20	650
D) *Crotalus scutulatus*	36	11,681	5,496	32	4,075
E) *Hypsiglena torquata*	27	27,202	6,857	25	599
F) *Lampropeltis getula*	35	12,219	8,236	34	3,622
G) *Masticophis flagellum*	30	14,443	5,901	29	4,610
H) *Pituophis catenifer*	41	13,264	6,351	37	4,466

I) *Rhinocheilus lecontei*	40	19,809	11,136	35	503
J) *Salvadora hexalepis*	15	32,154	18,291	14	2,584
K) *Sonora semiannulata*	13	37,607	21,259	12	4,988
L) *Thamnophis marcianus*	24	22,092	9,948	23	5,970
M) *Trimorphodon biscutatus*	15	46,444	21,073	14	3,251

841

842 Table 2: Results from generalized dissimilarity modeling analyses demonstrating the proportion of genomic divergence explained by
843 climate, geographic distance, and resistance surfaces, the statistical significance of the full GDM model, as well as the particular
844 variable that best explains genomic divergence in each of the full models and the most important climatic variable in predicting
845 ecological niche models. Model values in bold represent statistically significant ($p < 0.05$) values, those not in bold and followed by
846 (NS) are not statistically significant values.

Taxon	**GDM explained – IBD + IBE + IBR**	**GDM explained - Distance + Climate**	**GDM explained - Distance + Resistance**	**GDM explained - Climate + Resistance**	**GDM explained - Distance Only**	**GDM explained - Climate Only**	**GDM explained – Resistance Only**	**GDM Important Variables**	**ENM Most Important Variable**	**ENM AUC values**
A) *Arizona elegans*	**35.9**	**35.8**	**13.8**	**35.9**	**9**	**35.8**	**7.1**	Geographic Distance, Precipitation of Driest Month	Mean Temperature of Coldest Quarter	0.96
B) *Crotalus atrox*	**45.8**	**45.8**	**36.9**	**34.8**	**36.7**	**34.5**	1.28	Geographic Distance, Mean Temperature of Driest Quarter	Mean Temperature of Coldest Quarter	0.96

C) *Crotalus molossus*	**88.7**	**88.7**	**69**	**88.7**	**66.2**	**88.7**	N/A	Geographic Distance, Mean Temperature of Driest Quarter	Mean Temperature of Coldest Quarter	0.97
D) *Crotalus scutulatus*	**71.2**	**67.8**	**48.9**	**71.2**	**43.8**	**67**	**48.8**	Mean Diurnal Range, Annual Precipitation	Min Temperature of Coldest Month	0.96
E) *Hypsiglena torquata*	**70.9**	**70.9**	N/A	**70.5**	**29**	**70.4**	N/A	Annual Precipitation, Precipitation Seasonality	Mean Temperature of Driest Quarter	0.91
F) *Lampropeltis getula*	**37.9**	**34.5**	**36.7**	**36.5**	**30.3**	**33.3**	28.2	Geographic Distance, Elevation	Mean Temperature of Coldest Quarter	0.94
G) *Masticophis flagellum*	**67.8**	**67.5**	32.4	**66.7**	**31.5**	**66.2**	**27**	Geographic Distance, Mean Temperature of Driest Quarter	Annual Mean Temperature	0.92
H) *Pituophis*	**40.3**	**28.4**	**37.3**	**40.1**	**8.3**	**28.2**	**36.5**	Geographic	Mean	0.90

catenifer								Distance, ENM	Temperature of Warmest Quarter	
I) *Rhinocheilus lecontei*	**63**	**62.6**	**48.6**	**62.8**	**45.5**	**61.4**	**44.7**	Geographic Distance, Isothermality	Mean Temperature of Coldest Quarter	0.94
J) *Salvadora hexalepis*	**95.4**	**89.2**	**90**	**95.4**	N/A	**89.2**	**85**	Temperature Seasonality, Rivers	Annual Precipitation	0.96
K) *Sonora semiannulata*	**78.3**	**78.3**	**40.1**	**75.6**	**32.6**	**74.3**	**36.6**	Mean Temperature of Driest Quarter, Precipitation Seasonality	Temperature Seasonality	0.95
L) *Thamnophis marcianus*	**85**	**84.8**	**41.1**	**84.2**	**40.2**	**83.9**	**16.3**	Geographic Distance, Precipitation of Driest Month	Min Temperature of Coldest Month	0.96
M) *Trimorphodo*	**73.8**	**73.9**	**43.7**	73.3 (NS)	18.0 (NS)	**73.1**	**33**	Geographic Distance,	Precipitation of Driest	0.96

n biscutatus								Annual Precipitation	Month	

Supporting Information:

Appendix 1: Specimens used in these analyses, including the number of raw sequence reads per individual, and collecting latitude and longitude.

Appendix 2: Plots of geographic distance and genetic distances suggesting a strong pattern of isolation-by-distance in these data.

Appendix 3: Cross validation plots from conStruct analyses showing support for each value of K for both spatial and nonspatial clustering analyses. Results from spatial analyses are shown in red, which consistently outperform nonspatial analyses, shown in blue.

Appendix 4: Layer contribution for all possible tested layers in each taxon for both spatial and nonspatial analyses in conStruct. For each value of K, equal number of contributions are plotted.

Appendix 5: Alternative levels of *K*, inferred from conStruct, showing results from both spatial and nonspatial models.

Appendix 6: Projected ecological niche models for all 13 study species.

Appendix 7: Results of tests for environmental variation being explained by geographic distances.

Appendix 8: Variable importance values resulting from model permutations and statistical significance in GDM analyses.

Appendix 9: Results from GDM using Nei's D genetic distances as response variable.

Appendix 10: Plots of nucleotide diversity versus percent deviance explained and total number of genetic samples versus percent deviance explained in GDM models.