Predictors of genomic diversity within North American squamates

Ivy E. Larkin[1] (larkin.132@buckeyemail.osu.edu)

Edward A. Myers[2] (eddie.a.myers@gmail.com; https://orcid.org/0000-0001-5369-6329)

Bryan C. Carstens[1] (carstens.12@osu.edu; https://orcid.org/0000-0002-1552-227X)

Lisa N. Barrow[1,3*] (lnbarrow@unm.edu; https://orcid.org/0000-0001-7081-2432)

[1]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Ave, Columbus, OH 43210

[2]Department of Biological Sciences, Clemson University, Clemson, SC 29634

[3]Current Address: Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque NM 87131

*Corresponding author: lnbarrow@unm.edu

Running Title: GENOMIC DIVERSITY IN SNAKES

Abstract

Comparisons of intraspecific genetic diversity across species can reveal the roles of geography, ecology, and life history in shaping biodiversity. The wide availability of mitochondrial DNA (mtDNA) sequences in open-access databases makes this marker practical for conducting analyses across several species in a common framework, but patterns may not be representative of overall species diversity. Here, we gather new and existing mtDNA sequences and genome-wide nuclear data (genotyping-by-sequencing; GBS) for 30 North American squamate species sampled in the Southeastern and Southwestern United States. We estimated mtDNA nucleotide diversity for two mtDNA genes, COI (24 species alignments; average 16 sequences) and cytb (22 species; average 76 sequences), as well as nuclear heterozygosity and nucleotide diversity from GBS data for 119 individuals (30 species; four individuals and 6,820–44,309 loci per species). We showed that nuclear genomic diversity estimates were highly consistent across individuals for some species, while other species showed large differences depending on the locality sampled. Range size was positively correlated with both cytb diversity (Phylogenetically Independent Contrasts: $R^2 = 0.21$, $p = 0.03$) and GBS diversity ($R^2 = 0.19$; $p = 0.02$), while other predictors differed across the top models for each dataset. Mitochondrial and nuclear diversity estimates were not correlated within species, although sampling differences in the data available made these datasets difficult to compare. As new genome-wide datasets continue to be generated, mtDNA and nuclear diversity within species should be investigated to better understand the predictors of diversity across a variety of taxonomic groups.

Keywords: (3-6) genotyping-by-sequencing, mitochondrial DNA, nucleotide diversity, Serpentes, Florida, American Southwest

Introduction

Genetic information is key for investigating the evolutionary histories and demographic trajectories of populations and species. Since the advent of DNA sequencing in the late 1970s (Sanger et al. 1977), analysis of sequence variation has enabled biologists to discover species, determine relationships among lineages, reveal demographic histories, and provide assessments of population structure and diversity (Avise 2000; Hebert et al. 2003; Brito and Edwards 2009). Genetic diversity within species is especially critical for persistence in changing environments and is clearly of interest for global conservation (Laikre 2010; Laikre et al. 2020), but several challenges remain for gathering and comparing genetic information from large numbers of species. The extent to which genetic diversity is predictable given species characteristics such as range size or life history traits is an area of active interest (e.g., Leffler et al. 2012; Romiguier et al. 2014; Singhal et al. 2017; Pelletier and Carstens 2018; Barrow et al. 2021). These studies demonstrate that predictors often differ across different taxonomic groups and scales, prompting the need for additional investigations into understudied clades.

The type of genetic data chosen for comparative studies is important to consider. Most work to date has used widely available markers from organellar genomes, but the extent to which these data are representative of broader genomic patterns remains unclear. Mitochondrial DNA (mtDNA) is easily sequenced and has been used for decades, making it readily accessible from public databases for comparisons across large numbers of species (Zink and Barrowclough 2008; Miraldo et al. 2016). While certain characteristics of mtDNA, such as a high mutation rate and lack of

recombination, make it desirable for population-level studies and the inference of phylogenetic relationships at shallow scales (Avise et al. 1987), mtDNA may be poorly suited as a representation of genetic diversity (Galtier et al. 2009). Notably, several investigations comparing genetic diversity from mitochondrial and nuclear genomes within species have found conflicting results across different taxonomic groups (Bazin 2006; Mulligan et al. 2006; Singhal et al. 2017b). Bazin et al. (2006) showed that mtDNA did not correspond with expectations of population abundance when comparing across broad animal groups, e.g., invertebrate groups did not have higher mtDNA diversity than vertebrate groups, while nuclear sequence and allozyme datasets did meet this expectation. Comparisons within vertebrate groups such as mammals (Mulligan et al. 2006) and lizards (Singhal et al. 2017) have demonstrated a positive correlation between mtDNA and nuclear diversity estimates, however, suggesting that mtDNA is a useful marker for understanding patterns of genetic diversity in animals with smaller population sizes.

As a nonrecombinant unit that is matrilineally inherited, mtDNA represents a single coalescent history of sampled taxa, and it is well appreciated that any single genealogy may not accurately reflect population or species history (Hudson and Turelli 2003) particularly in cases where sex-biased dispersal or multiple divergences in a short timespan have occurred (Morin et al. 2004; Galtier et al. 2009). Sampling many nuclear loci enables more robust estimates of species relationships, population history, diversity estimates, and demographic parameters of interest (Edwards and Beerli 2000; Carling and Brumfield 2007). The growing availability of genome-scale nuclear datasets is promising for comparative studies (Garrick et al. 2015), but sampling many individuals

and populations for multiple species in a single study remains cost prohibitive. One solution implemented in previous comparative studies is to obtain genome-wide estimates of diversity from a single or few individuals as a representative of each species (e.g., Romiguier et al. 2014; Singhal et al. 2017; Grundler et al. 2019). It is not yet clear how consistent genomic diversity estimates are across populations of wide-ranging species that have experienced different histories across their ranges. We begin to address this question by evaluating genomic diversity from multiple individuals and localities within species in the present study.

In addition to considering appropriate sampling strategies, identifying potential predictors of genomic diversity within species is challenging because many species-level characteristics may correspond with diversity. Species with larger census population sizes are expected to have higher levels of neutral diversity (Kimura 1979). Total range size has been considered as a proxy for census population size (e.g., Leffler et al. 2012; Singhal et al. 2017) because species that occupy large areas are presumed to be more abundant, but range size and abundance can differ by several orders of magnitude and their relationship may be explained by a variety of mechanisms (Gaston et al. 1997). The size, shape, and characteristics of a range also impact how individuals move across the landscape, thus influencing rates of gene flow and whether population divergence has occurred (Wright 1943; Sexton et al. 2014; Myers et al. 2019).

Life history and ecological traits also correspond with genetic diversity across broad taxonomic groups (Romiguier et al. 2014; Chen et al. 2017). Body size is negatively correlated with genetic diversity and population structure for some groups

(e.g., mammals: Brüniche-Olsen et al. 2018; bees: López-Uribe et al. 2019; butterflies: Mackintosh et al. 2019), which could relate to limits on population size or to higher dispersal in larger-bodied species (White et al. 2007; Paz et al. 2015). Clutch size is positively correlated with genetic diversity across animals (Romiguier et al. 2014), where species that have many, small offspring have higher genetic diversity than long-lived species with few, large offspring. Aspects of mating system or reproductive mode have also been associated with genetic variation in different groups. In plants, genomic diversity is higher in outcrossing species compared to selfing species (Chen et al. 2017); in Panamanian frogs, species with direct development exhibit greater genetic structure compared to larval developing species (Paz et al. 2015); and in marine invertebrates, benthic direct-developing species have higher genetic structure than those with a pelagic larval stage (Collin 2001; Lee and Boulding 2009).

North American Squamates in Regions of Contrasting Topographic Complexity

This study focuses on the genetic diversity of squamates from the southern United States and north-central Mexico in two regions that were climatically suitable during the Last Glacial Maximum (LGM; Waltari et al. 2007), but differ in their topographic complexity (Badgley et al. 2017). We investigated species primarily from the North American Coastal Plain and the Chihuahuan Desert, both of which harbor high reptile diversity (Jenkins et al. 2015) and are currently under threat from habitat fragmentation, climate change, and change in wildfire regimes (Archer and Predick 2008; Noss et al. 2015; Briggs et al. 2020). The North American Coastal Plain includes the Gulf and Atlantic coasts in the southeastern United States and has been described

as a biodiversity hotspot (Noss et al. 2015). Although the region was never glaciated, when climatic and habitat shifts during Pleistocene glacial cycles occurred (Williams et al. 2004), the region likely provided stable refugial areas for many of the species in this study (Soltis et al. 2006; Weinell and Austin 2017). The Chihuahuan Desert occupies the southwestern United States and north-central Mexico and is considered among the 200 most biologically valuable ecoregions globally (Olson and Dinerstein 1998; Briggs et al. 2020). This region is also predicted to have provided refugial areas during the LGM (Waltari et al. 2007), but is more topographically complex with high environmental heterogeneity that plays an important role in population genetic structure and lineage divergence (Badgley et al. 2017; Myers et al. 2019). Both regions should retain much of the ancestral genetic diversity of the sampled species, since each is suspected to have harbored refugial populations during the Pleistocene. Furthermore, the high squamate diversity in these regions provides the opportunity to compare species with a variety of range sizes, body sizes, and life history traits.

In this study, we combined new and existing mtDNA and genome-scale nuclear sequence data for 30 North American squamate species to address three main questions. First, how consistent are nuclear genomic estimates across individuals and localities within species? Second, are measures of mtDNA and nuclear diversity within species correlated, suggesting that mtDNA is a useful proxy for within-species diversity? Third, are species geographic range size, body size, or life history traits (number of offspring and reproductive mode) associated with either mtDNA or nuclear diversity within squamates?

Methods

Sample Collection

We gathered and analyzed sequence data from 30 species representing three snake families (Viperidae, Colubridae, Natricidae) and two lizard families (Dactyloidae and Anguidae) (Fig. 1). Certain taxa (e.g., *Hypsiglena torquata*, *Trimorphodon biscutatus*, *Crotalus mitchelli*) were treated as species complexes, not because we disagree with previous species delimitation analyses of these groups, but because of uncertainty in the specific assignments of mtDNA sequence localities, range maps, and trait data available. In other cases where all data types were available for each species within a complex, we ensured sequences could be assigned to a single species based on their specific localities (e.g., *Pantherophis alleghaniensis*, *Pantherophis emoryi*). Tissue samples for 12 Southeastern U.S. species (4–7 individuals each) were collected in northern Florida and subsequently archived at the Museum of Southwestern Biology (Fig. 1; Table S1). We also included previously published data for 18 snake species distributed across the Southwestern U.S. (Myers et al. 2019, 2020; Myers 2021). For the nuclear dataset of these 18 species, we chose four individuals with a similar sampling strategy as the Southeastern U.S. species. Briefly, we chose two individuals each from two localities per species when possible, focusing on the Chihuahuan desert and a similar distance (~200 km) between localities.

Range Maps and Trait Data

Species range maps were downloaded as shapefiles from the IUCN Red List of Threatened Species Version 6.1 (www.iucnredlist.org/). Maps were edited in QGIS 3.12

as needed to exclude non-native ranges and combine taxa treated as species complexes. The R packages ‘rgdal’ (Bivand et al. 2021) and ‘geosphere’ (Karney 2013; Hijmans 2021) were used to visualize ranges and calculate total range area in kilometers squared. Species trait information was obtained primarily from Burbrink and Myers (2015), including maximum body size (log-transformed), average clutch size (log-transformed), and parity (viviparous or oviparous). Trait data for the two lizard species were obtained from field guides and online natural history accounts (Powell et al. 2016; https://www.virginiaherpetologicalsociety.com; https://animaldiversity.org).

Mitochondrial Sequences

For the 12 Southeastern U.S. species, genomic DNA was extracted from tail or liver tissue with the E.Z.N.A. DNA Tissue Kit (Omega Bio-Tek) according to the manufacturer’s instructions. We amplified a 658-base pair portion of the mtDNA cytochrome c oxidase 1 gene (COI) using primers ReptBCF_COI and ReptBCR_COI (Castañeda and de Queiroz 2011). The thermal profile included an initial denaturation step at 94° C for 3 minutes, 35 cycles of denaturing at 94° for 45 seconds, annealing at 53° for 30 seconds, elongating at 72° for 60 seconds, and a final extension step at 72° for 10 minutes. PCR products were visualized on agarose gels to ensure successful amplification, cleaned with EXOSAP-IT (Affymetrix Inc.), and sequenced at The Ohio State University (OSU) Comprehensive Cancer Center Genomics Shared Resource. Species-specific sequencing primers were designed when the original PCR primers did not perform adequately (Table S2). Additional mtDNA sequences for two genes, COI and cytochrome b (cytb), were obtained from NCBI GenBank for all 30 species with

data available. Sequences for each species were aligned in Geneious 2020.1.1 (https://www.geneious.com) using MAFFT 7.450 (Katoh and Standley 2013).

Nuclear Data Collection

Genotyping-by-sequencing (GBS) libraries were prepared following a modified version of the protocol described in Elshire et al. (2011). For four samples per Southeastern U.S. species (48 individuals), we digested 100 ng of input DNA with the enzyme Pst1, ligated 8 µL of adapter mix including unique barcodes, pooled libraries, and performed a bead cleanup with Sera-Mag Speedbeads (Rohland and Reich 2012). Final PCR amplification was conducted with 16 cycles in 8 replicate reactions, followed by pooling, bead cleanup, and quantification via a Qubit fluorometer (Life Technologies). Libraries were size selected to 200-500 bp using a Blue Pippen (Sage Science Inc.), quantified with a Bioanalyzer (Agilent Technologies), and sequenced at the OSU Comprehensive Cancer Center Genomics Shared Resource on an Illumina HiSeq 4000 with 150 bp paired-end sequencing. For four samples per Southwestern U.S. Species (72 individuals), we downloaded GBS reads from the NCBI Sequence Read Archive (Table S1).

Sequenced GBS reads from all 30 species were assembled through the ipyrad 0.7.28 pipeline (Eaton and Overcast 2020). After demultiplexing (no mismatches allowed in barcodes), the four individuals per species were assembled to generate within-species datasets. We trimmed reads for the Southeastern U.S. species to 125 bp prior to assembly. We used the de novo assembly method with a clustering similarity threshold of 85%, maximum clustering depth of 10000 reads, maximum of five low

quality base calls per read, minimum depth for base calling of six reads, strict filtering for adapters, and we retained reads longer than 35 bp. We allowed a maximum of eight heterozygous sites in consensus sequences, a maximum of 20 SNPs per locus, and required all four individuals (no missing data) to be included in a locus to retain that locus in output files.

Genetic Diversity Metrics

Measures of mtDNA diversity were calculated for each species using the R package 'pegas' (Paradis 2010). We calculated nucleotide diversity (pi or π Nei 1987) for COI and cytb alignments of each species with the nuc.div function. Nuclear diversity within species was estimated from each GBS dataset using the R packages 'adegenet' (Jombart 2008; Jombart and Ahmed 2011) and 'PopGenome' (Pfeifer et al. 2014). Expected heterozygosity (Hs) was computed from the unique SNPs outfile from ipyrad for each species using 'adegenet'. To calculate π, 'PopGenome' reads in a directory of FASTA files including an alignment for each locus; we used a custom Python script to generate this directory for each species using the alleles outfiles (full sequences) from ipyrad. We calculated π within each individual (two alleles per individual) and then determined the mean and standard deviation across individuals, hereafter referred to as "mean within-individual π". For comparison, we also calculated π within each species from the alignments including four individuals (eight alleles) per species, hereafter "within-species π". Species-level and within-population nuclear diversity can be estimated from a single individual when sufficient loci are sampled (e.g., Chen et al. 2017; Grundler et al. 2019). Where possible, we included two individuals per species

from each sampled locality to investigate how consistent genomic diversity estimates are across different individuals and localities within species.

Statistical Analyses

We compared mtDNA (COI π, cytb π) and nuclear (Hs, within-species π) diversity estimates within species using pairwise correlations in R. We then constructed phylogenetic generalized least square (PGLS) models to assess the importance of different predictors on genetic diversity, while accounting for the possibility that closely related species may share similar genomic diversity and species traits. We downloaded 100 phylogeny subsets for the 30 species in our study from VertLife.org, which includes the squamate relationships from Tonini et al. (2016), and we used the first tree from the posterior distribution for our analyses. We constructed a model set for each of the four responses: mtDNA COI π, mtDNA cytb π, nuclear Hs, and nuclear within-species π. For each response, predictors of interest included geographic range size, maximum body size, average clutch size, and parity. To assess effects of sample size on diversity metrics, we included the number of individuals sampled as a covariate in mtDNA models, and the number of GBS loci as a covariate for nuclear models.

Analyses were conducted using the R packages ‘ape’ (Paradis and Schliep 2019), ‘geiger’ (Harmon et al. 2008), and ‘nlme’ (Pinheiro et al. 2022). We constructed PGLS models using two models of evolution, Brownian Motion (BM) and Ornstein-Uhlenbeck (OU). The BM model assumes traits evolve according to a random walk, while the OU model incorporates a trait optimum towards which traits are pulled (Felsenstein 1985; Martins and Hansen 1997). For each model set, all combinations of

the predictors of interest were included and models were ranked based on the corrected Akaike Information Criterion (AICc). We ran model selection analyses using the R package 'MuMIn' (Bartón 2022) to determine which model of evolution had a better fit. We then constructed models incorporating all combinations of the predictors of interest and ran model selection analyses to assess which sets of predictors were included in the best-fit models for each diversity measure. We further assessed relationships between continuous variables using Spearman's rank correlation tests, phylogenetically independent contrasts implemented in 'ape' (Felsenstein 1985), and linear models.

Results

Data Summary

We generated 62 new COI sequences (GenBank accession numbers ON911378–ON911439) and combined these with available sequences on GenBank. Of the 30 squamates in our study, we were able to analyze COI alignments for 24 species (4–69 sequences per alignment; average 16) and cytb alignments for 22 species (5–136 sequences; average 76). Three species (*Crotalus oreganus*, *Hypsiglena torquata*, *Trimorphodon biscutatus*) did not have sufficient mtDNA data available for either COI or cytb (Table 1). We sequenced new GBS data for 47 individuals from 12 species (NCBI SRA ###–###) and calculated genomic diversity metrics for 119 individuals from 30 species. One *Thamnophis sirtalis* sample had poor sequencing success and was removed from subsequent analyses. The assembled GBS datasets included 2,923–25,086 loci for the unique SNP datasets and 6,820–44,309 loci for the full sequence datasets depending on the species (Table 1).

Mitochondrial and Nuclear Diversity

Mitochondrial diversity within species ranged from 0.0025 to 0.0812 for COI π and from 0.0005 to 0.0717 for cytb π. Diversity estimates for the two mtDNA genes were correlated within species (Spearman's rho (ρ) = 0.825, p = 8.147e-06). For the nuclear GBS data, within-species heterozygosity ranged from 0.3 to 0.346 and within-species π ranged from 0.0012 to 0.0083. Mean within-individual π ranged from 0.00076 to 0.0056 and there was substantial variation among individuals and localities for some species (Fig. 1). For species such as *Thamnophis marcianus* and *Masticophis flagellum*, the within-individual π estimates were consistent between the two individuals within a locality, but values for one locality were nearly double the values for the second locality. For other species such as *Crotalus adamanteus* and *Sonora semiannulata*, the within-individual π estimates were highly consistent for all four individuals. For still other species such as *Thamnophis sauritus* and *Lampropeltis getula*, there was large variation in π estimates even for individuals collected from the same locality. The within-species π estimates were strongly correlated with the mean within-individual π estimates (Spearman's ρ = 0.986, p < 2.2e-16), and we used the within-species estimate to represent overall species nuclear diversity for subsequent analyses.

Metrics of diversity for mtDNA and nuclear GBS data were not correlated for the species represented in both datasets (Fig. 2). mtDNA COI π was not correlated with nuclear heterozygosity (Spearman's ρ = 0.201, p = 0.345) or within-species π (Spearman's ρ = -0.104, p = 0.626). Similarly, mtDNA cytb π was not correlated with

nuclear heterozygosity (Spearman's ρ = -0.0096, p = 0.968) or within-species π (Spearman's ρ = -0.082, p = 0.717).

Predictors of Squamate Genetic Diversity

The OU model of evolution was the better fit for all four response variables (model weights > 0.986; Table 2). Different sets of predictors were included in the top models for the four response variables (Table 3). For mtDNA COI, body size and clutch size were consistently included in the top models, but their relationship with diversity was not significant with or without phylogenetic correction (Table 4; Supplementary Fig. S1). For mtDNA cytb, range size and parity were always included in the top models. Within-species cytb diversity had a positive relationship with range size (Fig. 3), suggesting that species with larger ranges tend to have higher mtDNA diversity. This relationship was consistent for both uncorrected (Spearman's ρ = 0.56; p = 0.007; Fig. 3) and phylogenetically corrected datasets (phylogenetically independent contrasts: R^2 = 0.21; p = 0.032; Supplementary Fig. S2). Oviparous species had higher cytb diversity compared to viviparous species (Fig. 3; t-test: t = 3.35, p = 0.003).

Models with different sets of predictors for nuclear heterozygosity were not distinguishable from one another and no predictors were significant (Table 3). For nuclear within-species π, however, range size and clutch size were consistently included in the top models. Nuclear within-species π had a positive relationship with both range size and average clutch size (Table 4; Fig. 3). The relationship between nuclear π and range size was consistent for both uncorrected (Spearman's ρ = 0.41; p = 0.024; Fig. 3) and phylogenetically corrected datasets (PIC: R^2 = 0.19; p = 0.016;

Supplementary Fig. 2). The relationship between nuclear π and average clutch size was significant for the uncorrected data (Spearman's $\rho = 0.54$; $p = 0.002$; Fig. 3), but not based on phylogenetically independent contrasts ($R^2 = 0.027$; $p = 0.382$; Supplementary Fig. S2).

Discussion

We compared genetic diversity estimates from mitochondrial and nuclear genomes for 30 North American squamate species. Our analyses demonstrated that range size is positively correlated with mtDNA and nuclear diversity in squamates and that life history traits are inconsistent predictors of diversity for these datasets. We found that mtDNA and nuclear diversity estimates were not correlated, but we note that several aspects of sampling made these datasets difficult to compare. Here, we discuss sampling considerations for this and future studies, including the number of individuals, localities, species, and loci available or feasible to sample. Despite these challenges, we provide empirical evidence that geography is a key element for understanding patterns of genomic diversity within and across species.

Comparative population genomics is a developing field that aims to uncover the forces underlying genomic diversity by comparing across multiple species (Ellegren and Galtier 2016; Edwards et al. 2022). Studies typically focus on the roles of selection, neutrality, and recombination in regulating genetic variation in a primarily non-geographic context, thus sampling usually consists of a single or few individuals to represent each species (e.g. (Romiguier et al. 2014; Chen et al. 2017). In contrast, comparative phylogeography takes a geography-centric view to understanding historical

processes underlying current genetic variation within species and lineages. A third emerging field, macrogenetics, focuses on an even broader geographic and taxonomic scale, typically using repurposed data from relatively few genetic markers to investigate large-scale patterns and predictors of genetic variation with species (Leigh et al. 2021). Ideally, species have been sampled across their entire range and as densely as possible, often revealing the presence of multiple lineages within nominal species (e.g., Schield et al. 2018). Until recently, genome-wide datasets were not common in phylogeographic studies (Garrick et al. 2015) and sampling many species thoroughly across geographic space remains cost-prohibitive for a single study. A pressing question for designing comparative studies to investigate determinants of genomic diversity remains: How representative are genomic diversity estimates from a few individuals or from a narrow part of a species' range? Our GBS data indicates that it depends – certain species (e.g., *Crotalus adamanteus*) had highly consistent genomic diversity estimates from localities sampled ~200 km apart, while others (e.g., *Thamnophis marcianus*) had two-fold differences in genomic diversity between localities (Fig. 1). This result indicates that using a single individual or locality to represent diversity for an entire species does not always provide a complete picture of diversity within species. We aimed to sample individuals from historically stable parts of species' ranges to better represent species-level diversity, thus it may seem surprising that such differences are already emerging. We predict even greater differences when populations are sampled from topographically complex regions or in areas with recent range expansion, a consideration that should continue to be addressed in future studies.

Our study provides an initial comparison of squamate genomic diversity from five families sampled in two species-rich regions, the Southwestern and Southeastern U.S. We found considerable variation in diversity among species within some families (e.g., Colubridae and Natricidae; Fig. 1), but further taxon sampling is needed for robust comparisons of genomic diversity across families. In general, the Southwestern U.S. has more heterogenous environments and landscapes compared to the Southeast, and Florida in particular, where populations were sampled (Noss et al. 2015; Badgley et al. 2017). Thus far, we did not detect any clear differences in average genomic diversity levels between the two regions, although several Southwestern species had high variation in genomic diversity among individuals and localities, an expected result given the high heterogeneity of the region (Myers et al. 2019).

One challenge that we could not fully address given the data available was how to consistently treat diversity estimates from known species complexes. When possible, we sampled individuals from within a single lineage to ensure that mtDNA sequences could be assigned correctly based on locality, that range sizes could be calculated based on the known range for that lineage, and so that other traits could be attributed to that lineage. However, in some cases, recent taxonomic revisions and limited geographic information about existing sequences made it impossible to assign data to a single lineage, and we treated that species as a complex instead. We acknowledge that these choices will influence our comparisons across species, if for example, range sizes are larger for those treated as species complexes, or mtDNA sequences for some species represent more genetic structure than others. We provide the alignments and GenBank accession numbers for the data we analyzed

(https://doi.org/10.5061/dryad.qfttdz0ks) such that they can be reassigned and reanalyzed in follow-up studies. Future efforts to link geographic coordinates with genetic sequences in open access databases will continue to improve the prospects of comparative studies (Pelletier et al. 2022).

We found that mtDNA and GBS diversity were not correlated, suggesting that mtDNA may not be a reliable proxy for within-species diversity in North American squamates. Several differences in our sampling for mtDNA and nuclear data made these datasets difficult to compare, however. The number of mtDNA sequences available for some species was limited, producing COI and cytb alignments for only 24 and 22 species, respectively. These mtDNA sequences most likely represent a much broader geographic area within each species compared to our GBS data, though it is not possible to quantify the influence of geographic distance because sample localities were not associated with the majority of GenBank sequences (Marques et al. 2013). On the other hand, the GBS data consist of several thousand loci and provide a better representation of genome-wide diversity within species compared to a single mtDNA gene. These differences in power and geographic representation, along with relatively few data points for comparison (22–24 species), may explain the lack of correlation between mtDNA and nuclear diversity metrics in our study. Previous studies have found a relationship between mtDNA and nuclear diversity in some taxonomic groups including mammals (47 species; Mulligan et al. 2006) and Australian lizards (60 species; Singhal et al. 2017). Further study of this topic is important for comparative studies because mtDNA is readily available for many more species, but reduced

representation nuclear genome datasets are increasing and have the potential to provide new insights on overall species diversity.

The approaches used to obtain and analyze nuclear genome-wide datasets are important to consider for comparative studies of genetic diversity. One appealing characteristic of mtDNA data for data repurposing and comparing species sequenced across different studies is the ease of aligning known genes. Restriction site associated methods such as GBS provide many more loci for analysis, but without an annotated reference genome for most species of interest, these loci are anonymous. Furthermore, the locus assemblies and resulting parameters are highly dependent on the parameter settings of the pipeline used for analysis (Harvey et al. 2016; Shafer et al. 2017). Our study used existing squamate datasets from the American Southwest that we reassembled to ensure the same ipyrad version and parameters were used for comparison with new GBS data for the Southeastern U.S. species. We found that heterozygosity values were highly sensitive to the parameter settings used. Prior to reassembly, heterozygosity estimates for the Southwestern species were an order of magnitude smaller than those of the Southeastern species (Table S3). These contrasting results demonstrate the additional challenges of comparing GBS and similar datasets collected from different studies. Targeted sequence capture approaches may be preferable to increase comparability across studies, though they remain more expensive and require more initial effort to design probe kits (Harvey et al. 2016; Singhal et al. 2017a; Hutter et al. 2022).

Range size was positively correlated with nucleotide diversity within squamate species for both mtDNA (cytb) and nuclear data, despite differences in geographic and

locus sampling. Previous studies have found mixed results in a variety of taxonomic groups and with different sampling strategies. For example, in a global study including single-gene sequences from more than 8,000 eukaryotic species found that total range size was one of the most important predictors of genetic structure within species (Pelletier and Carstens 2018). In contrast, (Romiguier et al. 2014) analyzed transcriptome data from 2-10 individuals of 76 animal species and found no correlation between genomic diversity and range size. Detailed focus within taxonomic groups demonstrated correlations between range size and genetic diversity in *Drosophila* (Leffler et al. 2012), Australian lizards (Singhal et al. 2017b), and cetaceans (Vachon et al. 2018), but not in butterflies (Mackintosh et al. 2019). To our knowledge, we provide the first evidence that total range size may be a useful predictor of snake genomic diversity. Future studies including additional species, biogeographic regions, and nuclear sampling across species' ranges would provide important insights about the generality of these findings and the potential influence of fluctuations in population size on genomic variation.

The life history and ecological traits we included were less consistent predictors of mitochondrial and nuclear diversity in squamates. Body size has negative associations with genetic diversity in various taxonomic groups including mammals (Brüniche-Olsen et al. 2018), frogs (Paz et al. 2015), and insects (López-Uribe et al. 2019; Mackintosh et al. 2019), but our results do not provide compelling evidence for squamates (Fig. S1). Our results for clutch size are consistent with the prediction that species with larger clutch sizes harbor higher genomic diversity (e.g., Romiguier et al. 2014), though the relationships were somewhat weak (Fig. S2). These patterns will

require further scrutiny with increased species sampling in future studies. The potential association of parity with mitochondrial diversity, where oviparous (egg-laying) species had higher diversity on average than viviparous (live-bearing) species, is intriguing and should also be investigated further with more comprehensive sampling. The evolution of viviparity is associated with colder climates and faster development times in cool temperatures (Ma et al. 2018), but our study lacked sampling across broad latitudinal or climatic gradients. Furthermore, all oviparous species included in this study belong to the family Colubridae, but viviparity in squamates has evolved from oviparity at least 34 and perhaps more than 100 times (Pyron and Burbrink 2014; Blackburn 2015). Squamates therefore present an excellent opportunity to continue investigating how life history traits and geographic range characteristics influence intraspecific genomic diversity.

Funding

This work was supported by The Ohio State University (OSU) Undergraduate Research Scholarship awarded to IEL and by the OSU President's Postdoctoral Scholars Program through an award to LNB. Computational resources were provided by the Ohio Supercomputer Center via a grant to BCC (PAA0202). LNB and BCC were supported via grants from the National Science Foundation (LNB: DEB-2112946 and BCC: DBI-1910623).

Acknowledgements

We thank Moses Michelsohn for providing tissues from Florida and Megan Smith for laboratory and analytical support. Specimens were obtained according to the appropriate collecting permits and animal care protocols.

Data Availability

We have deposited the primary data underlying these analyses as follows:

- Sampling locations, field numbers, museum catalogue numbers, and datasets used for analyses: Dryad link - https://doi.org/10.5061/dryad.qfttdz0ks
- DNA sequences: GenBank accession numbers ON911378–ON911439; NCBI SRA: ###

References

Archer S.R., Predick K.I. 2008. Climate Change and Ecosystems of the Southwestern United States. Rangelands. 30:23–28.

Avise J.C. 2000. Phylogeography: the history and formation of species. Cambridge, Massachussetts: Harvard University Press.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. 1987. Intraspecific Phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics. 18:489–522.

Badgley C., Smiley T.M., Terry R., Davis E.B., DeSantis L.R.G., Fox D.L., Hopkins S.S.B., Jezkova T., Matocq M.D., Matzke N., McGuire J.L., Mulch A., Riddle B.R., Roth V.L., Samuels J.X., Strömberg C.A.E., Yanites B.J. 2017. Biodiversity and Topographic Complexity: Modern and Geohistorical Perspectives. Trends in Ecology & Evolution. 32:211–226.

Barrow L.N., Masiero da Fonseca E., Thompson C.E.P., Carstens B.C. 2021. Predicting amphibian intraspecific diversity with machine learning: Challenges and prospects for integrating traits, geography, and genetic data. Molecular Ecology Resources. 21:2818–2831.

Bartón K. 2022. MuMIn: Multi-Model Inference. R package version 1.46.0. https://CRAN.R-project.org/package=MuMIn. .

Bazin E. 2006. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. Science. 312:570–572.

Bivand R., Keitt T., Rowlingson B. 2021. rgdal: Bindings for the “Geospatial” Data Abstraction Library. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal. .

Blackburn D.G. 2015. Evolution of viviparity in squamate reptiles: Reversibility reconsidered. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 324:473–486.

Briggs M.K., Lozano-Cavazos E.A., Mills Poulos H., Ochoa-Espinoza J., Rodriguez-Pineda J.A. 2020. The Chihuahuan Desert: A Binational Conservation Response to Protect a Global Treasure. Encyclopedia of the World’s Biomes. Elsevier. p. 126–138.

Brito P.H., Edwards S.V. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica. 135:439–455.

Brüniche-Olsen A., Kellner K.F., Anderson C.J., DeWoody J.A. 2018. Runs of homozygosity have utility in mammalian conservation and evolutionary studies. Conservation Genetics. 19:1295–1307.

Burbrink F.T., Myers E.A. 2015. Both traits and phylogenetic history influence community structure in snakes over steep environmental gradients. Ecography. 38:1036–1048.

Carling M.D., Brumfield R.T. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity. PLoS ONE. 2:e160–e160.

Castañeda M. del R., de Queiroz K. 2011. Phylogenetic relationships of the Dactyloa clade of Anolis lizards based on nuclear and mitochondrial DNA sequence data. Molecular Phylogenetics and Evolution. 61:784–800.

Chen J., Glémin S., Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. Molecular Biology and Evolution. 34:1417–1428.

Collin R. 2001. The effects of mode of development on phylogeography and population structure of North Atlantic Crepidula (Gastropoda: Calyptraeidae). Molecular Ecology. 10:2249–2262.

Eaton D.A.R., Overcast I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics. 36:2592–2594.

Edwards S.V., Beerli P. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution. 54:1839–1854.

Edwards S.V., Robin V.V., Ferrand N., Moritz C. 2022. The Evolution of Comparative Phylogeography: Putting the Geography (and More) into Comparative Population Genomics. Genome Biology and Evolution. 14:evab176.

Ellegren H., Galtier N. 2016. Determinants of genetic diversity. Nat Rev Genet. 17:13.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 6:1–10.

Felsenstein J. 1985. Phylogenies and the comparative method. American Naturalist. 125:1–15.

Galtier N., Nabholz B., Glémin S., Hurst G.D.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology. 18:4541–4550.

Garrick R.C., Bonatelli I.A.S., Hyseni C., Morales A., Pelletier T.A., Perez M.F., Rice E., Satler J.D., Symula R.E., Thomé M.T.C., Carstens B.C. 2015. The evolution of phylogeographic datasets. Molecular Ecology. 24:1164–1171.

Gaston K.J., Blackburn T.M., Lawton J.H. 1997. Interspecific Abundance-Range Size Relationships: An Appraisal of Mechanisms. Journal of Animal Ecology. 66:579–601.

Grundler M.R., Singhal S., Cowan M.A., Rabosky D.L. 2019. Is genomic diversity a useful proxy for census population size? Evidence from a species-rich community of desert lizards. Molecular Ecology.:1664–1674.

Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics. 24:129–131.

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Systematic Biology. 65:910–924.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proc Biol Sci. 270:313–321.

Hijmans R.J. 2021. geosphere: Spherical Trigonometry. R package version 1.5-14. https://CRAN.R-project.org/package=geosphere. .

Hudson R.R., Turelli M. 2003. Stochasticity Overrules the “Three-Times Rule”: Genetic Drift, Genetic Draft, and Coalescence Times for Nuclear Loci Versus Mitochondrial Dna. Evolution. 57:182–190.

Hutter C.R., Cobb K.A., Portik D.M., Travers S.L., Wood Jr. P.L., Brown R.M. 2022. FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. Molecular Ecology Resources. 22:1100–1119.

Jenkins C.N., Houtan K.S.V., Pimm S.L., Sexton J.O. 2015. US protected lands mismatch biodiversity priorities. PNAS. 112:5081–5086.

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

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

Karney C.F.F. 2013. Algorithms for geodesics. J Geod. 87:43–55.

Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 30:772–780.

Kimura M. 1979. The Neutral Theory of Molecular Evolution. Scientific American. 241:98–129.

Laikre L. 2010. Genetic diversity is overlooked in international conservation policy implementation. Conserv Genet. 11:349–354.

Laikre L., Hoban S., Bruford M.W., Segelbacher G., Allendorf F.W., Gajardo G., Rodríguez A.G., Hedrick P.W., Heuertz M., Hohenlohe P.A., Jaffé R., Johannesson K., Liggins L., MacDonald A.J., OrozcoterWengel P., Reusch T.B.H., Rodríguez-Correa H., Russo I.-R.M., Ryman N., Vernesi C. 2020. Post-2020 goals overlook genetic diversity. Science. 367:1083–1085.

Lee H.J., Boulding E.G. 2009. Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods: the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. Molecular Ecology. 18:2165–2184.

Leffler E.M., Bullaughey K., Matute D.R., Meyer W.K., Ségurel L., Venkat A., Andolfatto P., Przeworski M. 2012. Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? PLoS Biology. 10.

Leigh D.M., van Rees C.B., Millette K.L., Breed M.F., Schmidt C., Bertola L.D., Hand B.K., Hunter M.E., Jensen E.L., Kershaw F., Liggins L., Luikart G., Manel S., Mergeay J., Miller J.M., Segelbacher G., Hoban S., Paz-Vinas I. 2021. Opportunities and challenges of macrogenetic studies. Nat Rev Genet. 22:791–807.

López-Uribe M.M., Jha S., Soro A. 2019. A trait-based approach to predict population genetic structure in bees. Mol Ecol. 28:1919–1929.

Ma L., Buckley L.B., Huey R.B., Du W.-G. 2018. A global test of the cold-climate hypothesis for the evolution of viviparity of squamate reptiles. Global Ecology and Biogeography. 27:679–689.

Mackintosh A., Laetsch D.R., Hayward A., Charlesworth B., Waterfall M., Vila R., Lohse K. 2019. The determinants of genetic diversity in butterflies. Nature Communications. 10:1–9.

Marques A.C., Maronna M.M., Collins A.G. 2013. Putting GenBank Data on the Map. Science. 341:1341–1341.

Martins E.P., Hansen T.F. 1997. Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. The American Naturalist. 149:646–667.

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

Morin P.A., Luikart G., Wayne R.K., the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution. 19:208–216.

Mulligan C.J., Kitchen A., Miyamoto M.M. 2006. Comment on “Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals.” Science. 314:1390–1390.

Myers E.A. 2021. Genome-wide data reveal extensive gene flow during the diversification of the western rattlesnakes (Viperidae: Crotalinae: Crotalus). Molecular Phylogenetics and Evolution. 165:107313.

Myers E.A., McKelvy A.D., Burbrink F.T. 2020. Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). Mol Ecol. 29:797–811.

Myers E.A., Xue A.T., Gehara M., Cox C.L., Davis Rabosky A.R., Lemos-Espinal J., Martínez-Gómez J.E., Burbrink F.T. 2019. Environmental heterogeneity and not vicariant biogeographic barriers generate community-wide population structure in desert-adapted snakes. Mol Ecol. 28:4535–4548.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Noss R.F., Platt W.J., Sorrie B.A., Weakley A.S., Means D.B., Costanza J., Peet R.K. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Diversity Distrib. 21:236–244.

Olson D.M., Dinerstein E. 1998. The Global 200: A Representation Approach to Conserving the Earth's Most Biologically Valuable Ecoregions. Conservation Biology. 12:502–515.

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

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Paz A., Ibáñez R., Lips K.R., Crawford A.J. 2015. Testing the role of ecology and life history in structuring genetic variation across a landscape: A trait-based phylogeographic approach. Molecular Ecology. 24:3723–3737.

Pelletier T.A., Carstens B.C. 2018. Geographical range size and latitude predict population genetic structure in a global survey. Biol. Lett. 14:20170566.

Pelletier T.A., Parsons D.J., Decker S.K., Crouch S., Franz E., Ohrstrom J., Carstens B.C. 2022. phylogatR: Phylogeographic data aggregation and repurposing. Molecular Ecology Resources. n/a.

Pfeifer B., Wittelsbuerger U., Ramos-Onsins S.E., Lercher M.J. 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. Molecular Biology and Evolution. 31:1929–1936.

Pinheiro J., Bates D., DebRoy S., Sarkar D., R Core Team. 2022. nlme: Linear and Nonlinear Mixed Effects Models. .

Powell R., Conant R., Collins J.T. 2016. Peterson Field Guide to Reptiles and Amphibians of Eastern and Central North America 4th Edition. Boston, MA: Houghton Mifflin Harcourt.

Pyron R.A., Burbrink F.T. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecol Lett. 17:13–21.

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

Romiguier J., Gayral P., Ballenghien M., Bernard A., Cahais V., Chenuil A., Chiari Y., Dernat R., Duret L., Faivre N., Loire E., Lourenco J.M., Nabholz B., Roux C., Tsagkogeorga G., Weber A.A.T., Weinert L.A., Belkhir K., Bierne N., Glémin S., Galtier N. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 515:261–263.

Sanger F., Nicklen S., Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 74:5463–5467.

Schield D.R., Adams R.H., Card D.C., Corbin A.B., Jezkova T., Hales N.R., Meik J.M., Perry B.W., Spencer C.L., Smith L.L., García G.C., Bouzid N.M., Strickland J.L., Parkinson

C.L., Borja M., Castañeda-Gaytán G., Bryson R.W., Flores-Villela O.A., Mackessy S.P., Castoe T.A. 2018. Cryptic genetic diversity, population structure, and gene flow in the Mojave rattlesnake (Crotalus scutulatus). Molecular Phylogenetics and Evolution. 127:669–681.

Sexton J.P., Hangartner S.B., Hoffmann A.A. 2014. Genetic isolation by environment or distance: Which pattern of gene flow is most common? Evolution. 68:1–15.

Shafer A.B.A., Peart C.R., Tusso S., Maayan I., Brelsford A., Wheat C.W., Wolf J.B.W. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution. 8:907–917.

Singhal S., Grundler M., Colli G., Rabosky D.L. 2017a. Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. Molecular Ecology Resources. 17:e12–e24.

Singhal S., Huang H., Title P.O., Donnellan S.C., Holmes I., Rabosky D.L. 2017b. Genetic diversity is largely unpredictable but scales with museum occurrences in a species-rich clade of Australian lizards. Proceedings of the Royal Society B: Biological Sciences. 284.

Soltis D.E., Morris A.B., McLachlan J.S., Manos P.S., Soltis P.S. 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology. 15:4261–4293.

Tonini J.F.R., Beard K.H., Ferreira R.B., Jetz W., Pyron R.A. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation. 204:23–31.

Vachon F., Whitehead H., Frasier T.R. 2018. What factors shape genetic diversity in cetaceans? Ecology and Evolution. 8:1554–1572.

Waltari E., Hijmans R.J., Peterson a T., Nyári A.S., Perkins S.L., Guralnick R.P. 2007. Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PloS ONE. 2:e563–e563.

Weinell J.L., Austin C.C. 2017. Refugia and Speciation in North American Scarlet Snakes (*Cemophora*). Journal of Herpetology. 51:161–171.

White E.P., Ernest S.K.M., Kerkhoff A.J., Enquist B.J. 2007. Relationships between body size and abundance in ecology. Trends in Ecology and Evolution. 22:323–330.

Williams J.W., Shuman B.N., Webb T., Bartlein P.J., Leduc P.L. 2004. Late-Quaternary vegetation dynamics in north America: Scaling from taxa to biomes. Ecological Monographs. 74:309–334.

Wright S. 1943. Isolation by Distance. Genetics. 28:114–138.

Zink R.M., Barrowclough G.F. 2008. Mitochondrial DNA under siege in avian phylogeography: nuDNA VS. mtDNA in phylogeography. Molecular Ecology. 17:2107–2121.

Table and Figure Captions

Table 1. Sample sizes for mtDNA and GBS datasets for each species.

Table 2. Model comparison of PGLS models under the Brownian Motion (BM) or Ornstein-Uhlenbeck (OU) models of evolution.

Table 3. Model comparison of PGLS models with different sets of predictors. The top three models for each response are shown, which all had a delta AIC value of <= 2. Note that none of the models for Heterozygosity were substantially improved over others (delta AIC all < 4). The plus ("+") sign in the predictor columns indicates that predictor was included in that model.

Table 4. Model coefficients, standard errors (SE), t, and p-values for the top model listed in Table 3 for each response.

Figure 1. Individual sample localities and nucleotide diversity estimates for species included in the study. (a) Map of all samples for which genotyping-by-sequencing data were analyzed. Southwestern samples are shown in turquoise and southeastern species in brown. Dotted black boxes indicate the inset maps to the right, which depict localities sampled within two species as examples. Shapes indicate individuals were sampled from the same locality within a species, but do not necessarily indicate the

same locality across different species. (b) Phylogenetic relationships for the 30 species included in the study based on Tonini et al. (2016). Within-individual nucleotide estimates are shown to the right with colors and shapes corresponding to the maps in (a).

Figure 2. Comparison of mtDNA and nuclear diversity estimates within species. Spearman's rank correlation sample estimates and p-values are shown.

Figure 3. Predictors of mtDNA and nuclear diversity within species. Scatter plots and dotted lines depict simple linear relationships without phylogenetic correction. For additional predictors and phylogenetically corrected relationships, see Supplementary Figures 1 and 2.

Table 1. Sample sizes for mtDNA and GBS datasets for each species.

Species	Family	# COI seqs	# cytb seqs	# GBS loci (unique SNPs)	# GBS loci (sequences)
Ophisaurus ventralis	Anguidae	6	NA	22951	42875
Anolis carolinensis	Dactyloidae	27	NA	15708	28909
Agkistrodon piscivorus	Viperidae	12	116	5961	15540
Sistrurus miliarius	Viperidae	10	31	14356	23527
Crotalus molossus	Viperidae	5	60	9047	29251
Crotalus atrox	Viperidae	5	66	6926	19307
Crotalus scutulatus	Viperidae	4	74	14696	29502
Crotalus oreganus	Viperidae	NA	NA	11032	27615
Crotalus viridis	Viperidae	NA	8	15565	29627
Crotalus tigris	Viperidae	5	NA	9003	37903
Crotalus mitchellii	Viperidae	6	NA	3274	6820
Crotalus adamanteus	Viperidae	10	126	8424	27387
Nerodia fasciata	Natricidae	30	143	15130	25148
Thamnophis marcianus	Natricidae	NA	22	7467	17648
Thamnophis sauritus	Natricidae	7	NA	14540	17945
Thamnophis sirtalis	Natricidae	7	37	15664	25409
Hypsiglena torquata	Colubridae	NA	NA	19304	29950
Sonora semiannulata	Colubridae	8	44	2923	12181
Trimorphodon biscutatus	Colubridae	NA	NA	5566	26644
Salvadora hexalepis	Colubridae	8	31	3513	20475
Coluber flagellum	Colubridae	55	136	7512	18350
Coluber constrictor	Colubridae	51	22	18737	22267
Lampropeltis getula	Colubridae	11	273	6485	19095
Cemophora coccinea	Colubridae	69	9	15437	25907
Rhinocheilus lecontei	Colubridae	9	105	4713	15022
Arizona elegans	Colubridae	8	92	16879	37596
Pantherophis alleghaniensis	Colubridae	12	43	14352	26136
Pituophis catenifer	Colubridae	9	98	18205	32161
Pantherophis emoryi	Colubridae	NA	129	25086	44309
Pantherophis guttatus	Colubridae	11	11	14377	25517

Table 2. Model comparison of PGLS models under the Brownian Motion (BM) or Ornstein-Uhlenbeck (OU) models of evolution.

Response	Model	log Likelihood	AIC	ΔAIC	Weight
mtDNA COI π	OU	63.617	-111.2	0.00	0.998
	BM	56.158	-98.3	12.92	0.002
mtDNA cytb π	OU	63.818	-111.6	0.00	0.986
	BM	58.579	-103.2	8.48	0.014
Heterozygosity	OU	97.865	-179.7	0.00	1
	BM	88.002	-162.0	17.73	0
Within-species π	OU	157.317	-298.6	0.00	0.993
	BM	151.426	-288.9	9.78	0.007

Table 3. Model comparison of PGLS models with different sets of predictors. The top three models for each response are shown, which all had a delta AIC value of <= 2. Note that none of the models for Heterozygosity was substantially improved over others (delta AIC all < 4). The plus ("+") sign in the predictor columns indicates that predictor was included in that model.

Response	df	AIC	ΔAIC	Weight	Range size	Body size	Clutch size	Parity
mtDNA COI π (OU models)	8	-111.2	0.00	0.230	+	+	+	+
	7	-110.6	0.62	0.169		+	+	+
	7	-109.2	2.00	0.084	+	+	+	
mtDNA cytb π (OU models)	7	-113.3	0.00	0.402	+		+	+
	8	-111.6	1.65	0.176	+	+	+	+
	6	-111.5	1.81	0.163	+			+
Heterozygosity (OU models)	5	-183.4	0.00	0.136		+		
	5	-183.2	0.24	0.121				+
	5	-183.0	0.45	0.109			+	
Within-species π (OU models)	7	-300.2	0.00	0.455	+	+	+	
	6	-298.6	1.60	0.205	+		+	
	8	-298.6	1.61	0.203	+	+	+	+

Table 4. Model coefficients, standard errors (SE), t-, and p-values for the top model listed in Table 3 for each response.

Response	Coefficient	Value	SE	t	p-value
mtDNA COI π (OU model)	Intercept	-0.1012	0.0696	-1.4532	0.1634
	log(Range size)	0.0068	0.0047	1.4395	0.1672
	log(Body size)	0.0427	0.0200	2.1287	0.0473
	log(Clutch size)	-0.0534	0.0257	-2.0766	0.0524
	Parity_vivip	-0.0161	0.0089	-1.8074	0.0874
	N sequences	-0.0001	0.0003	-0.5354	0.5989
mtDNA cytb π (OU model)	Intercept	-0.0945	0.0645	-1.4655	0.1610
	log(Range size)	0.0108	0.0043	2.5183	0.0221*
	log(Clutch size)	-0.0322	0.0179	-1.7916	0.0910
	Parity_vivip	-0.0176	0.0069	-2.5499	0.0207*
	N sequences	0.0001	0.0001	1.7387	0.1001
Heterozygosity (OU model)	Intercept	0.3376	0.0175	19.2391	0.0000
	log(Body size)	-0.0064	0.0080	-0.7925	0.4350
	N loci	-0.0000	0.0000	-0.2187	0.8285
Within-species π (OU model)	Intercept	-0.0111	0.0047	-2.3341	0.0279
	log(Range size)	0.0010	0.0003	3.4052	0.0022*
	log(Body size)	-0.0023	0.0013	-1.7844	0.0865
	log(Clutch size)	0.0053	0.0017	3.1684	0.0040*
	N loci	0.0000	0.0000	0.3581	0.7233

Figure 1. Individual sample localities and nucleotide diversity estimates for species included in the study. (a) Map of all samples for which genotyping-by-sequencing data were analyzed. Southwestern samples are shown in turquoise and southeastern species in brown. Dotted black boxes indicate the inset maps to the right, which depict localities sampled within two species as examples. Shapes indicate individuals were sampled from the same locality within a species. (b) Phylogenetic relationships for the 30 species included in the study based on Tonini et al. (2016). Within-individual nucleotide estimates are shown to the right with colors and shapes corresponding to the maps in (a).

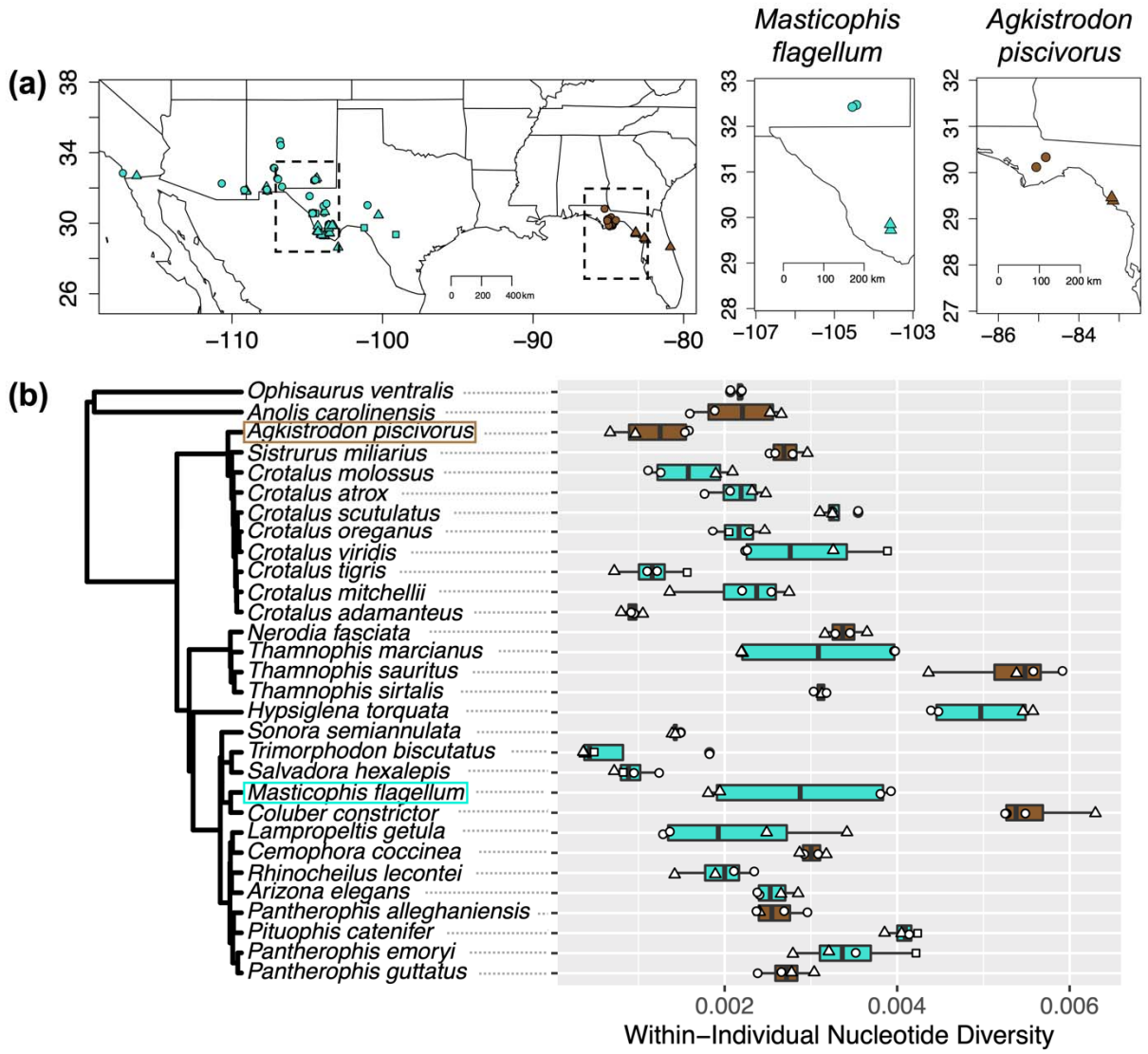


Figure 2. Comparison of mtDNA and nuclear diversity estimates within species. Spearman's rank correlation sample estimates and p-values are shown.

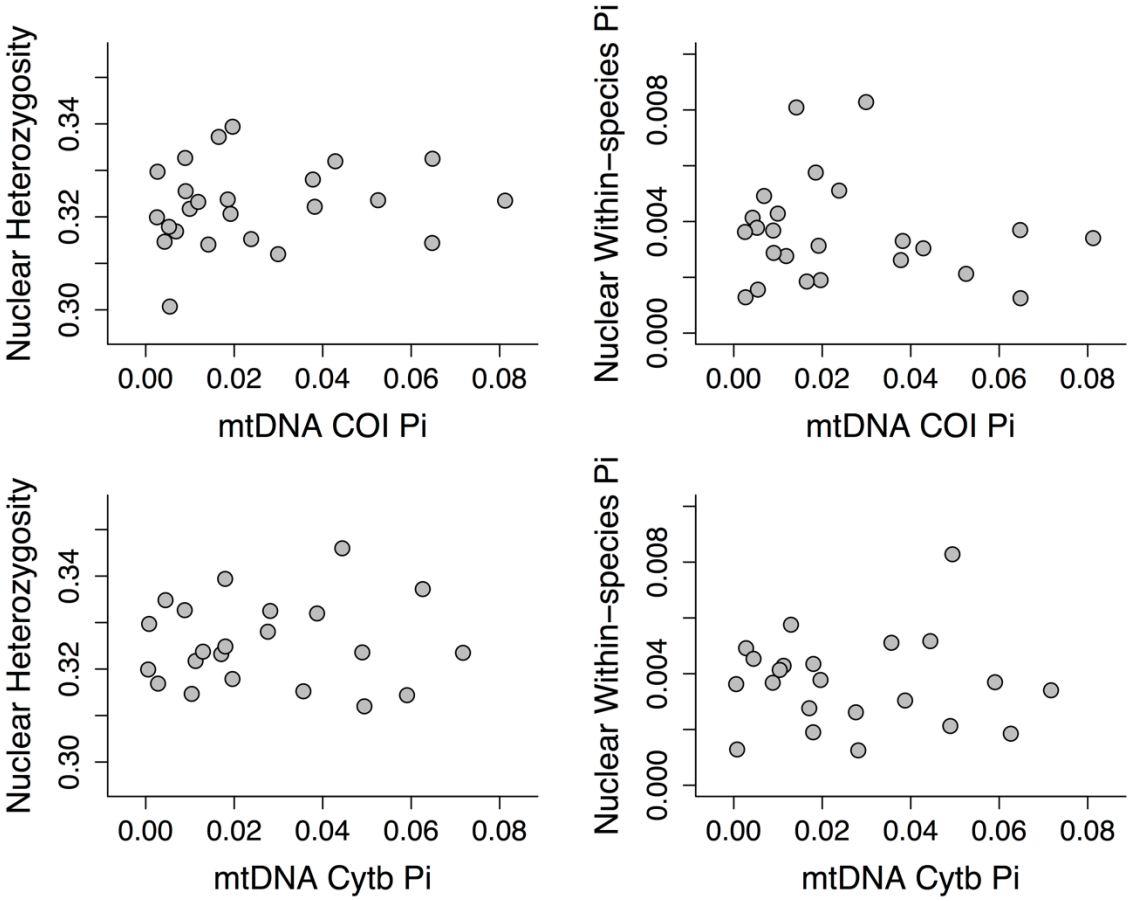
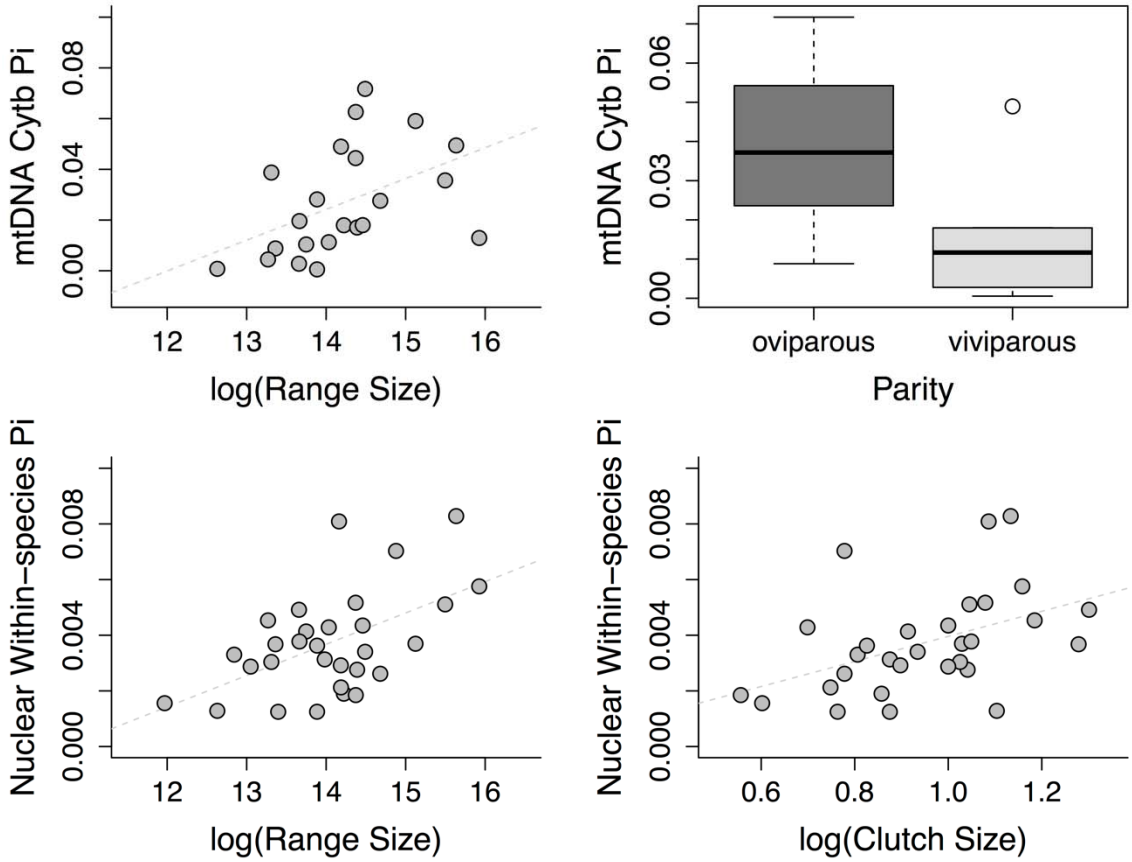


Figure 3. Predictors of mtDNA and nuclear diversity within species. Scatter plots and dotted lines depict simple linear relationships without phylogenetic correction. For additional predictors and phylogenetically corrected relationships, see Supplementary Figures 1 and 2.



Supplementary Materials

Table S1. List of all samples for which mtDNA COI was sequenced and GBS data were analyzed. Sample names, museum catalogue numbers, geographic coordinates, locality information, and total number of GBS reads sequenced are included.

[provided as Excel sheet – tab 1]

Table S2. Additional species-specific sequencing primers designed in this study to improve COI sequencing results.

[provided as Excel sheet – tab 2]

Table S3. Data for 30 squamate species used as input for analyses.

[provided as Excel sheet – tab 3 and with PGLS R script in Dryad package]

Table S4. Data for 119 squamate individuals used to make figures and summaries.

[provided as Excel sheet – tab 4 and with R script in Dryad package]

Additional files on Dryad (https://doi.org/10.5061/dryad.qfttdz0ks): Mitochondrial (COI and cytb) alignments including previous GenBank samples; GBS files; Range maps; Python/R scripts

Figure S1. Uncorrected and phylogenetically corrected plots of body size and mtDNA COI diversity.

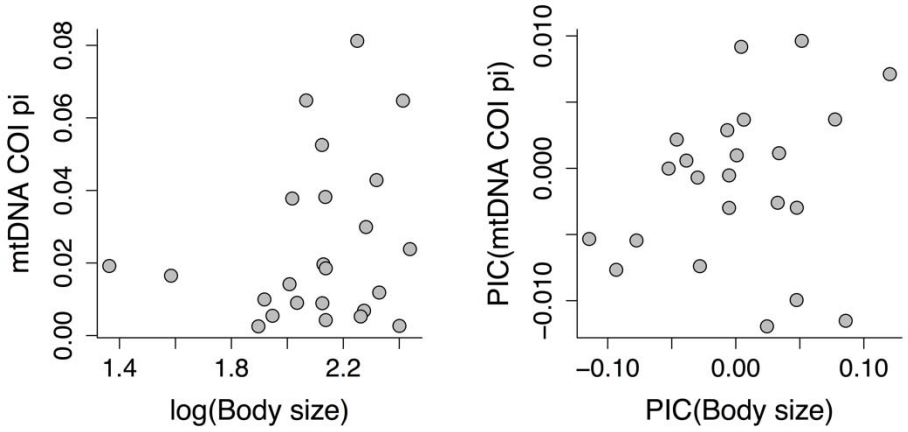
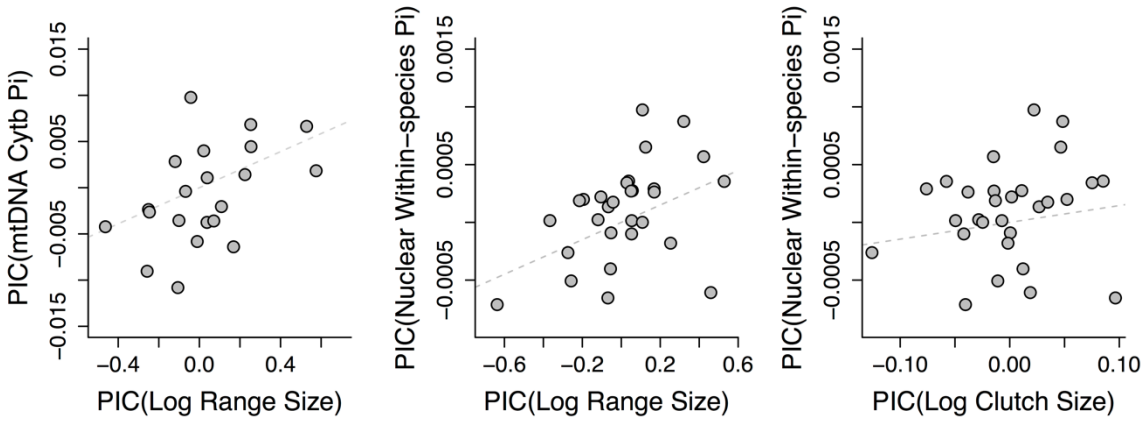



Figure S2. Relationships between predictors and genetic diversity based on phylogenetically independent contrasts.