

UCLA

UCLA Electronic Theses and Dissertations

Title

An Integrative Population and Landscape Genomic Approach to Conservation of a Threatened California Amphibian at Multiple Spatial Scales

Permalink

<https://escholarship.org/uc/item/4z18r4d8>

Author

Neal, Kevin M.

Publication Date

2019

Supplemental Material

<https://escholarship.org/uc/item/4z18r4d8#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

An Integrative Population and Landscape Genomic Approach to Conservation of a Threatened
California Amphibian at Multiple Spatial Scales

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy in Biology

by

Kevin Michael Neal

2019

© Copyright by
Kevin Michael Neal
2019

ABSTRACT OF THE DISSERTATION

An Integrative Population and Landscape Genomic Approach to Conservation of a Threatened
California Amphibian at Multiple Spatial Scales

by

Kevin Michael Neal

Doctor of Philosophy in Biology

University of California, Los Angeles, 2019

Professor Howard Bradley Shaffer, Chair

Amphibians are threatened world-wide, and due to the elusive, seasonal, and often nocturnal habits of adults, biological assessments of amphibian species are often best conducted via genetic analysis of easily-sampled pond-dwelling larvae. Genetic analysis of amphibian species can benefit their conservation in several ways, including identification of evolutionary lineages and subpopulations as fundamental units of conservation, genetic assessment of demography and diversity, and inference of patterns of gene flow among populations and how patterns are affected by environmental variation. In this dissertation I elucidated the evolutionary relationships and population genetic status of a threatened California amphibian (*Spea hammondi*) at multiple spatial scales using a combination of genetic, genomic, and environmental data. Chapter one utilized limited genetic data to determine phylogenetic

relationships of *Spea* species and used environmental niche modeling to examine ecological differentiation between two allopatric lineages identified within *S. hammondi*. Chapter two took advantage of a newer genomic-scale dataset of thousands of SNP markers to look at fine-scale patterns of genetic variation among natural and artificial *S. hammondi* ponds in a highly urbanized region of Southern California. Chapter three also made use of thousands of markers to validate species-level relationships in *Spea* and used the added genomic resolution to examine relationships within and among genetic clusters and quantified the potential impacts of urbanization on functional genetic connectivity. Broadly, I found that the nominal taxon *S. hammondi* likely comprises two species. Populations within each species were highly differentiated from one another and had exceptionally low effective population sizes, such that each species lacks sufficient adaptive potential to thrive without intervention. Overall, this dissertation applied a suite of phylogenetic, population genomic, and landscape genomic tools to analyze patterns of genetic variation in *S. hammondi* to guide ongoing and future conservation efforts.

The dissertation of Kevin Michael Neal is approved.

Robert Nathan Fisher

Thomas Welch Gillespie

Kirk Edward Lohmueller

Victoria Sork

Howard Bradley Shaffer, Committee Chair

University of California, Los Angeles

2019

Dedication

This dissertation is dedicated to Katy Veroski, who doesn't have the same weird enthusiasm I have for tiny slimy tetrapods but loves me enough to point out good logs to flip regardless.

Table of Contents

List of Figures	x
List of Tables	xvi
Acknowledgments.....	xvii
Vita.....	xx
Chapter 1: Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (<i>Spea hammondi</i>).....	1
Abstract	1
Introduction	3
Methods.....	5
Taxon and genetic sampling.....	5
Phylogenetic Analyses of <i>Spea</i> taxon sets.....	6
Clustering of <i>Spea hammondi</i>	7
Niche Modeling.....	8
Results.....	10
Taxon and genetic sampling.....	10
Phylogenetic analyses.....	10
Cluster analysis.....	12
Niche modeling.....	13
Discussion	14

Acknowledgments	17
Figures	19
Supplemental materials	24
References	25
Chapter 2: Genome-wide SNPs reveal fine-scale patterns of population structure and gene flow in a threatened Southern California amphibian	
Abstract	31
Introduction	33
Methods	36
Study area and sampling	36
DNA extraction and sequencing	37
DNA post-processing	38
Data analysis	39
Population structure	39
Genetic diversity and effective population size and their relationship to pond size	40
Genetic distance and gene flow	41
Landscape resistance	42
Results	43
RAD-seq data assembly	43
Genetic structure	44

Genetic distance and gene flow	46
Genetic diversity and effective population size.....	48
Landscape resistance	49
Discussion	50
Figures.....	56
Tables	64
Supplemental Figures and Tables	67
Appendix 2.1	81
References	86
Chapter 3: Comparative population and landscape genomics of two cryptic, allopatric clades	
suggests novel conservation strategies in a threatened amphibian	95
Abstract	95
Introduction	96
Methods.....	99
DNA extraction and sequencing.....	99
DNA post-processing	100
Data analysis.....	101
Results	107
RAD-seq data	107
Phylogenetic analysis	108

Analyses of <i>Spea hammondi</i> - North clade	109
Analyses of <i>Spea hammondi</i> - South clade	111
Comparative population and landscape genomics.....	113
Discussion	114
Figures.....	118
Supplemental Figures and Tables	134
References	143

List of Figures

- Fig. 1.1** Genetic sampling localities and species ranges (IUCN 2017) of the four nominal *Spea* species: *Spea hammondi* (shown as North [red] and South [purple]); *Spea bombifrons* (yellow); *Spea intermontana* (green); and *Spea multiplicata* (blue). Color figure available online 19
- Fig. 1.2** Tanglegram or cophylogeny of nuclear species tree generated in StarBEAST2 (left, A) and mitochondrial tree generated in MrBayes 3.2.6 (right, B) for *Spea*, with Bayesian posterior probabilities at nodes 20
- Fig. 1.3** STRUCTURE and SAMOVA genetic clusters for *S. hammondi*. A) Map of *S. hammondi* samples with pie charts displaying their STRUCTURE cluster membership proportions. Each pie chart shows proportions for a single individual, and pie charts are displaced from the actual geographic coordinate (connected by light gray lines) to allow display of all individuals at the same site. Inset shows STRUCTURE bar plots for K=2. B) Map of *S. hammondi* samples and their SAMOVA cluster membership; South mtDNA cluster corresponds to South nuDNA STRUCTURE cluster, and the two northern mtDNA clusters together correspond to the North nuDNA STRUCTURE cluster. Black dotted line on A) and B) shows phylogeographic break at the Transverse Ranges..... 21
- Fig. 1.4** Maxent niche models for *S. hammondi* based on current values of Bioclim variables, elevation, and slope. A) Niche model for North genetic cluster. B) Niche model for South genetic cluster. Scale corresponds to probability of occurrence. In both cases, the Transverse ranges show very low probabilities 22
- Fig. 2.1** Map of study area with sample sites colored by FastStructure genetic cluster membership for three hierarchically-determined clusters. Dot-dash outlines mark the boundary of

the Nature Reserve of Orange County. IM and SHOE ponds are artificial and were seeded with individuals from donor ponds at the marked site. Inset shows broader context of study area in Southern California. Green-shaded areas are public lands 56

Fig. 2.2 Hierarchical FastStructure barplots of 148 individuals using 6864 unlinked SNPs. Row 1 column 2 is K=2 for 148 Orange County samples. Row 2 column 1 is the Inland cluster (blue) at K=1. Row 2 column 2 is K=2 for Coast (Purple) + Santena (Green) samples. Row 3 shows Coast and Santena at K=1. 59

Fig. 2.3 Principle components analysis of 148 Orange County samples with 6864 unlinked SNPs 60

Fig. 2.4 NeighborNet distance network of Weir and Cockerham’s pairwise F_{ST} among 26 sites for OC_all_208 at up to 75% missing data per locus 61

Fig. 2.5 A) ResistanceGA-optimized resistance parameterization of depth to bedrock + slope raster. Blue is lower resistance, green-yellow is higher. B) Circuitscape current run through optimized depth to bedrock + slope raster, using native localities as focal nodes. Highest current is in yellow/orange, lowest current in purple..... 63

Fig. 2.S1 Heatmap and histogram of pairwise Weir-Cockerham F_{ST} between sites using the OC_all_208 dataset, with darker colors representing lower F_{ST} values and lighter being higher F_{ST} . Populations clustered using "ward.D2" method in cluster::hclust() in R. Heatmap visualized using gplots::heatmap.2(). Populations with fewer than two samples are excluded 67

Fig. 2.S2 Spearman’s Rho (lower triangle) and Kendall’s tau (upper triangle) correlations between measures of genetic diversity, effective population size, and pond area and perimeter. All values displayed are statistically significant ($p < 0.05$, two-tailed test) 69

Fig. 2.S3 Ancestry barplots for OCreduced148 run in FastStructure with ten cross-validations for K=1-10. K=1-5 from upper left to lower left; K=6-10 from upper right to lower right. Note at higher K values, K clusters typically do not appear in the barplots 71

Fig. 2.S4 fineRADstructure coancestry matrix of 148 spadefoots, from Orange County, California. Lighter/yellower colors are lower coancestry; darker/more purple colors are higher coancestry. Nodes in population tree are labeled with posterior population assignment probabilities..... 72

Fig. 2.S5 NeighborNet network of Jost's D for OC_all_208 at up to 75% missing data per locus 74

Fig. 2.S6 NeighborNet network of Nei's D_a for OC_all_208 at up to 75% missing data per locus 76

Fig. 2.S7 TreeMix tree of Orange County populations with migration parameter set to 8. The 8 edges were CCSP3 to CCSP1; CCSP3 to MORO; GREAT to FREM; SANCAN to BBEND; SADDLE to LOMA; CCSP1 to FREM; GREAT to the ancestor of MORO and CCSP2; and CCSP1 to the ancestor of LIME, IMSHOE, and FREM. 77

Fig. 3.1 Habitat suitability (measured as probability of occurrence) for SPHA-NORTH and SPHA-SOUTH (U.S.) modeled using Maxent 3.4.1, using the map extent of the state of California. Maxent models were run individually for both clades. Suitability surfaces were added together and rescaled to a maximum suitability value of 1 to form the combined suitability surface. Black areas show areas of zero suitability. Beige circles are presence points used in the Maxent models, compiled with a combination of this study's sampling locations and GBIF localities. Blue (SPHA-NORTH) and red (SPHA-SOUTH) points are the current study's genetic

sample locations. Note the band of extremely low suitability separating SPHA-NORTH and SPHA-SOUTH, representing the Transverse Ranges and Tehachapi Mountains 118

Fig. 3.2 IQ-TREE maximum likelihood phylogeny of 19 *Spea* individuals, rooted using 3 *Scaphiopus* individuals, with up to 50% missing data per locus, one SNP per locus, and using optimal model fit (TVMe+ASC+R2) determined by internal ModelFinder using Bayesian Information Criterion. Values at nodes are percent bootstrap support and percent SH-aLRT support, respectively 121

Fig. 3.3 Principle components analysis. A) *S. intermontana* ("SPIN", green dots), SPHA-NORTH ("SPHAN", blue), and SPHA-SOUTH ("SPHAS", red). The individual separated from the rest of the SPHA-SOUTH cluster is the sample from Mexico. B) SPHA-NORTH and SPHA-SOUTH only. PC2 splits SPHA-NORTH into two clusters north and south of the Sacramento-San Joaquin River Delta phylogeographic barrier 122

Fig. 3.4 IQ-TREE maximum likelihood phylogeny of 188 *S. hammondi* (SPHA-NORTH + SPHA-SOUTH) and *S. intermontana* individuals, with up to 50% missing data per locus, one SNP per locus, and using optimal model fit (GTR+F+ASC+R2) determined by internal ModelFinder using Bayesian Information Criterion. Values at nodes are percent bootstrap support and percent SH-aLRT support, respectively. Individuals with "SPHAN", "SPHAS", or "SPIN" in sample name refer to membership in SPHA-NORTH, SPHA-SOUTH, or *S. intermontana*, respectively 124

Fig. 3.5 sNMF ancestry coefficient barcharts for SPHA-NORTH, K= 2, 3, and 10..... 125

Fig. 3.6 TESS3 maps of interpolated cluster membership for K=3 clusters of SPHA-NORTH. Blue, red, and green indicate estimates of cluster membership for that particular cluster found in

a given location. More saturated color implies greater proportion of ancestry of that cluster in individuals at that location. Lower right map is a union of the ancestry maps for each cluster, indicating areas where each cluster is dominant. Black dots are genetic sample sites 126

Fig. 3.7 Best ResistanceGA-optimized landscape resistance surface (maxent suitability) for northern SPHA. Yellow is higher resistance; blue is lower. Genetic sample sites are labeled on map..... 127

Fig. 3.8 sNMF ancestry coefficient barcharts for SPHA-SOUTH, K=2, 3, and 7 128

Fig. 3.9 TESS3 maps of interpolated cluster membership for K=3 clusters of SPHA-SOUTH. Blue, red, and green indicate estimates of cluster membership for that particular cluster found in a given location. More saturated color implies greater proportion of ancestry of that cluster in individuals at that location. Lower right map is a union of the ancestry maps for each cluster, indicating areas where each cluster is dominant. Black dots are genetic sample sites 129

Fig. 3.10 Best ResistanceGA-optimized landscape resistance surface (climatic moisture index) for southern SPHA. Yellow is higher resistance; blue is lower. Genetic sample sites are labeled on map..... 130

Fig. 3.11 Boxplots of several population genetic parameters for the North and South clades of *S. hammondii*. P-value on each plot was calculated using a Wilcoxon rank sum test. H_o : observed heterozygosity; H_e : expected heterozygosity; F_{IS} : inbreeding coefficient; IBDg.FH: individual-level estimate of inbreeding based on expected homozygosity; N_{eb} : effective number of breeders; pi.pop: pond-level nucleotide diversity; pi.indiv.median: median individual nucleotide diversity within each pond; D_a .median: median Nei's D_a for each pond; wcfst.median: median pairwise Weir-Cockerham F_{ST} for each pond; maxentmedian.2km, elev.median.2km: median Maxent

suitability and elevation (respectively) within a 2 km buffer of each pond; imperv.max.2km:
 maximum % impervious surface within a 2 km buffer of each pond..... 131

Fig. 3.12 Maps of the best model of landscape resistance for each clade, with impervious surface masks applied. Black areas in each surface have been modeled as extremely high resistance, reflecting levels of urbanization. Black areas on the perimeters have also been masked due to extremely low Maxent habitat suitability (<0.1) to prevent unlikely alternative corridors when impervious surface masks are applied. Top three panels are the North clade of *S. hammondii*. Bottom three are the South clade..... 132

Fig. 3.13 Paired boxplots showing change in pairwise resistance distances with different % impervious surface masks applied. "impervious50" surfaces were masked such that pixels in the resistance surface with impervious surface greater than 50% were assigned a resistance value of 10,000; in "impervious20," this threshold was 20%. No impervious surface mask was applied in the "none" surface. Each point in a boxplot represents the resistance distance between two localities, with lines connecting the same locality pair across boxplots. Y-axis is in logarithmic scale..... 133

Fig. 3.S1 Choice of optimal number of clusters (K). Cross-entropy scores for SPHA-NORTH in A) aspatial sNMF clustering, B) spatial TESS3 clustering, C) snapclust BIC. Cross-entropy scores for SPHA-SOUTH in D) aspatial sNMF clustering, E) spatial TESS3 clustering, F) snapclust BIC..... 134

List of Tables

Table 1 Summary table of individual and site representation in genetic analyses in this study ..	23
Table 2.1 Genetic diversity values for individual ponds. Starred rows are values calculated on groupings of multiple ponds	64
Table 2.S1 ResistanceGA resistance surface optimization results using Nei's D_a as response variable.....	78
Table 3.S1 Population coordinates and summary statistics (further summarized in Fig. 3.11).	135
Table 3.S2 ResistanceGA model results for northern SPHA.....	141
Table 3.S3 ResistanceGA model results for southern SPHA.....	142

Acknowledgments

I would like to thank Brad Shaffer for his incredibly generous mentorship, advice, and enthusiasm over the years – it was truly a rare opportunity to find an advisor as eager as I was to study the evolution of spadefoots, and he went above and beyond to find funding opportunities to allow me to carry out this research. Brad's unwavering dedication to amphibian conservation is inspiring: no graduate school stint is without its moments of existential crisis, and when at my wit's end all it would take was a chat about peculiar observations of the goofy little animals we study to bring me back to seeing the privilege, significance, and joy of being able to do what we do to help save species. I thank the other members of my dissertation committee—Robert Fisher, Tom Gillespie, Kirk Lohmueller, and Victoria Sork—for their mentorship and feedback over the years and for serving as fantastic examples of enthusiastic study of the natural world. I'm grateful to Milan Mitrovich as well for securing funding, sharing his local expertise, and for helping to navigate the process of doing field work in Orange County. I thank my Master's advisors, Aaron Bauer and Todd Jackman, for their eagerness and patience in training me up from a geologist to an evolutionary biologist. I thank several undergrad advisors and mentors—Jim Russell, Jessica Whiteside, Mike Wyatt, and many others—for effecting that academic transition as well. Adam Leaché deserves immense credit as well, being the first actual herpetologist I worked for, for taking a chance on me when I walked into his office a year out of undergrad and asked for a job in his new lab, and for being the one to finally focus an eager young scientist with way too many interests.

The last two years of the PhD are nothing short of a crucible, and I am immensely grateful for my wonderful girlfriend Katy Veroski for her love, devotion, patience, encouragement, and grounding—sometimes you just need someone to tell you a figure is good

enough and definitely doesn't need to be re-made a 20th time. Our dog Mia deserves many treats and pets for dutifully napping beside me throughout the day while I coded and wrote. I thank my parents Heddy Neal and Brad Neal for raising me to be curious about the world around me, and for being good 90's parents who let me wander through the woods for hours catching salamanders. I thank my older brother Jeff for allowing me to learn from his being first in line for many of life's lessons, and for simply being a good friend.

Past and present members of the Shaffer Lab and others at UCLA have been instrumental in getting me to where I am. Their passion, wisdom, generosity, and camaraderie have carried me through the last six years of living and working in Los Angeles. I fondly thank: Joscha Beninde, Gary Bucciarelli, Samantha Cheng, Adam Clause, Mario Colon, Robert Cooper, Natalia Gallego-García, Müge Gidiş, Jesse Grismer, Ugur Kaya, Tara Luckau, Evan McCartney-Melstad, Genevieve Mount, Mark Phuong, Peter Scott, Phil Spinks, Erin Toffelmeier, Isolde van Riemsdijk, Jannet Vu, Ben Wielstra, and Alex Zarem. I also thank the staff in the UCLA Department of Ecology and Evolutionary Biology, especially Tessa Villaseñor and Jocelyn Yamadera for their above-and-beyond level of support in graduate school administrative logistics.

This work was supported by the UCLA Department of Ecology and Evolutionary Biology, the La Kretz Center for California Conservation Science, the UC Conservation Genomics Consortium, the U.S. Fish and Wildlife Service, the Sea and Sage Audubon, and the Natural Communities Coalition.

Chapter 1 is the final but unformatted manuscript version of Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*).

Conservation Genetics, 19(4), 937–946. doi: 10.1007/s10592-018-1066-7. A license for reuse in this dissertation was provided by Springer Nature. H. Bradley Shaffer was the Principle Investigator, Benjamin B. Johnson conducted the DNA extraction and sequence generation, and I analyzed the data and wrote the manuscript. For Chapters 2 and 3, I performed most of the lab work and most of the Orange County field sampling, generated and processed the sequence data, performed the analyses, and wrote the manuscript. H. Bradley Shaffer was the Principal Investigator, did additional field sampling, and edited the manuscript.

Vita

EDUCATION

Villanova University
MS, Biology (September 2015)
Villanova, PA

Brown University
B.Sc., Geology-Biology (May 2009)
Providence, RI

PUBLICATIONS

Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*). *Conservation Genetics*, 19(4), 937–946. doi: 10.1007/s10592-018-1066-7

Shaffer, H. B., Gidiş, M., McCartney-Melstad, E., Neal, K. M., Oyamaguchi, H. M., Tellez, M., & Toffelmier, E. M. (2015). Conservation Genetics and Genomics of Amphibians and Reptiles. *Annual Review of Animal Biosciences*, 3(1), 113–138. doi: 10.1146/annurev-animal-022114-110920

PRESENTATIONS

Neal, K. M., & Shaffer, H. B. (2017). Spatial genetic structure and niche modeling identify two conservation units for the western spadefoot (*Spea hammondi*). California-Nevada Amphibian Populations Task Force Meeting, Santa Barbara, CA. Oral presentation.

Neal, K. M. & Shaffer, H. B. (2016). Population genetic structure and conservation units of the western spadefoot toad, *Spea hammondi*. Joint Meeting of Ichthyologists and Herpetologists, New Orleans, LA. Oral presentation.

Neal, K. M., Johnson, B. B., Spinks, P. Q., & Shaffer, H. B. (2015). Nuclear-mitochondrial discordance of distinct phylogeographic lineages within the western spadefoot toad, *Spea hammondi*. Joint Meeting of Ichthyologists and Herpetologists, Reno, NV. Oral presentation.

Neal, K. M. & Shaffer, H. B. (2015). Using genomic tools for the conservation of the western spadefoot toad, *Spea hammondi*. Poster presentation, California-Nevada Amphibian Populations Task Force Meeting, Pepperdine University, Malibu, CA. Oral presentation.

Neal, K. M., Jackman, T. R., Sadler, R. A., & Bauer, A. M. (2013). Comparing population genetics in two co-distributed New Caledonian lizards: implications for conservation in a mining-impacted region. Joint Meeting of Ichthyologists and Herpetologists, Albuquerque, NM. Oral presentation.

Chapter 1: Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*)

Full citation:

Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*). *Conservation Genetics*, 19(4), 937–946. doi: 10.1007/s10592-018-1066-7

Abstract

The western spadefoot (*Spea hammondi*) is a Species of Special Concern in California and is now under review by the U.S. Fish and Wildlife Service for listing under the Endangered Species Act. We delineated potential conservation units within *S. hammondi* by analyzing spatial genetic structure across the species' range using five nuclear and one mitochondrial loci. For both nuclear and mitochondrial markers we found that *S. hammondi* consists of two genetically distinct, allopatric clusters divided by the Transverse Ranges. To corroborate the northern and southern genetic clusters as conservation units from an ecological perspective, we applied a niche identity test to environmental niche models of the two groups. We found that the niche models of the northern and southern clusters were significantly different, suggesting they may be ecologically non-exchangeable. Given our demonstration of significant genetic and ecological

differentiation between allopatric clusters of *S. hammondii*, we recommend that ongoing conservation efforts consider each as a separate unit with potentially unique management needs.

Introduction

Cataloging the Earth's biodiversity is of critical importance in our era of increasing extinction rates and rapid global change. The concept of evolutionarily significant units (ESUs) was pioneered as a way to pragmatically describe and catalog independently-evolving lineages below the nominal species level that capture the important evolutionary processes and outcomes of a species' history. Under the regulatory framework of the US Endangered Species Act, Distinct Population Segments (DPSs) often fulfill the same role and can be a powerful tool for protecting key elements of within-species vertebrate diversity. One of the limitations of both these approaches is that neutral genetic markers used in identifying ESUs fail to explicitly account for functional, environmental, or ecological differences among cryptic lineages, yet these facets of within-species diversity are of great consequence to evolution and conservation (Crandall et al. 2000; Fraser and Bernatchez 2001; Moritz et al. 2002; May et al. 2011). To incorporate these potential ecological differences into the identification of intraspecific conservation units, genetic analysis can be paired with environmental niche modeling (ENM) to identify genetically unique lineages that differ in ecological response to environmental variation and change (Rissler and Apodaca 2007; McCormack et al. 2010; May et al. 2011; Kalkvik et al. 2012; Fontanella et al. 2012; Hoisington-Lopez et al. 2012; Gutiérrez-Tapia and Palma 2016; Castellanos-Morales et al. 2016; Ikeda et al. 2017).

North American spadefoots of the genus *Spea* are an appealing system for studies of cryptic genetic and ecological diversity. The four nominal *Spea* species have geographic distributions that cover vast regions of western North America (Fig. 1.1), despite their life histories tending toward philopatry of ephemeral breeding pools throughout often xeric ranges (Buseck et al. 2005). In addition, spadefoots are unique among North American amphibians in

their ability to dig their own underground retreats. Combined with their apparent philopatry for relatively limited breeding sites over much of their range, *Spea* would seemingly have a high potential for geographically localized lineage diversification that is at odds with their current range sizes and lack of recognized within-species taxonomic diversity. To date, two phylogenetic studies using allozymes (Wiens and Titus 1991) and mitochondrial DNA (García-París et al. 2003) each recovered apparent non-monophyly within nominal *Spea* species (*S. intermontana* and *S. hammondi*, respectively) based on limited sampling at distant points within the species' ranges, but these taxa have yet to receive more thorough range-wide genetic examination.

Among *Spea*, the western spadefoot (*S. hammondi*) represents an immediate conservation concern. California has long identified *S. hammondi* as a Species of Special Concern (Jennings and Hayes 1994; Thomson et al. 2016). Recently, the U.S. Fish and Wildlife Service, after a preliminary review of evidence, initiated a formal review for listing under the Endangered Species Act. The bleak status of *S. hammondi* relative to its congeners is based largely on past and projected habitat loss (U.S. Fish and Wildlife Service 2005; Thomson et al. 2016). The rapid increase in commercial, residential, and especially agricultural development in 20th century California has led to an estimated 30% loss of *S. hammondi* native habitat in northern and central California and 80% in southern California (Fisher and Shaffer 1996; Thomson et al. 2016). In addition to ongoing habitat loss, climate change has significantly increased drought risk in the state (Diffenbaugh et al. 2015), leading to increased concern over the reliability of ephemeral vernal pool breeding habitat over much of the species' range.

The existence of two phylogeographic lineages within *S. hammondi* suggested by mitochondrial DNA (García-París et al. 2003), is consistent with the frequent observation of such lineages in other California amphibians and reptiles (Rissler et al. 2006) and emphasizes the

importance of evaluating ESUs for appropriate, lineage-specific conservation of *S. hammondi*. To do so, it is critical to determine the primary units of conservation with more thorough geographic and genetic sampling than has been done previously (Wiens and Titus 1991; García-París et al. 2003). To address questions of intraspecific variation within *S. hammondi*, we analyzed nuclear and mitochondrial phylogenetic relationships within *Spea* and intraspecific genetic clustering within *S. hammondi* to answer three primary questions. First, is *S. hammondi* non-monophyletic as suggested by previous mtDNA (García-París et al. 2003) analysis? Second, does *S. hammondi* show the genetic break into Northern and Southern California lineages across the Transverse Ranges that has been frequently recovered in other amphibians and reptiles (Rissler et al. 2006)? Third, are genetically recovered lineages of *S. hammondi* also distinct ecologically as might be predicted based on their different, isolated geographic ranges? We discuss and update the conservation status of *S. hammondi* based on the combined results from these genetic and environmental niche modeling analyses.

Methods

Taxon and genetic sampling

We targeted six genes for the four *Spea* species (*S. hammondi*, *S. bombifrons*, *S. intermontana*, and *S. multiplicata*), including five nuclear protein-coding (nuDNA: AKAP9, NTF3, RAG1, Rhod1, and SIA) and one mitochondrial (mtDNA: ND2) in this study. Tissue samples included whole tadpoles (from small individuals), tadpole tail snips (from larger individuals) and adult toe clips; most samples collected by us were immediately released at the point of capture. We extracted DNA using a salt extraction protocol and sequenced all PCR products in both directions on ABI 3730 sequencing platforms at the University of California,

Davis, Division of Biological Sciences sequencing facility. We aligned sequences using MUSCLE (Edgar 2004) in Geneious 6.1 (Kearse et al. 2012) and checked for nucleotide ambiguities. We used the implementation of PHASE (Stephens et al. 2001) in DnaSP 5.1 (Librado and Rozas 2009) with default settings to generate haplotypes for the nuclear sequences. Raw sequences are available on GenBank (MG137482 – MG138150). Detailed sample locality information is available online in Supplementary Table 1.

Before proceeding, we independently confirmed *Spea* as monophyletic (PP>0.994) in each of four gene trees (ND2, RAG1, Rhod1, and SIA – based on GenBank sequence availability for appropriate outgroups) (gene trees available in online supplement) generated with MrBayes 3.2.6 (Ronquist et al. 2012) via CIPRES (Miller et al. 2010). We used outgroups from other genera in the superfamily Pelobatoidea (*Scaphiopus* [sister genus to *Spea*], *Pelobates*, and *Pelodytes*) and rooted using *Leiopelma* or *Xenopus* (based on GenBank sequence availability). In order to continue with a complete matrix of the five nuclear genes we targeted for *Spea*, and because *S. hammondi* was the primary focus of the present study, our further analyses included data only from *Spea* species.

Phylogenetic Analyses of Spea taxon sets

We generated sequences for the remaining *Spea* species (*S. bombifrons*, *S. intermontana*, and *S. multiplicata*) to assess the monophyly of *S. hammondi* and to test existing phylogenetic hypotheses among these four species (Titus and Wiens 1991, Garcia-Paris et al. 2003). We used jModelTest2 (Darriba et al. 2012) via CIPRES (Miller et al. 2010) to determine the appropriate substitution models for each gene in the phylogenetic analyses. We generated a mtDNA tree using MrBayes 3.2.6 (Ronquist et al. 2012) via CIPRES (Miller et al. 2010), using 10 million

MCMC generations of burn-in followed by 50 million MCMC generations. We checked for MCMC convergence using Tracer 1.6 (Rambaut and Drummond 2003).

For the nuDNA, we used the multispecies coalescent model implemented in StarBEAST2 (Ogilvie et al. 2017) in BEAST 2.4.8 (Bouckaert et al. 2014) to generate individual gene trees and a species tree. StarBEAST requires that individuals be assigned *a priori* to “taxon sets” assumed to be monophyletic, and then estimates a species tree with those taxon sets as terminals. Because the monophyly of several *Spea* species has been questioned, we used gene trees from an initial StarBEAST run to identify potential taxon sets. We assigned *S. hammondii* samples to two monophyletic taxon sets based on Bayesian genetic clustering results (see STRUCTURE results, below) and on their individual monophyly in the mtDNA analysis. *Spea intermontana* and *S. bombifrons* were each non-monophyletic in the mtDNA tree and in some nuDNA gene trees (online supplement) in our preliminary StarBEAST run, and we assigned each monophyletic sub-lineage of these species to its own taxon set, resulting in two sets for *S. bombifrons* and three for *S. intermontana*. We ran StarBEAST2 for 1 billion generations with 200 million burn-in and assessed model convergence using Tracer 1.6.

To aid in comparing the topologies of the mtDNA and nuDNA phylogenetic trees, we used the *cophylo* function in the package phytools 0.6-44 (Revell 2017) in R 3.4.3 to create a tanglegram (i.e. a cophylogeny) between the StarBEAST nuDNA tree and the MrBayes mtDNA tree to more clearly visualize discordance in the position of clades between the two trees.

Clustering of Spea hammondii

We used the Bayesian assignment algorithm implemented in STRUCTURE 2.3.4 (Falush et al. 2003) to estimate the number of genetic clusters (K) within *S. hammondii*, using an

admixture ancestry model with correlated allele frequencies and K values between 1 and 10 for 1 million MCMC generations after 100,000 burn-in. We repeated this analysis independently five times and used the Evanno delta-K method (Evanno et al. 2005) in Structure Harvester (Earl and vonHoldt 2012) to determine the optimal value of K. To visualize cluster membership, we used CLUMPAK (Kopelman et al. 2015) to generate bar plots and ArcMap 10.4 to plot ancestry coefficients on a map. As an independent check of the Bayesian clustering results, we also used snapclust (Beugin et al. 2018) in the R package adegenet 2.1.1 (Jombart 2008; Jombart et al. 2018). Snapclust assesses clustering based on a combination of modeling Hardy-Weinberg equilibrium and maximum likelihood estimation based on the Expectation-Maximization algorithm, and it utilizes goodness-of-fit criteria to determine the optimal number of genetic clusters. To determine genetic clusters for the mtDNA sequences, we used Spatial Analysis of Molecular Variance (SAMOVA) (Dupanloup et al. 2002) implemented in SPADS 1.0 (Dellicour and Mardulyn 2014). SAMOVA uses simulated annealing to determine optimal clustering by maximizing differentiation among clusters, based on Φ_{CT} . We examined spatial distribution of SAMOVA ND2 clusters in ArcMap 10.4.

Niche Modeling

While our phylogenetic and clustering methods analyzed genetic differentiation, we used environmental niche modeling (ENM) to quantify ecological differentiation that would be consistent with functional or physiological distinctness. We used Maxent 3.3.3k (Phillips and Dudík 2008) to produce ENMs for (1) all *S. hammondii* sampling localities and for (2) localities from each of the inferred “North” and “South” genetic clusters which are completely allopatric. For sample inputs for Maxent, we first compiled presence points downloaded from GBIF and

then included the localities for our genetic samples. We then used the *gridSample* function in the R package *dismo* (Hijmans et al. 2015) to reduce the presence points to one per 0.1-degree grid cell to reduce effects of spatial sampling bias (Boria et al. 2014; Varela et al. 2014). This subsampling procedure left us with 68 sample localities from the North cluster and 49 from the South. We used 10,000 background points randomly sampled over the area of the input raster maps, and 30 arcsecond resolution rasters of bioclimatic variables 1-19 from Worldclim 1.4 (Hijmans et al. 2005) plus elevation and slope as predictors. We ran five replicates per run using cross-validation.

Given that the genetic data showed a clear difference between northern and southern *S. hammondii*, we used ENMTools (Warren 2017) in R 3.3.0 to conduct niche identity tests (Warren et al. 2008) to test the genetic results with ecological data. The niche identity test compares environmental conditions at presence points of the two clusters being tested while accounting for the environmental conditions available to the clusters in the vicinity of presence points (the “background”). For example, if the distribution of conditions at presence points of the two clusters do not overlap while their available background conditions do, the niche identity test takes this as evidence of niche differentiation and support for ecological distinctiveness that developed in allopatry. To determine significance for the niche identity test, we ran 100 permutations in which sample labels were randomized to generate a null distribution of the identity statistics, Schoener’s D (Schoener 1968) and Warren’s I (Warren et al. 2008). If the empirical value of these statistics fell below the 95% confidence interval of the null distribution, we considered the niches of the two clusters to be significantly differentiated.

Results

Taxon and genetic sampling

We sequenced 95 individuals of *S. hammondii* (North) from 45 sites, 20 *S. hammondii* (South) from 12 sites, six *S. intermontana* from two sites, two *S. bombifrons* from two sites, and two *S. multiplicata* from one site. Sequence lengths in the final alignments were 850bp for AKAP9, 605bp for NTF3, 676bp for RAG1, 683bp for Rhod1, 406bp for SIA, and 967bp for ND2. For our initial phylogenetic runs, we included all individuals and used the resulting gene trees to identify taxon sets. For our final phylogenetic analyses, we retained 45 individuals that included all major clades (Table 1) and had sequence data for all five nuclear genes for the StarBEAST analyses, and we used the same 45 individuals for a final MrBayes analysis of the mtDNA (a MrBayes tree using all individuals with mtDNA data is available in the online supplement). For genetic analyses within *S. hammondii* alone, we retained 63 individuals with sequence data for all five nuclear genes and converted these sequences to haplotypes for clustering analysis. Across both analyses, our *S. hammondii* dataset includes representation of the species' range-wide distribution except for Baja California, Mexico (Table 1, Fig. 1.1, and online supplement).

Phylogenetic analyses

Nuclear DNA phylogeny

The nuDNA gene trees (found in online supplement) from the StarBEAST run show generally high support for our chosen taxon sets, supporting these groups as operational taxonomic units. Across the five gene trees, the clades representing the taxon sets for the species tree all had posterior probability greater than 0.7 (median PP: *S. hammondii* North [0.7489]; *S.*

hammondii South [0.9969]; *S. bombifrons* [0.9985]; *S. intermontana* California [1.00]; *S. intermontana* Oregon [0.9816]; *S. multiplicata* [1.00]).

In the nuDNA species tree, StarBEAST returned *S. hammondii*, *S. bombifrons*, and *S. intermontana* as a monophyletic group to the exclusion of *S. multiplicata* (PP=0.8882). *Spea hammondii* North and South formed a monophyletic group (PP=0.9994), and *S. intermontana* was polyphyletic, with the Oregon clade placed sister to *S. bombifrons* (PP=0.9519) and the California clade recovered as sister to *S. hammondii* (PP=0.9638). Given the geographical locations of these *S. intermontana* samples (Fig. 1.1) we consider it extremely unlikely that these placements are the result of field misidentification.

Mitochondrial DNA phylogeny

The majority rule consensus mtDNA (ND2) MrBayes phylogeny had high support at most nodes (Fig. 1.2B) and very high support for the same monophyletic taxon sets identified in the nuDNA analysis (mtDNA tree PP: *S. hammondii* North [1.00]; *S. hammondii* South [1.00]; *S. intermontana* California [1.00]; *S. intermontana* Oregon [0.995]; *S. bombifrons* [1.00]; *S. multiplicata* [1.00]; *S. bombifrons* [PP=1.00]; full mtDNA tree in online supplement). The mtDNA tree recovered *S. multiplicata* as sister to all other *Spea* (PP=1.00). However, in sharp contrast to the nuDNA tree, the mtDNA recovered *S. hammondii* North as sister to all *Spea* except *S. multiplicata* (PP=1.00), including *S. hammondii* South. Within that group, *Spea hammondii* South and *S. intermontana* California were recovered as sister groups with modest support (PP=0.879), with *S. intermontana* Oregon and *S. bombifrons* recovered as sequential sister groups.

Cluster analysis

Nuclear clusters

The delta-K method for the nuDNA haplotype STRUCTURE results recovered two clusters as the best supported partitioning of *S. hammondii* (Fig. 1.3A and online supplement). These genetic clusters are geographically coherent and split by the Transverse Ranges in Central California (Fig. 1.3). The two clusters indicate essentially no admixture, with only one occurrence of a Northern haplotype of AKAP9 appearing in the South cluster, and none in the opposite direction. Using the Bayesian Information Criterion (Schwarz 1978), snapclust found K=3 as the most likely number of genetic clusters. At K=2, snapclust recovered the same partitioning as that found by STRUCTURE, and at K=3 the third cluster identified by snapclust is a split within the Northern cluster corresponding roughly to the Salinas Valley in Central California (online supplement). Given that 1) STRUCTURE and snapclust independently recovered the identical K=2 North/South clusters, 2) the BIC values for K values from 2-5 are very similar in snapclust (online supplement), 3) STRUCTURE supported K=2 clusters as the optimal K, and 4) our data set consists of relatively few nuclear genes, we favor the relatively conservative clustering of K=2 for our nuclear data set rather than K=3 identified in snapclust..

Mitochondrial clusters

The differentiation (Φ_{CT}) among *S. hammondii* clusters based on SAMOVA plateaus at K=3 (online supplement), with one cluster comprising all South individuals and two comprising the North individuals (Fig. 1.3B), with one of these North clusters identifying a unique cluster along the Central Coast of California in Santa Barbara and San Luis Obispo counties (off-white points in Fig. 1.3B), not seen in either nuDNA clustering analysis. The SAMOVA results at K=2

show the same South group as at $K=3$, indicating the two North clusters are mitochondrially-identified subgroups of the North set of populations.

Niche modeling

We ran three different *S. hammondii* Maxent models under present climate conditions: samples from the genetic North and South clusters each predicting their own ranges, and the combined North and South predicting the entire species. Based on the area under the curve (AUC) for the receiver operating characteristic (ROC), we found higher model fit for Maxent niche models generated for the North (Fig. 1.4A) (mean AUC for five-fold cross-validation: 0.934 ± 0.008) and South (Fig. 1.4B) (mean AUC: 0.968 ± 0.012) *S. hammondii* clusters than when the two were modeled as a single unit (mean AUC: 0.913 ± 0.007) (map not shown). For North alone, Bio18 (precipitation of warmest quarter) had the highest percent contribution at 54%, while Bio6 (minimum temperature of coldest month) had the highest permutation importance at 37%. For South alone, Bio11 (mean temperature of coldest quarter) had the highest percent contribution and permutation importance at 31% and 49%, respectively. For North and South combined, Bio15 (precipitation seasonality) had the highest percent contribution at 38%, while Bio9 (mean temperature of driest quarter) had the highest permutation importance at 14%. The niche identity test found that the niches of the North and South were more different than expected based on available habitat, suggesting ecological differentiation between the two groups (Schoener's D, 0.263, $p < 0.05$; Warren's I, 0.57, $p < 0.05$).

Discussion

In this study, our primary goal was to use genetic clustering and ecological differentiation as a combined means of determining potential taxonomic and conservation units within *S. hammondii*. Although we also included representatives of the other three nominal *Spea* taxa, our geographic sampling is sparse, and we primarily focus on our results for *S. hammondii*. However, despite our sparse sampling of the other *Spea* species, we observe and note that: 1) *S. multiplicata* is recovered as sister to the other *Spea* by both nuclear and mitochondrial DNA; and 2) our current concepts of species boundaries in *S. bombifrons* and *S. intermontana* may require re-examination with range-wide sampling and genomic techniques.

With the demonstration that *S. hammondii* populations north and south of the Transverse Ranges are both genetically and ecologically distinct, we recommend that each cluster be considered its own conservation unit with potentially unique management needs. The strength of the genetic isolation suggests the two clusters may in fact represent distinct species. These findings also corroborate the mitochondrial non-monophyly of *S. hammondii* identified by García-Paris et al. (2003), although our nuclear genetic analyses indicate that *S. hammondii*, while genetically differentiated between north and south clusters, is a monophyletic entity.

Phylogeography of Spea hammondii

The identification of two distinct lineages within a nominal species, separated by the Transverse Ranges, follows a phylogeographic pattern seen in a number of taxa in California (Rissler et al. 2006; Feldman and Spicer 2006; Spinks et al. 2010; Gottscho 2016). The striking result when comparing our two genetic data sets is that the mitochondria of the southern lineage of *S. hammondii* are more closely related to those of California *S. intermontana* than they are to

the northern clade of *S. hammondii*, while northern and southern *S. hammondii* comprise a well-supported monophyletic group based on nuclear evidence. Furthermore, no mitochondrial haplotypes of southern *S. hammondii* (south of the Transverse Ranges) appear north of the Transverse Ranges, and two methods of clustering of nuclear data strongly support this north-south split. With the StarBEAST nuclear species tree and all five individual nuclear gene trees demonstrating a highly-supported monophyletic *S. hammondii*, we rule out incomplete lineage sorting as a reasonable explanation of the mitochondrial result. Rather, a relatively ancient admixture event between southern *S. hammondii* and California *S. intermontana* serves as the best explanation of the discordance, with a replacement of southern *S. hammondii* mitochondrial DNA by *S. intermontana* that occurred after northern and southern *S. hammondii* lineages were reproductively isolated. Such mitochondrial replacements between currently allopatric taxa have been documented in other species (Gompert et al. 2008; Spinks and Shaffer 2009; Marshall et al. 2011; Zieliński et al. 2013; Good et al. 2015; Leavitt et al. 2017), reflecting past climatic conditions when sympatry and interbreeding presumably occurred.

Our present-time Maxent analyses further corroborate the results of the genetic analysis, with the occupation of unique climatic niches indicating ecological differentiation between the North and South clades of *S. hammondii*. Additionally, the niche models revealed the Transverse Ranges as a barrier of unsuitable habitat separating the north and south clades of *S. hammondii* (Fig. 1.4). Our genetic data indicate essentially no gene flow between these clades, and to our knowledge no *Spea* sample from the Transverse ranges has ever been documented. No morphological or call data comparisons between these two *S. hammondii* clades have been conducted, but the combination of genetic and ecological evidence pointing to pronounced

isolation makes morphological, and potentially mating call analyses (e.g. Feinberg et al. 2014) a logical next step in determining the species status of the two clades.

Conservation of Spea hammondii

With *S. hammondii* being considered for listing under the Endangered Species Act, our results demonstrating two distinct lineages within the nominal species indicate that each lineage should be considered as an independent conservation unit. Spadefoots in each region face varying degrees of threats that likely merit lineage-specific conservation actions. For example, in densely populated coastal southern California, *S. hammondii* has lost a substantially greater proportion of its native habitat than in northern California (Jennings and Hayes 1994; Thomson et al. 2016), and depending on how climate change ultimately affects California, vernal pool duration may be more of an issue in the south as pools tend to be smaller and more isolated compared to the sometimes vast pools that occur in the grasslands of the Central Valley (Bauder and McMillan 1998; Morey 1998). Additionally, given that the two lineages having evolved in isolation and occupy distinct ecological environments, we strongly advise against translocations between North and South populations that could introduce genotypes poorly suited to survival in the other lineage's environment or that may lead to reduced hybrid fitness.

Cryptic species within Spea: Is Spea hammondii a single taxon?

The four-species taxonomy of *Spea* has remained mostly stable since the mid-20th century based on a combination of morphology, calls, and distributions (Tanner 1939; Kluge 1966). However, a molecular phylogenetic hypothesis based on allozymes (Wiens and Titus 1991) suggested that *S. intermontana* may be non-monophyletic, while a mitochondrial

phylogeny by García-París et al. (2003) indicated a similar result for *S. hammondii*. Currently recognized, nominal *Spea* species generally have relatively wide distributions: *S. bombifrons*, *S. intermontana*, and *S. multiplicata* all have distributions that cover substantial portions of the climatologically and topographically complex interior of western North America (Fig. 1.1). The current xeric conditions of much of this region, combined with a life history indicating generally strong philopatry (Buseck et al. 2005), suggests that the potential for cryptic diversity in *Spea* driven by reproductive isolation during climatic dry periods may be high. Although we do not recommend any taxonomic changes until additional, more informative nuclear data are available, our current results demonstrate that *S. hammondii*, although it has by far the smallest geographic range, likely comprises two lineages that may be defensible species under a metapopulation lineage species concept (de Queiroz 2007; Shaffer and Thomson 2007). The potential for additional cryptic species within the more widely-distributed *Spea* species seems even more likely. We view well-sampled, multi-locus phylogeographic analyses of these taxa as an important direction for future research.

Acknowledgments

We thank Robert Fisher, Phil Spinks, and past and present members of the Shaffer Lab for help with field collections and analysis, and the California Department of Fish and Wildlife for permits. Grants from the National Science Foundation (DEB 1257648), US Fish and Wildlife Service (F16PX02290), the Natural Communities Coalition of Orange County (16-12), UCLA Department of Ecology and Evolutionary Biology, UCLA La Kretz Center for California Conservation Science, and Sea and Sage Audubon provided support to KMN and HBS.

Figures

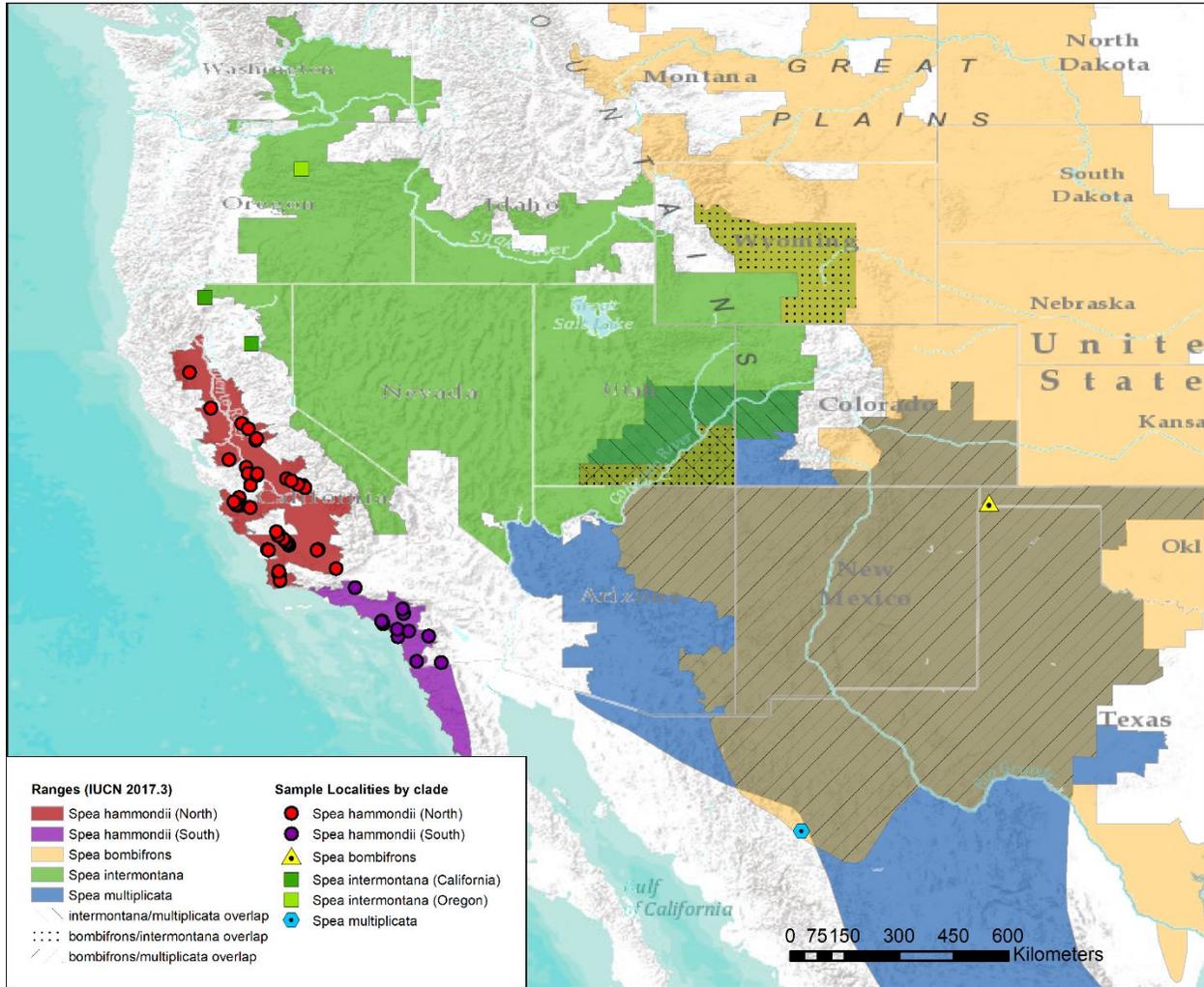


Fig. 1.1 Genetic sampling localities and species ranges (IUCN 2017) of the four nominal *Spea* species: *Spea hammondii* (shown as North [red] and South [purple]); *Spea bombifrons* (yellow); *Spea intermontana* (green); and *Spea multiplicata* (blue). Color figure available online

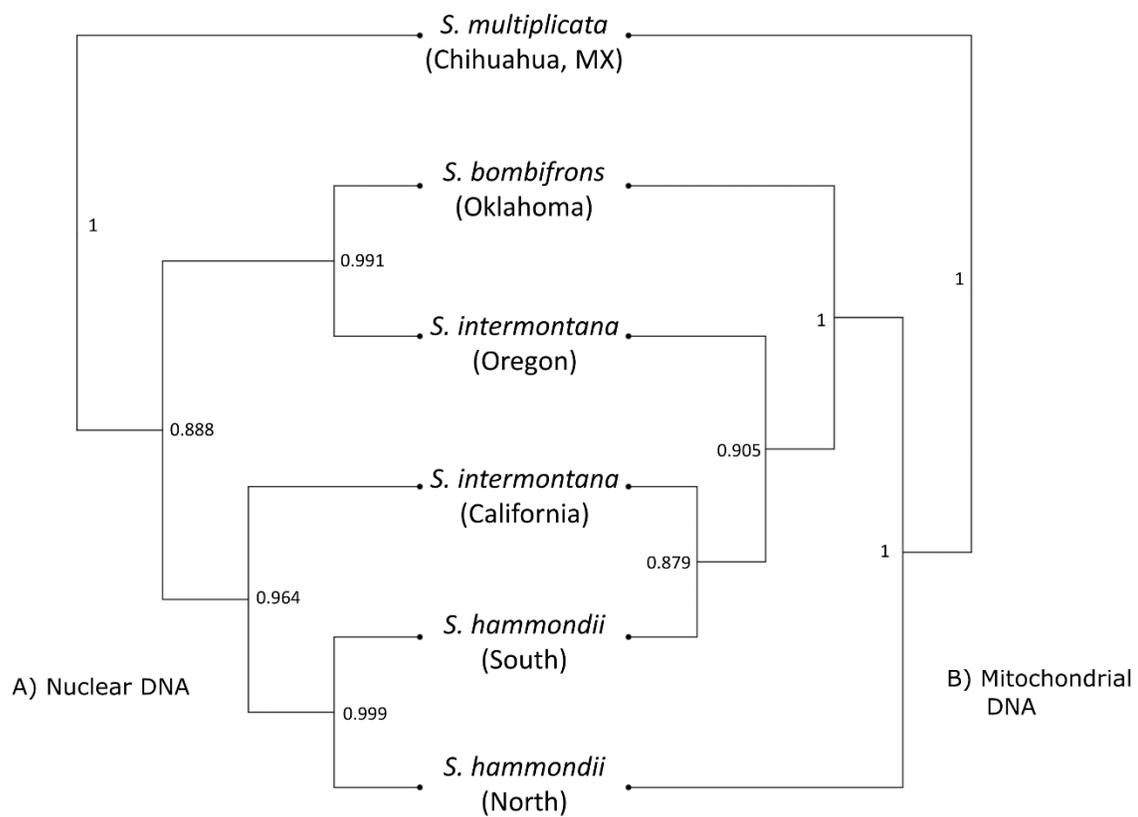


Fig. 1.2 Tanglegram or cophylogeny of nuclear species tree generated in StarBEAST2 (left, A) and mitochondrial tree generated in MrBayes 3.2.6 (right, B) for *Spea*, with Bayesian posterior probabilities at nodes

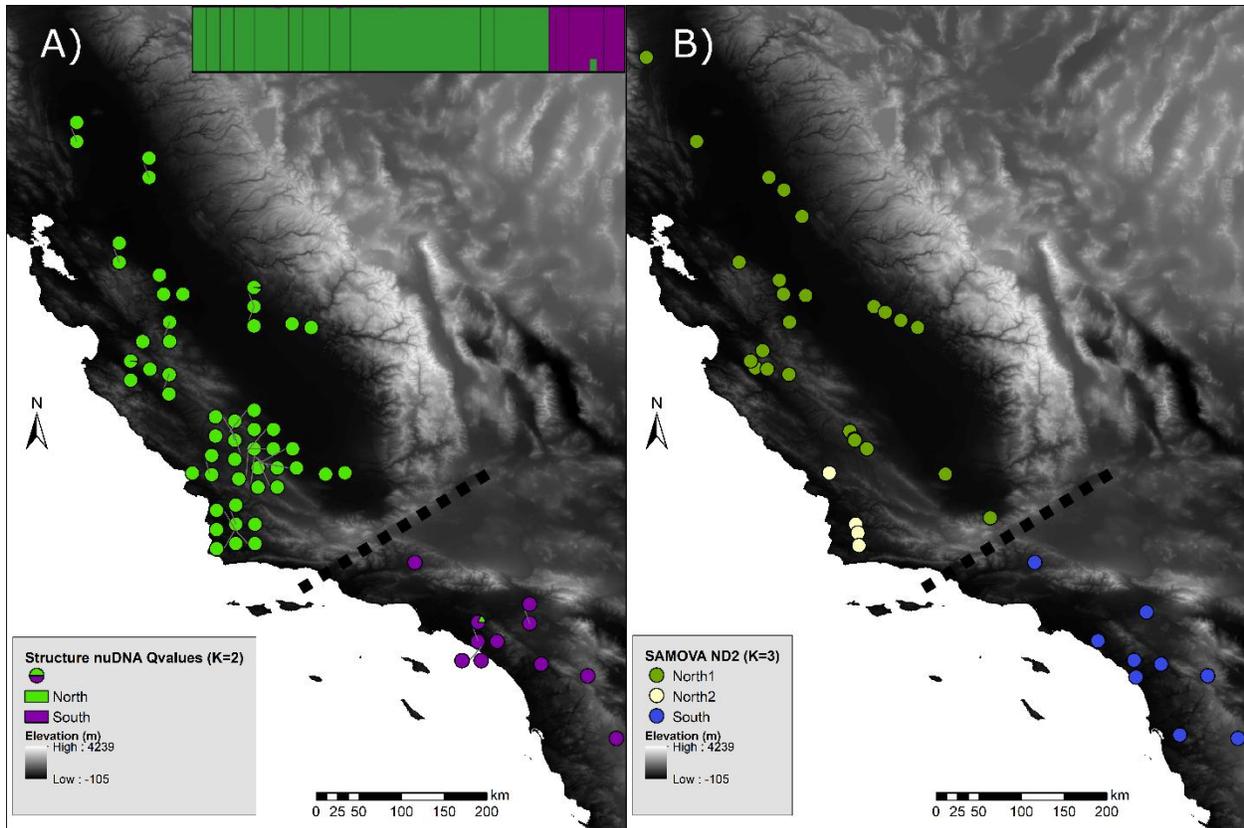


Fig. 1.3 STRUCTURE and SAMOVA genetic clusters for *S. hammondii*. A) Map of *S. hammondii* samples with pie charts displaying their STRUCTURE cluster membership proportions. Each pie chart shows proportions for a single individual, and pie charts are displaced from the actual geographic coordinate (connected by light gray lines) to allow display of all individuals at the same site. Inset shows STRUCTURE bar plots for K=2. B) Map of *S. hammondii* samples and their SAMOVA cluster membership; South mtDNA cluster corresponds to South nuDNA STRUCTURE cluster, and the two northern mtDNA clusters together correspond to the North nuDNA STRUCTURE cluster. Black dotted line on A) and B) shows phylogeographic break at the Transverse Ranges

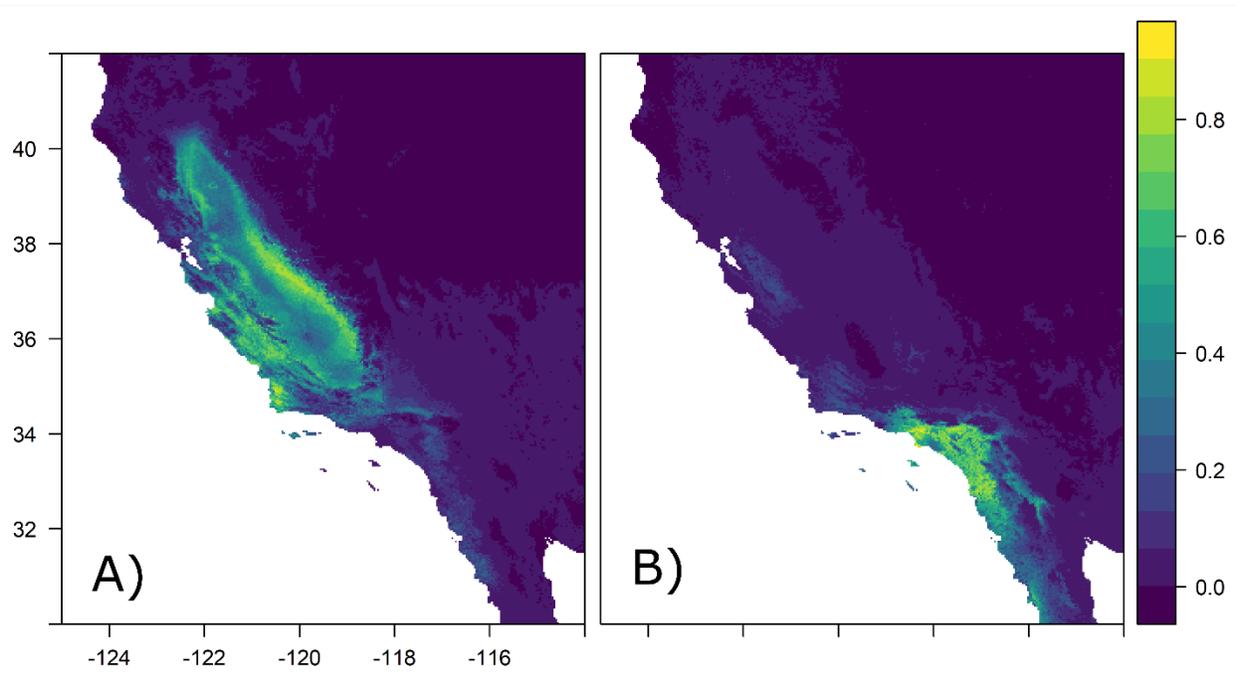


Fig. 1.4 Maxent niche models for *S. hammondii* based on current values of Bioclim variables, elevation, and slope. A) Niche model for North genetic cluster. B) Niche model for South genetic cluster. Scale corresponds to probability of occurrence. In both cases, the Transverse ranges show very low probabilities

Tables

Table 1 Summary table of individual and site representation in genetic analyses in this study

Species	Clade and Location	Number of individuals included (Number of unique sites represented)		
		Sequenced for any gene	StarBEAST	STRUCTURE
<i>Spea bombifrons</i>	Oklahoma	2 (2)	2 (2)	NA
<i>Spea hammondi</i>	California_South	20 (12)	10 (7)	11 (8)
<i>Spea hammondi</i>	California_North	95 (45)	26 (18)	52 (30)
<i>Spea intermontana</i>	Oregon	2 (1)	2 (1)	NA
<i>Spea intermontana</i>	California	4 (2)	3 (2)	NA
<i>Spea multiplicata</i>	Mexico	2 (1)	2 (1)	NA

Supplemental materials

Files for Chapter 1 Supplemental materials found on ProQuest page or on published manuscript's webpage at Conservation Genetics: Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018).

Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*). *Conservation Genetics*, 19(4), 937–946.

doi: 10.1007/s10592-018-1066-7

References

- Bauder ET, McMillan S (1998) Current distribution and historical extent of vernal pools in southern California and northern Baja California, Mexico. In: Ecology, Conservation and Management of Vernal Pool Ecosystems—Proceedings from a 1996 Conference. California Native Plant Society, Sacramento, CA
- Beugin M-P, Gayet T, Pontier D, et al (2018) A fast likelihood solution to the genetic clustering problem. *Methods Ecol Evol* 00:1–11. doi: 10.1111/2041-210X.12968
- Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol Model* 275:73–77. doi: 10.1016/j.ecolmodel.2013.12.012
- Bouckaert R, Heled J, Kühnert D, et al (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* 10:e1003537. doi: 10.1371/journal.pcbi.1003537
- Bryant D, Moulton V (2004) Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Mol Biol Evol* 21:255–265. doi: 10.1093/molbev/msh018
- Buseck R, Keinath D, Geraud M (2005) Species Assessment for Great Basin Spadefoot Toad (*Spea intermontana*) in Wyoming. U.S. Bureau of Land Management
- Castellanos-Morales G, Gutiérrez-Guerrero YT, Gámez N, Eguiarte LE (2016) Use of molecular and environmental analyses for integrated in situ and ex situ conservation: The case of the Mexican prairie dog. *Biol Conserv* 204:284–295. doi: 10.1016/j.biocon.2016.10.036
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol Evol* 15:290–295. doi: 10.1016/s0169-5347(00)01876-0
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. doi: 10.1038/nmeth.2109
- de Queiroz K (2007) Species Concepts and Species Delimitation. *Syst Biol* 56:879–886. doi: 10.1080/10635150701701083
- Dellicour S, Mardulyn P (2014) SPADS 1.0: a toolbox to perform spatial analyses on DNA sequence data sets. *Mol Ecol Resour* 14:647–651. doi: 10.1111/1755-0998.12200
- Diffenbaugh NS, Swain DL, Touma D (2015) Anthropogenic warming has increased drought risk in California. *Proc Natl Acad Sci* 112:3931–3936. doi: 10.1073/pnas.1422385112
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11:2571–2581. doi: 10.1046/j.1365-294X.2002.01650.x

- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. doi: 10.1007/s12686-011-9548-7
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164:1567–1587.
- Feinberg JA, Newman CE, Watkins-Colwell GJ, et al (2014) Cryptic Diversity in Metropolis: Confirmation of a New Leopard Frog Species (Anura: Ranidae) from New York City and Surrounding Atlantic Coast Regions. *PLOS ONE* 9:e108213. doi: 10.1371/journal.pone.0108213
- Feldman CR, Spicer GS (2006) Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. *Mol Ecol* 15:2201–2222. doi: 10.1111/j.1365-294X.2006.02930.x
- Fisher RN, Shaffer HB (1996) The decline of amphibians in California's Great Central Valley. *Conserv Biol* 10:1387–1397. doi: 10.1046/j.1523-1739.1996.10051387.x
- Fontanella FM, Feltrin N, Avila LJ, et al (2012) Early stages of divergence: phylogeography, climate modeling, and morphological differentiation in the South American lizard *Liolaemus petrophilus* (Squamata: Liolaemidae). *Ecol Evol* 2:792–808. doi: 10.1002/ece3.78
- Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol Ecol* 10:2741–2752. doi: 10.1046/j.0962-1083.2001.01411.x
- García-París M, Buchholz DR, Parra-Olea G (2003) Phylogenetic relationships of Pelobatoidea re-examined using mtDNA. *Mol Phylogenet Evol* 28:12–23. doi: 10.1016/S1055-7903(03)00036-8
- Gompert Z, Forister ML, Fordyce JA, Nice CC (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Mol Ecol* 17:5231–5244. doi: 10.1111/j.1365-294X.2008.03988.x
- Good JM, Vanderpool D, Keeble S, Bi K (2015) Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* 69:1961–1972. doi: 10.1111/evo.12712

- Gottscho AD (2016) Zoogeography of the San Andreas Fault system: Great Pacific Fracture Zones correspond with spatially concordant phylogeographic boundaries in western North America. *Biol Rev* 91:235–254. doi: 10.1111/brv.12167
- Gutiérrez-Tapia P, Palma RE (2016) Integrating phylogeography and species distribution models: cryptic distributional responses to past climate change in an endemic rodent from the central Chile hotspot. *Divers Distrib* 22:638–650. doi: 10.1111/ddi.12433
- Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978. doi: 10.1002/joc.1276
- Hijmans RJ, Phillips S, Elith JL and J (2017) *dismo: Species Distribution Modeling*. R package version 1.1-4. <https://CRAN.R-project.org/package=dismo>
- Hoisington-Lopez JL, Waits LP, Sullivan J (2012) Species limits and integrated taxonomy of the Idaho ground squirrel (*Urocitellus brunneus*): genetic and ecological differentiation. *J Mammal* 93:589–604. doi: 10.1644/11-MAMM-A-021.1
- Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73. doi: 10.1093/bioinformatics/14.1.68
- Ikeda DH, Max TL, Allan GJ, et al (2017) Genetically informed ecological niche models improve climate change predictions. *Glob Change Biol* 23:164–176. doi: 10.1111/gcb.13470
- IUCN 2017. The IUCN Red List of Threatened Species. Version 2017-3. <http://www.iucnredlist.org>
- Jennings MR, Hayes MP (1994) Amphibian and reptile species of special concern in California. California Department of Fish and Game, Inland Fisheries Division, Rancho Cordova, USA
- Jombart T (2008) *adeigenet: an R package for the multivariate analysis of genetic markers*. *Bioinformatics* 24:1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart T, Kamvar ZN, Collins C, et al (2018) *adeigenet: Exploratory Analysis of Genetic and Genomic Data*. R package version 2.1.1. <https://CRAN.R-project.org/package=adeigenet>
- Kalkvik HM, Stout IJ, Doonan TJ, Parkinson CL (2012) Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the *Peromyscus maniculatus* species group. *Ecography* 35:54–64. doi: 10.1111/j.1600-0587.2011.06994.x
- Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. doi: 10.1093/bioinformatics/bts199

- Kluge AG (1966) A new pelobatine frog from the Lower Miocene of South Dakota with a discussion of the evolution of the *Scaphiopus-Spea* complex. Los Angeles County Museum of Natural History, USA
- Kopelman NM, Mayzel J, Jakobsson M, et al (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191. doi: 10.1111/1755-0998.12387
- Leavitt DH, Marion AB, Hollingsworth BD, Reeder TW (2017) Multilocus phylogeny of alligator lizards (*Elgaria*, Anguillidae): Testing mtDNA introgression as the source of discordant molecular phylogenetic hypotheses. *Mol Phylogenet Evol* 110:104–121. doi: 10.1016/j.ympev.2017.02.010
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452. doi: 10.1093/bioinformatics/btp187
- Marshall DC, Hill KBR, Cooley JR, Simon C (2011) Hybridization, Mitochondrial DNA Phylogeography, and Prediction of the Early Stages of Reproductive Isolation: Lessons from New Zealand Cicadas (Genus *Kikihia*). *Syst Biol* 60:482–502. doi: 10.1093/sysbio/syr017
- May SE, Medley KA, Johnson SA, Hoffman EA (2011) Combining genetic structure and ecological niche modeling to establish units of conservation: A case study of an imperiled salamander. *Biol Conserv* 144:1441–1450. doi: 10.1016/j.biocon.2011.01.013
- McCormack JE, Zellmer AJ, Knowles LL (2010) Does Niche Divergence Accompany Allopatric Divergence in *Aphelocoma* Jays as Predicted Under Ecological Speciation?: Insights from Tests with Niche Models. *Evolution* 64:1231–1244. doi: 10.1111/j.1558-5646.2009.00900.x
- Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE), 2010. pp 1–8
- Morey SR (1998) Pool duration influences age and body mass at metamorphosis in the western spadefoot toad: implications for vernal pool conservation. In *Ecology, Conservation, and Management of Vernal Pool Ecosystems—Proceedings from a 1996 Conference*. California Native Plant Society, Sacramento, CA
- Moritz C, Funk V, Sakai AK (2002) Strategies to Protect Biological Diversity and the Evolutionary Processes That Sustain It. *Syst Biol* 51:238–254. doi: 10.1080/10635150252899752
- Ogilvie HA, Bouckaert RR, Drummond AJ (2017) StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of Substitution Rates. *Mol Biol Evol* 34:2101–2114. doi: 10.1093/molbev/msx126

- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175. doi: 10.1111/j.0906-7590.2008.5203.x
- Rambaut A, Drummond A (2003) Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. <http://tree.bio.ed.ac.uk/software/tracer/>.
- Revell LJ (2017) phytools: Phylogenetic Tools for Comparative Biology (and Other Things). R package version 0.6-44. <https://CRAN.R-project.org/package=phytools>
- Rissler LJ, Apodaca JJ (2007) Adding More Ecology into Species Delimitation: Ecological Niche Models and Phylogeography Help Define Cryptic Species in the Black Salamander (*Aneides flavipunctatus*). *Syst Biol* 56:924–942. doi: 10.1080/10635150701703063
- Rissler LJ, Hijmans RJ, Graham CH, et al (2006) Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *Am Nat* 167:655–666.
- Ronquist F, Teslenko M, Mark P van der, et al (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61:539–542. doi: 10.1093/sysbio/sys029
- Schoener TW (1968) The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology* 49:704–726. doi: 10.2307/1935534
- Schwarz G (1978) Estimating the Dimension of a Model. *Ann Stat* 6:461–464. doi: 10.1214/aos/1176344136
- Shaffer, HB, RC Thomson (2007) Delimiting species in recent radiations. *Syst Biol* 56:896–906. doi: 10.1080/10635150701772563
- Spinks PQ, Shaffer HB (2009) Conflicting Mitochondrial and Nuclear Phylogenies for the Widely Disjunct *Emys* (Testudines: Emydidae) Species Complex, and What They Tell Us about Biogeography and Hybridization. *Syst Biol* 58:1–20. doi: 10.1093/sysbio/syp005
- Spinks PQ, Thomson RC, Bradley Shaffer H (2010) Nuclear gene phylogeography reveals the historical legacy of an ancient inland sea on lineages of the western pond turtle, *Emys marmorata* in California. *Mol Ecol* 19:542–556. doi: 10.1111/j.1365-294X.2009.04451.x
- Stephens M, Smith NJ, Donnelly P (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. *Am J Hum Genet* 68:978–989. doi: 10.1086/319501
- Tanner VM (1939) A study of the genus *Scaphiopus*. *Gt Basin Nat* 1:3–20.
- Thomson RC, Wright AN, Shaffer HB (2016) California Amphibian and Reptile Species of Special Concern. UC Press

- U.S. Fish and Wildlife Service (2005) Recovery plan for vernal pool ecosystems of California and southern Oregon. Portland, Oregon
- Varela S, Anderson RP, García-Valdés R, Fernández-González F (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37:1084–1091. doi: 10.1111/j.1600-0587.2013.00441.x
- Warren D (2018) ENMTools R Package. GitHub repository:
<https://github.com/danlwarren/ENMTools>
- Warren DL, Glor RE, Turelli M (2008) Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. *Evolution* 62:2868–2883. doi: 10.1111/j.1558-5646.2008.00482.x
- Wiens JJ, Titus TA (1991) A phylogenetic analysis of *Spea* (Anura: Pelobatidae). *Herpetologica* 47:21–28. doi: 10.2307/3892812
- Zieliński P, Nadachowska-Brzyska K, Wielstra B, et al (2013) No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Mol Ecol* 22:1884–1903. doi: 10.1111/mec.12225

Chapter 2: Genome-wide SNPs reveal fine-scale patterns of population structure and gene flow in a threatened Southern California amphibian

Kevin Michael Neal

Abstract

Pond-breeding amphibians are under threat globally, with human-mediated habitat fragmentation as a leading factor in their declines. Patterns of fragmentation vary significantly at local scales, meaning assessment and management of these often-cryptic species require precise population and landscape genetic analysis specific to the species and populations in question. Populations of the western spadefoot (*Spea hammondi*) in Southern California bear the challenge of surviving in one of the most urbanized and fragmented landscapes on the planet and have lost up to 80% of their native habitat. In Orange County, ongoing restoration efforts targeting *S. hammondi* have involved the construction of artificial breeding ponds, but the performance of the introduced populations at these ponds relative to natural populations in the area has not been evaluated. Using thousands of genome-wide single-nucleotide polymorphisms in a landscape genetic framework, I characterized the population structure, genetic diversity, and functional connectivity of spadefoots in Orange County to guide ongoing and future management efforts. I identified two major genetic clusters with deeper substructure within clusters, with many individual ponds being genetically distinct. Estimates of historical gene flow and landscape resistance suggested ponds within and among genetic clusters were historically well-connected,

but the central low-resistance area has been largely destroyed by development, suggesting fragmentation has interrupted natural metapopulation dynamics. Resistance surfaces showed the existing artificial ponds were well-placed, being connected to natural populations by low-resistance corridors. Artificial ponds had a moderate level of genetic diversity, typically below the values of inland ponds but higher than those of coastal ponds. All ponds (natural and artificial) had extremely low estimates of effective population size, possibly due to a bottleneck caused by a recent multi-year drought. Management efforts should focus on maintaining gene flow among natural and artificial ponds by both manual exchange of spadefoots among ponds and construction of new ponds to bolster the existing pond network in the region.

Introduction

The decline of amphibians is a global crisis with habitat fragmentation as a major factor (Cushman, 2006; Hamer & McDonnell, 2008), but the degree of fragmentation and its effects on populations vary substantially at regional and local scales (Grant et al., 2016; Marsh & Trenham, 2001). This variability limits generalizability for conservation and management of amphibians at finer scales and demonstrates the importance of determining optimal management units.

Landscape genetics provides a useful framework for such fine-scale conservation efforts by elucidating historical isolation and functional connectivity among populations and the role of landscape features in shaping those patterns (Manel 2003, Storfer 2007). The role of environmental variables on the isolation and connectivity of populations can be inferred by comparing landscape resistance of environmental variables using spatially explicit models and data layers.

Pond-breeding amphibians are particularly attractive targets for landscape genetic analysis. Breeding pools function as discrete subunits for analysis, and the typically fixed positioning of ponds lends itself to relatively stable spatial sampling, maximizing the potential of localized, spatially explicit genetic models to explain genetic subdivision. Although adults are often exclusively nocturnal and are only active during heavy rainfall, tadpoles and larvae are often concentrated in large numbers at breeding pools and are easy to sample non-destructively for genetic material. For the same reasons, pond-breeding amphibians are uniquely vulnerable to local habitat disruption: their reliance on often ephemeral aquatic breeding sites and generally limited dispersal abilities means the destruction or disconnection of only a few key ponds can reduce functional connectivity amidst an unsuitable habitat matrix with an absence of stepping stones (Unglaub, Steinfartz, Drechsler, & Schmidt, 2015; Willson & Hopkins, 2013; Zamberletti,

Zaffaroni, Accatino, Creed, & De Michele, 2018; see Gabrielsen et al 2013 and refs therein as well), potentially leading to a collapse of preexisting metapopulation dynamics as gene flow and recolonization are disrupted. Identifying, maintaining, and restoring those levels of connectivity is a key way in which landscape genetics can and should inform amphibian conservation biology, particularly at the fine-scale level where most conservation actions occur (McCartney-Melstad, Vu and Shaffer, 2018).

Southern California is host to a number of sensitive, threatened, and endangered amphibians. Among them, the western spadefoot *Spea hammondi* stands out as a key species for management. Vernal pools utilized by the species occur in grasslands, coastal sage scrub, oak woodlands, and chaparral, and these areas have been decimated by urban and agricultural development. The western spadefoot is now extirpated across much of its Southern California range (Thomson, Wright, & Shaffer, 2016). Noting this dramatic decline, the California Department of Fish and Wildlife identified the western spadefoot as a Species of Special Concern (SSC) in both its original 1994 (Jennings & Hayes, 1994) and its current 2016 assessment (Thomson, Wright, & Shaffer, 2016). Additionally the species is currently under review by the U.S. Fish and Wildlife Service for listing under the Endangered Species Act (U.S. Fish and Wildlife Service, 2015). The western spadefoot is also a partially Covered Species under the Natural Community Conservation Plan (NCCP) and Habitat Conservation Plan (HCP) for Orange County (County of Orange, 1996), one of the primary strongholds for the species in Southern California: these conservation plans mandate active management of *S. hammondi* as if it were covered under the Federal and California Endangered Species Acts. However, management is complicated by the elusive life history of the species. As with other pond-breeding amphibians, direct observation of adult *S. hammondi* is rare: like all spadefoot species,

S. hammondi dig burrows up to 1 meter deep using the keratinous spade on each hindfoot, and they aestivate in these burrows until seasonal rains trigger their emergence (). The few existing observations of *S. hammondi* movements indicate very strong philopatry of adults to their breeding pools, but these observations are extremely limited and may not represent the actual species movement dynamics. Genetic sampling of tadpoles from breeding ponds has been suggested as the best way to determine patterns of gene flow and migration among ponds, and genetic assessment has been highlighted as an important conservation need (Thomson, Wright and Shaffer 2016).

My goal in this study was to provide a detailed landscape genomic analysis of *S. hammondi* that will simultaneously inform the basic landscape and molecular ecology of the species and help guide its management in a region with active and ongoing conservation efforts. The core of the study area is a roughly 38,000-acre open space co-managed for both wildlife protection and human recreation. The scale of the landscape presents a unique opportunity to focus landscape and population genomic inference at the spatial scale most appropriate for many amphibian species: within and among vernal pools separated by hundreds to a few thousand meters. To date, restoration efforts within Orange County targeting *S. hammondi* have included construction of several new seasonal pools in the northern portion of the county that were stocked with spadefoot tadpoles in 2005 and 2006, but genetic monitoring of these pools and surrounding, natural breeding sites has not yet occurred. I used restriction site-associated DNA sequencing (RAD-seq) to generate a large genome-wide dataset for essentially all *Spea hammondi* breeding sites in the region. RAD-seq is a reliable, widely-used, and inexpensive method of collecting genome-wide sequence data for population genomics (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Given the fine spatial scale of this study and the inability

of a previous effort with only 6 genes to resolve population-level relationships in *S. hammondi* (Neal, Johnson, & Shaffer, 2018), I sequenced thousands of markers to maximize precision and resolution of my genetic assessment (McCartney-Melstad, Vu, & Shaffer, 2018).

I analyzed the genomic data to address several questions: 1) Do spadefoot populations at this spatial scale segregate into distinct genetic clusters? 2) How do genetic diversity and effective population size vary among ponds and genetic clusters, and do they vary with pond surface area? 3) What are the genetic distances among ponds, and do genetic distances vary with geographic or environmental distance? 4) Are recently constructed artificial ponds performing adequately in preserving genetic diversity in the region? I then synthesize my findings to make specific recommendations for the management of the western spadefoot across this critical conservation landscape.

Methods

Study area and sampling

The study area comprises a roughly 38,000-acre open space in central Orange County, California including the eastern portion of the Los Angeles (LA) Basin, a relatively flat lowland region that includes western Orange County. Within the study region, the LA Basin is bounded by the coastal San Joaquin Hills to the southwest and the inland Santa Ana Mountains to the east/northeast. Historically spadefoots presumably occurred across the Los Angeles Basin, but the area is now urbanized to the extent that spadefoot localities occur almost exclusively on isolated upland ridgetops adjacent to the developed lowlands.

I sampled tadpoles at 26 sites within Orange County (Fig. 2.1) between 2015 and 2017, with the majority collected in 2017. Except for the constructed ponds at Irvine Mesa (IM) and

Shoestring Canyon (SHOE), sampled ponds are assumed to have been colonized naturally, even if some sites are man-made reservoirs or stock ponds. I refer to IM and SHOE as “artificial” or “constructed”, and naturally-colonized sites as “natural” or “native.” Most localities sampled for this study are situated on protected lands collectively known as the Nature Reserve of Orange County, managed by a consortium of public and private landowners. Sampling targeted known localities plus potential new sites based on Google Earth satellite imagery. Tadpoles were sampled using dipnets at different locations along the perimeter of each pool. For large tadpoles, roughly 1 cm of the tail tip was taken and the tadpole released at point of capture; for smaller tadpoles, whole individuals were sacrificed. Tissues were stored in 95% ethanol in the field and then in -20°C freezers until DNA extraction. I also utilized toes, liver, or muscle tissue from adults collected serendipitously, dating back to 1994. All collections were conducted via IACUC-approved protocols.

DNA extraction and sequencing

Genomic DNA was extracted using either a standard salt extraction protocol or the Kingfisher extraction protocol (Thermoscientific), followed by the 3RAD RAD-seq protocol (Bayona-Vásquez et al., 2019) to isolate genome-wide fragments of DNA for sequencing. Briefly, 3RAD adds a third enzyme to ddRAD (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) that cuts adapter dimers. I used Illumina iTru adapters as well as custom internal adapters (Integrated DNA Technologies) for multiplexing, and I used PstI-HF (dimer cutter), NsiI-HF (rare cutter), and MspI (common cutter) (New England BioLabs) for the three restriction enzymes. I pooled libraries with 200-300 ng of DNA per individual and used BluePippin (Sage Science) to size select 350-450 base pair fragments. Pooled libraries were sent to the Vincent J.

Coates Genomics Sequencing Laboratory at University of California, Berkeley, for sequencing on an Illumina HiSeq 4000 with either PE150, SR100, or SR150, although I used only the higher-quality first reads (R1's) for downstream protocols.

DNA post-processing

I first trimmed the internal adapters from the demultiplexed sequences using cutadapt (Martin, 2011). As an additional filtering step, I used fastp (Chen, Zhou, Chen, & Gu, 2018) to filter reads below 50% complexity (percentage of bases that are different from the next base in a sequence) and with a mean Phred quality score below Q30 (base call accuracy of 99.9%). To capture the highest-quality portions at the start of each read and because of differences of read lengths among SR100 and SR150/PE150 sequencing runs, I trimmed every read to 60 basepairs and removed reads less than 30 basepairs in length. I then used ipyrad 0.7.29 (D. Eaton, 2015/2019; D. A. Eaton, 2014) to assemble the RAD loci. I ran ipyrad using a reference draft genome sequenced from a *S. hammondii* metamorph from Santa Barbara County, California; this individual is a part of the northern clade of *S. hammondii*, the sister clade of the southern *S. hammondii* in the current study (Neal, Johnson, & Shaffer, 2018). The genome was sequenced by 10X Genomics, assembled using Supernova (10X Genomics) and analyzed using BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) to inspect the quality of the assembly (assembly results in Appendix 2-1).

From the ipyrad VCF file output, I used vcftools 1.1.5 (Danecek et al., 2011) to produce several genetic datasets for Orange County spadefoots with different subsets of individuals. OC_all_208 is the complete dataset including a maximum of 10 individuals from all sampled ponds. OC_reduced_148 reduced the sampling of the 9 artificial ponds from 79 individuals to 19

(using 2-3 individuals from every artificial pond); given that the spadefoots in all artificial ponds were derived from salvaged tadpoles from a single cluster of donor ponds, this reduction was aimed at more equal sampling parity between native and artificial ponds. OC_native_129 removed individuals from artificial ponds entirely to explore natural spatial connectivity as it occurred before the artificial ponds were created. For each dataset I removed singletons to filter noise from the dataset (Linck & Battey, 2019), included only biallelic SNPs, and retained only the first SNP in each RAD locus to reduce the effects of physical linkage among markers. I tested missing data thresholds from 20% to 75% but found no major differences in results, so I settled on allowing up to 75% missing data to maximize the size of the data matrices for pairwise and intrapopulation statistics. Parameters used in ipyrad and vcftools to produce each final VCF dataset can be found in the supplemental information.

Data analysis

Population structure

I evaluated population genetic structure of the OC_reduced_148 dataset using the Bayesian clustering method FastStructure 1.0 (Raj, Stephens, & Pritchard, 2014). I first converted the vcf to a structure file using PGDSpider (Lischer & Excoffier, 2012). I ran FastStructure using the simple prior with K values from 1 to 20, repeating each K value 10 times with different random seeds. I used the included chooseK.py or StructureSelector (Li & Liu, 2018) and selected the optimal K value based on the value that maximizes the marginal likelihood. To identify hierarchical genetic structure, I split the dataset into its component clusters based on the optimal K and re-ran FastStructure using K from 1 to 10 and 10 random seeds, stopping when K=1 had the highest marginal likelihood or when K equaled the number of

sampling localities in the subset. I also ran FastStructure with 10 cross-validation iterations with K from 1 to 20 to examine structure at K values above the optimal value.

To compare the model-based FastStructure against a non-model method, I ran a principle components analysis (PCA) in R 3.5.1 (R Core Team, 2018) using `adeigenet::glPca` (Jombart, 2008; Jombart & Ahmed, 2011; Jombart et al., 2018) to visualize genetic structure along the first eight principle components. I additionally ran `fineRADstructure` (Malinsky, Trucchi, Lawson, & Falush, 2018), an extension of `finestructure` (Lawson, Hellenthal, Myers, & Falush, 2012) for RAD-seq data, as an alternative examination of hierarchical genetic structure. `FineRADstructure` utilizes full haplotype sequence data in the `alleles.loci` file generated by `ipyrad` to create a matrix of individual pairwise coancestry. I ran `fineRADstructure` to assign individuals to populations with 1,000,000 Markov Chain Monte Carlo generations after 1,000,000 burn-in and constructed a population tree with 100,000 tree-building iterations. I calculated global F statistics with `hierfstat::fstat` (Goudet & Jombart, 2015) and conducted an analysis of molecular variance (AMOVA) in Arlequin 3.5 (Excoffier & Lischer, 2010) using default settings to partition genetic variation at hierarchical levels (individuals, breeding sites, FastStructure genetic clusters, and all samples combined).

Genetic diversity and effective population size and their relationship to pond size

Pond size has been implicated as an important determinant of population size in pond breeding salamanders (McCartney-Melstad, Vu, & Shaffer, 2018; Wang, Johnson, Johnson, & Shaffer, 2011), and I tested for this same potential effect at spadefoot breeding sites. I ran the `OC_all_208` VCF in the “populations” module of Stacks 2.3d (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to calculate pond-level and genetic cluster-level observed

heterozygosity (H_o), expected heterozygosity (H_E), nucleotide diversity (π), and inbreeding coefficient (F_{IS}), including invariant sites in the calculations. I required that loci be present in at least two populations, be present in at least 50% of individuals within a population and in 25% of all individuals, and minor allele count of ≥ 2 to remove singletons. In addition to these diversity measures, I used NeEstimator 2.1 (Do et al., 2014; Waples, Larson, & Waples, 2016) to calculate effective population size (N_e) for individual ponds using the linkage disequilibrium method (LD N_e) without singletons, and effective number of breeders (N_{eb}) using the molecular coancestry method (Nomura, 2008). I used the R package radiator (Gosselin, 2017/2019) to convert the VCF to GENEPOP format as input for NeEstimator. I used the ruler tool in Google Earth to estimate pond area and perimeter from historical satellite imagery and topography to make low- and high-end estimates for each pond. I then used `corrplot::cor.mtest` to calculate and test the significance of two non-parametric correlation coefficients (Spearman's rho and Kendall's tau) between measures of genetic diversity and N_e with pond area and perimeter.

Genetic distance and gene flow

I used `hierfstat::genet.dist` to calculate pairwise Nei's D_a (Nei, Tajima, & Tatenno, 1983; Takezaki & Nei, 1996) between individuals, sites, and clusters; I also used `hierfstat::pairwise.WCfst` to calculate pairwise Weir and Cockerham's F_{ST} (Weir & Cockerham, 1984) between sites and clusters, and `mmod::pairwise_D` (D. Winter, Green, Kamvar, & Gosselin, 2017; D. J. Winter, 2012) to calculate pairwise Jost's D (Jost, 2008; Jost et al., 2018). I additionally used `StAMPP::stampPhylip` (L. W. Pembleton, 2017; Luke W. Pembleton, Cogan, & Forster, 2013) to convert the Jost's D and Nei's D_a distance matrices to phylip distance files,

and visualized these as NeighborNet networks (Bryant & Moulton, 2004) in SplitsTree 4.14.8 (Huson & Bryant, 2006).

To examine relatively ancient admixture between sites, I used TreeMix 1.13 (Pickrell & Pritchard, 2012). TreeMix uses population allele frequencies to generate a tree and iteratively adds a given number of migration events as edges in the tree and tests for an increase in the model likelihood. I pooled individuals by site (and pooled artificial pools into as a single site), and used `dartR::gl2treemix` to convert the data as a `genlight` object into the appropriate input for TreeMix. I ran TreeMix using 10 different random seeds with 0 to 20 migration edges, and the SiZer method (Chaudhuri & Marron, 1999; Sonderegger, 2018) in the R package `OptM` (Fitak, 2019) to determine the optimal number of migration edges in the TreeMix results.

Landscape resistance

To quantify the role of the environment in hindering or facilitating gene flow on the landscape, I used ResistanceGA (Peterman, 2018). I chose ResistanceGA over other landscape resistance methods because of its relatively straightforward approach to the problem of parameterization of resistance surfaces. ResistanceGA takes a pairwise genetic distance matrix and a series of environmental rasters or surfaces as input and uses a genetic algorithm to transform the surfaces into resistance surfaces, using a linear mixed effect model to iteratively optimize the fit of resistance distance to genetic distance. The program returns a ranking of the optimized resistance surfaces. Using ResistanceGA removes the subjectivity inherent in using expert opinion to determine resistance values of an environmental variable and allows for exploration of a wide range of parameter space (Peterman, 2018; Peterman, Connette, Semlitsch, & Eggert, 2014). I calculated resistance distance between sites using

gdistance::commuteDistance, which determines distance as an average of a number of random walks between points on the environmental surface. I used pairwise Nei's D_a between the natural sites (OC_native_129 dataset) for the genetic distance and ran ResistanceGA using two sets of environmental variables. The first set included roughly 1 km (or 0.00833 decimal degrees) resolution bioclimatic layers from CHELSA 1.2 (Karger et al., 2017) and additional climatic and topographic layers from ENVIREM (Title & Bemmels, 2018). The second set included several soil and landcover layers from the 2011 National Landcover Database (Homer, Fry, & Barnes, 2012) and from SoilGrids (Hengl et al., 2017); these layers are much finer resolution (roughly 250 m, or 0.0020833 decimal degrees) to facilitate a fine scale analysis within Orange County. To enable comparison among rasters of all resolutions, I favored using marginal R^2 to rank surfaces as it provides a measure of absolute model fit across different datasets (Nakagawa & Schielzeth, 2013). I used Circuitscape (Shah & McRae, 2008) with the highest-ranked optimized resistance surface to visualize potential high-current corridors on the landscape.

Results

RAD-seq data assembly

The output of the OC_all_208 VCF file contained 6660 SNPs with up to 75% missing data per site. Per individual, the mean percent missingness was 34.28% (median 35.30%, range 16.30% to 50.86%). Mean depth per individual was 23.27 (median: 21.11; range: 9.182 to 59.93). The reduced dataset, OC_reduced_148, contained 6864 SNPs, with more SNPs than OC_all_208 presumably because I subsampled individuals from artificial sites based on those with the lowest individual missingness. OC_reduced_148 had a mean missingness per individual

of 35.24%, median of 35.50%, and range of 16.11% to 50.68%. Mean depth per individual in OC_reduced_148 was 21.06 (median of 19.41, range of 9.659 to 55.63).

Genetic structure

Run on OC_reduced_148, all ten replicates of FastStructure returned a maximum marginal likelihood at $K=2$, splitting Coastal Orange County (“Coast” cluster) from “Inland” sites (Fig. 2.2). FastStructure also reported the number of model components used to explain the structure in the data as 2 at 7 out of 10 replicates, and 3 at 3 out of 10. There was little admixture at $K=2$ at most sites; however, individuals at Sand Canyon Reservoir (SANCAN) site showed a roughly even mix of the two genetic clusters. Tenaja Pond (TENAJA), a pond directly adjacent to the Sand Canyon Reservoir, grouped with Coast. Both SANCAN and TENAJA sites are located adjacent to the LA Basin on the northern edge of the San Joaquin Hills, roughly 10 km from the other Coast sites. At $K=3$ (non-hierarchical), TENAJA was split from the other Coast sites in all replicates. Running FastStructure separately on the Coast+SANCAN+TENAJA and Inland clusters, $K=1$ had the highest marginal likelihood for all ten replicates for the Inland cluster; $K=2$ has the highest for all ten replicates of the Coast+SANCAN+TENAJA cluster, splitting off TENAJA as with the non-hierarchical dataset at $K=3$. SANCAN remained admixed but the majority of admixture for most individuals came from TENAJA, and due to their proximity to one another (satellite imagery showed the ponds connected during high rain years) I considered TENAJA and SANCAN together as their own subcluster (“Santena”) with no further subdivision. In the remaining Coast cluster, all ten replicates reported $K=1$ as the highest marginal likelihood, suggesting three major clusters: Inland, Santena, and Coast (Fig. 2.2). IM 01-14 (8 total pools) and SHOE—the artificial ponds—clustered with Inland. Although

FastStructure reported that higher values of K explain the structure in the data in the Coast and Inland subsets (between 3 and 5 for all replicates), marginal likelihoods dropped steeply and the modes of these replicates varied substantially at the higher K values, with sites moving in and out of different clusters and often fewer than K clusters appearing in a given replicate (Fig. 2.S3).

The PCA (Fig. 2.3) revealed additional structure than was indicated by the optimal FastStructure results. While discriminating between the three hierarchical FastStructure clusters on PC1, additional PCs indicated separation among individual sites. PC1 (6.42% of variance explained) separated Coast from Inland, with Santena falling in between the two and in particular SANCAN (the admixed site) stretched between the Coast and Inland clusters. PC2 (4.36% of variance) explains variation among the Inland sites, with STARR far removed from the rest. Inland sites SADDLE, LAUR, and THOM overlapped on PC1 and 2, corresponding to their geographic proximity to one another in southwestern Orange County. Northern Inland sites LOMA, LIME, GREAT, and FREM also overlapped, and the artificial ponds at IMSHOE clustered with these northern sites, corresponding to the geographic proximity of the IMSHOE source ponds to FREM, GREAT, and LOMA. PC3 (3.65% of variance) separated TENA, and PC4 (3.36%) further characterized genetic variation within the Inland cluster by distinguishing LOMA and LIME separately from the other sites. PCs 5-8 captured additional genetic variation within the Inland and Coast clusters; PC5 (2.88%) separated IMSHOE and FREM from the remainder, PC6 (2.51%) separated BBEND and CCSP1 individually, PC7 distinguished THOM, LAUR, SANCAN, TORO, SADDLE, FREM, and GREAT, while PC8 separated THOM, LAUR, SANCAN, TORO, and SADDLE, as well as FREM and GREAT. Ultimately, most

breeding sites, even those in very close geographic proximity, showed some genetic isolation, suggesting relatively restricted movement among sites and strong breeding philopatry.

FastStructure (non-hierarchical) and fineRADstructure, like PCA, confirmed these results, and showed substructure beyond the optimal FastStructure results using marginal likelihoods. FastStructure using 10 cross-validation iterations again showed marginal likelihood highest at $K=2$ with Coast+Santena split from Inland. Cross-validation error was lowest at $K=6$, although only five clusters are visible: Coast remained undifferentiated; LIME, STARR, and TENA split out into their own individual clusters; and SANCAN was admixed between Coast and the remaining undifferentiated Inland sites, rather than sharing TENA admixture (Fig. 2.S3). FineRADstructure identified most Inland sites as unique, cohesive clusters, with individuals within the same pond sharing more coancestry than with individuals at other sites. IMSHOE and Coast each formed independent clusters, but ponds within both were not readily distinguishable from one another. The population tree generated by fineRADstructure first split Coast from Inland+Santena. Within Inland+Santena, IMSHOE split off first, then STARR, THOM, and LAUR split off together, followed by the separation of SANCAN and TENAJA together from GREAT, FREM, SADDLE, TORO, LOMA, and LIME (Fig. 2.S4).

Genetic distance and gene flow

Considering the three FastStructure clusters (Inland, Coastal, Santena) as genetic units in the OC_reduced_148 dataset hierfstat returned a global F_{ST} of 0.111, F_{IT} of 0.278, and F_{IS} of 0.188. The OC_all_208 dataset, when individual pond breeding sites (including all artificial ponds) were treated populations, had a much higher F_{ST} and lower F_{IS} : global F_{ST} was 0.247, F_{IT} = 0.256, and F_{IS} = 0.012, consistent with the subtle, but consistent differentiation among ponds

in PC 2-8 and in the fineRADstructure results. In pairwise comparisons (Fig. 2.S1), I found lower F_{ST} values among the three clusters than among individual breeding sites. Pairwise F_{ST} among the three clusters was between 0.0789 and 0.142. Within the Coast cluster, pairwise F_{ST} among breeding pond sites ranged from 0.049 to 0.244 (mean: 0.142; median: 0.163). Within Inland (excluding TORO and SADDLE sites which had only one sample) the range was 0.119 to 0.37 (mean: 0.228; median: 0.218). Between the two sites of the Santena cluster, the pairwise F_{ST} was 0.214. Pairwise F_{ST} values among the artificial pools in IMSHOE range from 0.00453 between IM06 and IM07, to 0.297 between IM06 and IM12.

Hierarchical AMOVA results from Arlequin for OC_reduced_148 found F_{IS} of 0.0472; F_{SC} of 0.144; F_{CT} of 0.0130; and F_{IT} of 0.195. These AMOVA results indicated that a substantial amount of genetic variation occurred within individuals: 80.47% of variation occurred within individuals; 14.25% among sites within the three genetic clusters; 1.29% among clusters; and 3.98% among individuals within sites; all p-values were less than 0.0001.

NeighborNet networks of pairwise F_{ST} , Jost's D , and Nei's D_a all revealed similar patterns of genetic distance among the methods (Fig. 2.4, S4, S5), showing clear separation between Inland and Coastal, with Santena falling in between except in F_{ST} where it appeared within Inland. Reticulations in the networks were common among sites within genetic clusters. Notably, according to pairwise Jost's D and Nei's D_a , the artificial site SHOE appeared to share a high number of edges with natural Inland populations (Fig. 2.S5, S5), and SHOE is also the artificial site with the closest proximity to known natural sites. At higher values of K in FastStructure, SHOE also shows slight admixture with Inland sites, while the other artificial sites do not (Fig. 2.S3).

The topology of TreeMix trees was relatively unstable across replicates and number of migration edges. Considering the three FastStructure clusters, only Coast sites consistently form a monophyletic group. Although Inland is not always monophyletic, three groups within it consistently are: LIME+IMSHOE+FREM, STARR+THOM+LAUR, and SADDLE+TORO+GREAT. The SiZer method in OptM selected 8 as the optimal number of migration edges (Fig. 2.S7), but tree likelihood continued to increase through the maximum of 20 edges tested, suggesting regular historical migration between sites in Orange County.

Genetic diversity and effective population size

Among the three FastStructure genetic clusters, Inland (excluding artificial ponds) had the highest genetic diversity across measures, but also the highest F_{IS} value (Table 1). Coast had the lowest measures of genetic diversity—even lower than the artificial ponds (IMSHOE) when considered as a group—but had a slightly lower F_{IS} than IMSHOE (0.0213 versus 0.0257). At the site/pond level, natural sites within the Inland cluster tend to have higher values of genetic diversity ($\pi = 5.40E-4$) than Coast ($4.06E-4$), Santena ($4.36E-4$), and artificial sites ($4.20E-4$). Eight out of 17 natural sites with two or more individuals had slightly negative F_{IS} values, indicating that mild outbreeding may be taking place. Effective population size and effective number of breeders at sites were also generally low. Estimates of the effective number of breeders (N_{eb}) at breeding ponds based on the molecular coancestry method ranged from 1.2 to 12.2, while effective population size estimates based on the linkage disequilibrium method (LDN_e) ranged from 1.4 to 19.8, with three outliers: 62 (SHOE, an artificial pool), 64 (THOM) and 167 (LOMA). Given that these populations did not have high numbers of effective breeders (1.4, 1.2, 3.5), I view these outliers as likely statistical artifacts rather than large populations.

Across all sites combined, I found statistically significant correlations between genetic diversity and pond area, but no significant correlations with either N_e measure (Fig. 2.S2a). When I broke this down by breeding pool type, I found that among artificial sites, there was a strong and significant correlation between LDN_e and high-end estimates of pond area and perimeter (Fig. 2.S2b). Across natural sites, I found strong correlations between genetic diversity and pond area (Fig. 2.S2c). Coast sites had significant correlations of varying strength between genetic diversity and pond area, and, unexpectedly, strongly negative correlations between N_e and pond area. (Fig. 2.S2d). At Inland sites (excluding artificial pools), I found strong and significant correlations between genetic diversity measures and area (Fig. 2.S2e).

Landscape resistance

ResistanceGA results provided strong evidence of isolation by environment and distance. Using marginal R^2 (R^2_m) as the rank measure, ResistanceGA identified a transformation of depth to bedrock and slope (Fig. 2.5a) ranked highest among the environmental surfaces, explaining 77.0% ($R^2_m=0.770$) of the variance in pairwise genetic distance. Depth to bedrock alone ($R^2_m=0.761$) and landcover alone ($R^2_m=0.742$) ranked second and third, respectively (Table S1). For environmental surfaces available at both high-resolution and low-resolution, the high-resolution version always outperformed the low. Interestingly, the signal from depth to bedrock was strong enough that the low-resolution version ranked just behind the high-resolution version with a R^2_m of 0.741. Geographic distance alone ranked near the middle with a marginal R^2 of 0.562 (0.493 at low resolution), suggesting that isolation by distance is an important component of *Spea* landscape resistance. The depth to bedrock + slope surface tended to show the lowest resistance in areas where the spadefoots are now largely extirpated in the highly

developed lowlands of Orange County. Running current through the optimized depth to bedrock + slope resistance surface in Circuitscape confirmed low-lying valleys as generally the highest potential corridors between sample sites (Fig. 2.5b), consistent with the TreeMix and other results indicating historical gene flow across the flat, low-elevation central corridor separating Coastal and Inland populations.

Discussion

Although widely viewed as a key tool in the conservation biology toolkit, there is often a mismatch between the meaningful application of landscape genetic studies and the scale at which reasonable inferences can be drawn. The reason for this lies in the scale of most studies—evolutionary processes that are most recoverable with genetic approaches frequently occur at large spatial scales and across evolutionary time, while conservation management and planning occur on kilometer-level landscapes over recent time scales. In this study, I push the spatial and temporal boundaries of landscape genomics, and in so doing join a small set of studies that apply genome-level data sets across a small, kilometer-level landscape to make demographic and ecological inferences (McCartney-Melstad, Vu and Shaffer, 2018; see McCartney-Melstad and Shaffer, 2015 and Shaffer et al. 2015 for recent reviews). Even at this very limited spatial scale, my work reveals the presence of discrete population clusters that should function as management units, and more subtle population structure consistent with strong philopatry to individual breeding sites. Combined with small effective population sizes and severe habitat loss across the LA Basin, my work suggests that *Spea hammondi* in its Orange County stronghold is both at risk and a strong candidate for active management. Fortunately, constructed vernal pool breeding

sites appear to support relatively healthy populations, and I advocate for additional pond construction and upland restoration as reasonable strategies for future conservation actions.

Population structure

I identified two major genetic clusters in Orange County spadefoots, and these should comprise the primary management units for the species. A clear distinction exists between spadefoots in the coastal San Joaquin Hills of Orange County and those in the Los Angeles Basin and foothills of the Santa Ana Mountains. Based on marginal likelihood, the hierarchical FastStructure analyses revealed three clusters as the optimal clustering, but other methods, including FastStructure with higher values of K , PCA, fineRADstructure, and global and pairwise F -statistics all provided evidence of additional differentiation, sometimes among adjacent ponds.

Identifying pond-level structure is consistent with our understanding of spadefoot biology. Spadefoots are strongly philopatric, rarely moving more than a few hundred meters from their breeding pool (Baumberger, 2013). In contrast, a study of a congeneric spadefoot species, *S. multiplicata*, found it to have the lowest levels of genetic structure compared to two other pond-breeding amphibians in a floodplain in Arizona, suggesting spadefoots may actually be fairly vagile on flat, open terrain (Mims et al., 2014). The climate of Southern California tends to be more predictable and less extreme, however: while inland desert pools may be truly ephemeral, ponds in Southern California fill much more reliably in the rainy season and may select for spadefoot individuals that remain in place, while the substantial topographic and environmental complexity of the region produces strong resistance to gene flow.

Genetic diversity and effective population size

I found very low effective population sizes and numbers of effective breeders in sampled ponds—among the lowest recorded for pond-breeding amphibians, including several California species: red-legged frogs (*Rana draytonii*) (Richmond, Barr, Backlin, Vandergast, & Fisher, 2013), black toads (*Bufo exsul*) (Wang, 2009), and the federal- and state- listed California tiger salamanders (*Ambystoma californiense*) (Wang, Johnson, Johnson, & Shaffer, 2011)... I expected measures of genetic diversity and effective population size (LDN_e , N_{eb}) to be positively correlated with pond area. While I found significantly positive correlations with pond area for most measures of genetic diversity considered, I found no significant correlations between either measure of effective population size and pond area (Fig. 2.S2). I considered several possible reasons for this. First, even with thousands of SNPs, the N_e estimates are generally not very precise, especially for LD N_e . Parametric 95% confidence intervals for LDN_e estimates were typically reasonable, but jackknife 95% confidence intervals were often very large, and had an infinite upper bound for most sites. Jackknife 95% confidence intervals for molecular coancestry-estimated effective number of breeders (N_{eb}) were more reasonable, roughly 10% above and below the estimated values. Imprecision could potentially be driven by high relatedness among tadpoles in ponds, although removing putative siblings based on KING relatedness (Manichaikul et al., 2010) did not change the infinite upper bounds on jackknife confidence intervals, and removing putative siblings risks erasure of the true evolutionary signal of small populations (Waples & Anderson, 2017). Second, N_e may be better explained by other site characteristics such as age, hydroperiod, catchment/upland area, macroinvertebrates, or a wide variety of other hydrological or biological variables (e.g. Semlitsch, Peterman, Anderson, Drake, & Ousterhout, 2015). Third, California was recently in a multi-year drought from roughly

2011-2017 that likely reduced survival and fecundity for several consecutive breeding seasons (e.g. Fisher et al., 2018). In typical climate conditions, larger ponds should hold water longer and elongate the period before metamorphosis is necessary, leading to fitter metamorphs (Morey, 1998; Morey & Reznick, 2004) and potentially higher effective population sizes. However, if the drought was devastating enough that even large ponds rarely held water for more than the three weeks necessary for spadefoots to reach metamorphosis, all ponds could have experienced a similar bottleneck leading to small effective population sizes regardless of pond area. To narrow down the causes of the observed low effective population sizes, I encourage follow-up sampling to study temporal changes in effective population size in relation to climate and landscape change.

Landscape resistance and connectivity

The highest-ranked resistance surface in ResistanceGA identified relatively flat basins and valleys to have the lowest resistance to gene flow for spadefoots, in areas where suitable habitat (and spadefoots) by and large no longer exist. TreeMix also recovered regular migration events occurring among populations across this low-lying region of the Los Angeles Basin. A study of the historical extent of vernal pools in neighboring San Diego County revealed a once-extensive distribution of vernal pools and vernal pool complexes (Bauder & McMillan, 1998), and the same was likely true for the Los Angeles Basin (Mattoni, Longcore, George, & Rich, 1997) given that it is in the same ecoregion. A Maxent environmental niche model (with no land use variables included) for all southern-clade *S. hammondi* also showed some of the highest habitat suitability for the entire lineage in the Los Angeles Basin, and relatively low suitability in the currently-occupied hills (Neal et al., 2018). Human development has impacted nearly the

entirety of the Los Angeles Basin, eliminating the natural landscape and disrupting genetic connectivity for most native species. Most sampled pools in this study are restricted to the lower-suitability, higher-resistance uplands bordering the Basin, suggesting that most of the existing breeding pools in Orange County may represent refugial populations in marginal, but still intact, habitat spared from habitat destruction by land protections (County of Orange, 1996).

Artificial ponds' contribution to spadefoot conservation

The existing constructed breeding pools at Irvine Mesa (IM) and Shoestring Canyon (SHOE) were stocked with tadpoles from a single cluster of ponds within 1 km of one another (Glenn Lukos Associates, Inc., 2005 and 2006, unpublished reports). This source site shared similar characteristics with the natural ponds sampled in this study in that it was in hilly terrain at the edge of the Los Angeles Basin and isolated from other known breeding ponds by areas of high resistance. The placement of the constructed ponds is in generally low-resistance habitat that has the greatest potential for migration with nearby natural ponds, and in the 10 years since first tadpole introduction (2005-2006) and the current study's sampling (2015-2017), SHOE already showed small amounts of admixture with natural Inland populations. With only natural sites as focal nodes, Circuitscape (Fig. 2.5) reconstructed a low-resistance corridor between natural and these artificial sites, validating the choice of location for the construction of these ponds.

Stocking all of the constructed pools with individuals from only a single population limits their potential for bolstering Orange County spadefoot populations as new nodes for gene flow within the overall metapopulation. Other biologists and I observed hundreds of tadpoles in the artificial ponds during sampling from 2015-2017, but I found low levels of genetic diversity and

small effective population sizes in these ponds, consistent with their initial founding from a single gene pool. When combined and considered as a single population, the group of 9 artificial ponds still had lower genetic diversity than most natural inland ponds (but, notably, higher than Coast ponds). To take full advantage of the existing constructed ponds, or when considering the construction of any new ponds, I recommend stocking with tadpoles from sites throughout Orange County in an effort to recreate a genetically diverse population representative of the original, higher rates of gene flow that presumably existed before major human impacts in the LA Basin. Although I did find pond-level genetic isolation and structure, genetic distances among spadefoots in Orange County are probably low enough that outbreeding depression is probably much less a risk than inbreeding depression given the extremely low N_e estimates. I also found indications of natural admixture of Coast and Inland genotypes at SANCAN, suggesting that some additional admixture might replicate historical levels of movement and genetic mixing. A conservative step forward for the IM/SHOE ponds would be to first introduce tadpoles from more closely related Inland ponds and monitor the populations for outbreeding or other impacts. In addition to manual introduction of spadefoot tadpoles into constructed ponds, I strongly endorse the construction of new ponds to enhance natural metapopulation dynamics by establishing ponds or chains of ponds as stepping stones in low-resistance corridors within the migratory capacity of natural populations.

Figures

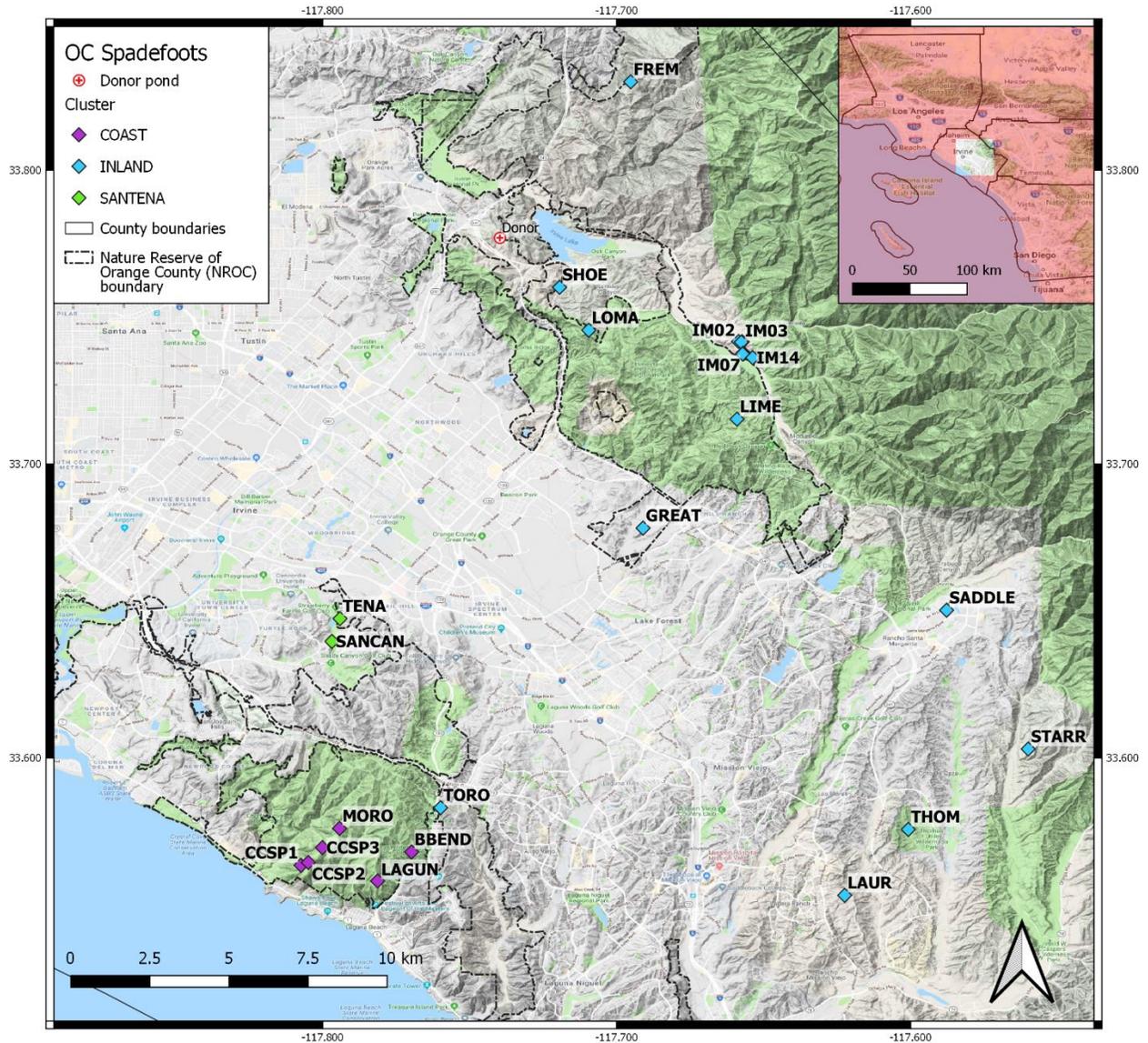


Fig. 2.1 Map of study area with sample sites colored by FastStructure genetic cluster membership for three hierarchically-determined clusters. Dot-dash outlines mark the boundary of the Nature Reserve of Orange County. IM and SHOE ponds are artificial and were seeded with

individuals from donor ponds at the marked site. Inset shows broader context of study area in Southern California. Green-shaded areas are public lands

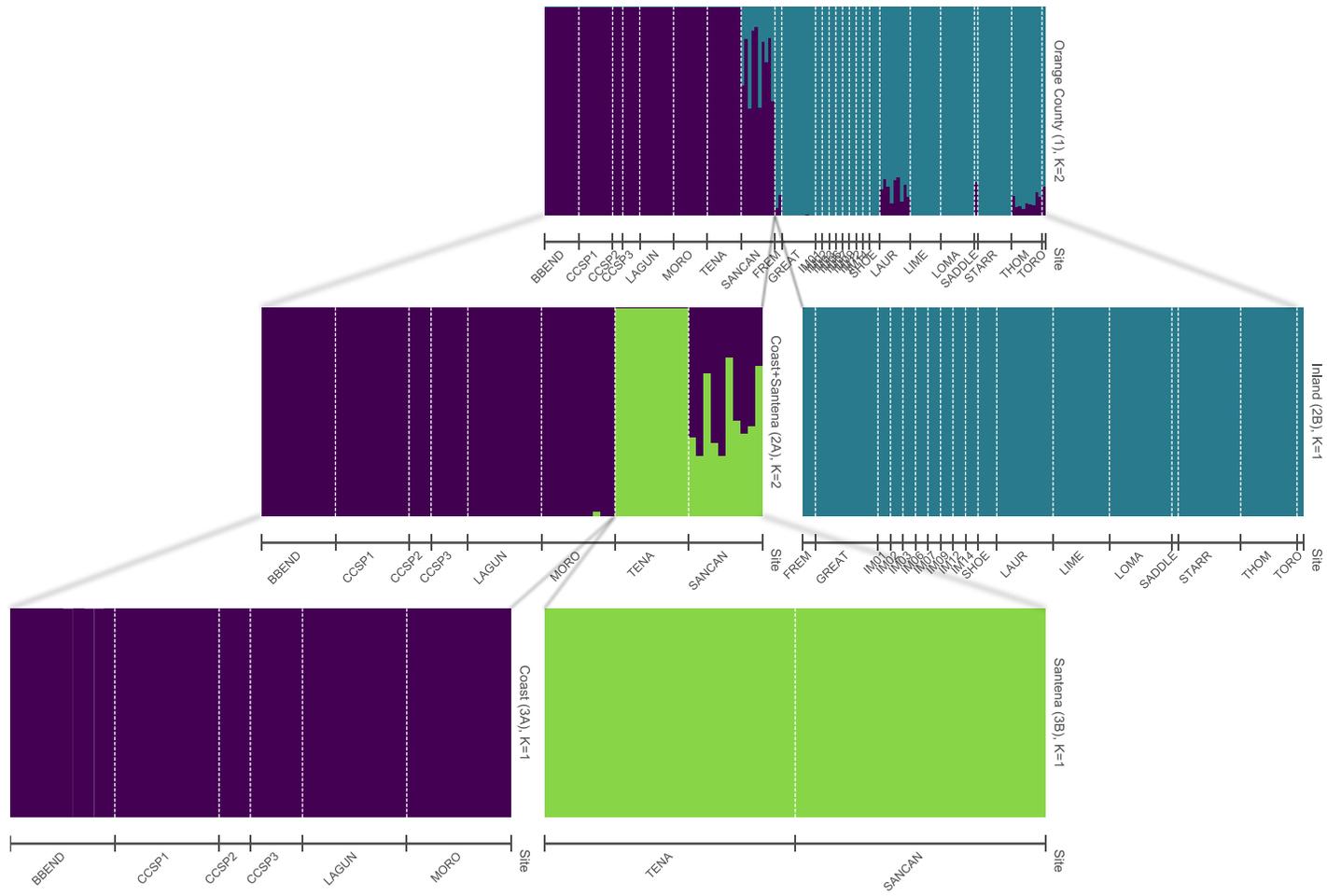


Fig. 2.2 Hierarchical FastStructure barplots of 148 individuals using 6864 unlinked SNPs. Row 1 column 2 is K=2 for 148 Orange County samples. Row 2 column 1 is the Inland cluster (blue) at K=1. Row 2 column 2 is K=2 for Coast (Purple) + Santena (Green) samples. Row 3 shows Coast and Santena at K=1.

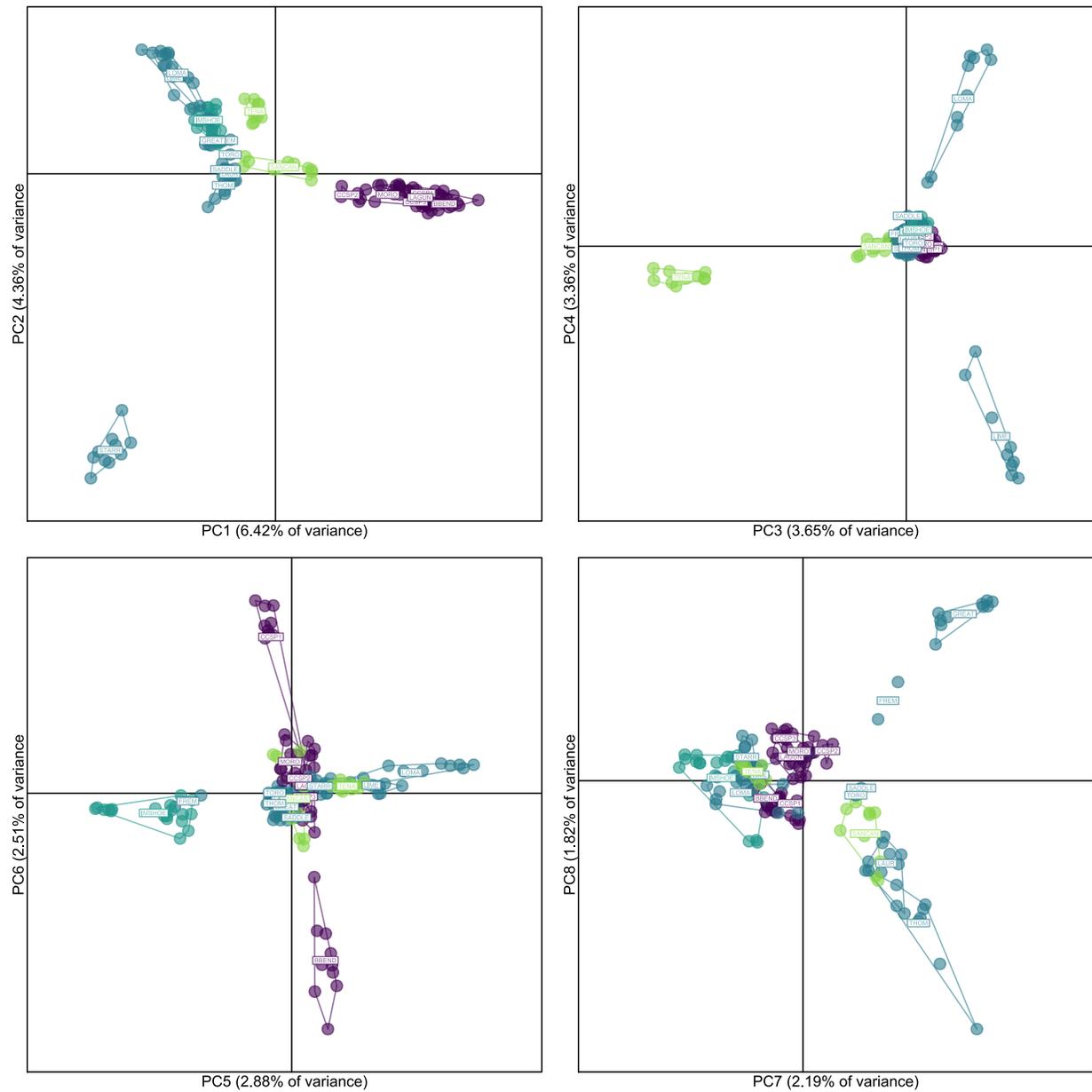


Fig. 2.3 Principle components analysis of 148 Orange County samples with 6864 unlinked SNPs

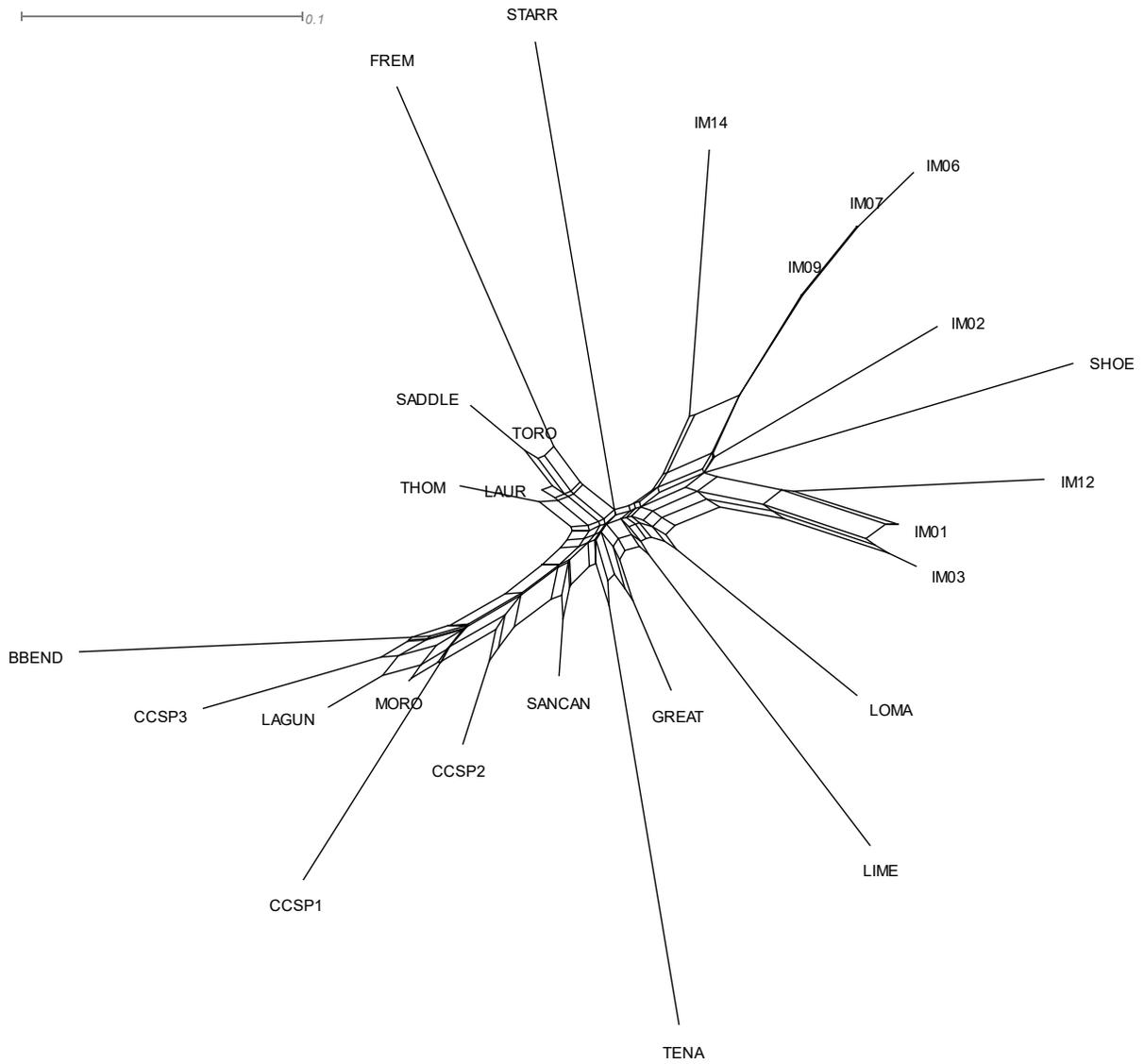
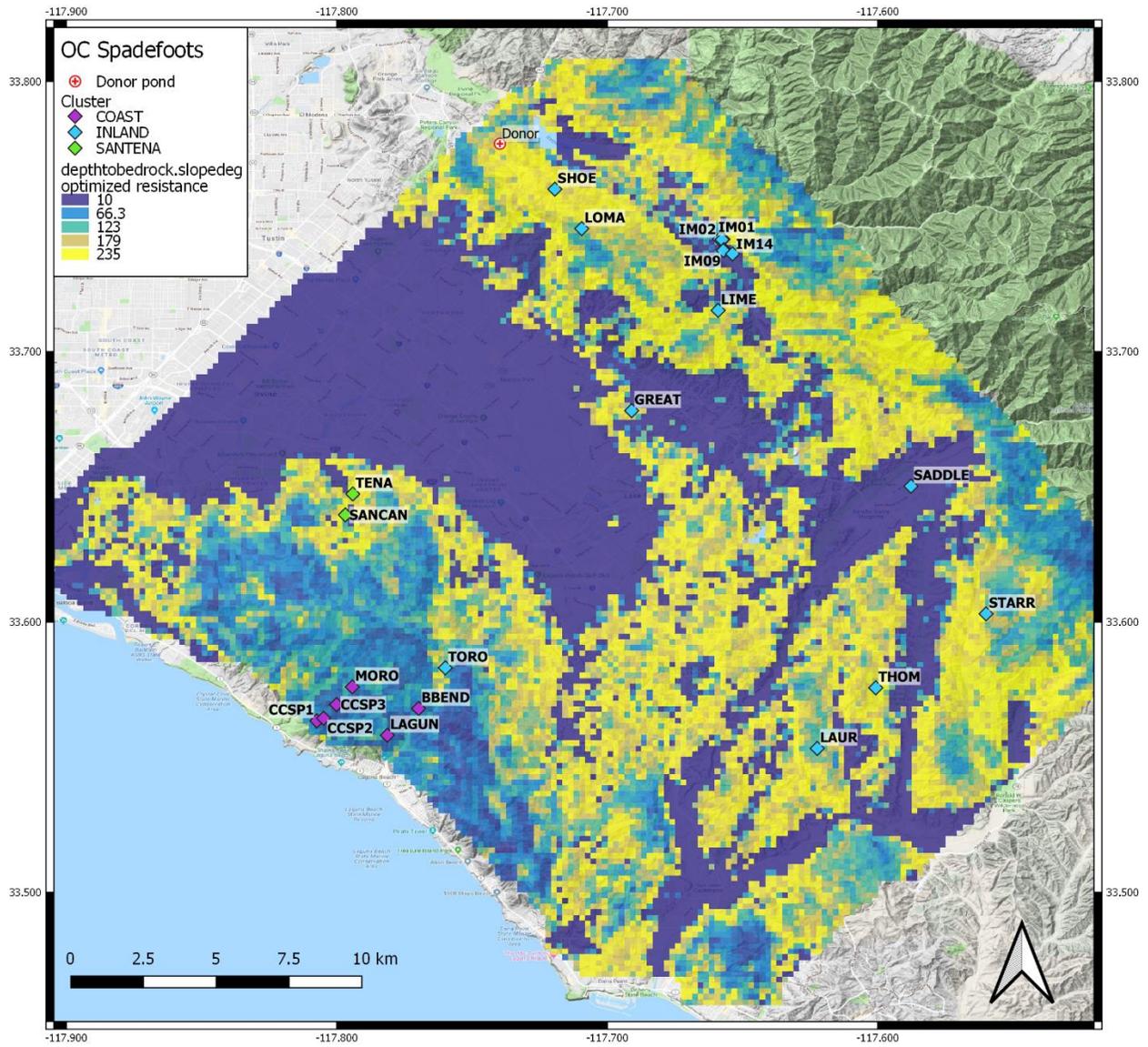


Fig. 2.4 NeighborNet distance network of Weir and Cockerham's pairwise F_{ST} among 26 sites for OC_all_208 at up to 75% missing data per locus



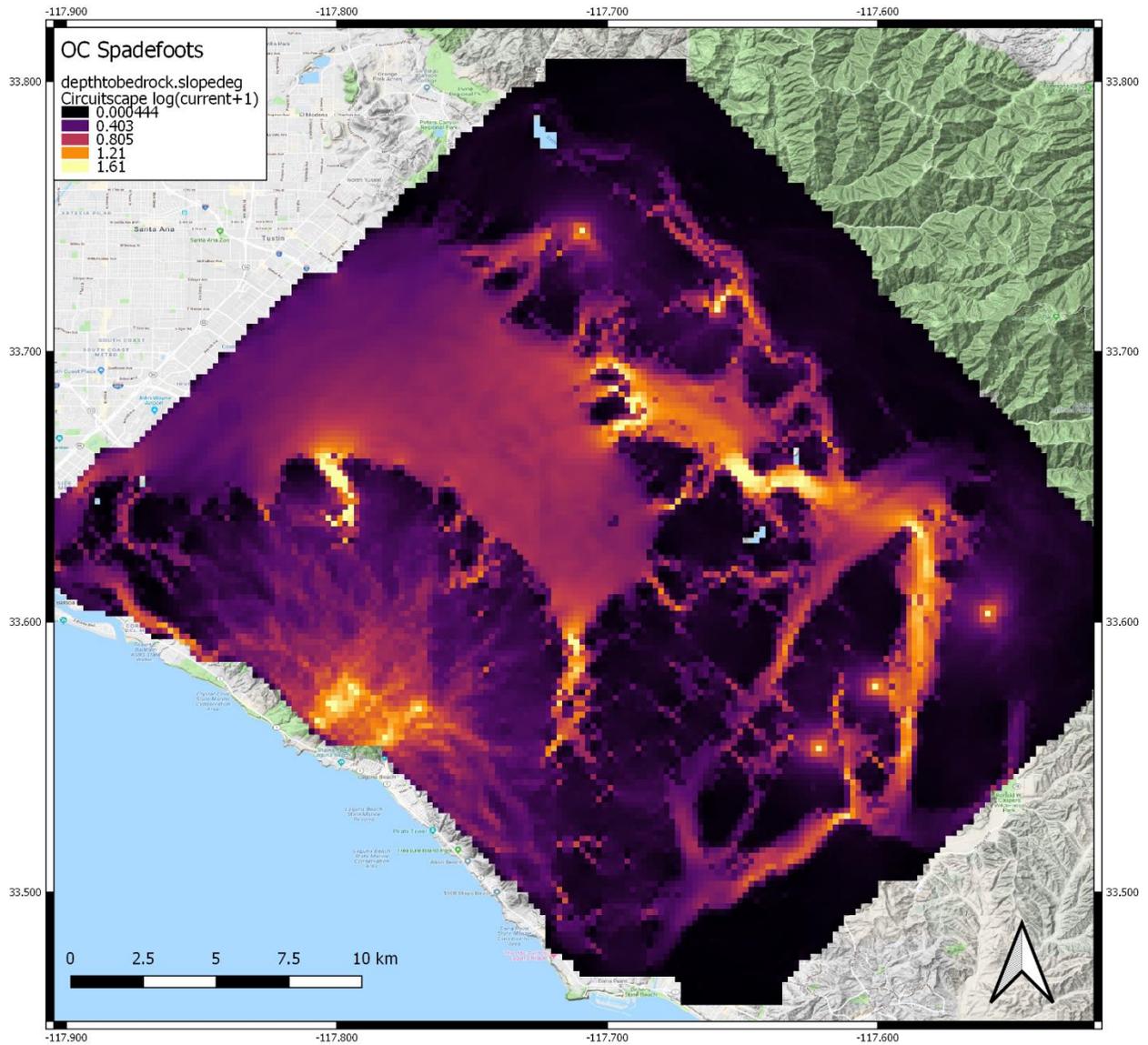


Fig. 2.5 A) ResistanceGA-optimized resistance parameterization of depth to bedrock + slope raster. Blue is lower resistance, green-yellow is higher. B) Circuitscape current run through optimized depth to bedrock + slope raster, using native localities as focal nodes. Highest current is in yellow/orange, lowest current in purple.

Tables

Table 2.1 Genetic diversity values for individual ponds. Starred rows are values calculated on groupings of multiple ponds

Pond	Cluster	n	H _O	H _E	Pi.allsites	F _{IS}	MolCoan	
							N _{eb}	LDN _e
IM01	Inland-artif	10	0.07826	0.07717	3.737E-04	0.0073	2	8.2
IM02	Inland-artif	8	0.07967	0.06579	3.268E-04	-0.01564	3.1	1.4
IM03	Inland-artif	7	0.06927	0.05873	3.213E-04	-0.011	1.7	19.8
IM06	Inland-artif	10	0.07798	0.06383	3.173E-04	-0.02134	2.2	6.1
IM07	Inland-artif	10	0.07617	0.0656	3.397E-04	-0.01256	1.7	2.9
IM09	Inland-artif	10	0.07297	0.07521	3.757E-04	0.02117	1.9	3
IM12	Inland-artif	6	0.06051	0.05421	2.792E-04	0.00152	4.9	3
IM14	Inland-artif	8	0.07426	0.06087	3.233E-04	-0.01898	2.1	8.7
SHOE	Inland-artif	10	0.09114	0.07622	3.532E-04	-0.01889	1.4	NA
*Inland-artif-all	Inland-artif	79	0.07624	0.09238	4.203E-04	0.09531		

*Inland-artif- sub	Inland-artif	19	0.08095	0.08981	4.208E-04	0.0425		
FREM	Inland-natural	2	0.07864	0.0555	3.347E-04	0.00783		
GREAT	Inland-natural	10	0.1098	0.09804	4.685E-04	-0.01399	3.7	2
LAUR	Inland-natural	9	0.11	0.11396	5.298E-04	0.03452	1.4	9.1
LIME	Inland-natural	9	0.09991	0.087	3.805E-04	-0.00832	2.1	11.4
LOMA	Inland-natural	10	0.09861	0.09243	3.972E-04	0.00565	3.5	NA
SADDLE	Inland-natural	1	0.10761	0.0538	4.350E-04			
STARR	Inland-natural	10	0.10897	0.08915	3.828E-04	-0.02632	4.4	2.2
THOM	Inland-natural	9	0.10429	0.10682	5.277E-04	0.02759	1.2	63.8
TORO	Inland-natural	1	0.12134	0.06067	5.150E-04			
*Inland-natural- all	Inland-natural	61	0.09935	0.12301	5.400E-04	0.14663		
BBEND	Coast	10	0.07391	0.06515	2.990E-04	-0.00286	3.2	11
CCSP1	Coast	9	0.08244	0.06979	3.267E-04	-0.01037	2.6	2.2

CCSP2	Coast	3	0.08748	0.06318	3.703E-04	-0.01516		
CCSP3	Coast	5	0.07735	0.05922	3.172E-04	-0.0184	12.2	NA
LAGUN	Coast	10	0.08403	0.0897	4.023E-04	0.03296	3	2.1
MORO	Coast	10	0.08974	0.09118	4.125E-04	0.02417	1.8	2
*Coast-all	Coast	48	0.07796	0.09164	4.060E-04	0.08156		
SANCAN	Santena	10	0.09819	0.10168	4.550E-04	0.02698	4.2	2.6
TENA	Santena	10	0.07938	0.06709	3.087E-04	-0.01351	9	NA
*Santena-all	Santena	20	0.08382	0.09476	4.360E-04	0.05048		
*Natural-all		129	0.08869	0.11671	5.093E-04	0.22937		
*All ponds		208	0.08523	0.11616	5.015E-04	0.28096		

Supplemental Figures and Tables

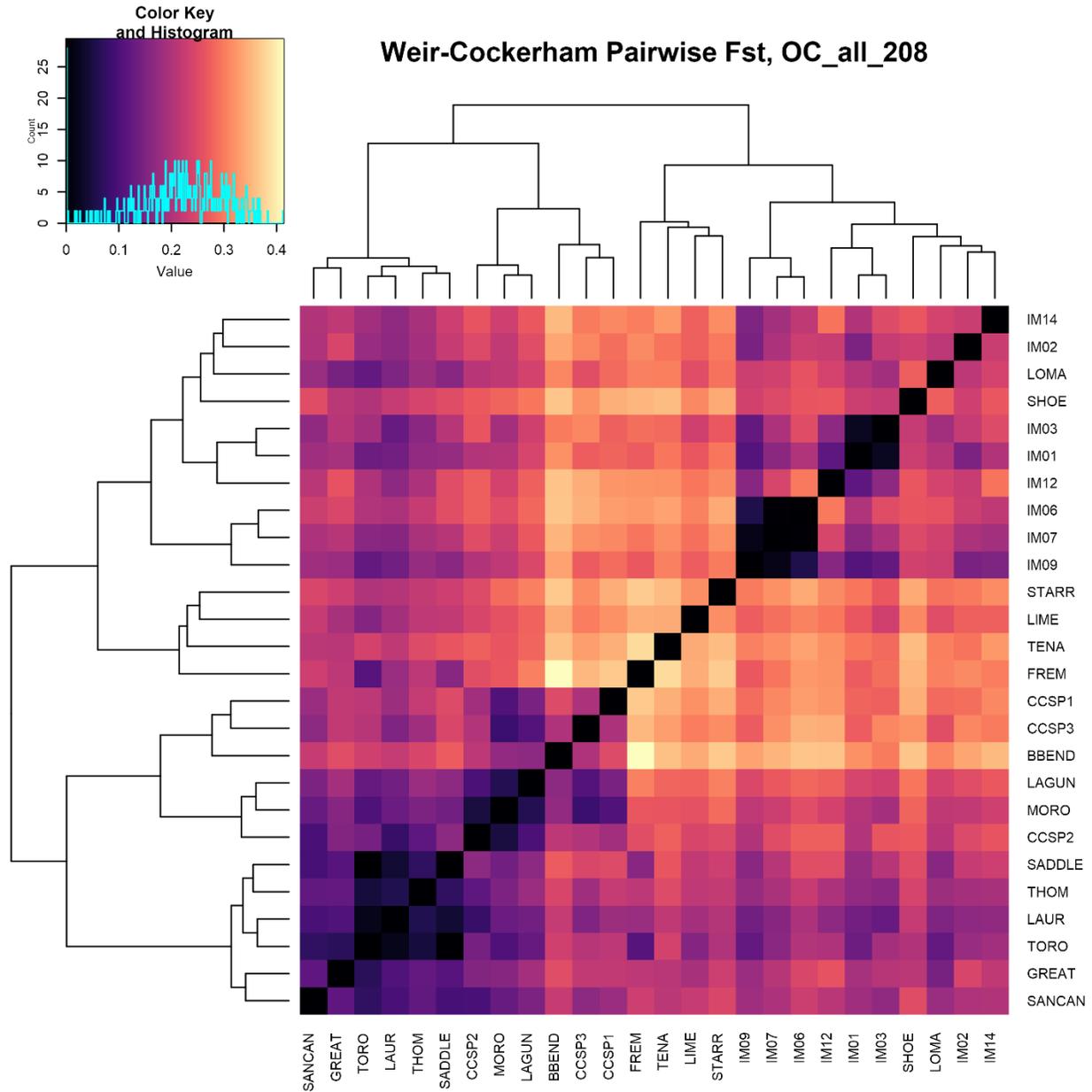


Fig. 2.S1 Heatmap and histogram of pairwise Weir-Cockerham F_{ST} between sites using the OC_all_208 dataset, with darker colors representing lower F_{ST} values and lighter being higher F_{ST} . Populations clustered using "ward.D2" method in `cluster::hclust()` in R. Heatmap visualized using `gplots::heatmap.2()`. Populations with fewer than two samples are excluded

Fig. 2.S2 Spearman's Rho (lower triangle) and Kendall's tau (upper triangle) correlations between measures of genetic diversity, effective population size, and pond area and perimeter. All values displayed are statistically significant ($p < 0.05$, two-tailed test)

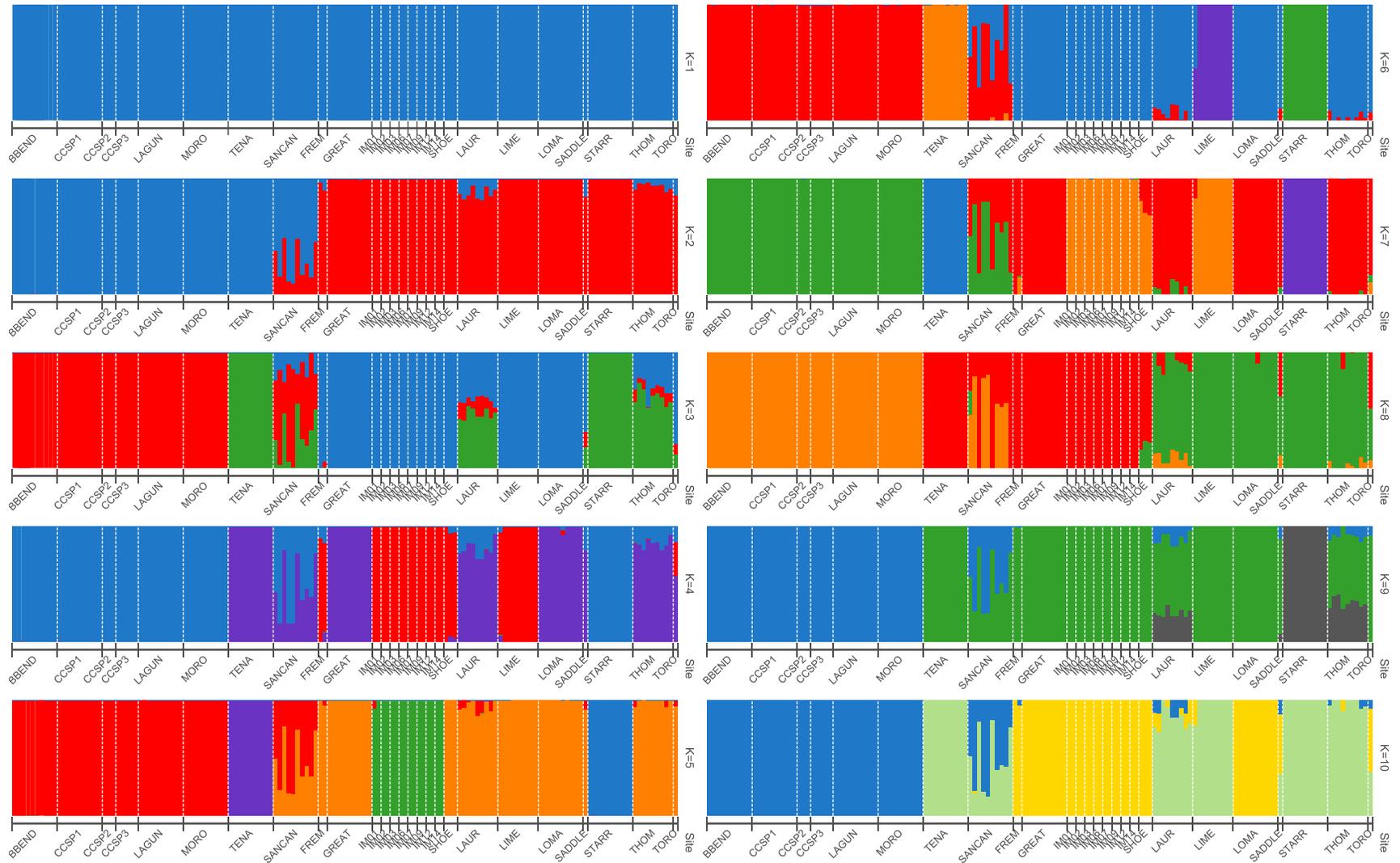


Fig. 2.S3 Ancestry barplots for OCReduced148 run in FastStructure with ten cross-validations for K=1-10. K=1-5 from upper left to lower left; K=6-10 from upper right to lower right. Note at higher K values, K clusters typically do not appear in the barplots

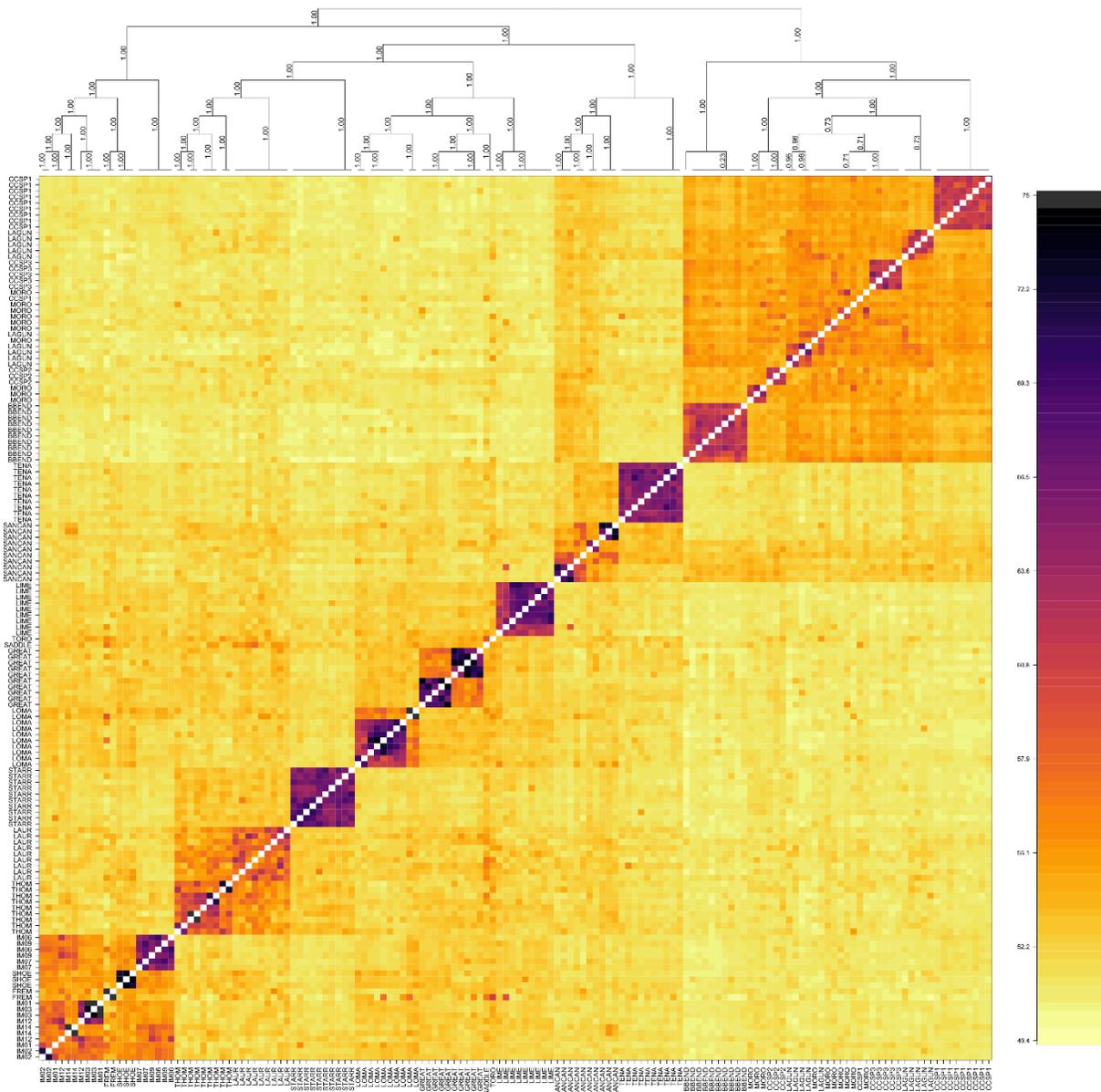


Fig. 2.S4 fineRADstructure coancestry matrix of 148 spadefoots, from Orange County, California. Lighter/yellower colors are lower coancestry; darker/more purple colors are higher coancestry. Nodes in population tree are labeled with posterior population assignment probabilities

0.01

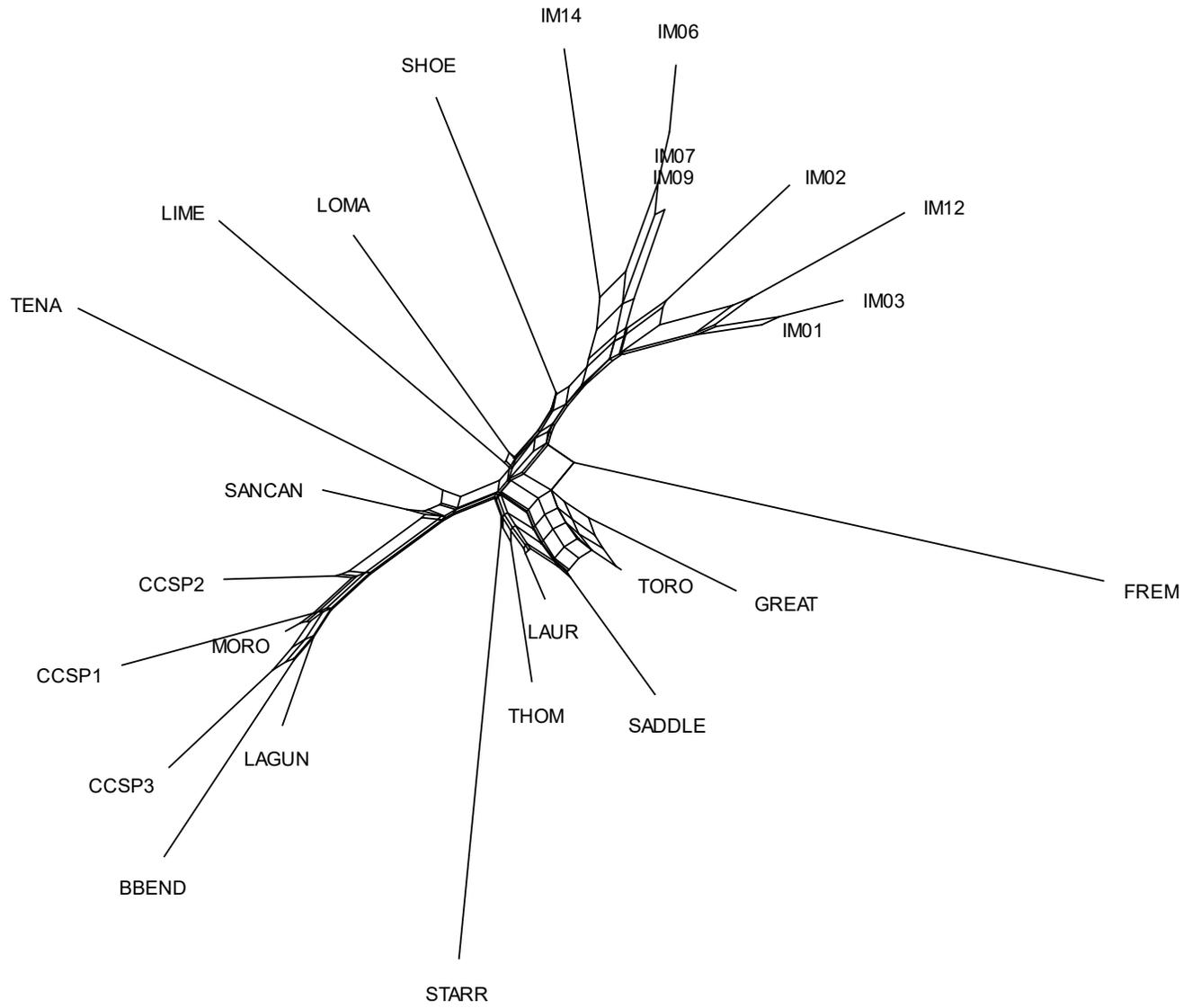


Fig. 2.S5 NeighborNet network of Jost's D for OC_all_208 at up to 75% missing data per locus

0.01

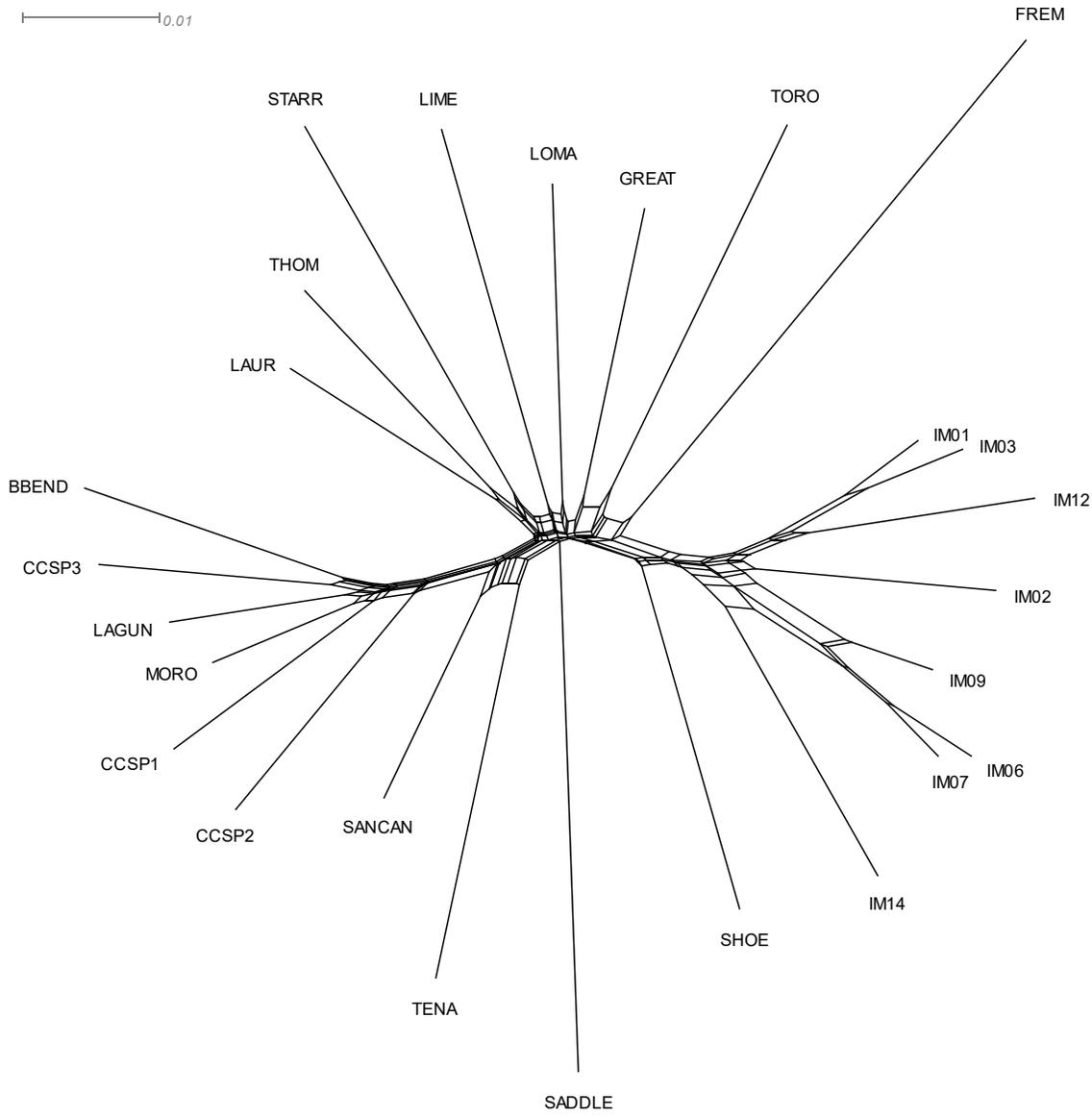


Fig. 2.S6 NeighborNet network of Nei's D_a for OC_all_208 at up to 75% missing data per locus

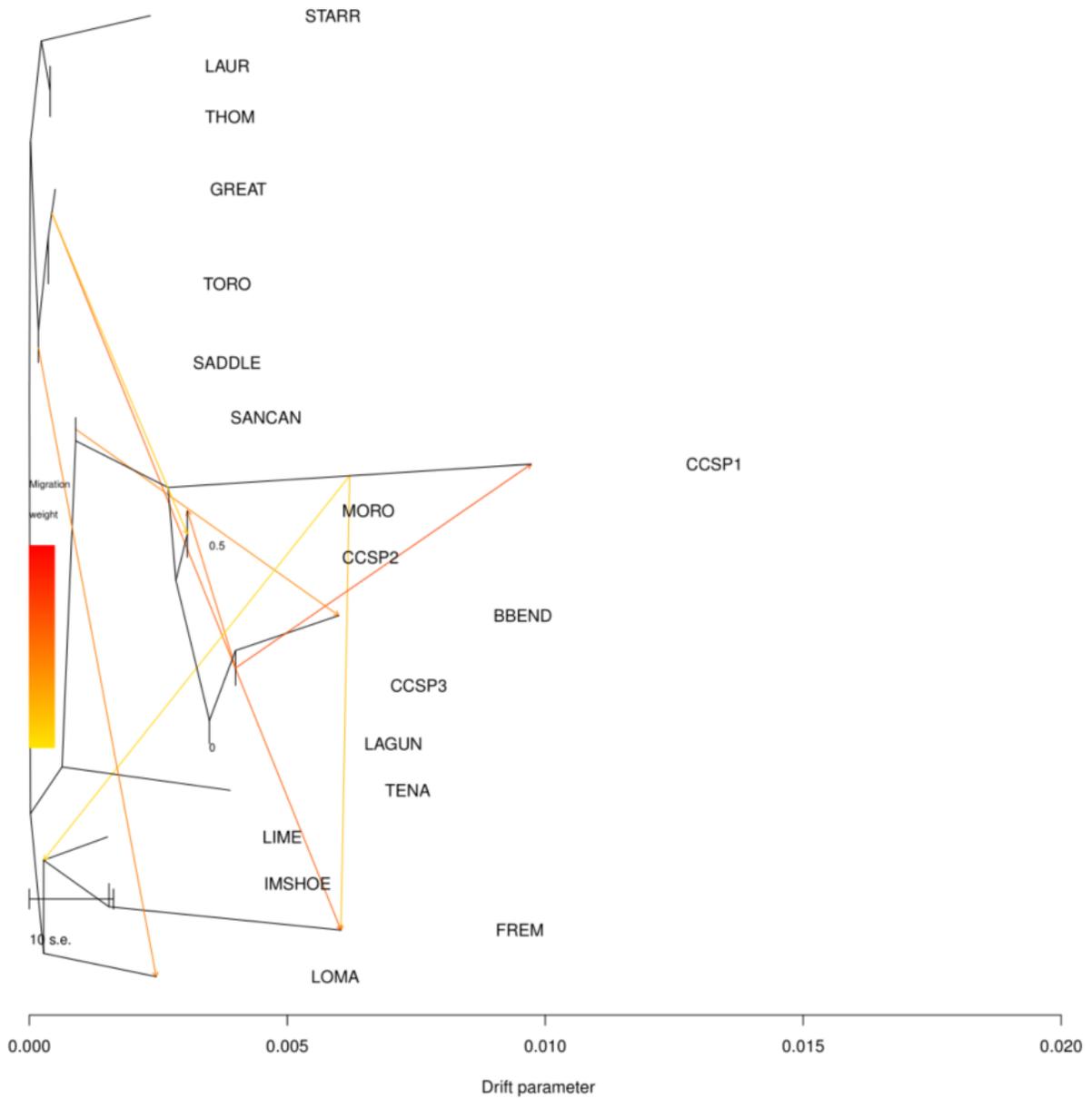


Fig. 2.S7 TreeMix tree of Orange County populations with migration parameter set to 8. The 8 edges were CCSP3 to CCSP1; CCSP3 to MORO; GREAT to FREM; SANCAN to BBEND; SADDLE to LOMA; CCSP1 to FREM; GREAT to the ancestor of MORO and CCSP2; and CCSP1 to the ancestor of LIME, IMSHOE, and FREM.

Table 2.S1 ResistanceGA resistance surface optimization results using Nei's D_a as response variable

Surface	resolution (°)	LL	k	AIC	AICc	R2m	R2c
depthtobedrock.slopedeg	0.0020833	320.049		7-632.098	-603.698	0.769	0.9713
depthtobedrock	0.0020833	318.580		4-629.160	-624.160	0.7613	0.9704
landcover	0.0020833	317.793		15-627.586	-765.586	0.742	10.8948
landcover.slopedeg	0.0020833	318.023		18-628.046	-714.046	0.7079	0.9007
bio12	0.00833	310.587		4-613.174	-608.174	0.6672	0.9555
bio19	0.00833	316.506		4-625.013	-620.013	0.6479	0.9415
maxentalOC	0.0020833	312.694		4-617.388	-612.388	0.6363	0.8773
TAXOUSDA	0.0020833	311.464		8-614.928	-570.928	0.6260	0.8683
terrainruggedness	0.00833	316.270		4-624.540	-619.540	0.6255	0.9734
bio13	0.00833	308.200		4-608.400	-603.400	0.6254	0.9418
claycontent	0.0020833	313.317		4-618.634	-613.634	0.6247	0.9207
bio18	0.00833	328.616		4-649.233	-644.233	0.6212	0.9878
slopedeg	0.0020833	310.751		4-613.502	-608.502	0.6183	0.9188
bio03	0.00833	320.002		4-632.004	-627.004	0.6117	0.9948
depthtobedrock.landcover	0.0020833	321.610		18-635.220	-721.220	0.5824	0.9333
impervious	0.0020833	310.120		4-612.240	-607.240	0.5791	0.8760

bulkdensity	0.0020833	308.627	4	-609.254	-604.254	0.5635	0.8585
bio17	0.00833	312.392	6	-616.784	-598.784	0.5620	0.9295
Distance	0.0020833	308.622	2	-609.245	-612.045	0.5615	0.8485
growingDegDays5	0.00833	302.778	4	-597.556	-592.556	0.5469	0.8259
growingDegDays0	0.00833	302.780	4	-597.560	-592.560	0.5460	0.8262
bpsprehumanvegmodel	0.0020833	313.466	13	-618.932	-964.932	0.5395	0.8842
bio09	0.00833	303.410	4	-598.819	-593.819	0.5369	0.8283
canopy	0.00833	302.744	4	-597.488	-592.488	0.5350	0.8494
bio10	0.00833	302.799	4	-597.598	-592.598	0.5306	0.8386
siltcontent	0.0020833	312.594	4	-617.188	-612.188	0.5253	0.9692
depthtobedrock.landcover.slopedeg	0.0020833	320.692	21	-633.384	-702.051	0.5228	0.9538
minTempWarmest	0.00833	302.142	4	-596.283	-591.283	0.5180	0.8387
bio15	0.00833	324.359	4	-640.719	-635.719	0.5177	0.9355
PETWettestQuarter	0.00833	302.129	4	-596.258	-591.258	0.5166	0.8363
PETColdestQuarter	0.00833	302.232	4	-596.465	-591.465	0.5163	0.8362
embergerQ	0.00833	302.250	4	-596.500	-591.500	0.5117	0.8409
bio05	0.00833	301.953	4	-595.906	-590.906	0.4952	0.8491
Distance	0.00833	301.652	2	-595.304	-598.104	0.4935	0.8440
bio16	0.00833	301.656	4	-595.312	-590.312	0.4931	0.8441
bio14	0.00833	301.613	4	-595.227	-590.227	0.4913	0.8290
monthCountByTemp10	0.00833	302.545	5	-597.091	-586.519	0.4899	0.8499
maxTempColdest	0.00833	303.242	4	-598.484	-593.484	0.4736	0.8494

climaticMoistureIndex	0.00833	303.719	4	-599.438	-594.438	0.4718	0.8577
aridityIndexThornthwaite	0.00833	306.502	4	-605.003	-600.003	0.4674	0.8898
bio08	0.00833	303.039	4	-598.078	-593.078	0.4562	0.8645
bio01	0.00833	304.894	4	-601.788	-596.788	0.4539	0.8821
bio07	0.00833	305.142	4	-602.284	-597.284	0.4522	0.9303
bio02	0.00833	305.178	4	-602.356	-597.356	0.4512	0.9298
PETseasonality	0.00833	304.397	4	-600.794	-595.794	0.4502	0.8780
thermicityIndex	0.00833	303.702	4	-599.404	-594.404	0.4499	0.8638
bio11	0.00833	303.372	4	-598.743	-593.743	0.4498	0.8699
PETDriestQuarter	0.00833	304.474	4	-600.948	-595.948	0.4481	0.8831
topoWet	0.00833	303.976	4	-599.952	-594.952	0.4472	0.8655
PETWarmestQuarter	0.00833	304.396	4	-600.792	-595.792	0.4465	0.8854
elevation	0.0020833	310.355	4	-612.710	-607.710	0.4459	0.8993
bio06	0.00833	303.546	4	-599.092	-594.092	0.4440	0.8781
continentality	0.00833	303.989	4	-599.977	-594.977	0.4438	0.8750
bio04	0.00833	305.236	4	-602.472	-597.472	0.4374	0.9095
annualPET	0.00833	304.995	4	-601.989	-596.989	0.4323	0.9098
Null	0.00833	259.377	1	-512.755	-516.391	0.0000	0.3478
Null	0.0020833	259.377	1	-512.755	-516.391	0.0000	0.3478

Appendix 2.1

Data filtering/parameters for datasets used in Chapters 2

I first trimmed the internal adapters from the demultiplexed sequences using cutadapt (Martin, 2011). As an additional filtering step, I used fastp (Chen, Zhou, Chen, & Gu, 2018) to filter reads below 50% complexity (percentage of bases that are different from the next base in a sequence) and with a mean quality score below Q30. To capture the highest-quality portions at the start of each read and because of differences of read lengths among SR100 and SR150/PE150 sequencing runs, I trimmed every read to 60 basepairs and removed reads less than 30 basepairs in length. I then used ipyrad 0.7.29 (Eaton, 2015/2019; Eaton, 2014) to assemble the RAD loci. I ran ipyrad using a reference draft genome that I sequenced in-house from a *S. hammondii* metamorph from Santa Barbara County, California; this individual is a part of the northern clade of *S. hammondii*, the sister clade of the southern *S. hammondii* in the current study (Neal, Johnson, & Shaffer, 2018). The genome was sequenced by 10X Genomics and assembled using Supernova (10X Genomics), and analyzed using BUSCO for quality (Simão et al., 2015) (assembly results in Appendix 2-1).

From the ipyrad output, I used vcftools 1.1.5 (Danecek et al., 2011) to produce several data subsets for Orange County spadefoots and clusters identified therein (see clustering results) to maximize the size of data matrices when analyzing the different subsets separately. For each dataset, I removed singletons as recommended to filter noise from the dataset (Linck & Battey, 2019); allowed only biallelic SNPs; and retained only the first SNP in each RAD locus to reduce effects of physical linkage among sites. I tested missing data thresholds from 20% to 75% but found no major differences in results, so I settled on allowing up to 75% missing data to

maximize the size of the data matrices for pairwise and intrapopulation statistics. Parameters used in ipyrad and vcftools 1.1.5 to produce each final dataset can be found in the supplemental information.

I ran the ipyrad pipeline on a dataset of 586 samples that includes members of all species of the genus *Spea* (*S. bombifrons*, *S. hammondii*, *S. intermontana*, and *S. multiplicata*), using a clustering threshold of 0.85, minimum depth of 6, and including loci missing from up to 90% of individuals. Additional parameters were a mindepth of 6; phred Q score offset of 33; max N's in consensus of 5; max heterozygotes in consensus of 8; max indels of 1; and max heterozygous sites per locus of 25%. After assessing individual missingness statistics from the multispecies ipyrad run, I removed individuals with >50% missing data unless they were one of two or fewer samples at a site. I branched the ipyrad pipeline at step 7 to include 267 individuals from across the range of *S. hammondii* (South), changing only the parameter to include loci missing from 90% to 75% of individuals. Using the vcf file output from ipyrad, I used vcftools 1.1.5 (Danecek et al., 2011) to subsample for only Orange County samples, plus additional data filtering. The Orange County subsample included 208 samples from 26 sites. I retained only the first SNP in each RAD locus to reduce effects of physical linkage among sites (thin=200); filtered out singletons (mac=2) as recommended to filter noise from the dataset (Linck & Battey, 2019); and allowed only biallelic SNPs (max-alleles=2). I produced datasets with up to 75% missing data to maximize the size of the data matrix for use with pairwise and intrapopulation statistics. I also produced datasets with up to 20% and 50% missing data and report those results when they differ substantially from the 75% dataset.

Following preliminary analyses on the full Orange County dataset of 208 individuals (OC_all_208), I used vcftools to produce several subsets of the 208 individuals. Because uneven

sampling can lead to inaccurate inference of genetic structure, and because the individuals in the artificial ponds in this study were derived from a single locality from 2005-2006, I produced a subset in which I retained only two or three individuals from each artificial pond, retaining those with the most complete genetic data based on percent missingness from vcftools. By subsampling individuals with the least amount of missing data from each artificial pond and recognizing that the resulting individuals were all derived from a single natural pond, my goal was to achieve better sampling parity with the ~10-20 individuals sampled from each native ponds. This reduced the total number of individuals from artificial ponds from 79 to 19, achieving much closer sampling parity with the native ponds, and the total number of individuals to 148 (OC_reduced_148). For explicitly spatial analyses where the goal was to explore natural connectivity, I also produced a subset of 129 individuals where I removed all individuals sampled from artificial ponds (OC_native_129).

To generate the alleles.loci file for fineRADstructure, I created an additional ipyrad branch using the same 148 individuals with 75% missingness and converted the file to fineRADstructure format using fineRADstructure-tools (Ortiz, 2017/2019), with the final input file containing 3099 RAD loci.

BUSCO parameters for S. hammondii draft genome assembly

Sample information:

Catalog number: HBS 132740

Species: *Spea hammondii*

Locality name: "Dominion Road"

County: Santa Barbara

State: California

Country: USA

Latitude: 34.842069

Longitude: -120.315529

Collection date: 11 September 2017

Remarks: "caught live 11Sept2017, maintained alive in captivity until sacrificed on 29Sept2017 for 10X genome tissue (liver and muscle)"

BUSCO input:

- 59.97 x = RAW COV = raw coverage; ideal ~56
- 36.03 x = EFFECTIVE COV = effective read coverage; ideal ~42 for nominal 56x
- 72.93 % = READ TWO Q30 = fraction of Q30 bases in read 2; ideal 75-85
- 1.46 Gb = EST GENOME SIZE = estimated genome size
- 21.55 % = REPETITIVE FRAC = estimated repetitive fraction
- 7.08 Kb = MOLECULE LEN = weighted mean molecule size; ideal 50-100
- 364.00 b = HETDIST = mean distance between heterozygous SNPs
- 21.77 K = LONG SCAFFOLDS = number of scaffolds \geq 10 kb
- 6.54 Kb = EDGE N50 = N50 edge size
- 29.66 Kb = CONTIG N50 = N50 contig size
- 2.05 Kb = PHASEBLOCK N50 = N50 phase block size
- 38.72 Kb = SCAFFOLD N50 = N50 scaffold size
- 24.38 % = MISSING 10KB = % of base assembly missing from scaffolds \geq 10 kb
- 662.81 Mb = ASSEMBLY SIZE 10K = assembly size (only scaffolds \geq 10 kb)

- 976.38 Mb = ASSEMBLY SIZE 1K = assembly size (only scaffolds >= 1kb)

BUSCO results:

INFO C:65.2%[S:61.6%,D:3.6%],F:21.9%,M:12.9%,n:3950

INFO 2576 Complete BUSCOs (C)

INFO 2435 Complete and single-copy BUSCOs (S)

INFO 141 Complete and duplicated BUSCOs (D)

INFO 866 Fragmented BUSCOs (F)

INFO 508 Missing BUSCOs (M)

INFO 3950 Total BUSCO groups searched

References

- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. doi: 10.1038/nrg2844
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. doi: 10.1038/nrg.2015.28
- Bauder, E. T., & McMillan, S. (1998). Current distribution and historical extent of vernal pools in southern California and northern Baja California, Mexico. *Ecology, Conservation and Management of Vernal Pool Ecosystems—Proceedings from a 1996 Conference*. Presented at the California Native Plant Society, Sacramento, CA.
- Baumberger, K. (2013). *Uncovering a fossorial species: Home range and habitat preference of the western spadefoot, Spea hammondii (Anura: Pelobatidae), in Orange County protected areas* (M.S., California State University, Fullerton). Retrieved from <https://search.proquest.com/docview/1503752887/abstract/FC4BEFD0FAF246EEPQ/1>
- Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *BioRxiv*, 205799. doi: 10.1101/205799
- Beebe, T. J. C., & Griffiths, R. A. (2005). The amphibian decline crisis: A watershed for conservation biology? *Biological Conservation*, *125*(3), 271–285. doi: 10.1016/j.biocon.2005.04.009
- Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., ... Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, *25*(13), 2967–2977. doi: 10.1111/mec.13647
- Blaustein, A. R., Wake, D. B., & Sousa, W. P. (1994). Amphibian Declines: Judging Stability, Persistence, and Susceptibility of Populations to Local and Global Extinctions. *Conservation Biology*, *8*(1), 60–71. doi: 10.1046/j.1523-1739.1994.08010060.x
- Bryant, D., & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, *21*(2), 255–265. doi: 10.1093/molbev/msh018
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354

- Chaudhuri, P., & Marron, J. S. (1999). SiZer for Exploration of Structures in Curves. *Journal of the American Statistical Association*, *94*(447), 807–823. doi: 10.1080/01621459.1999.10474186
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. doi: 10.1093/bioinformatics/bty560
- Corey, S. J., & Waite, T. A. (2008). Phylogenetic autocorrelation of extinction threat in globally imperilled amphibians. *Diversity and Distributions*, *14*(4), 614–629. doi: 10.1111/j.1472-4642.2007.00448.x
- County of Orange (1996). *Natural Community Conservation Plan and Habitat Conservation Plan*. Retrieved from <https://www.wildlife.ca.gov/Conservation/Planning/NCCP/Plans/Orange-Coastal>
- Cushman, S. A. (2006). Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation*, *128*(2), 231–240. doi: 10.1016/j.biocon.2005.09.031
- D’Amen, M., & Bombi, P. (2009). Global warming and biodiversity: Evidence of climate-linked amphibian declines in Italy. *Biological Conservation*, *142*(12), 3060–3067. doi: 10.1016/j.biocon.2009.08.004
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330
- Davidson, C., Shaffer, H. B., & Jennings, M. R. (2002). Spatial tests of the pesticide drift, habitat destruction, UV-B and climate-change hypotheses for California amphibian declines. *Conservation Biology*, *16*(6), 1588–1601. doi: 10.1046/j.1523-1739.2002.01030.x
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214. doi: 10.1111/1755-0998.12157
- Eaton, D. (2019). *Interactive assembly and analysis of RAD-seq data sets: dereneaton/ipyrad* [Jupyter Notebook]. Retrieved from <https://github.com/dereneaton/ipyrad> (Original work published 2015)
- Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. doi: 10.1093/bioinformatics/btu121
- Ellegren, H. (2013). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0169534713002310>

- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Fisher, R. N., Brehme, C. S., Hathaway, S. A., Hovey, T. E., Warburton, M. L., & Stokes, D. C. (2018). Longevity and population age structure of the arroyo southwestern toad (*Anaxyrus californicus*) with drought implications. *Ecology and Evolution*, *8*(12), 6124–6132. doi: 10.1002/ece3.4158
- Fitak, R. (2019). OptM: Estimating the Optimal Number of Migration Edges from “Treemix” (Version 0.1.1). Retrieved from <https://CRAN.R-project.org/package=OptM>
- Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., ... Luikart, G. (2016). Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. *Trends in Ecology & Evolution*, *31*(2), 81–83. doi: 10.1016/j.tree.2015.10.009
- Goldberg, C. S., & Waits, L. P. (2010). Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology*, *19*(17), 3650–3663. doi: 10.1111/j.1365-294X.2010.04673.x
- Gosselin, T. (2019). *RADseq Data Exploration, Manipulation and Visualization using R: thierrygosselin/radiator* [R]. Retrieved from <https://github.com/thierrygosselin/radiator> (Original work published 2017)
- Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics (Version 0.04-22). Retrieved from <https://CRAN.R-project.org/package=hierfstat>
- Grant, E. H. C., Miller, D. A. W., Schmidt, B. R., Adams, M. J., Amburgey, S. M., Chambert, T., ... Muths, E. (2016). Quantitative evidence for the effects of multiple drivers on continental-scale amphibian declines. *Scientific Reports*, *6*, 25625. doi: 10.1038/srep25625
- Hamer, A. J., & McDonnell, M. J. (2008). Amphibian ecology and conservation in the urbanising world: A review. *Biological Conservation*, *141*(10), 2432–2449. doi: 10.1016/j.biocon.2008.07.020
- Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., ... Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197–1211. doi: 10.1111/eva.12659
- Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. doi: 10.1371/journal.pone.0169748

- Homer, C. H., Fry, J. A., & Barnes, C. A. (2012). USGS Fact Sheet 2012–3020: The National Land Cover Database. Retrieved March 6, 2019, from <https://pubs.usgs.gov/fs/2012/3020/>
- Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254–267. doi: 10.1093/molbev/msj030
- Jennings, M. R., & Hayes, M. P. (1994). *Amphibian and Reptile Species of Special Concern in California*. USA: California Department of Fish and Game.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. doi: 10.1093/bioinformatics/btr521
- Jombart, T., Kamvar, Z. N., Collins, C., Lustrik, R., Beugin, M.-P., Knaus, B. J., ... Ewing, R. J. (2018). adegenet: Exploratory Analysis of Genetic and Genomic Data (Version 2.1.1). Retrieved from <https://cran.r-project.org/web/packages/adegenet/index.html>
- Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, 17(18), 4015–4026. doi: 10.1111/j.1365-294X.2008.03887.x
- Jost, L., Archer, F., Flanagan, S., Gaggiotti, O., Hoban, S., & Latch, E. (2018). Differentiation measures for conservation genetics. *Evolutionary Applications*, 11(7), 1139–1148. doi: 10.1111/eva.12590
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122. doi: 10.1038/sdata.2017.122
- Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLoS Genetics*, 8(1), e1002453. doi: 10.1371/journal.pgen.1002453
- Li, Y.-L., & Liu, J.-X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18(1), 176–177. doi: 10.1111/1755-0998.12719
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12995
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. doi: 10.1093/bioinformatics/btr642

- Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, 35(5), 1284–1290. doi: 10.1093/molbev/msy023
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22), 2867–2873. doi: 10.1093/bioinformatics/btq559
- Marsh, D. M., & Trenham, P. C. (2001). Metapopulation Dynamics and Amphibian Conservation. *Conservation Biology*, 15(1), 40–49. doi: 10.1111/j.1523-1739.2001.00129.x
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12. doi: 10.14806/ej.17.1.200
- Mattoni, R., Longcore, T., George, J., & Rich, C. (1997). Down memory lane: the Los Angeles coastal prairie and its vernal pools. *Poster Presentation at 2nd Interface Between Ecology and Land Development in California*. Presented at the Occidental College. Occidental College.
- McCartney-Melstad, E., Gidiş, M., & Shaffer, H. B. (2018). Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity*, 121(2), 112–125. doi: 10.1038/s41437-018-0097-7
- McCartney-Melstad, E., Vu, J. K., & Shaffer, H. B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Molecular Ecology*, 27(22), 4430–4443. doi: 10.1111/mec.14892
- McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66(2), 526–538. doi: 10.1016/j.ympev.2011.12.007
- McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9), 999–1007. doi: 10.1111/eva.12193
- Mims, M. C., Phillipsen, I. C., Lytle, D. A., Kirk, E. E. H., & Olden, J. D. (2014). Ecological strategies predict associations between aquatic and genetic connectivity for dryland amphibians. *Ecology*, 96(5), 1371–1382. doi: 10.1890/14-0490.1
- Morey, S. R. (1998). Pool duration influences age and body mass at metamorphosis in the western spadefoot toad: implications for vernal pool conservation. *Ecology, Conservation, and Management of Vernal Pool Ecosystems-Proceedings from a 1996 Conference*. California Native Plant Society, Sacramento, CA, 86–91.

- Morey, S. R., & Reznick, D. N. (2004). The relationship between habitat permanence and larval development in California spadefoot toads: Field and laboratory comparisons of developmental plasticity. *Oikos*, *104*(1), 172–190. doi: 10.1111/j.0030-1299.2004.12623.x
- Murphy, M. A., Dezzani, R., Pilliod, D. S., & Storfer, A. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, *19*(17), 3634–3649. doi: 10.1111/j.1365-294X.2010.04723.x
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x
- NCCP Plan Summary – County of Orange (Central/Coastal) NCCP/HCP. (n.d.). Retrieved February 25, 2019, from <https://www.wildlife.ca.gov/Conservation/Planning/NCCP/Plans/Orange-Coastal>
- Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*). *Conservation Genetics*, *19*(4), 937–946. doi: 10.1007/s10592-018-1066-7
- Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, *19*(2), 153–170.
- Nomura, T. (2008). Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications*, *1*(3), 462–474. doi: 10.1111/j.1752-4571.2008.00015.x
- Olson, D. H., Aanensen, D. M., Ronnenberg, K. L., Powell, C. I., Walker, S. F., Bielby, J., ... The Bd Mapping Group. (2013). Mapping the Global Emergence of *Batrachochytrium dendrobatidis*, the Amphibian Chytrid Fungus. *PLoS ONE*, *8*(2), e56802. doi: 10.1371/journal.pone.0056802
- Ortiz, E. M. (2019). *Tools for data conversion and results visualization for fineRADstructure* (<http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html>): *edgardomortiz/fineRADstructure-tools* [Python]. Retrieved from <https://github.com/edgardomortiz/fineRADstructure-tools> (Original work published 2017)
- Pembleton, L. W. (2017). StAMPP: Statistical Analysis of Mixed Ploidy Populations (Version 1.5.1). Retrieved from <https://CRAN.R-project.org/package=StAMPP>
- Pembleton, Luke W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. doi: 10.1111/1755-0998.12129

- Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9(6), 1638–1647. doi: 10.1111/2041-210X.12984
- Peterman, W. E., Connette, G. M., Semlitsch, R. D., & Eggert, L. S. (2014). Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, 23(10), 2402–2413. doi: 10.1111/mec.12747
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, 7(5), e37135.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet*, 8(11), e1002967. doi: 10.1371/journal.pgen.1002967
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16(3), 608–627. doi: 10.1111/1755-0998.12512
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from <https://www.R-project.org/>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Datasets. *Genetics*, genetics.114.164350. doi: 10.1534/genetics.114.164350
- Richmond, J. Q., Barr, K. R., Backlin, A. R., Vandergast, A. G., & Fisher, R. N. (2013). Evolutionary dynamics of a rapidly receding southern range boundary in the threatened California Red-Legged Frog (*Rana draytonii*). *Evolutionary Applications*, 6(5), 808–822. doi: 10.1111/eva.12067
- Rohr, J. R., & Raffel, T. R. (2010). Linking global climate and temperature variability to widespread amphibian declines putatively caused by disease. *Proceedings of the National Academy of Sciences*, 107(18), 8269–8274. doi: 10.1073/pnas.0912883107
- Scheele, B. C., Pasmans, F., Skerratt, L. F., Berger, L., Martel, A., Beukema, W., ... Canessa, S. (2019). Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science*, 363(6434), 1459–1463. doi: 10.1126/science.aav0379
- Semlitsch, R. D., Peterman, W. E., Anderson, T. L., Drake, D. L., & Ousterhout, B. H. (2015). Intermediate Pond Sizes Contain the Highest Density, Richness, and Diversity of Pond-Breeding Amphibians. *PLOS ONE*, 10(4), e0123055. doi: 10.1371/journal.pone.0123055
- Shah, V., & McRae, B. (2008). *Circuitscape: a tool for landscape ecology*. 7, 62–66.

- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smith, K. G., Lips, K. R., & Chase, J. M. (2009). Selecting for extinction: nonrandom disease-associated extinction homogenizes amphibian biotas. *Ecology Letters*, *12*(10), 1069–1078. doi: 10.1111/j.1461-0248.2009.01363.x
- Sonderegger, D. (2018). SiZer: Significant Zero Crossings (Version 0.1-5). Retrieved from <https://CRAN.R-project.org/package=SiZer>
- Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S., Fischman, D. L., & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, *306*(5702), 1783–1786.
- Takezaki, N., & Nei, M. (1996). Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA. *Genetics*, *144*(1), 389–399.
- Thomson, R. C., Wright, A. N., & Shaffer, H. B. (2016). *California Amphibian and Reptile Species of Special Concern*. UC Press.
- Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*(2), 291–307. doi: 10.1111/ecog.02880
- Unglaub, B., Steinfartz, S., Drechsler, A., & Schmidt, B. R. (2015). Linking habitat suitability to demography in a pond-breeding amphibian. *Frontiers in Zoology*, *12*. doi: 10.1186/s12983-015-0103-3
- U.S. Fish and Wildlife Service. (2015). Endangered and Threatened Wildlife and Plants; 90-Day Findings on 31 Petitions. *Federal Register*, *80*(50 CFR 17), 80583–80614.
- Wake, D. B., & Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0801921105
- Walls, S. C., Barichivich, W. J., & Brown, M. E. (2013). Drought, Deluge and Declines: The Impact of Precipitation Extremes on Amphibians in a Changing Climate. *Biology*, *2*(1), 399–418. doi: 10.3390/biology2010399
- Wang, I. J. (2009). Fine-scale population structure in a desert amphibian: landscape genetics of the black toad (*Bufo exsul*). *Molecular Ecology*, *18*(18), 3847–3856.
- Wang, I. J., Johnson, J. R., Johnson, B. B., & Shaffer, H. B. (2011). Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, *Ambystoma californiense*. *Conservation Genetics*, *12*(4), 911–920. doi: 10.1007/s10592-011-0194-0

- Wang, I. J., Savage, W. K., & Shaffer, H. B. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Molecular Ecology*, *18*(7), 1365–1374. doi: 10.1111/j.1365-294X.2009.04122.x
- Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, *117*(4), 233–240. doi: 10.1038/hdy.2016.60
- Waples, R. S., & Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: a cautionary view. *Molecular Ecology*, *26*(5), 1211–1224. doi: 10.1111/mec.14022
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. doi: 10.2307/2408641
- Willson, J. D., & Hopkins, W. A. (2013). Evaluating the Effects of Anthropogenic Stressors on Source-Sink Dynamics in Pond-Breeding Amphibians. *Conservation Biology*, *27*(3), 595–604. doi: 10.1111/cobi.12044
- Winter, D., Green, P., Kamvar, Z., & Gosselin, T. (2017). mmod: Modern Measures of Population Differentiation (Version 1.3.3). Retrieved from <https://CRAN.R-project.org/package=mmod>
- Winter, D. J. (2012). mmod: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, *12*(6), 1158–1160. doi: 10.1111/j.1755-0998.2012.03174.x
- Zamberletti, P., Zaffaroni, M., Accatino, F., Creed, I. F., & De Michele, C. (2018). Connectivity among wetlands matters for vulnerable amphibian populations in wetlandscapes. *Ecological Modelling*, *384*, 119–127. doi: 10.1016/j.ecolmodel.2018.05.008

Chapter 3: Comparative population and landscape genomics of two cryptic, allopatric clades suggests novel conservation strategies in a threatened amphibian

Kevin Michael Neal

Abstract

Proper identification of conservation units within cryptic species remains a challenge in biological conservation, and increasing levels of environmental change and population extirpation make it more important than ever to describe and understand biological diversity among species and populations before it is lost. The western spadefoot *Spea hammondi* (Anura:Scaphiopodidae) of California and Mexico was recently found to comprise two cryptic monophyletic lineages isolated by a strong phylogeographic barrier. However, that study was based on limited genetic data and lacked the resolution to make population-level inferences that would best support conservation status assessments and management. In this study I used thousands of genome-wide SNPs obtained via RADseq to examine evolutionary relationships of *S. hammondi*. I validated species-level relationships in the genus *Spea* using phylogenetic methods and PCA, and applied F-statistics, genetic clustering, and models of landscape resistance to characterize population differentiation and spatial genetic clustering. I confirmed two highly supported monophyletic clades within *S. hammondi*, and in both clades I found both subtle regional and strong pond-level structuring consistent with a pattern of isolation-by-distance (IBD). All ponds had exceptionally low genetic effective population sizes. Landscape resistance models supported the importance of IBD but found a much stronger influence of

environmental variation on the observed genetic variation in both northern and southern clades. Based on modeled resistance, I examined the effects of urbanization on functional connectivity of populations and found that urbanization in the South clade has completely isolated several populations that once naturally experienced gene flow. The results revealed high levels of interpopulation genetic differentiation and low effective population sizes that limit the adaptive potential of both clades. Western spadefoots will likely require active management to maintain connectivity among populations, with a particularly dire need for efforts in the South clade due to the degree of population isolation from widespread urbanization.

Introduction

The identification of distinct, non-interacting phylogenetic lineages as units of conservation within species or species complexes is of critical importance to biological conservation. Recently diverged lineages separated by space and time may differ in their environmental niches and interactions with the environment, and accordingly are subjected to differing evolutionary forces that may merit independently-derived management strategies specific to each lineage. Identification of these conservation units is a challenge in nominal species with cryptic lineages given the lack of morphological distinctiveness between the lineages (Bickford et al., 2007), but dense geographic sampling with genetic data has proved useful for clarifying these evolutionary relationships and identifying appropriate conservation units.

A recent study (Neal, Johnson, & Shaffer, 2018) confirmed two previously-identified cryptic, reciprocally monophyletic clades in the western spadefoot, *Spea hammondi* (García-París, Buchholz, & Parra-Olea, 2003), with each clade separated by an uninhabited geographic

barrier with extremely low predicted habitat suitability (Fig. 3.1). The Tehachapi Mountains and Transverse Ranges, a known phylogeographic barrier for other California herpetofauna (Rissler, Hijmans, Graham, Moritz, & Wake, 2006), have never produced a museum specimen of *S. hammondii* and may represent a barrier to gene flow sufficient for allopatric speciation. Although Neal et al. (2018) identified these clades as distinct based on high phylogenetic support values, clustering analyses, and niche differentiation, that analysis was based on limited genetic sampling that could have resulted in a false signal and painted an erroneous picture of the evolutionary history of *S. hammondii*. Robust identification of genetic conservation units within *S. hammondii* is of pressing importance, given that the nominal species is currently being considered for listing under the U.S. Endangered Species Act (ESA). If the nominal species as a whole can be found to warrant listing, distinct lineages within it may constitute the recovery units that guide management. If those units are considered separate species, then each may be at even greater risk of extinction due to their necessarily smaller ranges and numbers of populations (Bickford et al., 2007; Schönrogge et al., 2002). Regardless of their taxonomic status, each presumptive clade may be subject to unique combinations and levels of threats. Indeed, North and South clades of *S. hammondii* (referred to as "SPHA-NORTH" and "SPHA-SOUTH", respectively, through the rest of the manuscript) currently face differing threat levels in the form of habitat loss and fragmentation: Jennings and Hayes (1994) found a 30% loss of suitable habitat in the range of the SPHA-NORTH and an 80% loss in the range of SPHA-SOUTH. At about the same time, a summary of available field studies confirmed widespread extirpation of *S. hammondii* from several historically-occupied sites in the range of SPHA-NORTH (Fisher & Shaffer, 1996), although no published data on population trends existed for SPHA-SOUTH. Habitat loss certainly is ongoing as the human population of California continues to grow, from

an estimate of 31.3 million people in 1994 to 39.9 million in 2019 (California Department of Finance, 2019), , and comparative studies of the species across the Central Valley and Southern California populations has been called out as a particularly important research need for effective management (Thomson, Wright, & Shaffer, 2016).

Given the extent of habitat loss and population extirpation in both clades of *S. hammondii*, it is crucial to examine patterns of genetic variation within each clade to make optimal management recommendations and provide reliable results for the evaluation of *S. hammondii* as a candidate species for listing under the ESA. In this study I used restriction site-associated DNA sequencing (RAD-seq) to generate a large genome-wide dataset for *S. hammondii*. RAD-seq is a reliable, widely-used, and inexpensive method of collecting genome-wide sequence data with demonstrated utility for studying intraspecific genetic variation (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Catchen et al., 2017). The use of thousands of genome-wide markers enables accurate identification of cryptic lineages and structure (Linck et al., 2019; McCartney-Melstad, Gidiş, & Shaffer, 2018; Reyes-Velasco, Manthey, Bourgeois, Freilich, & Boissinot, 2018; Spinks, Thomson, & Shaffer, 2014) and detection of fine-scale, recent divergence among individual populations (Bryson, Savary, Zellmer, Bury, & McCormack, 2016; McCartney-Melstad, Vu, & Shaffer, 2018; Sánchez-Ramírez et al., 2018) at resolutions unobtainable with the previous range-wide analysis of *S. hammondii* using only five nuclear genes and one mitochondrial gene (Neal et al., 2018). Using this much larger genomic dataset, I address several questions: 1) How is *S. hammondii* related to other *Spea* species, and what are the major phylogenetic units and their interrelationships within *S. hammondii*? 2) What further genetic structuring emerges within clades of *S. hammondii*? 3) What environmental variables best explain that within-clade genetic variation? 4) How do

measures of genetic diversity and distance compare between clades? 5) Assuming that the effects of modern urbanization are not reflected in the assessments of interpopulation genetic distance, what effect can we expect urbanization to have on functional connectivity of populations?

Methods

DNA extraction and sequencing

I obtained tissue samples from all four nominal *Spea* species: from *S. hammondii* throughout California (localities in Fig. 3.1) and one location in Mexico; *Spea intermontana* from California and Oregon; *Spea bombifrons* from Wyoming and Oklahoma; and *Spea multiplicata* from Oklahoma. As an outgroup for phylogenetic analysis, I also obtained tissues from two of the three species in the sister genus *Scaphiopus* (Chen, Bever, Yi, & Norell, 2016; Pyron & Wiens, 2011; Zeng, Gomez-Mestre, & Wiens, 2014), with one sample of *Scaphiopus couchii* from Mexico and two *Scaphiopus holbrookii* from New York. Genomic DNA was extracted using either a standard salt extraction protocol or the Kingfisher extraction protocol (Thermoscientific), followed by the 3RAD RAD-seq protocol (Bayona-Vásquez et al., 2019) to isolate genome-wide fragments of DNA for sequencing. 3RAD adds a third enzyme to ddRAD (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) that cuts adapter dimers, increasing on-target sequencing efficiency (Bayona-Vásquez et al., 2019). I used Illumina iTru adapters as well as custom internal adapters (Integrated DNA Technologies) for multiplexing, and I used PstI-HF (dimer cutter), NsiI-HF (rare cutter), and MspI (common cutter) (New England BioLabs) for the three restriction enzymes. Libraries were pooled with 200-300 ng of DNA per individual and BluePippin (Sage Science) was used to size select 350-450 base pair fragments. Pooled libraries were sent to the Vincent J. Coates Genomics Sequencing Laboratory at University of California,

Berkeley, for sequencing on an Illumina HiSeq 4000 with either PE150, SR100, or SR150, although only the higher-quality first reads (R1's) were used for downstream protocols.

DNA post-processing

Internal adapters were trimmed from the demultiplexed sequences using cutadapt (Martin, 2011). As an additional filtering step, fastp (Chen, Zhou, Chen, & Gu, 2018) was used to filter reads below 50% complexity (percentage of bases that are different from the next base in a sequence) and with a mean phred quality score below Q30. To capture the highest-quality portions at the start of each read and because of differences of read lengths among SR100 and SR150/PE150 sequencing runs, every read was trimmed to a maximum of 60 basepairs and reads less than 30 basepairs in length were removed. I then used ipyrad 0.7.29 (Eaton, 2014, 2019) to assemble the RAD loci. I ran ipyrad using a reference draft genome that was sequenced from a North *S. hammondii* metamorph from Santa Barbara County, California. The genome was sequenced by 10X Genomics and assembled using Supernova (10X Genomics) and analyzed using BUSCO to inspect the quality of the assembly (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). BUSCO results can be found in the Appendix.

I ran ipyrad from steps 1 through 6 using all available sequences and created several branches at step 7 with different subsets of individuals specific to my analyses. After preliminary phylogenetic and genetic clustering analyses on the full dataset, I identified major clusters and subsampled individuals from those to create a reduced dataset of *Spea* and *Scaphiopus* samples to allow for faster and more evenly sampled phylogenetic analysis while still having representation of all major clusters. For North and South *S. hammondii* analyses, after preliminary clustering analyses on each clade, I subsampled sites within clusters when sites were

both well-sampled and in very proximal geographic groups in order to limit effects of spatial autocorrelation (Bradburd, Coop, & Ralph, 2018; Meirmans, 2012) and to limit the potential bias of uneven sampling in clustering analyses (Puechmaille, 2016). Loci missing in more than 50% of individuals were filtered from each dataset.

From the ipyrad-generated VCF files, I used either `vcftools` 1.1.5 (Danecek et al., 2011) or `filter_rad()` in `radiator` 1.1.0 (Gosselin, 2019) in R 3.5.1 (R Core Team, 2018) for additional filtering for SNP-based analyses (everything except `fineRADstructure`). Either with `vcftools` or `radiator`, all datasets were filtered to one SNP per locus, removed monomorphic loci, removed SNPs with a minor allele count less than three to ensure every SNP was found in at least two individuals, and removed SNPs with greater than 50% missing data. For phylogenetic analyses I then converted the VCF file to a PHYLIP using a python3 script (Ortiz, 2019). For population genetic analyses within clades, I used `radiator` for SNP filtering because of its easier integration into more elements of this study's mostly R-based population genetic pipeline and its superior data exploration and manipulation. Also, while `vcftools` selected the first SNP in each locus, I used `radiator` to select the SNP with the highest minor allele count in each locus.

Data analysis

Phylogeny

To examine the phylogenetic relationships of *Spea* species and to identify lineages for further downstream analysis, I ran a maximum likelihood analysis on concatenated SNPs (using one SNP per locus) with up to 20% missing data per locus in IQ-TREE 1.6.9 (Nguyen, Schmidt, von Haeseler, & Minh, 2015), a program with accuracy comparable to RAxML (Nguyen et al., 2015; Zhou, Shen, Hittinger, & Rokas, 2018). I ran IQ-TREE using the built-in ModelFinder

(Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermini, 2017) to select the optimal substitution model based on the Bayesian Information Criterion (BIC); applied ascertainment bias correction (Lewis, 2001); used ultrafast bootstrapping (Minh, Nguyen, & von Haeseler, 2013) with 1000 replicates; optimized bootstrap trees using nearest-neighbor interchange; and applied a second test of node support using 1000 replicates of the Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) (Guindon et al., 2010). I included *Scaphiopus holbrookii* and *Scaphiopus couchii* as outgroups. With higher level relationships established with 100% bootstrap support, I conducted another maximum likelihood analysis in IQ-TREE with up to 50% missing data per locus with representative individuals from all sampled *S. hammondii* ponds (SPHA-NORTH and SPHA-SOUTH together), with *S. intermontana* as the designated outgroup. To visualize the species-level distinctiveness found in the phylogenetic analyses, I also conducted principle components analyses (PCA) using pcadapt 4.1.0 (Luu, Bazin, & Blum, 2017), using the same dataset with SPHA-NORTH, SPHA-SOUTH, and *S. intermontana*, followed by a second PCA with only SPHA-NORTH and SPHA-SOUTH.

In SPHA-SOUTH, the single sample from Mexico emerged as the sister lineage to all other SPHA-SOUTH samples. Because of the extremely high genetic and geographic distance from all other samples and the fact it is only one individual, I removed this sample from the final population-level analyses and note the caveat that the population-level results apply specifically to *S. hammondii* within the boundaries of the United States.

Population structure

Following confirmation of two highly supported, reciprocally monophyletic clades in *S. hammondii* (Fig. 3.2), I used several complementary approaches to assess population genetic

structure within each. I used `radiator::fst_WC84()` to calculate global F_{ST} and F_{IS} values using the method of Weir and Cockerham (1984), as well as pairwise F_{ST} values among sites. Individual ancestry coefficients on the SNP data were calculated using the sparse non-negative matrix factorization (sNMF) method (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014) implemented in the R package LEA (Frichot & François, 2015). sNMF has been shown to perform as well as Structure and ADMIXTURE with a much faster run-time for use with genomic datasets, and sNMF is robust to deviations from Hardy-Weinberg equilibrium, having no a priori population genetic assumptions (Frichot et al., 2014). TESS3 (Caye, Deist, Martins, Michel, & François, 2016) in the R package `tess3r` was then used to assess population genetic structure, treating geographic coordinates as priors using an alternating projected least squares algorithm (Caye, Jay, Michel, & François, 2018). I used `tess3r::plot.tess3Q()` and `fields::Krig()` to map and interpolate the dominant genetic clusters for all values of K assessed. I ran both sNMF and TESS3 for K values from 1 to 20, with 10 repetitions for each K, a maximum of 200 iterations per run, and masked 10% of genotypes per run to compute the cross-entropy criterion. I selected the repetition for each K with the lowest cross-entropy to generate ancestry barcharts and maps. In addition to examining cross-entropy, I ran `adegenet::snapclust.choose.k` in `adegenet` 2.1.1 (Jombart, 2008) on the same genetic data input for sNMF and TESS3 to determine optimal K values using the Bayesian Information Criteria.

Landscape resistance

I used ResistanceGA 4.0-14 (Peterman, 2018, 2019) in R to determine relative contributions of geographic distance and environmental variation to genetic differentiation within both *S. hammondii* clades. While landscape genetic analyses have often used expert

opinion or trial and error to determine resistance values for environmental layers within a limited parameter space, ResistanceGA uses a genetic algorithm and linear mixed effect models to determine the optimal transformation of an environmental surface into a resistance surface, based on iterative improvements in the correlation of the genetic distance (response variable) between two sites with the resistance distance (predictor variable) between those sites. ResistanceGA then ranks the optimized resistance surfaces and returns the surfaces as raster maps. For the input environmental surfaces, I began with a set of 0.00833 decimal degree resolution variables from CHELSA 1.2 (Karger et al., 2017), ENVIREM (Title & Bemmels, 2018), the 2011 National Landcover Database (Homer et al., 2015), select SoilGrids (Hengl et al., 2017) layers. I went through a number of steps to reduce the number of variables due to prohibitively long computation times. I first ran ResistanceGA on all the layers with only two optimization iterations and used log-likelihood (LL) to select the top five or six best-fit layers in the final ResistanceGA run. I then used RStoolbox::rasterPCA() (Leutner, Horning, Schwalb-Willmann, & Hijmans, 2019) to reduce all of the continuous (i.e. excluding all categorical) layers to their principle components and retained the first five PCs for ResistanceGA. I also included habitat suitability layers generated individually for each clade in Maxent 3.4.1 (Phillips, Anderson, Dudík, Schapire, & Blair, 2017). In Maxent I included all CHELSA, ENVIREM, and SoilGrids layers, excluding layers related to human development (landcover and impervious surface) to better reflect pre-human habitat suitability. I ran Maxent with ten cross-validation replicates and applied a complementary log-log (cloglog) transform to produce an estimate of occurrence probability as a proxy for habitat suitability, and I used the mean of all cross-validation replicates as the final surface for ResistanceGA. For Maxent inputs, I combined GBIF (GBIF.org, 2017) localities with the current study's sample localities and used spThin::spThin() (Aiello-Lammens,

Boria, Radosavljevic, Vilela, & Anderson, 2015) to spatially subsample them to one site per 10 km radius. To reduce computation time for the large North layers, I reduced the resolution of the North layers using `raster::aggregate(fac=2)` to a resolution of 0.01667 decimal degrees for final ResistanceGA runs. Final Maxent habitat suitability models and the presence points for both clades are shown in Fig. 3.1.

In ResistanceGA, I calculated resistance distance using `gdistance::commuteDistance()` (Etten, 2017, 2018), which determines distance as an average of random walks between points on the environmental surface. I used `hierfstat::genet.dist()` (Goudet & Jombart, 2015) to calculate pairwise Nei's D_a between the sample sites for the genetic distance. ResistanceGA was run for a maximum of 200 iterations per surface, with the optimization ending if it failed to improve model log-likelihood for 50 consecutive iterations.

Comparative population and landscape genomics

In dividing *S. hammondii* into two reciprocally monophyletic clades, each clade merits unique conservation consideration given their independent evolutionary histories and exposure to different suites of environmental conditions (Neal et al., 2018). To optimize conservation and management decisions, I compared population-level (i.e. pond-level) genetic diversity, genetic distance, and landscape resistance within each clade. I used `hierfstat::basic.stats()` and `base::colMeans()` to calculate observed heterozygosity (H_o), expected heterozygosity or gene diversity (H_e), and F_{IS} for each pond. I used NeEstimator 2.1 to calculate effective number of breeders (N_{eb}) at each locality. I used `radiator::pi()` to calculate individual- and population-level nucleotide diversity (π), and I calculated the median individual π per population as a second measure of nucleotide diversity. For the calculation of π specifically, I created a separate dataset

in radiator that retained all SNPs in each locus (whereas for all others I retained one SNP per locus), with up to six SNPs per locus. I used `radiator::ibd_g_fh` to calculate individual-level estimates of inbreeding (F_H) based on excess SNP homozygosity (Kardos, Luikart, & Allendorf, 2015; Keller, Visscher, & Goddard, 2011; Purcell et al., 2007), and I calculated the median individual F_H per population. Additionally, I calculated the median Maxent suitability, median elevation, and maximum percent impervious surface within a 2 km radius of each sample site, using `raster::extract()` in R. With these population-level values for each clade, I used `ggplot2` to generate boxplots for each variable for each clade, and I used `ggpubr::stat_compare_means` with `stats::wilcox.test` to perform a Wilcoxon rank sum test for a significant difference in the values between the North and South.

I used the top-ranked resistance surface for each clade to look at the potential impacts of recent urbanization and development on functional connectivity in each region. To do this I used a National Land Cover Database (Homer et al., 2015) raster of percent impervious surface, ranging from 0 to 100%. I derived two rasters to represent urbanized areas as impassable terrain, assigning either >20% or >50% impervious surface with infinite resistance by setting these pixels to NA using the raster package in R 3.5.1. These thresholded impervious surface rasters were used to mask ResistanceGA-optimized resistance surfaces for each clade, and masked pixels on the resistance surfaces were assigned a resistance value of 10,000 – roughly 10 times higher than the highest value in all resistance surfaces assessed. Resistance surfaces were also masked with the clade-specific Maxent habitat suitability rasters (Fig. 3.1), setting any pixels below 0.1 suitability to a resistance value of 10,000 (intended to discourage highly unlikely alternate commute paths through biologically impassable areas like high mountains). I used `gdistance::commuteDistance()` to calculate commute distances among sites for the three levels of

urbanization and used `ggpubr::ggpaired()` to generate boxplots of the distances to visualize what effect urbanization had on functional connectivity. Paired Wilcoxon signed rank tests were used to test for significant differences between resistance distances with and without the impervious surface masks.

Results

RAD-seq data

For the phylogenetic dataset with all *Spea* and *Scaphiopus* as outgroup ("SCAPH22" dataset), `ipyrad` returned 295,584 loci across 22 individuals after filtering for duplicate sequences and maximum number of indels, SNPs, and heterozygous sites in a locus. After filtering for loci found in at least 50% of individuals and a maximum of two alleles per site, `ipyrad` returned 42,549 loci. Using `vcftools` 1.1.5 to filter for only the first SNP in a locus, variable sites, and SNPs with minor allele count (MAC) of at least 3 returned 28,198 SNPs for the final phylogenetic dataset of all *Spea* species run in IQ-TREE. For the *S. hammondii* (SPHA-NORTH + SPHA-SOUTH) + *S. intermontana* dataset ("SPHASPIN188"), `ipyrad` returned 547,625 loci across 188 individuals (29,352 in at least 50% of individuals) and `vcftools` returned 21,747 SNPs (one SNP per locus, variable sites only, $MAC \geq 3$).

`Ipyrad` returned 543,590 loci in 124 individuals for SPHA-NORTH alone, dropping to 29,762 after filtering for loci found across at least 50% of individuals and maximum two alleles per SNP. Filtering in `radiator` for the SNP with the highest allele frequency in each locus and MAC of at least 3 yielded 13,524 SNPs with 16.54% total missing data. For SPHA-SOUTH, `ipyrad` returned 545,431 loci across 178 individuals, dropping to 37,956 after filtering for loci found in at least 50% of individuals and maximum two alleles per SNP. Filtering in `radiator` for

the SNP with the highest allele frequency in each locus and a minimum minor allele count of 3 yielded 5,309 SNPs with 16.07% total missing data.

Phylogenetic analysis

In the SCAPH22 dataset, the highest-likelihood IQ-TREE identified all nominal species as monophyletic (although only 1 sample of *S. multiplicata* was available), all with maximum bootstrap support (100) and SH-aLRT support (100). The topology of this tree was the same as the nuclear gene topology recovered in Neal et al. (2018), with *S. multiplicata* sister to *S. bombifrons* + *S. intermontana* + *S. hammondii*, and *S. bombifrons* sister to *S. intermontana* + *S. hammondii* (Fig. 3.2). *Spea. hammondii* split into Northern (SPHA-NORTH) and Southern (SPHA-SOUTH) clades (bootstrap and SH-aLRT support = 100 for each), with the two *S. hammondii* clades representing geographically coherent entities divided by the Transverse Ranges and Tehachapi Mountains in California (Fig. 3.1); these distinctions are reinforced by *S. intermontana*, SPHA-NORTH, and SPHA-SOUTH all forming tightly-clustered groups in the PCAs (Fig. 3.3).

In the SPHASPIN188 tree (Fig. 3.4), *S. intermontana*, *S. hammondii*, SPHA-NORTH, and SPHA-SOUTH again formed reciprocally monophyletic clades with maximal support (100% bootstrap/100% SH-aLRT). SPHA-NORTH had high support values for most groups of individuals within ponds and for geographic clusters of nearby ponds, but poor resolution for the relationships among clusters. Two populations north of the Sacramento-San Joaquin River Delta (SSJRD), a phylogeographic break that has been observed in many California species (McCartney-Melstad et al., 2018; Rissler et al., 2006), formed a highly-supported clade (100% bootstrap/100% SH-aLRT support); although they appear highly divergent in the SPHA-

NORTH+SPHA-SOUTH PCA (Fig. 3.3B) and in clustering analyses (see below in Results), in the tree they are nested within a group of ponds at the northern end of the Central Valley (but south of the SSJRD phylogeographic break). These northern Central Valley populations with the north-of-SSJRD populations appeared as sister to the rest of SPHA-NORTH in the tree, but this relationship had very low support (43% bootstrap/33.8% SH-aLRT).

Like SPHA-NORTH, internal support values in SPHA-SOUTH had mostly high support values for groups of individuals within ponds and for geographic clusters of nearby ponds, but poor resolution in the relationships among clusters. The single sample from Mexico (UABC1148) emerged as sister to all other SPHA-SOUTH individuals and appeared distinct in the PCA. SPHA-SOUTH without the single Mexican sample had 100% bootstrap and 100% SH-aLRT support in both the SCAPH22 and SPHASPIN188 trees. Orange County, Los Angeles, Ventura, and Riverside County populations together formed a well-supported monophyletic group (95% bootstrap and 99.9% SH-aLRT support) nested within paraphyletic San Diego populations. Nested within the Orange/Los Angeles/Ventura/Riverside clade, Orange County Coastal populations formed a well-supported clade (95% bootstrap and 95.9% SH-aLRT support).

Analyses of Spea hammondi - North clade

Population structure

The global F_{ST} calculated from the SNPs was 0.358 (95% bootstrap confidence interval: 0.354-0.362), indicating high levels of structure in the northern clade of *S. hammondi*, and global F_{IS} was 0.0616. Pairwise F_{ST} and Nei's D_a among ponds were generally high, with the mean of all pairwise F_{ST} values being 0.334 (median = 0.331, 95% quantile: 0.0917 – 0.608) and

the mean of all pairwise Nei's D_a being 0.0785 (median = 0.0757, 95% quantile: 0.0488 – 0.122). Results of both sNMF and TESS3 clustering analyses pointed to similar patterns (Fig. 3.5, 6), and I here focus on results from sNMF, with TESS3 used to visualize clustering geographically. sNMF and TESS3 cross-entropy values consistently decreased with K , suggesting a pattern of isolation by distance (Fig. 3.S1A, B). The lowest cross-entropy value at $K=15$ nearly coincided with the number of ponds with three or more individuals sampled ($N=20$ such ponds), although sNMF cross-entropy plateaued around $K=10$. Values of K above 10 tended to split single ponds into clusters with little or no decrease in cross-entropy. Pond-level isolation is consistent with the philopatric life history exhibited by spadefoot species and may be exacerbated by the relatively discontinuous geographic sampling in this study. Lower values of K still showed relatively little admixture among identified clusters. In SPHA-NORTH, there is an arguable elbow in the cross-entropy curve at $K=3$ (Fig. 3.S1A), and BIC also selects $K=3$ as the optimum K value (Fig. 3.S1C). Additional clusters from $K=4$ and up typically identified single ponds. Combined, these results point to $K=3$ as the optimum. At $K=3$, clusters appeared at the northern and southern edges of the North clade range (Fig. 3.5, 6), with clusters separated from the core Central Valley cluster by known physiographic features. These features—the San Joaquin-Sacramento River Delta in the north portion of the range and the Coast Ranges in the southwest—have been observed as strong genetic barriers in other species as well (McCartney-Melstad et al., 2018; Rissler et al., 2006; Shaffer, Pauly, Oliver, & Trenham, 2004).

Landscape resistance

Results from ResistanceGA showed a strong effect of isolation by environment, with every included environmental variable ranking higher than geographic distance alone in

explaining variance in genetic distance. The transformed Maxent habitat suitability layer (Fig. 3.7) was the best fit model based on log-likelihood (LL), with LL = 1408.8 and marginal R^2 (R^2_m) of 0.872, indicating that the transformed Maxent surface was able to explain 87.2% of the variance in genetic distance among SPHA-NORTH populations (Nakagawa & Schielzeth, 2013). Values of suitability were transformed to resistance using an inverse monomolecular transform, such that the lowest suitability values had the highest resistance. Resistance exponentially decreased as suitability increased, suggesting that higher Maxent suitability facilitates gene flow. Climatic moisture index was the second-best model (LL = 1385.5, $R^2_m = 0.771$). Geographic distance (LL = 1240.3) explained only 26.8% ($R^2_m = 0.268$) of the variance among SPHA-NORTH populations.

*Analyses of *Spea hammondi* - South clade*

Population structure

The global F_{ST} calculated from the SNPs was 0.275 (95% confidence intervals: 0.269-0.281), indicating lower, but still substantial, levels of structure in SPHA-SOUTH compared with SPHA-NORTH, and global F_{IS} was near zero at -0.0288. Pairwise F_{ST} and Nei's D_a among ponds are generally high, with the mean of all pairs of F_{ST} being 0.280 (median = 0.288, 95% quantile: 0.0553 – 0.470) and the mean of all pairs of Nei's D_a being 0.0619 (median = 0.0626, 95% quantile: 0.0321 – 0.0872). As with SPHA-NORTH, results of both sNMF and TESS3 clustering analyses pointed to similar patterns (Fig. 3.8, 9), and I again focus on results from sNMF, with TESS3 used to visualize clustering geographically. sNMF and TESS3 cross-entropy values consistently decreased with K, suggesting a pattern of isolation by distance (Frichot et al., 2014) (Fig. 3.S1D,E). The lowest cross-entropy value was at K=16, nearly coinciding with the

number of ponds with three or more individuals sampled (19), Although not the lowest cross-entropy value, sNMF cross-entropy plateaued at $K=7$ (before forming another plateau at $K=10$). Values of K above 10 tended to split single ponds into clusters with little or no decrease in cross-entropy. Before plateauing, sNMF cross-entropy showed a subtle elbow at $K=2$, and BIC also supported $K=2$ as the highest level of clustering in the SPHA-SOUTH (Fig. 3.S1F). sNMF and TESS3 results at $K=2$ indicated a northern and southern cluster, with strong admixture at ponds located at the interface of the ranges of the two clusters in Riverside County (BOXSPR) and far northern San Diego County (FLORESRD, PENDLETONBRAVO2, and PENDLETONJULIETT). At $K=3$, the sites of the Orange County Coast emerged as a cluster, corroborating the distinctiveness of Orange County Coastal populations identified in Chapter 2. At $K=4$, sites in Riverside County formed a cluster. From $K=5$ and up, individual sites tended to be pulled out as clusters. The ponds in Los Angeles and Ventura Counties (BOUQ, HCC, and CAST), which are geographically very distant from the next closest sampled ponds in Orange County and separated by both the entire Los Angeles metropolitan area and the Santa Monica Mountains, remained clustered with inland Orange County populations until $K=6$.

Landscape resistance

ResistanceGA showed a strong effect of isolation by environment in SPHA-SOUTH. Whereas geographic distance alone explained only 26.8% of the variance in genetic distance in the northern clade, in the southern clade it explained 66.9%, despite being ranked at the bottom among surfaces tested ($R^2_m = 0.669$, LL = 1763.8) (Table 3.S2). Most environmental variables ranked higher than geographic distance, however, with a transformed climatic moisture index surface being the best fit landscape resistance model (LL = 1800.4, $R^2_m = 0.732$) (Fig. 3.10),

although topographic wetness actually had a higher marginal R^2 ($R^2_m = 0.798$, $LL = 1796.1$). The transform applied to the climatic moisture index surface (and to topographic wetness) was an inverse monomolecular transformation, with the lowest values of climatic moisture having the highest resistance. Resistance values exponentially decaying as climatic moisture increased, suggesting that higher climatic moisture facilitates gene flow. The transformed Maxent habitat suitability surface for SPHA-SOUTH ranked near the bottom in terms of model fit ($LL = 1764.8$), just above geographic distance, but still explained 63% of the variance ($R^2_m = 0.630$) in genetic distance among SPHA-SOUTH ponds.

Comparative population and landscape genomics

The individual inbreeding coefficient F_H , population p_i , median individual p_i , median Nei's D_a , and percent impervious surface within 2 km were significantly different between SPHA-NORTH and SPHA-SOUTH ($p < 0.05$). Median Nei's D_a was significantly lower in SPHA-SOUTH, while the other significant variables were higher in SPHA-SOUTH (Fig. 3.11). Effective number of breeders (N_{eb}) did not differ significantly between SPHA-NORTH and SPHA-SOUTH, but N_{eb} estimates for ponds in both clades were extraordinarily low, with a median of 5.25 (range: 2.3–18.3) in SPHA-NORTH and 4 (range: 1.4–20.7) in SPHA-SOUTH.

I found that percent impervious surface, as a proxy for urbanization, has a small but significant effect on functional connectivity in SPHA-NORTH (Fig. 3.12, 13A), with the median distance decreasing 0.5% (the decrease can be attributed to the random walks used to calculate commute distance) when masking >20% impervious surface pixels and a paired Wilcoxon test showing a small but significant difference between unmasked and impervious-masked surfaces ("none" versus "impervious50": $p = 0.013$; "none" versus "impervious20": $p = 0.152$;

"impervious50" versus "impervious20": $p = 0.013$). In SPHA-SOUTH, urbanization has a stark impact on range-wide functional connectivity (Fig. 3.12, 13B), with the median resistance distance increasing more than 3100% when masking pixels with more than 20% impervious surface. Paired Wilcoxon tests showed highly significant differences among unmasked and masked resistance distances ("none" versus "impervious20": $p < 2E-16$). The distribution of distances on the masked surfaces are highly skewed as some populations are effectively cut off from the others while closer populations tend to see little to no impact.

Discussion

The US Fish and Wildlife Service recently determined that listing *S. hammondi* as a federally threatened or endangered species may be warranted due to habitat destruction and fragmentation from urban and agricultural development (Davidson, Shaffer, & Jennings, 2002; Fisher & Shaffer, 1996; Morey, 1998; Morey & Guinn, 1992), vehicle-related mortality (Morey & Guinn, 1992), and the introduction of exotic species (Fisher & Shaffer, 1996). However, many of these earlier studies were based on relatively limited data, and there has been a general call for additional information to conduct a thorough status review (U.S. Fish and Wildlife Service, 2015). This study provides much-needed information for that status review, as well as results relevant to the life history and evolution of this cryptic and poorly-studied amphibian. Results of the current study, supported by those from Neal et al. (2018), indicate that nominal *S. hammondi* likely comprises two species. Adaptive potential of both clades appears to be limited, with ponds in both clades having exceptionally low effective population sizes, high global and pairwise F_{ST} , and high genetic distances among them. Currently the SPHA-SOUTH clade appears to be in more dire need of conservation actions: while the SPHA-NORTH and SPHA-SOUTH do not

appear to be significantly different in many population genetic parameters, functional connectivity among many populations in SPHA-SOUTH is essentially destroyed by the existing urbanization in Southern California. It stands to reason that the SPHA-NORTH will be similarly affected if urbanization or agricultural development lead to fragmentation levels found on the Southern California Coast.

Phylogeny, population structure, and landscape resistance

The phylogenomic analysis of *Spea* confirmed two distinct, reciprocally monophyletic clades within *S. hammondi* with 100% bootstrap and 100% SH-aLRT support, confirming the results of the five-gene nuclear tree of Neal et al. (2018). In addition, within both clades I found strong evidence of deeper population genetic structure. In the SPHA-NORTH, the sNMF clustering model returned a minimum of 3 clusters, consistent with known phylogeographic barriers for other California herpetofauna (McCartney-Melstad et al., 2018; Rissler et al., 2006; Shaffer et al., 2004). Glenn/Yolo County populations (RD2RD86 and DFG18) formed a cluster at the far north end of the Central Valley, separated from the core Central Valley cluster by the Sacramento-San Joaquin River Delta (SSJRD). Santa Barbara County populations (GUAD2, ZACR4) on the southern end of the Central Coast also formed a distinct cluster, mirroring the biogeographic uniqueness of the endangered California tiger salamanders (*Ambystoma californiense*) found in the same pools (Shaffer et al., 2004). Ponds in the Central Valley in close proximity to the peripheral clusters do show admixture with the peripheral clusters (e.g. GILL with the northern cluster; SITE213 and BITTER224 with the southern). The best model of landscape resistance confirmed this model of three SPHA-NORTH clusters, with the SSJRD, Transverse Ranges, and Coast Ranges standing out as high resistance barriers. The resistance

surface identified narrow, low-resistance corridors among the clusters that prevent complete isolation, consistent with observed admixture.

The South clade (United States only) showed a minimum of two clusters based on cross-entropy and BIC, but here the distinction appears to be driven by isolation-by-distance. The clusters are geographically separated, but they do not align with known or obvious candidate phylogeographic barriers. The best model of landscape resistance shows only narrow bands of slightly elevated resistance separating the populations. While the climatic moisture index was the highest ranked resistance surface, distance alone was able to explain a high proportion of the variance in genetic distance in SPHA-SOUTH. Ponds sampled along the interface of these two apparent clusters (e.g. BOXSPR, FLORESRD, PENDLETONBRAVO2) also showed consistently high admixture between both clusters. Phylogenetically, branch lengths and bootstrap support values within SPHA-SOUTH tend to be very low, consistent with relatively shallow divergence among putative groups. The Los Angeles/Ventura/Orange/Riverside cluster did however emerge as monophyletic with high support, nested within a paraphyletic San Diego cluster, possibly indicating a recent recolonization of these northerly counties from southern refugia. The single Mexican specimen appeared in the phylogeny as sister to all other individuals in the SPHA-SOUTH with 100% bootstrap support, but given the geographic separation and the strength of IBD in the rest of the clade, I am hesitant to draw conclusions concerning its genetic (or taxonomic) uniqueness. Rather, I strongly encourage additional analysis with more samples from Mexico that fill in the substantial sampling gap.

*Conservation of *Spea hammondi* in a highly urbanized landscape*

Given the strong philopatry of spadefoots and recency of mega-urbanization in California, the models of landscape resistance in this study should largely reflect historical gene flow with only a weak signal of the effects of modern human development. In both North and South, models showed ample low resistance corridors connecting most populations, and particularly in the South very few ponds appear to be isolated by resistance. Urbanization in Southern California is far more extensive than in areas north of the Transverse Ranges (Fig. 3.12), and it comes as no surprise that species with distributions with more urbanization will be more susceptible to habitat fragmentation and its effects. My attempt to explore the impact of current urbanization used resistance surface masking to quantify the potential effects of existing urbanization on functional connectivity for both SPHA-NORTH and SPHA-SOUTH and emphasizes the extent to which urbanization in SPHA-SOUTH has already fragmented many populations. Existing urbanization in Northern and Central California leaves open much more space for corridors, with only a slightly significant effect when masking out >20% impervious surfaces in SPHA-NORTH. In SPHA-SOUTH, several populations become completely isolated. Although I only tested explicitly for the effects of urbanization, analysis of the effects of agricultural development—a much bigger concern in the range of SPHA-NORTH—and more explicit testing of the effects of different types of urban development (e.g. roads) is critical for determining impacts to the functional connectivity of these clades and for guiding conservation efforts focusing on maintaining natural migration among ponds. Between the observed interpopulation genetic differentiation, exceptionally low pond effective population sizes, and threat of a rapidly-growing human population to functional connectivity, active management will be key in sustaining *S. hammondi* into the future.

Figures

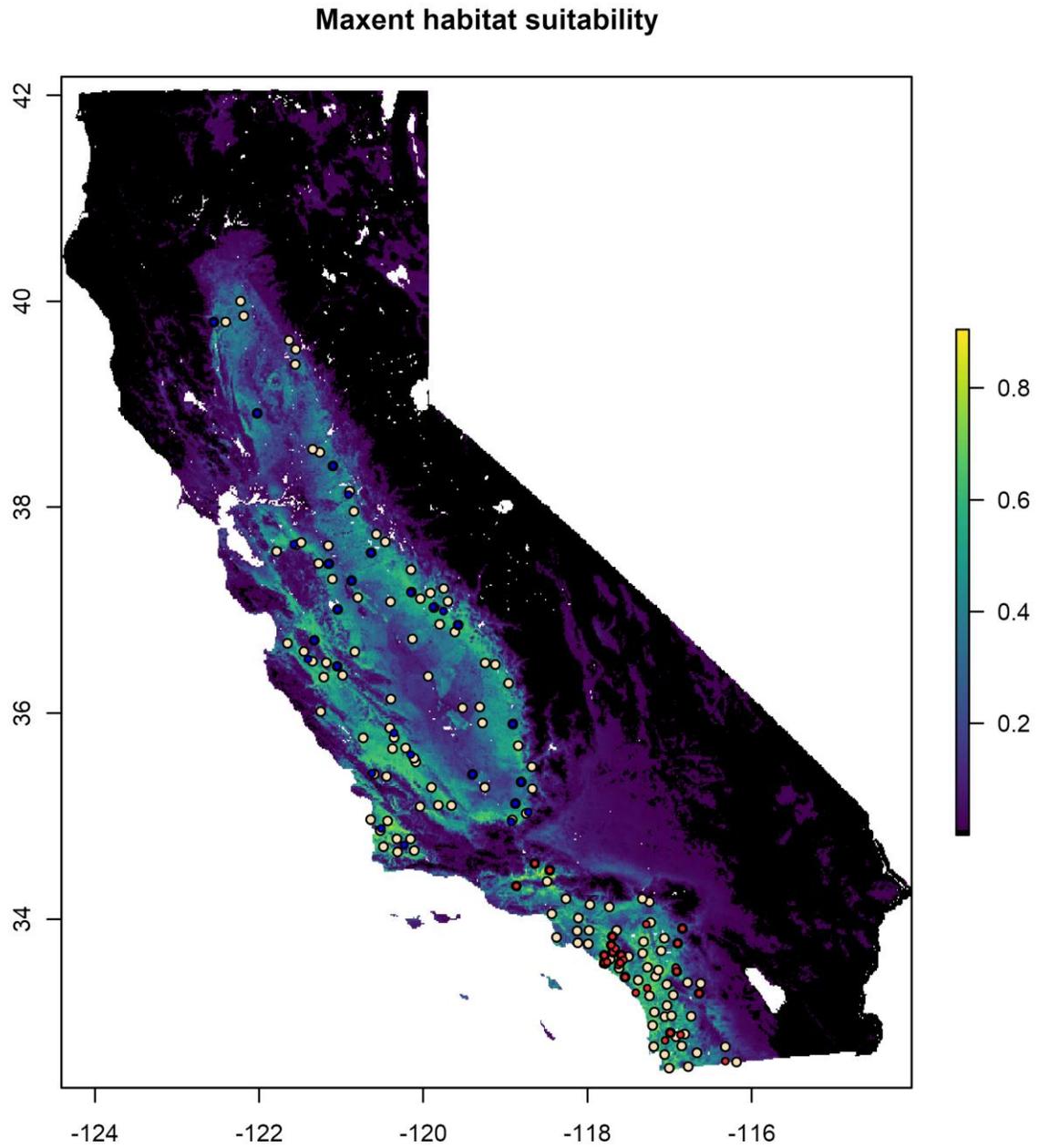


Fig. 3.1 Habitat suitability (measured as probability of occurrence) for SPHA-NORTH and SPHA-SOUTH (U.S.) modeled using Maxent 3.4.1, using the map extent of the state of

California. Maxent models were run individually for both clades. Suitability surfaces were added together and rescaled to a maximum suitability value of 1 to form the combined suitability surface. Black areas show areas of zero suitability. Beige circles are presence points used in the Maxent models, compiled with a combination of this study's sampling locations and GBIF localities. Blue (SPHA-NORTH) and red (SPHA-SOUTH) points are the current study's genetic sample locations. Note the band of extremely low suitability separating SPHA-NORTH and SPHA-SOUTH, representing the Transverse Ranges and Tehachapi Mountains

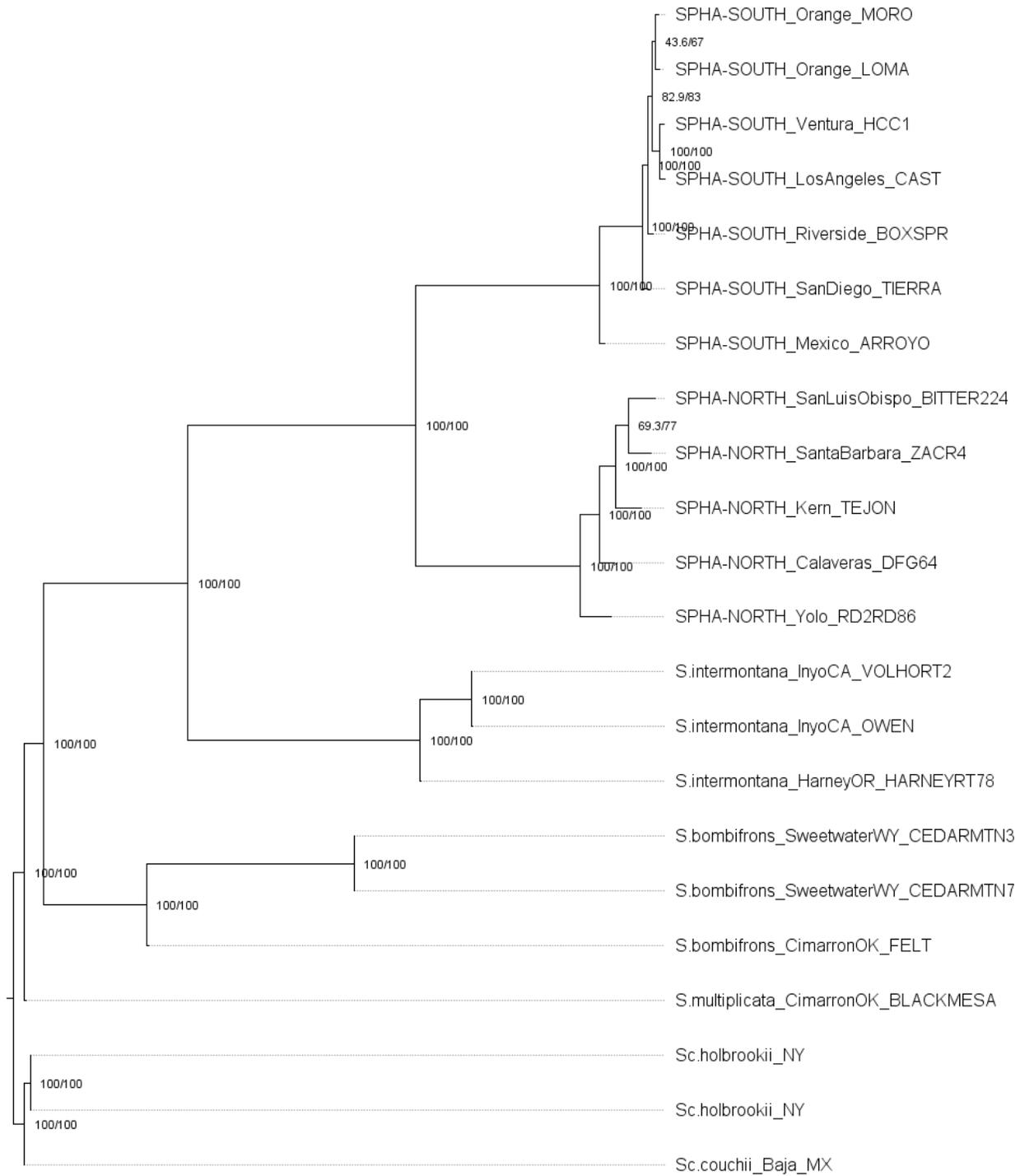


Fig. 3.2 IQ-TREE maximum likelihood phylogeny of 19 *Spea* individuals, rooted using 3 *Scaphiopus* individuals, with up to 50% missing data per locus, one SNP per locus, and using optimal model fit (TVMe+ASC+R2) determined by internal ModelFinder using Bayesian Information Criterion. Values at nodes are percent bootstrap support and percent SH-aLRT support, respectively

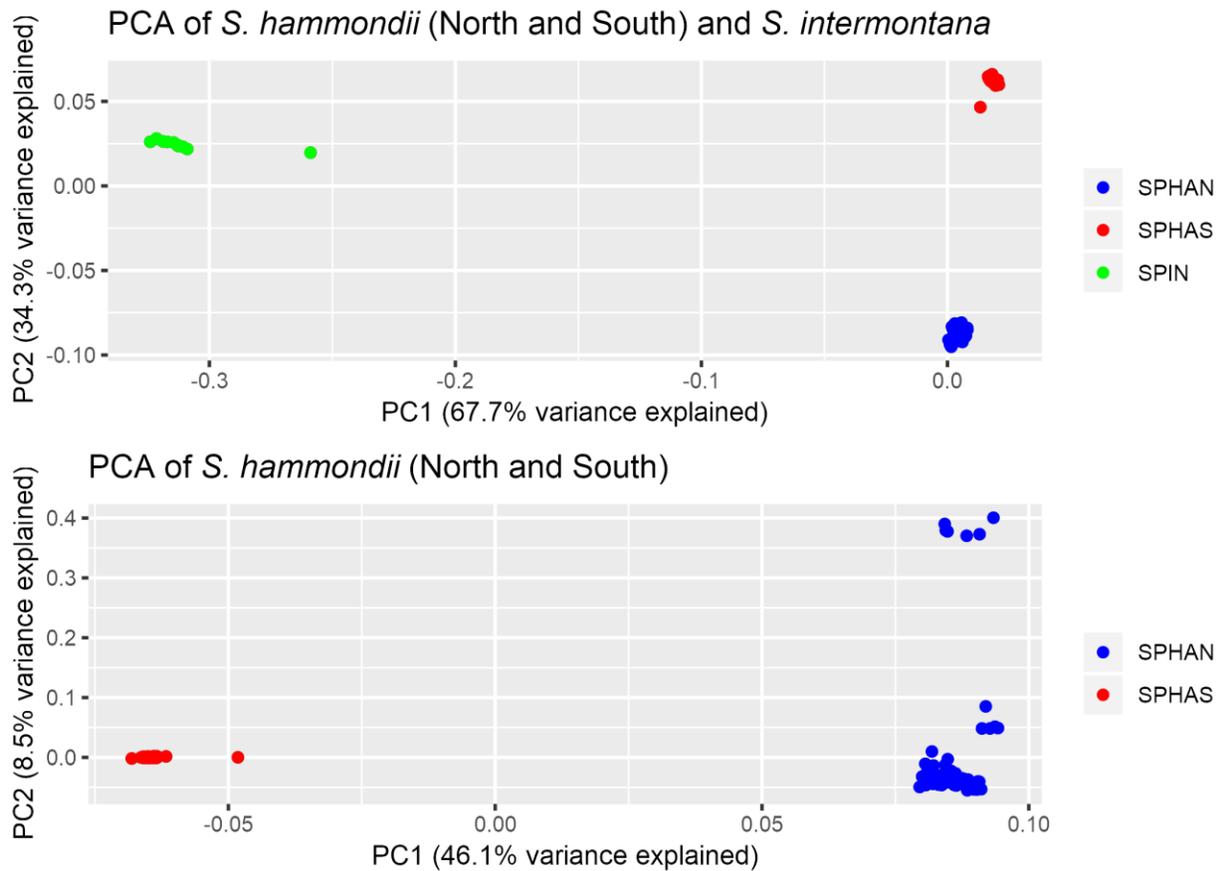


Fig. 3.3 Principle components analysis. A) *S. intermontana* ("SPIN", green dots), SPHA-NORTH ("SPHAN", blue), and SPHA-SOUTH ("SPHAS", red). The individual separated from the rest of the SPHA-SOUTH cluster is the sample from Mexico. B) SPHA-NORTH and SPHA-SOUTH only. PC2 splits SPHA-NORTH into two clusters north and south of the Sacramento-San Joaquin River Delta phylogeographic barrier

Fig. 3.4 IQ-TREE maximum likelihood phylogeny of 188 *S. hammondii* (SPHA-NORTH + SPHA-SOUTH) and *S. intermontana* individuals, with up to 50% missing data per locus, one SNP per locus, and using optimal model fit (GTR+F+ASC+R2) determined by internal ModelFinder using Bayesian Information Criterion. Values at nodes are percent bootstrap support and percent SH-aLRT support, respectively. Individuals with "SPHAN", "SPHAS", or "SPIN" in sample name refer to membership in SPHA-NORTH, SPHA-SOUTH, or *S. intermontana*, respectively

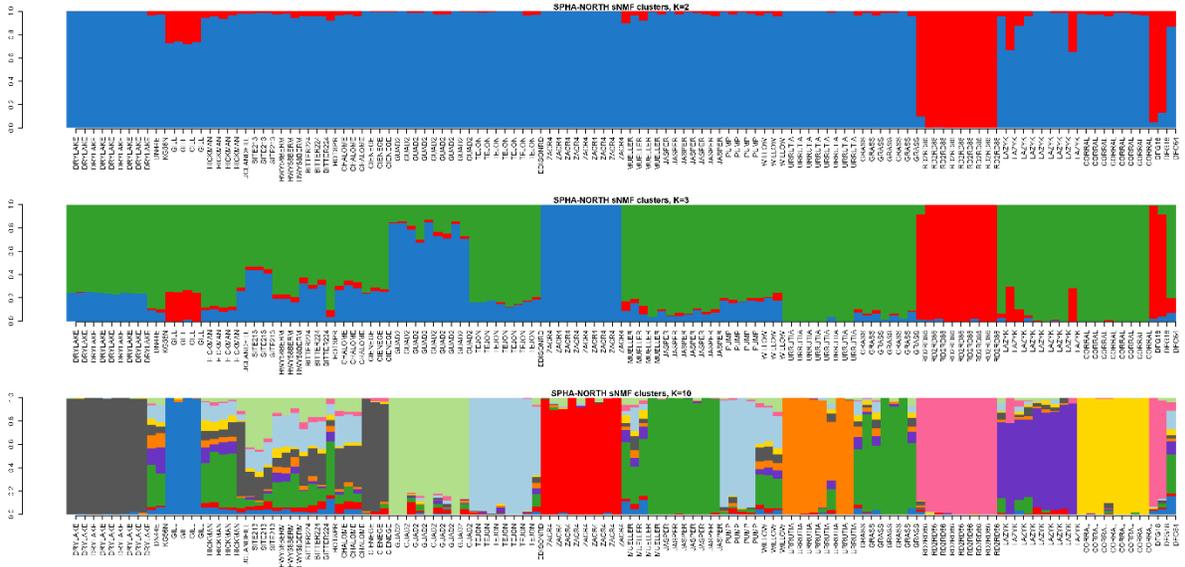


Fig. 3.5 sNMF ancestry coefficient barcharts for SPHA-NORTH, K= 2, 3, and 10

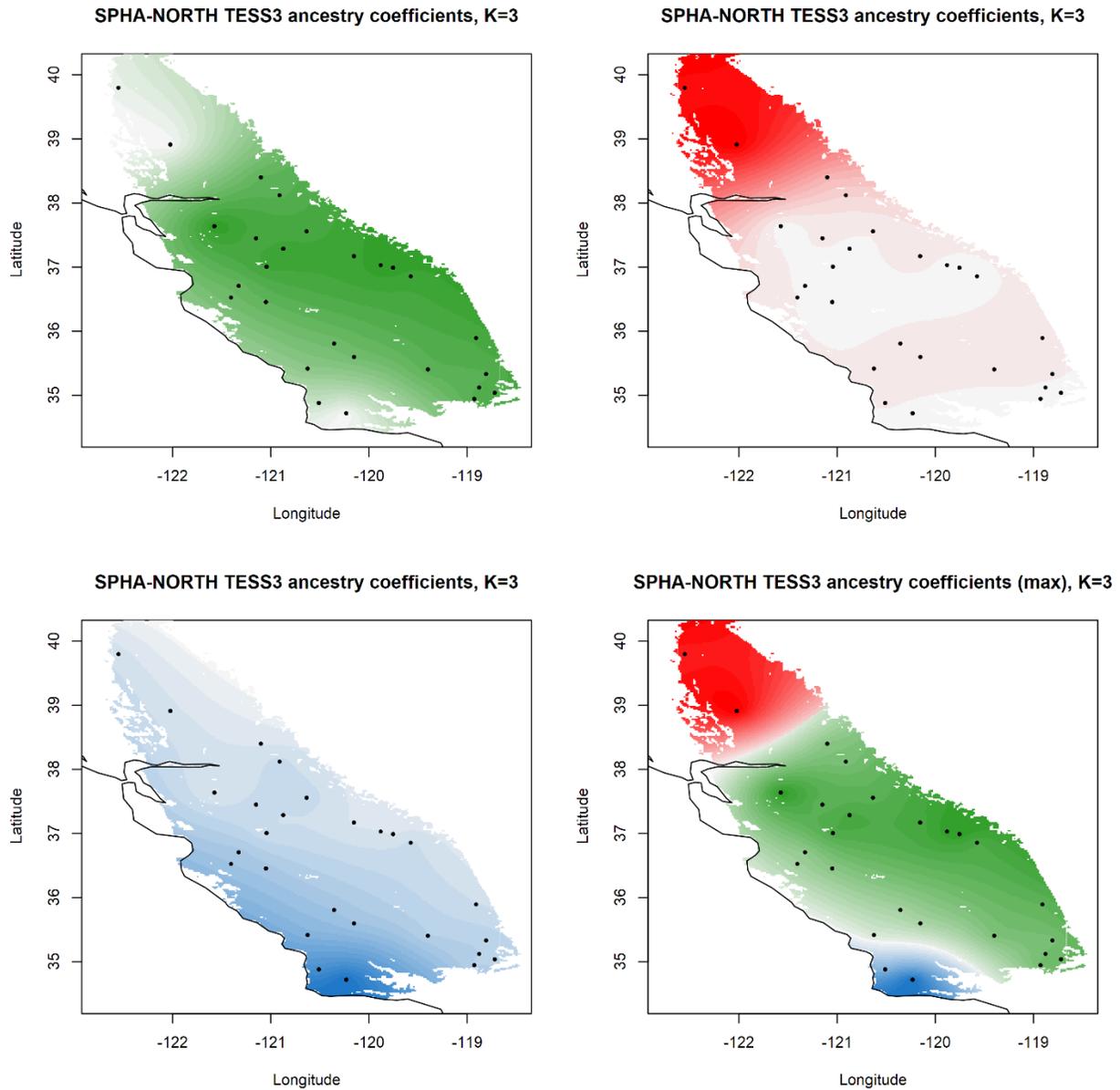


Fig. 3.6 TESS3 maps of interpolated cluster membership for K=3 clusters of SPHA-NORTH. Blue, red, and green indicate estimates of cluster membership for that particular cluster found in a given location. More saturated color implies greater proportion of ancestry of that cluster in individuals at that location. Lower right map is a union of the ancestry maps for each cluster, indicating areas where each cluster is dominant. Black dots are genetic sample sites

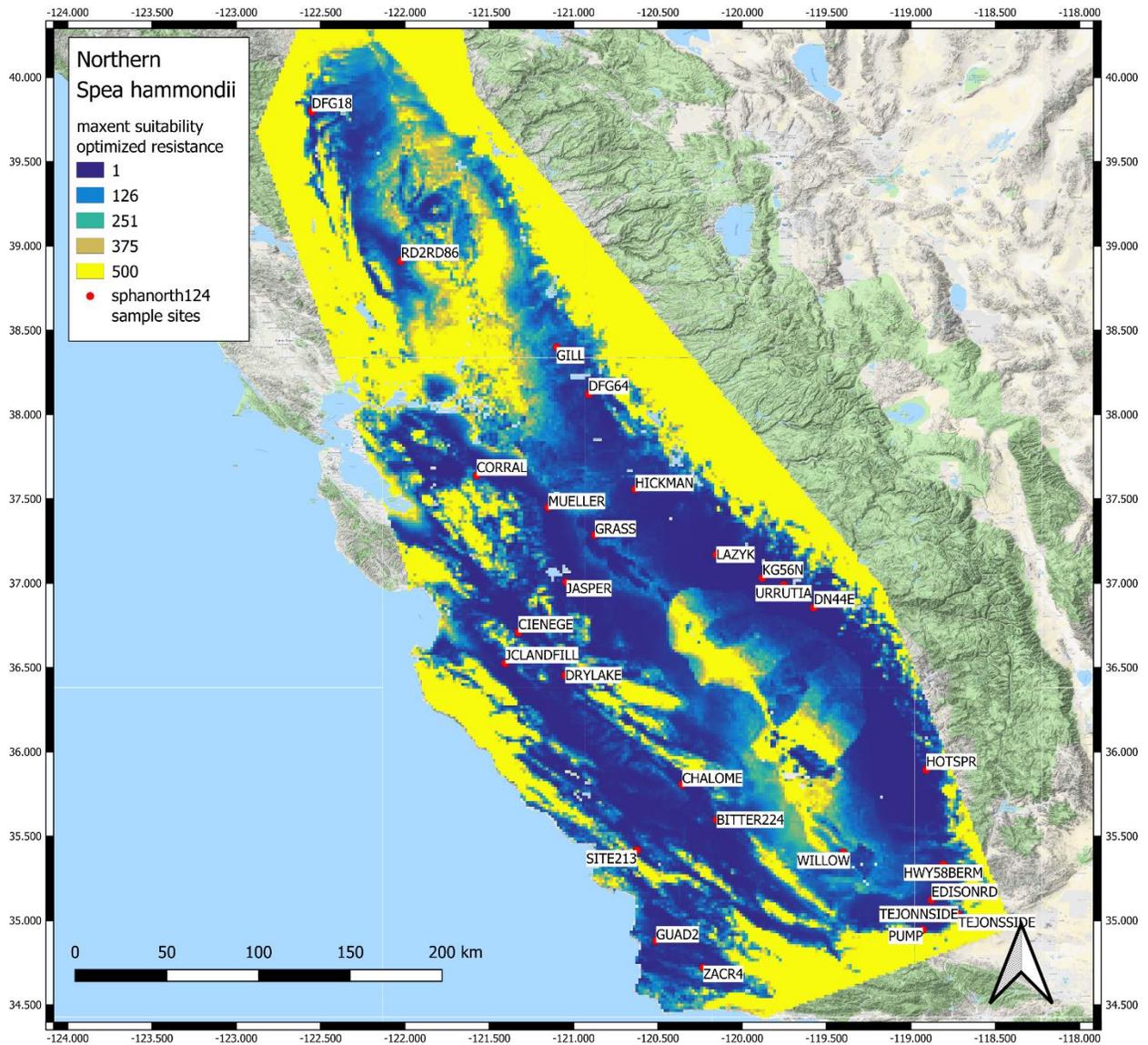


Fig. 3.7 Best ResistanceGA-optimized landscape resistance surface (maxent suitability) for northern SPHA. Yellow is higher resistance; blue is lower. Genetic sample sites are labeled on map

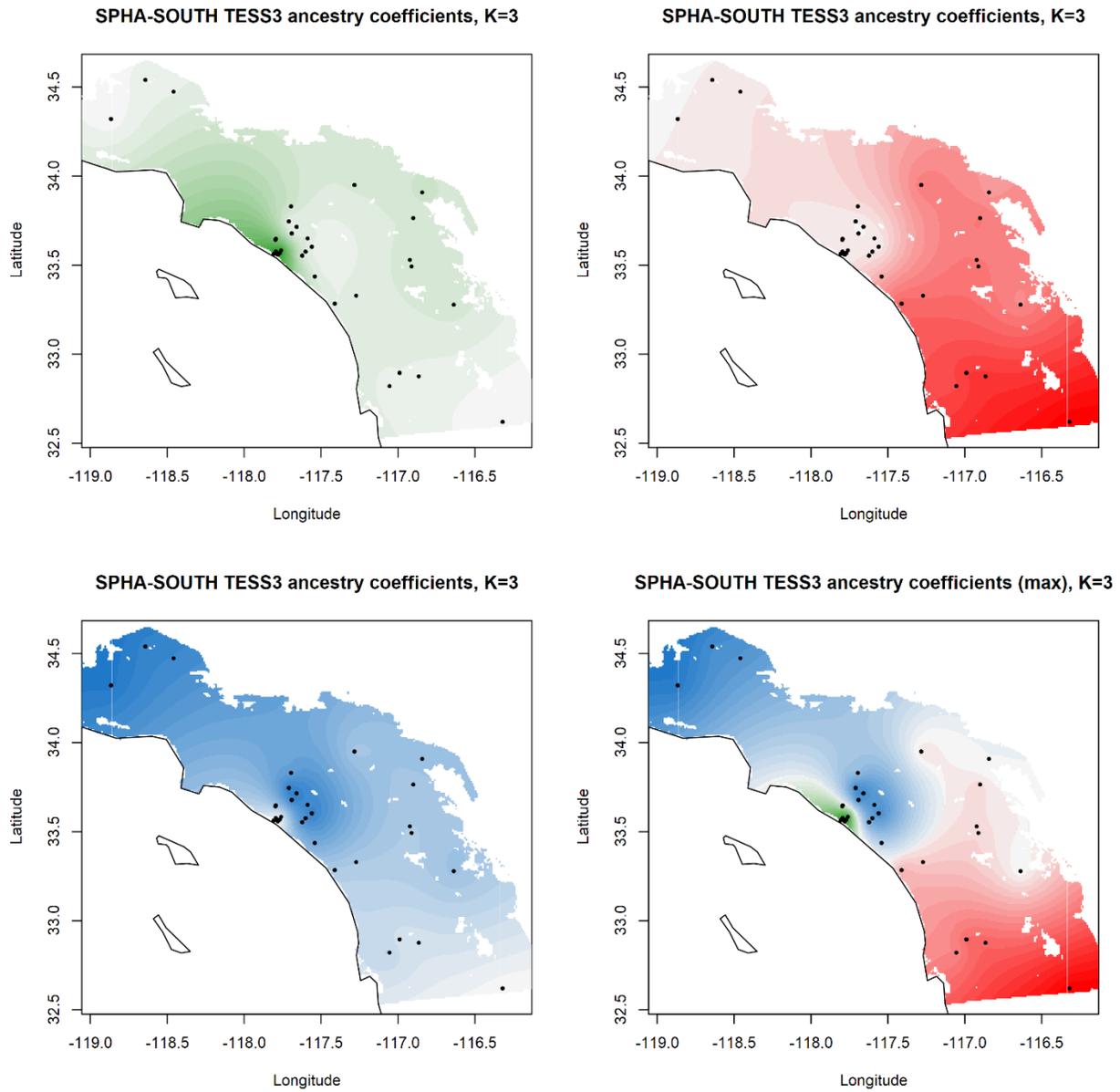


Fig. 3.9 TESS3 maps of interpolated cluster membership for K=3 clusters of SPHA-SOUTH. Blue, red, and green indicate estimates of cluster membership for that particular cluster found in a given location. More saturated color implies greater proportion of ancestry of that cluster in individuals at that location. Lower right map is a union of the ancestry maps for each cluster, indicating areas where each cluster is dominant. Black dots are genetic sample sites

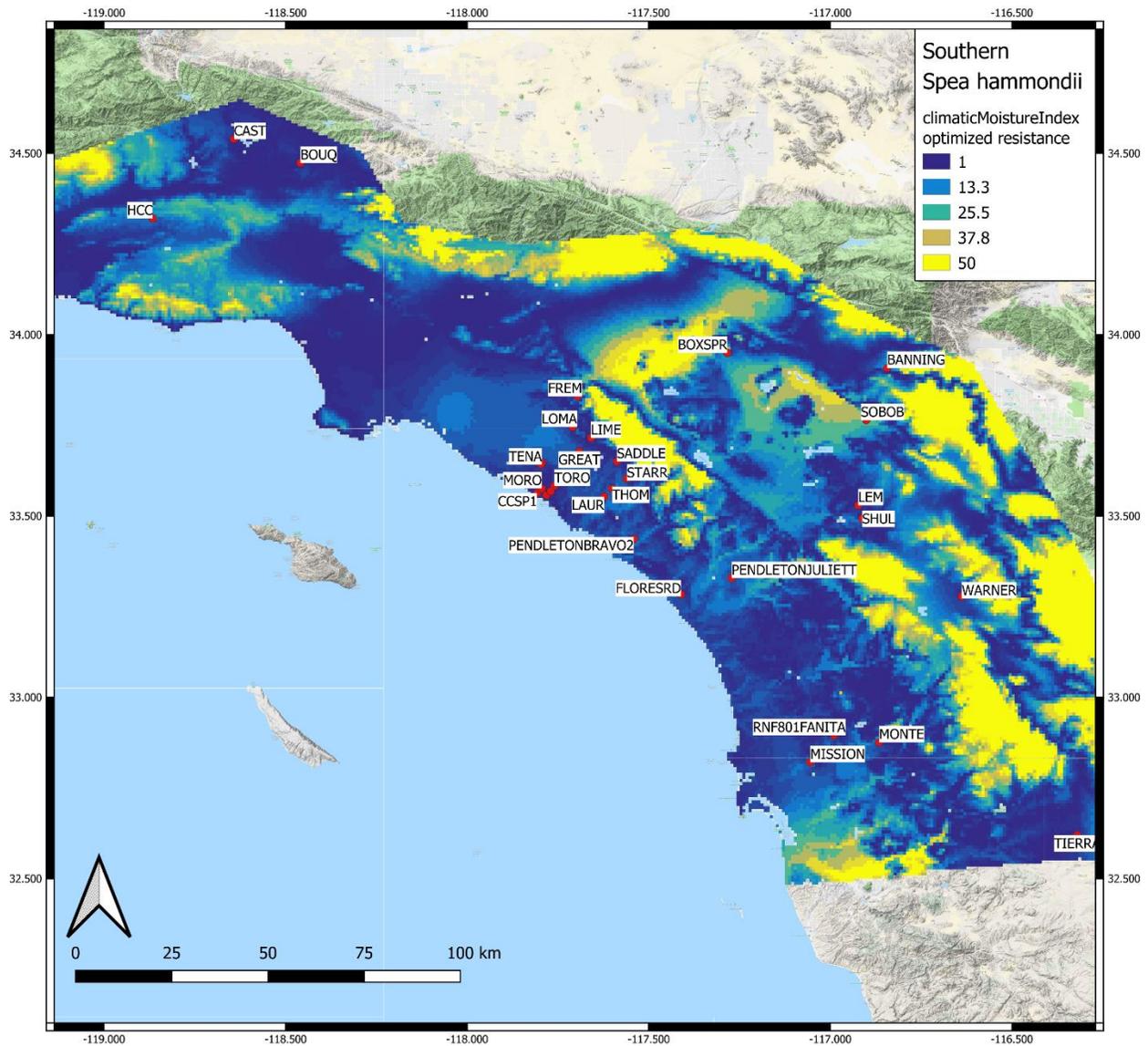


Fig. 3.10 Best ResistanceGA-optimized landscape resistance surface (climatic moisture index) for southern SPHA. Yellow is higher resistance; blue is lower. Genetic sample sites are labeled on map

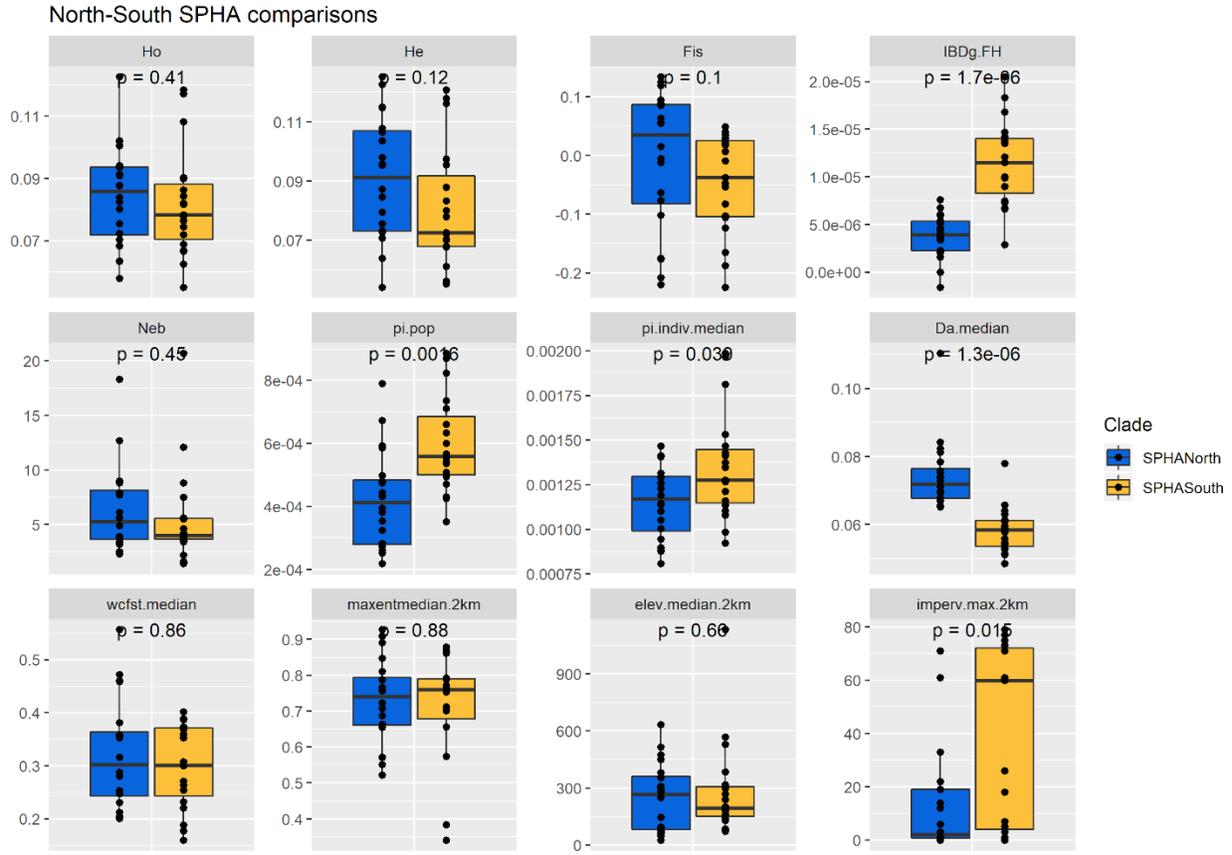


Fig. 3.11 Boxplots of several population genetic parameters for the North and South clades of *S. hammondii*. P-value on each plot was calculated using a Wilcoxon rank sum test. H_o : observed heterozygosity; H_e : expected heterozygosity; F_{IS} : inbreeding coefficient; IBDg.FH: individual-level estimate of inbreeding based on expected homozygosity; N_{eb} : effective number of breeders; pi.pop: pond-level nucleotide diversity; pi.indiv.median: median individual nucleotide diversity within each pond; D_a .median: median Nei's D_a for each pond; wcfst.median: median pairwise Weir-Cockerham F_{ST} for each pond; maxentmedian.2km, elev.median.2km: median Maxent suitability and elevation (respectively) within a 2 km buffer of each pond; imperv.max.2km: maximum % impervious surface within a 2 km buffer of each pond

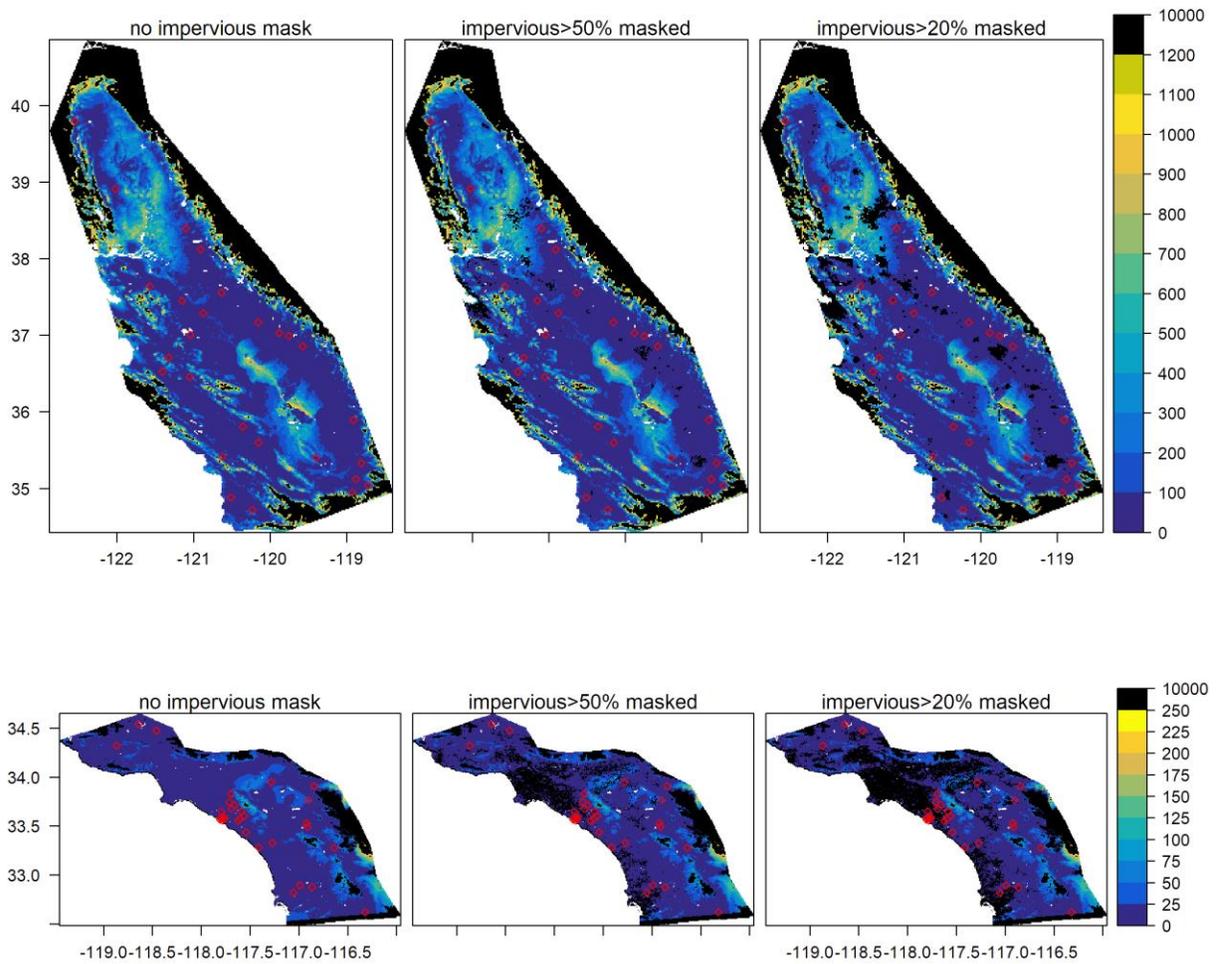


Fig. 3.12 Maps of the best model of landscape resistance for each clade, with impervious surface masks applied. Black areas in each surface have been modeled as extremely high resistance, reflecting levels of urbanization. Black areas on the perimeters have also been masked due to extremely low Maxent habitat suitability (<0.1) to prevent unlikely alternative corridors when impervious surface masks are applied. Top three panels are the North clade of *S. hammondii*. Bottom three are the South clade

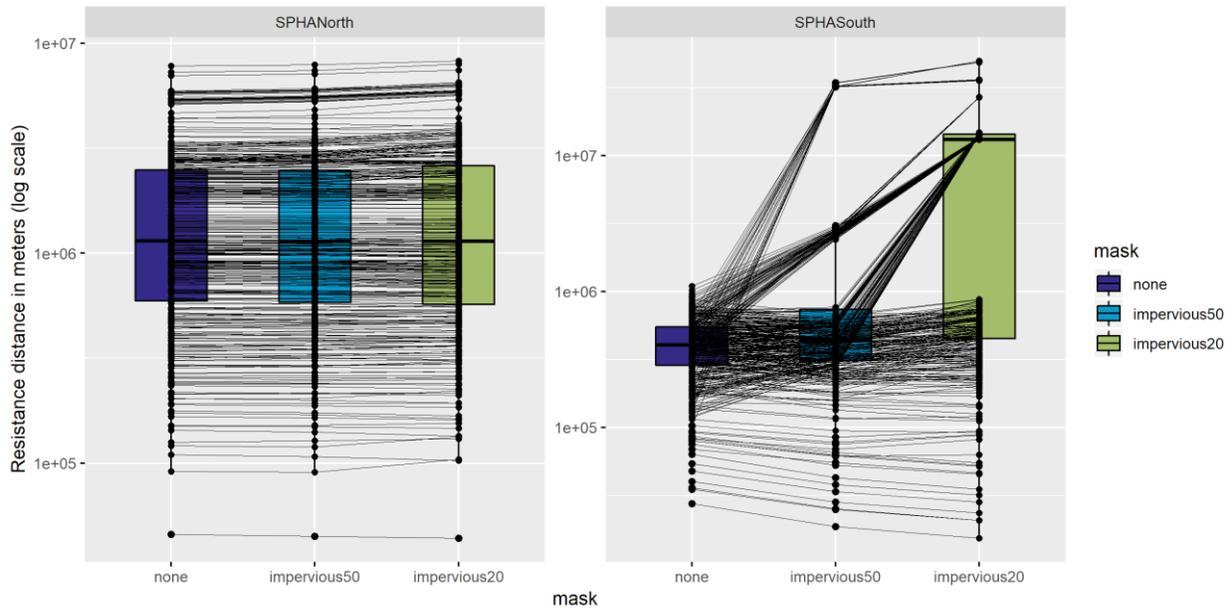


Fig. 3.13 Paired boxplots showing change in pairwise resistance distances with different % impervious surface masks applied. "impervious50" surfaces were masked such that pixels in the resistance surface with impervious surface greater than 50% were assigned a resistance value of 10,000; in "impervious20," this threshold was 20%. No impervious surface mask was applied in the "none" surface. Each point in a boxplot represents the resistance distance between two localities, with lines connecting the same locality pair across boxplots. Y-axis is in logarithmic scale

Supplemental Figures and Tables

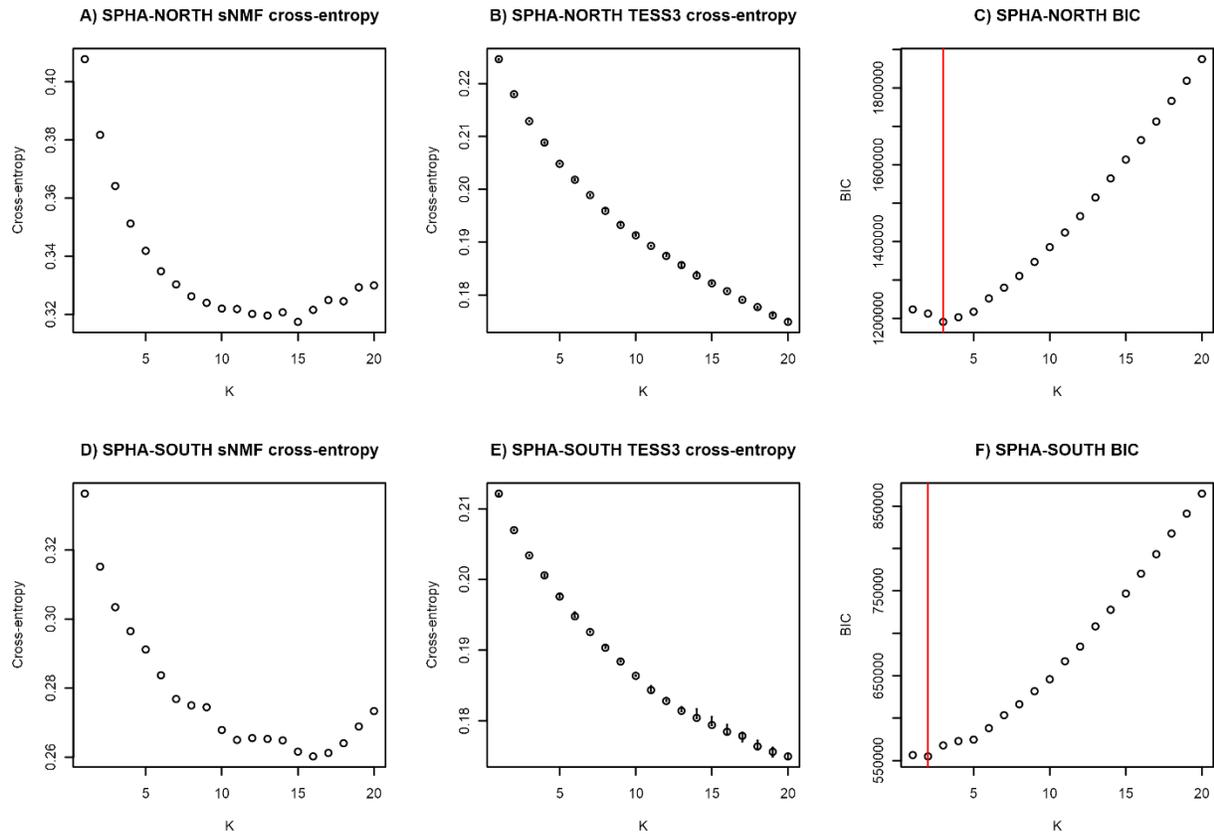


Fig. 3.S1 Choice of optimal number of clusters (K). Cross-entropy scores for SPHA-NORTH in A) aspatial sNMF clustering, B) spatial TESS3 clustering, C) snapclust BIC. Cross-entropy scores for SPHA-SOUTH in D) aspatial sNMF clustering, E) spatial TESS3 clustering, F) snapclust BIC.

Table 3.S1 Population coordinates and summary statistics (further summarized in Fig. 3.11)

clade	pop	Longitude	Latitude	n	Ho	He	Fis	IBDg.FH	Neb
SPHA-NORTH	BITTER224	-120.15213	35.597775	3	0.07033	0.08723	0.0856	2.13E-06	NA
SPHA-NORTH	CHALOME	-120.355393	35.807266	3	0.07545	0.07321	-0.10182	2.27E-06	NA
SPHA-NORTH	CIENEGE	-121.32698	36.707153	3	0.06344	0.05416	-0.20786	-1.07E-08	9
SPHA-NORTH	CORRAL	-121.574083	37.63837	8	0.0826	0.07078	-0.1751	2.27E-06	7.9
SPHA-NORTH	DRYLAKE	-121.049458	36.454872	9	0.08015	0.09585	0.11908	4.07E-06	4.9
SPHA-NORTH	GILL	-121.10063	38.40042	4	0.10212	0.08463	-0.22002	3.37E-06	12.7
SPHA-NORTH	GRASS	-120.873109	37.286801	7	0.10051	0.11481	0.06351	6.76E-06	3.7
SPHA-NORTH	GUAD2	-120.510367	34.880424	9	0.06833	0.07562	0.05523	3.52E-06	3.9
SPHA-NORTH	HICKMAN	-120.635	37.558	4	0.0941	0.11501	0.08883	6.05E-06	8.8
SPHA-NORTH	HWY58BERM	-118.807	35.332	3	0.08394	0.10362	0.08585	4.62E-06	NA
SPHA-NORTH	JASPER	-121.043079	37.006403	7	0.08767	0.10787	0.12534	5.33E-06	7.7
SPHA-NORTH	LAZYK	-120.153702	37.170502	9	0.12267	0.12521	-0.00508	7.60E-06	2.3
SPHA-NORTH	MUELLER	-121.15	37.45	4	0.09349	0.12262	0.13414	5.93E-06	3.4
SPHA-NORTH	PUMP	-118.927778	34.945394	4	0.09131	0.07955	-0.17666	3.73E-06	18.3

SPHA-NORTH	RD2RD86	-122.022862	38.911533	9	0.05787	0.06396	0.05636	-1.59E-06	3.7
SPHA-NORTH	SITE213	-120.624853	35.416697	3	0.07238	0.07281	-0.06354	2.05E-06	6.1
SPHA-NORTH	TEJON	-118.719436	35.038524	7	0.09086	0.10658	0.09477	5.10E-06	5.6
SPHA-NORTH	URRUTIA	-119.754186	36.99044	8	0.10185	0.09539	-0.07649	5.35E-06	2.5
SPHA-NORTH	WILLOW	-119.39975	35.405107	3	0.08784	0.09803	0.01543	4.38E-06	NA
SPHA-NORTH	ZACR4	-120.23106	34.71991	9	0.06335	0.06393	-0.01206	1.59E-06	3.2
SPHA-SOUTH	BBEND	-117.769774	33.568042	9	0.06246	0.05622	-0.10575	6.72E-06	5.5
SPHA-SOUTH	BOXSPR	-117.283333	33.950001	9	0.09027	0.09546	0.02212	1.35E-05	4.2
SPHA-SOUTH	CAST	-118.641941	34.539786	4	0.05502	0.05532	-0.05312	2.89E-06	12.1
SPHA-SOUTH	CCSP1	-117.807412	33.563498	10	0.07195	0.06126	-0.12326	9.00E-06	1.4
SPHA-SOUTH	FLORESRD	-117.411	33.284	6	0.10824	0.11791	0.02867	1.83E-05	3.9
SPHA-SOUTH	GREAT	-117.691	33.67822	10	0.08635	0.08004	-0.08247	1.38E-05	4
SPHA-SOUTH	HCC	-118.865715	34.320768	7	0.06881	0.06776	-0.00947	6.65E-06	4
SPHA-SOUTH	LAGUN	-117.78134	33.558171	10	0.06656	0.07256	0.04879	9.95E-06	3.4
SPHA-SOUTH	LAUR	-117.622406	33.553285	9	0.08999	0.09751	0.03986	1.68E-05	5.6

SPHA-SOUTH	LIME	-117.659055	33.715315	9	0.07831	0.07046	-0.10283	9.88E-06	2.2
SPHA-SOUTH	LOMA	-117.709481	33.745564	10	0.07647	0.07228	-0.04761	1.15E-05	1.6
SPHA-SOUTH	MORO	-117.794304	33.576053	10	0.07441	0.07801	0.01797	1.21E-05	3.6
SPHA-SOUTH	PENDLETON BRAVO2	-117.541	33.4364	5	0.08212	0.07009	-0.16542	7.49E-06	20.7
SPHA-SOUTH	RNF801 FANITA	-116.98962	32.89526	9	0.11728	0.12085	0.00679	2.05E-05	3.8
SPHA-SOUTH	SANCAN	-117.796639	33.64281	10	0.0778	0.08337	0.02965	1.35E-05	3.9
SPHA-SOUTH	STARR	-117.559973	33.603068	10	0.08427	0.06804	-0.2247	1.00E-05	7.5
SPHA-SOUTH	TENA	-117.794122	33.647421	10	0.06682	0.05604	-0.18779	7.29E-06	8.8
SPHA-SOUTH	THOM	-117.600762	33.575703	9	0.08173	0.08791	0.03537	1.47E-05	3.7
SPHA-SOUTH	TIERRA	-116.32	32.62	7	0.11853	0.11619	-0.03738	1.42E-05	4.6

Table 3.S1 cont.

clade	pop	Population	Median	Da	FST	Maxent	Elevation	Impervious
--------------	------------	-------------------	---------------	-----------	------------	---------------	------------------	-------------------

		pi	individual pi	median	median	Median	Median	Surface Max
						2km	2km	2km
SPHA-NORTH	BITTER224	0.00059359	0.000898	0.075753	0.28053	0.70816	634	6
SPHA-NORTH	CHALOME	0.00025171	0.0010507	0.074095	0.35624	0.9272	381	19
SPHA-NORTH	CIENEGE	0.00032413	0.000882	0.082283	0.47228	0.75852	296	3
SPHA-NORTH	CORRAL	0.00049831	0.0011374	0.08426	0.46037	0.5711	354	0
SPHA-NORTH	DRYLAKE	0.00047518	0.001102	0.065125	0.2878	0.66358	514.5	0
SPHA-NORTH	GILL	0.00027349	0.0014086	0.081202	0.38111	0.52243	65	1
SPHA-NORTH	GRASS	0.0003547	0.0014119	0.065486	0.21242	0.78665	26	0
SPHA-NORTH	GUAD2	0.00028298	0.0009458	0.069762	0.35392	0.68726	72	12
SPHA-NORTH	HICKMAN	0.00044403	0.0013152	0.06702	0.2007	0.81013	83	19
SPHA-NORTH	HWY58BERM	0.00027323	0.0011458	0.069264	0.2303	0.70851	251	14
SPHA-NORTH	JASPER	0.00039484	0.0011905	0.071038	0.25293	0.84786	282.5	1
SPHA-NORTH	LAZYK	0.00043399	0.0014672	0.067014	0.21188	0.90892	96	1
SPHA-NORTH	MUELLER	0.00047921	0.0012906	0.06781	0.20451	0.78765	46.5	61

SPHA-NORTH	PUMP	0.0005852	0.0012576	0.075179	0.35799	0.76684	449	33
SPHA-NORTH	RD2RD86	0.00067305	0.0008069	0.11041	0.55668	0.57186	58	0
SPHA-NORTH	SITE213	0.00038142	0.0010044	0.073675	0.35236	0.72341	309.5	22
SPHA-NORTH	TEJON	0.00079054	0.0012607	0.066877	0.24705	0.75705	475	1
SPHA-NORTH	URRUTIA	0.00043203	0.0014069	0.072807	0.31604	0.89106	146	1
SPHA-NORTH	WILLOW	0.00021822	0.0012281	0.069666	0.25338	0.65471	92	71
SPHA-NORTH	ZACR4	0.00026032	0.0008777	0.078434	0.45866	0.55126	280	0
SPHA-SOUTH	BBEND	0.00043046	0.0009829	0.061077	0.38691	0.76611	194	60
SPHA-SOUTH	BOXSPR	0.00071067	0.0015316	0.058444	0.25378	0.7019	530.5	75
SPHA-SOUTH	CAST	0.00035227	0.0009225	0.065722	0.40162	0.57383	567.5	1
SPHA-SOUTH	CCSP1	0.00047098	0.0011612	0.057724	0.35906	0.8617	140.5	60
SPHA-SOUTH	FLORESRD	0.00082333	0.001812	0.055827	0.15969	0.87876	83	7
SPHA-SOUTH	GREAT	0.0005675	0.0014267	0.052928	0.27075	0.86466	191	77
SPHA-SOUTH	HCC	0.00049401	0.0011339	0.062943	0.36893	0.38359	317	5
SPHA-SOUTH	LAGUN	0.00056017	0.0011037	0.054596	0.30034	0.79247	161	60

SPHA-SOUTH	Laur	0.00073529	0.0014661	0.048437	0.17745	0.77229	178	71
SPHA-SOUTH	LIME	0.00054281	0.0012753	0.058776	0.35294	0.65606	385.5	26
SPHA-SOUTH	LOMA	0.00055539	0.0012668	0.054091	0.30715	0.71282	299	18
SPHA-SOUTH	MORO	0.00060019	0.001213	0.052726	0.2632	0.75375	191	0
SPHA-SOUTH	PENDLETON BRAVO2	0.00050774	0.0013469	0.064071	0.37294	0.78883	129.5	3
SPHA-SOUTH	RNF801 FANITA	0.00088643	0.0019662	0.061223	0.18775	0.86969	239.5	0
SPHA-SOUTH	SANCAN	0.00063412	0.0012701	0.051303	0.23177	0.65667	85	73
SPHA-SOUTH	STARR	0.00053628	0.0014134	0.059474	0.37315	0.7595	269	61
SPHA-SOUTH	TENA	0.00042597	0.0010797	0.060805	0.38844	0.71012	72	73
SPHA-SOUTH	THOM	0.0006602	0.0013744	0.051149	0.22013	0.76521	200.5	79
SPHA-SOUTH	TIERRA	0.00086989	0.0019874	0.077923	0.30072	0.34075	1133	1

Table 3.S2 ResistanceGA model results for northern SPHA

Surface	k	AIC	AICc	R2m	R2c	LL
maxentcloglog	4	-2810	-2808	0.8717	0.9766	1408.8
climaticMoistureIndex	4	-2763	-2761	0.7712	0.993	1385.5
PC1ofrasters	4	-2635	-2633	0.7234	0.9758	1321.5
PC4ofrasters	4	-2565	-2563	0.7169	0.9353	1286.6
aridityIndexThornthwaite	4	-2535	-2533	0.7092	0.9513	1271.4
PC3ofrasters	4	-2795	-2794	0.6885	0.9961	1401.7
PC5ofrasters	4	-2531	-2529	0.6803	0.9752	1269.5
continentality	4	-2610	-2609	0.6793	0.9697	1309.2
PC6ofrasters	4	-2557	-2556	0.6588	0.9137	1282.7
PC2ofrasters	4	-2493	-2491	0.57	0.7743	1250.6
terrainruggedness	4	-2520	-2518	0.5687	0.8996	1264.1
topoWet	4	-2520	-2519	0.5484	0.9121	1264.2
landcover	16	-2553	-2475	0.533	0.8516	1280.6
Distance	2	-2473	-2476	0.2684	0.676	1240.3
Null	1	-2356	-2360	0	0.6097	1181.1

Table 3.S3 ResistanceGA model results for southern SPHA

Surface	k	AIC	AICc	R2m	R2c	LL
topoWet	4	-3584	-3583	0.7978	0.9373	1796.1
continentality	4	-3535	-3534	0.7628	0.932	1771.6
terrainruggedness	4	-3571	-3569	0.7591	0.9349	1789.5
PC2ofrasters	4	-3523	-3522	0.7325	0.9287	1765.6
climaticMoistureIndex	4	-3593	-3591	0.7323	0.9488	1800.4
PC5ofrasters	4	-3535	-3533	0.7046	0.9182	1771.4
landcover	16	-3544	-3478	0.6856	0.9294	1776
PC1ofrasters	4	-3529	-3528	0.6824	0.9251	1768.7
taxousda	13	-3536	-3495	0.671	0.9257	1772
Distance	2	-3520	-3523	0.6688	0.9226	1763.8
PC4ofrasters	4	-3578	-3577	0.6647	0.9627	1793.1
PC3ofrasters	4	-3520	-3518	0.6415	0.9182	1763.9
maxentcloglog	4	-3522	-3520	0.6301	0.9197	1764.8
Null	1	-2962	-2966	0	0.5549	1483.8

References

- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545. doi: 10.1111/ecog.01132
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. doi: 10.1038/nrg.2015.28
- Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *BioRxiv*, 205799. doi: 10.1101/205799
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. doi: 10.1016/j.tree.2006.11.004
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring Continuous and Discrete Population Genetic Structure Across Space. *Genetics*, *210*(1), 33–52. doi: 10.1534/genetics.118.301333
- Bryson, R. W., Savary, W. E., Zellmer, A. J., Bury, R. B., & McCormack, J. E. (2016). Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. *Molecular Ecology*, *25*(15), 3731–3751. doi: 10.1111/mec.13707
- California Department of Finance. (2019, May 1). Demographic Estimates. Retrieved June 7, 2019, from <http://www.dof.ca.gov/Forecasting/Demographics/Estimates/>
- Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. doi: 10.1111/1755-0998.12669
- Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. doi: 10.1111/1755-0998.12471
- Caye, K., Jay, F., Michel, O., & François, O. (2018). Fast inference of individual admixture coefficients using geographic data. *The Annals of Applied Statistics*, *12*(1), 586–608. doi: 10.1214/17-AOAS1106
- Chen, J., Bever, G. S., Yi, H.-Y., & Norell, M. A. (2016). A burrowing frog from the late Paleocene of Mongolia uncovers a deep history of spadefoot toads (Pelobatoidea) in East Asia. *Scientific Reports*, *6*, 19209. doi: 10.1038/srep19209

- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. doi: 10.1093/bioinformatics/btr330
- Davidson, C., Shaffer, H. B., & Jennings, M. R. (2002). Spatial Tests of the Pesticide Drift, Habitat Destruction, UV-B, and Climate-Change Hypotheses for California Amphibian Declines. *Conservation Biology*, 16(6), 1588–1601. doi: 10.1046/j.1523-1739.2002.01030.x
- Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. doi: 10.1093/bioinformatics/btu121
- Eaton, D. A. R. (2019). *Interactive assembly and analysis of RAD-seq data sets: dereneaton/ipyrad* [Jupyter Notebook]. Retrieved from <https://github.com/dereneaton/ipyrad>
- Etten, J. van. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, 76(1), 1–21. doi: 10.18637/jss.v076.i13
- Etten, J. van. (2018). gdistance: Distances and Routes on Geographical Grids (Version 1.2-2). Retrieved from <https://CRAN.R-project.org/package=gdistance>
- Fisher, R. N., & Shaffer, H. B. (1996). The decline of amphibians in California's Great Central Valley. *Conservation Biology*, 10(5), 1387–1397. doi: 10.1046/j.1523-1739.1996.10051387.x
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. doi: 10.1111/2041-210X.12382
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, 196(4), 973–983. doi: 10.1534/genetics.113.160572
- García-París, M., Buchholz, D. R., & Parra-Olea, G. (2003). Phylogenetic relationships of Pelobatoidea re-examined using mtDNA. *Molecular Phylogenetics and Evolution*, 28(1), 12–23. doi: 10.1016/S1055-7903(03)00036-8
- GBIF.org. (2017). *Spea hammondi* (Baird, 1859) GBIF Occurrence Download. doi: <http://doi.org/10.15468/dl.uggrh3>
- Gosselin, T. (2019). *RADseq Data Exploration, Manipulation and Visualization using R: thierrygosselin/radiator* [R]. Retrieved from <https://github.com/thierrygosselin/radiator>
- Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics (Version 0.04-22). Retrieved from <https://CRAN.R-project.org/package=hierfstat>

- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. doi: 10.1093/sysbio/syq010
- Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. doi: 10.1371/journal.pone.0169748
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., ... Megown, K. (2015, May). Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information [Text]. Retrieved June 8, 2019, from <https://www.ingentaconnect.com/content/asprs/pers/2015/00000081/00000005/art00002>
- Jennings, M. R., & Hayes, M. P. (1994). *Amphibian and reptile species of special concern in California*. USA: California Department of Fish and Game, Inland Fisheries Division, Rancho Cordova.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. doi: 10.1038/nmeth.4285
- Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity*, *115*(1), 63–72. doi: 10.1038/hdy.2015.17
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122. doi: 10.1038/sdata.2017.122
- Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics*, *189*(1), 237–249. doi: 10.1534/genetics.111.130922
- Leutner, B., Horning, N., Schwalb-Willmann, J., & Hijmans, R. J. (2019). RStoolbox: Tools for Remote Sensing Data Analysis (Version 0.2.4). Retrieved from <https://CRAN.R-project.org/package=RStoolbox>
- Lewis, P. O. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology*, *50*(6), 913–925. doi: 10.1080/106351501753462876

- Linck, E., Epperly, K., Van Els, P., Spellman, G. M., Bryson, R. W., McCormack, J. E., ... Klicka, J. (2019). Dense Geographic and Genomic Sampling Reveals Paraphyly and a Cryptic Lineage in a Classic Sibling Species Complex. *Systematic Biology*. doi: 10.1093/sysbio/syz027
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77. doi: 10.1111/1755-0998.12592
- McCartney-Melstad, E., Gidiş, M., & Shaffer, H. B. (2018). Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity*, 121(2), 112–125. doi: 10.1038/s41437-018-0097-7
- McCartney-Melstad, E., Vu, J. K., & Shaffer, H. B. (2018). Genomic data recover previously undetected fragmentation effects in an endangered amphibian. *Molecular Ecology*, 27(22), 4430–4443. doi: 10.1111/mec.14892
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. doi: 10.1111/j.1365-294X.2012.05578.x
- Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. doi: 10.1093/molbev/mst024
- Morey, S. R. (1998). *Pool duration influences age and body mass at metamorphosis in the western spadefoot toad: implications for vernal pool conservation*. Retrieved from ://ZOOREC:ZOOR13600083765
- Morey, S. R., & Guinn, D. A. (1992). Activity patterns, food habits, and changing abundance in a community of vernal pool amphibians. *Endangered and Sensitive Species of the San Joaquin Valley, California: Their Biology, Management, and Conservation. The California Energy Commission, Sacramento, California, and the Western Section of the Wildlife Society*, 149–158.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x
- Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondi*). *Conservation Genetics*, 19(4), 937–946. doi: 10.1007/s10592-018-1066-7
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. doi: 10.1093/molbev/msu300

- Ortiz, E. M. (2019). *Convert SNPs in VCF format to PHYLIP, NEXUS, binary NEXUS, or FASTA alignments for phylogenetic analysis: edgardomortiz/vcf2phylip* [Python]. Retrieved from <https://github.com/edgardomortiz/vcf2phylip>
- Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9(6), 1638–1647. doi: 10.1111/2041-210X.12984
- Peterman, W. E. (2019). *Optimize resistance surfaces using Genetic Algorithms: wpeterman/ResistanceGA* [R]. Retrieved from <https://github.com/wpeterman/ResistanceGA>
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. *Ecography*, 40(7), 887–893. doi: 10.1111/ecog.03049
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16(3), 608–627. doi: 10.1111/1755-0998.12512
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. doi: 10.1086/519795
- Pyron, R., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, 61(2), 543–583. doi: 10.1016/j.ympev.2011.06.012
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from <https://www.R-project.org/>
- Reyes-Velasco, J., Manthey, J. D., Bourgeois, Y., Freilich, X., & Boissinot, S. (2018). Revisiting the phylogeography, demography and taxonomy of the frog genus *Ptychadena* in the Ethiopian highlands with the use of genome-wide SNP data. *PLOS ONE*, 13(2), e0190440. doi: 10.1371/journal.pone.0190440
- Rissler, L. J., Hijmans, R. J., Graham, C. H., Moritz, C., & Wake, D. B. (2006). Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *The American Naturalist*, 167(5), 655–666.
- Sánchez-Ramírez, S., Rico, Y., Berry, K. H., Edwards, T., Karl, A. E., Henen, B. T., & Murphy, R. W. (2018). Landscape limits gene flow and drives population structure in Agassiz's desert tortoise (*Gopherus agassizii*). *Scientific Reports*, 8(1), 11231. doi: 10.1038/s41598-018-29395-6

- Schönrogge, K., Barr, B., Wardlaw, J. C., Napper, E., Gardner, M. G., Breen, J., ... Thomas, J. A. (2002). When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnean Society*, 75(3), 291–300. doi: 10.1046/j.1095-8312.2002.00019.x
- Shaffer, H. B., Pauly, G. B., Oliver, J. C., & Trenham, P. C. (2004). The molecular phylogenetics of endangerment: cryptic variation and historical phylogeography of the California tiger salamander, *Ambystoma californiense*. *Molecular Ecology*, 13(10), 3033–3049. doi: 10.1111/j.1365-294X.2004.02317.x
- Spinks, P. Q., Thomson, R. C., & Shaffer, H. B. (2014). The advantages of going large: genome-wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. *Molecular Ecology*, 23(9), 2228–2241. doi: 10.1111/mec.12736
- Thomson, R. C., Wright, A. N., & Shaffer, H. B. (2016). *California Amphibian and Reptile Species of Special Concern*. UC Press.
- Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, 41(2), 291–307. doi: 10.1111/ecog.02880
- U.S. Fish and Wildlife Service. (2015). Endangered and Threatened Wildlife and Plants; 90-Day Findings on 31 Petitions. *Federal Register*, 80(126), 37568–37579.
- Zeng, C., Gomez-Mestre, I., & Wiens, J. J. (2014). Evolution of Rapid Development in Spadefoot Toads Is Unrelated to Arid Environments. *PLOS ONE*, 9(5), e96637. doi: 10.1371/journal.pone.0096637
- Zhou, X., Shen, X.-X., Hittinger, C. T., & Rokas, A. (2018). Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Molecular Biology and Evolution*, 35(2), 486–503. doi: 10.1093/molbev/msx302