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 (22 species alignments; average 16 sequences) and cytb (22 species; average 58 sequences), as well as nuclear heterozygosity and nucleotide diversity from GBS data for 118 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: $R^2 = 0.31$, $p = 0.007$) and GBS diversity ($R^2 = 0.21$; $p = 0.006$), 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. Further study of mtDNA and nuclear diversity sampled across species' ranges is needed to evaluate the roles of geography and life history in structuring 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

67 may be poorly suited as a representation of genomic diversity in a given species (Galtier et al.
68 2009). As a nonrecombinant unit that is matrilineally inherited, mtDNA represents a single
69 realization of many possible coalescent histories of sampled taxa, and it is well appreciated that
70 any single genealogy may not accurately reflect population or species history (Hudson and
71 Turelli 2003) particularly in cases where sex-biased dispersal or multiple divergences in a short
72 timespan have occurred (Morin et al. 2004; Galtier et al. 2009). Notably, several investigations
73 comparing genetic diversity from mitochondrial and nuclear genomes within species have
74 found conflicting results across different taxonomic groups (Bazin 2006; Mulligan et al. 2006;
75 Singhal et al. 2017b). Bazin et al. (2006) showed that mtDNA did not correspond with
76 expectations of population abundance when comparing across broad animal groups, e.g.,
77 invertebrate groups did not have higher mtDNA diversity than vertebrate groups, while nuclear
78 sequence and allozyme datasets did meet this expectation. Comparisons within vertebrate
79 groups such as mammals (Mulligan et al. 2006) and lizards (Singhal et al. 2017) have
80 demonstrated a positive correlation between estimates of mtDNA and nuclear diversity,
81 however, suggesting that mtDNA may be a useful marker for understanding patterns of genetic
82 diversity in animals with smaller population sizes. Taken together, the contrasting results of
83 these studies also illustrate how the taxonomic scale of a given study can lead to different
84 findings.
85 Sampling many nuclear loci enables more robust estimates of species relationships,
86 population history, diversity estimates, and demographic parameters of interest (Edwards and
87 Beerli 2000; Carling and Brumfield 2007). The growing availability of genome-scale nuclear
88 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 (but see Nazareno et al. (2017) for an example in plant populations separated by 20 km). We begin to address this question in squamates 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 within populations, diversity between populations (i.e., genetic structure), or both. Species with larger census population sizes are expected to have larger effective population sizes and therefore higher levels of neutral genetic diversity within populations (Kimura 1979). This assumed relationship may also extend to total range size, which 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 locally abundant. The abundance-range size relationship may be explained by a variety of mechanisms (Gaston et al. 1997). In analyses of some taxonomic groups, population density appears to be unrelated to range size (Novosolov et al. 2017), while empirical examples of island and mainland bird species have demonstrated the expected positive relationship between genomic diversity and range size (Brüniche-Olsen et al. 2019; Leroy et al. 2021). Furthermore, the size, shape, and characteristics of a species' range can impact how individuals

111 move across the landscape, thus influencing rates of dispersal, gene flow, and the maintenance
112 of genetic structure and overall diversity within species (Wright 1943; Sexton et al. 2014; Myers
113 et al. 2019).

114 Life history and ecological traits also correspond with genetic diversity across broad
115 taxonomic groups (Romiguier et al. 2014; Chen et al. 2017). Body size is negatively correlated
116 with genetic diversity within and between populations for some groups (e.g., mammals:
117 Brüniche-Olsen et al. 2018; bees: López-Uribe et al. 2019; butterflies: Mackintosh et al. 2019),
118 which could relate to limits on population size (leading to lower within-population diversity) or
119 to higher dispersal (leading to lower genetic structure) in larger-bodied species (White et al.
120 2007; Paz et al. 2015). Fecundity, or clutch size, is also expected to relate to abundance, where
121 species that have more offspring will have larger and more stable population sizes through time
122 and therefore higher neutral diversity (Kimura 1979). Indeed, clutch size is positively correlated
123 with genetic diversity across animals (Romiguier et al. 2014), where species that have many,
124 small offspring have higher within-population genetic diversity than long-lived species with few,
125 large offspring. Aspects of mating system or reproductive mode have also been associated with
126 genetic variation in different groups. In plants, within-population genomic diversity is higher in
127 outcrossing species compared to selfing species (Chen et al. 2017), which is expected because
128 inbreeding reduces effective population size.

129 Different reproductive strategies may also lead to differences in dispersal distance and
130 resulting genetic structure. For example, in Panamanian frogs, species with direct development
131 exhibit greater genetic structure compared to larval developing species (Paz et al. 2015); and in
132 marine invertebrates, benthic direct-developing species have higher genetic structure than

133 those with a pelagic larval stage (Collin 2001; Lee and Boulding 2009). Another aspect of
134 reproduction that has not been thoroughly addressed is parity mode, whether species lay eggs
135 (oviparous) or have live young (viviparous). Oviparity could be associated with higher dispersal,
136 and less genetic structure, if females must travel long distances to locate suitable nest sites,
137 while viviparous species have reduced movements because of increased energetic costs (Shine
138 2015). In the common lizard (*Zootoca vivipara*), greater genetic structure was found in an
139 oviparous lineage compared to a viviparous lineage, but in this case, the pattern could be
140 explained by differences in demographic history (Cornetti et al. 2015). Comparisons of species-
141 wide genetic diversity across multiple oviparous and viviparous species are lacking thus far.
142
143 *North American Squamates in Regions of Contrasting Topographic Complexity*
144 This study focuses on the genetic diversity of squamates from North America in two
145 regions that were climatically suitable during the Last Glacial Maximum (LGM; Waltari et al.
146 2007), but differ in their topographic complexity (Badgley et al. 2017). We investigated species
147 from the North American Coastal Plain and the Desert Southwest (Fig. 1), both of which harbor
148 high reptile diversity (Jenkins et al. 2015) and are currently under threat from habitat
149 fragmentation, climate change, and change in wildfire regimes (Archer and Predick 2008; Noss
150 et al. 2015; Briggs et al. 2020). The North American Coastal Plain includes the Gulf and Atlantic
151 coasts in the southeastern United States and has been described as a biodiversity hotspot (Noss
152 et al. 2015). Although the region was never glaciated, when climatic and habitat shifts during
153 Pleistocene glacial cycles occurred (Williams et al. 2004), the region likely provided stable
154 refugial areas for many of the species in this study (Soltis et al. 2006; Weinell and Austin 2017).

155 Within the Desert Southwest, located in the southwestern United States and north-central
156 Mexico, the Chihuahuan Desert is the most biologically diverse desert and is considered among
157 the 200 most biologically valuable ecoregions globally (Olson and Dinerstein 1998; Briggs et al.
158 2020). This region is also predicted to have provided refugial areas during the LGM (Waltari et
159 al. 2007), but is more topographically complex with high environmental heterogeneity that
160 plays an important role in population genetic structure and lineage divergence (Badgley et al.
161 2017; Myers et al. 2019). Both regions should retain much of the ancestral genetic diversity of
162 the sampled species since each is suspected to have harbored refugial populations during the
163 Pleistocene. Furthermore, the high squamate diversity in these regions provides the
164 opportunity to compare species with a variety of range sizes, body sizes, and life history traits.
165 In this study, we combined new and existing mtDNA and genome-scale nuclear
166 sequence data for 30 North American squamate species to address three main questions. First,
167 how consistent are nuclear genomic estimates across individuals and localities within species?
168 Second, are measures of mtDNA and nuclear diversity within species correlated, suggesting that
169 mtDNA is a useful proxy for within-species diversity? Third, are species geographic range size,
170 body size, or life history traits (number of offspring and reproductive mode) associated with
171 either mtDNA or nuclear diversity within squamates?
172
173 **Methods**
174 *Sample Collection*
175 We gathered and analyzed sequence data from 30 species representing four snake
176 families (Viperidae, Colubridae, Natricidae, Dipsadidae) and two lizard families (Anolidae and

177 Anguidae) (Fig. 1). We followed the most recent taxonomic revisions in the literature, ensuring
178 mtDNA sequences could be assigned to a single species based on their specific localities, and
179 excluding sequences without locality information (Table S1). Tissue samples for 12 Coastal Plain
180 species (4–7 individuals each) were collected in northern Florida from 2009 to 2018, preserved
181 in either 95% ethanol or DMSO tissue buffer, frozen at or below -20 °C, and subsequently
182 archived at the Museum of Southwestern Biology (Fig. 1; Table S2). We also included previously
183 published data for 18 snake species distributed across the Desert Southwest (Myers et al. 2019,
184 2020; Myers 2021). For the nuclear dataset of these 18 species, we chose four individuals with
185 a similar sampling strategy as the Coastal Plain species. Briefly, we chose two individuals each
186 from two localities per species when possible, focusing primarily on the Chihuahuan Desert and
187 a similar distance (~200 km) between localities. This standardized distance enabled us to make
188 comparisons among species without the potentially confounding effects of geographic distance
189 on genetic structure.
190
191 *Range Maps and Trait Data*
192 Species range maps were downloaded as shapefiles from the IUCN Red List of
193 Threatened Species Version 6.1 (www.iucnredlist.org/). Maps were edited in QGIS 3.12 as
194 needed to exclude non-native parts of the range where species have been introduced and to
195 edit species ranges according to recent taxonomic changes (Table S1). The R packages 'rgdal'
196 (Bivand et al. 2021) and 'geosphere' (Karney 2013; Hijmans 2021) were used to visualize ranges
197 and calculate total range area in kilometers squared. Species trait information was obtained
198 primarily from Burbrink and Myers (2015), including maximum body size (log-transformed),

199 average clutch size (log-transformed), and parity (viviparous or oviparous). Trait data for the
200 two lizard species were obtained from field guides and online natural history accounts (Powell
201 et al. 2016; https://www.virginiaherpetologicalsociety.com; https://animaldiversity.org).
202
203 *Mitochondrial Sequences*
204 For the 12 Coastal Plain species, genomic DNA was extracted from tail or liver tissue
205 with the E.Z.N.A. DNA Tissue Kit (Omega Bio-Tek) according to the manufacturer's instructions.
206 We amplified a 658-base pair portion of the mtDNA cytochrome c oxidase 1 gene (COI) using
207 primers ReptBCF_COI and ReptBCR_COI (Castañeda and de Queiroz 2011). The thermal profile
208 included an initial denaturation step at 94° C for 3 minutes, 35 cycles of denaturing at 94° for 45
209 seconds, annealing at 53° for 30 seconds, elongating at 72° for 60 seconds, and a final extension
210 step at 72° for 10 minutes. PCR products were visualized on agarose gels to ensure successful
211 amplification, cleaned with EXOSAP-IT (Affymetrix Inc.), and sequenced at The Ohio State
212 University (OSU) Comprehensive Cancer Center Genomics Shared Resource. Species-specific
213 sequencing primers were designed when the original PCR primers did not perform adequately
214 (Table S3). Additional mtDNA sequences for two genes, COI and cytochrome b (cytb), were
215 obtained from NCBI GenBank for species with data available. Sequences for each species were
216 aligned in Geneious 2020.1.1 (https://www.geneious.com) using MAFFT 7.450 (Katoh and
217 Standley 2013).
218
219 *Nuclear Data Collection*

220 Genotyping-by-sequencing (GBS) libraries were prepared following a modified version of
221 the protocol described in Elshire et al. (2011). For four samples per Coastal Plain species (48
222 individuals), we digested 100 ng of input DNA with the enzyme Pst1, ligated 8 µL of adapter mix
223 including unique barcodes, pooled libraries, and performed a bead cleanup with Sera-Mag
224 Speedbeads (Rohland and Reich 2012). Final PCR amplification was conducted with 16 cycles in
225 8 replicate reactions, followed by pooling, bead cleanup, and quantification via a Qubit
226 fluorometer (Life Technologies). Libraries were size selected to 200-500 bp using a Blue Pippen
227 (Sage Science Inc.), quantified with a Bioanalyzer (Agilent Technologies), and sequenced at the
228 OSU Comprehensive Cancer Center Genomics Shared Resource on an Illumina HiSeq 4000 with
229 150 bp paired-end sequencing. For four samples per Desert Southwest species (72 individuals),
230 we downloaded GBS reads from the NCBI Sequence Read Archive (Table S2).
231 Sequenced GBS reads from all 30 species were assembled through the ipyrad 0.7.28
232 pipeline (Eaton and Overcast 2020) with the following settings. After demultiplexing (no
233 mismatches allowed in barcodes), the R1 reads for the four individuals per species were
234 assembled to generate within-species datasets. We trimmed reads for the Southeastern U.S.
235 species to 125 bp prior to assembly using the built-in ipyrad option which uses the software
236 tool 'cutadapt' (trim_reads = 0, -25). We used the de novo assembly method with a clustering
237 similarity threshold of 85% (clust_threshold = 0.85), maximum clustering depth of 10000 reads
238 (maxdepth = 10000), maximum of five low quality base calls per read (max_low_qual_bases = 5;
239 with Q<20), minimum depth for base calling of six reads (mindepth_statistical = 6 and
240 mindepth_majrule = 6, strict filtering for adapters (filter_adapters = 2), and we retained reads
241 longer than 35 bp after trimming adapters (filter_min_trim_len = 35). We allowed a maximum

242 of eight heterozygous sites in consensus sequences (max_Hs_consens = 5), a maximum of 20
243 SNPs per locus (max_SNPs_locus = 20) and required all four individuals (no missing data) to be
244 included in a locus to retain that locus in output files (min_samples_locus = 4).
245
246 *Genetic Diversity Metrics*
247 Measures of mtDNA diversity were calculated for each species using the R package
248 'pegas' (Paradis 2010). We calculated nucleotide diversity (pi or π, Nei 1987) for COI and cytb
249 alignments of each species with the nuc.div function. Nuclear diversity within species was
250 estimated from each GBS dataset using the R packages 'adegenet' (Jombart 2008; Jombart and
251 Ahmed 2011) and 'PopGenome' (Pfeifer et al. 2014). Expected heterozygosity (Hs) was
252 computed from the unique SNPs outfile from ipyrad for each species using 'adegenet'. To
253 calculate π, 'PopGenome' reads in a directory of FASTA files including an alignment for each
254 locus; we used a custom Python script (available on Dryad) to generate this directory for each
255 species using the alleles outfiles (full sequences) from ipyrad. We calculated π within each
256 sample (two alleles per individual) and then determined the mean and standard deviation
257 across individuals, hereafter referred to as "mean within-individual π". For comparison, we also
258 calculated π within each species from the alignments including four individuals (eight alleles)
259 per species, hereafter "within-species π". Species-level and within-population nuclear diversity
260 can be estimated from a single individual when sufficient loci are sampled (e.g., Chen et al.
261 2017; Grundler et al. 2019). Where possible, we included two individuals per species from each
262 sampled locality to investigate how consistent genomic diversity estimates are across different
263 individuals and localities within species.

264

265 *Statistical Analyses*

266 We compared mtDNA (COI π, cytb π) and nuclear (mean within-individual π, within-
267 species π) diversity estimates within species using pairwise correlations in R. We then
268 constructed phylogenetic generalized least square (PGLS) models to assess the importance of
269 different predictors on genetic diversity, while accounting for the possibility that closely related
270 species may share similar genomic diversity and species traits. We downloaded 100 phylogeny
271 subsets for the 30 species in our study from VertLife.org, which includes the squamate
272 relationships from Tonini et al. (2016), and generated a least-squares consensus tree using
273 phytools (Revell 2012) for our analyses. We constructed a model set for each of three
274 responses: mtDNA COI π, mtDNA cytb π, and nuclear within-species π. For each response,
275 predictors of interest included geographic range size, maximum body size, average clutch size,
276 and parity. To assess effects of sample size on diversity metrics, we included the number of
277 individuals sampled as a covariate in mtDNA models, and the number of GBS loci as a covariate
278 for nuclear models. We included the region (Coastal Plain or Desert Southwest) in our initial
279 PGLS models for nuclear within-species π and used Welch's two-sample t-tests to check for
280 differences between the GBS datasets generated in this study and those generated previously.

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

286 each model set, all combinations of the predictors of interest were included and models were
287 ranked based on the corrected Akaike Information Criterion (AICc). We ran model selection
288 analyses using the R package 'MuMIn' (Bartón 2022) to determine which model of evolution
289 had a better fit. We then constructed models incorporating all combinations of the predictors
290 of interest and ran model selection analyses to assess which sets of predictors were included in
291 the best-fit models for each diversity measure. We further assessed relationships between
292 continuous variables using Spearman's rank correlation tests, phylogenetically independent
293 contrasts implemented in 'ape' (Felsenstein 1985), and linear models.
294
295 **Results**
296 *Data Summary*
297 We generated 62 new COI sequences (GenBank accession numbers ON911378–
298 ON911439) and combined these with available sequences on GenBank. Of the 30 squamates in
299 our study, we were able to analyze COI alignments for 22 species (4–69 sequences per
300 alignment; average 16) and cytb alignments for 22 species (8–143 sequences; average 58).
301 Three species (*Crotalus oreganus*, *Hypsiglena jani*, *Trimorphodon vilkinsonii*) did not have
302 sufficient mtDNA data available for either COI or cytb and 17 species had alignments for both
303 genes (Table 1). We sequenced new GBS data for 47 individuals from 12 species (NCBI SRA
304 accession numbers SAMN31800422–SAMN31800468; PRJNA903383) and calculated genomic
305 diversity metrics for 118 individuals from 30 species. One *Thamnophis sirtalis* sample had poor
306 sequencing success and was removed from subsequent analyses. The assembled GBS datasets

307 included 2,923–25,086 loci for the unique SNP datasets and 6,820–44,309 loci for the full
308 sequence datasets depending on the species (Table 1).
309
310 *Mitochondrial and Nuclear Diversity*
311 Mitochondrial diversity within species ranged from 0.0015 to 0.0812 for COI π and from
312 0.0005 to 0.0717 for cytb π. Diversity estimates for the two mtDNA genes were correlated
313 within species (Fig. S1; Spearman's rho (ρ) = 0.897, $p < 2.2e-16$). For the nuclear GBS data,
314 within-species heterozygosity ranged from 0.3 to 0.346 and within-species π ranged from
315 0.0005 to 0.0083. Mean within-individual π ranged from 0.0004 to 0.0056 and there was
316 substantial variation among individuals and localities for some species (Fig. 1). For species such
317 as *Thamnophis marcianus* and *Masticophis flagellum*, the within-individual π estimates were
318 consistent between the two individuals within a locality, but values for one locality were nearly
319 double the values for the second locality. For other species such as *Crotalus adamanteus* and
320 *Sonora episcopa*, the within-individual π estimates were highly consistent for all four
321 individuals. For still other species such as *Thamnophis sauritus* and *Lampropeltis splendida*,
322 there was large variation in π estimates even for individuals collected from the same locality.
323 The within-species π estimates were strongly correlated with the mean within-individual
324 π estimates (Fig. S1; Spearman's ρ = 0.986, $p < 2.2e-16$), and we used the within-species
325 estimate to represent overall species nuclear diversity for subsequent analyses. Metrics of
326 diversity for mtDNA and nuclear GBS data were not correlated for the species represented in
327 both datasets (Fig. 2). Nuclear within-species π was not correlated with mtDNA COI π
328 (Spearman's ρ = 0.16, p = 0.476) or mtDNA cytb π (Spearman's ρ = 0.293, p = 0.185). We did not

329 detect any clear differences in nuclear diversity between datasets sampled from the Desert
330 Southwest compared to the Coastal Plain (Fig. S2).
331
332 *Predictors of Squamate Genetic Diversity*
333 The OU model of evolution was the better fit for all three response variables (model
334 weights > 0.986; Table 2). Different sets of predictors were included in the top models for the
335 three response variables (Table 3). For mtDNA COI, parity was the only variable consistently
336 included in the top models. Diversity between viviparous and oviparous species was not
337 significantly different according to the top PGLS model (Table 4), although oviparous species did
338 have higher COI diversity than viviparous species without phylogenetic correction (Fig. 3a; t-
339 test: t = 2.34, p = 0.03). For mtDNA cytb, range size and parity were always included in the top
340 models. Within-species cytb diversity had a positive relationship with range size (Fig. 3),
341 suggesting that species with larger ranges tend to have higher mtDNA diversity. This
342 relationship was consistent for both uncorrected (Spearman's ρ = 0.64; p = 0.002; Fig. 3c) and
343 phylogenetically corrected datasets (phylogenetically independent contrasts: R^2 = 0.31; p =
344 0.007; Fig. 3d). Oviparous species had higher cytb diversity compared to viviparous species (Fig.
345 3b; t-test: t = 3.11, p = 0.008).
346 Range size and clutch size were consistently included in the top models for nuclear
347 within-species π (Table 2). Nuclear within-species π had a positive relationship with both range
348 size and average clutch size (Table 4; Fig. 3). The relationship between nuclear π and range size
349 was consistent for both uncorrected (Spearman's ρ = 0.56; p = 0.001; Fig. 3e) and
350 phylogenetically corrected datasets (PIC: R^2 = 0.21; p = 0.006; Fig. 3f). The relationship

351 between nuclear π and average clutch size was significant for the uncorrected data (Spearman's
352 ρ = 0.53; p = 0.003; Fig. 3g), but not based on phylogenetically independent contrasts (R^2 =
353 0.04; p = 0.288; Fig. 3h).
354
355 **Discussion**
356 We compared genetic diversity estimates from mitochondrial and nuclear genomes for
357 30 North American squamate species. Our analyses demonstrated that range size is positively
358 correlated with mtDNA and nuclear diversity in squamates and that life history traits are
359 inconsistent predictors of diversity for these datasets. We found that mtDNA and nuclear
360 diversity estimates were not correlated, but we note that several aspects of sampling made
361 these datasets difficult to compare. Here, we discuss sampling considerations for this and
362 future studies, including the number of individuals, localities, species, and loci available or
363 feasible to sample. Despite these challenges, we provide empirical evidence that geography is a
364 key element for understanding patterns of genomic diversity within and across species.
365
366 *Sampling Considerations Across Geography, Taxonomy, and Genomes*
367 Comparative population genomics is a developing field that aims to uncover the forces
368 underlying genomic diversity by comparing across multiple species (Ellegren and Galtier 2016;
369 Edwards et al. 2022). Studies typically focus on the roles of selection, neutrality, and
370 recombination in structuring genetic variation in a primarily non-geographic context, thus
371 sampling usually consists of a single or few individuals to represent each species (e.g. Romiguier

372 et al. 2014; Chen et al. 2017). In contrast, comparative phylogeography takes a geography-
373 centric view to understanding historical processes underlying current genetic variation within
374 species and lineages. Ideally, species have been sampled across their entire range and as
375 densely as possible, often revealing the presence of multiple lineages within nominal species
376 (e.g., Schield et al. 2018). A third emerging field, macrogenetics, focuses on an even broader
377 geographic and taxonomic scale, typically using repurposed data from relatively few genetic
378 markers to investigate large-scale patterns and predictors of genetic variation within species
379 (Leigh et al. 2021). Macrogenetic studies rely on sampling choices and data from previous
380 studies, but until relatively recently, genome-wide datasets were not common in
381 phylogeographic studies (Garrick et al. 2015). The gradual accumulation of these data over time
382 and the introduction of new data aggregators (e.g. Pelletier et al. 2022) potentially increases
383 the scale of comparative studies that use opportunistic sampling, but the novel data collected
384 here offer a direct comparison between two ecoregions.

385 Thoroughly sampling many individuals from many species across geographic space
386 remains costly, thus a pressing question relevant to the design of comparative studies is: How
387 representative are genomic diversity estimates from a few individuals or from a narrow part of
388 a species' range? Our GBS data indicate that this answer depends on several factors – certain
389 species (e.g., *Crotalus adamanteus*) had highly consistent genomic diversity estimates from
390 localities sampled ~200 km apart, while others (e.g., *Thamnophis marcianus*) had two-fold
391 differences in genomic diversity between localities (Fig. 1). This result indicates that using a
392 single individual or locality to represent diversity for an entire species does not always provide a
393 complete picture of diversity within species. We expected to find similar levels of genetic

394 diversity across individuals because we sampled them from historically stable regions. Thus, it
395 may seem surprising that such differences are already emerging. We predict even greater
396 differences when populations are sampled from topographically complex regions or in areas
397 that have been influenced by historical climate change (e.g., Howes and Lougheed 2008; Nali et
398 al. 2020), a consideration that should continue to be addressed in future studies.

399 Our study provides an initial comparison of squamate genomic diversity from five
400 families sampled in two species-rich regions, the North American Desert Southwest and North
401 American Coastal Plain. We found considerable variation in diversity among species within
402 some families (e.g., Colubridae and Natricidae; Fig. 1), but further taxon sampling is needed for
403 robust comparisons of genomic diversity across families. In general, the Desert Southwest has
404 more heterogenous environments and landscapes compared to the Coastal Plain, and Florida in
405 particular, where populations were sampled (Noss et al. 2015; Badgley et al. 2017). Thus far, we
406 did not detect any clear differences in average genomic diversity levels between the two
407 regions (Fig. S2), although several Desert Southwest species had high variation in genomic
408 diversity among individuals and localities (Fig. 1). This result is expected given the
409 heterogeneous nature of the region, where large differences in climate, elevation, and habitat
410 on relatively small spatial scales can conceivably lead to differences in population sizes and
411 standing genomic diversity among nearby localities (Myers et al. 2019).

412 We found that mtDNA and GBS diversity were not correlated, suggesting that mtDNA
413 may not be a reliable proxy for within-species diversity in North American squamates. Several
414 differences in our sampling for mtDNA and nuclear data made these datasets difficult to
415 compare, however. The number of mtDNA sequences available for some species was limited,

416 producing COI and cytb alignments for only 22 species each. These mtDNA sequences most
417 likely represent a much broader geographic area within each species compared to our GBS
418 data, though it is not possible to quantify the influence of geographic distance because sample
419 localities were not associated with the majority of GenBank sequences (Marques et al. 2013).
420 On the other hand, the GBS data consist of several thousand loci and provide a better
421 representation of genome-wide diversity within species compared to a single mtDNA gene.
422 These differences in power and geographic representation, along with relatively few data
423 points for comparison (22 species), may explain the lack of correlation between mtDNA and
424 nuclear diversity metrics in our study. Previous studies have found a relationship between
425 mtDNA and nuclear diversity in some taxonomic groups including mammals (47 species;
426 Mulligan et al. 2006) and Australian lizards (60 species; Singhal et al. 2017). Further study of this
427 topic is important for comparative studies because mtDNA is readily available for many more
428 species, but reduced representation nuclear genome datasets are increasing and have the
429 potential to provide new insights on overall species diversity.
430
431 *Additional Challenges for Repurposed Data*
432 One challenge for comparative studies such as ours is the difficulty of gathering
433 different data types from species complexes that have undergone recent taxonomic revisions.
434 When possible, we manually edited mtDNA alignments and IUCN range maps to reflect data
435 from a single described lineage (Table S1). In some cases, however, recent taxonomic revisions
436 and limited geographic information about existing sequences and traits made it impossible to
437 assign data to newly-described species. For those taxa, we excluded mtDNA data and assumed

438 trait information from former species designations would be representative for the whole
439 complex. We provide the alignments and GenBank accession numbers for the data we analyzed
440 (https://doi.org/10.5061/dryad.qfttdz0ks) such that they can be reassigned and reanalyzed in
441 follow-up studies. Future efforts to link geographic coordinates with genetic sequences in open
442 access databases will continue to improve the prospects of comparative studies (Pelletier et al.
443 2022). Frequent updates to geographic distribution and trait databases will also be needed to
444 facilitate large-scale comparative studies.

445 The approaches used to obtain and analyze nuclear genome-wide datasets are
446 important to consider for comparative studies of genetic diversity. One appealing characteristic
447 of mtDNA data for data repurposing and comparing species sequenced across different studies
448 is the ease of aligning known genes. Restriction site associated methods such as GBS provide
449 many more loci for analysis, but without an annotated reference genome for most species of
450 interest, these loci are anonymous. Furthermore, the locus assemblies and resulting
451 parameters are highly dependent on the parameter settings of the pipeline used for analysis
452 (Harvey et al. 2016; Shafer et al. 2017). Our study used existing squamate datasets from the
453 Desert Southwest that we reassembled to ensure the same ipyrad version and parameters were
454 used for comparison with new GBS data we collected for the Coastal Plain species. We found
455 that heterozygosity values were highly sensitive to the parameter settings used. Prior to
456 reassembly, heterozygosity estimates for the Desert Southwest species were an order of
457 magnitude smaller than those of the Coastal Plain species (Table S4). These contrasting results
458 demonstrate the additional challenges of comparing GBS and similar datasets collected from
459 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).

Predictors of Squamate Genomic Diversity

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, 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 squamate 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),

482 frogs (Paz et al. 2015), and insects (López-Uribe et al. 2019; Mackintosh et al. 2019). One
483 possible explanation is that larger-bodied animals can disperse over longer distances, leading to
484 less genetic structure and overall diversity across a species range. Our results do not provide
485 compelling evidence for squamates, perhaps indicating that body size is not a good indicator of
486 dispersal for this group. Our results for clutch size are consistent with the prediction that
487 species with larger clutch sizes harbor higher genomic diversity (e.g., Romiguier et al. 2014),
488 though this relationship did not remain after phylogenetic correction (Fig. 3h). These patterns
489 will require further scrutiny with increased species sampling in future studies.

490 The potential association of parity with mitochondrial diversity, where oviparous (egg-
491 laying) species had higher diversity on average than viviparous (live-bearing) species, is
492 intriguing and should also be investigated further with more comprehensive sampling. Since the
493 mitochondrial sequences were sampled across species ranges, our results may reflect greater
494 genetic structure for oviparous species, in contrast with the prediction that oviparous species
495 must disperse to find suitable nest sites (Shine 2015). The evolution of viviparity is associated
496 with colder climates and faster development times in cool temperatures (Ma et al. 2018), but
497 our study lacked sampling across broad latitudinal or climatic gradients. Furthermore, most of
498 the oviparous species included in this study belong to the family Colubridae, but viviparity in
499 squamates has evolved from oviparity at least 34 and perhaps more than 100 times (Pyron and
500 Burbrink 2014; Blackburn 2015). Squamates therefore present an excellent opportunity to
501 continue investigating whether life history traits such as clutch size or parity are associated with
502 intraspecific genomic diversity while taking phylogeny into account.

503 Overall, we provide new insights into the predictors of diversity within squamates
504 considering both mitochondrial and nuclear genome patterns. We showed that geographic
505 range size corresponds with intraspecific genomic diversity even when a relatively small part of
506 the range is sampled. We also demonstrate that genomic diversity can vary widely between
507 populations on small spatial scales, suggesting that sampling a few individuals may not
508 adequately represent diversity levels within species. As genome-wide datasets continue being
509 generated, comparative studies of intraspecific diversity across species ranges will increase
510 understanding of the roles of geography and life history in structuring diversity across a variety
511 of taxonomic groups.

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. Samples were obtained in accordance with collecting permits from the Florida Fish and Wildlife Conservation Commission and animal care protocols approved by the Florida State University Animal Care and Use Committee.

Data Availability

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

- Sampling locations, field numbers, museum catalogue numbers, R/Python scripts, and datasets used for analyses and visualization: Dryad link - https://doi.org/10.5061/dryad.qfttdz0ks
- DNA sequences: GenBank accession numbers ON911378–ON911439; NCBI SRA accession numbers: SAMN31800422–SAMN31800468 (PRJNA903383)

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.

Brüniche-Olsen A., Kellner K.F., DeWoody J.A. 2019. Island area, body size and demographic history shape genomic diversity in Darwin’s finches and related tanagers. Molecular Ecology. 28:4914–4925.

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

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

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

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

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

576 Cornetti L., Ficetola G.F., Hoban S., Vernesi C. 2015. Genetic and ecological data reveal species
577 boundaries between viviparous and oviparous lizard lineages. Heredity. 115:517–526.

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

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

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

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

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

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

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

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

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

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

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

601 Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2016. Sequence capture versus
602 restriction site associated DNA sequencing for shallow systematics. Systematic Biology. 65:910–
603 924.

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

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

608 Howes B.J., Lougheed S.C. 2008. Genetic diversity across the range of a temperate lizard. Journal of
609 Biogeography. 35:1269–1278.

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

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

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

617 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics.
618 24:1403–1405.

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

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

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

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

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

627 Laikre L., Hoban S., Bruford M.W., Segelbacher G., Allendorf F.W., Gajardo G., Rodríguez A.G., Hedrick
628 P.W., Heuertz M., Hohenlohe P.A., Jaffé R., Johannesson K., Liggins L., MacDonald A.J.,
629 OrozcoterWengel P., Reusch T.B.H., Rodríguez-Correa H., Russo I.-R.M., Ryman N., Vernesi C.
630 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.

Leroy T., Rousselle M., Tilak M.-K., Caizergues A.E., Scornavacca C., Recuerda M., Fuchs J., Illera J.C., De Swardt D.H., Blanco G., Thébaud C., Milá B., Nabholz B. 2021. Island songbirds as windows into evolution in small populations. Current Biology. 31:1303-1310.e4.

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.

Nali R.C., Becker C.G., Zamudio K.R., Prado C.P.A. 2020. Topography, more than land cover, explains genetic diversity in a Neotropical savanna tree frog. Diversity and Distributions. 26:1798–1812.

Nazareno A.G., Bemmels J.B., Dick C.W., Lohmann L.G. 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources. 17:1136–1147.

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.

Novosolov M., Rodda G.H., North A.C., Butchart S.H.M., Tallowin O.J.S., Gainsbury A.M., Meiri S. 2017. Population density–range size relationship revisited. Global Ecology and Biogeography. 26:1088–1097.

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. 22:2830–2843.

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. R package version 3.1-155. https://CRAN.R-project.org/package=nlme. .

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.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 3:217–223.

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.

Shine R. 2015. The evolution of oviparity in squamate reptiles: An adaptationist perspective. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 324:487–492.

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.

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

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

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

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

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

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

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

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

753

754 **Table and Figure Captions**

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

756

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

759

760 Table 3. Model comparison of PGLS models with different sets of predictors. The top three
761 models for each response are shown, which all had a delta AIC value of <= 2. Note that none of
762 the models for mtDNA COI π were substantially improved over others (delta AIC all < 4.4). The
763 plus ("+") sign in the predictor columns indicates that predictor was included in that model.

764

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

767

768 Figure 1. Individual sample localities and nucleotide diversity estimates for species included in
769 the study. (a) Map of all samples for which GBS data were analyzed. Desert Southwest samples
770 are shown in turquoise and Coastal Plain species in brown. Dotted black boxes indicate the
771 inset maps to the right, which depict localities sampled within two species as examples. Shapes
772 indicate individuals were sampled from the same locality within a species, but do not
773 necessarily indicate the same locality across different species. (b) Phylogenetic relationships for

774 the 30 species included in the study based on Tonini et al. (2016). Within-individual nucleotide
775 estimates are shown to the right with colors and shapes corresponding to the maps in (a).

776

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

779

780 Figure 3. Predictors of mtDNA (a-d) and nuclear diversity (e-h) within species. Scatter plots and
781 dotted lines depict simple linear relationships without phylogenetic correction. a, b) Welch's
782 two-sample t-test statistics and p-values are shown. c, e, g) Spearman's rank correlation sample
783 estimates and p-values are shown. d, f, h) Phylogenetically-corrected R^2 values and p-values are
784 shown.

785

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

Species	Family	# COI seqs	# cytb seqs	# GBS samples	# GBS loci (unique SNPs)	# GBS loci (seqs)
Ophisaurus ventralis	Anguidae	6	NA	4	22951	42875
Anolis carolinensis	Anolidae	16	NA	4	15708	28909
Agkistrodon conanti	Viperidae	5	46	4	5961	15540
Sistrurus miliarius	Viperidae	10	31	4	14356	23527
Crotalus ornatus	Viperidae	NA	27	4	9047	29251
Crotalus atrox	Viperidae	5	66	4	6926	19307
Crotalus scutulatus	Viperidae	4	74	4	14696	29502
Crotalus oreganus	Viperidae	NA	NA	4	11032	27615
Crotalus viridis	Viperidae	NA	8	4	15565	29627
Crotalus tigris	Viperidae	5	NA	4	9003	37903
Crotalus pyrrhus	Viperidae	6	NA	4	3274	6820
Crotalus adamanteus	Viperidae	10	126	4	8424	27387
Nerodia fasciata	Natricidae	30	143	4	15130	25148
Thamnophis marcianus	Natricidae	NA	22	4	7467	17648
Thamnophis sauritus	Natricidae	7	NA	4	14540	17945
Thamnophis sirtalis	Natricidae	7	37	3	15664	25409
Hypsiglena jani	Dipsadidae	NA	NA	4	19304	29950
Sonora episcopa	Colubridae	7	20	4	2923	12181
Trimorphodon vilkinsonii	Colubridae	NA	NA	3	5566	26644
Salvadora hexalepis	Colubridae	8	31	4	3513	20475
Masticophis flagellum	Colubridae	55	136	4	7512	18350
Coluber constrictor	Colubridae	51	22	4	18737	22267
Lampropeltis splendida	Colubridae	NA	25	4	6485	19095
Cemophora coccinea	Colubridae	69	9	4	15437	25907
Rhinocheilus lecontei	Colubridae	9	105	4	4713	15022
Arizona elegans	Colubridae	8	92	4	16879	37596
Pantherophis quadrivittatus	Colubridae	12	20	4	14352	26136
Pituophis catenifer	Colubridae	9	98	4	18205	32161
Pantherophis emoryi	Colubridae	NA	129	4	25086	44309
Pantherophis guttatus	Colubridae	11	11	4	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	57.452	-98.9	0.00	0.998
	BM	50.320	-86.6	12.26	0.002
mtDNA cytb π	OU	66.965	-117.9	0.00	0.997
	BM	60.257	-106.5	11.42	0.003
Within-species π	OU	157.485	-297.0	0.00	0.994
	BM	151.435	-286.9	10.1	0.006

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 mtDNA COI π were substantially improved over others (delta AIC all < 4.4). The plus ("+") sign in the predictor columns indicates that predictor was included in that model.

Response	df	log Like	AIC	ΔAIC	Weight	Range size	Body size	Clutch size	Parity
mtDNA COI π (OU models)	6	55.745	-99.5	0.00	0.129	+			+
	5	54.731	-99.5	0.03	0.127				+
	6	55.729	-99.5	0.03	0.127		+		+
mtDNA cytb π (OU models)	7	66.402	-118.8	0.00	0.379	+	+		+
	8	66.964	-117.9	0.88	0.245	+	+	+	+
	6	64.615	-117.2	1.58	0.172	+			+
Within-species π (OU models)	7	156.994	-300.0	0.00	0.342	+	+	+	
	6	155.531	-299.1	0.92	0.215	+		+	
	8	157.310	-298.6	1.37	0.172	+	+	+	+

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	-0.0565	0.0662	-0.8540	0.4043
	log(Range size)	0.0062	0.0047	1.3184	0.2039
	Parity_vivip	-0.0165	0.0100	-1.6530	0.1157
	N sequences	-0.0001	0.0003	-0.2777	0.7844
mtDNA cytb π (OU model)	Intercept	-0.1522	0.0539	-2.8252	0.0117
	log(Range size)	0.0078	0.0033	2.3328	0.0322*
	log(Body size)	0.0315	0.0182	1.7320	0.1014
	Parity_vivip	-0.0179	0.0061	-2.9503	0.0090*
	N sequences	0.0001	0.0001	0.8222	0.4223
Within-species π (OU model)	Intercept	-0.0099	0.0041	-2.4112	0.0236
	log(Range size)	0.0010	0.0003	3.8158	0.0008*
	log(Body size)	-0.0022	0.0014	-1.6001	0.1221
	log(Clutch size)	0.0043	0.0017	2.5383	0.0177*
	N loci	0.0000	0.0000	0.0665	0.9475

Figure 1. Individual sample localities and nucleotide diversity estimates for species included in the study. (a) Map of all samples for which GBS data were analyzed. Desert Southwest samples are shown in turquoise and Coastal Plain samples 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).

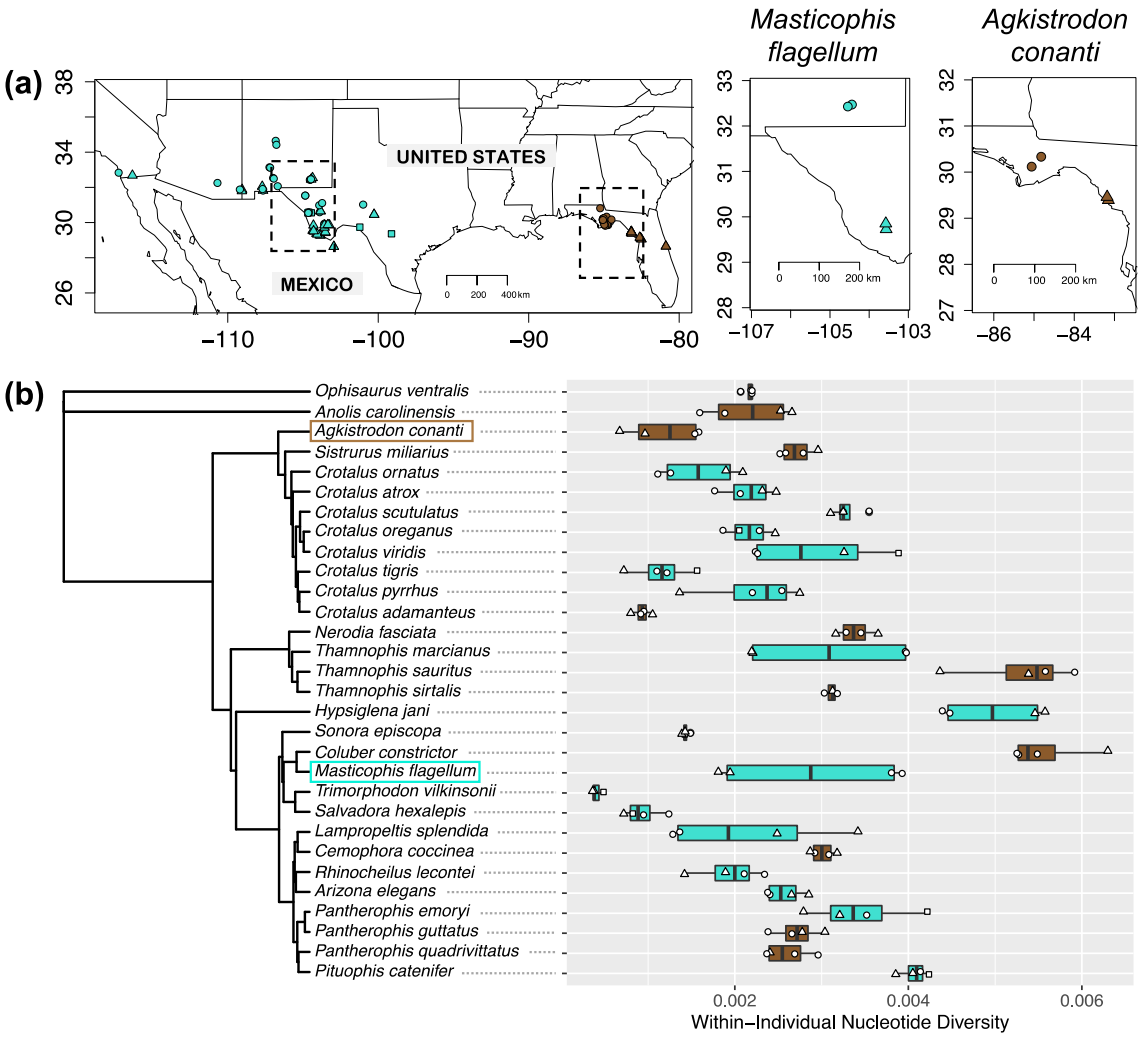


Figure 2. Lack of correlation between mtDNA and nuclear diversity estimates within species. Spearman's rank correlation sample estimates and p-values are shown.

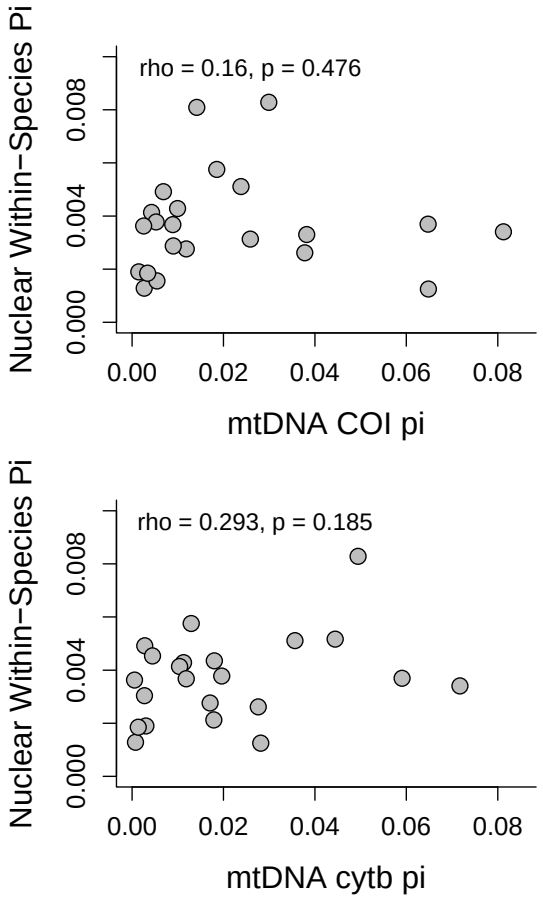
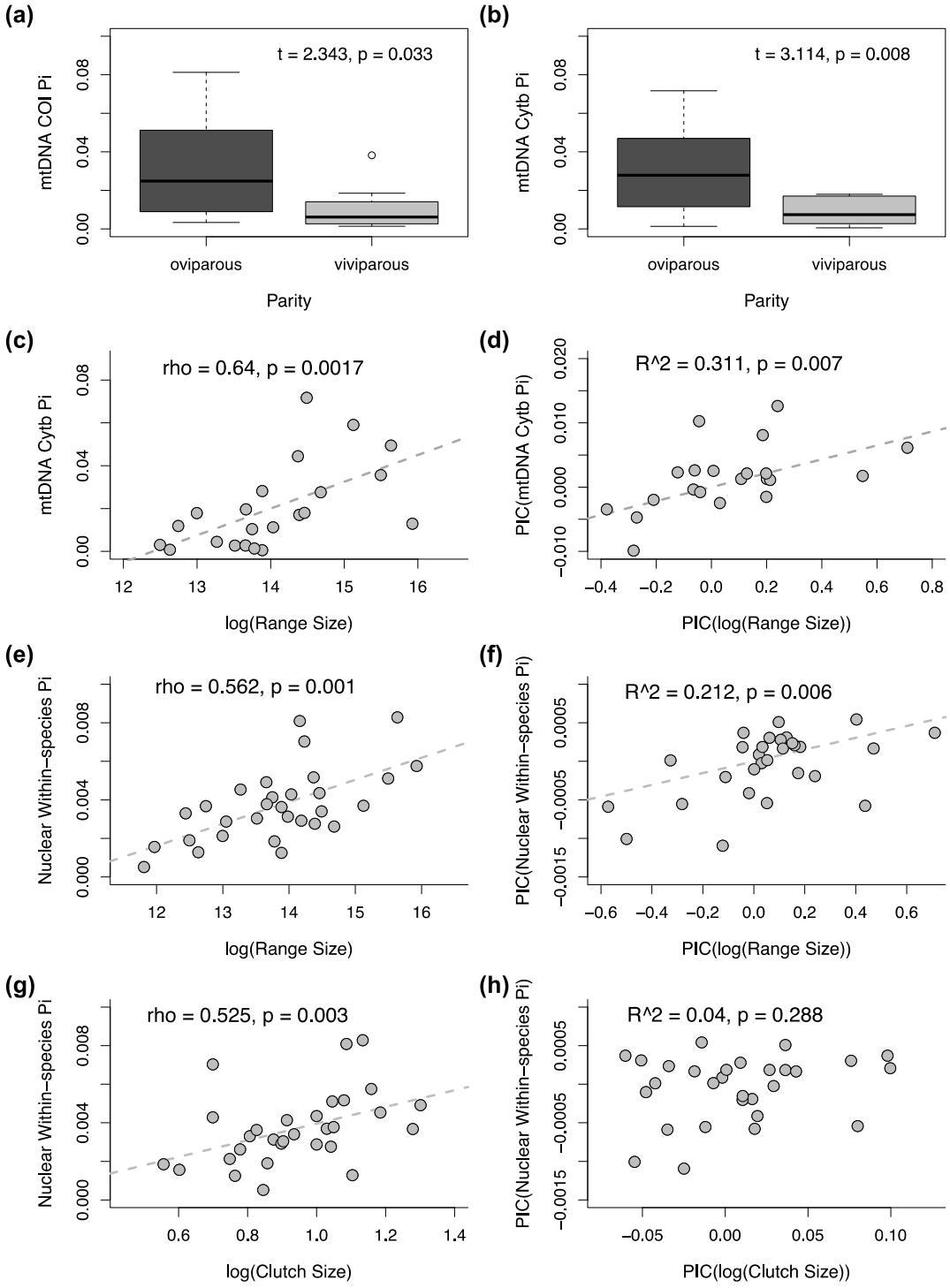


Figure 3. Predictors of mtDNA (a-d) and nuclear diversity (e-h) within species. a, b) Welch's two-sample t-test statistics and p-values are shown. c, e, g) Scatter plots and dotted lines depict simple linear relationships without phylogenetic correction. Spearman's rank correlation sample estimates and p-values are shown. d, f, h) Phylogenetically-corrected R^2 values and p-values are shown.



826 **Supplementary Materials**

827 Table S1. Taxonomy used for this study and details about how different datasets were handled
828 to reflect recent taxonomic changes. References that were used to manually edit mitochondrial
829 alignments and range maps are provided.

830 *[provided as Excel sheet – tab 1]*

831

832 Table S2. List of all samples for which mtDNA COI was sequenced and GBS data were analyzed.
833 Sample names, museum catalogue numbers, geographic coordinates, locality information, and
834 total number of GBS reads sequenced are included.

835 *[provided as Excel sheet – tab 2]*

836

837 Table S3. Additional species-specific sequencing primers designed in this study to improve COI
838 sequencing results.

839 *[provided as Excel sheet – tab 3]*

840

841 Table S4. Data for 30 squamate species used as input for analyses.

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

843

844

845 Table S5. Data for 118 squamate individuals used to make figures and summaries.

846 *[provided as Excel sheet – tab 5 and with R script in Dryad package]*

847

848 Additional files on Dryad (https://doi.org/10.5061/dryad.qfttdz0ks): Mitochondrial (COI and

849 cytb) alignments including previous GenBank samples; GBS files; Range maps; Python/R scripts

Figure S1. (a) Diversity estimates for mtDNA genes COI and cytb are correlated. (b) Mean within-individual pi estimates are strongly correlated with within-species pi estimates.

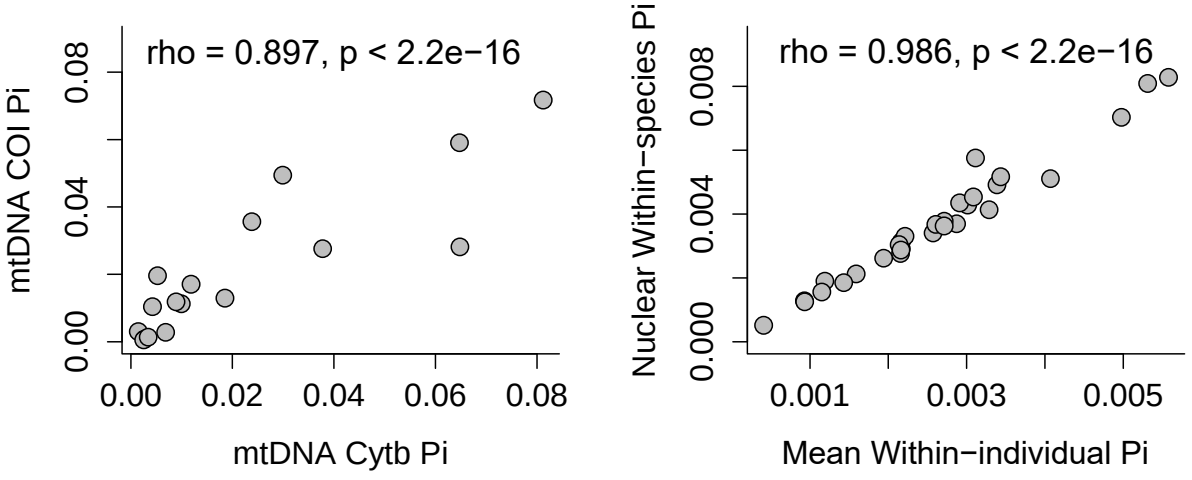
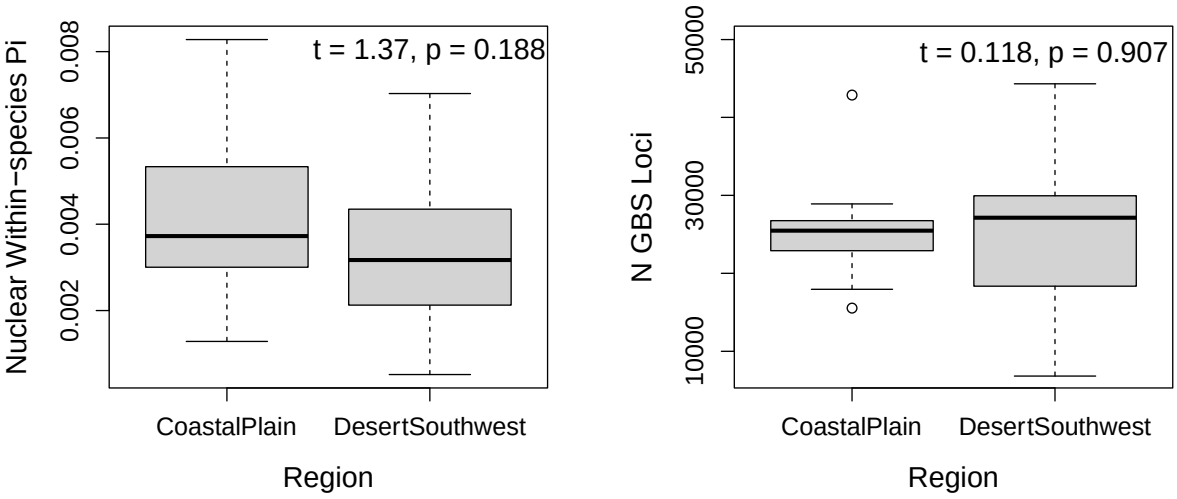



Figure S2. There are no clear differences in nuclear diversity estimates or the GBS datasets collected from the two different regions.