Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (Crotalus ruber) in Baja California

Sean M. Harrington1 | Bradford D. Hollingsworth2 | Timothy E. Higham3 | Tod W. Reeder1

1Department of Biology, San Diego State University, San Diego, CA, USA
2San Diego Natural History Museum, San Diego, CA, USA
3Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA, USA

Correspondence
Sean M. Harrington, Department of Biology, San Diego State University, San Diego, CA, USA.
Email: seanharrington256@gmail.com

Abstract

Aim: Many studies have investigated the phylogeographic history of species on the Baja California Peninsula, and they often show one or more genetic breaks that are spatially concordant among many taxa. These phylogeographic breaks are commonly attributed to vicariance as a result of geological or climatic changes, followed by secondary contact when barriers are no longer present. We use restriction-site associated DNA sequence data and a phylogeographic model selection approach to explicitly test the secondary contact hypothesis in the red diamond rattlesnake, Crotalus ruber.

Location: Baja California and Southern California.

Methods: We used phylogenetic and population clustering approaches to identify population structure. We then used coalescent methods to simultaneously estimate population parameters and test the fit of phylogeographic models to the data. We used ecological niche models to infer suitable habitat for C. ruber at the Last Glacial Maximum (LGM).

Results: Crotalus ruber is composed of distinct northern and southern populations with a boundary near the town of Loreto in Baja California Sur. A model of isolation followed by secondary contact provides the best fit to the data, with both divergence and contact occurring in the Pleistocene. We also identify a genomic signature of northern range expansion in the northern population, consistent with LGM niche models showing that the northern-most portion of the range of C. ruber was not suitable habitat during the LGM.

Main conclusions: We provide the first explicitly model-based test of the secondary contact model in Baja California and show that populations of C. ruber were isolated before coming back into contact near Loreto, a region that shows phylogeographic breaks for other taxa. Given the timing of divergence and contact, we suggest that climatic fluctuations have driven the observed phylogeographic structure observed in C. ruber and that they may have driven similar patterns in other taxa.

KEYWORDS
Baja California, model selection, phylogeography, population genomics, population structure, range expansion
Baja California has been the focus of many phylogenetic and phylogeographic studies due to its unique geologic history and high endemism (Dolby, Bennett, Lira-Noriega, Wilder, & Munguía-Vega, 2015; Grismer, 2002; Lindell, Méndez-De La Cruz, & Murphy, 2008; Riddle, Hafner, Alexander, & Jaeger, 2000; Savage, 1960; Zink, 2002). Many of the taxa that have been examined have shown spatially concordant phylogenetic or phylogeographic breaks at one, two or three key regions in Baja California: the mid-peninsula, near the town of Loreto and at the Isthmus of La Paz (Figure 1; Riddle, Hafner, Alexander, & Jaeger, 2000). These breaks have been found in diverse sets of taxa, including, but not limited to, spiders (Crews & Hedin, 2006), reptiles (Lindell, Méndez-de la Cruz, & Murphy, 2005; Lindell et al., 2008; Upton & Murphy, 1997), birds (Zink, Kessen, Line, & Blackwell-Rago, 2001), cacti (Nason, Hamrick, & Fleming, 2002) and mammals (Riddle, Hafner, & Alexander, 2000; Whorley, Alvarez-Castañeda, & Kenagy, 2004). A common explanation for the mid-peninsular and La Paz splits, which are more common than the Loreto break, are seaways proposed to have crossed the peninsula and since receded (Aguirre, Morafka, & Murphy, 1999; Murphy, 1983). However, this hypothesis has been weakened by a general lack of direct geological evidence and the finding that mid-peninsular divergences indicate at least two episodes of divergence (Crews & Hedin, 2006; Dolby et al., 2015; Leaché, Crews, & Hickerson, 2007). No strong evidence has been presented to suggest that a seaway crossed the peninsula in the vicinity of Loreto, leaving the drivers of spatially concordant phylogenetic and phylogeographic breaks unclear at all of these regions.

The phylogeographic patterns observed could be produced by at least two different processes: (1) isolation with ongoing migration between diverging populations (the IM, isolation migration model), (2) vicariance followed by secondary contact (the secondary contact model), as well as more complex scenarios. Despite the absence of strong evidence for barriers to dispersal, the secondary contact model has often been invoked to explain patterns of genetic structure (e.g., Crews & Hedin, 2006; Lindell et al., 2008). However, tools to statistically compare IM and secondary contact models have only recently become available (e.g., Beaumont, 2010; Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Jackson, Morales, Carstens, & O’Meara, 2017), and these models have not yet been explicitly tested in Baja California taxa.

Here, we sought to examine the phylogeography of mainland populations of Crotalus ruber (the red diamond rattlesnake), a large rattlesnake that ranges throughout Baja California and into southern California (Figure 1), using genomic data and explicit testing of alternative phylogeographic models. We also investigated the effect that past climatic fluctuations may have had on populations of C. ruber using an ecological niche modelling approach to determine if geographic shifts in suitable habitat could explain phylogeographic patterns. Mainland populations of C. ruber were previously recognized as two subspecies on the basis of morphological characters including scale counts and coloration, with C. ruber ruber in the north and C. ruber lucasensis in the south, and with the boundary between these subspecies occurring somewhere in the vicinity of Loreto (Klauber, 1949; Van Denburgh, 1920). The subspecies were subsequently synonymized due to the presence of individuals with intermediate phenotypes at the boundary near Loreto (Grismer, McGuire, & Hollingsworth, 1994; Klauber, 1949). Despite the presence of morphological variation between northern and southern populations of C. ruber, a previous study that included individuals sampled from southern California, northernmost Baja California, and one sample from southernmost Baja California Sur showed very low mitochondrial DNA (mtDNA) divergence across the peninsula (0.3%, Douglas, Douglas, Schuett, & Porras, 2006). Similarly, Murphy et al. (1995) found very low mtDNA divergence between individuals sampled from Cedros Island, Baja California and from Riverside, CA. Although C. ruber shows low mtDNA divergence across the peninsula, many other squamate reptiles (lizards and snakes) show strong phylogenetic or phylogeographic mtDNA and/or nuclear breaks (e.g., Leaché & McGuire, 2006; Lindell et al., 2005, 2008; Melk, Streicher, Lawing, Flores-Villela, & Fujita, 2015; Mulcahy, 2008; Upton & Murphy, 1997). We used a restriction site-associated DNA sequencing

**FIGURE 1** Map showing the Baja California Peninsula with positions of the mid-peninsula, Loreto, and La Paz breaks found across many taxa shown. The range of *Crotalus ruber* is outlined in black. Black circles show sampling localities of sequenced individuals.
(RADseq) approach to determine if the increased power of a large genome-wide dataset would be able to identify phylogeographic breaks even in the face of the low observed mtDNA divergence. We used these data to evaluate two primary hypotheses: (1) C. ruber will show a phylogeographic break at one or more of the mid-peninsula, Loreto, or La Paz regions, as found in other taxa; and (2) observed phylogeographic breaks will be the result of isolation followed by secondary contact, rather than isolation with migration or strict isolation.

2 | MATERIALS AND METHODS

2.1 | Sequencing and bioinformatics processing

We sequenced 35 individuals of C. ruber from across the range of the species (sampling localities shown in Figure 1) and one individual each from the outgroup taxa Crotalus atrox, Crotalus horridus, Crotalus cerastes, Crotalus scutulatus and Crotalus molossus. Samples used in this study were obtained from specimens in the collections of the Universidad Autónoma de Baja California in Ensenada, Centro de Investigaciones Biológicas del Noroeste, Universidad Nacional Autónoma de México, San Diego Natural History Museum, and San Diego State University collections (Appendix S1). In the text, we refer to all samples by field collection numbers, and refer readers to Appendix S1 for the corresponding specimen numbers. We followed the double digest RADseq protocol of Peterson, Weber, Kay, Fisher, and Hoekstra (2012) with modifications following Gottscho et al. (2017), with the exception that we used the enzymes SbfI and Sau3AI as in Schield et al. (2015). We sequenced the final libraries (100 bp single-end reads) on one half flow-cell lane of an Illumina HiSeq 2500 at the University of California, Riverside Institute of Integrative Genome Biology. We used the PrRAD 3.0.5 pipeline (Eaton, 2014) for data processing using a clustering threshold of 0.85 and requiring at least 10x coverage with 10 or fewer Ns. We included outgroups in phylogenetic analyses for rooting purposes but not population clustering and demographic analyses; thus, different outgroup taxa were excluded or included in different PrRAD runs, resulting in datasets with differing numbers of taxa that were used for different analyses. We also used different thresholds for the number of individuals that must have data for a given locus for that locus to be retained in the dataset due to differences in the ability of analyses to accommodate missing data. One individual that recovered low quality sequence was removed from all datasets during PrRAD processing (SD 506). The presence-absence of outgroups and minimum number of individuals that a locus was required to be present in are shown in Table S1.

2.2 | Concatenated phylogenetic analysis

We estimated the relationships among individuals on datasets including all outgroup taxa using RAxML 8.2.4 (Stamatakis, 2006, 2014) on the CIPRES Science Gateway (Miller, Pfeiffer, & Schwartz, 2010). We acknowledge that phylogenetic analyses that force a bifurcating tree such as RAxML are not the most appropriate for intraspecific data, but they provide a useful heuristic to see how individuals cluster, even if gene flow likely makes many of the inferred relationships uninformative. To test that excluding missing data did not have a large effect on our analyses (see Huang & Knowles, 2016), we ran RAxML on two datasets differing only in the number of individuals required to retain each locus in PrRAD processing, using thresholds of either 10 or 20 individuals. We did not partition our data and performed rapid bootstrapping (using automatic stopping under the autoMRE criterion) and a search for the best tree under the GTRCAT model with final optimization of trees using GTR + I. We also inferred the relationships among ingroup individuals using the program SplitsTree 4.14.2 (Huson & Bryant, 2006), which infers phylogenetic networks, and therefore does not force a strict bifurcating topology. We used the Jukes Cantor model (Jukes & Cantor, 1969) when calculating distances and used the NeighborNet algorithm for constructing the network.

2.3 | Population clustering and isolation by distance

We used the model-based Bayesian clustering method STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) and its maximum likelihood analog ADMIXTURE 1.23 (Alexander, Novembre, & Lange, 2009) to assign individuals to populations and detect admixed individuals. We ran STRUCTURE with correlated allele frequencies for 300k post-burn-in generations with 300k generations of burn-in. For STRUCTURE, the optimal number of populations (K) was determined by running STRUCTURE with K = 1–5 for 15 replicates each and using the Evanno method (Evanno, Regnaut, & Goudet, 2005) in STRUCTURE HARVESTER 0.6.94 (Earl & vonHoldt, 2012). The optimal number of populations in ADMIXTURE was determined using the cross-validation approach (Alexander & Lange, 2011), testing K = 1–5, as in STRUCTURE. We also explored population structure using the non-model-based methods spatial principal components analysis (sPCA) and discriminant analysis of principal components (DAPC) in the R package “adegenet” 2.0.1 (Jombart, 2008; Jombart & Ahmed, 2011) in R 3.3.2 (R Core Team, 2016). We used k-means clustering and the Bayesian information criterion to determine the optimum number of clusters for DAPC, and retained all principal components. R code for performing these analyses is deposited on Dryad (doi:10.5061/dryad.7hsOn). To test the effect of dataset completeness on population clustering, we performed ADMIXTURE analyses using two datasets of differing completeness that required a locus to be present in either 26 or 17 individuals to be retained. All population clustering analyses were highly concordant, and so for all subsequent analyses that required assigning individuals to populations a priori, we based these assignments on the most probable population assignment from ADMIXTURE using the more complete dataset.

To ensure that our population clustering analyses were not erroneously interpreting isolation by distance as population structure, we ran a Mantel test on genetic and geographic distances among individuals using the mantel.ran.test function in the “Adegenet” package. Significance of the Mantel test was assessed using 99,999 permutations. We plotted genetic and geographic distances to
visualize if isolation by distance exists as a continuous cline or is the result of patches of distant, divergent individuals. Plots were coloured by point density as measured by 2-dimensional kernel density estimation using the kde2d function in the R package "MASS" 7.3-47 (Venables & Ripley, 2002). We also ran a Mantel test and plotted genetic against geographic distance for the northern population to test if patterns of isolation by distance within this population are consistent with a northern population expansion suggested by other analyses.

2.4 Coalescent phylogenetic analysis and population modelling

We utilized three coalescent-based approaches to estimate divergence times and migration rates between populations of *C. ruber* identified by our population clustering analyses. For all of these analyses, we assigned individuals to populations based on the results from STRUCTURE and ADMIXTURE (which are highly concordant with each other and other analyses), with admixed individuals assigned to whichever population makes up the majority of their ancestry.

We used the program FASTSIMCOAL2 2.5.2.21 (FSC2; Excoffier et al., 2013) to perform phylogeographic model selection and estimate the parameters of each model. FSC2 is a coalescent-based method that takes site frequency spectra as input and uses a simulation approach to approximate the likelihood of any arbitrarily complex phylogeographic model that can be specified and estimate the demographic parameters specified in each model. We generated joint site frequency spectra from each of 10 downsampled SNP datasets using code developed by Jordan Satler (https://github.com/jordansatler/SNPtoAFS) following Thomé and Carstens (2016) to account for the effects of missing data using a missing data threshold of 50%. We compared a set of nine phylogeographic models, summarized in Figure 2. For each model, we calculated the Akaike information criterion (AIC) from the approximated likelihood and number of parameters in order to rank models. For folded SFS, FSC2 requires that the mutation rate is specified. We assumed a genome-wide mutation rate of $2.2 \times 10^{-9}$ mutations/site/year based on similar rates calculated across a large pool of mammalian loci (Kumar & Subramanian, 2002) and a few lizard loci (Gottscho, Marks, & Jennings, 2014). We used a generation time of 3.3 years/generation based on the estimated generation time of the sister species of *C. ruber*, *C. atrox* (Castoe, Spencer, & Parkinson, 2007). Combining these, we get a mutation rate of $7.26 \times 10^{-9}$ mutations/site/generation. We acknowledge that use of a mutation rate from a more closely related taxon would be preferable, but we do not expect our main conclusions to change unless the true mutation rate is very different from what we have used, and we address this in our discussion section. We ran 100 replicate FSC2 analyses under each model.
on each of the 10 downsampled datasets to ensure that we found
the optimum likelihood. Each individual FSC2 run included 100,000
simulations for estimation of the composite likelihood and 10 ECM
cycles for parameter optimization. For each downsampled dataset,
we retained only the FSC2 run that obtained the highest likelihood.
We report parameter estimates and AIC scores as averages across
the 10 downsampled datasets.

As an alternative method to estimate demographic parameters,
we used the program G-PHOCS 1.2.2 (Gronau, Hubisz, Gulko, Danko,
& Siepel, 2011) to estimate divergence time, population size, and
migration rates between populations of C. ruber under an IM model.
We ran two replicate G-PHOCS runs using automatic fine-tuning for
a total of 300k generations each, sampling every 50 generations,
and discarding the first 10% of samples as burn-in. We allowed
asymmetric migration between the northern and southern populations
and set the two populations to coalesce into a single ancestral
population in the past. We set \( \alpha = 1.0 \) and \( \beta = 10,000 \)
for the gamma distribution used for priors on \( \tau \) and \( \Theta \) parameters,
and \( \alpha = 0.002 \) and \( \beta = 0.00001 \) for the gamma distribution used for priors
on migration rates. We checked for convergence between runs
using TRACER 1.6 (Rambaut, Suchard, Xie, & Drummond, 2014).
To convert divergence times we used the program SNAPP 1.2.5 (Bryant,
Bockaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) imple-
mented in BEAST 2.3.1 (Bouckaert et al., 2014), which models a bifu-
crating tree topology in the absence of gene flow, to estimate the
divergence time between populations. SNAPP analyses included one
individual of C. atrox as an outgroup, such that divergence times ini-
tially estimated in substitutions/site could be roughly converted to
absolute time on the basis of assuming a split between C. ruber and
C. atrox c. 3 Ma based on previous fossil-calibrated phylogenies
(Blair & Sánchez-Ramírez, 2016; Reyes-Velasco, Meik, Smith, & Car-
toe, 2013). SNAPP does not currently support node calibrations, and
so we estimated the divergence between C. ruber populations as the
ratio of the height of the node uniting C. ruber populations and the
height of the node uniting C. ruber and C. atrox multiplied by the
assumed 3 Ma divergence between C. atrox and C. ruber. This is an
admittedly coarse procedure, but we are only seeking to determine
degree similarity between estimates from FSC2, G-PHOCS, and
SNAPP. Because this procedure does not rely on the mutation rate
assumed for FSC2 and G-PHOCS analyses, it helps provides indepen-
dent validation of the dates produced using this rate. Although the
estimated divergence is likely to be inaccurate if there is gene flow
among populations, radically different divergence times from SNAPP
and demographic modelling approaches could indicate problems
with our assumed mutation rate. Two replicate SNAPP analyses were run
for a total of 300k generations with 10% of generations discarded
as burn-in. Convergence was assessed using TRACER and trees from
both runs were combined and summarized using the LOGCOMBINER
and TREEANNOTATOR programs distributed with BEAST2, respectively.

### 2.5 Niche modelling

We used MAXENT 3.3.3k (Phillips, Anderson, & Schapire, 2006) to
estimate current and Last Glacial Maximum (LGM) niches for C. ruber
to determine if phylogeographic patterns could be related to changes
in suitable habitat caused by Pleistocene climatic fluctuations. GPS
coordinates were obtained from the VertNet database. Localities
outside the range of C. ruber were removed and localities were
down-sampled so that only localities at least ~20 km apart were
included to alleviate issues associated with highly uneven sampling
across the range of C. ruber, retaining a total of 101 locality points
(southern California is sampled far more densely than most of Baja
California). We used a total of 11 climatic variables from the BioClim
dataset (Bio 2, 3, 5,7–9, 15–19; Hijmans, Cameron, Parra, Jones, & Car
and Jarvis, 2005) after removing variables that are highly correlated
(\( -0.8 \geq r \geq 0.8 \)). We used two models of LGM climate, the Com-
unity Climate System Model (CCSM4) and the Model for Interdisci-
plinary Research on Climate (MIROC), to estimate the distribution of
C. ruber at the LGM. To limit over-prediction in areas well outside
the range of C. ruber, we clipped all environmental layers to a
300 km buffer around the minimum convex polygon containing all
occurrence points used for niche modelling. MAXENT analyses were
run using default setting and models were evaluated using the area
under the curve (AUC) statistic.

### 3 RESULTS

#### 3.1 Sequence data

From our Illumina sequencing, we obtained a total of \( 20 \) M reads
across all samples (including outgroups). After PrRAD processing, the
number of loci retained ranged from 998 for the dataset including C.
atrox as an outgroup and requiring each locus to be present in at
least 31 individuals to 2,878 for the dataset including all outgroup taxa
and requiring each locus to be present in at least 20 individuals.
The number of parsimony informative sites per dataset ranged from
717 (with 629 unlinked SNPs) when all outgroups were excluded and
loci were required to be present in at least 30 individuals to 3,834
(with 2,634 unlinked SNPs) for the dataset including all outgroup taxa
requiring each locus to be present in at least 20 individuals.

#### 3.2 Concatenated phylogenetic analysis

Results from RAxML analyses on the two data matrices that we
tested were highly similar, with the exception that support was
lower for many nodes in the analysis of the more incomplete dataset
(Figure 3, Figure S1). We therefore focus only on the RAxML results
using the more complete dataset. The relationships among individu-
als inferred by RAxML revealed a basal split between a northern and
southern clade (Figure 3). Most of the individuals are strongly
supported as members of the northern or southern clade, with the exception of the southern-most individual of the northern clade. The genetic break between the northern and southern clades of individuals is geographically located near the town of Loreto in Baja California Sur. Within the northern clade, the northern-most individuals are the most highly nested, with more southern individuals in the clade recovered as less nested and branching earlier, consistent with a pattern of northern range expansion. Results from SPLITSTREE are largely concordant with RAXML results and also show groupings of northern and southern individuals, corresponding to individuals north and south of Loreto, respectively (Figure S2). Several individuals that are near Loreto are also placed in intermediate positions on the phylogenetic network, suggesting that these individuals are genetically admixed between the northern and southern groups.

3.3 Population clustering and isolation by distance

STRUCTURE and ADMIXTURE both indicate that a two-population model provides the best fit to the data (Figure S3, Table S2). Population assignments of individuals by STRUCTURE and ADMIXTURE are nearly identical, as are the results of ADMIXTURE analyses using datasets of differing completeness, so we discuss only the results from Admixture, with the results from STRUCTURE and ADMIXTURE with the more incomplete dataset shown in Figures S4 and S5, respectively. Admixture identified northern and southern populations of C. ruber, with a boundary between these populations at Loreto (Figure 3), concordant with the groupings identified by RAxML and SPLITSTREE. As suggested by SPLITSTREE, the individuals nearest the population boundary are recovered as admixed. Admixture of southern ancestry in northern individuals is detectable much farther from the population boundary than is northern ancestry in southern individuals. We examined the results of Admixture analyses under a three-population model to determine if any other common biogeographic/phylogeographic breaks may be present, and found that this analysis recovered the admixed individuals identified in the two population model as a distinct population that admixes with individuals to the north and south (Figure S6). Given these results and the lower cross-validation error in Admixture and higher \( D_K \) score in Structure for the two-population model, we do not discuss other \( K \) values further.

Spatial principal components analysis and DAPC support the results of concatenated phylogenetic analyses and STRUCTURE and ADMIXTURE (Figure 4, Figure S7). sPCA shows strong differentiation along PC1 into two distinct clusters of individuals corresponding to northern and southern populations. Intermediate individuals again correspond to the individuals within each population that are closest to the boundary between the populations at Loreto. DAPC supports partitioning of individuals into two populations, similar to Admixture and Structure (Figure S7). The Mantel test we performed indicated the presence of significant isolation by distance across the range of C. ruber (\( p = .00001 \)). However, plotting geographic and genetic distances shows that is not the result of a single cline, but of two distant populations that are genetically distinct (Figure 5), consistent with the identification of two geographically and genetically distinct populations identified by clustering methods. Significant isolation by distance was also detected within the northern population when southern individuals were excluded from the analysis (\( p = .00001 \)). Plotting genetic and geographic distances within the northern population showed a single cloud of points consistent with
a continuous genetic cline resulting from range expansion (Figure S8).

3.4 | Coalescent phylogenetic analysis and population modelling

We used FSC2 to select among the phylogeographic models shown in Figure 2. Rankings of these models by AIC scores are shown in Table 1. The best fit model is model 6, which is a secondary contact model and has an AIC weight more than twice that of any other single model. Several other models provided reasonably good fits to the data, as evidenced by AIC weights ranging from 0.19 to 0.07. These included an IM model, a secondary contact model with expansion of the northern population, an IM model with migration from south to north only, and an IM model with expansion of the northern population (models 2, 9, 3, and 5 in Figure 2, respectively).

Across all five of the best-fit FSC2 models, the northern population is approximately half as large as the southern population (Table 2). Migration rates between populations are estimated to be highly asymmetric, with a much higher rate for migration from south to north (0.77–1.36 individuals/generation across models) than the opposite (0.26–0.67 individuals/generation across models). Divergence between populations was estimated to have occurred c. 450–510 ka, with models incorporating secondary contact estimating contact around 80 ka.

The estimated effective population size of the northern population estimated from FSC2 is similar to the estimate from G-PHOCS (Table 2; see Table S3 for uncertainty around estimated values). However, across all of the best models tested in FSC2, higher population sizes were recovered for the southern population and lower population sizes were recovered for the ancestral population compared to G-PHOCS estimates. FSC2 also estimates considerably older divergence times for the northern and southern populations than G-PHOCS (~450–510 ka compared to ~280 ka, respectively). The two best-fit IM models with bidirectional migration yielded migration rates somewhat similar to the estimated rates from G-PHOCS, but with higher rates from south to north. Migration rates for other models are considerably different from G-PHOCS estimates, with both secondary contact models recovering more than twice as many migrants/generation moving in each direction.
SNAPP analyses recovered the divergence between the northern and southern  _C. ruber_ populations as 0.19 times that of the root height of the tree (Figure S9). Assuming a divergence of 3 Ma between _C. atrox_ and _C. ruber_ (Blair & Sánchez-Ramírez, 2016; Reyes-Velasco et al., 2013), we converted this to an approximate divergence between the northern and southern _C. ruber_ populations of 570 ka.

### 3.5 Niche modelling

The high AUC value of 0.907 suggests that the niche model generated from present-day climatic conditions captures the current distribution of _C. ruber_ well. LGM projections are similar whether using MIROC or CCSM climatic reconstructions (Figure 5, Figure S10). LGM projections suggest that a reasonably large portion of the Baja California Peninsula was suitable habitat for _C. ruber_. However, the northernmost portion of the current range of _C. ruber_ is estimated to have been unsuitable habitat.

### 4 DISCUSSION

All of our phylogenetic and population clustering analyses concordantly identify two populations within _C. ruber_, a northern and southern population with a boundary near Loreto. In the vicinity of Loreto, there is a large zone of admixture that extends a linear distance of c. 350 km between the northernmost and southernmost admixed individuals present in our dataset (Figure 3). This admixed zone extends farther to the north of Loreto than to the south, consistent with the higher migration rates estimated in this direction between populations from G-PhoCS and FSC2. The population break is concordant with the historical subspecific taxonomy of _C. ruber_, including the zone of admixture, indicating that the genetic differentiation of the populations matches previously observed morphological differentiation among traditional subspecies (Klauber, 1949; Van Denburgh, 1920). This phylogeographic break has also been observed in zebra-tailed lizards (_Callisaurus draconoides_; Lindell et al., 2005). However, _C. draconoides_ also shows breaks at the mid-peninsula and Isthmus of La Paz, leaving _C. ruber_ as unique among squamate reptiles in showing the Loreto break alone.

The estimated divergence times for the northern and southern populations of _C. ruber_ can provide some insight into possible processes that have resulted in population divergence centred at Loreto. Divergence dates estimated by SNAPP and the best-fit models in FSC2 are all roughly congruent in the range of ~450–570 ka. These dates are somewhat older than the divergence time estimated by
on taxa not particularly closely related to C. ruber populations in Baja California and southern California subsampled datasets for models fit in FSC2 to Crotalus ruber.

TABLE 2 Average AIC, ΔAIC, and AIC weights across 10 replicate subsampled datasets for models fit in FSC2 to Crotalus ruber populations in Baja California and southern California

<table>
<thead>
<tr>
<th>Model</th>
<th>Avg. AIC</th>
<th>ΔAIC</th>
<th>AICw</th>
</tr>
</thead>
<tbody>
<tr>
<td>SecCon</td>
<td>3,545.7</td>
<td>0</td>
<td>0.44</td>
</tr>
<tr>
<td>IM</td>
<td>3,547.3</td>
<td>1.6</td>
<td>0.19</td>
</tr>
<tr>
<td>Sec Con N exp</td>
<td>3,547.7</td>
<td>2.0</td>
<td>0.16</td>
</tr>
<tr>
<td>IM S to N mig</td>
<td>3,548.1</td>
<td>2.4</td>
<td>0.13</td>
</tr>
<tr>
<td>IM N exp</td>
<td>3,549.3</td>
<td>3.5</td>
<td>0.07</td>
</tr>
<tr>
<td>IM N to S mig</td>
<td>3,569.6</td>
<td>23.9</td>
<td>2.8 x 10^{-6}</td>
</tr>
<tr>
<td>Sec Con N to S</td>
<td>3,570.7</td>
<td>25.0</td>
<td>1.6 x 10^{-6}</td>
</tr>
<tr>
<td>No M</td>
<td>3,570.9</td>
<td>25.2</td>
<td>1.5 x 10^{-6}</td>
</tr>
<tr>
<td>Sec Con S to N</td>
<td>3,574.8</td>
<td>29.0</td>
<td>2.1 x 10^{-7}</td>
</tr>
</tbody>
</table>

Abbreviations of models are: secondary contact (Sec Con), isolation migration (IM), secondary contact with expansion of the northern population size (Sec Con N Exp), IM with migration from the southern population into the northern population only (IM S to N), IM with expansion of the northern population size (IM N Exp), IM with migration from the northern population into the southern population only (IM N to S), secondary contact with migration from the northern population into the southern population only (Sec Con N to S), no migration (No M), and secondary contact with migration from the southern population into the northern population only (Sec Con S to N).

G-PhoCS (280 ka). The discrepancy between divergence times estimated by FSC2 and G-PhoCS may be partially attributable to some combination of the different data types used in each analyses (FSC2 takes site frequency spectra as input and G-PhoCS takes full sequence data and integrates over gene trees) and the higher ancestral population sizes estimated by G-PhoCS (Oswald, Overcast, Mauck, Andersen, & Smith, 2017), but the overall cause of these discrepancies is unclear. Due to the congruence between divergence time estimates from SNAPP and FSC2, we prefer parameter estimates from FSC2 over those from G-PhoCS. Given that G-PhoCS includes migration, it would be expected that the divergence date would be similar to or older than the estimate from SNAPP, because the failure to account for ongoing migration should cause SNAPP to be biased towards younger rather than older divergence times (Leaché, Harris, Ranala, & Yang, 2014). Although the mutation rate that we have used to convert estimates into absolute time is based on taxa not particularly closely related to C. ruber, the general congruence in divergence times between SNAPP and FSC2 increases our confidence in the estimated dates. Furthermore, all dates estimated from FSC2 and G-PhoCS would have to be off by a factor of more than 5 to fall outside of the Pleistocene, which would require the true mutation rate to be very different from the rate we have used.

Population divergence and secondary contact in the Pleistocene suggests the possibility that climatic cycles have played a role in past isolation of C. ruber, as has been suggested for several other desert southwest taxa (e.g., Jezkova et al., 2016; Riddle, Hafner, Alexander, & Jaeger, 2000; Schield et al., 2015). There are no obvious current barriers to dispersal by C. ruber in the vicinity of Loreto, strongly suggesting that temporary, rather than ongoing, barriers to dispersal have resulted in the observed differentiation between northern and southern populations. No known or proposed geographic barriers, such as seaways, were present during the Pleistocene, further suggesting that climatic fluctuations may have led to initial population isolation. Unfortunately, we are unable to estimate climatic suitability of areas near Loreto at the timing of divergence due to a lack of climatic data available for this time period. Our niche models suggest that areas near Loreto were suitable habitat for C. ruber at the LGM, but climatic conditions at the LGM may not be analogous to conditions during other glacial cycles.

FSC2 analyses indicate that a model of isolation followed by secondary contact provides the best fit to the data. This model provides a better fit than IM, secondary contact with an expanding northern population, IM with northward migration only, and IM with an expanding northern population models, but these models all had reasonably high AIC weights (0.19–0.07; Table 1). All of these models contain high levels of migration between populations and two of them include secondary contact, with the combined AIC weight of these two secondary models totaling 0.6, lending support to the hypothesis that these populations were temporarily isolated by climatic conditions and have subsequently come back into contact and resumed exchanging genes. If Pleistocene climate cycles have produced the patterns of genetic structure observed in C. ruber, then it is plausible that older divergences in other taxa might also have been caused by climatic fluctuations rather than historically proposed seaways.

TABLE 2 Parameter estimates from G-PhoCS and FastSimCoal2 for populations of Crotalus ruber in Baja California and southern California

<table>
<thead>
<tr>
<th></th>
<th>N North</th>
<th>N South</th>
<th>N ancestral</th>
<th>N north preExp</th>
<th>T_div</th>
<th>T_cont</th>
<th>2Nm N → S</th>
<th>2Nm S → N</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-PhoCS IM</td>
<td>38,223</td>
<td>79,201</td>
<td>63,361</td>
<td>2,818</td>
<td>283,182</td>
<td>0.21</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>FSC2 Sec Con</td>
<td>45,870</td>
<td>93,555</td>
<td>36,211</td>
<td>467,417</td>
<td>86,291</td>
<td>0.59</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>FSC2 IM</td>
<td>42,765</td>
<td>92,798</td>
<td>37,641</td>
<td>494,006</td>
<td></td>
<td>0.27</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>FSC2 Sec Con N Exp</td>
<td>45,356</td>
<td>93,684</td>
<td>37,126</td>
<td>27,274</td>
<td>462,786</td>
<td>82,880</td>
<td>0.67</td>
<td>1.36</td>
</tr>
<tr>
<td>FSC2 IM S to N</td>
<td>39,501</td>
<td>97,612</td>
<td>44,214</td>
<td>446,922</td>
<td></td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSC2 IM N Exp</td>
<td>43,274</td>
<td>91,672</td>
<td>36,478</td>
<td>23,842</td>
<td>508,330</td>
<td>0.26</td>
<td>0.78</td>
<td></td>
</tr>
</tbody>
</table>

The first row shows the parameters estimated in G-PhoCS, while remaining rows show parameter estimates from models fit in FSC2 ordered by AIC score (Table 1). FSC2 models follow abbreviations in Table 1. Estimated parameters are the population sizes of the current, ancestral, and pre-expansion northern populations, divergence time, timing of secondary contact, and migration rates in numbers of individuals between populations.
The difference between the migration rates estimated under IM models and secondary contact models highlights the importance of proper model specification for parameter estimation (Thomé & Carstens, 2016). The number of migrants from north to south was just over twice as high when estimated under secondary contact models as when estimated under IM models, and the number of migrants moving from south to north was also higher under secondary contact than bidirectional IM models. Such differences in estimated migration rates could be critical for researchers seeking to estimate rates of migration to support or reject species delimitation hypotheses (e.g., Gottscho et al., 2017), as use of a mis-specified model could result in biased estimates of migration that could lead to erroneous conclusions about species limits.

A northern range expansion of the northern population of *C. ruber* is suggested by the pattern of more northern individuals being successively nested in the RAxML results (Figure 3), the much smaller population size of the northern population relative to the southern population, despite occupying a much larger geographic range, the strong signal of isolation by distance within this population, and that two of the five best-fit models evaluated with FSC2 included expansion of the northern population (Excoffier, Foll, & Petit, 2009; Hewitt, 1996). FSC2 models that include expansion of the northern population estimate the pre-expansion effective population size to have been ~24–27k individuals, compared to ~40–45k individuals at present (across all five best models), representing a considerable increase. A northern population expansion is consistent with our niche models, which show that a large portion of the current distribution of the northern population was not climatically suitable during the LGM. The northern population of *C. ruber* may therefore have been forced to retreat to southern climatic refugia, only to expand to the north once climatic conditions in these regions became suitable once again. Although a pattern of Pleistocene/post-Pleistocene northern range expansion has only been demonstrated for a small number of taxa in the region (e.g., Nason et al., 2002; Whorley et al., 2004), many peninsular taxa have similar distributions. In particular, many lizards and snakes have similar distributions to *C. ruber* (Grismer, 1999), and likely similar climatic requirements, suggesting that it is likely that other Baja California reptiles may be found to have experienced range contractions and northern expansions. For instance, the mitochondrial phylogenies of *C. draconoides* (Lindell et al., 2005) and *Urosaurus nigricaudus* (Lindell et al., 2008) both show a general pattern in which more northern individuals tend to be highly nested, suggesting the possibility that these taxa have also experienced northern range expansions in response to changing climates.

5 | CONCLUSIONS

Genomic data strongly support the presence of two populations within mainland *C. ruber* that correspond to historically recognized subspecies with an extensive zone of admixture where the populations contact in a zone of morphological intermediates. Using demographic modelling approaches and phylogeographic model selection, we have shown that migration between populations is extensive, and migration rates are higher from south to north than from north to south. We demonstrate that the population structure within *C. ruber* can be explained by a model of isolation followed by secondary contact during the Pleistocene. There is also evidence for a recent range expansion in the northern population of *C. ruber*. Climatic fluctuations during the Pleistocene are the most plausible driver of the observed phylogeographic patterns in *C. ruber*, given that there are no known geographic barriers in the Pleistocene. We suggest that climatic fluctuations may have produced similar genetic structure in additional species. Finally, we reiterate previous findings that accurate estimation of demographic parameters, such as migration rates, is contingent on proper model selection.

ACKNOWLEDGEMENTS

This work was supported by a UCR Newell Award to SMH and startup funds from UCR to TEH. We thank Andrew Gottscho, John Andermann and Dean Leavitt for laboratory and analytical assistance. The 2003 binational expedition, including Oscar Flores-Villela, Dustin Wood, Jorge H. Valdez Villavicencio, Anny Peralta Garcia, Cynthia Jauregui, Tom Myers and Robert Hill, to the Agua Verde-Punta Mechudo conservation corridor contributed to the collection of specimens used in this study. We thank Jordan Satler for analytical advice.

ORCID

Sean M. Harrington http://orcid.org/0000-0002-7784-0070

REFERENCES


DATA ACCESSIBILITY

All data matrices and input files used for analyses in this study are deposited in the Dryad Digital Repository: doi:10.5061/dryad.7hs0n. Raw sequence data are deposited in the NCBI Sequence Read Archive, SRA SRP119491. Accession numbers for individual samples can be found in Appendix S1.

BIOSKETCH

Sean Harrington is a postdoctoral fellow broadly interested in phylogenetics and biogeography of lizards and snakes. He is specifically interested in the processes that drive biodiversity patterns, ranging from phylogeographic scales to drivers of large-scale variation in speciation rates among clades.

Author contributions: All authors contributed to the writing of the manuscript. SMH and TWR conceived the project. SMH collected molecular data and performed all data analyses. BH collected specimens.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Harrington SM, Hollingsworth BD, Higham TE, Reeder TW. Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (Crotalus ruber) in Baja California. J Biogeogr. 2017:00:1–12. https://doi.org/10.1111/jbi.13114