Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation

DREW R. SCHIELD[1†], BLAIR W. PERRY[1†], RICHARD H. ADAMS[1], DAREN C. CARD[1,2,3], TEREZA JEZKOVA[4], GIULIA I. M. PASQUESI[1], ZACHARY L. NIKOLAKIS[1], KRISTOPHER ROW[1], JESSE M. MEIK[5], CARA F. SMITH[6], STEPHEN P. MACKESSY[6] and TODD A. CASTOE[1*]

[1]*Department of Biology, 501 South Nedderman Drive, University of Texas at Arlington, Arlington, TX 76019, USA*

[2]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 012138, USA*

[3]*Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA*

[4]*Department of Biology, Miami University of Ohio, 700 East High Street, Oxford, OH 45056, USA*

[5]*Department of Biological Sciences, Tarleton State University, 1333 West Washington Street, Stephenville, TX 76402, USA*

[6]*School of Biological Sciences, 501 20th Street, University of Northern Colorado, Greeley, CO 80639, USA*

The study of recently diverged lineages whose geographical ranges come into contact can provide insight into the early stages of speciation and the potential roles of reproductive isolation in generating and maintaining species. Such insight can also be important for understanding the strategies and challenges for delimiting species within recently diverged species complexes. Here, we use mitochondrial and nuclear genetic data to study population structure, gene flow and demographic history across a geographically widespread rattlesnake clade, the western rattlesnake species complex (*Crotalus cerberus*, *Crotalus viridis*, *Crotalus oreganus* and relatives), which contains multiple lineages with ranges that overlap geographically or contact one another. We find evidence that the evolutionary history of this group does not conform to a bifurcating tree model and that pervasive gene flow has broadly influenced patterns of present-day genetic diversity. Our results suggest that lineage diversity has been shaped largely by drift and divergent selection in isolation, followed by secondary contact, in which reproductive isolating mechanisms appear weak and insufficient to prevent introgression, even between anciently diverged lineages. The complexity of divergence and secondary contact with gene flow among lineages also provides new context for why delimiting species within this complex has been difficult and contentious historically.

ADDITIONAL KEYWORDS: *Crotalus cerberus* – *Crotalus oreganus* – *Crotalus scutulatus* – *Crotalus viridis* – demographic inference – incipient speciation – introgression – population genomics – population structure – RADseq.

INTRODUCTION

Mechanisms that promote reproductive isolation between lineages will accumulate as lineages diverge (Coyne & Orr, 2004). One common assumption is that such mechanisms accumulate proportionally with divergence time (Wu, 2001; Wu & Ting, 2004; Barton & De Cara, 2009) and may reinforce lineage diversification if they drive pre- or postzygotic reproductive isolation between diverging lineages that come into secondary contact (Butlin, 1995; Orr & Presgraves, 2000; Orr & Turelli, 2001; Butlin & Smadja, 2018). Studies examining the evolutionary consequences of hybridization and introgression between historically isolated lineages have been instrumental in our understanding of the early stages

*Corresponding author. Email: todd.castoe@uta.edu
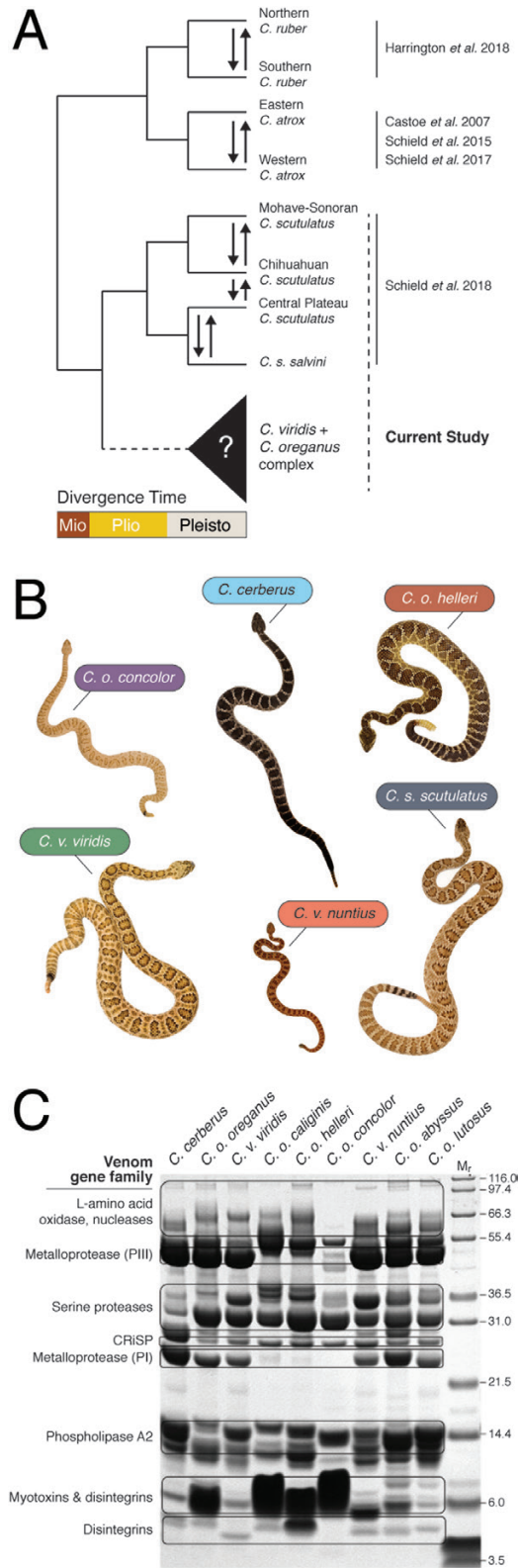†These authors contributed equally to this study.

of speciation (Orr, 1995; Coyne & Orr, 2004; Mallet *et al.*, 2007; Abbott *et al.*, 2013; Payseur & Rieseberg, 2016). Despite the wealth of studies, the relative roles of allopatric divergence and reproductive isolation mechanisms, and specifically, reinforcement through postzygotic isolation mechanisms after secondary contact (Butlin & Smadja, 2018), are still poorly understood and debated (Paterson, 1978; Butlin, 1995; Orr & Turelli, 2001; Coyne & Orr, 2004; Hoskin *et al.*, 2005; Matute, 2010). The prevalence of speciation modes and reproductive isolation mechanisms across diverse animal lineages also remains an open question, because most previous studies have focused on a narrow set of organisms (e.g. mammals, birds, fishes, flies; reviewed by Payseur & Rieseberg, 2016).

Recent studies of multiple North American rattlesnake species, including *Crotalus atrox* (Castoe *et al.*, 2007; Schield *et al.*, 2015, 2017), *Crotalus ruber* (Harrington *et al.*, 2018) and *Crotalus scutulatus* (Schield *et al.*, 2018), suggest that allopatric divergence followed by gene flow in secondary contact might be a common mode of speciation among temperate rattlesnake lineages (Fig. 1A). These studies suggest that isolation occurred before or during the Pleistocene, followed by Late Pleistocene or Holocene range expansions, leading to secondary contact with gene flow, and they raise the question of whether weak (or non-existent) reproductive isolating mechanisms might be a common feature across rattlesnake species. However, given that each of these studies was limited to a small number of comparisons and lineages, it remains unknown whether these features broadly characterize patterns of speciation and evolution in rattlesnakes.

Here, we interrogate patterns of diversification in a widespread and diverse clade of rattlesnakes by testing for evidence of allopatric divergence and secondary contact. We evaluate whether barriers to gene flow have evolved between lineages with sympatric or adjacent modern geographical distributions. We combine mitochondrial gene sequences with nuclear reduced-representation genomic data generated using restriction site-associated DNA sequencing (RADseq) to characterize lineage diversity, population genetic

bar depicts the relative times since lineages split from common ancestors (Mio, Miocene; Pleisto, Pleistocene; Plio, Pliocene). Citations to previous studies focused on each group are provided to the right of the tree. B, representative colour and body size phenotypes of a subset of western rattlesnake species complex (CVOS) lineages, with sizes to scale. C, comparison of phenotypes of venom composition between CVOS lineages, redrawn from Mackessy (2010). Abbreviations: CRiSP, cysteine-rich secretory proteins; M_r, Novex Mark 12 molecular weight standards.

Figure 1. A, phylogeny of rattlesnake groups examined previously and in the present study for evidence of gene flow in secondary contact. The divergence time scale

structure, patterns of gene flow and demographic history in the western rattlesnake species complex and its sister group, the Mojave rattlesnake. This clade of rattlesnake lineages includes four recognized species: the Arizona black rattlesnake (*Crotalus cerberus*), the prairie rattlesnake (*Crotalus viridis*), the western rattlesnake (*Crotalus oreganus*) and the Mojave rattlesnake (*Crotalus scutulatus*); for brevity, we refer to this clade as the CVOS hereafter. Lineages within the CVOS exhibit substantial phenotypic variation in coloration (Klauber, 1956), body size (Ashton, 2001; Fig. 1B) and venom composition (Mackessy, 2010; Strickland *et al.*, 2018; Fig. 1C), which has led to the description of numerous subspecies within *C. viridis*, *C. oreganus* and *C. scutulatus*: the prairie rattlesnake (*C. viridis viridis*), Hopi rattlesnake (*C. viridis nuntius*), northern Pacific rattlesnake (*C. oreganus oreganus*), southern Pacific rattlesnake (*C. oreganus helleri*), Coronado Island rattlesnake (*C. oreganus caliginis*), Great Basin rattlesnake (*C. oreganus lutosus*), Grand Canyon rattlesnake (*C. oreganus abyssus*), midget-faded rattlesnake (*C. oreganus concolor*), Mojave rattlesnake (*C. scutulatus scutulatus*) and the Huamantlán rattlesnake (*C. scutulatus salvini*). This phenotypically diverse group of snakes collectively occupies an extensive range in western North America encompassing nearly all of the western USA, regions of southern Canada and much of northern and central Mexico (Campbell & Lamar, 2004; Crother *et al.*, 2012; Fig. 2A).

The systematics of the CVOS has been notoriously problematic and contentious, with persistent disagreement regarding species limits; various authors have proposed recognizing between two and 11 species within this group (e.g. Klauber, 1956; Pook *et al.*, 2000; Ashton & de Queiroz, 2001; Douglas *et al.*, 2002; Campbell & Lamar, 2004; Davis, 2016; Davis *et al.*, 2016). Estimates of the evolutionary relationships among CVOS lineages have also differed among studies, and previous molecular analyses have included only mitochondrial DNA sequence data (e.g. Pook *et al.*, 2000; Ashton & de Queiroz, 2001; Douglas *et al.*, 2002). Here, we leverage this system and a large set of nuclear loci to address unresolved questions about the phylogeny, population structure, patterns of diversification and the roles of allopatric divergence and gene flow.

Is there genetic evidence for gene flow between CVOS lineages?

Are patterns of genetic diversity consistent with allopatric divergence followed by introgression in secondary contact?

Is there genetic evidence for reproductive isolation between historically isolated lineages?

Do climatic niche models corroborate inferences of allopatric divergence and secondary contact based on genomic data?

MATERIAL AND METHODS

DATA COLLECTION

We generated reduced-representation genomic sequencing libraries for samples from throughout the range of the CVOS ($N = 247$; Fig. 2A; Supporting Information, Table S1), including representatives of each recognized subspecies and individuals from four distinct *C. scutulatus* lineages identified by Schield *et al.* (2018; 'Mohave–Sonoran', 'Chihuahuan', 'Central Mexican Plateau' and '*C. s. salvini*'; Supporting Information, Fig. S1). For the purposes of this study, we refer to CVOS lineages according to subspecies designations and use the names above when referring to *C. scutulatus* lineages.

We extracted DNA from tissue that was either snap frozen in ethanol or in RNAlater and stored at –80 °C, using a standard phenol–chloroform–isoamyl extraction. We then constructed sequencing libraries using a double restriction digest reduced-representation sequencing technique (ddRADseq) described by Peterson *et al.* (2012), with modifications described by Schield *et al.* (2015, 2018). Briefly, we digested genomic DNA using *Sbf*I and *Sau*3aI overnight at 37 °C, then purified and concentrated the digested DNA using Serapure beads. Samples were normalized by concentration and ligated barcoded, double-stranded Illumina DNA adapters, which also included an 8 bp unique molecular identifier for polymerase chain reaction (PCR) clone filtering post-sequencing. We then pooled samples with ligated barcodes into groups of six to eight, and purified pooled libraries with Serapure beads. Size selection was performed using a Blue Pippin Prep (Sage Science) for fragments 575–655 bp in length, based on predictions that this range should target ~20 000 loci based on in silico digests of the prairie rattlesnake (CroVir2.0) genome (NCBI BioSample SAMN07738522). We PCR amplified size-selected libraries using library-specific indexed primers, quantified amplified libraries, and combined libraries in equimolar ratios into final pools containing a maximum of 96 samples each. These were then sequenced using 100 bp single-end reads on multiple Illumina HiSeq lanes.

We used the Stacks v.1.42 (Catchen *et al.*, 2013) clone_filter module to filter PCR clones from raw sequence data, then trimmed 8 bp of adapter sequence from all reads using Fastx-Toolkit (Gordon, 2010). We then demultiplexed trimmed sequences into individual samples using the Stacks process_radtags module, which demultiplexes samples according to their 6 bp individual barcode and checks for intact restriction cut sites. Processed and demultiplexed data are available from the NCBI SRA (PRJNA548132). We mapped read data for all samples to the *Crotalus viridis* version 2 reference genome using BWA v.0.7.10 (Li & Durbin,
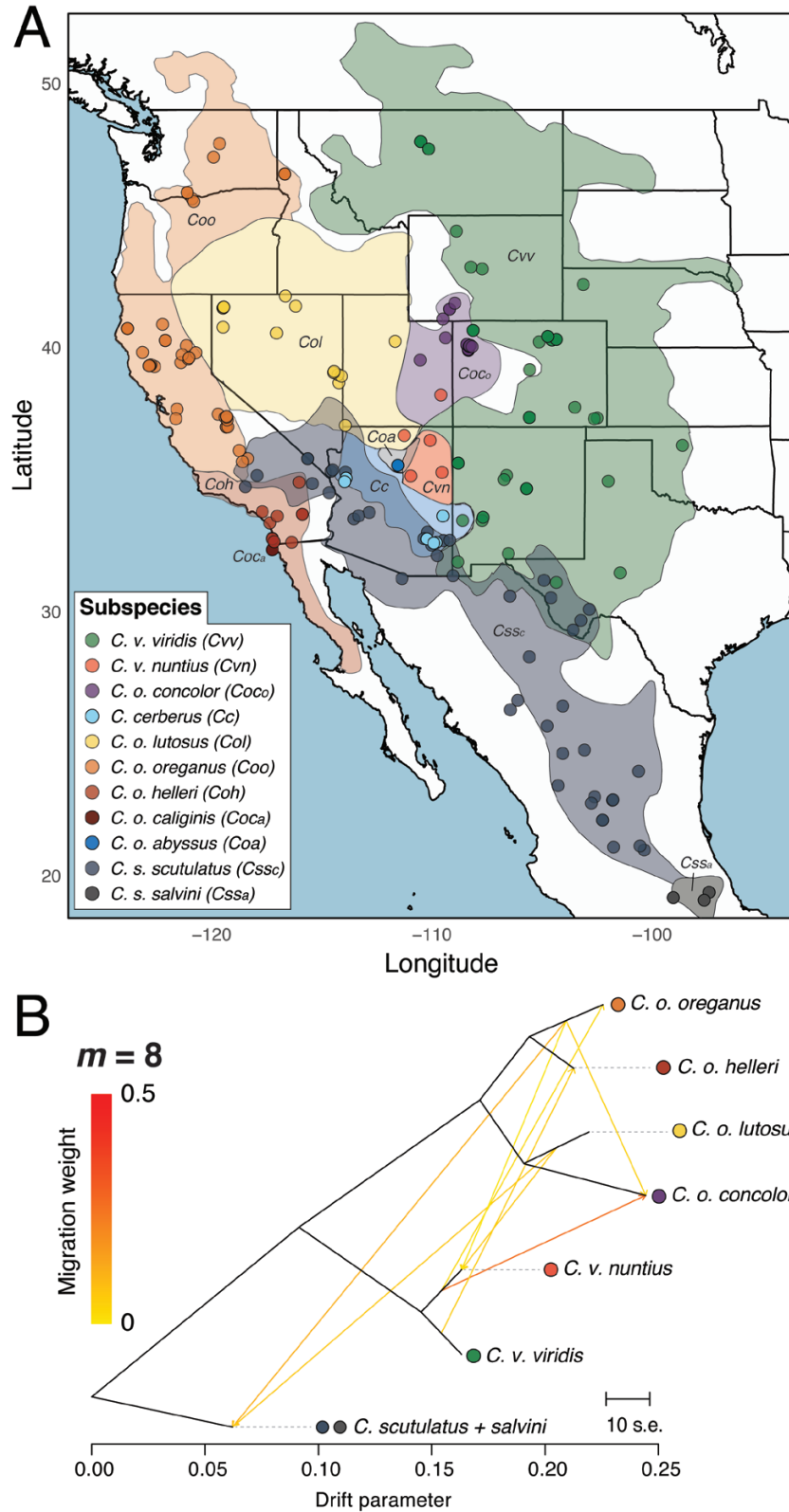
Figure 2. A, range map of the western rattlesnake species complex (CVOS) and sampling in the present study. Approximate distributions of recognized taxa are shown as differently shaded regions, which are also labelled with three-letter

2009) using default settings. We then sorted mappings, generated a 'pile-up' of all sample alignments, and called variants using a combination of SAMtools and BCFtools v.1.2 (Li *et al.*, 2009). We recoded all raw variant calls within individuals with a read depth fewer than five as missing data using the BCFtools filter function. We filtered variants to retain biallelic single nucleotide polymorphisms (SNPs) with a minor allele frequency > 0.05 that also had data from ≥ 20% of samples across the dataset using VCFtools (Danecek *et al.*, 2011). Our processing and filtering procedure for the nuclear RADseq dataset resulted in 22 459 SNPs for downstream analyses.

For phylogenetic and divergence time estimates, we sequenced a fragment of the mitochondrial gene *NADH4* (*ND4*) from a subset of samples in our nuclear dataset using the PCR primers ND4 and Leu (Arevalo *et al.*, 1994). Amplified PCR products were purified using Serapure beads and quantified using a Qubit flourometer (Life Technologies, Grand Island, NY, USA). Purified PCR products were sequenced using amplification primers and BigDye on the ABI 3730 capillary sequencing platform. We then evaluated sequence chromatograms and generated consensus sequences using Geneious v.8.0 (Biomatters Ltd, Auckland, New Zealand). Additional *ND4* sequences for outgroup taxa were obtained from GenBank (Supporting Information, Table S1).

MITOCHONDRIAL PHYLOGENY AND DIVERGENCE TIME ESTIMATION

We estimated phylogenetic relationships and divergence times from our mitochondrial dataset using BEAST2 v.2.4.7 (Bouckaert *et al.*, 2014). *ND4* sequences for the CVOS and outgroup taxa (Supporting Information, Table S1) were aligned using MUSCLE v.3.8 (Edgar, 2004), with minor manual edits. The final alignment included 810 nucleotides. We identified the best-fit partitioning scheme (three separate partitions representing each codon position) and set of substitution models using PartitionFinder v.2.1.1 (Lanfear *et al.*, 2012), which were GTR+Γ, TN93+Γ and TN93+invariant Γ for first, second and third codon positions, respectively. We calibrated the tree using several most recent common ancestor priors, including priors for the ancestor of *Agkistrodon* (6 Mya offset, with a mean of 0.01 and SD

of 0.42) and ancestor of *Sistrurus* (8 Mya offset, with a mean of 0.01 and SD of 0.76), which were specified using lognormal distributions and constraints according to Reyes-Velasco *et al.* (2013). We set the prior for the ancestral node for the rattlesnakes (*Crotalus* + *Sistrurus*) to a normal distribution with a mean of 0, SD of 3, and an offset of 11.2 Mya. We also set the prior for the common ancestor of *C. atrox* and *C. ruber* using a normal distribution, an offset of 3.2 Mya and SD of 1, following Castoe *et al.*, (2007). Finally, we set a lognormal prior for the ancestor of *C. viridis* + *C. cerberus* + *C. oreganus* to an offset of 5 Mya based on Late Miocene fossils (Holman, 2000), using a mean of 0.01 and SD of 0.6. Remaining priors were set using BEAUTi default values. We ran two independent runs for 5 × 10⁹ generations per run, then evaluated convergence, stationarity and effective parameter sample sizes using Tracer v.1.6 (Drummond & Rambaut, 2007). We discarded the initial 10% of samples as burn-in and generated a maximum clade credibility tree using LogCombiner and TreeAnnotator (Bouckaert *et al.*, 2014).

TESTING THE HYPOTHESIS OF A STRICTLY BIFURCATING TREE USING NUCLEAR DATA

We used TreeMix (Pickrell & Pritchard, 2012) and our nuclear dataset to determine whether the evolutionary history of the CVOS adheres to a bifurcating tree model or if instead migration between branches and tips of the tree has occurred during diversification. TreeMix generates a hypothesized population graph (i.e. topology) and allows the user to generate population graphs with successive numbers of migration events between lineages. We ran TreeMix on a subset of individuals that excluded groups with few individuals (*C. cerberus*, *C. o. abyssus* and *C. o. caliginis*) and samples with > 50% missing data and loci with > 5% missing data; these groups also did not correspond to distinct nuclear genetic clusters in downstream analyses (Fig. 3). TreeMix was run without sample size correction on blocks of 1000 SNPs to account for linkage disequilibrium. We rooted the population graph using our combined *C. scutulatus* sampling and generated graphs allowing for 0–18 migration events. We then compared results across models by calculating the proportion of variance in relatedness between groups explained by each migration model.

abbreviations (see key). Coloured circles within each region show localities from which the various taxa were sampled. A detailed map of the *Crotalus scutulatus* distribution, with previously identified phylogeographical breaks, is shown in the Supporting Information (Fig. S1). B, results of TreeMix analysis on a pruned subset of our nuclear RADseq dataset, showing the population graph with eight migration edges (*m*). The scale of relative weight (i.e. amount of migration) of migration edges is shown on the left, with darker colours representing greater migration. Coloured circles adjacent to taxon names correspond to the legend in panel A.

NUCLEAR ESTIMATES OF POPULATION GENETIC STRUCTURE AND DIVERSITY

We used ADMIXTURE v.1.3 (Alexander *et al.*, 2009) to estimate population structure and admixture within the CVOS using our nuclear dataset, which uses the likelihood model implemented in STRUCTURE (Pritchard *et al.*, 2000) to determine ancestry probabilities across individuals. Here, we excluded individuals with a large amount of missing data (i.e. < 1000 SNPs) to reduce the potential effects of missing genotypes on ancestry estimates, leaving a total of 200 individuals that we analysed across a range of values of K (K = 2–13). We used the default cross-validation method in ADMIXTURE to determine the value of K that produced the best-fitting model. Results of ADMIXTURE analyses were plotted using 'Pophelper' v.2.2.6 (https://github.com/royfrancis/pophelper, last accessed in 9 July 2018) and 'viridis' (https://github.com/sjmgarnier/viridis, last accessed in 9 July 2018) packages in R (R Core Team, 2019).

We also evaluated the potential for population structure and isolation by distance to account for genetic variation within the CVOS using the program conStruct (Bradburd *et al.*, 2018). Given that it is particularly sensitive to missing data, we ran conStruct on a subset of our nuclear RADseq dataset that excluded individuals with > 50% missing data and loci with > 5% missing data. We used the cross-validation approach implemented in the conStruct package to test spatial and non-spatial models with K = 1–12 and compared the resulting predictive accuracy of each model with the individual layer contributions to determine the value of K that was sufficient to describe our data (K = 3; see Results). We then ran three independent conStruct analyses for both spatial and non-spatial models at this value of K with 10 000 iterations each and confirmed convergence of analyses using trace plots in conStruct.

We evaluated genetic diversity within CVOS lineages by estimating the parameter θ (4 $N_e\mu$; effective population size = $N_e\mu$) for each population in a Bayesian framework using *ThetaMater* (Adams *et al.*, 2018), which uses an infinite-sites likelihood function to estimate a posterior probability distribution of θ. To obtain sets of complete RAD loci, including variant and invariant sites, we first generated a concatenated BAM file from all individual mappings and extracted RAD locus regions that had a minimum of 6000 × coverage summed across all samples, corresponding to an average of 30 times per individual. We then sampled up to five individuals (i.e. ten alleles) with the lowest amount of missing data per population and used the BCFtools (Li *et al.*, 2009) filter function to subset individual variant information and filter to remove any variants called based on less than five times coverage per individual.

These variant sites were then cross-referenced with the RAD locus region coordinates described above to retain only regions for which an individual had variant data. Finally, we used the BCFtools consensus function to extract full sequence information per RAD locus and to input variant information for both alleles per individual, and combined biallelic information from each individual per locus as input. We then ran the *ThetaMater*M1 MCMC simulation model to estimate the posterior distribution of θ using a total of 1×10^6 generations per analysis and discarded the first 10% of generations per analysis as burn-in.

ANALYSIS OF TWO-POPULATION DEMOGRAPHIC HISTORIES

We performed a series of pairwise analyses of populations to identify the best-fitting models of demographic history based on our nuclear data. These analyses were designed to test the hypothesis that lineages with overlapping or adjoining distributions show evidence of gene flow (as suggested by migration edges in TreeMix analysis and individuals with mixed population assignments in ADMIXTURE analyses). Specifically, we excluded lineages that did not correspond to mitochondrial clades or demonstrate distinct nuclear genetic structure (based on ADMIXTURE inferences). We tested hypotheses of population demographic history by analysing two-dimensional site frequency spectra (2D-SFS) from our nuclear data for a total of nine two-population comparisons with lineages that have overlapping or adjoining ranges (Table 1; Fig. 2), using the diffusion approximation approach implemented in $\delta a \delta i$ (Gutenkunst *et al.*, 2009). For each analysis, we tested eight putative models of two-population demographic history, including divergence models with and without gene flow and that allowed for inference of current and ancestral relative effective population size (Supporting Information, Dataset S1; Fig. S2). We formatted $\delta a \delta i$ input site frequency spectra by converting variants in VCF format detailed in 'Data collection' above using the 'vcf2dadi' conversion tool implemented in the R (R Core Team, 2019) package 'radiator' (Gosselin, 2019). We projected down to smaller sample sizes within populations to minimize missing data and maximize the number of segregating sites for each analysis (Table 1), with an average of 4585.77 segregating sites per analysis.

We performed initial parameter optimizations by generating 50 sets of randomly perturbed parameters (for three grid sizes per analysis) over a maximum of 20 iterations using the 'optimize_log_fmin' method (described as the 'amoeba' method in the $\delta a \delta i$ manual). This simulated a 2D-SFS for each set of parameters,

Table 1. Summary of two-population divergence analyses performed using the two-dimensional allele frequency spectrum (2D-AFS) between western rattlesnake species complex (CVOS) populations, including the number of alleles sampled after downsampling to balance missing data and the number of segregating sites, and the best-fitting model per analysis

Analysis	Population 1 alleles	Population 2 alleles	Segregating sites	Best-fitting model
C. o. concolor and *C. o. lutosus*	20	10	5560.68	Divergence with asymmetrical secondary contact, size change
C. o. concolor and *C. v. viridis*	30	30	12 384.74	Divergence with asymmetrical secondary contact, size change
C. o. helleri and *C. o. oreganus*	10	24	3888.53	Divergence with asymmetrical secondary contact, size change
C. o. helleri and *C. s. scutulatus**	12	16	3635.57	Divergence with asymmetrical secondary contact, size change
C. o. lutosus and *C. o. oreganus*	14	20	3941.41	Divergence with asymmetrical secondary contact, size change
C. s. scutulatus and *C. v. viridis**	14	16	3698.03	Divergence with asymmetrical secondary contact, size change
C. v. nuntius and *C. v. viridis*	6	10	1680.62	Divergence with asymmetrical migration, size change
C. s. scutulatus (MS) and *C. s. scutulatus* (Ch)	16	10	5685.35	Divergence with asymmetrical secondary contact, size change
C. s. scutulatus (Ch) and *C. s. scutulatus* (CP)	10	10	5843.59	Divergence with asymmetrical secondary contact, size change

Model testing and parameter details are provided in the Supporting Information (Dataset S1).
Abbreviations: Ch, Chihuahuan; CP, Central Mexican Plateau; MS, Mohave–Sonoran.
*Analysis resulted in two models with similar support. The alternative model was divergence with asymmetrical migration. For additional details, see Supporting Information (Fig. S7; Dataset S1).

and we used the multinomial approach to calculate the log-likelihood of the SFS given each model and assessed model likelihood using the Akaike information criterion (AIC). We performed two additional rounds of parameter optimization using starting parameter values with the best-scoring AIC value per model from the previous round of optimization. The final optimizations included 100 replicates with a maximum of 50 iterations per replicate. We identified the replicate with the highest likelihood per model, then compared AIC values and Akaike weights (Burnham & Anderson, 2003) across models to determine the best-fitting model per analysis. Models and scripts used for these analyses were modified from those used by Portik *et al.* (2017).

Several pairwise analyses included non-sister lineages. To account for the potential influence of gene flow from other more closely related lineages, we also performed three-population analyses specifically designed to confirm results from pairwise analyses that included *C. o. concolor*, *C. o. helleri*, *C. v. viridis* and *C. s. scutulatus* (see Results). Here, we tested six alternative demographic models (Supporting Information, Fig. S2), again including a no-migration model, and multiple models to estimate the relative timing of gene flow between lineages. In all pairwise and three-population analyses, our principal aim was to determine best-fitting models, thus we did not transform unscaled parameter estimates to biological estimates, and report on relative unscaled population size and migration parameters within pairwise and three-population sets of analyses.

ECOLOGICAL NICHE MODELLING

We reconstructed suitable climatic niches within the CVOS for current climatic conditions and those of the Last Glacial Maximum (LGM) across the range of each taxon using ecological niche modelling. For occurrence data, we used occurrence records from the VertNet database (www.vertnet.org). Here, we sampled occurrence records for eight CVOS taxa; data for all *C. scutulatus* lineages were combined, *C. o. abyssus* records were combined with *C. o. lutosus*, and *C. o. caliginis* lacked sufficient locality records for analysis. All records from VertNet were inspected visually and records outside the known distribution of each taxon removed. We then filtered the occurrence records using the R package spThin (Aiello-Lammens *et al.*, 2015) to include only one occurrence record per 10 km. This filtering alleviated potential bias caused by unequal sampling effort (Merow *et al.*, 2013).

We derived the current climatic niche of the species using bioclimatic variables at a 2.5′ (~5 km) resolution from the WorldClim dataset v.2.0 (Hijmans *et al.*, 2005). We derived the LGM climatic niche per taxon using two simulation models: community climate system model (CCSM v.3; Otto-Bliesner *et al.*, 2006) with a resolution of 1°, and the model for interdisciplinary research on climate (MIROC v.3.2; Sugiyama *et al.*, 2010) with an original spatial resolution of 1.4° × 0.5° (Braconnot *et al.*, 2007). These climatic variables were downscaled to the spatial resolution of 2.5′ (under the assumption of high spatial autocorrelation) and converted to bioclimatic variables (Hijmans *et al.*, 2005; Peterson & Nyári, 2008). The two models both indicate colder and wetter climate during the LGM; however, the CCSM model predicts lower values across temperature variables, whereas the MIROC model predicts higher values across precipitation variables (see Jezkova *et al.* 2016). Although machine-learning algorithms (i.e. MaxEnt) have been shown to be robust to correlation among variables, we nevertheless removed highly correlated variables among the 19 bioclimatic variables by selecting a subset of variables with all pairwise correlation coefficients < |0.9|. When one of two correlated variables needed to be removed, we selected seasonal averages to be kept over monthly and yearly averages. The resulting models included 12 variables for each climatic dataset.

We constructed climatic niche models for each climatic dataset in the program MaxEnt v.3.3.3k (Phillips *et al.*, 2006) using the R packages 'ENMeval' (Muscarella *et al.*, 2014) and 'dismo' (Hijmans *et al.*, 2015). For each taxon, we used 1000 background points randomly extracted from a polygon drawn around the occurrence records for that taxon and expanded by 10° in all directions. This selection of background points was used to exclude distant areas with very different environmental conditions, following recommendations by Merow *et al.* (2013). We explored values for the regularization multiplier (rm) between 0.5 and 4 (in increments of 0.5) and all combinations of available features (i.e. linear, quadratic, product, threshold and hinge). We ran threefold cross-validation replicates to choose a model with the best fit, as assessed by the lowest corrected Akaike information criterion (AICc) value. The best-fitting models for each climatic dataset were visualized using logistic probability values (Merow *et al.*, 2013) and converted to presence–absence maps using the equal training sensitivity and specificity threshold.

RESULTS

EVIDENCE AGAINST A STRICTLY BIFURCATING TREE MODEL

TreeMix generated a population graph in which a model without migration events explained the majority (~96%) of the variance in relatedness among CVOS lineages (Fig. 2B; Supporting Information, Fig. S3). We also tested models with a range of migration events (i.e. 'migration edges', *m*), in which migration between population branches may explain greater genetic variance than a model without migration. The proportion of variance explained by models allowing for successively greater numbers of migration events increased and plateaued at ~99.9% (Supporting Information, Fig. S3). Here, we focus on the results with *m* = 8 migration edges, which indicate that the evolutionary history of the CVOS does not fit a strictly bifurcating tree model and that divergence with migration and admixture has been prevalent (Fig. 2B), with divergence in the absence of gene flow contributing the most to the variance in relatedness among lineages. Migration edges were also concentrated at the tips of the population graph, consistent with most major migration events occurring after initial lineage divergence.

MITOCHONDRIAL PHYLOGENY AND DIVERGENCE TIME ESTIMATES

Although tests of our nuclear dataset revealed evidence against a strictly bifurcating tree, our mitochondrial gene analyses yielded strong support for distinct mitochondrial clades within the CVOS. Our Bayesian phylogenetic analysis inferred 100% posterior probability (PP) for monophyly of the CVOS group, which diverged from the *C. atrox* + *C. ruber* clade during the Miocene (~7.32 Mya; Fig. 3A; Supporting Information, Fig. S4). Lineages within the CVOS diverged from a common ancestor between the Late Miocene and early Pliocene (~5–7 Mya; Fig. 3A). These estimates are consistent with previous analyses that included more limited sampling of CVOS lineages (e.g. Reyes-Velasco *et al.*, 2013; Zheng & Wiens, 2016). Within the CVOS, we found support (100% PP) for the divergence between *C. scutulatus* and *C. viridis* + *oreganus* + *cerberus* clades and for the more recent split between *C. viridis* and *C. oreganus* + *cerberus*, which occurred ~5.47 and ~5.24 Mya, respectively (Fig. 3A).

We inferred numerous splits within the CVOS that occurred at various stages during the Pleistocene (i.e. within ~2.5 Myr), including four clades within *C. scutulatus* previously identified by Schield *et al.* (2018), and clades corresponding to *C. viridis*, *C. v. nuntius*, *C. cerberus*, *C. o. oreganus*, *C. o. lutosus*, *C. o. helleri* and *C. o. concolor*. We infer that *C. cerberus* and *C. oreganus* diverged ~2 Mya, with more recent diversification evident within *C. oreganus* (i.e. within the last 1 Myr; Fig. 3A). Other CVOS subspecies were supported as being nested within these clades: *C. o. abyssus* is nested within *C. o. lutosus*, and

C. o. caliginis is nested within *C. o. helleri*. Several individuals identified as *C. o. concolor* were nested within the *C. viridis* clade, suggesting mitochondrial introgression between *C. viridis* and *C. o. concolor*.

NUCLEAR ESTIMATES OF POPULATION STRUCTURE AND GENETIC DIVERSITY

ADMIXTURE analysis of 200 individuals that passed missing data threshold requirements (see Material and Methods) found support for a best-fitting model of *K* = 4 genetic clusters using the cross-validation approach (Fig. 3B; Supporting Information, Fig. S5). Under this model, we found evidence for genetic clusters corresponding to *C. scutulatus*, *C. viridis*, *C. o. concolor* and a fourth major cluster corresponding to *C. cerberus* and the remaining *C. oreganus* lineages. The second-best-supported model, *K* = 5, showed similar patterns, but identified a distinct *C. o. lutosus* genetic cluster. Models with higher *K* values (e.g. *K* = 11, equal to the number of recognized subspecies included in this study), identified additional clusters corresponding to *C. v. nuntius* and the *C. oreganus* subspecies *C. o. oreganus*, *C. o. helleri*, *C. o. lutosus* and *C. o. concolor*, whereas other subspecies (i.e. *C. s. salvini*, *C. o. caliginis* and *C. o. abyssus*) were not defined as distinct clusters (Fig. 3B). We observed several consistent patterns across models of *K*. First, there was consistent evidence for genomic similarity between *C. o. abyssus* and *C. o. lutosus* and between *C. o. caliginis* and *C. o. helleri* (Fig. 3B). Second, no models of *K* inferred a unique *C. cerberus* genetic cluster. Third, we found consistent evidence for individuals with ancestry from multiple genetic clusters across all models, suggesting potential gene flow and admixture between structured populations across the CVOS.

To interrogate population structure in the context of the geographical distribution of the CVOS, we performed conStruct analyses using a subset of individuals and loci with low missing data (because conStruct is highly sensitive to missing data). Predictive accuracy of spatial models plateaued at *K* = 3 (Supporting Information, Fig. S6A), and individual layer contributions indicated that additional layers contributed little to explaining covariance among samples (Supporting Information, Fig. S6B); we thus determined that *K* = 3 was sufficient to describe our data, as suggested by Bradburd *et al.* (2018). These three layers corresponded broadly to *C. oreganus*, *C. viridis* and *C. scutulatus*; Supporting Information, Fig. 6C, D). Notably, individuals with mixed assignment were identified in regions where admixture would be predicted based on contact of adjoining ranges, including the contact zones between *C. oreganus* (*concolor*) and *C. viridis* near the northern Colorado–Utah border and in north-eastern Arizona, and between *C. viridis* and *C. scutulatus* at the

southern Arizona–New Mexico border and in south-western Texas (Supporting Information, Fig. S6D).

Effective population sizes based on the *θ* parameter estimated using *ThetaMater* varied considerably across CVOS lineages, with median *θ* ranging from 0.00127 to 0.00328 (Fig. 4). The majority of lineages had posterior *θ* distributions between 0.002 and 0.003. As predicted, *C. s. salvini* and *C. o. caliginis* had substantially lower posterior *θ* distributions; *C. s. salvini* occupies a relatively small geographical region at the southern extreme of the *C. scutulatus* distribution, and *C. o. caliginis* is found only on Coronado Island off the coast of Baja California. Genetic diversity was inferred to be comparatively much higher in the closest relatives of each of these lineages. In contrast, the two lineages with the greatest posterior *θ* distributions were *C. s. scutulatus* 'Chihuahuan' and *C. o. lutosus*, both of which occupy large, continuous geographical distributions. However, we observed relatively high *θ* for *C. o. abyssus*, despite its restricted range within the Grand Canyon. Higher than expected genetic diversity in *C. o. abyssus* might be explained by gene flow into the Grand Canyon from multiple adjacent lineages (i.e. *C. cerberus*, *C. o. concolor*, *C. o. lutosus* and *C. v. nuntius*).

NUCLEAR ESTIMATES OF DEMOGRAPHIC HISTORY AND PATTERNS OF GENE FLOW

Analyses of the 2D-SFS in *δaδi* supported demographic models of allopatric divergence followed by secondary contact with gene flow for eight of the nine two-population comparisons (Fig. 5; Supporting Information, Fig. S7; Table 1; Dataset S1). These models also highlighted repeated patterns of relative effective population size change through time. The eight comparisons that fitted secondary contact models consisted of focal lineages with a range of mitochondrial divergence time estimates. For example, we found support for secondary contact between lineages that diverged in the Miocene (e.g. *C. s. scutulatus* and *C. o. helleri*; ~5.47 Mya; Fig. 5A), Pliocene (e.g. *C. o. concolor* and *C. v. viridis*; ~5.24 Mya; Fig. 5C) and Pleistocene (e.g. *C. o. helleri* and *C. o. oreganus*; ~1.19 Mya; Fig. 5E). These results indicate that secondary contact between lineages has occurred across a continuum of population divergence. Support for this model was mostly unambiguous across analyses, with the exception of *C. o. helleri* and *C. s. scutulatus* and between *C. v. viridis* and *C. s. scutulatus*. In both cases, models involving asymmetric continuous gene flow were also well supported (Supporting Information, Fig. S7; Dataset S1). Given that ambiguity could be the result of analysis of non-sister lineages that also share gene flow with closer relatives, we also performed three-population analyses to verify evidence of secondary
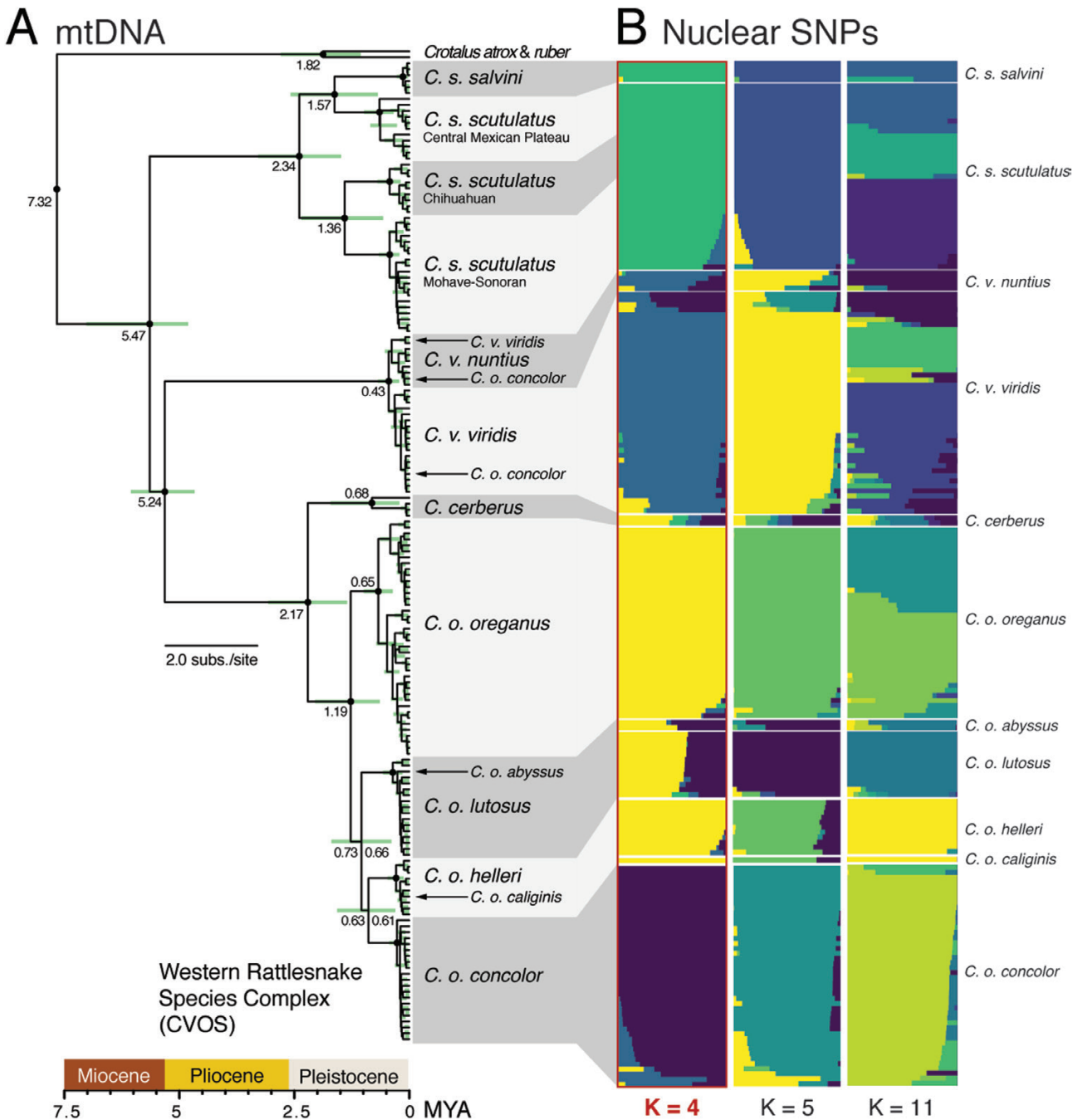
Figure 3. A, Bayesian mitochondrial gene phylogeny of the western rattlesnake species complex (CVOS) and the outgroup clade (*Crotalus atrox* + *Crotalus ruber*); all outgroups are shown in the Supporting Information (Fig. S4). Nodes with > 0.95 posterior support are shown with a black dot. For nodes with less posterior support, the posterior probability is labelled on the right of the node. Median estimated divergence times (in millions of years ago) are labelled on the left of each node. The 95% highest posterior density (HPD) for each divergence time estimate is depicted with a green bar. The scale bar at the bottom shows the timescale of divergence (in millions of years ago). B, results of ADMIXTURE analysis on the nuclear RADseq dataset of 200 individuals that met filtering requirements (see Material and Methods), under three genetic cluster models (K = 4, 5 and 11). Each bar represents an individual, with the relative probability of assvignment to genetic clusters shown as the proportion of the bar shaded with different colours. The K = 4 model was the best-supported model based on the cross-validation approach to compare increasing values of K. Correspondence between panels A and B is shown with alternating grey shaded regions.
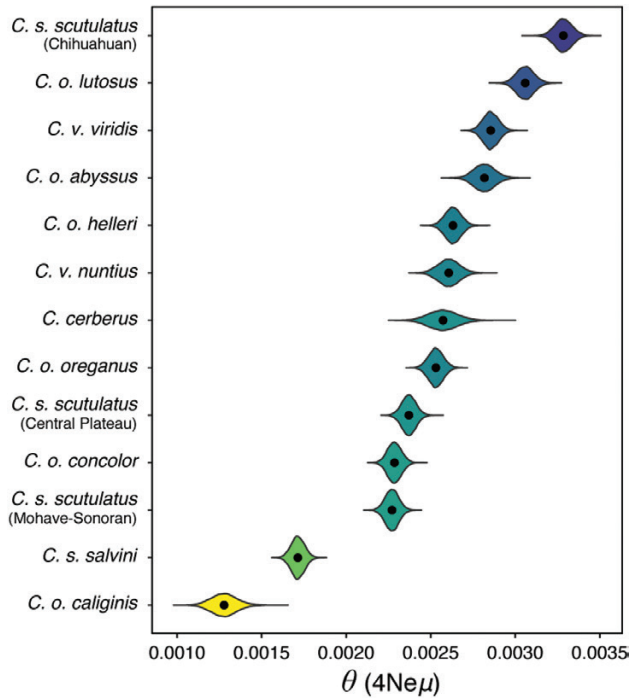
Figure 4. Violin plot of posterior distributions of genetic diversity (θ), estimated using *ThetaMater*. Distributions were estimated for each taxon independently. Wider areas within each violin represent greater posterior density, and black dots represent the median estimate per taxon. Violins are shaded and shown in ascending order based on the median posterior estimate, with darker blues showing higher θ.

contact between *C. s. scutulatus* and *C. o. helleri* and between *C. s. scutulatus* and *C. v. viridis* (Supporting Information, Fig. S8); results from these analyses indeed yielded support for gene flow in secondary contact between these non-sister lineages (Supporting Information, Dataset S2). We therefore dismiss the continuous migration model as less biologically plausible, with the caveat that the tested models (even three-population models) probably do not capture the complexity of the demographic history between these lineages (see Discussion).

The majority of the lineages with overlapping or adjoining current distributions diverged from common ancestors during the Pleistocene (Fig. 5D–I). Analyses including these lineages each yielded a best-supported demographic model of divergence in isolation followed by secondary contact with gene flow, with the exception of *C. v. nuntius* and *C. v. viridis*, for which a model of divergence with continuous asymmetric gene flow was supported (Fig. 5G). This inference is notable because it is consistent with multiple other lines of evidence (e.g. mitochondrial haplotype sharing and evidence for admixed individuals between *C. v. nuntius* and

C. v. viridis; Fig. 3), suggesting that these two lineages have shared gene flow and diverged relatively recently (i.e. ~0.43 Mya). Divergence between *C. v. viridis* and *C. v. nuntius* might thus be a product of isolation by distance, in which case these lineages provide the most compelling example of divergence without a period of physical isolation within the CVOS. To test explicitly that isolation by distance has driven divergence between *C. v. nuntius* and *C. v. viridis*, we performed a Mantel test to test for a significant relationship between pairwise genetic distances and pairwise geographical distances using the mantel. randtest function in the R package 'adegenet' (Jombart, 2008). We performed 99 999 permutations to generate a random distribution of values to compare our empirical pairwise values. We plotted genetic and geographical distances to inspect visually for evidence of a continuous cline of genetic distance vs. distinctive genetic breaks between *C. v. nuntius* and *C. v. viridis* (Supporting Information, Fig. S9). The Mantel test was significant ($P = 0.0284$), indicating support for divergence between *C. v. nuntius* and *C. v. viridis* owing to isolation by distance.

Owing to limitations of Mantel tests in identifying isolation by distance (Legendre *et al.*, 2015) and because environmental factors may play a substantial role in clinal variation in addition to physical distance, we performed redundancy analyses on *C. v. viridis* and *C. v. nuntius* individuals as described by Myers *et al.* (2016) to examine how much variation in genetic distances is explained by geographical distance vs. environmental distance. We used abiotic factors associated with our sampling localities from the WorldClim dataset and performed principal components analyses that included geographical and environmental distances as predictor variables and a normalized genetic distance matrix as the response variable. We conducted three analyses that incorporated combinations of all variables to quantify which combination was the best predictor of genetic distances and tested for statistical significance using ANOVA on the r^2 values of each analysis. The model incorporating full environmental distance (i.e. all climatic variables) and geographical distance variables was the best predictor of genetic distance, but all models were significant (i.e. $P < 0.001$), and the amount of genetic variance explained ranged from 14 to 51% (Table 2).

ECOLOGICAL NICHE MODELLING AND PROJECTIONS OF PAST SUITABLE HABITAT

Ecological niche models generated for the eight taxa from VertNet occurrence records predicted current climatic niches largely overlapping with the approximate distributions of each of the CVOS lineages (Figs 2A, 6A). This was consistent across climatic models used to project suitable habitat
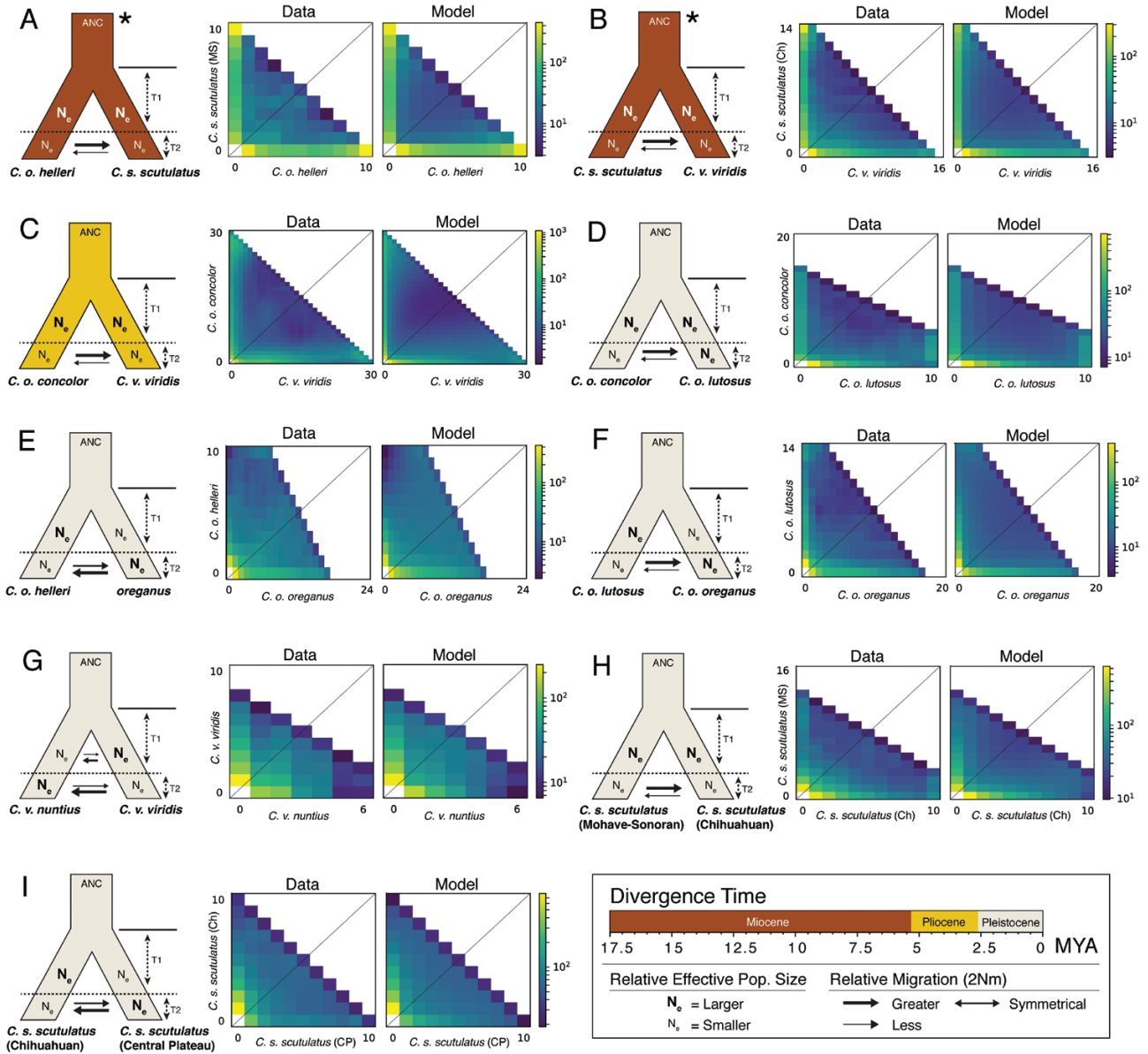
Figure 5. Results of demographic model testing in $\delta a \delta i$ for two-population demographic histories, showing the best-supported model per analysis after final optimization steps. A, B, lineages with estimated Miocene divergence. C, lineages with estimated Pliocene divergence. D–I, lineages with estimated Pleistocene divergence. For each analysis, the schematic of the best-fitting model is shown on the left, followed by the two-dimensional allele frequency spectra for the data (middle) and simulated model (right). The scale of relative allele frequencies is shown to the right of each simulated model, with lighter colours representing higher frequencies. Unscaled parameter estimates (nμ1a, nμ2a, nμ1b, nμ2b, m12, m21, t1 and t2) are shown for each model, and correspond to the Supporting Information (Dataset S1). We did not scale parameters to biological estimates; however, we show greater relative effective population size as larger, bold N_e labels, and greater migration rates as larger, bolder arrows. *The *Crotalus oreganus helleri* and *Crotalus scutulatus scutulatus* and the *C. s. scutulatus* and *Crotalus viridis viridis* analyses yielded similar support for two models; see the Supporting Information (Dataset S1; Fig. S7).

(Supporting Information, Fig. S10). Values of the area under the curve for all models were > 0.88, with a mean area under the curve across various models and analyses of 0.94. Projected reconstructions of

current climatic niches onto LGM conditions indicate that distributions of suitable habitat were more restricted and shifted strongly to the south compared with present-day distributions under both CCSM and

MIROC LGM models (Fig. 6B). Considering the CVOS distribution as a whole, we infer from these models that much of the currently occupied landscape was likely to be unsuitable during glacial periods, similar to the pattern observed by Lawing & Polly (2011).

Comparing current vs. LGM projections, we found evidence for large northward expansions of suitable habitat across several CVOS lineages from past to present (Fig. 6B). For example, the inferred LGM range for *C. v. viridis* was apparently restricted to the Chihuahuan Desert, and this lineage has since expanded northwards into the Great Plains region. Likewise, the climatic niche of *C. o. lutosus* was historically restricted to the Death Valley and western regions of the Mohave Desert and has since expanded into the entire Great Basin region. Furthermore, we infer that the northern extent of the *C. o. oreganus* distribution was historically unoccupied, and this lineage has since expanded northwards into Cascadia. Inferred LGM habitat across other lineages was fragmentary and shifted southwards relative to current distributions (e.g. *C. cerberus*, *C. scutulatus*, *C. o. helleri* and *C. v. nuntius*).

A broad pattern across LGM projections is evidence of historically uninhabited regions where lineages currently come into contact with one another (e.g. *C. o. concolor* and *C. o. lutosus*, *C. o. helleri* and *C. s. scutulatus*, and *C. v. viridis* and *C. s. scutulatus*) or more restricted overlap between currently overlapping lineages (e.g. *C. cerberus* and *C. s. scutulatus*; Fig. 6). These results provide a historical geographical context for patterns of repeated secondary contact observed in our genetic data (Fig. 5). We infer that ancestrally separated lineages have expanded into secondary contact, facilitating gene flow between historically isolated populations.

DISCUSSION

DIVERGENCE, POPULATION STRUCTURE AND GENE FLOW ACROSS THE SPECIES COMPLEX

Using our range-wide sampling of CVOS, we first investigated genetic evidence for lineage diversity and relationships among groups. We found consistent evidence for multiple divergent lineages within the CVOS (Figs 2B, 3A); however, these lineages do not strictly correspond to subspecies as currently recognized. Based on analyses of mitochondrial DNA, we find evidence that some subspecies represent well-differentiated lineages (e.g. *C. o. concolor*), evidence of well-differentiated lineages within *C. scutulatus* not ascribed to subspecies, and evidence that yet other recognized subspecies (e.g. *C. o. abyssus* and *C. o. caliginis*) do not appear genetically distinct. Previous studies have used mitochondrial DNA to interrogate relationships within *C. viridis* + *C. oreganus* (Pook *et al.*, 2000; Ashton & de Queiroz, 2001; Douglas *et al.*, 2002); our study improves upon previous estimates by including a robust sampling of additional mitochondrial haplotypes ($N = 168$ unique haplotypes) from diverse localities and populations and the first analyses of *C. viridis* + *C. oreganus* together with its sister taxon, *C. scutulatus*.

Overall, our mitochondrial tree topology is similar to the estimate of Pook *et al.* (2000). This includes the inferred sister relationship between *C. o. helleri* and *C. o. concolor*, although this node had low posterior support (PP = 0.61). The monophyly of *C. o. concolor* and *C. o. helleri* is difficult to reconcile with geography and patterns of nuclear genomic variation; the ranges of the two subspecies are separated by hundreds of kilometres, and we found consistent support for *C. o. helleri* sharing more recent common ancestry with *C. o. oreganus* in nuclear analyses (Figs 2B, 3B). As suggested by Pook *et al.* (2000), a true sister relationship between *C. o. concolor* and *C. o. helleri* would require the extinction of a very wide-ranging intermediate ancestral population. This is inconsistent with our estimates of past suitable climatic niches (Fig. 6B), which indicate that there was unlikely to be suitable habitat between the current ranges of these two subspecies that could have sustained an intermediate population during glacial periods of the Pleistocene.

Divergence estimates placed the initial divergence of the CVOS from a common ancestor during the Late Miocene or early Pliocene ~5.5 Mya.

Table 2. Results of redundancy analysis to investigate the roles of geographical distance and environmental distances (i.e. climatic variables from the WorldClim database) in genetic distances among *Crotalus viridis viridis* and *Crotalus viridis nuntius*

Model	Variance (adjusted r^2)	Variance (r^2)	*P*-value
Full (geographical distance + environment)	0.19	0.51	< 0.001
Environment	0.19	0.48	< 0.001
Geographical distance	0.1	0.14	< 0.001

The 'full' model, which accounted for the combined effects of geographical and environmental distances, explained the greatest variance in the genetic distance response variable.
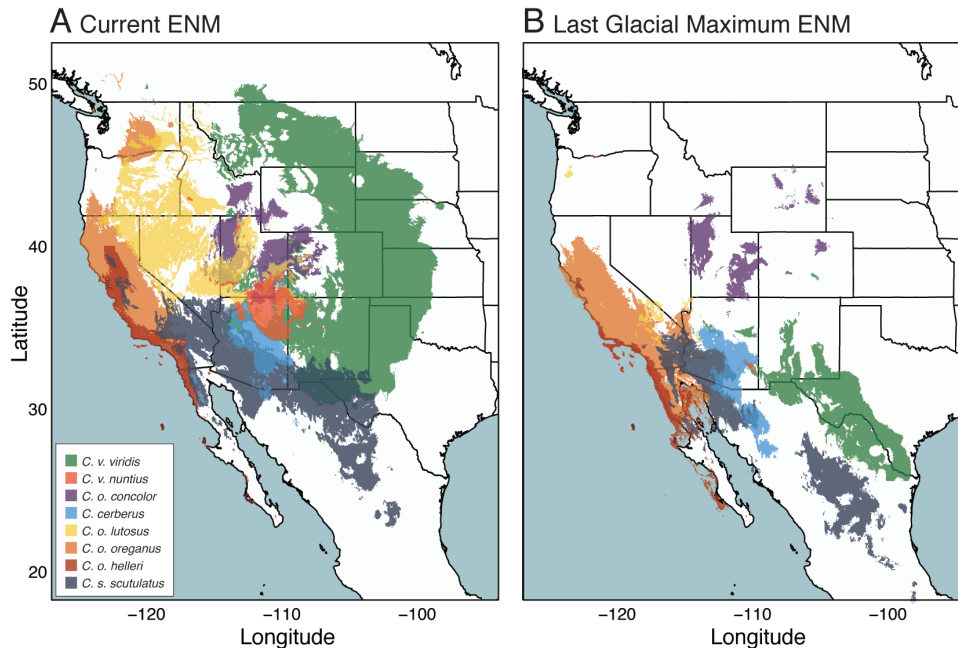
Figure 6. A, B, current (A) and Last Glacial Maximum (LGM; B) projections of suitable habitat for western rattlesnake species complex (CVOS) lineages using ecological niche modelling and the community climate system model (CCSM) model. Current and LGM presence maps for each taxon correspond to areas with a high logistic probability of suitable habitat (i.e. ≥ 0.8). Inferred regions of potentially suitable habitat with lower logistic probabilities, in addition to projections using the model for interdisciplinary research on climate (MIROC) model, are shown in the Supporting Information (Fig. S10).

Diversification within the species complex has thus been influenced by dynamic shifts in climate, habitat and ecological communities during the Pliocene and Pleistocene, as suggested for other rattlesnake species inhabiting this region (Douglas *et al.*, 2006). We infer that the major divergence event between *C. viridis* and *C. cerberus* + *C. oreganus* lineages occurred soon after the split between their common ancestor and *C. scutulatus* (Fig. 3A), during the Pliocene. This would have coincided with the uplift of the Colorado Plateau and bisection of Mojavia (Axelrod, 1958; Morafka, 1977), as emphasized by Douglas *et al.* (2002) and Castoe *et al.* (2007). Douglas *et al.* (2002) inferred Pliocene origins for the majority of subspecies within *C. oreganus* and *C. viridis* by applying various molecular clocks to levels of pairwise sequence divergence. Although there is uncertainty associated with divergence time estimation regardless of the method, we expect that our divergence time inferences are robust because they are derived from a highly diverse and fairly complete set of populations and lineages and were explicitly calibrated using numerous outgroup taxa with informative priors from fossil data. We find evidence contradicting the hypothesis from Douglas *et al.* (2002), in favour of more recent Pleistocene diversification (i.e. < 2.6 Mya) within major lineages, including *C. oreganus*, *C. viridis* and *C. scutulatus*.

Inferred shifts in lineage ranges from our ecological niche modelling results provide context for Pleistocene diversification within the CVOS and the complexity of gene flow dynamics through time (Fig. 6). In general, projections of suitable habitat onto Pleistocene glacial conditions suggest that the distributions of all CVOS taxa were shifted southwards, and more constrained and fragmented during glacial periods. This phenomenon is widely implicated as a major driver of diversification in temperate regions during the Pleistocene, especially in the geographical region that the CVOS inhabits (e.g. Castoe *et al.*, 2007; Leaché & Mulcahy, 2007; Mulcahy, 2008; Graham *et al.*, 2013; Jezkova *et al.*, 2015, 2016; Myers *et al.*, 2016). Although LGM projections indicate broad regions of allopatry between some CVOS lineages (e.g. *C. o. concolor* and *C. v. viridis*, *C. o. concolor* and *C. o. lutosus*, and *C. scutulatus* and *C. v. viridis*), historical distributions of other lineages indicate potential co-localization into restricted southern refugia (e.g. *C. o. oreganus* and *C. o. helleri*, and *C. cerberus* and *C. scutulatus*). Further detailed examination of the temporal history of range shifts, combined with analyses of more complex demographic models, would improve our understanding of the complex patterns and impacts of gene flow in sympatry and allopatry through time.

Despite numerous distinct mitochondrial clades within the CVOS (Fig. 3A), the precise number of

distinct lineages is less clear based on our nuclear genomic dataset. TreeMix analyses demonstrated that models of diversification that include migration events explain genetic variation in the CVOS better than a strictly bifurcating tree topology (Fig. 2B; Supporting Information, Fig. S3). Furthermore, the concentration of migration edges towards the tips of the tree suggests that gene flow between populations has occurred after a period of divergence. Based on these same data, analyses of population genetic structure demonstrated that a model of $K = 11$ genetic clusters (the number of recognized subspecies) is a poorer fit to our data than lower values of K (e.g. $K = 4$ or 5; Fig. 3B; Supporting Information, Fig. S5). Notably, as we tested higher numbers of genetic clusters (e.g. $K = 11$), changes to population assignments did not result in fitting discrete clusters to all recognized subspecies, but rather inferred genetic substructure within widely ranging lineages (e.g. *C. o. oreganus* and *C. v. viridis*) or splitting population assignments within individuals to multiple populations. ConStruct analyses indicated that the predictive accuracy of spatial models consistently explained variation in relatedness better than non-spatial models, indicating that both discrete population structure and isolation by distance contribute to patterns of genetic variation in the CVOS (Supporting Information, Fig. S6A). In conStruct analyses, a K of 3 was sufficient to describe the data and, notably, did not distinguish *C. o. concolor* as a distinct genetic cluster, in contrast to our ADMIXTURE analysis (Supporting Information, Fig. S6D). These analyses identified introgression between these three groups (representing divergent lineages) in regions of geographical contact independent of current subspecies designations. Thus, regardless of the K and model used to fit the data, all analyses identified a consistent pattern of individuals with admixed population assignments (Fig. 3B). These results collectively underscore the pervasive existence of admixture between multiple differentiated lineages within the CVOS. We note, however, that although the results of ADMIXTURE and conStruct analyses indicate models that are sufficient to describe patterns of variation in our data (Alexander *et al.*, 2009; Bradburd *et al.*, 2018), they should not necessarily be taken out of context to represent a proposal for the number of species that should be recognized within this complex.

DEMOGRAPHIC HISTORY AND SECONDARY CONTACT BETWEEN HISTORICALLY ISOLATED LINEAGES

Given evidence for gene flow between CVOS lineages, we leveraged the model selection framework in δαδi (Gutenkunst *et al.*, 2009) to compare alternative demographic histories using two-dimensional allele

frequency spectra from pairs of CVOS lineages. Given that characterization of the relative timing of introgression between lineages is important to understanding the historical and geographical context of diversification, we tested models with ancestral gene flow, continuous gene flow, and gene flow in secondary contact. We also tested models that parameterized ancestral and current effective population sizes, because historical shifts in lineage ranges are predicted to result in changes in population size through time. We found widespread evidence for gene flow between CVOS lineages that occurred after periods of allopatric divergence (Fig. 5; Supporting Information, Fig. S8; Table S1; Datasets S1 and S1) and evidence that most CVOS lineages have experienced reductions in effective population size through time. Evidence of secondary contact was found across the geographical distribution of the species complex between lineages that collectively span a broad range of divergence times (from < 1 to > 5 Mya). These findings argue that allopatry has been a major driver of CVOS diversification and that secondary contact with gene flow has occurred repeatedly between many lineages. In contrast, demographic estimates for *C. v. nuntius* and *C. v. viridis* suggest divergence with continuous gene flow in the absence of complete physical isolation, and we infer that differentiation between populations has evolved owing to isolation by distance (Table 2; Supporting Information, Fig. S9) and environmental heterogeneity (i.e. isolation by environment); these findings align with Klauber's hypothesis of clinal variation between *C. v. nuntius* and *C. v. viridis* (Klauber, 1935).

Although our approach allowed us to test hypotheses regarding the relative timing of gene flow, it is important to consider some key limitations of our inferences. As with any model-based approach, it is likely that our models do not capture the natural complexity of the system fully, and instead approximate its historical demography (Johnson & Omland, 2004). For example, Quaternary climate change involved multiple glacial and interglacial cycles (Davis & Shaw, 2001; Hewitt, 2004), probably resulting in repeated population range contraction and subsequent range expansion over time. This would, in turn, lead to episodic periods of allopatry and periods of secondary contact. Thus, evidence for secondary contact between populations might reflect the culmination of multiple contact events through time. In two cases (*C. o. helleri* and *C. s. scutulatus*, and *C. s. scutulatus* and *C. v. viridis*), pairwise demographic model selection yielded support for a 'secondary contact' model and a 'continuous gene flow' model. Continuous gene flow between *C. scutulatus* and either *C. o. helleri* or *C. v. viridis* is unlikely given their deep mitochondrial divergence (Fig. 3A); however, it is possible that multiple secondary contact events

have resulted in an allele frequency signature that is difficult to distinguish from consistent migration during divergence (Supporting Information, Fig. S7). Our three population analyses were designed to investigate these demographic histories further, and provided additional support for secondary contact between *C. s. scutulatus* and *C. o. helleri* and *C. v. viridis*, respectively (Supporting Information, Fig. S8; Dataset S2). As stated above, our focus in this study was to understand the relative timing of gene flow between divergent lineages, and future studies that test more complex demographic models could be useful for quantifying in more detail the number and frequency of secondary contact events.

INTROGRESSION ACROSS A CONTINUUM OF LINEAGE DIVERGENCE

Our results indicate that distinct lineages within the CVOS experience introgression, regardless of the apparent divergence between lineages, which in some cases is ancient (e.g. ~5 Mya for the split between *C. s. scutulatus* and *C. v. viridis*). This broad evidence for introgression between lineages across a continuum of divergence highlights a surprising apparent lack of pre- or postzygotic reproductive isolation mechanisms in this group. Our genetic evidence does not, however, rule out the possibility of partial reproductive isolation between lineages. For example, in the western diamondback rattlesnake (*C. atrox*), we previously identified patterns implicating cytonuclear incompatibilities as a potential source of partial reproductive isolation during hybridization of divergent lineages in secondary contact (Schield *et al.*, 2015, 2017). Such incompatibilities are common in nature and reduce the frequency of low-fitness combinations of alleles in hybrid populations (Ortiz-Barrientos *et al.*, 2004; Burton & Barreto, 2012; Sloan *et al.*, 2016).

An intriguing question that stems from our findings is the degree to which hybrid zones in this system are relatively static or in an approximate equilibrium state, or if they remain dynamic (i.e. are continuing to expand and/or contract). Although our results suggest that reproductive isolating mechanisms between lineages are weak, it is possible that local selection within populations inhabiting distinct habitats could stabilize the spread of hybrid zones by combating gene flow into the core ranges of distinct lineages. This hypothesis stems from the remarkable phenotypic and habitat preference diversity among lineages within the species complex and from genetic evidence that many lineages maintain non-admixed populations at the core of their geographical ranges. Maintenance of lineage diversity via selection is not necessarily mutually exclusive with the establishment of partial

reproductive isolation, and it is possible that both positive and negative selection linked to hybrid fitness and local adaptation may maintain lineage-specific variation in CVOS phenotypes and venom composition. It is also possible that gene flow may generate new complements of alleles within hybrid populations that are well suited to intermediate environments at ecotones (Gompert *et al.*, 2006; Mallet, 2007; Abbott *et al.*, 2013), which could also favour a stable equilibrium for hybrid zones. Although the findings of our study raise more questions about the evolutionary history of this group than they answer, our results demonstrate that the CVOS is an enriched system for investigating how genetic drift, gene flow and multiple forms of selection may generate and maintain lineage diversity within species complexes.

THE COMPLEXITY OF SPECIES DELIMITATION IN SPECIES COMPLEXES WITH GENE FLOW

Our results provide new insight into the dynamic evolutionary history of the CVOS complex and help to explain why the taxonomy of this group has remained unresolved for centuries (Rafinesque, 1818; Klauber, 1935, 1936, 1956; Pook *et al.*, 2000; Ashton & de Queiroz, 2001; Douglas *et al.*, 2002; Davis, 2016; Davis *et al.*, 2016). The remarkable phenotypic and habitat preference variation among the numerous distinct lineages of this group (Figs 1B, C, 2A; Klauber, 1956; Campbell & Lamar, 2004; Mackessy, 2010), combined with rampant gene flow among lineages, pose substantial challenges for delimiting and diagnosing species. Beyond the scope of the present study, resolving a stable taxonomy for species complexes with similar evolutionary histories to the CVOS will require confronting how the presence of introgression among distinct lineages should be interpreted in the context of species delimitation.

CONCLUSION

Our analysis of divergence and gene flow across the CVOS complex of rattlesnakes highlights the challenges such systems may pose to our fundamental philosophies for delimiting and defining species, yet also illustrates the value of the system for studying gene flow, speciation and the roles of selection in hybrid zones. Our analyses provide evidence that repeated allopatric divergence and introgression in secondary contact have broadly shaped the evolution of this polytypic species complex. Our study, combined with other recent studies, suggest that this mode of speciation is surprisingly common in rattlesnakes (Meik *et al.*, 2015; Schield *et al.*, 2015, 2017, 2018; Harrington *et al.*, 2018). These results further showcase the surprising weakness of reproductive

isolating mechanisms to inhibit gene flow between lineages, which seems to be a widespread feature of snakes (Fitzpatrick *et al.*, 2008; Wood *et al.*, 2008, 2014; Burbrink & Guiher, 2015; Kindler *et al.*, 2017).

With the recent accumulation of genome-scale datasets from diverse organisms, we are faced with the reality that the evolutionary histories of many systems are poorly characterized by a strictly bifurcating model of divergence and that observed patterns of diversity are surprisingly often a product of both divergence and introgression (Mallet *et al.*, 2007; Abbott *et al.*, 2013; Payseur & Rieseberg, 2016). A major lingering puzzle is how prevalent gene flow is in speciation across the diversity of life, and increases in larger genome-scale datasets will, hopefully, shed light on this question. Such patterns of speciation may be particularly common in young species complexes in temperate regions that, like the CVOS system, were impacted by Pleistocene range contractions and Quaternary range expansions. These revisions to our view of speciation in nature challenge our current views of how to delimit species and illustrate the need to reconcile the practice of how we recognize species with the biological process of speciation, which appears often to incorporate gene flow.

ACKNOWLEDGEMENTS

REFERENCES

Abbott R, **Albach D**, **Ansell S**, **Arntzen JW**, **Baird SJ**, **Bierne N**, **Boughman J**, **Brelsford A**, **Buerkle CA**, **Buggs R**, **Butlin RK**, **Dieckmann U**, **Eroukhmanoff F**, **Grill A**, **Cahan SH**, **Hermansen JS**, **Hewitt G**, **Hudson AG**, **Jiggins C**, **Jones J**, **Keller B**, **Marczewski T**, **Mallet J**, **Martinez-Rodriguez P**, **Möst M**, **Mullen S**, **Nichols R**, **Nolte AW**, **Parisod C**, **Pfennig K**, **Rice AM**, **Ritchie MG**, **Seifert B**, **Smadja CM**, **Stelkens R**, **Szymura JM**, **Väinölä R**, **Wolf JB**, **Zinner D. 2013.** Hybridization and speciation. *Journal of Evolutionary Biology* **26:** 229–246.

Adams RH, **Schield DR**, **Card DC**, **Card DC**, **Corbin A**, **Castoe TA. 2018.** *ThetaMater*: Bayesian estimation of population size parameter θ from genomic data. *Bioinformatics* **34:** 1072–1073.

Aiello-Lammens ME, **Boria RA**, **Radosavljevic A**, **Vilela B**, **Anderson RP. 2015.** spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38:** 541–545.

Alexander DH, **Novembre J**, **Lange K. 2009.** Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19:** 1655–1664.

Arevalo E, **Davis SK**, **Sites JW Jr**. **1994.** Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Biology* **43:** 387–418.

Ashton KG. 2001. Body size variation among mainland populations of the western rattlesnake (*Crotalus viridis*). *Evolution* **55:** 2523–2533.

Ashton KG, **de Queiroz A. 2001.** Molecular systematics of the western rattlesnake, *Crotalus viridis* (Viperidae), with comments on the utility of the D-loop in phylogenetic studies of snakes. *Molecular Phylogenetics and Evolution* **21:** 176–189.

Axelrod DI. 1958. Evolution of the Madro-Tertiary geoflora. *The Botanical Review* **24:** 433–509.

Barton NH, **De Cara MAR. 2009.** The evolution of strong reproductive isolation. *Evolution* **65:** 1171–1190.

Bouckaert R, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

Braconnot P, **Otto-Bliesner B**, **Harrison S**, **Joussaume S**, **Peterchmitt J-Y**, **Abe-Ouchi A**, **Crucifix M**, **Driesschaert E**, **Fichefet Th.**, **Hewitt CD**, **Kageyama M**, **Kitoh A**, **Laîné A**, **Loutre M-F**, **Marti O**, **Merkel U**, **Ramstein G**, **Valdes P**, **Weber SL**, **Yu Y**, **Zhao Y. 2007.** Results of PMIP2 coupled simulations of the Mid-Holocene and last glacial maximum – part 1: experiments and large-scale features. *Climate of the Past* **3:** 261–277.

Bradburd GS, **Coop GM**, **Ralph PL. 2018.** Inferring continuous and discrete population genetic structure across space. *Genetics* **210:** 33–52.

Burbrink FT, **Guiher TJ. 2015.** Considering gene flow when using coalescent methods to delimit lineages of North American pitvipers of the genus *Agkistrodon*. *Zoological Journal of the Linnean Society* **173:** 505–526.

Burnham KP, **Anderson DR. 2003.** *Model selection and multimodel inference: a practical information-theoretic approach.* New York: Springer Science, Business Media.

Burton RS, **Barreto FS. 2012.** A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology* **21:** 4942–4957.

Butlin RK. 1995. Reinforcement: an idea evolving. *Trends in Ecology & Evolution* **10:** 432–434.

Butlin RK, **Smadja CM. 2018.** Coupling, reinforcement, and speciation. *The American Naturalist* **191:** 155–172.

Campbell JA, **Lamar WW. 2004.** *The venomous reptiles of the Western Hemisphere.* Ithaca: Cornell University Press.

Castoe TA, **Spencer CL**, **Parkinson CL. 2007.** Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): a perspective on North American desert biogeography. *Molecular Phylogenetics and Evolution* **42:** 193–212.

Catchen J, **Hohenlohe PA**, **Bassham S**, **Amores A**, **Cresko WA. 2013.** Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22:** 3124–3140.

Coyne JA, **Orr HA. 2004.** *Speciation.* Sunderland: Sinauer Associates.

Crother BI, Boundy J, Burbrink FT, Campbell JA, de Queiroz K, Frost DR, Green DM, Highton R, Iverson JB, Kraus F, McDiarmid RW, Mendelson III JR, Meylan PA, Pyron A, Reeder TW, Seidel ME, Tilley SG, Wake DB. 2012. Squamata - snakes. In: Crother BI, ed. *Scientific and standard English names of amphibians and reptiles in North America North of Mexico, with comments regarding confidence in our understanding*. Lawrence, Kansas: Society for the Study of Amphibians and Reptiles

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R**; 1000 Genomes Project Analysis Group. **2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

Davis MA. 2016. The western rattlesnake complex: 200 years of intrigue and change. In: Schuett GW, Feldner MJ, Smith CF, Reiserer RS, eds. *Rattlesnakes of Arizona.* Rodeo, New Mexico: EcoBooks.

Davis MA, **Douglas MR**, **Collyer ML**, **Douglas ME. 2016.** Deconstructing a species-complex: geometric morphometric and molecular analyses define species in the Western Rattlesnake (*Crotalus viridis*). *PLoS ONE* **11:** e0146166.

Davis MB, **Shaw RG. 2001.** Range shifts and adaptive responses to Quaternary climate change. *Science* **292:** 673–679.

Douglas ME, **Douglas MR**, **Schuett GW**, **Porras LW, Holycross AT. 2002.** Phylogeography of the western rattlesnake (*Crotalus viridis*) complex, with emphasis on the Colorado Plateau. In: Schuett GW, Hoggren M, Douglas ME, Greene HW, eds. *Biology of the vipers.* Eagle Mountain, Utah: Eagle Mountain Publishing.

Douglas ME, **Douglas MR**, **Schuett GW**, **Porras LW. 2006.** Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Molecular Ecology* **15:** 3353–3374.

Drummond AJ, **Rambaut A. 2007.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Fitzpatrick BM, **Placyk JS Jr**, **Niemiller ML**, **Casper GS**, **Burghardt GM. 2008.** Distinctiveness in the face of gene flow: hybridization between specialist and generalist gartersnakes. *Molecular Ecology* **17:** 4107–4117.

Gompert Z, **Fordyce JA**, **Forister ML**, **Shapiro AM**, **Nice CC. 2006.** Homoploid hybrid speciation in an extreme habitat. *Science* **314:** 1923–1925.

Gordon, A. 2010. FASTX-Toolkit. Available at: http://hannonlab.cshl.edu/fastx_toolkit/. Accessed 20 August 2018.

Gosselin T. 2019. *radiator: RADseq Data Exploration, Manipulation and Visualization using R*. doi:10.5281/zenodo.1475182. Available at: https://thierrygosselin.github.io/radiator/.

Graham MR, **Jaeger JR**, **Prendini L**, **Riddle BR. 2013.** Phylogeography of Beck's Desert Scorpion, *Paruroctonus becki*, reveals Pliocene diversification in the Eastern California shear zone and postglacial expansion in the Great Basin Desert. *Molecular Phylogenetics and Evolution* **69:** 502–513.

Gutenkunst RN, **Hernandez RD**, **Williamson SH**, **Bustamante CD. 2009.** Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5:** e1000695.

Harrington SM, **Hollingsworth BD**, **Higham TE**, **Reeder TW. 2018.** Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. *Journal of Biogeography* **45:** 64–75.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 183–195.

Hijmans RJ, **Cameron SE**, **Parra JL**, **Jones PG**, **Jarvis A. 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Hijmans RJ, **Phillips SJ**, **Leathwick J**, **Elith J. 2015.** *dismo: species distribution modeling. R package version 1.0-12*. Available at: http://CRAN.R-project.org/package=dismo. Accessed 20 November 2018.

Holman JA. 2000. *Fossil snakes of North America: origion, evolution, distribution, paleoecology*. Bloomington: Indiana University Press.

Hoskin CJ, **Higgie M**, **McDonald KR**, **Moritz C. 2005.** Reinforcement drives rapid allopatric speciation. *Nature* **437:** 1353–1356.

Jezkova T, **Jaeger JR**, **Oláh-Hemmings V**, **Jones KB**, **Lara-Resendiz RA**, **Mulcahy DG**, **Riddle BR. 2016.** Range and niche shifts in response to past climate change in the desert horned lizard *Phrynosoma platyrhinos*. *Ecography* **39:** 437–448.

Jezkova T, **Riddle BR**, **Card DC**, **Schield DR**, **Eckstut ME**, **Castoe TA. 2015.** Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. *Molecular Ecology* **24:** 83–97.

Johnson JB, **Omland KS. 2004.** Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19:** 101–108.

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

Kindler C, **Chèvre M**, **Ursenbacher S**, **Böhme W**, **Hille A**, **Jablonski D**, **Vamberger M**, **Fritz U. 2017.** Hybridization

patterns in two contact zones of grass snakes reveal a new Central European snake species. *Scientific Reports* **7:** 7378.

Klauber LM. 1935. A new subspecies of *Crotalus confluentus*, the prairie rattlesnake. *Transactions of the San Diego Society of Natural History* **8:** 75–90.

Klauber LM. 1936. A key to the rattlesnakes: with summary of characteristics. *Transactions of the San Diego Society of Natural History* **8:** 185–276.

Klauber LM. 1956. *Rattlesnakes: their habits, life histories, and influence on mankind.* Berkeley: University of California Press.

Lanfear R, **Calcott B**, **Ho SY**, **Guindon S. 2012.** PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29:** 1695–1701.

Lawing AM, **Polly PD. 2011.** Pleistocene climate, phylogeny, and climate envelope models: an integrative approach to better understand species' response to climate change. *PLoS ONE* **6:** e28554.

Leaché AD, **Mulcahy DG. 2007.** Phylogeny, divergence times and species limits of spiny lizards (*Sceloporus magister* species group) in western North American deserts and Baja California. *Molecular Ecology* **16:** 5216–5233.

Legendre P, **Fortin M**, **Borcard D. 2015.** Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6:** 1239–1247.

Li H, **Durbin R. 2009.** Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**; 1000 Genome Project Data Processing Subgroup. **2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Mackessy SP. 2010. Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis sensu lato*): toxicity vs. tenderizers. *Toxicon* **55:** 1463–1474.

Mallet J. 2007. Hybrid speciation. *Nature* **446:** 279–283.

Mallet J, **Beltrán M**, **Neukirchen W**, **Linares M. 2007.** Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7:** 28.

Matute DR. 2010. Reinforcement can overcome gene flow during speciation in *Drosophila*. *Current Biology* **20:** 2229–2233.

Meik JM, **Streicher JW**, **Lawing AM**, **Flores-Villela O**, **Fujita MK. 2015.** Limitations of climatic data for inferring species boundaries: insights from speckled rattlesnakes. *PLoS ONE* **10:** e0131435.

Merow C, **Smith MJ**, **Silander JA. 2013.** A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36:** 1058–1069.

Morafka DJ. 1977. *A biogeographical analysis of the Chihuahuan desert through its herpetofauna.* The Hague: W. Junk.

Mulcahy DG. 2008. Phylogeography and species boundaries of the western North American Nightsnake (*Hypsiglena*

torquata): revisiting the subspecies concept. *Molecular Phylogenetics and Evolution* **46:** 1095–1115.

Muscarella R, **Galante PJ**, **Soley-Guardia M**, **Boria RA**, **Kass JM**, **Uriarte M**, **Anderson RP. 2014.** ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution* **5:** 1198–1205.

Myers EA, **Hickerson MJ**, **Burbrink FT. 2016.** Asynchronous diversification of snakes in the North American warm deserts. *Journal of Biogeography* **44:** 461–474.

Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139:** 1805–1813.

Orr HA, **Presgraves DC. 2000.** Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays* **22:** 1085–1094.

Orr HA, **Turelli M. 2001.** The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55:** 1085–1094.

Ortiz-Barrientos D, **Counterman BA**, **Noor MAF. 2004.** The genetics of speciation by reinforcement. *PLoS Biology* **2:** e416.

Otto-Bliesner BL, **Brady EC**, **Clauzet G**, **Tomas R**, **Levis S**, **Kothavala Z. 2006.** Last Glacial Maximum and Holocene climate in CCSM3. *Journal of Climate* **19:** 2526–2544.

Paterson HEH. 1978. More evidence against speciation by reinforcement. *South African Journal of Science* **74:** 369.

Payseur BA, **Rieseberg LH. 2016.** A genomic perspective on hybridization and speciation. *Molecular Ecology* **25:** 2337–2360.

Peterson AT, **Nyári ÁS. 2008.** Ecological niche conservatism and Pleistocene refugia in the thrush-like Mourner, *Schiffornis* sp., in the neotropics. *Evolution* **62:** 173–183.

Peterson BK, **Weber JN**, **Kay EH**, **Fisher HS**, **Hoekstra HE. 2012.** Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* **7:** e37135.

Phillips SJ, **Anderson RP**, **Schapire RE. 2006.** Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Pickrell JK, **Pritchard JK. 2012.** Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8:** e1002967.

Pook CE, **Wüster W**, **Thorpe RS. 2000.** Historical biogeography of the Western Rattlesnake (Serpentes: Viperidae: *Crotalus viridis*), inferred from mitochondrial DNA sequence information. *Molecular Phylogenetics and Evolution* **15:** 269–282.

Portik DM, **Leaché AD**, **Rivera D**, **Barej MF**, **Burger M**, **Hirschfeld M**, **Rödel M**, **Blackburn DC**, **Fujita MK. 2017.** Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology* **26:** 5245–5263.

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

R Core Team. **2019.** *R: A language and environment for statistical computing*. R foundation for Statistical Computing. Vienna, Austria. Available at: https://www.R-project.org.

Rafinesque CS. 1818. Further account of discoveries in natural history in the western states. *American Monthly Magazine and Critical Review*. **4**: 39–42.

Reyes-Velasco J, **Meik JM**, **Smith EN**, **Castoe TA. 2013.** Phylogenetic relationships of the enigmatic longtailed rattlesnakes (*Crotalus ericsmithi*, *C. lannomi*, and C. *stejnegeri*). *Molecular Phylogenetics and Evolution* **69**: 524–534.

Schield DR, **Adams RH**, **Card DC**, **Perry BW**, **Pasquesi GM**, **Jezkova T**, **Portik DM**, **Andrew AL**, **Spencer CL**, **Sanchez EE**, **Fujita MK**, **Mackessy SP**, **Castoe TA. 2017.** Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and Evolution* **7**: 3951–3966.

Schield DR, **Adams RH**, **Card DC**, **Corbin AB**, **Jezkova T**, **Hales NR**, **Meik JM**, **Perry BW**, **Spencer CL**, **Smith LL**, **García GC**, **Bouzid NM**, **Strickland JL**, **Parkinson CL**, **Borja M**, **Castañeda-Gaytán G**, **Bryson RW Jr**, **Flores-Villela OA**, **Mackessy SP**, **Castoe TA. 2018.** Cryptic genetic diversity, population structure, and gene flow in the Mojave Rattlesnake (*Crotalus scutulatus*). *Molecular Phylogenetics and Evolution* **127**: 669–681.

Schield DR, **Card DC**, **Adams RH**, **Jezkova T**, **Reyes-Velasco J**, **Proctor FN**, **Spencer CL**, **Herrmann HW**, **Mackessy SP**, **Castoe TA. 2015.** Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution* **83**: 213–223.

Sloan DB, **Havird JC**, **Sharbrough J. 2016.** The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* **26**: 2212–2236.

Strickland JL, **Smith CF**, **Mason AJ**, **Schield DR**, **Borja M**, **Castañeda-Gaytán G**, **Spencer CL**, **Smith LL**, **Trápaga A**, **Bouzid NM**, **Campillo-García G**, **Flores-Villela OA**, **Antonio-Rangel D**, **Mackessy SP**, **Castoe TA**, **Rokyta DR**, **Parkinson CL. 2018.** Evidence for divergent patterns of local selection driving venom variation in Mojave Rattlesnakes (*Crotalus scutulatus*). *Scientific Reports* **8**: 17622.

Sugiyama M, **Shiogama H**, **Emori S. 2010.** Precipitation extreme changes exceeding moisture content increases in MIROC and IPCC climate models. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 571–575.

Wood DA, **Fisher RN**, **Reeder TW. 2008.** Novel patterns of historical isolation, dispersal, and secondary contact across Baja California in the Rosy Boa (*Lichanura trivirgata*). *Molecular Phylogenetics and Evolution* **46**: 484–502.

Wood DA, **Fisher RN**, **Vandergast AG. 2014.** Fuzzy boundaries: color and gene flow patterns among parapatric lineages of the western shovel-nosed snake and taxonomic implication. *PLoS ONE* **9**: e97494.

Wu CI. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851–865.

Wu CI, **Ting CT. 2004.** Genes and speciation. *Nature Reviews Genetics* **5**: 114–122.

Zheng Y, **Wiens JJ. 2016.** Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* **94**: 537–547.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Specimen data for samples used in this study. Where noted, samples were used previously by Schield *et al.* (2018).

Figure S1. Distribution of *Crotalus scutulatus* and localities of sampled individuals for analysis in the present study. Differently coloured circles depict individuals belonging to four *C. scutulatus* groups identified by Schield *et al.* (2018). Major phylogeographical breaks identified in that study are depicted as dashed lines.

Figure S2. Schematics of two-population and three-population demographic models tested in the present study. The 'no divergence' two-population model is not shown, because it does not involve any unscaled parameters. Unscaled parameter estimates per model are shown, as follows: nμA, ancestral population effective size; nμ1, population 1 effective size; nμ2, population 2 effective size; nμ3, population 3 effective size; 2NmA, ancestral migration; 2Nm1, migration between populations 1 and 2; 2Nm2, migration between populations 2 and 3; 2Nm12, migration from population 2 to 1; 2Nm21, migration from population 1 to 2; T1, time since population split; T2, time since secondary contact, population split or isolation interval; T3, time since secondary contact or isolation interval. All models were modified and redrawn from those used by Portik *et al.* (2017).

Figure S3. Plot of variance explained in our nuclear RADseq dataset with increasing migration (*m*) events. The inferred population graph in the absence of migration explained ~96% of the variance, and additional migration events explained increasing amounts of variation. The amount of variation plateaued at $m = 8$ and asymptotically approached 100% with increasing values of *m*. The dashed line depicts the variance explained at $m = 8$.

Figure S4. Results of Bayesian mitochondrial gene tree and divergence time analysis, showing all outgroup taxa. Nodes with > 0.95 posterior support are shown with a black dot. Posterior probabilities of nodes with less support

are labelled on the right of respective nodes. Median divergence time estimates are labelled on the right of nodes. The 95% highest posterior density for divergence times are shown as green bars.

Figure S5. Cross-validation error values from ADMIXTURE analyses across models with increasing values of *K* genetic clusters.

Figure S6. Results of spatially aware analysis of population structure in conStruct. A, cross-validation measures of model predictive accuracy for *K* of 1–12 plateau around *K* = 3. B, layer contributions shown for *K* = 1–8 show that additional layers beyond *K* = 3 contribute little to explaining covariance in the data. C, admixture proportions based on a *K* = 3 spatial model for all individuals included in conStruct analyses. D, map of admixture proportions placed at the individual localities, with approximated ranges of *Crotalus oreganus* (yellow), *Crotalus viridis* (green) and *Crotalus scutulatus* (grey) overlaid.

Figure S7. Alternative demographic models for *Crotalus oreganus helleri* and *Crotalus scutulatus scutulatus* (A, B) and for *C. s. scutulatus* and *Crotalus viridis viridis* (C, D) that had similar support after final optimizations in $\delta a \delta i$. A, C, models of divergence with asymmetric gene flow in secondary contact, with population size change. B, D, models of divergence with continuous asymmetric gene flow, with population size change. The schematics of each model are shown on the left, with unscaled parameter estimates. Two-dimensional allele frequency spectra from the empirical data and simulated model data are shown in the centre; brighter colours represent higher-frequency alleles. Heatmaps and histograms of residuals between empirical and simulated distributions are shown on the right. For additional details, see the Supporting Information (Dataset S1).

Figure S8. A, B, best-fitting demographic models for three-population analyses of [(*Crotalus scutulatus*, (*Crotalus oreganus helleri*, *Crotalus oreganus oreganus*)] (A) and [(*C. scutulatus*, (*Crotalus viridis viridis*, *Crotalus oreganus concolor*)] (B) based on the diffusion approximation model selection approach in $\delta a \delta i$. Unscaled parameters for each model are labelled, and the relative ancestral population sizes (NμA) compared with inferred current effective population sizes are represented by larger labels. For additional details, see the Supporting Information (Dataset S2).

Figure S9. The relationship between pairwise geographical and genetic distances between *Crotalus viridis nuntius* and *Crotalus viridis viridis*. A Mantel test yielded significant evidence of divergence attributable to isolation by distance. The black line shows the slope of the positive relationship between geographical and genetic distance, and black dots show independent pairwise distances across samples. The shaded region of the plot shows the density of pairwise distances, with greater density represented by red and yellow, and lower density represented by blues.

Figure S10. Results of ecological niche modelling and projections of suitable habitat for western rattlesnake species complex (CVOS) lineages onto current and Last Glacial Maximum conditions, and a comparison of results using the community climate system model (CCSM) and model for interdisciplinary research on climate (MIROC) models. In each panel, the logistic probability of habitat suitability is shown in shaded regions, with green representing the highest probability of suitable habitat, and yellow and red depicting regions of lower probability. A, *Crotalus cerberus*. B, *Crotalus oreganus concolor*. C, *Crotalus oreganus helleri*. D, *Crotalus oreganus lutosus + abyssus*. E, *Crotalus oreganus oreganus*. F, *Crotalus viridis nuntius*. G, *Crotalus viridis viridis*. H, *Crotalus scutulatus + salvini*.

Dataset S1. Results of demographic model testing of pairwise lineages using the folded site frequency spectrum in $\delta a \delta i$. Each analysis included two lineages that share a range boundary. The best-supported model is highlighted in bold for each analysis. All parameters are unscaled and thus represent only relative values and are not directly comparable across analyses. Abbreviations: AIC, Akaike information criterion; Akweight, Akaike weight; nμ1a, initial effective population size of population 1; nμ2a, initial effective population size of population 2; nμ1b, present effective population size of population 1; nμ2b, present effective population size of population 2; 2Nm12, migration rate from population 2 to population 1; 2Nm21, migration rate from population 1 to population 2; param_set, number of parameters in model; RL, relative likelihood; θ, the effective mutation rate of the reference population (4NrefμL); T1, scaled time between population split and the present or T2, the scaled time of secondary contact or isolation interval.

Dataset S2. Results of demographic model testing of three populations using diffusion approximation of the site frequency spectrum in $\delta a \delta i$. The best-supported model is highlighted in bold for each analysis. All parameters are unscaled and thus represent only relative values and are not directly comparable across analyses. Abbreviations are as follows: AIC, Akaike information criterion; Akweight, Akaike weight; nμ1, effective population size of population 1; nμA, effective ancestral population size, nμ2, effective population size of population 2; nμ3, effective population size of population 3; 2Nm1, migration rate between population 1 and 2; 2Nm2, migration rate between population 2 and 3; 2NmA, ancestral migration between population 1 and the ancestor of populations 2 and 3; param_set, number of parameters in model; RL, relative likelihood; θ, the effective mutation rate of the reference population (4NrefμL); T1, scaled time between population split and the present or T2, the scaled time of secondary contact or isolation interval.