Resolving spatial complexities of hybridization in the context of the gray zone of speciation in North American ratsnakes (*Pantherophis obsoletus* complex)

Frank T. Burbrink[1*]

Marcelo Gehara[1,2]

Edward A. Myers[1,3]

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

[2]Department of Biological Sciences, Rutgers University Newark, 195 University Ave, Newark, NJ 07102

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

***Corresponding author** – fburbrink@amnh.org

Author contributions: FTB and EAM contributed to data collection, all authors contributed to analyses, FTB wrote the initial draft, and all authors edited subsequent versions.

Abstract:

Inferring the history of divergence between species in a framework that permits the presence of gene flow has been crucial for characterizing the gray zone of speciation, which is the period of time where lineages have diverged but have not yet achieved strict reproductive isolation. However, estimates of both divergence times and rates gene flow often ignore spatial information, for example the formation and shape of hybrid zones. Using population genomic data from the eastern ratsnake complex (*Pantherophis obsoletus*), we infer phylogeographic groups, gene flow, changes in demography, the timing of divergence, and hybrid zone widths. We examine the spatial context of diversification by linking migration and timing of divergence to the location and widths of hybrid zones. Artificial neural network approaches are applied to understand how landscape features and past climate have influenced population genetic structure among these lineages prior to hybridization. Rates of migration between lineages are associated with the width and shape of hybrid zones. Timing of divergence is not related to migration rate across species pairs and is therefore a poor proxy for inferring position in the gray zone. However, timing of divergence is related to the number of loci weakly introgressing through hybrid zones.

Key Words: Eastern Nearctic, migration, isolation, neural networks, reproductive isolation, cline, hybrid zone

Wide-ranging species complexes that cross numerous biogeographic barriers provide opportunities to better understand how changing landscapes affect diversification, gene flow, demography, and the formation of hybrid zones. Lineages within species complexes ranging

60 across heterogeneous landscapes are likely at different stages of the speciation process. These
61 differences may reflect how specific environmental and biogeographic barriers have uniquely
62 altered changes in gene flow and other demographic processes in a particular complex (Myers et
63 al. 2020). Even at the same barrier within related groups of organisms, the timing of divergence
64 and degree of gene flow can be species specific and dependent on when they encountered the
65 barrier (Riddle 2016; Myers et al. 2019b). Alternatively, generalist species not as constrained to
66 specific habitats may not show a correlation between genetic variation and landscape features
67 (Joseph and Wilke 2007; Makowsky et al. 2009; Lourenço et al. 2017).

68 Lineage divergence can be placed in the context of the gray zone of speciation, which
69 defines the range of time where speciation proceeds from early population differentiation with
70 unfettered gene flow to nearly reproductive isolation defined by low rates of gene flow (de
71 Queiroz 2007; Hewitt 2008; Roux et al. 2016; Jackson et al. 2017). The ability to delimit species
72 therefore may be correlated with their position in the gray zone, which has previously been
73 assessed by relating measures of genetic isolation and migration (the genealogical divergence
74 index; GDI) with species-delimitation probabilities (Roux et al. 2016; Leaché et al. 2019).
75 Moreover, the position within the gray zone could be simply predicted by the age of lineage
76 formation relative to population size if the evolution of Dobzhansky-Muller incompatibilities
77 via genetic drift alone occurred after divergence (Orr 1995; Gavrilets 2004; Singhal and Moritz
78 2013). Deviations from a negative correlation between rates of gene flow and age of divergence
79 could occur where increased gene flow between the oldest divergences in a group may cause
80 lineages to rapidly collapse, such as in the case where physical barriers to gene flow disappear.
81 On the other hand, strong selection may prevent young lineages from exchanging alleles at
82 particular loci, shortening the time in the gray zone (Mayr 1963; Barton 2010; Feder et al. 2012;

83 Roux et al. 2016; Edwards et al. 2020). Therefore, finding a threshold that can determine if two
84 species are unique given only divergence times may fail where rates of gene flow between
85 lineages drastically change over time and space (Gourbière and Mallet 2010; Nosil et al. 2017).
86 Understanding how divergence has occurred over the landscape in the context of hybrid
87 zones should provide a better understanding of the location of diverging lineages in the gray
88 zone. As first point of inquiry, examining gene flow in space determines if lineages were
89 generated via isolation and migration or are simply artificially partitioned groups with
90 continuous gene flow over the landscape defined by isolation by distance (IBD; Wright 1943;
91 Frantz et al. 2010; Bradburd et al. 2018). Barring the latter, if species pairs rapidly adapt to
92 unique habitats across their distribution, then age of divergence and degree of gene flow may be
93 unrelated relative to neutral processes when compared across taxa (Nosil and Crespi 2006;
94 Agrawal et al. 2011; Nosil 2012; Karrenberg et al. 2019). Therefore, divergence over landscape
95 can show how selection at a barrier can be discordant from the timing of origin of the lineages
96 (Barton and Hewitt 1985; Harrison 1993; Jiggins and Mallet 2000; Gay et al. 2008; Seehausen et
97 al. 2014; Stankowski et al. 2017).
98 To understand population divergence many studies use coalescent-derived, isolation-
99 migration models (Hey and Nielsen 2004; Hey 2010) to estimate gene flow throughout time,
100 reflecting position in the gray zone of speciation. These coalescent models have been used to
101 detect unique species (Rannala and Yang 2020), however, they lack any spatial component
102 necessary to understand the underlying geographic and environmental processes influencing
103 contemporary migration and degree of reproductive isolation reflected within any particular
104 locus. Coalescent estimates for understanding gene flow, timing, and species delimitation are
105 also affected by geographic sampling and connectivity among populations (Pante et al. 2015;

106 Coates et al. 2018; Mason et al. 2020). Studying this connectivity can determine if hybridization
107 among lineages are enhancing or reducing reproductive isolation (Abbott et al. 2013) and can
108 reveal if contemporary gene flow is restricted to a few loci (genic species) or most of the genome
109 (biological species; Nosil and Feder 2013). Therefore, studying hybrid zones should enhance
110 traditional coalescent estimators of divergence by determining: 1) the location and size where
111 lineages meet and migration occurs, 2) how genes respond to selection across these zones, 3) the
112 degree of contemporary reproductive isolation, and 4) if parental lineages exist in unique
113 habitats.

114 To understand the factors shaping reproductive isolation in a species complex, we ask
115 whether divergence time predicts rates of gene flow or cline width across space in the North
116 American ratsnakes (*Pantherophis obsoletus* complex). This complex is a wide-ranging group of
117 four species found throughout the forested regions of the Eastern Nearctic and nearby
118 Chihuahuan Desert that have diverged at three unique biogeographic barriers. The primary
119 divergence in this complex occurred at the Mississippi River, which has consistently been
120 identified as one of the main biogeographic barriers in the ENA (Robison 1986; Burbrink et al.
121 2000; Soltis et al. 2006; Brandley et al. 2010; Zellmer et al. 2012; Myers et al. 2020). Further
122 east, the Appalachian Mountains and Apalachicola/Chattahoochee River System (AARS) have
123 noted barrier effects in other organisms and likely contributed to the divergence between *P.*
124 *alleghaniensis* and *P. spiloides* (Walker and Avise 1998; Burbrink et al. 2000; Soltis et al. 2006).
125 West of the Mississippi River, divergences occurred at the transition between temperate forests
126 to the western edge of the Edwards Plateau into the Chihuahuan Desert and isolated the species
127 *P. obsoletus* and *P. bairdi* (Fig.1; Lawson and Lieb 1990; Burbrink et al. 2000; Burbrink 2001).
128 Since the Pliocene, the ENA has experienced numerous climate change events associated with

129 glacial cycles that forced species into refugia and compressed populations; upon climate
130 amelioration populations expanded into formerly glaciated areas (Hewitt 2000; Bintanja and van
131 de Wal 2008; Burbrink et al. 2016). Using dense population sampling and genomic-scale data,
132 we combine isolation-migration estimates with spatial information to understand species
133 divergence in the context of geographic distance, contemporary and historical environments, and
134 biogeographic boundaries. We examine how migration rates and divergence dates relate to the
135 widths of hybrid zones at biogeographic boundaries and investigate demographic changes
136 through time to understand if population expansion occurs among all lineages as predicted given
137 glacial cycling. Addressing these questions provides an integrated view of phylogeographic
138 history over a physically and historically complex landscape and yields a clearer understanding
139 of speciation processes and delimitation in the gray zone.
140
141
142 **Methods**
143
144 *Dataset*
145 We sampled 288 individuals liberally covering the range of all species within the
146 *Pantherophis obsoletus* complex (Fig.1; Dryad XXX). DNA was extracted from all samples
147 using Qiagen DNeasy Blood & Tissue Kits and samples were screened for quality using broad-
148 range Qubit Assays. We used services from RAPiD Genomics (https://www.rapid-
149 genomics.com/services/) to generate 5472 baits and to sequence 5060 conserved elements
150 (UCEs) loci following the protocols from (Faircloth et al. 2012) and (Sun et al. 2014). These
151 markers have been used to address phylogeographic/population genetic and deeper phylogenetic

152 questions (Myers et al. 2019a; Younger et al. 2019). We mapped UCE reads to a Chromium 10x
153 *Pantherophis spiloides* genome (in prep), removed loci containing >50% missing samples, and
154 removed individual specimens missing >30% of all alignments (details on the assembly of UCE
155 loci are available in Supporting Information Material 1). We produced both phased locus
156 datasets, used in our coalescent estimators, and single SNPs/locus for all other analyses. We
157 filtered the unlinked SNP dataset to remove alleles with a minor allele frequency <0.1 and used
158 these data for population structuring analyses (Linck and Battey 2019).
159
160 *Geographic groupings*
161 We estimated population structure by comparing Discriminant Analysis of Principal
162 Components (DAPC; Jombart et al., 2010) in the R package (R Core Team 2015) adegenet
163 v2.1.2 (Jombart 2008). Additionally, we estimated effective migration surfaces (EEMS; Petkova
164 et al., 2016) three times with different starting seeds to insure consistency. This method models
165 effective migration rates over geography to represent regions where migration is low in cases
166 where genetic dissimilarity increases rapidly (details on these population grouping methods and
167 other methods used for assessing structure are available in Supporting Information Material 1).
168
169 *Isolation, migration, and historical demography*
170 We generated a coalescent-based species tree with SNAPP v1.3.0 (Bryant et al. 2012) to
171 test species limits and understand relationships among the four groups identified from clustering
172 analyses. Because generating a tree in SNAPP using all individuals was computationally
173 intractable, we used four individuals per taxon that were sampled from different regions of the
174 taxon's distribution. We estimated SNAPP trees using BEAST v2.5.2 (Bouckaert et al. 2014;

175 details on parameter setting for SNAPP are available in Supporting Information Material 1). We
176 assessed four alternative species delimitation models including four taxa, three taxa (collapsing
177 *P. bairdi/P. obsoletus* or *P. alleghaniensis/P. spiloides*), and two taxa (collapsing *P. bairdi/P.*
178 *obsoletus* and *P.alleghaniensis/P. spiloides).*

179 Using the four inferred lineages, we examined which demographic models of species
180 divergence best described the origins of these groups given divergence time, historical
181 demographic change, and migration between spatially adjacent taxa. We tested 2,300 candidate
182 isolation-migration models including all possible topologies with or without migration between
183 all spatially adjacent pairs using the program PHRAPL (Jackson et al. 2017) implemented in R.
184 To further test that four genetic lineages exist, we ran pairwise comparisons between the best
185 selected model and models of three populations where sister species were collapsed into a single
186 entity; four *Pantherophis* species against one model where *P. obsoletus* and *P. bairdi* were
187 collapsed, and another where *P. spiloides* and *P. alleghaniensis* were collapsed (details on
188 PHRAPL runs are available in Supporting Information Material 1).

189 To estimate gene flow, timing of divergence, and demographic change, we used
190 PipeMaster (Gehara et al., 2017; Gehara et al. in review) to simulate genetic data and perform
191 approximate Bayesian computation (ABC) and supervised machine-learning. Here we used the
192 best model selected in the PHRAPL analysis as a template (same topology and migration
193 parameters) to generate three competing models: (i) an isolation migration model with constant
194 population size for each lineage and constant migration, (ii) demographic change along each
195 lineage and constant migration through time and demographic change, and (iii) population size
196 change with migration occurring after the Last Glacial Maximum. We simulated 100,000 data
197 sets of 54 summary statistics and performed ABC rejection with 0.01 tolerance level to select the

198 best of these three models (Table S1). We then took the selected model and performed rejection
199 using a 0.1 tolerance level with a neural network regression using a single layer and 20 nodes to
200 estimate the model parameters. We also performed this without hybrid individuals (admixture
201 proportions between 0.5-0.7) to explore how including these samples may change estimates of
202 divergence timing, demographic change, timing of demographic change, and migration within
203 these lineages.
204
205 *Spatial population genetics*
206 We used redundancy analyses (RDA; Legendre et al. 2011; Diniz-Filho et al. 2013) and
207 artificial neural networks (ANN; Lek and Guégan 1999; Legendre and Fortin 2010; Legendre et
208 al. 2011) to understand how contemporary and historical aspects of the landscape and climate
209 generated genetic diversity. To set up our data for these analyses, we used as our response
210 variable Euclidian genetic distance and transformations of this (see below) among all individuals
211 in R. For predictor variables, we used 1) geographic distances measured as pairwise great-circle
212 distances among all points to account for isolation-by-distance using the R package fossil
213 (Vavrek 2020), 2) binary categorization of isolation east and west of the Mississippi River, 3)
214 elevation, and 4) bioclimatic variables at 2.5 mins representing current climate, Last Glacial
215 Maximum (LGM, 21kya), Last interglacial (LIG, 130 kya), and Pleistocene Marine Isotope
216 Stage 19 (MS19, 787 kya) all obtained from the PaleoClim database (Brown et al. 2018). We
217 also extracted level II ecoregions for all individuals. Ecoregion variables delineate unique
218 ecological areas and are well characterized across North America (Bailey 1995). For the
219 bioclimatic data and elevation, we extracted parameter values given the latitude and longitude for
220 each individual sample using the R package raster (Hijmans et al. 2014). To reduce

221 dimensionality of the bioclimatic variables for each type (current, LGM, etc.), we eliminated
222 correlated variables (>0.90) and transformed remaining variables using centered principal
223 components analyses in R. We transformed categorical ecoregion data into dummy variables and
224 reduced dimensionality using logistic PCA for binary variables (Landgraf and Lee 2015). To
225 account for the effect of geographic distances on genetic structure, we transformed spatial
226 distance between all samples using principal coordinates neighbor matrices (PCNM) in the R
227 package vegan (Dixon 2003). This method allowed us to account for multiple spatial structures
228 (neighborhoods) on heterogenous ecological structures and barriers, which better clarified non-
229 spatial effects on genetic structure (Borcard and Legendre 2002).

230 We first used RDA, asymmetric canonical analysis, to determine the significance for each
231 spatial predictor axis (Legendre et al. 2011) on genetic structure (McGaughran et al. 2014;
232 Noguerales et al. 2016). We then used the reduced set of significant spatial axes to examine their
233 effect with environmental and barrier variables on genetic structure using RDA and ANN (see
234 Supporting Information Material 1 for details). Machine learning methods, such ANN, can infer
235 non-linear interactions among many predictor variables, regardless of the distribution or variable
236 type and can generate a range of models of varying complexity (Lek et al. 1996; Zhang 2010;
237 Libbrecht and Noble 2015; Sheehan et al. 2016). We transformed Euclidian genetic distances
238 into principal coordinates (PCoA; (Gower 1966) in adegenet (Jombart 2008) to produce
239 uncorrelated orthogonal axes to summarize genetic variation and assessed accuracy and variable
240 importance on the first 10 PCoA axes. We used regression-based ANN to determine which
241 variables predict genetic distances and rank those with the highest model importance. To conduct
242 the ANN regressions we used the R package *caret* (Kuhn 2008) and partitioned the data into the
243 standard 70% training and 30% test sets (Lek et al. 1996; Zhang 2010; Burbrink et al. 2020).

244 This analysis was run using 1,000 maximum iterations to ensure convergence. We resampled the
245 data using the default 25 bootstrap replicates to reach convergence over the following
246 parameters: weight decay, root mean squared error, r^2, and mean absolute error. Accuracy was
247 compared to an ANN run with random shuffling of the response variable. We also compared
248 results from our ANN analysis to those using redundancy analyses (RDA, see Supporting
249 Information Material 1 for details).
250
251 *Hybrid-zone dynamics*
252 Using the program HZAR v.2-9 (Derryberry et al. 2014) we used admixture proportions
253 to determine if cline widths differ among adjacent species. We used TESS3r to estimate
254 admixture for each species pair using SNPs and geographic data (see Supporting Information
255 Material 1; Caye et al. 2018). As required by HZAR, we reduced 2-dimensional space (latitude
256 and longitude) into 1-dimensional distance from the hybrid zone. We first determined the
257 location of the hybrid zone by mapping admixed individuals (0.5-0.6 admixture), calculating all
258 sample distances to the hybrid zone, and then fixing a sign to these distances where 0 was the
259 hybrid zone, positive distances trended towards one parental taxon, and negative distances
260 trended towards the other. Using the Gaussian cline model, we estimated the center and width of
261 the cline and determined if these sigmoidal distributions have significant tails by fitting the
262 following models: 1) no tails, 2) right tail only, 3) left tail only, 4) mirrored tails, and 5) both
263 tails estimated independently (see Derryberry et al. 2014). We fit these models to our data using
264 AICc and ran the MCMC chains for $5x10^6$ generations, thinned by $5x10^3$ generations, and
265 estimated stationarity using ESS >200 in the R package CODA (Plummer et al. 2006).

266 To examine the width of these clines over environmental space, we transformed
267 environmental space to a single dimension and measured distance relative to the hybrid zone. We
268 filtered correlated bioclim data (>0.90), reduced dimensionality using PCA, estimated absolute
269 environmental distance between all individual points and hybrid zones, and provided a positive
270 or negative sign depending on the side of the hybrid zone. We ran HZAR described above for
271 spatial distance data.

272 To examine the width of allelic clines through hybrid zones relative to overall admixture,
273 we also estimated cline widths for each locus over geographic and environmental distance. Here
274 we used allele frequencies for each individual and applied the same distances as described above
275 using HZAR. To better understand the effects of selection over geographic space for each
276 species-pair comparison (see Stankowski et al. 2017), we determined which loci were fixed for
277 each parental taxa (>0.80) in the tails of the cline and also showed sharp clines (reduced widths).
278 The difference in the frequency of fixation (ΔP) between each tail of the cline (sampling 5% of
279 each most distant tail) was estimated and widths were compared to neutral expectations of size
280 (see below).

281 Using equations from Barton and Gale (1993) and applied in Bailey et al. (2015) we
282 evaluated selection against hybridization for each cline. With estimates of the root mean squared
283 (RMS) maximum lifetime dispersal (σ) for ratsnakes, we predicted the width of the cline under
284 neutrality over a range of generation times (*T*) since the origin of each sister pair and group using
285 the equation: $w = 2.51\sigma\sqrt{T}$. Unfortunately RMS of dispersal is unknown for ratsnakes, but we
286 bound this from 1.02 to 4.03 km given mean and maximum distances of nesting sites to
287 hibernaculum (Blouin-Demers and Weatherhead 2002). The maximum width was likely an

288 underestimate of lifetime dispersal, particularly in areas in the southern United States where
289 ratsnakes are not constrained by hibernacula.
290
291 All scripts, genetic, and spatial data are available on Dryad XX.
292
293 **Results**
294 *Data and population genetic structure*
295 After filtering loci for presence in 70% or more individuals, we generated 2,491 UCE
296 loci. For all downstream analyses we randomly sampled one SNP per locus yielding a total of
297 846 SNPs for 238 individuals for an average of 13.96% missing data.
298 Both EEMS and DAPC found population structure that generally matched the geographic
299 ranges of *P. alleghaniensis*, *P. spiloides, P. obsoletus*, and *P. bairdi* (Fig. 1 and Fig. S1;
300 Burbrink, 2001). All three EEMS generated similar acceptance proportions for all proposal types
301 (12-51%). Total DAPC assignment probabilities without a priori species groupings were 0.975
302 (*P. bairdi* = 1.0, *P. obsoletus* = 0.96; *P. spiloides* = 1.0, and *P. alleghaniensis* = 0.97). Similarly,
303 all three EEMS replicates showed low migration at the Mississippi River (MR) and the area
304 separating the *P. bairdi* and *P. obsoletus* in west Texas. East of the MR, estimated low migration
305 occurred near the Appalachian Mountains, though this area is a complex mixture of isolation and
306 gene flow (Fig. 1).
307 These estimates of population structure all showed *P. alleghaniensis* as occurring mostly
308 in the Florida Peninsula and north along the east coast to Virginia/Delaware, *P. spiloides* was
309 found from western Florida to the Mississippi River up through the Midwest and east to the

310 northeastern US, *P. obsoletus* was found west of the Mississippi River in the forested regions of
311 the Midwest, and *P. bairdi* was distributed in the Chihuahuan Desert of west Texas (Fig. 1).
312
313 *Isolation-migration processes*
314 Our SNAPP analyses produced a phylogeny with a root separating the taxa east and west
315 of the Mississippi River with a sister relationship between *P. bairdi* and *P. obsoletus* and then *P.*
316 *spiloides* and *P. alleghaniensis* (Fig. 2), consistent with Burbrink et al. (2000). Bayes factors
317 (BF) showed the four-taxon model as decisively superior to the three-taxon (BF =35.40) and
318 two-taxon model (BF=2022.46).
319 Using those four lineages we filtered models of divergence using PHRAPL. A model
320 delimiting all four taxa, incorporating migration between spatially adjacent lineages was
321 preferred with the same topology as found using SNAPP. ΔAIC between the best ranked model
322 and next model was > 20, suggesting high confidence in model selection (Table S2).
323 The demographic model estimated from PipeMaster that incorporated historical
324 population size change and constant migration (IMD) best fit the data (posterior probability 1.0;
325 see PCA plots in Fig. S2; Fig. 2). Median migration rates were highest between eastern groups
326 (*P. alleghaniensis* to *P. spiloides* = 2.79 and *P. spiloides* to *P. alleghaniensis* = 3.79
327 individuals/generation), lower in the central groups across the Mississippi River (*P. spiloides* to
328 *P. obsoletus* = 1.95 and *P. obsoletus* to *P. spiloides* = 1.28 individuals/generation), and lowest in
329 western groups (*P. bairdi* to *P. obsoletus*= 0.55 and *P. obsoletus* to *P. bairdi* = 0.65
330 individuals/generations; Table 1 & Fig. 3). Divergence occurred earliest in the central lineages at
331 the Mississippi River at 3.181 My (95% quantile =1.65 to 5.02 My), then the eastern lineages at
332 1.95 My (95% quantile = 0.674 to 2.95 My) and western lineages at 1.10 My (95% quantile

333 0.407 to 2.62 My; Table 1 & Fig. 3). Divergence dates and average rates of migration were
334 uncorrelated (ρ=0.264, *P*=0.83, n=3). For all taxa we inferred population expansion from
335 ancestral N_e sizes ranging from 5,200 to 5,600 to modern sizes at 138,000 to 281,000 (Fig. 3).
336 Median expansion times were similar for all taxa (0.230-0.260 Ky) with lower 95% quantiles
337 always occurring in the Pleistocene (0.024-0.38 Ky; Fig S3). Posterior distributions for all
338 parameters were different from the uniform priors, though wide distributions of several
339 parameters suggested some uncertainty in these estimates (Fig. 3). Similarly, rerunning
340 PipeMaster without heavily admixed individuals resulted in similar predictions but with lower
341 estimates of migration for eastern and central lineages (Fig. S3).
342
343 *Spatial population genetics and hybrid zones*
344 We used ANN to determine what features of the landscape and environmental layers
345 through time best predict genetic structure (Fig.4). Estimates of accuracy using ANN were high
346 (>90%) for the first PCoA axis of genetic distance and declined rapidly thereafter approaching
347 random shuffling of the response variable by axis 4-7 (Fig. 4). Most genetic structure was
348 predicted by the Mississippi River showing 100% variable importance, when examining
349 population structure across all four species. Spatial components alone also played a role in
350 structuring these data for all species-pair comparisons, although they were partialed out in RDA
351 analyses showing the significant effect of environmental and biogeographic variables (see
352 Supporting Information Material 1). For those lineages east of the Mississippi River, population
353 structure was predicted by climate at the Last interglacial and current environment for the first
354 PCoA axis. For those lineages west of the Mississippi River, elevation, Last Glacial Maximum
355 and eceoregions were important for structuring these lineages in the first PCoA axis. We note

356 that while accuracy decreased on the 2nd and 3rd PCoA, other environmental variables still
357 showed importance.
358 Using TESS admixture estimates and one-dimensional spatial distances, HZAR showed
359 support for a model with no tails for all lineage-pair comparisons (eastern ΔAIC =-4 .08 – -8.34;
360 central ΔAIC = -3.55 – -8.17; western ΔAIC = -2.12 – -8.48; Fig. 5, Table 1). Median width of
361 clines decreased from the eastern lineages (256 km), to central (147 km) to the western lineages
362 (21 km). Using environmental distances, both the eastern (ΔAIC = -4.39 – -9.45) and central
363 lineages (ΔAIC = -4.32– -346.38) showed support for a model with no tails for both, whereas the
364 western lineages supported a left-tailed model (ΔAIC = -6.12 – -11.55). As a percentage of the
365 entire environmental distance for a species-pair comparison, the environmental cline width
366 decreased from the central (5.0%), eastern (4.0%), and western lineages (2.1%; Table 1).
367 Percentages of widths per species-pair comparison were lower as a measure of total
368 environmental distance than spatial distance.
369 We examined cline widths by locus for each species pair (Fig.6). For each comparison a
370 large number of loci were uninformative about cline widths, due to a lack of fixed differences
371 among parental taxa. However, as a measure of expected cline width per locus under neutrality
372 from the last glacial maximum, we showed that widths should be much larger (~633-733 km)
373 than found using admixture coefficients and across many loci (Fig. S5). We estimated that 128
374 loci between the central lineages, 8 between the eastern lineages, and 4 between the western
375 lineages were both fixed in the tails of a cline for each parental taxon(ΔP>0.80) and presented
376 cline widths lower than neutral expectations,(Fig. 6). Median cline centers for all loci, where the
377 direction of selection on an allele was expected to change, was close to zero, the predicted center
378 based on admixture proportions, for the central and western lineages, -8.32 and 4.95 km,

379 respectively. However, center estimates ranged from -28.35 to 1766 for the eastern lineages,
380 though a large number of loci (n=66) estimated the center closest to zero (ranging 20 km on each
381 side; Fig. 6).

382 When using a single SNP per locus we showed that 30% of loci from central lineages,
383 10.9% of loci for western lineages, and only 1.9% of loci for eastern lineages attained an F_{st}
384 >0.35. Given demographic histories between comparisons, this threshold for F_{st} was generally
385 considered extreme population disconnection beyond which advantageous loci will be exchanged
386 between populations (Wright 1978; Frankham et al. 2010).

387

388 **Discussion**

389 *Grey Zone Dynamics*

390 To understand the factors shaping reproductive isolation in a species complex, we show
391 that divergence time does not predict rates of migration or cline width across space in eastern
392 ratsnakes. While reproductive isolation should increase with time (Orr 1995; Singhal and Moritz
393 2013), our results indicate that overall genomic estimates of gene flow may be inadequate to
394 assess reproductive isolation with time. In contrast, the number of loci showing steep cline shifts
395 in contact zones is related to timing of divergence. Disconnection between episodes of repeated
396 secondary contact through time differing among species-pair comparisons or variable strength of
397 ecological divergence may account for the dissociation between time and overall migration rate.

398 Using multiple methods to examine genetic structure and migration over space and
399 accounting for isolation by distance, we show that the eastern ratsnakes are composed of four
400 divergent lineages, with *P. alleghaniensis* and *P. spiloides* meeting in the southeastern US at the
401 intersection of subtropical and temperate areas, *P. spiloides* and *P. obsoletus* separated at the

402 Mississippi River, and *P. obsoletus* and *P. bairdi* meeting at the connection on the forested and
403 rocky habitats on the western edge of the Edwards Plateau. The western lineages (*P. obsoletus*,
404 *P. bairdi*) have low migration rates, a cline equivalent to only 1.6% of the parental range, large
405 fixation indexes ($F_{st} > 0.35$) in 10.9% of loci, and likely geographic selection in only 4 loci. This
406 is in comparison with central lineages (*P. obsoletus, P. spiloides*) showing more migration, a
407 cline occupying 7.2% of the parental ranges, but with 35% of loci showing high F_{st}, and
408 geographic selection in 128 loci. Finally, eastern lineages (*P. alleghanienis, P. spiloides*) have
409 the highest rate of migration, with only 1.9% of loci showing high fixation, and only 8 loci
410 revealing geographic selection (Figs. 3 & 6). The oldest species-pair at the Mississippi River
411 have moderate migration/generation but yet have the highest number of loci showing steep clines
412 at the hybrid zones. Similar to other studies where many loci showing differential introgression
413 or variance in cline slope (Nosil et al. 2009; Carneiro et al. 2013; Stankowski et al. 2019), our
414 results lend support that the degree of reproductive isolation may be defined at the level of the
415 genes and not the genome (Macholán et al. 2011; Harrison 2012). Therefore, it is likely that
416 these species-pairs of ratsnakes are traversing different phases of reproductive isolation, from the
417 genic to genomic, and therefore overall coalescent estimates of migration alone may not useful
418 proxies for understanding position in the gray zone.

419 Hybrid zone dynamics help to understand the interactions between pairs of taxa as
420 characteristics of the hybrid zone reveal whether there is selection against hybrids (tension
421 zones), selection for hybrids (bounded superiority), or ecological gradients with reduced fitness.
422 They also reveal whether selection is endogenous (genomic) and/or exogenous (environmental;
423 Barton and Hewitt 1985; Barton and Gale 1993). Our results that combine estimates of when and
424 how divergence occurred with contemporary dynamics of hybrid zones provide a comprehensive

425 view of how lineage structure is maintained through time. Demographic models for ratsnakes
426 indicate that some migration at the Mississippi River and among all species pairs occurred during
427 the Pleistocene suggesting primary or secondary contact existed through major climate change
428 cycles. Population expansion in all lineages occurred at a minimum before the LGM, likely
429 indicating ancient contact and formation of these hybrid zones. Given this age and assuming
430 selective neutrality, we demonstrate that even after the last glacial maximum (LGM) that these
431 hybrid zones should be much larger, in most cases exceeding the ranges of parental taxa (Fig.
432 S5). The widths of the current hybrid zones have likely been maintained or continuously
433 reformed through time after reemerging from glacial refugia, as discussed for other temperate
434 taxa (Hewitt 1996, 2011). Despite constant connectivity, the historical and geographic identity of
435 the parental taxa are preserved. While there is likely some selection against hybrids as estimated
436 from hybrid zone width and dispersal, depending on species-pair comparison, it is possible that
437 other factors may be constraining these hybrid zones, for example fitness changes along an
438 ecological gradient or isolation by environment (Case and Taper 2000; Case et al. 2005;
439 McEntee et al. 2018). For all groups, our ANN show that either past or current environmental
440 differences or barriers to dispersal are important for structuring overall genetic differences
441 among species pairs. Among contemporary lineages and analyses, we show that admixture clines
442 change sharply at environmental transitions relative to just spatial distances (Fig. 5), which
443 suggests that fitness is likely changing across these ecological areas. Similarly, we find several
444 loci in each species-pair comparison that show similar or smaller cline shifts in these areas found
445 for overall admixture predictions. Unless these loci represent the outcomes of isolation by
446 distance and fixation in each ends of the cline, there is likely some selection on these genes.
447 Therefore, barriers that isolated these lineages initially, along with combinations of current and

448 past environmental changes across the ranges of all these species pairs, likely helped form and
449 maintain these hybrid zones.
450
451 *Influence of barriers and environment on population structure*
452 These ratsnake lineages are structured by a complex of physical barriers, historical
453 climate profiles, and geographic distance. Machine learning approaches show that the
454 Mississippi River has served as a strong barrier for isolating lineages of ratsnakes, forming
455 groups east and west of this river during the late Pleistocene/Pliocene (Figs 1&3,4, Table 1).
456 These results are consistent with previous mtDNA and morphological conclusions (Burbrink,
457 2001; Burbrink et al., 2000). The Mississippi River has served as a strong barrier to gene flow
458 given evidence from historical estimates of increased drainage (Cox et al. 2014) and numerous
459 unrelated organisms show genetic disconnection across the river (Robison 1986; Burbrink 2002;
460 Soltis et al. 2006; Burbrink et al. 2008; Pyron and Burbrink 2009; Brandley et al. 2010; Satler
461 and Carstens 2017; Myers et al. 2020). Hybridization, while low (1-2 individuals/generation,
462 Fig.3), does occur in a zone (~100km) east of the Mississippi River between the central lineages
463 (*P. obsoletus* and *P. spiloides*). Furthermore, genetic structure is also predicted by geographic
464 distance and mid-Pleistocene climate change (787kya).
465 The Mississippi River is responsible for the earliest divergence in this complex but other
466 geographic features are also important. East of the Mississippi River, divergence between *P.*
467 *alleghaniensis* and *P. spiloides* was influenced by climate at the last interglacial, current climate,
468 and geographic distance (Fig 4). For these taxa, divergence occurred at the Appalachian
469 Mountains and Apalachicola/Chattahoochee River System. *Pantherophis alleghaniensis* may
470 have diverged in an isolated Florida, due to sea level changes during interglacials, as suggested

471 by our ANN, and subsequently expanded north and west. This isolation is reflected by the unique
472 morphology there showing a high concentration of the yellow, striped color patterns in the range
473 of *P. alleghaniensis* in Florida and parts of the southeastern coast of the US (Schultz 1996;
474 Burbrink 2001). This suture zone between the Florida peninsula and the continental US has been
475 found for many taxa (Remington 1968; Swenson and Howard 2005; Burbrink et al. 2008). While
476 these two forms can be delimited with genetic data and occupy distinctly different geographic
477 regions, reproductive isolation may not be complete. Although, individual loci likely show
478 varying levels of reproductive isolation. Niches are also different between these taxa; *P.*
479 *alleghaniensis* is typically found in the Florida peninsula and coastal plains environments (Bailey
480 1995; Burbrink 2001), whereas *P. spiloides* is found throughout the remainder of the forested
481 habitats east of the Mississippi River including the ecoregions defined as interior river valleys
482 and hills, interior plateau, Appalachian habitats, southeastern and southcentral plains, and
483 Mississippi River valley. Uncertainty in the identity of taxa using morphology where these taxa
484 meet is likely due to extensive gene flow in hybrid zones at barriers following postglacial range
485 expansion. Finer-scale testing is necessary to tease apart the effects of these river systems,
486 ancient embayments, and connections to uplifted areas. Although the timing of divergence
487 predates the last interglacial, it is possible that repeated 100,000 year glacial cycles isolated these
488 taxa into refugia, as has been suggested previously for ratsnakes and other organisms (Fig.3;
489 Waltari et al. 2007; Noss et al. 2015).

490 Divergence between *P. bairdi* and *P. obsoletus* occurred in the western edge of the
491 Edwards Plateau, with the former extending westward into xeric forested or Chihuahuan desert
492 scrub associated with rocky habitats, whereas the latter extends east into Texas, preferring
493 forested habitats, (Lawson and Lieb 1990; Werler and Dixon 2000). Population genetic structure

494 between these taxa is associated with elevation, the last glacial maximum, ecoregion differences,
495 and geographic distance (Fig. 4). Dates of origin of these taxa are within the mid-Pleistocene,
496 suggesting that past climate played a strong role in promoting population divergence. During this
497 period, the Chihuahuan Desert was much drier relative to the late-Pleistocene (Graham and Mead
498 1987; Metcalfe et al. 2002; Metcalfe 2006) and this increased aridity could have driven
499 ecological divergence between this species pair. While divergence time is not the oldest here,
500 migration rates are lowest, and the hybrid zone is narrow. Previous research shows limited
501 hybridization with backcrosses between these taxa occurring in a narrow region of the
502 southwestern Edwards Plateau (Lawson and Lieb 1990; Vandewege et al. 2012).
503 Despite each lineage occupying uniquely different habitats, population responses to
504 Pleistocene climate change were similar (Fig. 3). Given the divergence time for each group, N_e
505 has expanded synchronously by 50 times since the Pleistocene. These estimates are consistent
506 with other predictions in the ENA, where 75% of the tested vertebrates show population size
507 expansion and 75% of tested snakes show synchronous expansion of N_e, likely coinciding with
508 the retreat of glaciers at various times in the Pleistocene or during the Holocene (Bintanja and
509 van de Wal 2008; Burbrink et al. 2016). However, *Pantherophis bairdi* has a range only in a
510 small area of the west Texas and isolated populations in northeastern Mexico, where habitats
511 may have been indirectly affected by glacial cycles.
512
513 *Taxonomy and delimitation in the genomic age*
514 Delimiting species using genetic data, while useful, has proved controversial (Bauer et al.
515 2011; Sukumaran and Knowles 2017; Leaché et al. 2019). In particular, most coalescent
516 delimitation methods fail to account for gene flow (but see Flouris et al. 2019), do not consider

517 spatial information, and place individuals into predefined groups prior to testing (O'Meara
518 2010). These methods all assume species diverge at an evolutionary time scale where mutations
519 accumulate through drift or selection (Rannala 2015). And while simulations show that a
520 threshold above one migrating individual/10 generations will cause model-based delimitation
521 methods to fail at finding two species (Zhang et al. 2011), it is not clear how pulses or
522 inconsistent migration affect species transitioning through the gray zone. Importantly, Wright
523 (1931) indicated that up to 1,000 individuals migrating/generation may not prevent populations
524 from drifting given effective population size and spatial segregation. Also, this does not consider
525 the effects of selection and recombination producing variable introgression among loci.
526 Modern thresholds for species delimitation (Jackson et al. 2017; Leaché et al. 2019) use
527 GDI (>0.7), estimates of species divergence time over population size ($2\tau/\theta>1$), absolute
528 divergence time (10^4) and maximum migration rates ($M=Nm<1$), metrics that are all likely
529 influenced by sampling effort, location of samples sequenced, and detection of hybrids. For
530 example, if large hybrid zones exist but remain poorly sampled then estimates of migration will
531 be reduced, GDI will be high, and the probability of MSC estimators will suggest multiple
532 species. Conversely, if only hybrid zones are sampled then all metrics will predict only a single
533 species. For all pairs of ratsnakes $2\tau/\theta>1$ and divergence time is greater than 10^4. However,
534 overall migration rates remain high between adjacent pairs in the eastern lineages. This forces us
535 to address how we are defining a species in the context of hybrid zones, how we infer
536 reproductive isolation beyond MSC methods, and how selection is maintaining hybrid zones. For
537 all lineages of ratsnakes, hybrid zones are smaller than predicted by neutral selection if these
538 hybrid zones formed in the Pleistocene. However, geographic lineage identity has been
539 maintained for >133,000 generations in all pairs, regardless of migration. Moreover, where

540 hybrid zones extend over large areas, gene flow likely varies throughout space and may have
541 changed over time (Barton and Hewitt 1985; Mallet 2007), particularly in response to glacial
542 cycles. Thus, identifying and delimiting species in the context of genetic admixture alone may be
543 difficult without understanding the history of gene flow, and as shown here, specific locus
544 introgression.

545 We have demonstrated that these four species, delimited previously using mtDNA and
546 morphology (Burbrink, 2001; Burbrink et al., 2000) can be recognized as distinct taxa given
547 methods of spatial grouping and species delimitation, though ranges in the eastern-most lineage
548 are restricted further south than previously thought. However, migration rates between *P.*
549 *spiloides* and *P. alleghaniensis* (Fig. 3) may be high enough that they do not represent
550 independent evolutionary trajectories (de Queiroz, 2007; Rannala, 2015; Zhang et al., 2011),
551 though differential introgression among some might suggest that reproductive isolation is
552 maintained for some loci. Unique niches in the subtropical areas of the southeastern US and
553 morphology provides some evidence for isolation between these taxa: the striped forms (with
554 variable ground colors of orange, yellow, green, and gray) are indicative of *P. alleghaniensis*
555 whereas adults with dark saddle patterns on grey, olive, brown, or black ground colors (or
556 completely black with no pattern) are found west of Florida and east of the Mississippi River,
557 throughout the Midwest and up to the northeastern US and Canada represent *P. spiloides*.
558 However, hybridization north and west of Florida is likely represented by a dulling of this yellow
559 ground color and mixed patterned individuals (Schultz 1996; see Supporting Information
560 Material 1 for more information on taxonomy).

561 Modern genomic tools have advanced our understanding of timing of divergence,
562 population isolation, lineage connectivity, and the nature of hybrid zones. As we show with the

eastern ratsnakes, we can better comprehend processes of speciation by integrating studies of hybrid zones to understand changes in admixture and differential introgression with coalescent estimates of isolation and migration. We also show that timing, rates of gene flow, and differential introgression at hybrid zones are unique to each species pair, likely traversing different genic and genomic scales of divergence in the gray zone across changing environments in the Eastern Nearctic. Future studies should examine clines with more detailed sampling using whole genomes to connect weakly introgressing loci with traits.

Literature Cited

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Agrawal, A. F., J. L. Feder, and P. Nosil. 2011. Ecological divergence and the origins of intrinsic postmating isolation with gene flow. Int. J. Ecol. 2011:1–15. Hindawi.

Bailey, R. G. 1995. Description of the ecoregions of the United States. 2nd ed. USDA Forest Service.

Bailey, R. I., M. R. Tesaker, C. N. Trier, and G.-P. Saetre. 2015. Strong selection on male plumage in a hybrid zone between a hybrid bird species and one of its parents. J. Evol. Biol.

28:1257–1269. John Wiley & Sons, Ltd (10.1111).

Barton, H. H., and K. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* Hybrid Zones and the Evolutionary Process. Oxford University Press, New York.

Barton, N. H. 2010. What role does natural selection play in speciation? Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365:1825–40. The Royal Society.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annu. Rev. Ecol. Syst. 16:113–148. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA .

Bauer, A. M., J. F. Parham, R. M. Brown, B. L. Stuart, L. Grismer, T. J. Papenfuss, W. Böhme, J. M. Savage, S. Carranza, J. L. Grismer, P. Wagner, A. Schmitz, N. B. Ananjeva, and R. F. Inger. 2011. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. Proc. R. Soc. B Biol. Sci. 278:490. The Royal Society.

Bintanja, R., and R. S. W. van de Wal. 2008. North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. Nature 454:869–72. Nature Publishing Group.

Blouin-Demers, G., and P. J. Weatherhead. 2002. Implications of movement patterns for gene flow in black rat snakes (*Elaphe obsoleta*). Can. J. Zool. 80:1162–1172. NRC Research Press Ottawa, Canada .

Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Modell. 153:51–68. Elsevier.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537. Public Library of Science.

Bradburd, G. S., G. M. Coop, and P. L. Ralph. 2018. Inferring continuous and discrete

population genetic structure across space. Genetics 210:33–52. Genetics.

Brandley, M. C., T. J. Guiher, R. A. Pyron, C. T. T. Winne, and F. T. Burbrink. 2010. Does dispersal across an aquatic geographic barrier obscure phylogeographic structure in the diamond-backed watersnake (*Nerodia rhombifer*)? Mol. Phylogenet. Evol. 57:552–560.

Brown, J. L., D. J. Hill, A. M. Dolan, A. C. Carnaval, and A. M. Haywood. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Sci. Data 5:180254. Nature Publishing Group.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Burbrink, F. T. 2002. Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. Mol. Phylogenet. Evol. 25:465–476.

Burbrink, F. T. 2001. Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). Herpetol. Monogr. 1–53.

Burbrink, F. T., Y. L. Chan, E. A. Myers, S. Ruane, B. T. Smith, and M. J. Hickerson. 2016. Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. Ecol. Lett. 19:1457–1467.

Burbrink, F. T., F. Fontanella, R. A. Pyron, T. J. Guiher, C. Jimenez, and R. Alexander Pyron. 2008. Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes : Colubridae : *Coluber constrictor*). Mol. Phylogenet. Evol. 47:274–288.

Burbrink, F. T., F. G. Grazziotin, R. A. Pyron, D. Cundall, S. Donnellan, F. Irish, J. S. Keogh, F.

632 Kraus, R. W. Murphy, B. Noonan, C. J. Raxworthy, S. Ruane, A. R. Lemmon, E. M.
633 Lemmon, and H. Zaher. 2020. Interrogating genomic-scale data for Squamata (lizards,
634 snakes, and amphisbaenians) shows no support for key traditional morphological
635 relationships. Syst. Biol. 69:502–520.
636 Burbrink, F. T., R. Lawson, and J. B. Slowinski. 2000. Mitochondrial DNA phylogeography of
637 the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies
638 concept. Evolution (N. Y). 54:2107–2118.
639 Carneiro, M., S. J. E. Baird, S. Afonso, E. Ramirez, P. Tarroso, H. Teotónio, R. Villafuerte, M.
640 W. Nachman, and N. Ferrand. 2013. Steep clines within a highly permeable genome across
641 a hybrid zone between two subspecies of the European rabbit. Mol. Ecol. 22:2511–25. Mol
642 Ecol.
643 Case, T. J., R. D. Holt, M. A. McPeek, and T. H. Keitt. 2005. The community context of species'
644 borders: ecological and evolutionary perspectives. Oikos 108:28–46.
645 Case, T. J., and M. L. Taper. 2000. Interspecific competition, environmental gradients, gene
646 flow, and the coevolution of species' borders. Am. Nat. 155:583–605.
647 Caye, K., F. Jay, O. Michel, and O. François. 2018. Fast inference of individual admixture
648 coefficients using geographic data. Ann. Appl. Stat. 12:586–608. Institute of Mathematical
649 Statistics.
650 Coates, D. J., M. Byrne, and C. Moritz. 2018. Genetic diversity and conservation units: dealing
651 with the species-population continuum in the age of genomics. Front. Ecol. Evol. 6:165.
652 Frontiers.
653 Cox, R. T., D. N. Lumsden, and R. B. Van Arsdale. 2014. Possible relict meanders of the
654 Pliocene Mississippi River and their implications. J. Geol. 122:609–622.

de Queiroz, K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. hzar : hybrid zone analysis using an R software package. Mol. Ecol. Resour. 14:652–663.

Diniz-Filho, J. A. F., T. N. Soares, J. S. Lima, R. Dobrovolski, V. L. Landeiro, M. P. D. C. Telles, T. F. Rangel, and L. M. Bini. 2013. Mantel test in population genetics. Genet. Mol. Biol. 36:475–485.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14:927–930. Blackwell Publishing.

Edwards, S., R. Hopkins, and J. Mallet. 2020. Speciation. Pp. 296–318 *in* S. Scheiner and D. Mindell, eds. The Theory of Evolution. University of Chicago Press, Chicago.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726. Narnia.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350. Elsevier Current Trends.

Flouris, T., X. Jiao, B. Rannala, and Z. Yang. 2019. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Mol. Biol. Evol., doi: 10.1093/molbev/msz296.

Frankham, R., J. D. (Jonathan D. . Ballou, and D. A. (David A. Briscoe. 2010. Introduction to conservation genetics. Cambridge University Press.

Frantz, A. C., L. C. Pope, T. R. Etherington, G. J. Wilson, and T. Burke. 2010. Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. Mol. Ecol. 19:1663–1674. John Wiley & Sons, Ltd

(10.1111).

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton University Press.

Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution (N. Y). 62:2789–2806. John Wiley & Sons, Ltd (10.1111).

Gehara, M., A. A. Garda, F. P. Werneck, E. F. Oliveira, E. M. da Fonseca, F. Camurugi, F. d. M. Magalhães, F. M. F. M. Lanna, J. W. Sites, R. Marques, R. Silveira-Filho, V. A. São Pedro, G. R. Colli, G. C. Costa, and F. T. Burbrink. 2017. Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. Mol. Ecol. 26:4756–4771.

Gourbière, S., and J. Mallet. 2010. Are species real? The shape of the species boundry with exponential failure, reinforcement, and the "missing snowball." Evolution (N. Y). 64:1–24. John Wiley & Sons, Ltd.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338. Oxford Academic.

Graham, R. W., and J. I. Mead. 1987. Environmental fluctuations and evolution of mammalian faunas during the last deglaciation in North America. Pp. 371–402 *in* North America and Adjacent Oceans During the Last Deglaciation. Geological Society of America, Boulder, Colorado 80301.

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press.

Harrison, R. G. 2012. The language of speciation. Evolution (N. Y). 66:3643–3657. John Wiley & Sons, Ltd.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

701 Hewitt, G. M. 2011. Quaternary phylogeography: the roots of hybrid zones. Genetica 139:617–
702 638. Springer.
703 Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and
704 speciation. Biol. J. Linn. Soc. 58:247–276.
705 Hewitt, G. M. 2008. Speciation, hybrid zones and phylogeography - or seeing genes in space and
706 time. Mol. Ecol. 10:537–549. John Wiley & Sons, Ltd (10.1111).
707 Hey, J. 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol.
708 27:905–920.
709 Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration
710 rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura*
711 and *D-persimilis*. Genetics 167:747–760.
712 Hijmans, R. J., J. van Etten, and J. Checg. 2014. Geographic data analysis and modeling [R
713 package raster version 2.5-8]. Comprehensive R Archive Network (CRAN).
714 Jackson, N. D., B. C. Carstens, A. E. Morales, and B. C. O'Meara. 2017. Species delimitation
715 with gene flow. Syst. Biol. 66:799–812. Oxford Academic.
716 Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol. Evol.
717 15:250–255. Elsevier Current Trends.
718 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
719 Bioinformatics 24:1403–1405.
720 Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a
721 new method for the analysis of genetically structured populations. BMC Genet. 11:94.
722 BioMed Central.
723 Joseph, L., and T. Wilke. 2007. Lack of phylogeographic structure in three widespread

Australian birds reinforces emerging challenges in Australian historical biogeography. J. Biogeogr. 34:612–624.

Karrenberg, S., X. Liu, E. Hallander, A. Favre, J. Herforth-Rahmé, and A. Widmer. 2019. Ecological divergence plays an important role in strong but complex reproductive isolation in campions (Silene). Evolution (N. Y). 73:245–261. John Wiley & Sons, Ltd (10.1111).

Kuhn, M. 2008. Caret: Classification and Regression Training Package. R package version 6.0-77.

Landgraf, A. J., and Y. Lee. 2015. Dimensionality Reduction for Binary Data through the Projection of Natural Parameters.

Lawson, R., and C. S. Lieb. 1990. Variation and hybridization in *Elaphe bairdi* (Serpentes: Colubridae). J. Herpetol. 24:280.

Leaché, A. D., T. Zhu, B. Rannala, and Z. Yang. 2019. The spectre of too many species. Syst. Biol. 68:168–181. Narnia.

Legendre, P., and M. J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol. Ecol. Resour. 10:831–844.

Legendre, P., J. Oksanen, and C. J. F. ter Braak. 2011. Testing the significance of canonical axes in redundancy analysis. Methods Ecol. Evol. 2:269–277. John Wiley & Sons, Ltd (10.1111).

Lek, S., M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga, and S. Aulagnier. 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Modell. 90:39–52.

Lek, S., and J. F. Gue. 1999. Artificial neural networks as a tool in ecological modelling , an introduction. 120:65–73.

Libbrecht, M. W., and W. S. Noble. 2015. Machine learning applications in genetics and genomics. Nat. Rev. Genet. 16:321–332.

Linck, E., and C. J. Battey. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol. Ecol. Resour. 19:639–647. John Wiley & Sons, Ltd (10.1111).

Lourenço, C. R., K. R. Nicastro, C. D. McQuaid, R. M. Chefaoui, J. Assis, M. Z. Taleb, and G. I. Zardi. 2017. Evidence for rangewide panmixia despite multiple barriers to dispersal in a marine mussel. Sci. Rep. 7:10279. Nature Publishing Group.

Macholán, M., S. J. E. Baird, P. Dufková, P. Munclinger, B. V. Bímová, and J. Piálek. 2011. Assessing multilocus introgression patterns: a case study on the mouse X chromosome in Central Europe. Evolution (N. Y). 65:1428–1446. John Wiley & Sons, Ltd.

Makowsky, R., J. Chesser, and L. J. Rissler. 2009. A striking lack of genetic diversity across the wide-ranging amphibian *Gastrophryne carolinensis* (Anura: Microhylidae). Genetica 135:169–183. Springer Netherlands.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283. Nature Publishing Group.

Mason, N. A., N. K. Fletcher, B. A. Gill, W. C. Funk, and K. R. Zamudio. 2020. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Syst. Biodivers. 18:269–280. Taylor & Francis.

Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University Press, Cambridge,.

McEntee, J. P., J. G. Burleigh, and S. Singhal. 2018. Dispersal predicts hybrid zone widths across animal diversity: Implications for species borders under incomplete reproductive isolation. bioRxiv 472506. Cold Spring Harbor Laboratory.

770 McGaughran, A., K. Morgan, and R. J. Sommer. 2014. Environmental variables explain genetic
771 structure in a beetle-associated nematode. PLoS One 9:e87317. Public Library of Science.
772 Metcalfe, S. E. 2006. Late Quaternary environments of the northern deserts and central
773 transvolcanic belt of Mexico. Ann. Missouri Bot. Gard. 93:258–273. Missouri Botanical
774 Garden Press.
775 Metcalfe, S., A. Say, S. Black, R. McCulloch, and S. O'Hara. 2002. Wet Conditions during the
776 Last Glaciation in the Chihuahuan Desert, Alta Babicora Basin, Mexico. Quat. Res. 57:91–
777 101.
778 Myers, E. A., R. W. Bryson, R. W. Hansen, M. L. Aardema, D. Lazcano, and F. T. Burbrink.
779 2019a. Exploring Chihuahuan Desert diversification in the gray-banded kingsnake,
780 *Lampropeltis alterna* (Serpentes: Colubridae). Mol. Phylogenet. Evol. 131:211–218.
781 Myers, E. A., A. D. McKelvy, and F. T. Burbrink. 2020. Biogeographic barriers, Pleistocene
782 refugia, and climatic gradients in the southeastern Nearctic drive diversification in
783 cornsnakes (*Pantherophis guttatus* complex). Mol. Ecol. 29:797–811. John Wiley & Sons,
784 Ltd.
785 Myers, E. A., A. T. Xue, M. Gehara, C. L. Cox, A. R. Davis Rabosky, J. Lemos-Espinal, J. E.
786 Martínez-Gómez, and F. T. Burbrink. 2019b. Environmental heterogeneity and not vicariant
787 biogeographic barriers generate community-wide population structure in desert-adapted
788 snakes. Mol. Ecol. 28:4535–4548.
789 Noguerales, V., P. J. Cordero, and J. Ortego. 2016. Hierarchical genetic structure shaped by
790 topography in a narrow-endemic montane grasshopper. BMC Evol. Biol. 16:96. BioMed
791 Central.
792 Nosil, P. 2012. Ecological Speciation. Oxford University Press, Oxford.

793 Nosil, P., and B. J. Crespi. 2006. Ecological divergence promotes the evolution of cryptic
794 reproductive isolation. Proc. R. Soc. B-Biological Sci. 273:991–997.
795 Nosil, P., and J. L. Feder. 2013. Genome evolution and speciation: toward quantitative
796 descriptions of pattern and process. Evolution 67:2461–7.
797 Nosil, P., J. L. Feder, S. M. Flaxman, and Z. Gompert. 2017. Tipping points in the dynamics of
798 speciation. Nat. Ecol. Evol. 1:1. Nat Ecol Evol.
799 Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous
800 genomic divergence. Mol. Ecol. 18:375–402.
801 Noss, R. F., W. J. Platt, B. A. Sorrie, A. S. Weakley, D. B. Means, J. Costanza, and R. K. Peet.
802 2015. How global biodiversity hotspots may go unrecognized: lessons from the North
803 American Coastal Plain. Divers. Distrib. 21:236–244.
804 O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree
805 inference. Syst. Biol. 59:59–73.
806 Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities.
807 Genetics 139:1805–13. Genetics Society of America.
808 Pante, E., N. Puillandre, A. Viricel, S. Arnaud-Haond, D. Aurelle, M. Castelin, A. Chenuil, C.
809 Destombe, D. Forcioli, M. Valero, F. Viard, and S. Samadi. 2015. Species are hypotheses:
810 avoid connectivity assessments based on pillars of sand. Mol. Ecol. 24:525–544. John
811 Wiley & Sons, Ltd.
812 Petkova, D., J. Novembre, and M. Stephens. 2016. Visualizing spatial population structure with
813 estimated effective migration surfaces. Nat. Genet. 48:94–100. Nature Publishing Group.
814 Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and
815 output analysis for MCMC. R News 6:7–11.

816 Pyron, R. A., and F. T. Burbrink. 2009. Lineage diversification in a widespread species: roles for
817 niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. Mol.
818 Ecol. 18:3443–3457.
819 R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation
820 for Statistical Computing., Vienna, Austria.
821 Rannala, B. 2015. The art and science of species delimitation. Curr. Zool. 61:846–853.
822 Rannala, B., and Z. Yang. 2020. Species Delimitation. Pp. 5.5:1--5.5:18 *in* C. Scornavacca, F.
823 Delsuc, and N. Galtier, eds. Phylogenetics in the Genomic Era. No commercial publisher |
824 Authors open access book.
825 Remington, C. L. 1968. Suture-zones of hybrid interaction between recently joined biotas. Pp.
826 321–428 *in* Evolutionary Biology. Springer US, Boston, MA.
827 Riddle, B. R. 2016. Comparative phylogeography clarifies the complexity and problems of
828 continental distribution that drove A. R. Wallace to favor islands. Proc. Natl. Acad. Sci. U.
829 S. A. 113:7970. National Academy of Sciences.
830 Robison, H. W. 1986. Zoogeographic implications of the Mississippi River Basin. Pp. 267–285
831 *in* The Zoogeography of North American Freshwater Fishes. Wiley, New York.
832 Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light
833 on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol.
834 14:e2000234. Public Library of Science.
835 Satler, J. D., and B. C. Carstens. 2017. Do ecological communities disperse across biogeographic
836 barriers as a unit? Mol. Ecol. 26:3533–3545. John Wiley & Sons, Ltd (10.1111).
837 Schultz, K. D. 1996. A monograph of the colubrid snakes of the genus *Elaphe* Fitzinger. Koelz
838 Scientific Books, Havlicuv.

839 Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L.
840 Peichel, G.-P. Saetre, C. Bank, A. Brannstrom, A. Brelsford, C. S. Clarkson, F.
841 Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C.
842 Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews,
843 J. I. Meier, M. Most, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T.
844 Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat
845 Rev Genet 15:176–192. Nature Publishing Group.
846 Sheehan, S., Y. S. Song, E. Buzbas, D. Petrov, A. Boyko, and A. Auton. 2016. Deep learning for
847 population genetic inference. PLOS Comput. Biol. 12:e1004845. Public Library of Science.
848 Singhal, S., and C. Moritz. 2013. Reproductive isolation between phylogeographic lineages
849 scales with divergence. Proceedings. Biol. Sci. 280:20132246. The Royal Society.
850 Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative
851 phylogeography of unglaciated eastern North America. Mol. Ecol. 15:4261–4293.
852 Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld.
853 2019. Widespread selection and gene flow shape the genomic landscape during a radiation
854 of monkeyflowers. PLOS Biol. 17:e3000391. Public Library of Science.
855 Stankowski, S., J. M. Sobel, and M. A. Streisfeld. 2017. Geographic cline analysis as a tool for
856 studying genome-wide variation: a case study of pollinator-mediated divergence in a
857 monkeyflower. Mol. Ecol. 26:107–122.
858 Sukumaran, J., and L. L. Knowles. 2017. Multispecies coalescent delimits structure, not species.
859 Proc. Natl. Acad. Sci. U. S. A. 114:1607–1612. National Academy of Sciences.
860 Sun, K., K. A. Meiklejohn, B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball. 2014.
861 The evolution of peafowl and other taxa with ocelli (eyespots): a phylogenomic approach.

Proc. R. Soc. B Biol. Sci. 281:20140823. The Royal Society.

Swenson, N. G., and D. J. Howard. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am. Nat. 166:581–591.

Vandewege, M. W., D. Rodriguez, J. P. Weaver, T. D. Hibbitts, M. R. J. Forstner, and L. D. Densmore. 2012. Evidence of Hybridization between *Elaphe bairdi* and *Elaphe obsoleta lindheimeri* Including Comparative Population Genetics Inferred from Microsatellites and Mitochondrial DNA. J. Herpetol. 46:56–63. Society for the Study of Amphibians and Reptiles.

Vavrek, M. J. 2020. Fossil. https://cran.r-project.org/web/packages/fossil/index.html.

Walker, D., and J. C. Avise. 1998. Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. Annu. Rev. Ecol. Syst. 29:23–58.

Waltari, E., R. J. Hijmans, A. T. Peterson, Á. S. Nyári, S. L. Perkins, and R. P. Guralnick. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS One 2:e563. Public Library of Science.

Werler, J. E., and J. R. Dixon. 2000. Texas snakes : identification, distribution, and natural history. 1st ed. University of Texas Press, Austin.

Wright, S. 1978. Evolution and the Genetics of Populations. Vol. 4. Variability within and among Natural Populations. University Chicago Press.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159. Genetics Society of America.

Wright, S. 1943. Isolation by distance. Genetics 28:114–38. Genetics Society of America.

Younger, J. L., P. Dempster, Á. S. Nyári, T. O. Helms, M. J. Raherilalao, S. M. Goodman, and S. Reddy. 2019. Phylogeography of the Rufous vanga and the role of bioclimatic transition

zones in promoting speciation within Madagascar. Mol. Phylogenet. Evol. 139:106535. Academic Press.

Zellmer, A. J., M. M. Hanes, S. M. Hird, and B. C. Carstens. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant Sarracenia alata. Syst. Biol. 61:763–777.

Zhang, C., D.-X. Zhang, T. Zhu, and Z. Yang. 2011. Evaluation of a bayesian coalescent method of species delimitation. Syst. Biol. 60:747–761.

Zhang, W. 2010. Computational Ecology: Artificial Neural Networks and Their Applications. World Scientific.

Acknowledgments: We thank R. Glor, R. Brown, and L. Welton at KU and C. Austin and D. Dittman at LSUMN and K. Krysko and FLMNH, D. Shepard at Louisiana Tech University, and A. McKelvy for providing tissues. We also thank L. and G. Derryberry for help with HZAR. We also thank E. Chen for help extracting DNA. EAM was funded by the Gerstner Scholar/Theodore Roosevelt and Peter Buck/Walter Rathbone Bacon fellowships. We thank D. Frost, A. Leaché, A. Bauer, and K. deQuieroz for helpful discussions on species delimitation, type specimens, and admixed type specimens.

Data Accessibility Statement: All data and unique scripts are available on Dryad XXX.

Title: Resolving spatial complexities of hybridization in the context of the gray zone of

950 Table 1. Measures of divergence dates in millions of years (My), migration rates (first number is
951 first taxon migrating into second taxon, e.g., *Pantherophis alleganiensis* into *P.spiloides*, and
952 second represents second species migrating into the first), and cline widths for both spatial
953 distance (Km) and environmental distance for all species-pair comparisons in the *Pantherophis*
954 *obsoletus* complex. Percentage of median width over total sampled distance are shown for both
955 spatial and environmental clines
956

Group	Age (My)	Migration Rate (ind/gen)	Cline Width (Km)	Cline Width (Env)
Eastern (*P. alleghaniensis* x *P.spiloides*)	1.95 (0.674-0.2.95)	2.79/3.79	256 (168-356); 14.8%	0.81 (0.32-1.45); 4.0%
Central (*P.spiloides* x *P.obsoletus*)	3.18 (1.65-5.02)	1.96/1.28	147 (40-283); 7.2%	0.13 (0.06-0.24); 5.0%
Western (*P.obsoletus* x *P. bairdi*)	1.10 (0.407-02.62)	0.65/0.55	21 (2-46); 1.6%	4.4 (3.2-12.53); 2.1%

957

Figures

Fig.1. Distribution of species within the North American ratsnake complex. (A) Map showing range and geographic features: Chihuahuan Desert Province, the Mississippi River, and the Appalachian Mountains and Apalachicola/Chattahoochee River System (AARS). (B) Estimated effective migration surfaces (EEMS) showing areas of both high (blue) and low migration (brown). (C) Population structure using Discriminant Analysis of Principal Components (DAPC) shows the location of four geographically distinct lineages. (D) Bi-plots of PC space shows relative distances among the four taxa using DAPC.

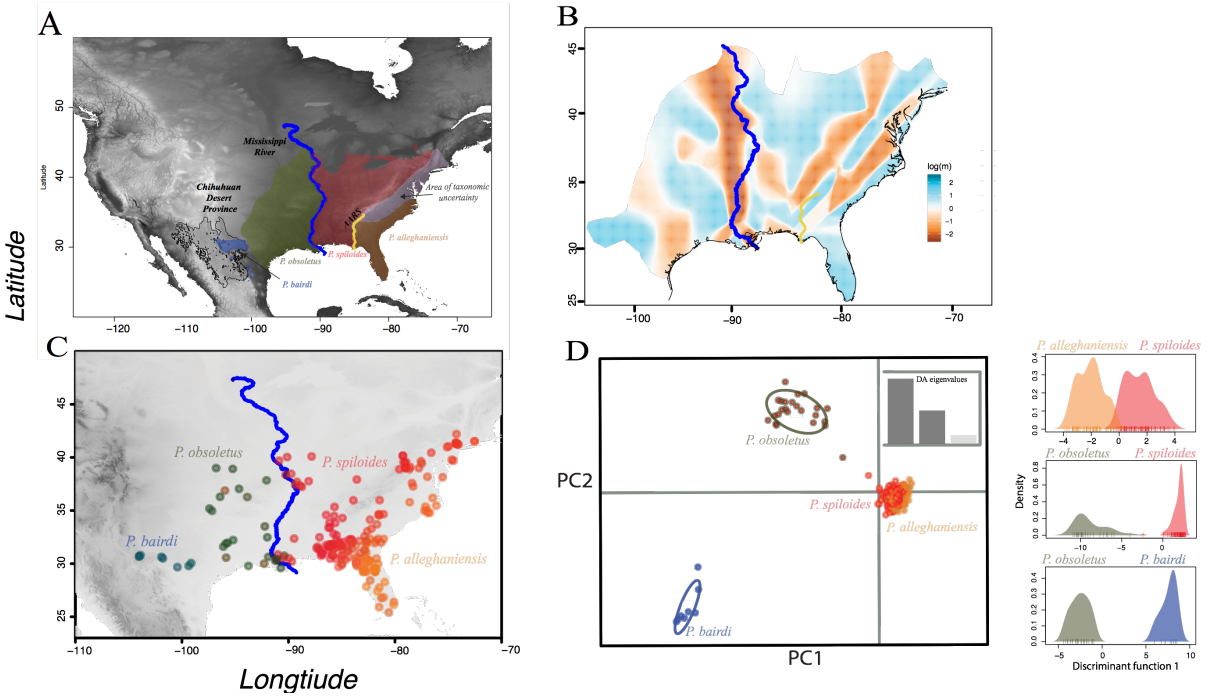
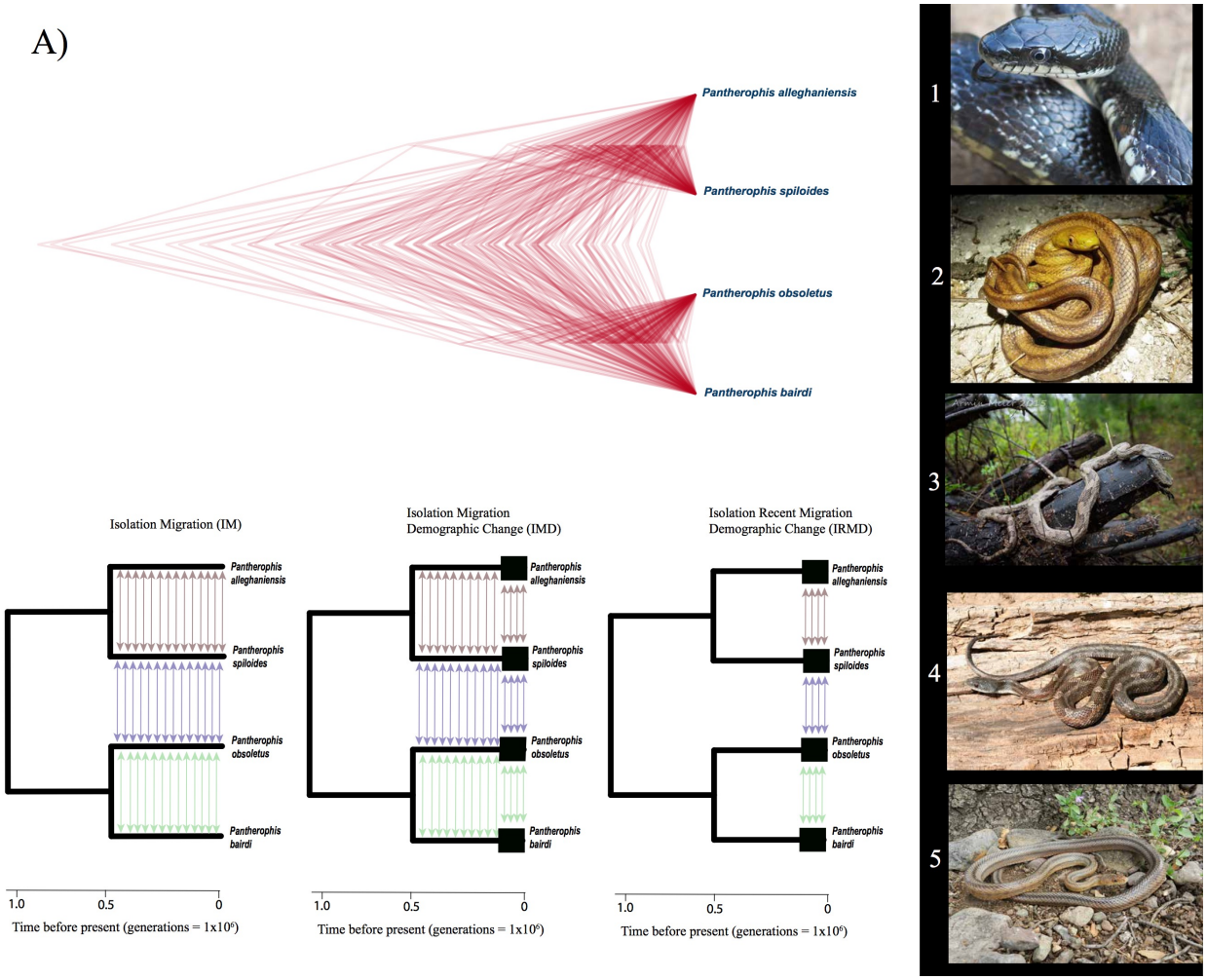


Fig.2. The results of SNAPP species-tree estimation for the four taxa over all loci (A) and models showing Isolation Migration (IM) and Isolation Migration with Demographic Change (IMD; best-fit model) and Isolation with Recent Migration and Demographic Change (IMRD) estimated in PipeMaster. Colored arrows represent migration among spatially adjacent pairs. (B). Photographs representing some of the typical color patterns seen in ratsnakes: 1) *Pantherophis alleghaniensis* (photo F. Burbrink), 2) *P. alleghaniensis* (photo N. Claunch), 3) *P. spiloides* (photo A. Meier), 4) *P. obsoletus* (photo D. Shepard), and 5) *P. bairdi* (E. Myers).

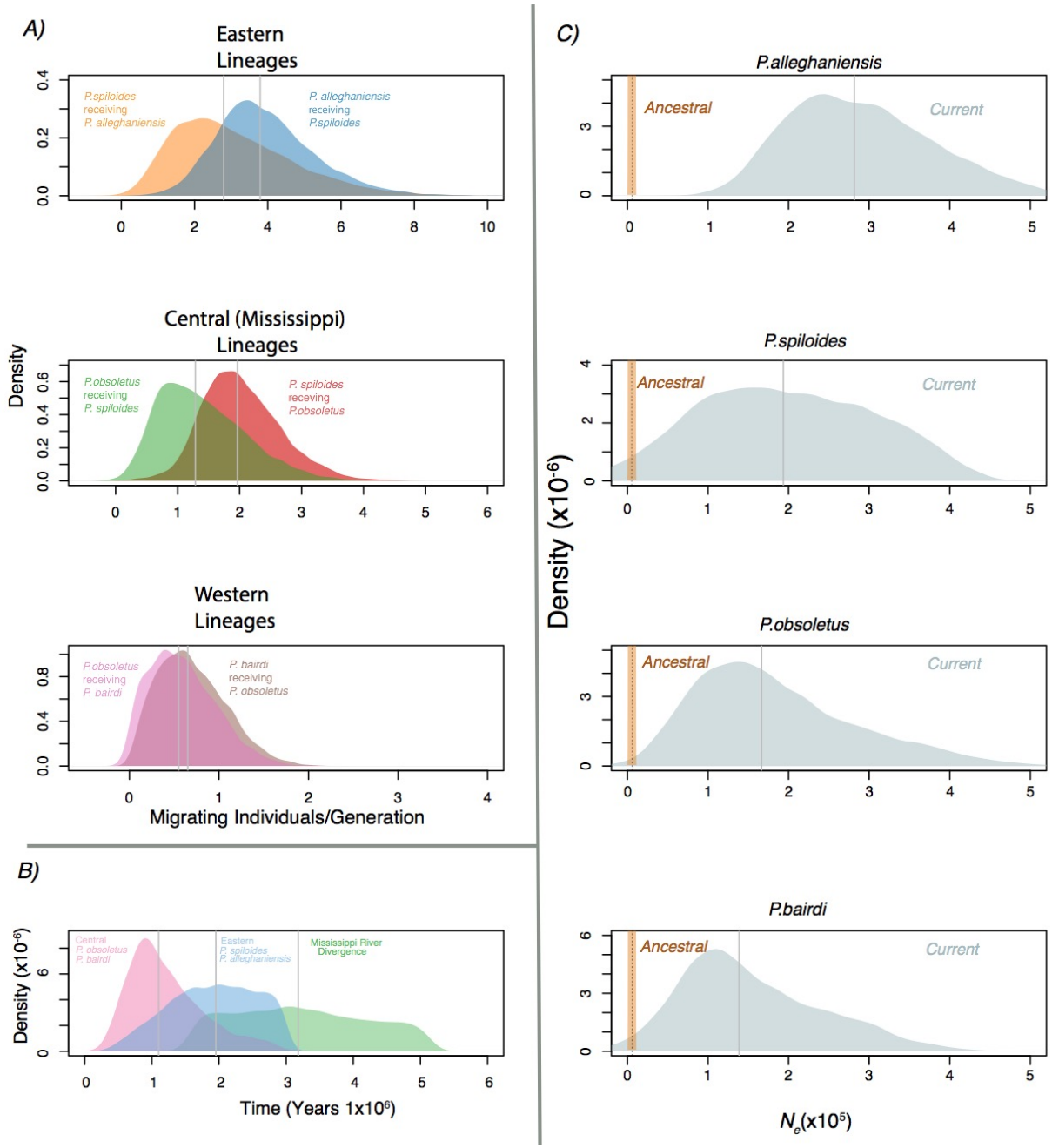


979 Fig.3. Results from PipeMaster showing estimates and directionality of migration (2Nm) (A),

980 divergence times between taxon pairs and across the Mississippi River (*P. alleghaniensis/ P.*

981 *spiloides* vs *P. obsoletus/P. bairdi*) (B), and changes in population sizes over time (C).

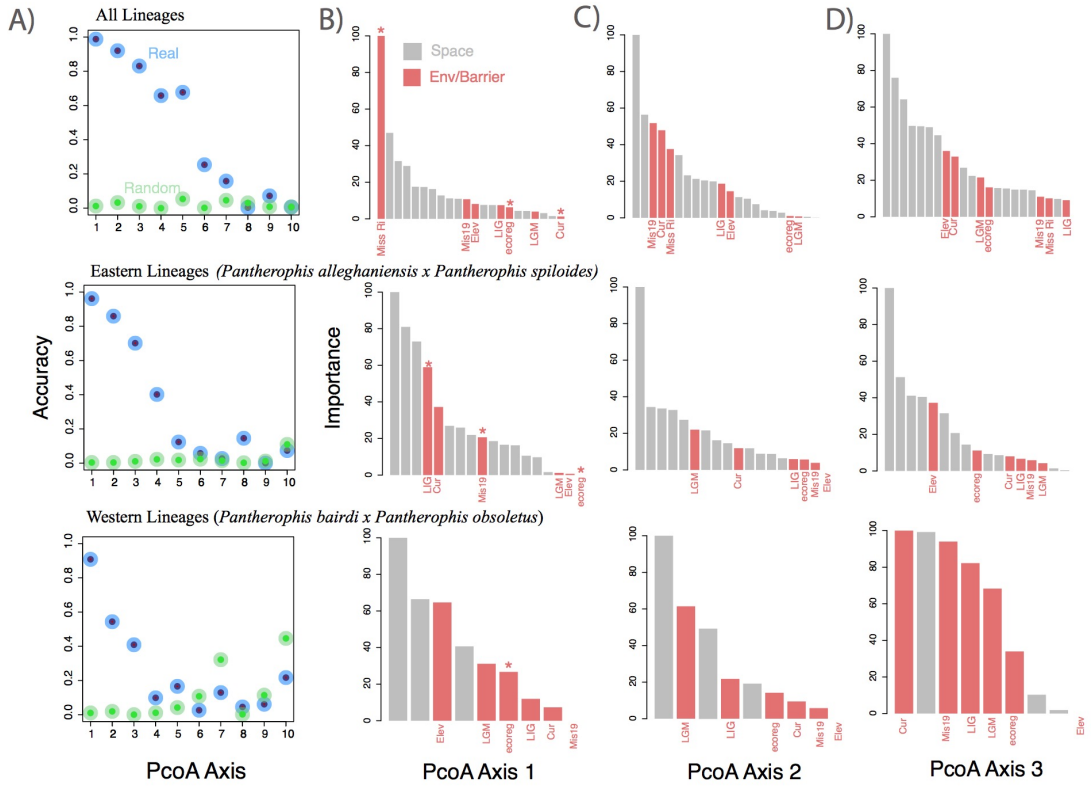
982



983

984

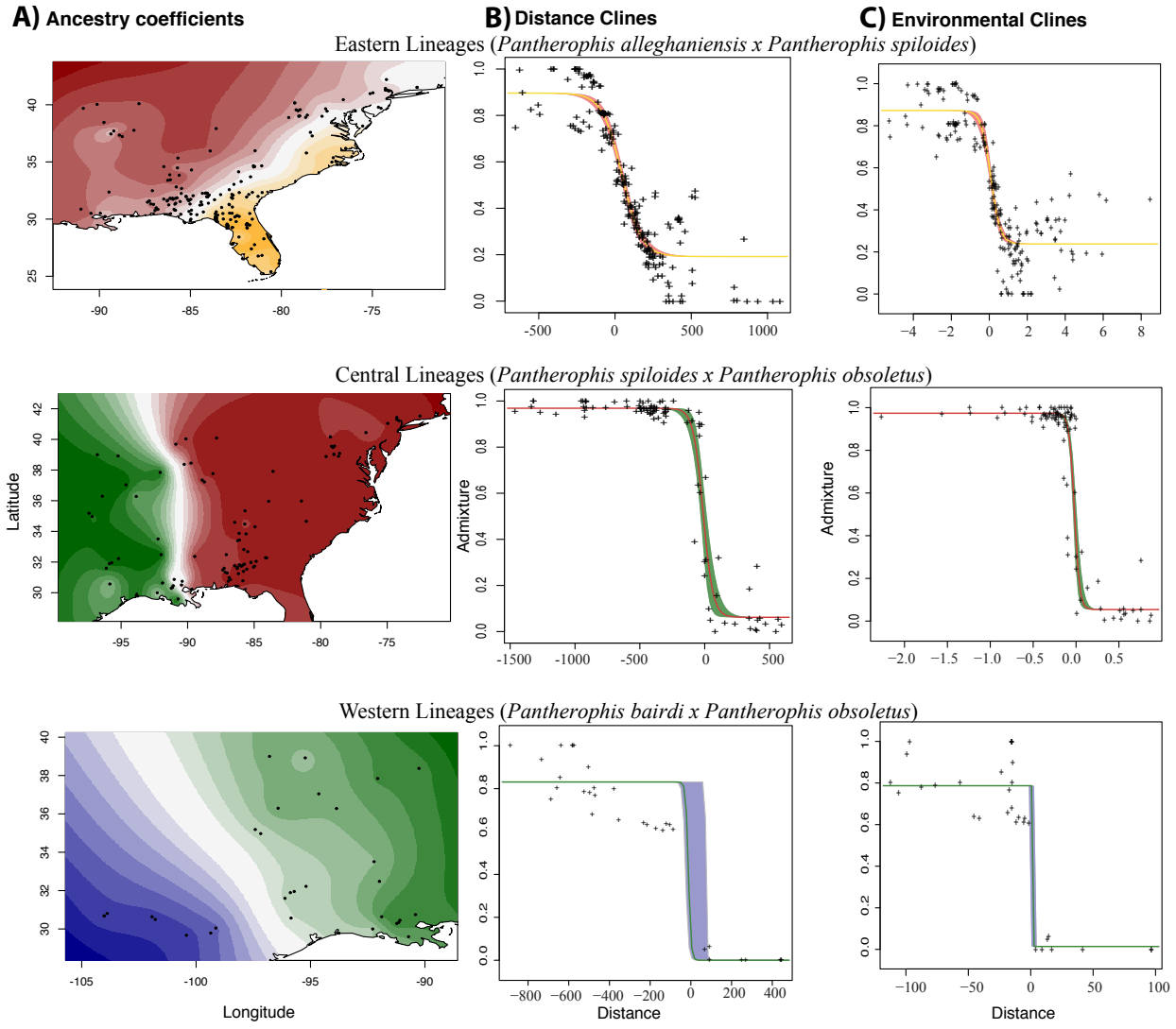
Fig.4. Artificial neural network results showing accuracy predicting genetic structure for 10 PCoA axes among all lineages and species pairs for real and randomly shuffled dependent variables (A). Variable importance are shown for spatial data (grey; principal coordinates neighbor matrices) and environmental charcteristics (red; B-D). Variable abbreviations are: Miss Ri (Mississippi River), Elev (elevation), ecoreg (ecoregions), Mis19 (Pleistocene Marine Isotope Stage 19 - 787 Kya), LIG (Last interglacial – 130 Kya), LGM (Last glacial maximum – 21Kya), and Cur (Current climate data). Asterisks above red columns are significant after corrected for space using redundancy analyses (RDA).



997 Fig.5. Using admixture from TESS3r (A), Gaussian cline (HZAR) model predictions are shown

998 for both spatial (B) and environmental distances (C) for all three species pairs.

999



1000

1001

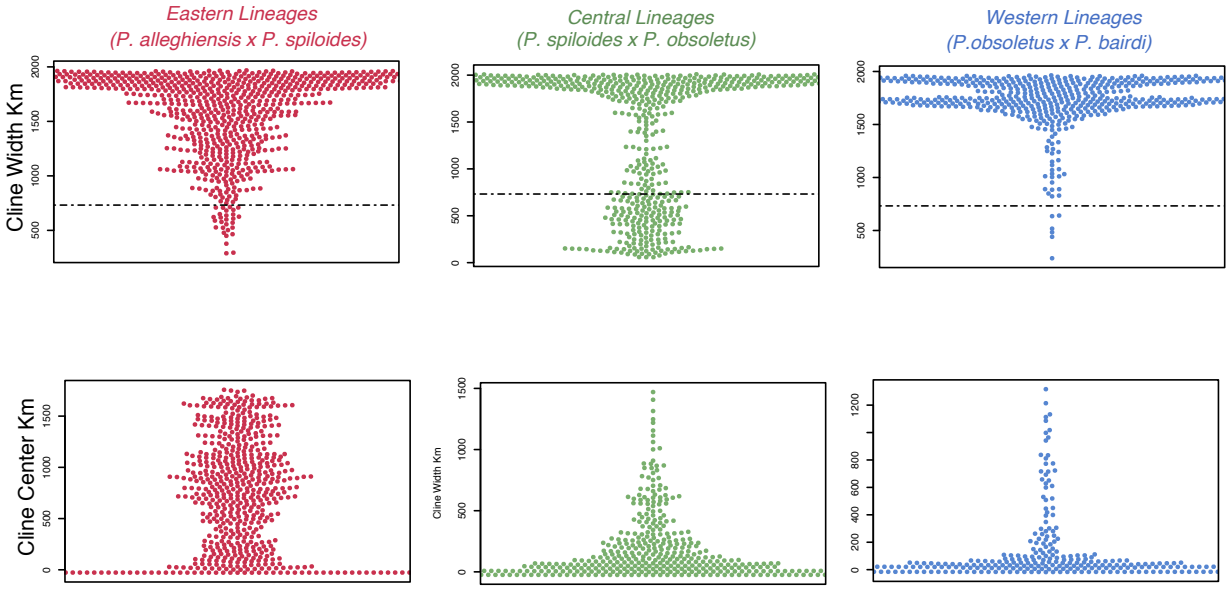
1002

1003

1004

1005

Fig.6. Beeswarm plots showing predicted cline widths and centers for all three species pairs over all loci using HZAR. Cline widths for loci lower than the predictions for neutrally forming hybrid zones since the last glacial maximum fall below the indicated dashed line.



1021

1022

1023

1024 **Supplemental Material**

1025 **Supporting Information Material 1**

1026

1027

1028 **Supplemental Methods and Results**

1029

1030 *Dataset*

1031 We sampled 288 individuals liberally covering the range of all species within the
1032 *Pantherophis obsoletus* complex (Fig.1; Dryad XXX). DNA was extracted from all samples
1033 using Qiagen DNeasy Blood & Tissue Kits and samples were screened for quality using broad-
1034 range Qubit Assays (https://www.thermofisher.com/us/en/home/industrial/spectroscopy-
1035 elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-assays.html). We
1036 used services from RAPiD Genomics (https://www.rapid-genomics.com/services/) to generate
1037 5472 baits and to sequence 5060 conserved elements (UCEs) loci following the protocols from
1038 (Faircloth et al. 2012) and (Sun et al. 2014).

1039 Raw sequence reads were trimmed of adapter contamination using illumiprocessor
1040 (Faircloth 2013), a wrapper around the trimmomatic package (Bolger et al. 2014). These
1041 sequence capture data were then assembled by mapping reads to all UCEs found within a
1042 Chromium 10x assembled *Pantherophis spiloides* genome (Burbrink and Myers, *in prep*) after
1043 2,500 base pairs of flanking DNA around UCE baits were pulled from this genome assembly

1044 using the Phyluce commands *phyluce_probe_run_multiple_lastzs_sqlite* and
1045 *phyluce_probe_slice_sequence_from_genomes* (Faircloth 2016). Sequence reads for each sample
1046 were then mapped to the 'reference' UCE dataset using bwa v0.7.17 (Li and Durbin 2009).
1047 Samtools was used to convert sam format to bam files and alignments were soft clipped using the
1048 CleanSam tool in Picard (http://broadinstitute.github.io/picard/). To phase these data, we mapped
1049 reads back to each individual sample's alignments. To do this, bam files were converted to fastq
1050 using samtools mpileup (Li et al. 2009), fastq files were converted to fasta format via seqtk
1051 (https://github.com/lh3/seqtk) and all missing sites were removed using seqkit (Shen et al 2016).
1052 These newly created fasta files were then used as a reference sequence for phasing the data
1053 following the seqcap_pop pipeline (Harvey et al. 2016). Briefly we followed this pipeline and
1054 used GATK (McKenna et al. 2010) to mark duplicates, call, realign, and annotate/mask indels,
1055 call and annotate SNPs via GATK, restrict SNP calling to high quality SNPs, and then conducted
1056 read backed phasing. Finally, we created a sample specific fasta file of all phased loci for every
1057 sample using 'add_phased_snps_to_seqs_filter.py' available via the seqcap_pop pipeline
1058 (Harvey et al. 2016). All samples were combined into locus specific fasta files using a perl script
1059 developed in (Myers et al. 2019). Locus specific fasta files were aligned using muscle as
1060 implemented in the R package ape (Paradis and Schliep 2019) after removing individuals from
1061 alignments that fell below the 25% quantile of the distribution of sequence lengths for each
1062 locus. All fasta files were conservatively trimmed of missing data resulting in alignments with no
1063 missing sites (i.e., trimming to the shortest sequence in the alignment). We filtered these
1064 alignments by removing all loci that were missing >50% sequenced samples and removed all
1065 individuals that were missing from >30% of all alignments. Finally, we created a vcf file from
1066 these alignments and using vcftools (Danecek et al. 2011) and filtered this for minor allele

1067 frequency >10% retaining only 1 SNP/locus for population assignment analyses (Linck and
1068 Battey 2019).
1069
1070
1071 *Geographic groupings*
1072 We estimated population structure by comparing Discriminant Analysis of Principal
1073 Components (DAPC; Jombart et al., 2010) and spatial Principal Component analysis (sPCA;
1074 Jombart et al., 2008) in adegenet v2.1.2(Jombart 2008), sparse nonnegative matrix
1075 factorization (SNMF; Frichot et al., 2014) in LEA v1.4.0 (Frichot and François 2015), and
1076 estimated effective migration surfaces (EEMS v; Petkova et al., 2016), which accounts for IBD .
1077 Each of these methods use different assumptions for grouping individuals and we compare
1078 congruence among them.The model-free method DAPC sequentially estimates K-means for
1079 clustering and selects the best fit to infer genetic clusters in the absence of prior group
1080 identification. Using the package adegenet (Jombart 2008) in R (R Core Team 2015) we first
1081 transformed data using PCA by choosing a large number of PC axes (n=200), then picking the
1082 number of discriminant functions yielding large F-statistics (>2000) and optimal number of
1083 groups using BIC (Bayesian inference criterion). Because using a large number of the PCs may
1084 yield arbitrary solutions, we estimated the optimal *a-score*, which measures this bias by
1085 calculating the difference between the actual cluster assignment probabilities and randomly
1086 assigned probabilities (predicting 15-23 PCs). We also used cross-validation with a 90% training
1087 and 10% test dataset to estimate the average predicted success for each group.
1088 Because DAPC does not account for spatial structure, such as IBD, we also used sPCA.
1089 Similar to DAPC, this method does not require data to be in Hardy-Weinberg or linkage

1090 equilibrium. sPCA also adds an element of space by isolating global structure, representing
1091 disconnected groups or clines, from local structure, representing repulsion (selection against
1092 similar genetic types co-occurring), and random noise. Using adegenet, we estimated sPCA
1093 given SNPs and locations of each sample using up to 15 positive eigenvalues (global structures)
1094 and 15 negative eigenvalues (local structures). To connect individual locations to build a
1095 connection network, we used Delaunay triangulation. We also implemented Monte-Carlo tests
1096 with 100 permutations using the function *spca_randtests* to determine if significant local or
1097 global spatial structures exist.

1098 We also assessed ancestral coefficients using SNMF in the package LEA (Frichot and
1099 François 2015). This methodology is comparable to ADMIXTURE (Alexander et al. 2009) and
1100 Structure (Pritchard et al. 2000) methods, but runs 10-30x faster using genomic-scale data. This
1101 method produces a least-squares estimate of ancestry populations given *K* ancestral populations
1102 We estimated ancestry coefficients over 1-6 *K* using 100 repetitions per *K*, 100 iterations per
1103 algorithm, and a tolerance of 0.05. To determine the appropriate number of groups, we used a
1104 cross-validation technique with the entropy criterion, which masks genotypes to fit a particular
1105 model to each *K*.

1106 Finally, to compare these methods of grouping individuals into populations, we estimated
1107 migration rates over the landscape using EEMS (Petkova et al. 2016). This method models
1108 effective migration rates over geography to represent regions where migration is low in cases
1109 when genetic dissimilarity increases rapidly, thus providing a uniquely distinct view of the
1110 location of population clusters relative to biogeographic barriers as compared to DAPC, sPCA,
1111 and SNMF. We ran EEMS 3 times for $3x10^6$ generations with burnin set at $1x10^6$ generations,

1112 thinned by 9,999 generations, with 1,000 demes covering the known range of the *P. obsoletus*
1113 complex (Burbrink 2001).
1114 Both EEMS and DAPC find population structure, generally matched the geographic
1115 ranges of *P. alleghaniensis*, *P. spiloides, P. obsoletus*, and *P. bairdi* (Fig. 1 and Fig. S1;
1116 Burbrink, 2001). Total DAPC assignment probabilities were 0.975 (*P. bairdi* = 1.0, *P. obsoletus*
1117 = 0.96; *P. spiloides* = 1.0, and *P. alleghaniensis* = 0.97). Comparing these assignments to those
1118 from SMNF, we found that only *P. alleghaniensis* was misclassified as *P. spiloides* 8.5% of the
1119 time. Similarly, SNMF cross validation predicted four groups (cv = 0.599) as compared to two
1120 (cv=0.627), three (cv=0.607), and five (cv=0.601). These four groups generally match the same
1121 ranges as those found in DAPC. sPCA showed the presence of four groups, with PC1 showing
1122 the separation of groups at the MR and the two lineages east of this river, whereas PC2-5 shows
1123 divergence in all taxa. Importantly, global spatial structures were significant (observed spatial
1124 structure test statistic = 75.47, random expectation = 28.14; $P < 0.009$). Similarly, all three
1125 EEMS runs suggest low migration at the MR and the area separating the *P. bairdi* and *P.*
1126 *obsoletus* in west Texas. East of the MR, estimated low migration occurred near the Appalachian
1127 Mountains, though this area is a complex mixture of isolation and gene flow (Fig. 1). All three
1128 EEMS generated similar acceptance proportions for all proposal types (12-51%).
1129
1130 *Running SNAPP*
1131 We set forward and reverse mutation rates to 1. We applied a gamma distribution ($\alpha = 2$,
1132 $\beta = 200$) for the speciation rate (λ) prior and the *snapprior* was set to $\alpha = 1$, $\beta = 250$, and $\kappa = 1$. To
1133 test the four alternative species delimitation models including four taxa, three taxa (collapsing *P.*
1134 *bairdi/P. obsoletus* or *P. alleghaniensis/P. spiloides*), and two taxa (collapsing *P. bairdi/P.*

1135 *obsoletus* and *P.alleghaniensis/P. spiloides* we used a stepping-stone analysis with the
1136 *PathSampleAnalyser* with 72 steps and a chain length of 200000, burnin percentage at 50%, and
1137 a preBurnin of 10,000. We checked for stationarity using Tracer v1.7.1 (Drummond and
1138 Rambaut 2007) determining that estimated sample size (ESS) were > 200 for all parameters.
1139 *Running PHRAPL*
1140 PHRAPL uses gene-tree distances between simulated and empirical data to rank models using
1141 approximate likelihoods and AICs. To generate the empirical observations we estimated
1142 unrooted gene trees for each loci using IQ-Tree 1.6.12 (Nguyen et al. 2015) and testing all
1143 substitution models per locus prior to estimation. We rooted the trees using midpoint rooting and
1144 performed 100 subsamples of two individuals per population. To reduce model space, we
1145 allowed only one population size parameter (all populations have equal size) and one migration
1146 parameter (all migration rates are equal). We then ran 1000 simulations per model, searching
1147 over a grid of four CollapseStarts values and five MigrationStarts values.
1148
1149 *Spatial genetics using RDA*
1150 We predicted genetic distances from environmental, elevation, river, ecoregion, and
1151 distance variables using RDA in the R package Vegan (Dixon 2003), which avoided problems
1152 with spatial data using Mantel tests (Legendre et al. 2015). This method was used to compare
1153 results from the ANN analyses. RDA is an asymmetric canonical analysis here used to determine
1154 the significance of predictor axes (Legendre et al. 2011) at generating predict genetic structure
1155 over a landscape (McGaughran et al. 2014; Noguerales et al. 2016). We converted spatial
1156 distances using principal coordinates of neighbor matrices (PCNM) in the r package vegan
1157 (Dixon 2003) keeping all axes with positive eigenvectors. We ran this using the *capscale*

1158 function with only spatial variables alone to determine which significantly predict genetic
1159 distance to be passed to out ANN analyses. We also ran this with the significant spatial axes
1160 partialed out and environmental, elevational, ecoregion, and river variables. All variable
1161 distances were used to predict genetic distance and significance of each variable was tested using
1162 anova-like permutation tests to examine the joint effect of predictor variables given partial
1163 effects from spatial distance to account for IBD.

1164 Estimates of accuracy using ANN were high (>90%) for all comparisons and yielded
1165 similar conclusions to RDA in that environmental and geographic barriers and not just spatial
1166 variables predicted genetic structure. Here, RDA demonstrated that most of the structure for all
1167 taxa can be predicted by the Mississippi River (P=0.001), ecoregion (P =0.008), and current
1168 climate (P=0.019). Distance also played a role in structuring these data, though they were
1169 partialed out in RDA. For those lineages east of the Mississippi River, population structure was
1170 predicted by ecoregion (P=0.014), Pleistocene Marine Isotope Stage 19 (787 Kya; P=0.001), last
1171 interglacial (P=0.007) and ecoregion (P=0014). For those lineages west of the Mississippi
1172 River, we demonstrated predicted by ecoregion (P= 0.001).

1173

1174 *Genomics, the ICZN, and Taxonomy*

1175 One problem likely to be encountered by other researchers using genomic data is that
1176 previously named type specimens at particular type localities may be from regions that show
1177 some admixture. For instance, the type locality of the eastern-most species, *P. alleghaniensis* is
1178 "the summit of the Blue Ridge in Virginia and the Highlands of the Hudson" (Holbrook 1836;
1179 Schultz 1996). Both areas are likely composed of admixed individuals and the International Code
1180 of Zoological Nomenclature (ICZN) forbids naming species based on hybrids (Article

1.3.3;ICZN, 1999) though the intention of this ICZN article did not consider proportions of admixture. One possible solution would be to use the next available name from Florida, where no admixture is present (eliminating *P. quadrivittata*), which would be *P.deckerti* (type locality: Lower Matecumber Key, FL; Brady 1932). Another solution would be to designate a Neotype from a locality showing no admixture and retain *P. alleghaniensis.* If *P. alleghaniensis/P. spiloides* were not considered unique, then *P. alleghaniensis* has priority by 18 years (Holbrook 1836; Duméril et al. 1854).

The type locality for the remaining taxa, *P. spiloides (*New Orleans, Louisiana; Duméril et al. 1854), *P. obsoletus* (On the Missouri River from the Vicinity Isle au Vache to Council Bluff; Say, 1823), and *P. bairdi* (Fort Davis, Apache Mountains, Jeff Davis Co.: Texas; Yarrow 1880) are all from regions that likely represent individuals not admixed (Fig.1 & 5, Burbrink, 2001). Species delimitation methods, estimates of fixation for loci, and observation that these four organisms have remained geographically distinct in the face of gene flow throughout the Pleistocene suggests that these taxa should be considered unique species though hybrid zones exist. We acknowledge difficulties recognizing the eastern lineages as distinct and could argue for recognizing them as a single taxon, *P. alleghaniensis*.

Literature Cited

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–64. Cold Spring Harbor Laboratory Press.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics.

Brady, M. K. 1932. A new snake from Florida. Proc. Biol. Soc. Washingt. 5–8.

Burbrink, F. T. 2001. Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). Herpetol. Monogr. 1–53.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. Narnia.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. hzar : hybrid zone analysis using an R software package. Mol. Ecol. Resour. 14:652–663.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14:927–930. Blackwell Publishing.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. Bmc Evol. Biol. 7.

Duméril, A. M. C., G. Bibron, and A. H. A. Duméril. 1854. Erpétologie Genérale ou Histoire Naturelle Complète des Reptiles Vol. 7. Paris: Librarie Enclyclopedique de Roret., Paris.

Faircloth, B. C. 2013. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788. Narnia.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Syst. Biol. 61:717–726. Narnia.

Frichot, E., and O. François. 2015. LEA : An R package for landscape and ecological association studies. Methods Ecol. Evol. 6:925–929. John Wiley & Sons, Ltd (10.1111).

1227 Frichot, E., F. Mathieu, T. Trouillon, G. Bouchard, and O. François. 2014. Fast and Efficient
1228 Estimation of Individual Ancestry Coefficients. Genetics 196:973–983.
1229 Harvey, M. G., B. T. Smith, T. C. Glenn, B. C. Faircloth, and R. T. Brumfield. 2016. Sequence
1230 Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. Syst.
1231 Biol. 65:910–924. Narnia.
1232 Holbrook, J. E. 1836. North American herpetology : or, A description of the reptiles inhabiting
1233 the United States / by John Edwards Holbrook. J. Dobson, Philadelphia :
1234 ICZN: 1999. International Code of Zoological Nomenclature. Fourth. London, England.
1235 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
1236 Bioinformatics 24:1403–1405.
1237 Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a
1238 new method for the analysis of genetically structured populations. BMC Genet. 11:94.
1239 BioMed Central.
1240 Jombart, T., S. Devillard, A. B. Dufour, and D. Pontier. 2008. Revealing cryptic spatial patterns
1241 in genetic variability by a new multivariate method. Heredity (Edinb). 101:92–103.
1242 Legendre, P., M.-J. Fortin, and D. Borcard. 2015. Should the Mantel test be used in spatial
1243 analysis? Methods Ecol. Evol. 6:1239–1247. John Wiley & Sons, Ltd (10.1111).
1244 Legendre, P., J. Oksanen, and C. J. F. ter Braak. 2011. Testing the significance of canonical axes
1245 in redundancy analysis. Methods Ecol. Evol. 2:269–277. John Wiley & Sons, Ltd
1246 (10.1111).
1247 Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler
1248 transform. Bioinformatics 25:1754–1760. Narnia.
1249 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R.

1250 Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics
1251 25:2078–2079. Narnia.
1252 Linck, E., and C. J. Battey. 2019. Minor allele frequency thresholds strongly affect population
1253 structure inference with genomic data sets. Mol. Ecol. Resour. 19:639–647. John Wiley &
1254 Sons, Ltd (10.1111).
1255 McGaughran, A., K. Morgan, and R. J. Sommer. 2014. Environmental variables explain genetic
1256 structure in a beetle-associated nematode. PLoS One 9:e87317. Public Library of Science.
1257 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D.
1258 Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a
1259 MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.
1260 20:1297–303. Cold Spring Harbor Laboratory Press.
1261 Myers, E. A., R. W. Bryson, R. W. Hansen, M. L. Aardema, D. Lazcano, and F. T. Burbrink.
1262 2019. Exploring Chihuahuan Desert diversification in the gray-banded kingsnake,
1263 *Lampropeltis alterna* (Serpentes: Colubridae). Mol. Phylogenet. Evol. 131:211–218.
1264 Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and
1265 effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. Mol. Biol.
1266 Evol. 32:268–274. Oxford University Press.
1267 Noguerales, V., P. J. Cordero, and J. Ortego. 2016. Hierarchical genetic structure shaped by
1268 topography in a narrow-endemic montane grasshopper. BMC Evol. Biol. 16:96. BioMed
1269 Central.
1270 Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and
1271 evolutionary analyses in R. Bioinformatics 35:526–528. Narnia.
1272 Petkova, D., J. Novembre, and M. Stephens. 2016. Visualizing spatial population structure with

estimated effective migration surfaces. Nat. Genet. 48:94–100. Nature Publishing Group.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News 6:7–11.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing., Vienna, Austria.

Say, T. 1823. Coluber obsoletus. P. 140 *in* Account of an expedition from Pittsburgh to the Rocky Mountains, Performed in the Years 1819 and ’20. Logman, Hurst, Rees, Ovme, and Brown, London.

Schultz, K. D. 1996. A monograph of the colubrid snakes of the genus *Elaphe* Fitzinger. Koelz Scientific Books, Havlicuv.

Sun, K., K. A. Meiklejohn, B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball. 2014. The evolution of peafowl and other taxa with ocelli (eyespots): a phylogenomic approach. Proc. R. Soc. B Biol. Sci. 281:20140823. The Royal Society.

Yarrow, H. C. 1880. Coluber bairdi sp. nov. P. *in* E. D. Cope, ed. On the zoological position of Texas Bull. US Natl. Mus.

Fig.S1 – Delimiting populations using sparse nonnegative matrix factorization (SNMF) and spatial Principal Component analysis (sPCA) showing the location of four geographically distinct lineages.

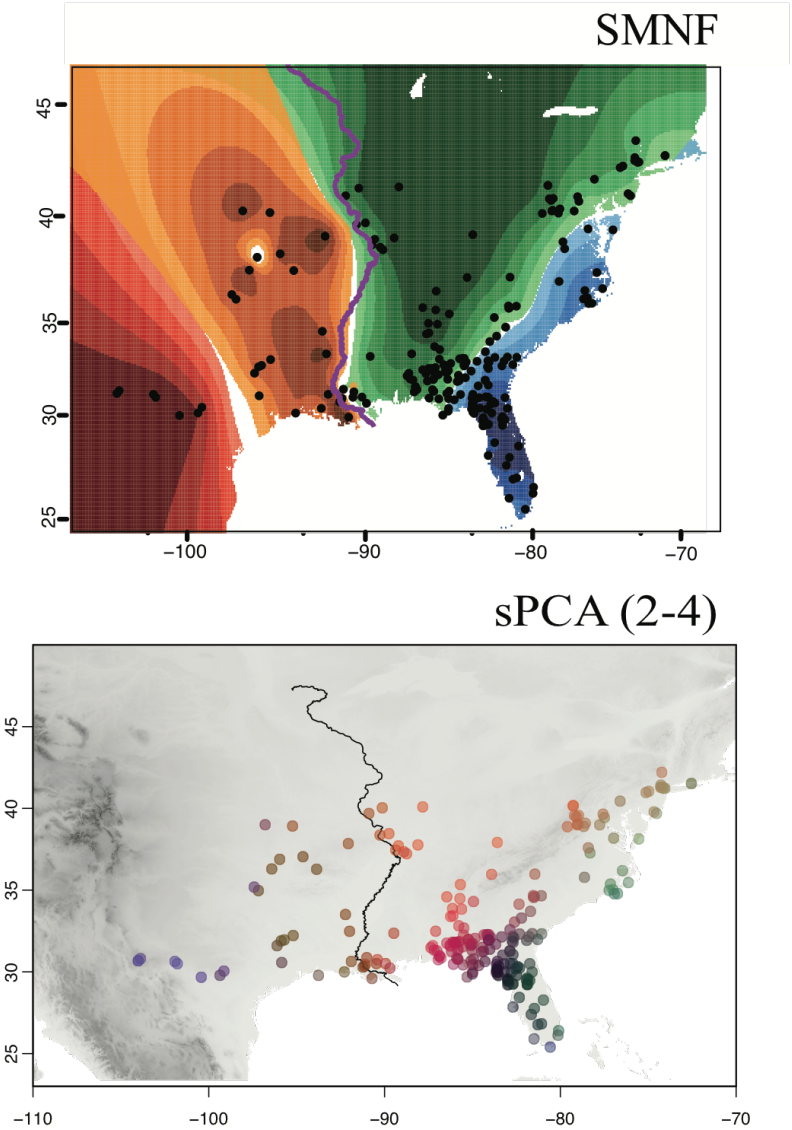


Fig. S2 – PipeMaster model fit of observed data to simulations under IM, IMD, and IMD-LGM models for PC1-PC10.

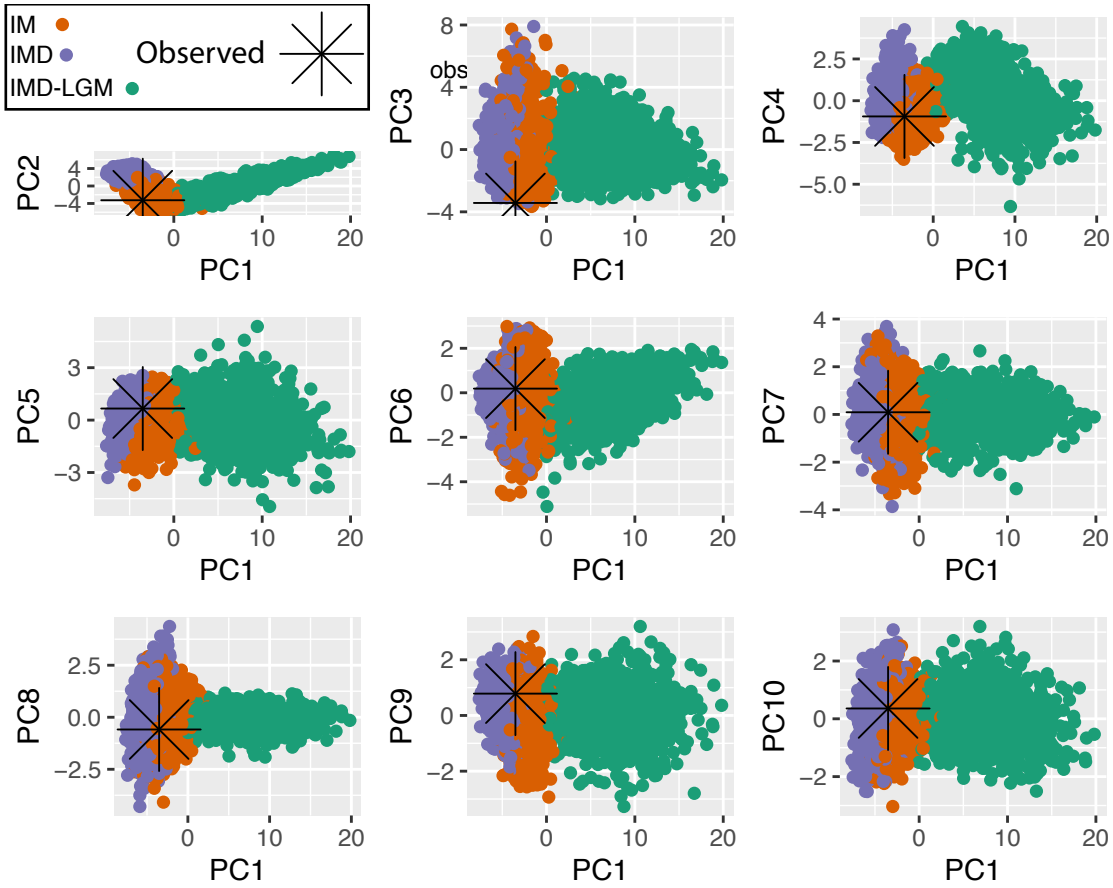
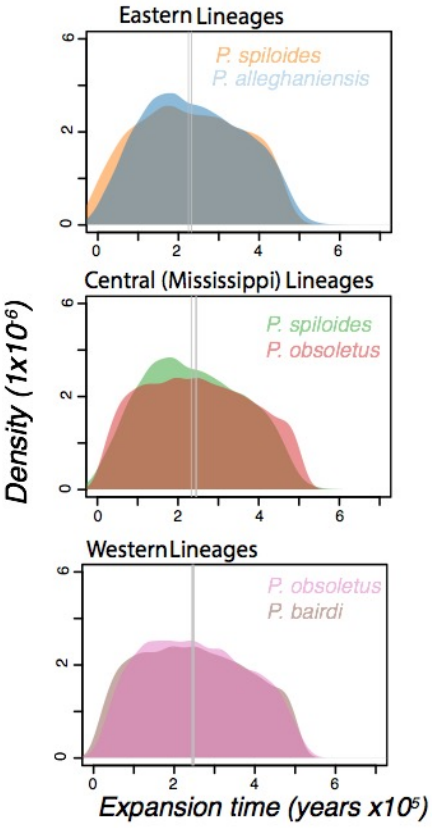


Fig. S3 – Expansion times estimated from PipeMaster for all adjacent taxon pairs.



1327 Fig. S4– Results from PipeMaster after removing hyrbids showing estimates and directionality of

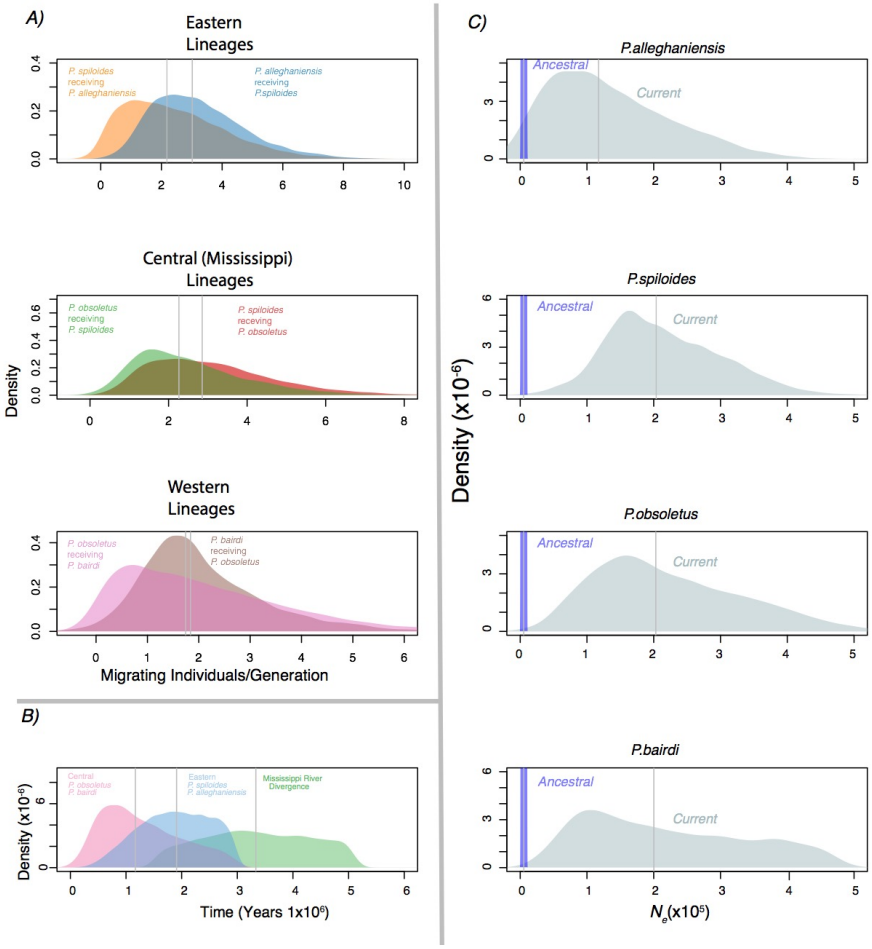
1328 gene flow (A), divergence times between groups and taxon pairs and the Mississippi River

1329 (*P.alleghaniensis/ P.spiloides* vs *P.obsoletus/P.bairdi*) (B), and changes in population sizes over

1330 time (C).

1331

1332



1333

1334

1335

1336

Fig. S5 – Prediction of cline width under neutrality (black line through time) given divergence times for each species pair: A) *Pantherophis alleghaniensis/P. spiloides*, B) *P. spiloides/P.obsoletus*, and C) *P. obsoletus/P.bairdi.* Colored vertical lines represent times of divergence between species pairs and horizontal colored lines indicate actual cline widths.

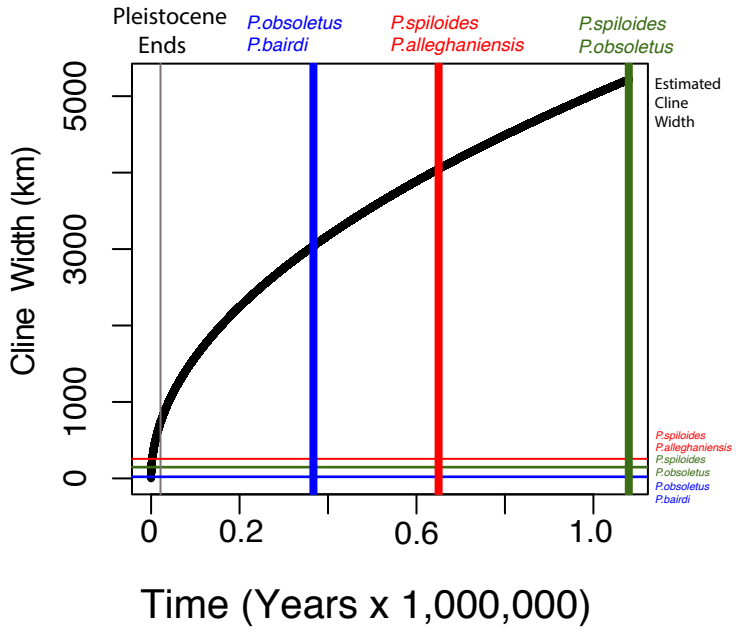


Table S1. Priors and summary stats for PipeMaster for the IM, IMD, IMD-LGM models; estimating migration, timing of divergence, and historical demography.

IM

PARAMETER	PRIOR.1	PRIOR.2	DISTRIBUTION	
1	Ne0.pop1	20000	500000	runif
2	Ne0.pop2	20000	5e+05	runif
3	Ne0.pop3	20000	5e+05	runif
4	Ne0.pop4	20000	5e+05	runif
5	mig0.1_2	2	5	runif
6	mig0.1_3	0	0	runif
7	mig0.1_4	0	0	runif
8	mig0.2_1	2	5	runif
9	mig0.2_3	2	5	runif
10	mig0.2_4	0	0	runif
11	mig0.3_1	0	0	runif
12	mig0.3_2	2	5	runif
13	mig0.3_4	2	5	runif
14	mig0.4_1	0	0	runif
15	mig0.4_2	0	0	runif
16	mig0.4_3	2	5	runif
17	join1_2	30000	1000000	runif
18	join3_4	30000	1000000	runif
19	join2_4	500000	1700000	runif

IMD

PARAMETER	PRIOR.1	PRIOR.2	DISTRIBUTION	
1	Ne0.pop1	10000	500000	runif
2	Ne0.pop2	10000	500000	runif
3	Ne0.pop3	10000	500000	runif
4	Ne0.pop4	10000	500000	runif
5	mig0.1_2	2	5	rtnorm
6	mig0.1_3	0	0	rtnorm
7	mig0.1_4	0	0	rtnorm
8	mig0.2_1	2	5	rtnorm
9	mig0.2_3	2	5	rtnorm
10	mig0.2_4	0	0	rtnorm
11	mig0.3_1	0	0	rtnorm
12	mig0.3_2	2	5	rtnorm
13	mig0.3_4	2	5	rtnorm
14	mig0.4_1	0	0	rtnorm
15	mig0.4_2	0	0	rtnorm
16	mig0.4_3	2	5	rtnorm
17	Ne1.pop1	1000	10000	runif
18	Ne1.pop2	1000	10000	runif
19	Ne1.pop3	1000	10000	runif

20	Ne1.pop4	1000	10000	runif
21	t.Ne1.pop1	3000	170000	runif
22	t.Ne1.pop2	3000	170000	runif
23	t.Ne1.pop3	3000	170000	runif
24	t.Ne1.pop4	3000	170000	runif
25	join1_2	30000	1000000	runif
26	join3_4	30000	1000000	runif
27	join2_4	500000	1700000	runif

IMD-LGM

PARAMETER	**PRIOR.1**	**PRIOR.2**	**DISTRIBUTION**	
1	Ne0.pop1	10000	500000	runif
2	Ne0.pop2	10000	500000	runif
3	Ne0.pop3	10000	500000	runif
4	Ne0.pop4	10000	500000	runif
5	mig0.1_2	2	5	rtnorm
6	mig0.1_3	0	0	rtnorm
7	mig0.1_4	0	0	rtnorm
8	mig0.2_1	2	5	rtnorm
9	mig0.2_3	2	5	rtnorm
10	mig0.2_4	0	0	rtnorm
11	mig0.3_1	0	0	rtnorm
12	mig0.3_2	2	5	rtnorm
13	mig0.3_4	2	5	rtnorm
14	mig0.4_1	0	0	rtnorm
15	mig0.4_2	0	0	rtnorm
16	mig0.4_3	2	5	rtnorm
17	Ne1.pop1	1000	10000	runif
18	Ne1.pop2	1000	10000	runif
19	Ne1.pop3	1000	10000	runif
20	Ne1.pop4	1000	10000	runif
21	t.Ne1.pop1	3000	170000	runif
22	t.Ne1.pop2	3000	170000	runif
23	t.Ne1.pop3	3000	170000	runif
24	t.Ne1.pop4	3000	170000	runif
25	mig1.1_2	0	0	runif
26	mig1.1_3	0	0	runif
27	mig1.1_4	0	0	runif
28	mig1.2_1	0	0	runif
29	mig1.2_3	0	0	runif
30	mig1.2_4	0	0	runif
31	mig1.3_1	0	0	runif
32	mig1.3_2	0	0	runif
33	mig1.3_4	0	0	runif
34	mig1.4_1	0	0	runif
35	mig1.4_2	0	0	runif

36	mig1.4_3	0	0	runif
37	t.mig1.1_2	0	8700	runif
38	t.mig1.1_3	0	8700	runif
39	t.mig1.1_4	0	8700	runif
40	t.mig1.2_1	0	8700	runif
41	t.mig1.2_3	0	8700	runif
42	t.mig1.2_4	0	8700	runif
43	t.mig1.3_1	0	8700	runif
44	t.mig1.3_2	0	8700	runif
45	t.mig1.3_4	0	8700	runif
46	t.mig1.4_1	0	8700	runif
47	t.mig1.4_2	0	8700	runif
48	t.mig1.4_3	0	8700	runif
49	join1_2	30000	1000000	runif
50	join3_4	30000	1000000	runif
51	join2_4	500000	1700000	runif