Journal of Heredity, 2024, **115**, 57–71
https://doi.org/10.1093/jhered/esad073
Advance access publication November 20, 2023
Original Article

Original Article

Repeated patterns of reptile diversification in Western North America supported by the Northern Alligator Lizard (*Elgaria coerulea*)

Adam D. Leaché[1,*], Hayden R. Davis[1], Chris R. Feldman[2], Matthew K. Fujita[3], Sonal Singhal[4]

[1]Department of Biology & Burke Museum of Natural History and Culture, University of Washington, Seattle, WA, United States
[2]Department of Biology and Program in Ecology, Evolution and Conservation Biology, University of Nevada, Reno, NV, United States
[3]Department of Biology, The University of Texas at Arlington, Arlington, TX, United States
[4]Department of Biology, California State University – Dominguez Hills, Carson, CA, United States

*Corresponding author: Email: leache@uw.edu

Corresponding Editor: Christopher Blair

Abstract

Understanding the processes that shape genetic diversity by either promoting or preventing population divergence can help identify geographic areas that either facilitate or limit gene flow. Furthermore, broadly distributed species allow us to understand how biogeographic and ecogeographic transitions affect gene flow. We investigated these processes using genomic data in the Northern Alligator Lizard (*Elgaria coerulea*), which is widely distributed in Western North America across diverse ecoregions (California Floristic Province and Pacific Northwest) and mountain ranges (Sierra Nevada, Coastal Ranges, and Cascades). We collected single-nucleotide polymorphism data from 120 samples of *E. coerulea*. Biogeographic analyses of squamate reptiles with similar distributions have identified several shared diversification patterns that provide testable predictions for *E. coerulea*, including deep genetic divisions in the Sierra Nevada, demographic stability of southern populations, and recent post-Pleistocene expansion into the Pacific Northwest. We use genomic data to test these predictions by estimating the structure, connectivity, and phylogenetic history of populations. At least 10 distinct populations are supported, with mixed-ancestry individuals situated at most population boundaries. A species tree analysis provides strong support for the early divergence of populations in the Sierra Nevada Mountains and recent diversification into the Pacific Northwest. Admixture and migration analyses detect gene flow among populations in the Lower Cascades and Northern California, and a spatial analysis of gene flow identified significant barriers to gene flow across both the Sierra Nevada and Coast Ranges. The distribution of genetic diversity in *E. coerulea* is uneven, patchy, and interconnected at population boundaries. The biogeographic patterns seen in *E. coerulea* are consistent with predictions from co-distributed species.

Keywords: admixture, biogeography, gene flow, phylogeography, population structure

Introduction

Phylogeography aims to identify the spatial and temporal patterns of genetic diversity within species by estimating the structure, connectivity, and phylogenetic history of populations (Edwards et al. 2016). Spatial patterns of genetic variation are shaped by a broad array of contemporary forces and historical events, including geographic features, climate, habitat fragmentation, and ecological interactions with other species (Zamudio et al. 2016). Understanding the processes that shape genetic diversity by either promoting or preventing population divergence can help identify key geographic areas that contain high genetic diversity or unique evolutionary lineages.

Western North America provides an ideal setting for investigating phylogeographic patterns and understanding the forces that shape and maintain genetic diversity. The region includes the California Floristic Province (a biodiversity hotspot) and Pacific Northwest, diverse ecoregions with high levels of species richness, geomorphic complexity, variable climate, and dramatic geological history. Wide-ranging species that occur across such diverse landscapes can help us understand how historical and present day biogeographic and environmental heterogeneity can shape spatial patterns of divergence. One such widespread species is the Northern Alligator Lizard (*Elgaria coerulea*). The species inhabits coniferous forest and other cool mesic habitats that extend from the southern portion of British Columbia, Canada to Northern California with a ring-like distribution southward along the Sierra Nevada Mountains in the east, and Coast Ranges in the west (Lais 1976; Stebbins and McGinnis 2018). This geographic distribution is broadly overlapping with a diverse community of other squamates endemic to the region (Stebbins and McGinnis 2018), including lizards (e.g. *E. multicarinata*, *Plestiodon skiltonianus/gilberti* complex, *Sceloporus graciosus*, and *Sceloporus occidentalis*) and snakes (e.g. *Charina bottae*, *Contia tenuis*, *Diadophis punctatus*, *Lampropeltis zonata*, and *Thamnophis elegans*).

Received July 07, 2023; Accepted November 09, 2023

In this study, we use genomic data from 120 *E. coerulea* individuals spanning the species range to test hypotheses related to population structure, species delimitation, and biogeography. First, many squamate species—especially those found in dynamic regions like Western North America—consist of multiple, highly divergent populations, some of which exhibit levels of divergence on par with recognized species. Thus, we test whether *E. coerulea* is composed of a single wide-ranging metapopulation (our null model), or whether spatially segregated populations represent species-level lineages. Previous morphological and color pattern datasets identified four subspecies (Fitch 1938) or two pattern classes (Good 1985) across *E. coerulea*. However, allozyme data (Good 1985) and mitochondrial DNA (mtDNA) sequence data (Lavin et al. 2018) found these groups do not correspond to clades, suggesting color patterns may reflect responses to local selection pressures rather than phylogenetic relationships and biogeography (Lavin et al. 2018). We expand on this work by using single-nucleotide polymorphism (SNP) data to determine population structure and phylogenetic history in *E. coerulea*. Through population clustering analyses, we estimate the number and composition of populations in *E. coerulea*, as well as their potential admixture. We then test opposing hypotheses about diversity in *E. coerulea* by analyzing these data in a species delimitation framework to determine whether the genomic data support: 1) a single metapopulation, 2) the subspecific taxonomy, or 3) the mtDNA groupings.

Second, many Western North American squamate species show common patterns of diversification, suggestive of shared responses to this region's dynamic glacial history (Feldman and Spicer 2006; Bouzid et al. 2022; Grismer et al. 2022). These species exhibit a contrasting pattern of high persistence in the southern end of their ranges and recent colonization in the northern end of their ranges. Southern populations in these species show evidence of demographic stability, and are genetically quite divergent, especially in the southern Sierra Nevada. In the north, populations display reduced genetic diversity indicative of recent post-glacial expansion into the Pacific Northwest, and some species like *S. occidentalis* exhibit admixture between populations indicative of shifting, unstable demographics (Bouzid et al. 2022). Thus, we test the hypothesis that the phylogeography of *E. coerulea* conforms to diversification patterns identified in other co-distributed squamate reptiles. To do so, we first test the hypothesis that the population structure seen in *E. coerulea* is spatially concordant with the boundaries of major ecological regions and mountain ranges as seen in other co-distributed species. Then, we conduct a phylogenetic analysis to determine the branching order of populations and their divergence times. Based on the Pleistocene climatic restrictions to colonization at high latitudes, we predict that the phylogeny will display deep divergences in the Sierra Nevada and relatively shallow divergence in the Pacific Northwest. Finally, we test if the northern populations show evidence of reduced genetic diversity relative to southern populations, reflective of post-Pleistocene range expansion (Hewitt 1999), and if northern populations exhibit high admixture, reflective of shifting population geographic boundaries.

Methods

Sampling

We sampled 120 *E. coerulea* from throughout their geographic distribution in the Western US and Canada (Fig. 1). Regions with high sampling density include the Sierra Nevada Mountains and Coast Ranges in California. Two large sampling gaps where tissue samples were not available include Oregon and the Northeastern portion of the distribution in Idaho and Montana (Fig. 1). We acquired tissue samples from the Museum of Vertebrate Zoology (MVZ), California Academy of Sciences (CAS), Burke Museum of Natural History and Culture (UWBM), and Central Washington University (CWU). We included two *E. multicarinata* to serve as outgroups for the phylogenetic analyses (Leavitt et al. 2015; Lavin et al. 2018), but excluded them from the population genetic analyses of *E. coerulea*. Voucher specimen information and locality data are provided in Supplementary Table S1.

Data collection

We performed DNA extractions using QIAGEN DNeasy extraction kits (QIAGEN Inc.). We collected ddRADseq data following standard library protocols (Peterson et al. 2012) with the restriction enzymes *SbfI* and *MspI* and with size selection between 415 and 515 bp (see Leaché et al. 2015). All P1 adapters included a unique molecular identifier (UMI; 8 bp) to identify PCR duplicates, a barcode sequence (6 bp) for multiplexing, and a restriction enzyme overhang sequence (5 bp). We obtained sequence reads (100 bp) using single-end sequencing with an Illumina NovaSeq 6000 at the QB3 facility at UC Berkeley.

Bioinformatics

We demultiplexed, filtered, and assembled the raw Illumina reads using ipyrad v.0.7.30 (Eaton and Overcast 2020). To demultiplex samples, we used their unique barcode and adapter sequences, and we changed sites with Phred quality scores under 99.95% (Phred score = 33) into "N" characters and discarded reads with ≥ 10% N's. We clustered the filtered reads using a threshold of 90%. Additionally, we discarded consensus sequences that had low coverage (<10 reads), excessive undetermined or heterozygous sites (>8), or too many haplotypes (>2 for diploids). We removed putative paralogs by filtering out loci with excessive shared heterozygosity among samples (paralog filter = 0.5). The final data assembly allowed a maximum of 10% missing data at a locus. Given that stringency of data filtering can affect downstream inferences (O'Leary et al. 2018), we also also conducted population structure analyses with up to 50% missing data. To filter the assembly for subsequent population genetic analyses, we used VCFtools v0.1.15 (Danecek et al. 2011) to apply a minor-allele frequency threshold (MAF = 0.01) and to select one SNP per locus. The demultiplexed data are available at the NCBI Sequence Read Archive (BioProject ID: PRJNA903883). Data assemblies and analysis input files are deposited on Dryad (https://doi.org/10.5061/dryad.sj3tx96b9).

Population structure and genetic diversity

We determined the optimal number of populations (K) and admixture proportions of samples using the maximum likelihood method ADMIXTURE v1.3.0 (Alexander et al. 2009). ADMIXTURE is an efficient and accurate method for ancestry and population structure estimation from SNP data that is three to four times faster than other popular methods (Alexander and Lange 2011). To determine the best-fit model, we compared analyses for K = 1 through K = 12 and selected the analysis that minimized group assignment error; considering

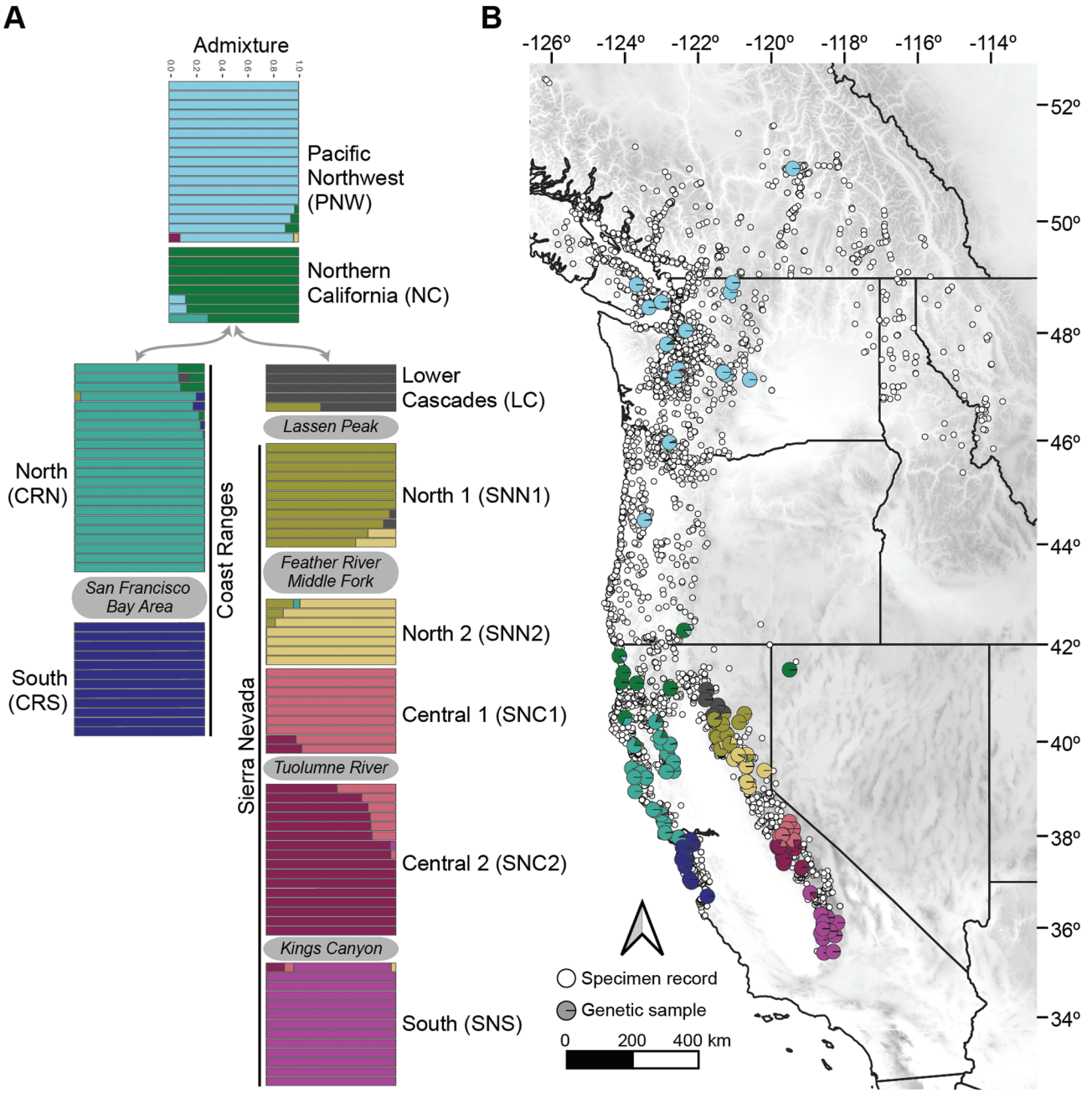

Fig. 1. Population structure of the Northern Alligator Lizard (*Elgaria coerulea*) in Western North America. A) Admixture analysis results shown as barplots for 120 specimens using the SNP data assuming a model with 10 populations (K = 10). The populations are organized into geographic groups located in the Coast Ranges and Sierra Nevada Mountains. Geographic features that coincide with population boundaries are shown between populations. B) Map of the west coast of North America and the sampling locations of the 120 *E. coerulea* included in the study (shown as admixture pie charts) along with additional museum specimen records from GBIF (white circles).

the K with the lowest cross-validation (CV) error as the best-fit model (Alexander et al. 2009). The analyses were repeated 10 times to measure uncertainty in CV error estimation. These analyses were repeated with datasets allowing up to 10% and 50% missing data to compare model selection results with varying numbers of loci. We also generated a principal component analysis (PCA) scatterplot of genetic variation (assuming no grouping information) for both datasets using the R package Adegenet (Jombart and Ahmed 2011).

We used the filtered SNP data to calculate F_{ST} (Weir and Cockerham 1984) among populations with VCFtools. The F_{ST} values are unweighted averages for all SNPs. We used PAUP v.4.0a (Swofford 2002) to calculate pairwise genetic distances (uncorrected p-distances). These calculations used the concatenated ddRADseq loci (with 10% missing data), which combines all variable and constant sites.

Phylogenetic network, species tree, and species delimitation

We estimated the genealogical relationships for *E. coerulea* using phylogenetic methods that make different assumptions about the evolutionary process of divergence. First, because

the genealogical relationships for independently segregating nuclear loci are not expected to conform to a single bifurcating tree for a population sample, we used a phylogenetic network to visualize the data. Phylogenetic networks are useful for displaying relationships that are not bifurcating, and networks can also help identify admixed samples (Blair and Ané 2020). We constructed a genetic network from the concatenated SNP data (uncorrected "*p*" distances; all constant and variable sites were included) using the NeighborNet algorithm (Bryant and Moulton 2004) in SplitsTree v4.6 (Huson and Bryant 2006).

Second, we conducted species tree estimation to obtain a bifurcating species tree topology for the major population clusters, which we then used as a guide tree for subsequent species delimitation analyses (see below). To estimate a species tree, we used SNAPP v1.5.0 (Bryant et al. 2012) implemented in BEAST v2.5.2 (Bouckaert et al. 2019). We assigned samples to populations using the results of the ADMIXTURE analysis corresponding to $K = 10$ populations, with four samples selected per population to decrease computational times. We selected samples to maximize geographic coverage without reference to admixture levels. Because the model does not estimate gene flow between populations it is expected to underestimate branch lengths and overestimate population sizes, although the severity of these biases depend on the amount of gene flow occurring (Leaché et al. 2014; Jiao et al. 2020). The SNAPP model uses biallelic loci, and we filtered the data for biallelic SNPs using the phrynomics package (Leaché et al. 2015). We set the mutation rates u and v to 1.0. For the species tree prior, we used a gamma distribution $\sim G(2, 200)$ with a mean of alpha*beta = 400. We set the prior for the expected genetic divergence theta (θ) using a gamma distribution $\theta \sim G(1, 250)$ with a mean of alpha/beta = 0.004. We ran two separate analyses for 5 million generations each, sampling every 50 generations. We combined the posterior distributions using Logcombiner, and summarized an MCC tree using Treeannotator after discarding the first 20% of samples as burn-in. We visualized the posterior probability distribution of species tree topologies using DensiTree (Bouckaert 2010). Lastly, we transformed branch lengths to divergence times assuming a lizard-specific nuclear mutation rate of 3.17e–9 (mutations per site per year), which is a genome-wide estimate from a breeding study of the gecko *Coleonyx brevis* (Bergeron et al. 2023). Given the phylogenetic distance between alligator lizards and geckos (approx. 200 million yr) and the uncertainty of mutation rate estimates, these divergence times should be interpreted with caution.

Finally, we conducted species delimitation analysis to test opposing hypotheses about taxonomic diversity within *E. coerulea*. Specifically, tested whether populations represent a single species. Alternatively, the number of species could equal the number of subspecies (4) or mtDNA groups (10) (Lavin et al. 2018). For these analyses, we used the multispecies coalescent with migration (MSC-M) (Flouri et al. 2023) implemented in the program BPP v.4.6.2 (Flouri et al. 2018). We used heuristic criteria to delimit species based on population parameters estimated under the MSC-M model (e.g. population/species divergence times, population sizes, and migration rates). The genealogical divergence index (*gdi*) (Jackson et al. 2017) is a metric that measures the probability that two alleles from a population/species will coalesce before either coalesces with an allele from another population/species, and before the population/species divergence time. The *gdi* ranges from 0 to 1, where values close to 1 indicate a high level of population divergence (resembling that of species), and values close to 0 reflect a panmictic population (Jackson et al. 2017). A hierarchical merging procedure allows for recursive species delimitation using *gdi* to traverse large species trees (Leaché et al. 2019). We conducted hierarchical heuristic species delimitation under the MSC-M model using the program HHSD v.0.9.8 (Kornai et al. 2023) to automate the estimation of *gdi* in BPP. We used the species tree topology estimated using SNAPP as the guide tree for species delimitation, and we allowed bidirectional migration between all sister-species pairs in the phylogeny, including historical migration between common ancestors and their sister lineages. We assigned the migration rate prior using a gamma distribution $G(0.1, 10)$ with mean 0.1/10 = 0.01 migrant individuals per generation. Populations were merged if either descendant from a common ancestor failed to exceed $gdi > 0.7$, which is a conservative empirical threshold for the population-species boundary (Jackson et al. 2017).

Migration surfaces and gene flow

We used three methods to investigate migration across the range of *E. coerulea*. These analytical methods differ in their underlying assumptions and the spatial and temporal resolution at which they can measure gene flow. Thus, comparing results across methods can provide complimentary perspectives on gene flow patterns. First, to estimate the spatial patterns of migration across the range of *E. coerulea*, we used Estimating Effective Migration Surfaces (EEMS; Petkova et al. 2016). EEMS compares genetic and geographic distances to those expected under an isolation-by-distance (IBD) model to identify corridors of spatial connectivity and geographical barriers to migration (effective migration rates, m). The method also estimates genetic diversity (effective diversity rates, q) in a spatial context. We ran EEMS for 10 million iterations (2,000,000 burn-in) sampling every 25,000 iterations assuming 1,000 demes.

Second, to assess gene flow between populations, we used *D*-statistic tests and *F*4-ratio statistics (Malinsky et al. 2021). The tests use small asymmetric species trees with four tips to identify an excess of shared derived alleles between populations to distinguish gene flow or ancestral population structure from incomplete lineage sorting (Durand et al. 2011; Patterson et al. 2012). We examined patterns of excess allele sharing using the filtered SNP data using DSuite (Malinsky et al. 2021). We computed *F*-statistics (tests of excess allele sharing) across all possible sets of three populations on the SNAPP species tree using Dtrios. We then used Fbranch to construct a matrix of *f*-branch statistics for all branches on the tree (tips and internal branches), and then used dtools.py to visualize patterns of excess allele sharing between all populations.

Finally, we used the MSC-M (Flouri et al. 2023) to analyze the genomic sequence data and estimate the migration rates between population pairs of *E. coerulea*. We conducted MSC-M analyses to estimate the bidirectional migration rates between geographically adjacent population pairs using BPP. Within the MSC-M framework, it is possible to estimate migration rates with low information content loci (e.g. RADseq/ddRADseq loci) while accounting for incomplete lineage sorting and phylogenetic uncertainty at each locus (Xu and Yang 2016; Jiao et al. 2021). The MSC-M model measures migration as continuous gene flow over the history of a population/species and does not distinguish ancient from recent gene flow (Jiao et al. 2021). The migration rate $M = mN$ is

measured in the expected number of migrants from the donor population to the recipient population per generation; N is the (effective) population size of the donor population, and m is the proportion of immigrants in the recipient population from the donor population every generation under the forward-in-time perspective (Flouri et al. 2023). We assigned the migration rate prior using a gamma distribution $G(0.1, 10)$ with mean 0.1/10 = 0.01 migrant individuals per generation. Population size (θ) and tree height (τ) are measured in units of substitutions per site, and as such we assigned the θ prior an inverse gamma distribution IG(3, 0.002) with mean 0.002/2 = 0.001 (corresponding to 0.1% sequence divergence within a population) and the τ prior IG(3, 0.01) with mean 0.01/2 = 0.005 (corresponding to 0.5% sequence divergence from root to tip). We ran each analysis twice to check for consistency in estimated parameters using a burn-in period of 10,000 samples before collecting 50,000 samples (sampling every other step).

Results

Genetic data

After demultiplexing and filtering the ddRADseq data, we obtained an average of 2.58 million raw reads for each of the 120 *Elgaria coerulea* included in the study (Table 1). The high number of retained reads (99.97%) and high coverage (41.99X) resulted in an assembly with an average of 3,072 (10% missing data) or 7,925 (50% missing data) loci (Table 1).

Table 1. Summary of the ddRADseq data assembly for 120 *Elgaria coerulea* samples using the iPyrad pipeline.

Variable	Mean
Raw reads	2,582,015
Reads retained	99.97%
Clusters	57,121
Coverage	41.99X
Heterozygosity	0.00924
Error	0.00128
Loci (10% missing data)	3,072
Loci (50% missing data)	7,925

The libraries were sequenced with an Illumina NovaSeq 6000 (100 bp single-end reads).

Population structure and genetic diversity

Population structure analyses of the filtered SNP data suggest high levels of geographic structure in *E. coerulea* (Fig. 1). Using CV scores as a metric for selecting the best-fit model, the 10 replicate ADMIXTURE analyses selected population models ranging from K = 8 to K = 12 (Supplementary Fig. S1). Patterns of CV scores were similar across both high and low levels of filtering for missing data (Supplementary Fig. S1). Models $K = 9$ and $K = 10$ had the highest CV scores, and we present the results for $K = 10$ (Fig. 1). The PCA analysis of the SNP data (with no grouping information) separates the samples into five geographic groups, two of which are overlapping (Supplementary Fig. S2).

The 10 populations show regional geographic structure (Fig. 1) conforming to the Pacific Northwest (PNW), Northern California (NC), Lower Cascades (LC), two populations in the Coast Ranges that are separated north–south by the San Francisco Bay Area (CRN = Coast Range North; CRS = Coast Range South), and five populations in the Sierra Nevada with two in the north separated by the Feather River Middle Fork (SNN1 = Sierra Nevada North 1, SNN2 = Sierra Nevada North 2), two in the central Sierra Nevada separated by the Tuolumne River (SNC1 = Sierra Nevada Central 1, SNC2 = Sierra Nevada Central 2), and one in the south that is separated by Kings Canyon (SNS = Sierra Nevada South).

Samples with mixed ancestry belonging to ≥ 2 populations are located at population boundaries (Fig. 1). These admixed samples are found between the Pacific Northwest, NC, and Coast Range populations. Admixed samples are also found throughout the Sierra Nevada and Lower Cascades, where they effectively link together geographically proximate populations (Fig. 1). The one exception is that the adjacent northern and central Sierra Nevada populations show no evidence of admixture; however, there is a sizeable sampling gap between these populations (Fig. 1).

Measurements of population differentiation due to genetic structure (F_{ST}) range from a low of 0.079 to 0.151 for comparisons between geographically proximate populations in NC, Coast Ranges, and Lower Cascades, and a maximum of 0.314 to 0.453 between the Pacific Northwest and the Sierra Nevada, populations at opposite ends of the distribution (Table 2). The average F_{ST} across all populations is 0.239. Pairwise genetic distances are lowest between Northern California with nearby populations in the Coast Range, Lower Cascades, and Pacific Northwest (0.095% to

Table 2. Sequence divergence and F_{ST} values for *Elgaria coerulea* populations calculated using the SNP data.

	CRN	CRS	LC	NC	PNW	SNC1	SNC2	SNN1	SNN2	SNS
CRN	**0.059%**	0.108	0.143	0.079	0.171	0.236	0.201	0.156	0.174	0.219
CRS	0.105%	**0.082%**	0.226	0.151	0.279	0.326	0.266	0.220	0.238	0.301
LC	0.124%	0.166%	**0.059%**	0.129	0.358	0.363	0.291	0.182	0.226	0.334
NC	0.095%	0.133%	0.121%	**0.085%**	0.225	0.289	0.242	0.169	0.187	0.275
PNW	0.136%	0.172%	0.177%	0.125%	**0.043%**	0.453	0.347	0.314	0.367	0.398
SNC1	0.216%	0.247%	0.233%	0.225%	0.270%	**0.044%**	0.190	0.287	0.288	0.323
SNC2	0.205%	0.234%	0.224%	0.216%	0.259%	0.117%	**0.068%**	0.242	0.240	0.233
SNN1	0.139%	0.174%	0.130%	0.147%	0.195%	0.211%	0.200%	**0.064%**	0.132	0.274
SNN2	0.158%	0.191%	0.167%	0.166%	0.215%	0.197%	0.187%	0.115%	**0.091%**	0.286
SNS	0.222%	0.255%	0.242%	0.233%	0.272%	0.207%	0.176%	0.223%	0.214%	**0.054%**

F_{ST} values (above the diagonal) are unweighted averages for all loci. Sequence divergences (below diagonal; uncorrected p-distances) are population averages. Mean sequence divergence within a population is shown along the diagonal (in bold). Population codes are as follows: CRN = Coast Range North, CRS = Coast Range South, LC = Lower Cascades, NC = Northern California, PNW = Pacific Northwest, SNC1 = Sierra Nevada Central 1, SNC2 = Sierra Nevada Central 2, SNN1 = Sierra Nevada North 1, SNN2 = Sierra Nevada North 2, SNS = Sierra Nevada South.

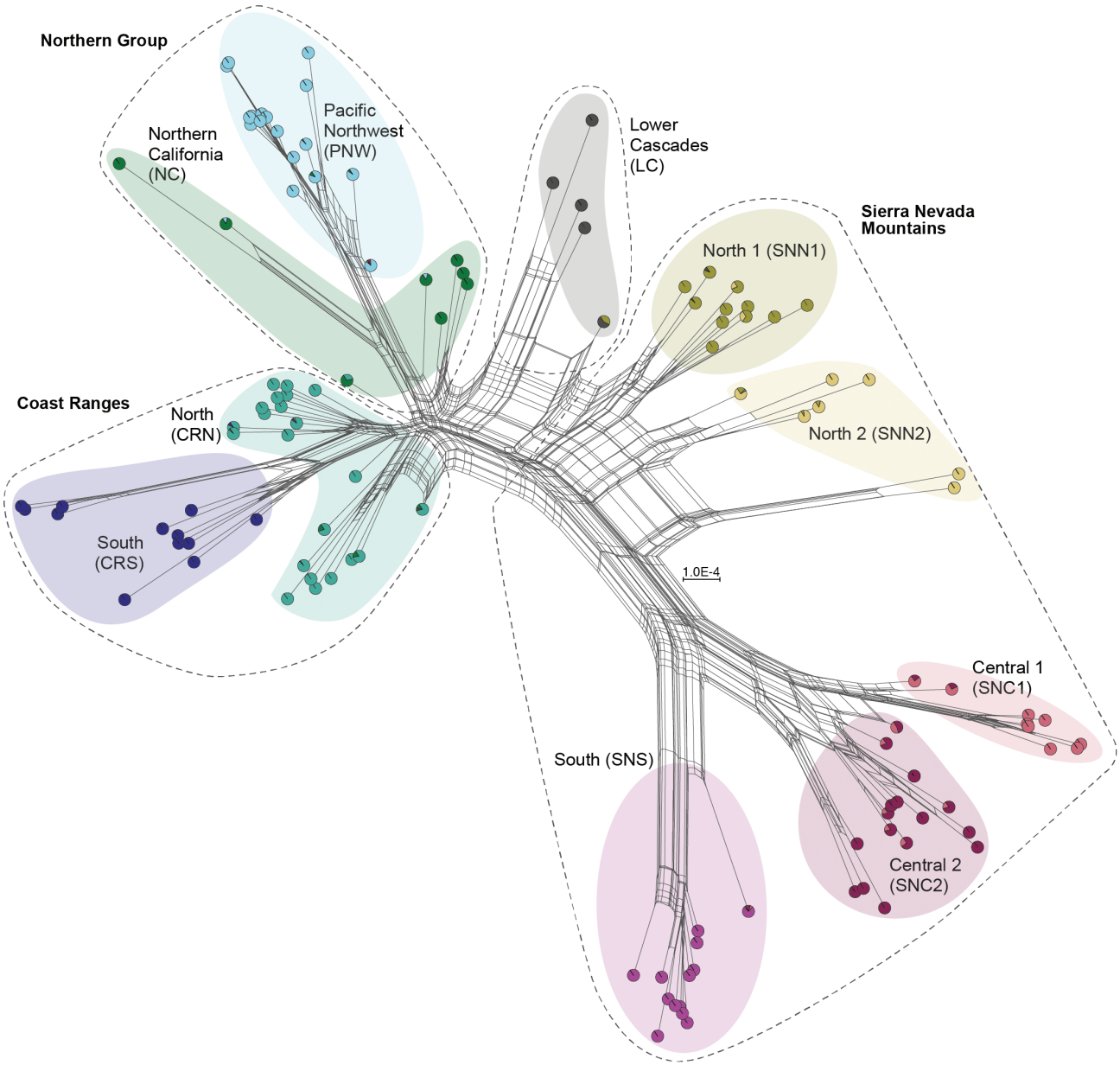

Fig. 2. Phylogenetic network analysis for *Elgaria coerulea* with admixture proportions assigned to each sample. Three major geographical regions show clustering (dashed lines), including a northern group, Coast Ranges, and Sierra Nevada Mountains. The Lower Cascades is positioned between the Northern Group and Sierra Nevada Mountains.

0.177%), and a maximum of 0.272% between the Pacific Northwest and the Sierra Nevada (Table 2). The average pairwise genetic distances (uncorrected *p*-distances) among all populations is 0.187%. The average genetic distance—or, genetic diversity—within populations ranges from a low of 0.043% in the Pacific Northwest and a high of 0.091% in the Sierra Nevada (population SNN2; Table 2).

Phylogenetic network, species tree, and species delimitation

The phylogenetic network analysis of *E. coerulea* clusters the samples into geographical groups that closely match the population structure results (Fig. 2). Examples of geographic clustering include the Sierra Nevada Mountains and the Coast Ranges. Distinct clusters are not supported in the phylogenetic network for NC and northern Coast Ranges (CRN), indicating that these groups could contain additional population structure (Fig. 2). Genetic diversity is higher for southern populations and the Sierra Nevada Mountains, and lowest in the northern populations (Fig. 2). As expected for network analyses, samples with high admixture proportions are positioned between populations (Fig. 2).

The species tree analysis of *E. coerulea* provides strong support (posterior probability = 1.0) for relationships among the Sierra Nevada and Lower Cascade populations, while the specific branching order within the clade containing the Pacific Northwest, NC, and Coast Ranges populations is uncertain (Fig. 3). Superimposing all species trees in the posterior distribution using DENSITREE illustrates that this phylogenetic uncertainty is related to the placements of the NC

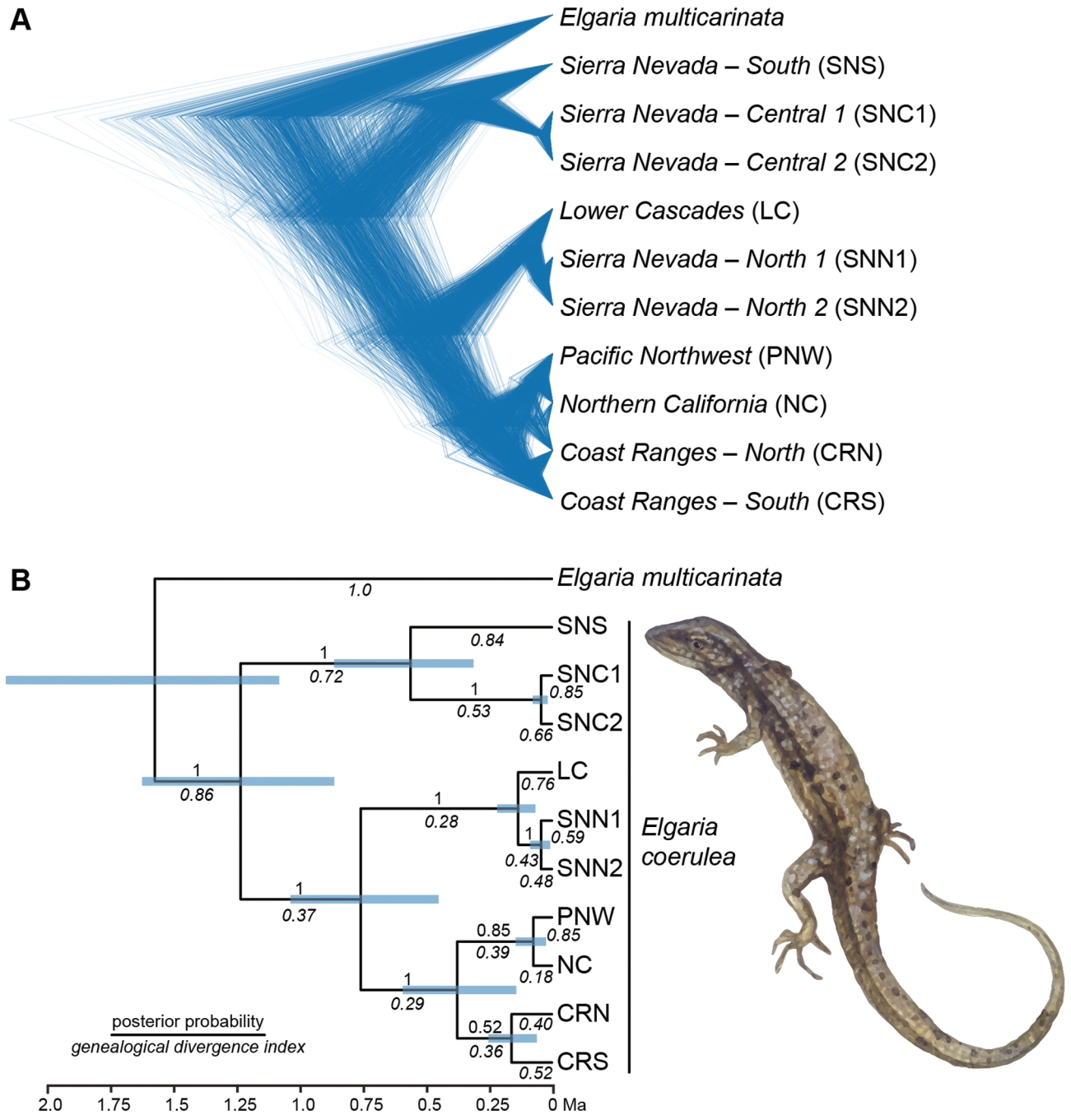

Fig. 3. Species tree analysis of *Elgaria coerulea* populations (K = 10) using the SNP data in a multispecies coalescent framework using SNAPP. A) Densitree plot of the posterior distribution of species trees illustrates uncertainty in the phylogenetic placement of the NC population, as well as branch length uncertainty across the species tree. B) Maximum clade credibility (MCC) species tree with node bars showing branch length uncertainty (95% HPD) and posterior probability values above branches. Results of species delimitation analysis using *gdi* are shown below branches, which supports two species *E. coerulea* and *E. multicarinata*. Divergence times were calculated assuming a lizard-specific nuclear mutation rate of 3.17e–9 mutations/site/yr (Bergeron et al. 2023). Illustration of *E. coerulea* courtesy of Simone Des Roches.

and northern Coast Ranges populations (Fig. 3A). The stable relationships among the Sierra Nevada samples include an initial split between the southern and central Sierra Nevada with all other populations, and a close relationship between the northern Sierra Nevada and the Lower Cascades (Fig. 3).

Divergence time estimation using a fixed mutation rate places the initial divergence of *E. coerulea* populations from the Sierra Nevada (southern and central) from all others in the Pleistocene at 1.28 mya, but with wide confidence intervals (95% highest posterior density; HPD) ranging from 0.86 to 1.61 mya (Fig. 3B). The next divergence event at 0.75 mya (95% HPD = 0.44 to 1.03 mya) further subdivides populations from the Sierra Nevada (northern) and Lower Cascades from those in the Coast Ranges, NC, and Pacific Northwest. At shallowest levels of the species tree, divergence events between geographically proximate populations in the Sierra

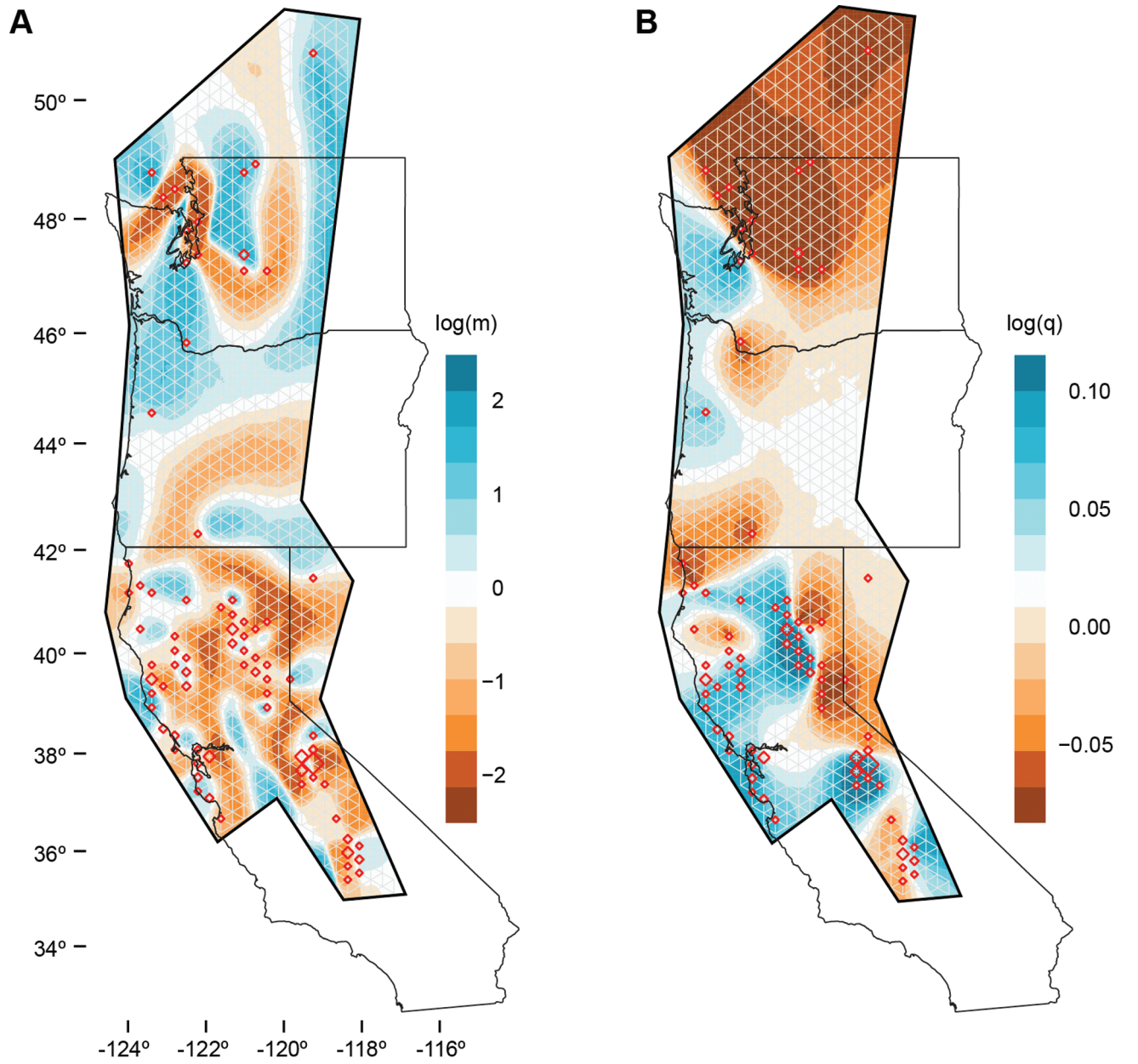

Fig. 4. Contour maps for effective migration (m) and effective diversity (q) estimated for *Elgaria coerulea* using the SNP data in EEMS. A) Compared with an isolation-by-distance (IBD) model (areas shown in white), regions with higher and lower m are shown in blue and brown, respectively. B) Compared with an IBD model, regions with higher and lower q are shown in blue and brown, respectively. The analysis assumes 1,000 demes with sample fit to demes (shown in red).

Nevada or the Pacific Northwest and NC occurred as recently as 75 kya (95% HPD = 22 to 141 kya) or 157 kya (95% HPD = 59 to 249 kya), respectively.

Species delimitation analysis using *gdi* under the MSC-M supports all populations of *E. coerulea* as a single species (Fig. 3). Starting with a 10-population-species tree, the HHSD approach merges all populations into one species before rejecting the final merge between *E. coerulea* and *E. multicarinata* (*gdi* > 0.7 for both descendants; Fig. 3).

Migration surfaces and gene flow

Estimated migration (m) and diversity (q) surfaces across the range of *E. coerulea* are presented in Fig. 4. In comparison to an IBD model, barriers to gene flow are more pronounced in the southern portion of the range (i.e. California) than the northern portion of the range (i.e. Oregon and Washington) (Fig. 4A). For example, the regions of high migration in California are small and tightly clustered, while barriers to gene flow are numerous across both the Sierra Nevada and Coast Ranges (Fig. 4A). The opposite pattern is seen in the Pacific Northwest, where there are only a few barriers to gene flow (Fig. 4A). Likewise, genetic divergence is relatively higher in the southern portion of the range in California, with pockets of high genetic diversity along the Coast Ranges and Sierra Nevada, in contrast to an extensive area of low genetic diversity further north in the Pacific Northwest (Fig. 4B).

Tests for gene flow between populations using the *D*-statistic and *f*4-ratio provide evidence for gene flow between

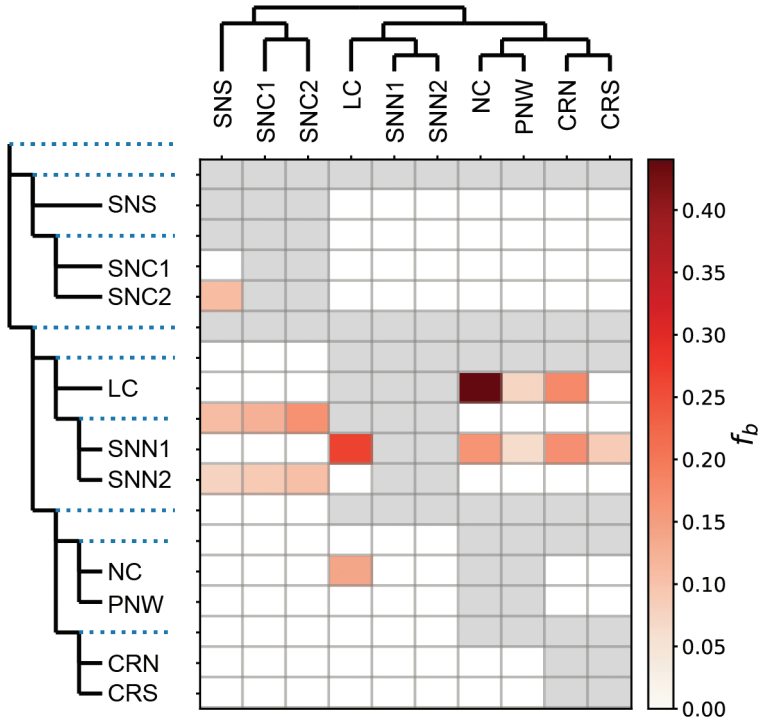

Fig. 5. Matrix of *f*-branch values from the Dsuite analysis of the SNP data for populations of *Elgaria coerulea*. The branch-specific statistic *f*-branch measures the proportion of alleles shared between populations (populations on columns are allele donors, populations on rows are allele recipients). The SNAPP species tree with $K = 10$ populations was used. The outgroup (*E. multicarinata*) is not shown. The species tree is expanded on the *y* axis to include internal branches. Gray entries in the matrix correspond to tests that could not be conducted because they are inconsistent with the species tree. Population abbreviations follow those used in Fig. 1.

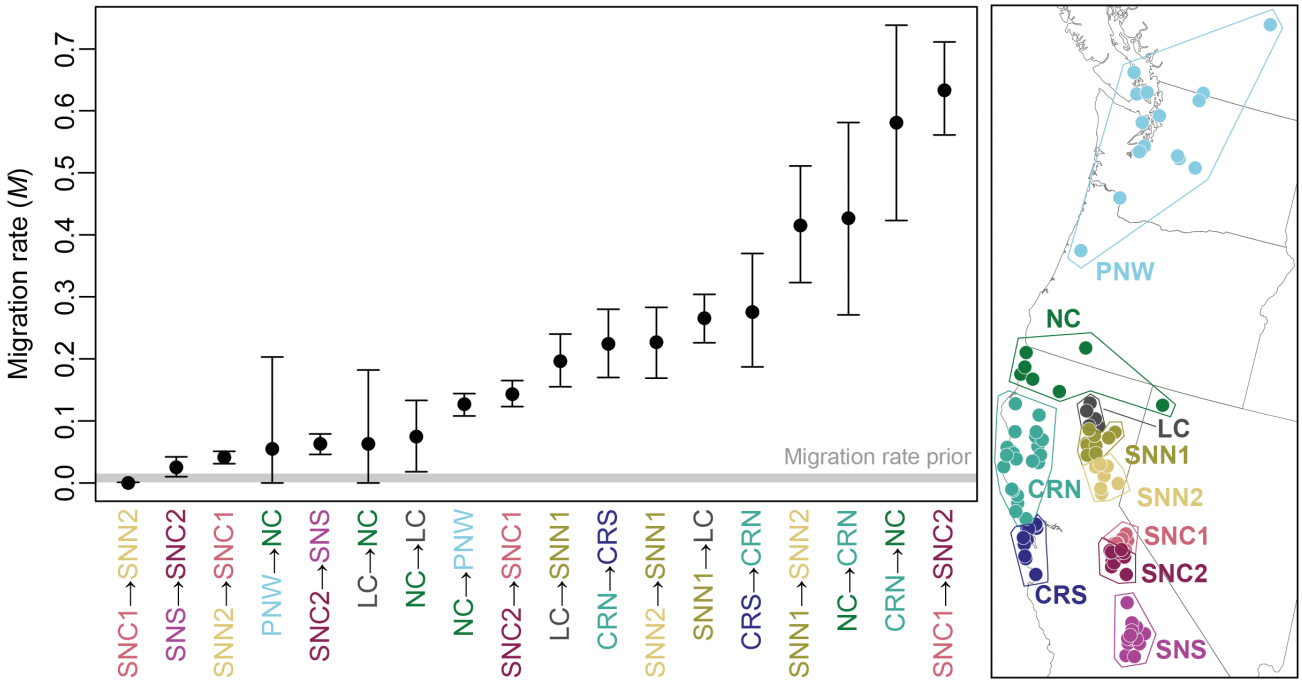

Fig. 6. Migration rate estimates ($M = mN$) estimated using the MSC-M model in BPP. Each bar shows the unidirectional migration rate between geographically adjacent populations (mean and 95% confidence interval). Map displays ten major population clusters.

several adjacent populations (Fig. 5, Supplementary Table S2). Most instances of gene flow are localized around the Lower Cascades and northern Sierra Nevada, which show evidence for gene flow with all surrounding populations, regardless of their phylogenetic relationships (Fig. 5). Additional gene flow is detected in the central and southern Sierra Nevada, which effectively links all populations located along the Sierra Nevada and Lower Cascades (Fig. 5, Supplementary Table S2).

Migration rates estimated with the MSC-M model in BPP support a continuum of gene flow between geographically adjacent populations of *E. coerulea* (Fig. 6). The estimated

migration rates range from a low of 0 to a high of 0.63 migrants per generation (Fig. 6, Supplementary Table S3). Some unidirectional migration rates include 0 in the 95% confidence interval or overlap broadly with the prior distribution for migration, suggesting that migration is absent or too weak to detect; these cases each include the populations separated by large gaps in geographical sampling (SNC1 to SNN2; SNS to SNC2; PNW to NC; LC to NC). The highest migration rates are found between population in the same or similar geographic areas, including the central Sierra Nevada (SNC1 to SNC2), northern Sierra Nevada (SNN1 to SNN2), Coast Ranges (CRS to CRN), and between the Coast Ranges and NC (Fig. 6, Supplementary Table S3).

Discussion

Population structure, phylogeography, and systematics

Phylogeographic investigations are crucial for characterizing the spatial and temporal components of genetic diversity that produce and maintain contemporary patterns of genetic variation in widespread species. Particularly for relatively low-dispersing species like squamates, phylogeographic studies often reveal multiple, deeply divergent populations, some of which exhibit as much divergence as between recognized species. Thus, as a first step, we conducted phylogeographic analyses of the Northern Alligator Lizard (*E. coerulea*) using genomic data to understand the population structure and connectivity of this species.

Estimating population structure is an important yet nuanced step in phylogeographic studies that can be prone to misinterpretation (Lawson et al. 2018). We adopted a common procedure to determine the best-fit population structure model for *E. coerulea* by evaluating multiple models and selecting the one with the lowest CV error. Replicating this procedure 10 times yielded ambiguous results for the optimal model, which ranged somewhere between $K = 8$ and $K = 12$, and $K = 10$ was often the optimal model (Supplementary Fig. S1). However, the discrete populations estimated under this model might be better represented by an IBD model (Bradburd et al. 2018). Additionally, population structure is often hierarchical, and some structure likely remains to be detected. The $K = 10$ model is a useful approximation of population structure, because it produces clear geographic structure that is biologically sensible.

We conducted additional analyses to further evaluate the implications of the population structure model. For example, the population structure model suggests that individuals at population boundaries are of mixed ancestry (Fig. 1), and the gene flow and migration analyses complement this result by estimating significant gene flow across population boundaries, in some cases even between non-sister phylogenetic lineages (Fig. 5). We also conducted a spatial analysis of gene flow that does not require any predefined population groupings, and this analysis also identified high levels of connectivity indicative of gene flow as well as significant barriers, some of which map to geographical barriers separating population clusters (Figs. 1 and 4A). Taken together, these analyses indicate that, while *E. coerulea* exhibits population structure, the species maintains connectivity across an expansive geographic distribution.

From a phylogenetic perspective, *E. coerulea* populations are structured into three well-supported groups that occupy distinct geographic regions: 1) southern and central Sierra Nevada; 2) northern Sierra Nevada and Lower Cascades; and 3) Coast Ranges, NC, and Pacific Northwest (Fig. 3). The phylogenetic relationships among populations in the first two groups are robust, while the relationships among populations belonging to the last group are accompanied by substantial discordance and weak support, which is an expected outcome when populations exchange genes (Leaché et al. 2014). Specifically, this species tree uncertainty implies that the NC population contains signatures of gene flow with neighboring populations in the Pacific Northwest and/or northern Coast Range (Fig. 3). This interpretation is consistent with the migration and population structure inferences that also support gene flow between these geographic regions (Figs. 1 and 5).

A previous phylogeographic study of *E. coerulea* using mtDNA (Lavin et al. 2018) provides an independent dataset for comparing the phylogeographic patterns that we detected using SNP data. Discordance among independent loci is expected under some circumstances, such as incomplete lineage sorting and gene flow (Jiao et al. 2021; Petit and Excoffier 2009). The mtDNA genealogy supports 10 major clades, and while the composition of several of these clades match the nuclear data, there are some notable differences (Fig. 7). Concordant patterns of population structure are supported at the extremes of the geographic range of *E. coerulea*, with matching mtDNA and SNP boundaries found in the Pacific Northwest, southern Coastal Range, and southern Sierra Nevada populations (Fig. 7). The central and northern Sierra Nevada populations also share concordant boundaries, although in relation to the mtDNA groupings, the SNP data splits each region into two distinct populations (Fig. 7C), while the mtDNA data suggest these populations are only minimally subdivided (Fig. 2 of Lavin et al. 2018). The most substantial discordance is located at the boundary between NC and northern Coast Range populations, which are broadly dissimilar between mtDNA and SNP data. Gene flow is the likely mechanism for this discordance, as the SNP data support admixture in this geographic area (Figs. 1 and 5).

The subspecific taxonomy of *E. coerulea* is not supported by mtDNA (Lavin et al. 2018) or the SNP data presented here. There are four currently recognized subspecies of *E. coerulea* (Stebbins and McGinnis 2018), based on a handful of meristic traits, body proportions, and color patterns (Fitch 1938; Lais 1976) with some support for intergradation between groups (Fitch 1938). The geographic distributions of two of these subspecies coarsely match the genetically defined groups described here; the range of the Northwestern Alligator Lizard (*E. c. principis*) circumscribes the Pacific Northwest population, and the range of the Sierra Alligator Lizard (*E. c. palmeri*) contains the five Sierra Nevada populations (see Map 124 in Stebbins and McGinnis 2018; Fig. 2 in Lavin et al. 2018). However, the Northwestern subspecies (as currently defined) also includes a portion of the NC population, and the Sierra subspecies would define a paraphyletic assemblage that lumps together distantly related populations that span the Sierra Nevada. Further data collection on morphological and color pattern variation across *E. coerulea* could help us understand what structures morphological variation and if morphological intergradation correlates with regions of genetic admixture (Fig. 1).

In terms of the taxonomic ramifications of the new genomic data presented here, we find no clear evidence for splitting

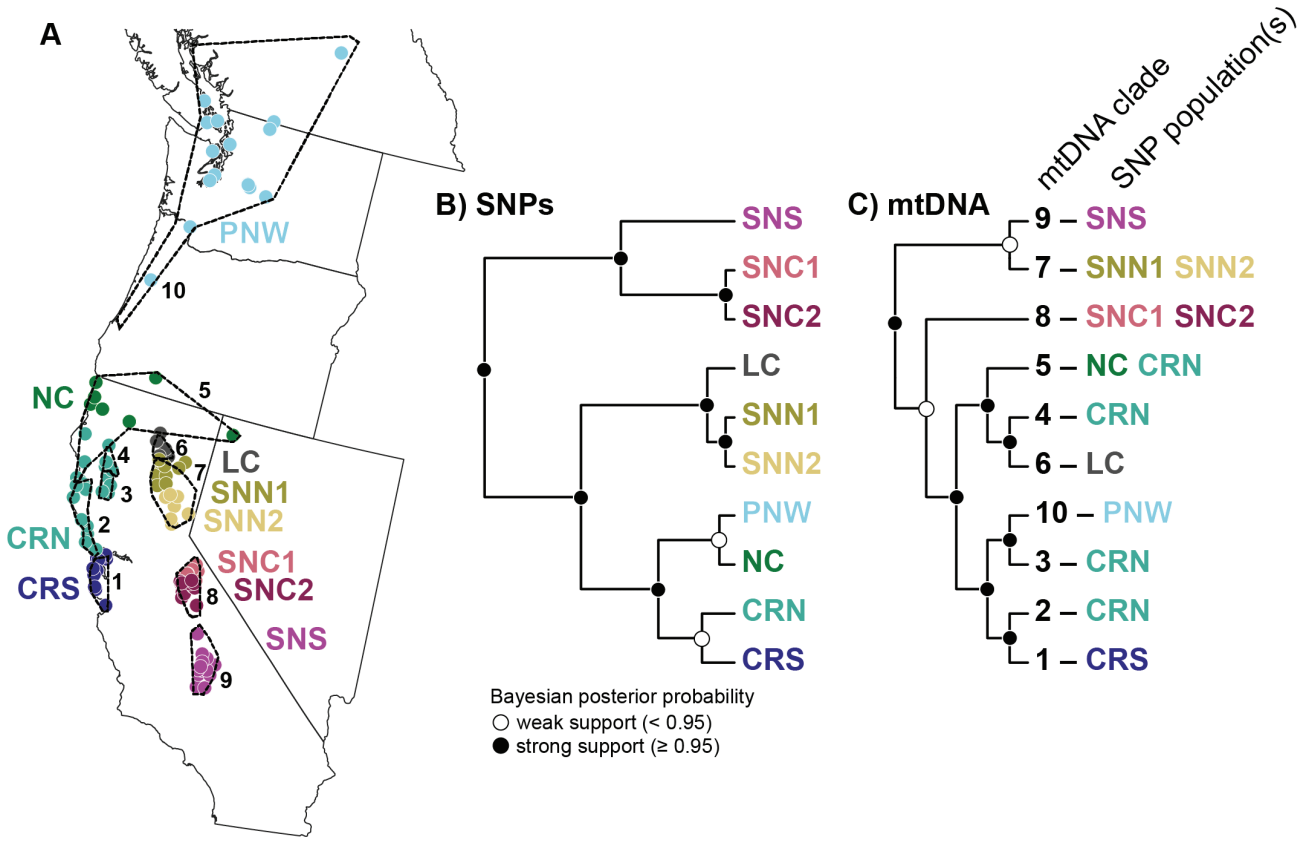

Fig. 7. Comparison of phylogeographic patterns for *Elgaria coerulea* based on mtDNA and SNP data. A) The geographic distributions of the 10 major mtDNA clades are numbered (1 to 10) and outlined with dashed lines. The 10 SNP-based populations (colored circles) are superimposed and color coded to match Fig. 1. B) SNP-based species tree reproduced from Fig. 3. (C) mtDNA gene tree from Lavin et al. (2018) depicted as a cladogram.

E. coerulea into multiple species or subspecies. Species delimitation analyses using the *gdi* failed to support more than one species within *E. coerulea*. These analyses were conducted using an multispecies coalescent model that included migration (Flouri et al. 2023) between all sister lineages in the phylogeny. Genetic data are often used as evidence for delineating species boundaries (Fujita et al. 2012) under the assumption that prolonged periods of reproductive isolation will drive population divergence, potentially leading to speciation (de Queiroz 2007). Although *E. coerulea* is characterized by relatively high levels of population structure, these populations remain connected by gene flow. These data provide evidence that *E. coerulea* populations lack reproductive isolation and are therefore not on independent evolutionary trajectories, similar to other western species that display substantial geographic structure, yet maintain genetic cohesion (*Ensatina eschscholtzii*, Pereira et al. 2011; *P. skiltonianus/gilberti*, Richmond and Jockusch 2007; *S. occidentalis*, Bouzid et al. 2022).

Shared biogeographic patterns

Western North America contains high levels of species richness, ecological diversity, and geological complexity, making it an ideal setting for investigating comparative biogeography. This region includes the California Floristic Province, which is among the world's most biodiverse and endangered terrestrial ecosystems (Myers et al. 2000; Mittermeier et al. 2011). Many species of squamate reptiles that occupy upland habitats (e.g. chaparral, oak woodland, mixed pine, and oak forest) and are co-distributed in this region, sharing a ring-like distribution that follows the foothills and mountains surrounding the Great Central Valley of California, which is a large area of unsuitable habitat (prairie, marshlands, and agricultural fields). These species include *C. bottae* (Rubber Boa), *C. tenuis* (Sharp-tailed Snake), *D. punctatus* (Ringneck Snake), *L. zonata* (California Mountain Kingsnake), *E. coerulea* (Northern Alligator Lizard), *E. multicarinana* (Southern Alligator Lizard), *P. skiltonianus/gilberti* complex (Western Skink), *S. graciosus* (Sagebrush Lizard), *S. occidentalis* (Western Fence Lizard), and *Phrynosoma blainvillii* (Blainville's Horned Lizard). While these species all share a characteristic ring-like distribution in California, their northern and southern distributional limits differ with some species extending as far south as Baja California, Mexico (*D. punctatus*, *L. zonata*, *E. multicarinata*, *P. skiltonianus/gilberti* complex, and *S. occidentalis*), and others extending north into Canada (*C. bottae*, *C. tenuis*, *D. punctatus*, *E. coerulea*, and *S. occidentalis*). Within the overlapping portions of these distributions in the Sierra Nevada Mountains, Coast Ranges, NC, and Pacific Northwest, previous biogeographic work has identified shared phylogenetic patterns that correspond to geological events (Lapointe and Rissler 2005; Feldman and Spicer 2006; Rissler et al. 2006; Mulcahy 2008; Myers et al. 2013; Lavin et al. 2018; Hallas et al. 2021; Bouzid et al. 2022). Informed by the patterns seen in other co-distributed species, we test three hypotheses to explain the phylogeographic structure in *E. coerulea*.

First, given that western Northern American species are responding to similar topographic, ecogeographic, and historical dynamics, we hypothesize that the phylogeographic breaks within *E. coerulea* map to those seen in other species. In almost all cases, we find support for this hypothesis. For example, many species show evidence of two to three distinct populations in the Sierra Nevada (Feldman and Spicer 2006; Rissler et al. 2006); we find three primary populations further structured into five subpopulations (Fig. 1A). While habitat is currently suitable for *E. coerulea* throughout the Sierra Nevada (Fig. 1B), historical climate changes and resulting habitat shifts have likely driven some of this population structuring. Ecological niche modeling of *E. coerulea* during the Last Glacial Maximum identified two primary gaps in habitat suitability in the Sierra Nevada (Lavin et al. 2018). The first gap falls between the northern Sierra Nevada populations (SNN1 and SNN2) and the central Sierra Nevada populations (SNC1 and SNC2), and the second falls between SNC1 and SNC2 and the southern Sierra Nevada population (SNS) (Fig. 1B). These phylogeographic breaks appear roughly concordant with those seen in other species, such as woodrats (*N. fuscipes* complex; Matocq 2002), shrews (*Sorex ornatus*; Maldonado et al. 2001), Ensatina salamanders (*E. eschscholtzii*; Kuchta et al. 2009), Foothill Yellow-legged Frogs (*R. boylii*; McCartney-Melstad et al. 2018), and trapdoor spiders (*Aliatypus janus* complex; Starrett et al. 2018; Nemesiidae; Leavitt et al. 2015). This part of the Sierra Nevada experienced multiple, dramatic glaciation events (Gillespie and Douglas 2011) that would have displaced and segregated populations, leading to the shared patterns of genetic fragmentation in the central Sierra Nevada. These primary breaks are further structured by major river barriers in the northern and central Sierra Nevada populations (Fig. 1). Some breaks are identical to those seen in other species, such as the break in Yosemite National Park, where both *E. coerulea* and *S. occidentalis* are subdivided into populations on either side of the Tuolumne River (Bouzid et al. 2022), or the break in Kings Canyon, which divides populations of *E. coerulea* (Fig. 1) and also separates *R. muscosa* and *R. sierrae* (Byrne et al. 2023).

Other breaks correspond to known phylogeographic splits in other species, such as the structure across the San Francisco Bay Area that splits the northern Coastal Range (CRN) population from the southern (CRS). Also called the Golden Gate break, this divide has been seen in several amphibian, reptile, and mammal species (Rissler et al. 2006; Davis et al. 2008). The northernmost populations of *E. coerulea* (NC and PNW) are found across the Coast Range and Cascade Range divide from the Coastal Range populations. This split has also been observed in some amphibian species in the region (Rissler et al. 2006). Finally, the southern limits of *E. coerulea* fall at two major biogeographic transitions in California: the Monterey Bay in the west and the Transverse Ranges in the east (Feldman and Spicer 2006; Rissler et al. 2006; Lavin et al. 2018).

Thus, the population structure of *E. coerulea* tracks well-known biogeographic features, though the ages of splits across these features is typically younger than the hypothesized formation of the feature. Whether divergence across these features was in complete isolation or occurred with gene flow remains unknown. Our estimates of gene flow (Fig. 6) and the presence of admixed individuals at nearly all population transitions (Fig. 1A) suggest low, geographically structured migration across these features, likely in the past reaching into the present. However, resolving the time-span of migration remains analytically difficult and imprecise, especially in the absence of haplotype data (Roux et al. 2016; Fraïsse et al. 2021).

Second, for species that share a ring-like distribution in western North America, there is no consensus as to where along the ring species are thought to have originated. Where species originated is thought to vary depending on the species. Based on previous data for *E. coerulea* (Peabody and Savage 1958; Lavin et al. 2018), we hypothesized that the species originated in the southern end of its range in the southern Sierra Nevada. As predicted under this hypothesis, the central and southern populations represent the earliest divergence in the phylogeny (Fig. 3), and robustly demonstrate a Sierra Nevada origin for the taxon (Lavin et al. 2018). This pattern of early divergence of southern Sierra Nevada populations is shared with other squamate species, including *C. tenuis* and *E. multicarinata* (Feldman and Spicer 2006; Feldman and Hoyer 2010), and several other sympatric animals, such as woodrats (*Neotoma fuscipes* complex) (Matocq 2002), yellow-legged frogs (*Rana muscoca* complex) (Vredenburg et al. 2007; Byrne et al. 2023), newts (*Taricha torosa* complex) (Kuchta and Tan 2006), and harvestmen (*Calicina* complex) (Emata and Hedin 2016). Divergence time estimates from mtDNA suggest that these events occurred approximately 2 mya (Feldman and Spicer 2006), consistent with the SNP-based estimates for *E. coerulea* divergence between the southern and central Sierra Nevada (1.28 mya [95% HPD = 0.86 to 1.61 mya]; Fig. 3B). Similar splits in other reptiles and amphibians have been dated to an overlapping range of divergence times (1.33 to 3.75 mya) (Calsbeek et al. 2003).

Third, in contrast to the stable southern end, the northern end of their range in the Pacific Northwest was more dynamic for many Western North American species. The Pacific Northwest was heavily glaciated through the Pleistocene, and squamate reptile expansion into the region would have been unlikely until after the retreat of the Cordilleran Ice Sheet (Darvill et al. 2018). As such, we hypothesize that the northern populations of *E. coerulea* found in the Pacific Northwest would show evidence of that region's more dynamic history. Broad-scale analyses of genetic variation in woodland and forest squamates show a decline in genetic diversity at northern latitudes consistent with population expansion into the region (Feldman and Spicer 2006; Bouzid et al. 2022; Davis et al. 2022). Patterns in *E. coerulea* further bolster this model. Recent divergence time estimates support post-Pleistocene expansions into the Pacific Northwest in *E. coerulea* (Fig. 3); the Pacific Northwest population is the most recently diverged population at 0.022 to 0.14 mya (Fig. 3B). Further, as expected, populations in the northern part of the range have reduced levels of genetic diversity relative to the those in the southern region (Fig. 4B). Additionally, dynamic range boundaries can bring formerly isolated populations into contact, thus allowing gene flow across populations (Cutter and Gray 2016; Folk et al. 2018). This biogeographic scenario predicts that populations in recently colonized areas, like the Pacific Northwest, will be more likely to exhibit admixture and gene flow, as seen in *S. occidentalis* (Bouzid et al. 2022). In *E. coerulea*, estimated migration rates are higher in the northern part of the range relative to the south (Fig. 4A), genealogical discordance is high among northern populations (Fig. 3A and 7), and northern populations exhibit evidence

for gene flow (Figs. 5 and 6). All together, these data support that *E. corulea*—like *S. occidentalis*—shows evidence of increased gene flow among northern populations, reflective of these populations' dynamic history. More detailed sampling of woodland and forest squamate species is needed to help us understand how common these patterns are in other species and to further refine our understanding of the timing and origins of post-glacial reptile expansion into the Pacific Northwest.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

For tissue sample loans, we thank L. Scheinberg and R. Bell (California Academy of Sciences), C. Spencer and J. McGuire (Museum of Vertebrate Zoology), J. Irwin (Central Washington University), and S. Birks (Burke Museum of Natural History and Culture). We thank B. Lavin for sharing his mtDNA alignment and phylogeny of *E. coerulea*. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank two anonymous reviewers and the Associate Editor for their insightful comments on a previous version of the manuscript.

Funding

This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. This work is supported in part by funds from the National Science Foundation (DEB-SBS-2023723, NSF-DEB-SBS-2024014, and NSF-DEB-SBS-2023979).

Conflict of interest statement. None declared.

Author contributions

Adam D. Leaché (Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Visualization, Supervision, Project administration, Funding acquisition), Hayden R. Davis (Validation, Investigation, Data curation, Writing – review and editing), Chris R. Feldman (Conceptualization, Resources, Writing – review and editing), Matthew K. Fujita (Conceptualization, Writing – review and editing, Supervision, Funding acquisition), and Sonal Singhal (Conceptualization, Writing – review and editing, Supervision, Project administration, Funding acquisition).

Literature Cited

Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011:12:1–6.

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

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St. Leger J, Shao C, et al. Evolution of the germline mutation rate across vertebrates. Nature. 2023:615:285–291.

Blair C, Ané C. Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. Syst Biol. 2020:69:593–601.

Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010:26:1372–1373.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019:15:e1006650.

Bouzid NM, Archie JW, Anderson RA, Grummer JA, Leaché AD. Evidence for ephemeral ring species formation during the diversification history of western fence lizards (*Sceloporus occidentalis*). Mol Ecol. 2022:31:620–631.

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

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

Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004:21:255–265.

Byrne AQ, Rothstein AP, Smith LL, Kania H, Knapp RA, Boiano DM, Briggs CJ, Backlin AR, Fisher RN, Rosenblum EB. Revisiting conservation units for the endangered mountain yellow-legged frog species complex (*Rana muscosa*, *Rana sierrae*) using multiple genomic methods. Conserv Genet. 2023:1–16. https://doi.org/10.1007/s10592-023-01568-5

Calsbeek R, Thompson JN, Richardson JE. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Mol Ecol., 2003:12:1021–1029.

Cutter AD, Gray JC. Ephemeral ecological speciation and the latitudinal biodiversity gradient. Evolution. 2016:70:2171–2185.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011:27:2156–2158.

Darvill C, Menounos B, Goehring B, Lian OB, Caffee M. Retreat of the western Cordilleran Ice Sheet margin during the last deglaciation. Geophys Res Lett. 2018:45:9710–9720.

Davis EB, Koo MS, Conroy C, Patton JL, Moritz C. The California Hotspots Project: identifying regions of rapid diversification of mammals. Mol Ecol. 2008:17:120–138.

Davis HR, Des Roches S, Anderson RA, Leaché AD. Population expansion, divergence, and persistence in Western Fence Lizards (*Sceloporus occidentalis*) at the northern extreme of their distributional range. Sci Rep. 2022:12:6310.

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

Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011:28:2239–2252.

Eaton DA, Overcast I. ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics. 2020:36:2592–2594.

Edwards SV, Potter S, Schmitt CJ, Bragg JG, Moritz C. Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proc Natl Acad Sci USA. 2016:113:8025–8032.

Emata K, Hedin M. From the mountains to the coast and back again: ancient biogeography in a radiation of short-range endemic harvestmen from California. Mol Phylogenet Evol. 2016:98:233–243.

Feldman CR, Hoyer RF. A new species of snake in the genus *Contia* (Squamata: Colubridae) from California and Oregon. Copeia. 2010:2010:254–267.

Feldman CR, Spicer GS. Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. Mol Ecol. 2006:15:2201–2222.

Fitch HS. A systematic account of the alligator lizards (*Gerrhonotus*) in the western United States and lower California. Am Midl Nat. 1938:20:381–424.

Flouri T, Jiao X, Huang X, Rannala B, Yang Z. Efficient Bayesian inference under the multispecies coalescent with migration. Proc Natl Acad Sci USA. 2023:120:e2310708120
Flouri T, Jiao X, Rannala B, Yang Z. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol Biol Evol., 2018:35:2585–2593.
Folk RA, Soltis PS, Soltis DE, Guralnick R. New prospects in the detection and comparative analysis of hybridization in the tree of life. Am J Bot. 2018:105:364–375.
Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire E, Simon A, Galtier N, Duret L, et al. Dils: demographic inferences with linked selection by using abc. Mol Ecol Res. 2021:21:2629–2644.
Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol. 2012:27:480–488.
Gillespie AR, Douglas HC. Glaciations of the sierra Nevada, California, USA. Developments in Quaternary Sciences. Vol. 15. Elsevier; 2011. p. 447–462.
Good DA. Studies of interspecific and intraspecific variation in the alligator lizards (Lacertilia: Anguidae: Gerrhonotinae). Berkeley, California, USA; 1985.
Grismer J, Scott P, Toffelmier E, Hinds B, Klabacka R, Stewart G, White V, Oaks J, Shaffer HB. Genomic data reveal local endemism in Southern California Rubber Boas (serpentes: Boidae, *Charina*) and the critical need for enhanced conservation actions. Mol Phylogenet Evol. 2022:174:107542.
Hallas JM, Parchman TL, Feldman CR. The influence of history, geography and environment on patterns of diversification in the Western Terrestrial Garter Snake. J Biogeogr. 2021:48:2226–2245.
Hewitt GM. Post-glacial re-colonization of European biota. Biol J Linn Soc. 1999:68:87–112.
Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006:23:254–267.
Jackson ND, Carstens BC, Morales AE, O'Meara BC. Species delimitation with gene flow. Syst Biol. 2017:66:799–812.
Jiao X, Flouri T, Rannala B, Yang Z. The impact of cross-species gene flow on species tree estimation. Syst Biol. 2020:69:830–847.
Jiao X, Flouri T, Yang Z. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. Natl Sci Rev. 2021:8:nwab127.
Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011:27:3070–3071.
Kornai D, Flouri T, Yang Z. Hierarchical heuristic species delimitation under the multispecies coalescent model with migration, bioRxiv, 2023, p. 2023–09, https://doi.org/10.1101/2023.09.10.557025
Kuchta SR, Parks DS, Mueller RL, Wake DB. Closing the ring: historical biogeography of the salamander ring species *Ensatina eschscholtzii*. J Biogeogr. 2009:36:982–995.
Kuchta SR, Tan AM. Lineage diversification on an evolving landscape: phylogeography of the California newt, *Taricha torosa* (Caudata: Salamandridae). Biol J Linn Soc. 2006:89:213–239.
Lais PM. Gerrhonotus coeruleus. Catalog of American Amphibians and Reptiles. Vol. 178. 1976. p. 1–4
Lapointe F-J, Rissler LJ. Congruence, consensus, and the comparative phylogeography of codistributed species in California. Am Nat. 2005:166:290–299.
Lavin BR, Wogan GO, McGuire JA, Feldman CR. Phylogeography of the Northern Alligator Lizard (Squamata, Anguidae): hidden diversity in a western endemic. Zool Scr. 2018:47:462–476.
Lawson DJ, Van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commu. 2018:9:3258.
Leaché AD, Banbury BL, Felsenstein J, De Oca AN-M, Stamatakis A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst Biol. 2015:64:1032–1047.
Leaché AD, Harris RB, Rannala B, Yang Z. The influence of gene flow on species tree estimation: a simulation study. Syst Biol. 2014:63:17–30.
Leaché AD, Zhu T, Rannala B, Yang Z. The spectre of too many species. Syst Biol. 2019:68:168–181.
Leavitt DH, Starrett J, Westphal MF, Hedin M. Multilocus sequence data reveal dozens of putative cryptic species in a radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). Mol Phylogenet Evol. 2015:91:56–67.
Maldonado JE, Vilà C, Wayne RK. Tripartite genetic subdivisions in the Ornate Shrew (*Sorex ornatus*). Mol Ecol. 2001:10:127–147.
Malinsky M, Matschiner M, Svardal H. Dsuite-fast *D*-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021:21:584–595.
Matocq MD. Phylogeographical structure and regional history of the Dusky-footed Woodrat, *Neotoma fuscipes*. Mol Ecol. 2002:11:229–242.
McCartney-Melstad E, Gidiş M, Shaffer HB. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining Foothill Yellow-legged Frog. Heredity. 2018:121:112–125.
Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C. Global biodiversity conservation: the critical role of hotspots. In: Biodiversity hotspots: distribution and protection of conservation priority areas. Berlin, Heidelberg: Springer; 2011. p. 3–22.
Mulcahy DG. Phylogeography and species boundaries of the western North American Nightsnake (*Hypsiglena torquata*): revisiting the subspecies concept. Mol Phylogenet Evol. 2008:46:1095–1115.
Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000:403:853–858.
Myers E, Rodríguez-Robles J, Denardo D, Staub R, Stropoli A, Ruane S, Burbrink F. Multilocus phylogeographic assessment of the California Mountain Kingsnake (*Lampropeltis zonata*) suggests alternative patterns of diversification for the California Floristic Province. Mol Ecol. 2013:22:5418–5429.
O'Leary S, Puritz J, Willis S, Hollenbeck C, Portnoy D. These aren't the loci you'e looking for: principles of effective SNP filtering for molecular ecologists. Mol Ecol. 2018:27:1–24.
Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012:192:1065–1093.
Peabody FE, Savage JM. Evolution of a coast range corridor in California and its effect on the origin and dispersal of living amphibians and reptiles. Zoogeography. 1958:51:159–186.
Pereira RJ, Monahan WB, Wake DB. Predictors for reproductive isolation in a ring species complex following genetic and ecological divergence. BMC Evol Biol. 2011:11:1–15.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE. 2012:7:e37135.
Petit RJ, Excoffier L. Gene flow and species delimitation. Trends Ecol Evol. 2009:24:386–393.
Petkova D, Novembre J, Stephens M. Visualizing spatial population structure with estimated effective migration surfaces. Nat Genet. 2016:48:94–100.
Richmond JQ, Jockusch EL. Body size evolution simultaneously creates and collapses species boundaries in a clade of scincid lizards. Proc R Soc B Biol Sci. 2007:274:1701–1708.
Rissler LJ, Hijmans RJ, Graham CH, Moritz C, Wake DB. Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. Am Nat. 2006:167:655–666.
Roux C, Fraisse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 2016:14:e2000234.
Starrett J, Hayashi CY, Derkarabetian S, Hedin M. Cryptic elevational zonation in trapdoor spiders (Araneae, Antrodiaetidae, *Aliatypus janus* complex) from the California southern Sierra Nevada. Mol Phylogenet Evol. 2018:118:403–413.
Stebbins RC, McGinnis SM. Peterson field guide to western reptiles & amphibians. New York, New York, USA: Houghton Mifflin; 2018.

Swofford DL. Paup*. phylogenetic analysis using parsimony (*and other methods). Sunderland, Massachusetts, USA; 2002.

Vredenburg VT, Bingham R, Knapp R, Morgan JA, Moritz C, Wake D. Concordant molecular and phenotypic data delineate new taxonomy and conservation priorities for the endangered mountain yellow-legged frog. J Zool. 2007:271:361–374.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984:38:1358–1370.

Xu B, Yang Z. Challenges in species tree estimation under the multispecies coalescent model. Genetics. 2016:204:1353–1368.

Zamudio KR, Bell RC, Mason NA. Phenotypes in phylogeography: species' traits, environmental variation, and vertebrate diversification. Proc Natl Acad Sci USA. 2016:113:8041–8048.