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

DREW R. SCHIELD1,1, BLAIR W. PERRY1,3, RICHARD H. ADAMS1, DAREN C. CARD1,2,3, TEREZA JEZKOVA4, GIULIA I. M. PASQUESI1, ZACHARY L. NIKOLAKIS1, KRISTOPHER ROW1, JESSE M. MEIK5, CARA F. SMITH6, STEPHEN P. MACKESSY6 and TODD A. CASTOE1,*

1Department of Biology, 501 South Nedderman Drive, University of Texas at Arlington, Arlington, TX 76019, USA
2Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
3Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA
4Department of Biology, Miami University of Ohio, 700 East High Street, Oxford, OH 45056, USA
5Department of Biological Sciences, Tarleton State University, 1333 West Washington Street, Stephenville, TX 76402, USA
6School of Biological Sciences, 501 20th Street, University of Northern Colorado, Greeley, CO 80639, USA

Received 15 February 2019; revised 1 May 2019; accepted for publication 1 May 2019

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...
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 diversity, and species boundaries. We test for evidence of allopatric divergence and secondary contact by comparing patterns of gene flow in these species.

**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 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; Mr, Novex Mark 12 molecular weight standards.
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 Huamantlan 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 SbfI and Sau3AI 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, 2009) and Samtools (http://samtools.sourceforge.net) to determine the number of reads mapping to each PCR clone and thereby reconstructing diploid genome representations of individual samples. 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 SbfI and Sau3AI 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, 2009) and Samtools (http://samtools.sourceforge.net) to determine the number of reads mapping to each PCR clone and thereby reconstructing diploid genome representations of individual samples.
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 ≥ single nucleotide polymorphisms (SNPs) with a minor filter function. We filtered variants to retain biallelic fewer than five as missing data using the BCFTools variant calls within individuals with a read depth.

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 fluorometer (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 x 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 approach implemented 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_eμ; effective population size = N_eμ) 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 ThetaMaterM1 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 θα0i (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 θα0i 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 θα0i manual). This simulated a 2D-SFS for each set of parameters,
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.byssus 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).

### 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

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

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).
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 inter disciplinary 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 or 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 Ma; Fig. 3A; Supporting Information, Fig. S4). Lineages within the CVOS diverged from a common ancestor between the Late Miocene and early Pliocene (~5–7 Ma; Fig. 3A). Other CVOS subspecies were supported as being nested within these clades: C. v. nuntius 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**

ADIMIXTURE 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 and C. v. nuntius).

**Nuclear estimates of demographic history and patterns of gene flow**

Analyses of the 2D-SFS in ḍabcı 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 assignment 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.
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 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).

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...
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

<table>
<thead>
<tr>
<th>Divergence Time</th>
<th>Relative Effective Pop. Size</th>
<th>Relative Migration (2Nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.5</td>
<td>Larger</td>
<td>Greater</td>
</tr>
<tr>
<td>15.0</td>
<td>Smaller</td>
<td>Greater</td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td>Greater, Symmetrical</td>
</tr>
<tr>
<td>10.0</td>
<td></td>
<td>Greater, Symmetrical</td>
</tr>
<tr>
<td>7.5</td>
<td></td>
<td>Greater, Symmetrical</td>
</tr>
<tr>
<td>5.0</td>
<td></td>
<td>Greater, Symmetrical</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>Greater, Symmetrical</td>
</tr>
</tbody>
</table>

Figure 5. Results of demographic model testing in $\delta_\alpha$ $\delta_i$ 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 ($n_1^a$, $n_2^a$, $n_1^b$, $n_2^b$, $m_1$, $m_2$) 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).
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. c. cerberus, C. s. 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 Ma.

<table>
<thead>
<tr>
<th>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</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Full (geographical distance + environment)</td>
</tr>
<tr>
<td>Environment</td>
</tr>
<tr>
<td>Geographical distance</td>
</tr>
</tbody>
</table>

The ‘full’ model, which accounted for the combined effects of geographical and environmental distances, explained the greatest variance in the genetic distance response variable.

© 2019 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2019, **XX**, 1–21
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 \( \delta \alpha \delta \) (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. a. 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; Burbink & Guicher, 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 introgression (Mallet et al., 2007; Abbott et al., 2013; Payseur & Rieseberg, 2016). A major lingering puzzle is how prevalent gene flow is in species 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
We thank Jens Vindum and the California Academy of Sciences, Robert Murphy and the Royal Ontario Museum, Gordon Schuett, Matthew Holding, Corey Roelke, Andrew Holycross, Hans-Werner Herrmann, James McCabe, Joshua Parker and Stephen Spear for tissue loans. We are grateful for the suggestions and constructive feedback from two anonymous reviewers. D.R.S. and B.W.P. contributed equally to this work. Support for this work was provided by National Science Foundation grants DEB-1655571 to T.A.C. and S.P.M., DEB-1501886 to D.R.S. and T.A.C., and DEB-1501747 to D.C.C. and T.A.C.

REFERENCES


© 2019 The Linnean Society of London, Biological Journal of the Linnean Society, 2019, XX, 1–21


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


Results of Bayesian mitochondrial gene tree and divergence time analysis, showing all outgroup taxa. The dashed line depicts the variance explained at $m = 8$. 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 $m = 8$.

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: $n_1\alpha$, ancestral population effective size; $n_1\beta$, population 1 effective size; $n_2\alpha$, population 2 effective size; $n_2\beta$, population 3 effective size; $2Nm\alpha$, ancestral migration; $2Nm\beta$, migration between populations 1 and 2; $2Nm\gamma$, migration between populations 2 and 3; $2Nm\delta$, migration from population 2 to 1; $2Nm\epsilon$, migration from population 1 to 2; $T_1$, time since population split; $T_2$, time since secondary contact; $T_3$, 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 $m = 8$. 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 (green), 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 $\delta\alpha\delta$. 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 $\delta\alpha\delta$. Unscaled parameters for each model are labelled, and the relative ancestral population sizes ($N\mu A$) 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 (CASM) 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 oreganus. D, Crotalus oreganus lutosus + abyssus. E, Crotalus oreganus oreganus. F, Crotalus viridis nuntius. G, Crotalus viridis viridis. H, Crotalus scutulatus + salvinii.

**Dataset S1.** Results of demographic model testing of pairwise lineages using the folded site frequency spectrum in $\delta\alpha\delta$. 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; $\mu 1a$, initial effective population size of population 1; $\mu 2a$, initial effective population size of population 2; $\mu 1b$, present effective population size of population 1; $\mu 2b$, 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; 0, the effective mutation rate of the reference population ($4N\mu_f L$); 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 $\delta\alpha\delta$. 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; $\mu 1$, effective population size of population 1; $\mu A$, effective ancestral population size, $\mu 2$, effective population size of population 2; $\mu 3$, 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; 0, the effective mutation rate of the reference population ($4N\mu_f L$); T1, scaled time between population split and the present or T2, the scaled time of secondary contact or isolation interval.

© 2019 The Linnean Society of London, Biological Journal of the Linnean Society, 2019, XX, 1–21