![USGS logo - science for a changing world]

Spatial structuring and temporal shifts in the genetic composition of Blainville's horned lizard (*Phrynosoma blainvillii*) populations in southern California

By Jonathan Q. Richmond and Robert N. Fisher

Data summary report prepared for the California Department of Fish and Wildlife. Please do not distribute or cite without permission from authors.
U.S. Department of the Interior
U.S. Geological Survey

Cover Photo by Jonathan Richmond, USGS: (adult female *Phrynosoma blainvilli* from Mission trails, San Diego, California)

Spatial structuring and temporal shifts in the genetic composition of Blainville's horned lizard (*Phrynosoma blainvillii*) populations in southern California

By Jonathan Q. Richmond[1] and Robert N. Fisher

U.S. Geological Survey, Western Ecological Research Center, 4165 Spruance Rd Suite 200, San Diego CA 92101

[1]correspondence: jrichmond@usgs.gov, 619-225-6434

INTRODUCTION

Sprawling urban development presents significant challenges to conservation biologists and wildlife managers given that remaining non-converted lands often persist as small, disconnected patches with no conveyance to larger core habitat areas. Patch size, degree of connectivity, and sensitivity to disturbed edges are a few of the many factors that determine whether remnant open space can still support wildlife (Lande 1987; Fahring 2003), and mitigating for these factors on a species-by-species basis is not feasible in areas with high biodiversity. Furthermore, human population growth guarantees that remaining natural lands will continue to be encroached upon, particularly in areas near large urban epicenters.

One approach to mitigating the 'development-conservation' dilemma is to identify ways of maintaining, retaining, or restoring habitat linkages among the largest remaining tracts of conserved lands as possible (Crooks and Sanjayan 2006; Hilty *et al.* 2006), without requiring a moratorium on development. Instead of targeted management for individual species, this approach aims at preserving the same preferred habitats across multiple species, with an emphasis on connectivity (Gonzalez *et al.* 1998; Haddad *et al.* 2003; Damschen et al. 2006; Gilbert-Norton *et al.* 2010). One area where this strategy is being implemented is in southern California, a region known for its high biodiversity and consumption by urban sprawl. Although biodiversity is high, numerous species are in decline, particularly in the southwestern corner of California in San Diego County, and many are now threatened or endangered at the federal and state levels (Dobson *et al.* 1997).

To reduce the impacts of habitat loss on the region's diversity, while at the same time recognizing that reality of economic and human population growth, the County of San Diego joined with private citizens, developers and various agencies (e.g. U.S. Fish and Wildlife Service and California Department of Fish and Wildlife) to approve the Multiple Species Conservation Plan (MSCP) in 1998. The MSCP operates within the 'conserve habitat first' paradigm, with an emphasis on linkage habitat, and streamlines the regulatory process associated with protecting state and federally listed species affected by development (which would otherwise need to be completed on a case by case basis). The intended outcome is to establish a large, inter-connected series of preserves that accommodates the habitat requirements of as many species as possible, while configuring future development in a manner that best conserves biodiversity (http://www.sandiegocounty.gov/pds/mscp/; accessed on 15 March 2016).

While the MSCP is intended to promote habitat and wildlife self-sustainability, changes in the amounts and configuration of remaining natural lands are to the extent that, when combined with high numbers of sensitive species, continuous monitoring of select species and habitat may be

the only way to ensure ecosystem longevity. Functional linkages between them must also continue to be acquired, maintained, or restored to ensure that this capacity is upheld. Fortunately, recent advances in DNA sequencing technology have dramatically improved the efficiency of biological monitoring and the accuracy of identifying corridors or barriers to gene movement (Allendorf and Luikart 2012; Funk *et al.* 2012). This is mainly due to the exponential increase in the number of genetic markers that these technologies provide, typically at a fraction of the cost and effort that molecular biologists faced as recently as 8–10 years ago (McCormack *et al.* 2013; Andrews *et al* 2016).

With powerful genetic monitoring tools but limited management resources, the question then becomes, which species should be prioritized for monitoring and assessment? One option is to focus on 'umbrella species', whose own ecological requirements encompass those of many other species occurring in the same area (Roberge and Angelstam 2004; Caro 2010). Conserving an umbrella species therefore engenders protection to other species with the same habitat preference (Andelman and Fagan 2000; Breckheimer *et al.* 2014). A second option might be to focus on so-called 'ecological indicator species' with high sensitivity to habitat disturbance, given their potential to serve as early warning signals for distressed environments (Landres *et al.* 1998; McGeoch 2007). As the most sensitive species are predicted to respond to the earliest forms of disturbance, monitoring such species could buy time in counteracting the disturbance before it substantially impacts the ecosystem. The most judicious approach of course would build towards monitoring some combination of umbrella and indicator species.

Enter Blainville's horned lizard Phrynosoma blainvillii

Blainville's horned lizard (*Phrynosoma blainvillii*) has numerous attributes that meet the criteria for an ecological indicator species (summarized in Caro 2010), earmarking it as a worthy target for monitoring. These attributes include a wide geographic distribution (Stebbins 2003; Brattstrom 2013; Thompson *et al.* 2016), high specificity to shrublands with loose soils (Stebbins 2003; Lemm 2006), ease of being located within these habitats and high probability of being sampled during monitoring (if present; Fisher *et al.* 2002), low mobility, small body size, small home ranges (Suarez *et al.* unpub. data), dietary specialization (Suarez *et al.* 2000), and predictable responses to disturbance (Suarez and Case 2002). It is also listed as a 'species of special concern' by the California Department of Fish and Wildlife due to its shrinking distribution, over-collecting, and pressure from non-native species (Jennings 1987, 1988; Thompson *et al.* 2016), and is a covered species under the San Diego MSCP. Of its life history attributes, dietary specialization on native harvester ants is a main driver of environmental sensitivity in *P. blainvillii* – where harvester ants decline or become extirpated, so do horned lizards (Fisher *et al.* 2002; Suarez and Case 2002).

Members of the ant genera *Pogonomyrmex*, *Veromessor*, and *Crematogaster* are the main components of the *P. blainvillii* diet and are critical seed dispersers for many of the region's native plant species. These native harvester ants have been displaced by the invasive Argentine Ant *Linepithema humile* in large parts of southern California and elsewhere (Holway 1995; Suarez *et al.* 1998, 2001). *Linepithema humile* cannot substitute for native ants as a food source for *P. blainvillii* and hatchling *P. blainvillii* are incapable of persisting on diets of arthropods that are typical of invaded communities (Suarez et al. 2000; Suarez and Case 2002). In fact, presence of *L. humile* is considered evidence of degraded habitat because they thrive in areas with increased non-native vegetation and

higher soil moisture content relative to the surrounding natural environment (Holway 1998; Suarez *et al.* 1998; Bolger *et al.* 2000).

In addition to being a high-value monitoring target, the broad range of *P. blainvillii* across southern California (including numerous preserve systems) suggests that their population genetics might be useful in guiding the acquisition or restoration of linkage habitat, given that patterns of genetic admixture among different localities can signal where gene interchanges are currently being made, or where they occurred prior to major habitat conversion. To this end, we conducted a population genomic study on *P. blainvillii* across a large portion of the range to (1) generate baseline data on population genetic structure; (2) identify where current and/or historical genetic linkages occurred in southern California to guide acquisition, maintenance, or restoration efforts for linkage habitat; (3) to test whether certain populations have experienced declines or shifts in genetic diversity based on temporal sampling at certain localities; and (4) to compare population structuring with other available datasets from geographically overlapping taxa. An ancillary objective was to generate a genomic dataset for designing sequence capture probes that will provide an easy, cost-effective assay for future genetic monitoring of this species.

MATERIALS AND METHODS
Data collection

We extracted DNA from tissue samples spanning 45 different localities in five counties across southern California (San Diego, Orange, Riverside, San Bernardino and Los Angeles counties; Table 1). Samples were acquired through USGS pitfall trapping surveys (Rochester *et al.* 2010; Fisher *et al.* 2002), field efforts from this study, and from contributions by research colleagues. The USGS surveys covered 15 of the 45 localities, four of which had temporal samples spanning more than three generations. The oldest of these samples date to 1995 and the most recent were from 2012, although the spacing between sampling was not consistent across all sites, and sufficient sample sizes for statistical comparison were not always available in a single year's worth of sampling. For cases involving the latter, we combined samples across successive years but never exceeded more than two years between sampling events for any given temporal cohort (Table 1).

Tissues consisted of toe or tail clips stored in 95% ethanol, and/or muscle or liver tissue from a small number of salvaged specimens (typically road mortalities). We extracted genomic DNA using a Qiagen® DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia CA) and generated multiple datasets consisting of two marker types: single nucleotide polymorphisms (i.e. SNPs) and mtDNA sequences. The SNP datasets were derived from restriction-site associated DNA sequences (i.e. one SNP per 100 bp read) and coded as alleles for frequency-based population genetic analyses. The second dataset was an alignment of the mitochondrial ND1 protein coding gene for a subset of the samples used for RAD sequencing. All 45 localities were represented in the mtDNA dataset, whereas 26 were represented in the SNP dataset. In nearly all cases, the same individuals were used for mtDNA and RAD sequencing (Fig. 1). There were fewer localities represented in the SNP dataset because not all sites had sufficient sampling for population-level analysis, and for many samples the DNA concentrations were sufficient for mtDNA sequencing but too low for RAD sequencing (which requires substantially higher DNA concentrations as starting material).

RAD genome library preparation followed the protocol of Peterson *et al.* (2012), using 20 units each of the *Sbf*I and *Msp*I enzymes to digest the genomic DNA from individual lizards (New

England Biolabs, Ipswich MA). We used Agencourt AMPure beads to ligate barcoded Illumina adaptors onto the digested DNA fragments, and then size-selected fragments in the 415-515 bp range using a Pippin Prep size fractionator. These fragments served as templates for the final library amplification, which included Illumina indexed primers and proofreading *Taq* polymerase. We pooled equimolar amounts of 12 libraries (eight samples/library) for a single lane of sequencing (100bp single end reads) on an Illumina HiSeq 4000 at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

We processed the Illumina sequencing reads using iPYRAD v.0.5.15 (http://ipyrad.readthedocs.org). This software uses a series of data filtering steps to first assign the many millions of sequencing reads to individual samples based on indexes and barcodes, and then removes data that fail to meet quality standards specified by the user. It then implements several executable dependencies to identify homologous sequences and aligns the sequence reads from the same locus within and across individuals. For this study, the iPYRAD output consisted of two datasets comprised of concatenated alignments of single SNPs isolated from each RAD sequence read. Assembly parameters for both datasets included the following: minimum percent sequence similarity required for reads at a final locus = 0.90 ('Wclust' parameter in PYRAD); minimum number of reads to make a statistical base call at a particular site per locus = 10 ('Mindepth' parameter). The two datasets differed in the proportion of samples allowed to have missing data in a final locus (minimum taxon coverage, or 'MinCov'), which was set to 0.10 and 0.20, respectively.

The goal of generating these different datasets is to examine how the iPYRAD filtering parameters affected the final dataset assemblies, and in turn how this affected our analyses. Allowing more missing data in a final locus can increase the number of SNPs recovered from the RAD sequencing library, which in turn can increase the resolution and precision of the analyses. However, previous work has shown that there is a threshold at which allowing too much missing data can confound certain types of analyses. The 10-20% missing data thresholds chosen here strikes a balance between maximizing the amount of informative data recovered from the library and minimizing the incorporation of non-informative data for the particular research questions and analyses pursued here.

For mtDNA, we sequenced 1–8 lizards for ND1 protein coding gene (1081 bp) and a small fragment of the 16s rDNA gene. We selected the ND1 gene for compatibility with previous studies (Leaché and McGuire 2006; Leaché *et al.* 2009) and included sequences from the Cedros Island horned lizard *P. cerroense* as an outgroup, as well as additional *P. blainvilli* sequences from Ventura, Los Angeles, and Kern counties from Leaché *et al.* 2009. Primers and PCR conditions are detailed in Leaché and McGuire 2006 and Leaché *et al.* 2009. Sanger sequencing for mtDNA was performed on a 3730xl DNA Analyzer at Genewiz (La Jolla, CA). The final dataset also included some additional published ND1 sequences from Los Angeles (LA), Ventura (VNTR), and Kern counties (KRN: Leaché *et al.* 2009).

Ancestry membership assignments using SNP data

A goal of this work was to identify genetic structuring in *P. blainvillii* populations across southern California and to evaluate the regional distinctiveness of that structure across the landscape. Bayesian clustering methods provide a powerful exploratory tool to address both subjects. To this end, we used STRUCTURE V2.3 (Pritchard *et al.* 2000; Falush et al 2003) to infer ancestry membership

proportions for different samples based on the SNP data. We limited these analyses to only one temporal cohort per sampling site (usually the most recent sample), and then used the additional temporal samples for a separate set of analyses focused only on sites where time series were available. The STRUCTURE software uses Bayesian statistical methods to generate a membership coefficient matrix, Q_{ik}, which contains for each individual i the membership coefficients of that individual in each of K clusters. The coefficients are interpreted as the fraction of the individual's genome represented in a particular cluster. Analyses were conducted using admixture and correlated allele frequency models, allowing individuals to have mixed ancestry in multiple clusters. We used STRAUTO v1.0 to parallelize the STRUCTURE analysis on the Yeti high performance computing cluster supported by the USGS Advanced Research Computing (ARC) program.

To visualize the estimated cluster membership coefficients, we used CLUMPP (Jakobsson and Rosenberg 2007) to align the assignment matrices for the top 10 STRUCTURE runs with the best likelihood scores at each K using the Greedy algorithm, and DISTRUCT (Rosenberg 2004) to generate assignment plots. We considered results across a range of K settings because useful information about demographic, historical and environmental processes is often gained by doing so and because of the large uncertainty of K estimates in general (Pritchard *et al.* 2000; Meirmans 2015).

Phylogeography using RAD and mtDNA sequences

We used MRBAYES 3.2 (Ronquist *et al.* 2012) to estimate a phylogenetic tree based on the mtDNA sequences. Tree searches consisted of two Markov Chain Monte Carlo (MCMC) analyses for 10 x 10^6 generations each, sampling every 1000 steps and discarding 25% of the initial samples. We analyzed four data partitions (tRNA, 1st, 2nd, and 3rd codon positions for the ND1 gene) in the gene sequences and identified the best-fit substitution models using Bayes Information Criterion in PARTITIONFINDER v1.1.0 (Lanfear *et al.* 2012). We summarized the posterior distribution of trees as a 50% majority consensus tree and considered branch support values of ≥0.95 to be statistically well-supported.

Genetic diversity and differentiation estimates through time

From SNP data, we calculated the observed heterozygosity H_o, gene diversity H_s, inbreeding coefficient G_{is}, and among-site differentiation G_{st} within a cluster using GENODIVE (Meirmans and Van Tienderen 2004). Statistical comparison of diversity indices required that samples from all sites be grouped into a specific cluster K even though lizards at many of the sites had admixed genotypes. For admixed sites, we used a cluster membership coefficient of >0.50 to place lizards from those sites into a particular cluster, and then used permutation tests to measure statistical differences in the diversity indices among clusters. We removed sites with two or fewer samples from this analysis.

Long term population viability can be studied by assessing changes in population genetic diversity and structure over time (Allendorf *et al.* 2012). Also, genetic structure evolves as a result of processes that operate over multiple generations; thus analysis of samples collected through a time series provides the opportunity to distinguish contemporary from historic influences on population structure. Sufficient sampling for statistical comparisons of temporal samples was available for the Elliot Reserve, San Diego National Wildlife Refuge, Little Cedar Ridge, and San Pasqual Valley.

For these samples, we first computed pairwise F_{st} values between or among cohorts to test for differentiation and assessed significance using non-parametric permutations. We then conducted an Analysis of Molecular Variance (AMOVA: Excoffier 1992; Michalakis and Excoffier 1996) performed on a matrix of squared Euclidean distances based on F-statistics. We assumed an Infinite Allele Model such that the F-statistics corresponded to those defined by Weir and Cockerham (1984). Standard deviations and confidence intervals were calculated by jacknifing and bootstrapping over loci (10,000 bootstraps), respectively. We also tested for differences in the observed heterozygosity H_o and gene diversity H_s by averaging the two statistics within each group and using a permutation test to evaluate any significant differences between or among them.

RESULTS

Data Summary

We collected RAD data for 192 individual lizards at 26 different localities. The mean number of RAD sequence reads across all samples that passed filtering was 2,785,670 and the mean number of homologous loci was 72,165. As expected, the more stringent minimum taxon coverage (i.e. MinCov = 0.10) implemented in the first dataset assembly yielded many fewer SNPs than did the second dataset with more relaxed coverage (i.e. MinCov = 0.20; 600 vs. 1429 SNPs, respectively). However, patterns of population structuring generated from either dataset were similar.

The mtDNA dataset (1081 characters) had 191 polymorphic sites, 138 of which were parsimony informative. The best-fit substitution model included three subsets, one for 3rd codon positions (GTR + Γ; 302 sites), one for 1st positions and the tRNA fragment combined (K80 + I; 478 sites), and one for the 2nd codon position (GTR + I + Γ; 301 sites).

Population structure

Cluster assignments reveal substantial population structuring that is consistent with regionally cohesive units. Two metrics used to approximate the number of clusters K, the ΔK statistic (Evanno et al. 2005) and the $\ln(\Pr(X|K))$ values, favored $K = 6$ for the first SNP dataset (10% missing data in a final locus) and $K = 7$ for the second dataset (20% missing data; Fig. 2). Because of this slight discrepancy in identifying an 'optimal' K, and because of the uncertainty associated with K estimates in general, we present results for $K = 6$-7 from both datasets (Fig. 3).

Some amount of genetic admixture was a general property of all inferred clusters, with the exception of the southernmost population at the Tijuana Estuary (cluster 1; Fig. 3). Here, lizards were highly distinctive from other sampling areas (although one mtDNA haplotype from the Tijuana Estuary was also recovered in lizards at Little Cedar Ridge and the Elliot Reserve). A second group included populations extending from Little Cedar Ridge north to Santa Ysabel, but with some distinction for populations north of the San Diego National Wildlife Refuge (*i.e.* clusters 2-3; Fig. 3). Populations in the Temecula Basin and Perris Plain area of Riverside County east of the Santa Ana Mountains form a cluster (cluster 5; Fig. 3), with admixture south of the basin in the Tenaja Corridor and Camp Pendleton. Santa Ana Mountain populations are also distinctive, with some admixture between the northernmost sites in Weir Canyon and Chino Hills and others to the east in the Perris Plain (cluster 6; Fig. 3). A final break, albeit poorly defined, is shown for the northernmost samples at Silverwood Lake and El Segundo Dunes.

Phylogeography

Phylogenetic analysis of mtDNA sequences revealed little fine-scale geographic structure, with substantial mixing of the same haplotypes and/or highly similar haplotypes across many sampling sites. Shared identical haplotypes were always from populations in close geographic proximity, as expected in a low-vagility lizard.

Branches leading to three major lineages had high posterior probabilities (i.e. $PP \geq 1.0$), with each having some additional well-supported substructure within them (Fig. 4). The two most divergent mtDNA lineages are in geographic contact in San Pasqual Valley, San Diego County, a division that also emerged in the SNP dataset (see paragraph below). From south to north: All haplotypes from the Tijuana Estuary north to San Pasqual Valley formed a well-supported lineage (i.e. the Northern Baja CA lineage, following the nomenclature of Leaché *et al.* 2009; Fig. 4) and all sites from San Pasqual Valley north to Silverwood Lake SRA formed a second well-supported lineage (Southern CA lineage; Fig. 4). We note the presence of more well-supported subunits within the Northern Baja CA lineage compared to the Southern CA lineage, although these subunits do not show any clear association with geography at local scales. Samples from Aliso Canyon in the northern San Gabriel Mountains and the El Segundo Dunes in the western Los Angeles Basin were members of a third lineage (i.e. Northern CA) that extends northward into Kern County, California.

To better test for signal concordance in the SNP and mtDNA datasets, we fit the SNP data to $K = 2$ in STRUCTURE to examine whether sampling sites clustered in a manner consistent with the Northern Baja CA and Southern CA mtDNA lineages (Fig. 4). These results showed a good correspondence between the two datasets, although the geographic extent of admixture spans more broadly to the north and south in the SNP vs. the mtDNA dataset (*i.e.* in SNPs, from as far south as El Monte Valley in Lakeside, San Diego County to North Hills in Riverside County; Fig. 3). In contrast, sympatric mtDNA haplotypes were restricted to San Pasqual Valley. Regardless, the geographic boundaries between northern and southern groups occurred at essentially the same latitude for both SNP and mtDNA datasets.

Genetic diversity and differentiation

Most but not all pairwise F_{st} estimates between sampling localities indicated significant genetic differentiation (table not shown). For sites with two or more samples, F_{st} values ranged from a low of 0.006 between Rancho Jamul and Hollenbeck Canyon ($P = 0.291$) to a high of 0.256 between Silverwood Lake and Covington Flats ($P = 0.002$), and as expected the lowest F_{st} estimates were always between geographically proximate sites. We found no statistically significant differences among the inferred clusters in any genetic diversity indices or in the levels of genetic differentiation among populations within clusters.

For all populations with temporal sampling, the main source of genetic variation stems from differences among individual lizards at a site (range = 0.69-0.83), as expected, followed by among individuals nested within temporal samples (range = 0.17-0.25; Table 2). Only the Little Cedar Ridge and the San Diego NWR populations showed evidence of genetic differentiation between temporal samples, with the former being substantially higher ($F_{st} = 0.056$, $P = 0.001$) than the latter ($F_{st} = 0.018$, $P = 0.015$; Table 2). These results indicate significant shifts in the allele frequencies at these

sites. At the same time, we found no statistical evidence for changes in the observed heterozygosity H_o or gene diversity H_s in either population between the earliest and most recently collected samples.

DISCUSSION

Our results show a high degree of genetic structuring in *P. blainvilli* across southern California and a consistency with the sampling geography. The inferred clusters in most cases extend across broad geographic areas, even though their constituent populations are no longer connected by continuous habitat. This genetic cohesion in the presence of contemporary isolation indicates historically widespread population connectivity in *P. blainvilli*, a natural history that is at odds with the current fragmented landscape in southern California.

Lag times between the interruption of gene flow due to habitat loss or fragmentation (or otherwise) and the genetic responses associated with that process are common and can vary due to a variety of factors (Sumner *et al.* 2004; Richmond *et al.* 2008; Landguth *et al.* 2010). In general, larger population sizes and longer generation times can prolong this inertia by buffering against the effects of genetic drift, which leads to the random loss or fixation of alleles over time in a closed population. Even small amounts of migration with gene flow can further decelerate or counteract the process (Busch *et al.* 2007; Shama *et al.* 2011). In *P. blainvilli*, a combination of the recency of habitat fragmentation and sustained, large population sizes (or at least sizes above some critical threshold at which drift would cause local divergence) may explain why the genomes of these lizards have retained signals of their former inter-population connectivity.

Despite this continuity across large tracts of land, notable genetic breaks exist over the sampling area. The starkest break, in terms of the exclusivity of a single cluster, is the Tijuana Estuary population. Although one mtDNA haplotype from this site was also recovered at Little Cedar Ridge and the Elliot Reserve, we find no evidence of any recent gene exchange outside of this population. It is possible, and in fact likely, that the 'Tijuana cluster' extends south of the international border, as indicated by Leaché *et al.* 2009, and may even represent the northern extension of the closely related and endemic Baja species *P. cerroense*. Further sampling in northern Baja California is necessary to test either hypothesis; however, the fact remains that the Tijuana Estuary population is highly distinctive from all others in the US and persists on a very narrow, isolated stretch of unique coastal dune habitat.

A second major break in the SNP dataset occurs in central coastal San Diego County between Carlsbad/San Marcos and Camp Pendleton (clusters 3-5; Fig. 3). There is no obvious historical movement barrier that would have disrupted population linkages across this area, although the break appears to coincide with the east-west lying San Luis Rey Watershed. The sampling gap between Rancho La Costa and Camp Pendleton could be accentuating this signal given the low movement capabilities of the lizard (*i.e.* genetic isolation by distance; Fig. 1), whereas intervening populations (e.g. Fallbrook Naval Weapons Station) may have revealed more of a clinal shift in the allele frequencies than that shown here.

Further north, a third break distinguishes Santa Ana Mountain populations in Orange County from those in the Temecula Basin and Perris Plain in Riverside County (clusters 5-6; Fig. 3). Again, there is a sampling gap in the coastal plains to the southwest of the Santa Ana Mountains that could be amplifying the distinctions between Santa Mountain and Temecula Basin populations (Fig. 1), although this same break in gene continuity has been detected in other vertebrate taxa in this same area (see '*Comparisons of population structure with other taxa*' below).

A final break distinguishes populations north of Chino Hills (cluster 7; Fig. 3), but sampling density is much lower and more dispersed beyond the Santa Ana Mountains. As such, the limits and integrity of this cluster cannot be assessed with the current data. The northernmost samples in this study, Aliso Canyon, were collected opportunistically during other field work and were sequenced only for mtDNA given the small sample size and the large geographic distance separating this site from all others. These northern sites were also not the focus of this study.

Comparisons of population structure with other taxa

In terms of sampling density and the amounts and types of data, few genetic studies exist for other vertebrate taxa occurring across a similar portion of southern California. Of the three that are most comparable, two are bird species (California gnatcatcher *Polioptila californica* and the cactus wren *Camphylorynchus brunneicapillus*) and one is a large mammal (puma *Felis concolor*). Flight obviously engenders greater movement capacity in the two bird species compared to horned lizards, but both are habitat restricted, occur in localized aggregations, and prefer the same types of habitats occupied by *P. blainvilli* (primarily coastal sage scrub and chaparral). Like *P. blainvilli*, linkage habitat is considered critical for long-term population persistence for these bird species due to their fidelity to these declining habitats (Atwood 1993).

While the gnatcatcher shows no discernable genetic structure across the species range (Vandergast *et al.* 2014), clustering patterns for the cactus wren are remarkably similar to those for *P. blainvilli* (Barr *et. al* 2013; Barr *et al.* 2015). In fact, the only real differences between the two is as follows: (1) the Tijuana Estuary population for cactus wren clusters with other populations to the immediate east at Otay and Encanto, whereas in *P. blainvilli* it does not; and (2) *P. blainvilli* populations from Spring Canyon north to Santa Ysabel are genetically cohesive but with some distinction in the amount of admixture north and south of the San Diego NWR (clusters 2-3; Fig. 3), whereas cactus wrens separate into two more distinctive groups over this same general area (SD and PASQ in Barr *et. al* 2013). Further north, cactus wrens have a Riverside County assemblage similar to cluster 5 in *P. blainvilli*, and Orange County/Santa Ana Mountains populations (OCPN in Barr *et al.* 2013) group in a manner similar to cluster 6 for *P. blainvilli* (Fig. 3). In cactus wren, the OCPN cluster also includes populations throughout Camp Pendleton, whereas *P. blainvilli* populations at Camp Pendleton are highly admixed and at least somewhat unique in their allelic composition relative to adjacent areas (cluster 6; Fig. 3)

A major break in gene continuity also occurs for cactus wren and *P. blainvilli* at the northern end of the Santa Ana Mountains, although for cactus wren the break is slightly further south than in *P. blainvilli*. In *P. blainvilli*, the Santa Ana Mountains cluster (i.e. #6; Fig. 3) includes Chino Hills State Park, whereas in cactus wren, Chino Hills populations belong to a separate cluster with geographic affinities further to the north and west. In fact, the boundary for cactus wren closely adheres to the 91 freeway. The puma *Felis concolor* also displays a high degree of genetic exclusivity for Santa Ana Mountain populations; however for puma, much of this exclusivity is due to a recent and severe demographic decline combined with genetic isolation. The gene isolation is driven mainly by contemporary urbanization that impedes movement into and out of the Santa Ana Mountains.

Shifts in diversity through time

We found no evidence of any significant shift in allele frequencies, observed heterozygosity, or gene diversity over a ~15-year time period for *P. blainvilli* populations at the Elliot Reserve and San

Pasqual Valley; however, we did find significant shifts at the San Diego NWR and Little Cedar Ridge. Interestingly though, this shift did not translate to significant reduction or increase in observed heterozygosity or gene diversity at either site. Typically, allele shifts over such a short time period and of the magnitude shown here would be caused by a sudden change in demography, either a 'bottleneck' due to a catastrophic event (e.g. fire or flood) or a rapid influx of migrants introducing different genetic variation. The former would purge alleles from the population, driving diversity down, whereas the latter would introduce them through gene flow, driving diversity up. However, we see neither of these patterns at the San Diego NWR and Little Cedar Ridge, despite significant shifts in the allele frequencies.

One possible explanation is that both populations experienced bottlenecks after the first sampling took place. This could have altered the allele frequencies and reduced genetic diversity before the second sampling, resulting in the significant F statistics. However, rapid immigration would have had to occur for there to be no differences in heterozygosity or gene diversity between samples, given that immigrants would represent the source of new alleles that replenished the lost variation caused by the bottlenecks.

At least some evidence supports this scenario. Little Cedar Ridge experienced two severe fires in 2003 and 2007, which could have negatively impacted horned lizard densities. At the San Diego NWR, a major urban development (Rancho San Diego Towne Center) was constructed immediately adjacent to the sampling sites after the first sample was collected, but was completed before acquiring the second sample. Both events could have feasibly caused *P. blainvilli* declines between sampling events. However, the possibility that immigration from surrounding, unaffected areas would have been sufficient to counteract the loss of variation from bottlenecks seems less likely given (1) the low movement capacity of *P. blainvilli* and (2) the fact that would-be sources of immigrants would probably have similar allele frequencies to the bottlenecked populations.

Nonetheless, rapid rebounds with similar findings to this study have been well-documented in populations of the banner-tailed kangaroo rat *Dipodomys spectabilis* (Busch *et al.* 2008) and alpine caddisfly *Allogamus uncatus* (Shama *et al.* 2011). Both examples involve large metapopulation systems, where local aggregations are spatially connected in a manner that allows rapid recolonization into areas affected by demographic bottlenecks. These studies speak to the importance of maximizing connectivity in contemporary landscapes, given that without adequate linkage habitat these species would not be able to respond as quickly or as effectively to local demographic declines.

We caution against making strong interpretations about these results at the present time given that further SNP data exploration are necessary to rule out whether data artifacts may be driving some of the signal. The most recent temporal sample from Little Cedar Ridge contains two individuals with substantial missing data, and the most recent sample from the San Diego NWR consists of only five individuals. The second sample for the San Diego NWR was also obtained from a slightly different part of the refuge than the first sample, and as such measures of genetic differentiation would also need to be corrected for the geographic distance separating the two sampling areas. Effects of missing data and small sample size on these analyses has yet to be thoroughly vetted, although simulation studies have shown that smaller sample sizes for SNP datasets can still produce reliable results for similar types of analyses due to the high number of markers for individual samples.

Implications for land acquisition, restoration, and conservation

Our results provide evidence for historically widespread gene connectivity of *P. blainvilli* populations in southern California, emphasizing the importance of the MSCP approach – maximizing habitat linkages between core areas known to function as movement and gene exchange corridors for wildlife – in managing the changing landscape in southern California. If horned lizards are viewed as a suitable ecological indicator species, and if their genetic data are to aid in guiding land acquisition, restoration and preservation efforts, one approach would be to target non-converted lands between and among populations represented within the different clusters identified in this study. This would preserve connections that we know were important for maintaining population viability in the most recent configuration of the unurbanized landscape. If connections can be preserved or restored at this geographic level (i.e. within clusters), a second tier management effort could then focus on connecting lands between and among the major genetic lineages.

As we have shown here, the strong similarities in population gene structuring for cactus wren, *P. blainvilli*, and even the puma, all point to a common set of core habitat areas and linkages that are historically important for maintaining population viability in a breadth of taxa.

Next steps

This work has narrowed the geography of different contact zones between genetically differentiated groups of *P. blainvilli*, yet sampling gaps still leave some important questions unanswered. Sampling gaps of interest are as follows:

- Fallbrook area between sites sampled on Camp Pendleton and Rancho La Costa: Filling this gap would provide insight on the width of the allelic cline between Riverside County populations and those to the south in San Diego County. Including samples from this area may also better resolve the structure for Camp Pendleton and Tenaja Corridor populations, which are highly admixed in the current sample.
- Coastal properties in the NCCP/HCP Nature Reserve of Orange County (Aliso and Wood Canyons; Laguna Coast; Crystal Cove; City of Irvine Open Space Preserve): We were unable to incorporate these sites, yet their inclusion is critical to evaluate *P. blainvilli* genetic structuring and diversity across as many conserved lands as possible in southern California. Knowing the genetic affinities of these populations would also be valuable for targeting linkage corridors that were historically important for maintaining population viability and for informing translocation or gene rescue efforts should they become necessary.
- Garner Valley in the southwestern San Jacinto Mountains, Riverside County: Lizards from this area would bridge a large sampling gap between Temecula Basin/Perris Plain and Covington Flats in Joshua Tree National Monument. In this study, data from Garner Valley were limited to mtDNA given that there was insufficient tissue to generate the necessary DNA concentrations for RAD sequencing.
- Northern Baja California: The Tijuana Estuary population is genetically exclusive with respect to other populations sampled for this study – what remains unclear is whether these lizards are an isolated segment of a larger cluster that persists in northern Baja California. If they are a single isolated entity restricted to the sand dunes of the Tijuana Estuary, this would have important, immediate management implications.

Sample sizes for some sites were also too low to provide robust estimates of genetic diversity and differentiation. Any sites represented by fewer than eight samples in the current RAD dataset are candidates for additional data collection, particularly if they occur in potential contact zones. One site of particular interest is Spring Canyon given its location between the Tijuana Estuary and the Otay Mesa area. A sample of 8-10 lizards from this site, assuming the species still exists there, would allow us to better test whether there was any east-west gene connectivity between the now-isolated Tijuana Estuary population and Otay Mesa. Because of the signals of temporal genetic differentiation at the San Diego NWR and Little Cedar Ridge sites, additional contemporary sampling from both sites also merits further consideration.

ACKNOWLEDGEMENTS

LITERATURE CITED

Allendorf FW, Luikart GH, Aitken SN (2012). Conservation and the Genetics of Populations. John Wiley and Sons, West Sussex.

Andelman SJ, Fagan WF (2000) Umbrellas and flagships: Efficient conservation surrogates or expensive mistakes? *Proceedings of the National Academy of Sciences*, **97**, 5954–5959.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews in Genetics*, **17**, 81-92.

Archer FI, Adams PE, Schneiders BB (2016) STRATAG: An R package for manipulating, summarizing and analyzing population genetic data. *Molecular Ecology Resources*, **17**, 5-11.

Atwood, J. L. 1993. California Gnatcatchers and coastal sage scrub: the biological basis for endangered species listing. Pages 149-169 *in* J. E. Keeley, editor. Interface between ecology and land development in California. Southern California Academy of Science, Los Angeles.

Brattstrom BH (2103) Distribution of the coast horned lizard, *Phrynosoma coronatum*, in southern California. Bulletin of the Southern California Academy of Sciences, **112**, 206-216.

Breckheimer I, Haddad NM, *et al* (2014) Defining and evaluating the umbrella species concept for conserving and restoring landscape connectivity. *Conservation Biology*, **28**, 1584-1593.

Busch JD, Waser PM, DeWoody JA (2007) Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology*, **16,** 2450-2462.

Caro TM (2010) Conservation by proxy: indicator, umbrella, keystone, flagship, and other surrogate species. Island Press, Washington, DC

Crooks KR, Sanjayan MA (2006) Connectivity conservation. Cambridge University Press, Cambridge, United Kingdom.

Damschen EI, Haddad NM, Orrock JL, Tewksbury JJ, Levey DJ (2006) Corridors increase plant species richness at large scales. *Science*, **313**, 1284–1286.

Dobson AP, Rodriguez JP, Roberts WM, Wilcove DS (1997) Geographic distribution of endangered species in the United States. *Science*, **275**, 550–553.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molecular Biology And Evolution*, **29**, 1969–1973.

Eaton DAR (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analysis. *Bioinformatics*, **30**, 1844–1849.

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

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics*, **131**, 479-491.

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.

Fahring L (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology and Systematics*, **34**, 487–515.

Fisher RN, Suarez AV, Case TJ (2002) Spatial patterns in the abundance of the coastal horned lizard. *Conservation Biology*, **16**, 205-215

Fordham DA, Brook BW, Moritz C, Nogués-Bravo D (2014) Better forecasts of range dynamics using genetic data. *Trends in Ecology and Evolution*, **29**, 436–443.

Funk WC, McKay JC, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**, 489–495.

Gilbert-Norton L, Wilson R, Stevens JR, Beard KH (2010) A meta-analytic review of corridor effectiveness. *Conservation Biology*, **24**, 660–668.

Gonzalez A, Lawton JH, Gilbert FS, Blackburn TM, Evans-Freke I (1998) Metapopulation dynamics, abundance, and distribution in a microecosystem. *Science*, **281**, 2045–2047.

Haddad NM, Bowne DR, *et al.* (2003) Corridor use by diverse taxa. *Ecology*, **84**, 609–615.

Hilty JA, Lidicker Jr WZ, Merenlender AM (2006) Corridor ecology: the science and practice of linking landscapes for biodiversity conservation. Island Press, Washington, DC.

Holway DA (1995) The distribution of the Argentine ant in northern California. *Conservation Biology*, **9**, 1634-1637.

Holway DA (1998) Factors governing the rate of invasion: a natural experiment using Argentine ants. *Oecologia*, **115**, 206-212.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jennings MR (1987) Impact of the curio trade for San Diego horned lizards (*Phrynosoma coronatum blainvillii*) in the Los Angeles basin, California: 1885-1930. *Journal of Herpetology*, **21**, 356-358.

Jennings MR (1988) *Phrynosoma coronatum. Catalogue of American Amphibians and Reptiles*, **428**, 1-5.

Stebbins, R.C. 2003. *A Field Guide to Western Reptiles and Amphibians.* Third Edition. Houghton Mifflin Company, Boston, Massachusetts.

Lande R (1987) Extinction thresholds in demographic models of territorial populations. *The American Naturalist*, **130**, 624-635.

Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179-4191.

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

Leaché AD, McGuire JA (2006) Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial DNA: Evidence for a misleading mitochondrial gene tree. *Molecular Phylogenetics and Evolution*, **39**, 628-644.

Leaché AD, Koo MS, Spencer CL, Papenfuss TJ, Fisher RN, McGuire JA (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences*, **106**, 12418-12423.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032–1047.

Lemm J (2006) Field Guide to Amphibians and Reptiles of the San Diego County Region. University of California Press, Berkeley.

Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.

Meirmans PG, Van Tienderen (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792-794.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

McGeoch MA (2007) Insects and bioindication: theory and progress. In: *Insect Conservation Biology*, (eds Stewart AJA, New TR, Lewis OT), pp. 144–74. London: The Royal Entomological Society.

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, **24**, 3223–3231.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 1061-1064.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135. DOI: 10.1371/journal.pone.0037135

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

Rambaut A, Drummond AJ (2007) Tracer v1.4, available from http://beast.bio.ed.ac.uk/Tracer

Richmond JQ, Reid DT, Ashton KG, Zamudio KR (2009) Delayed genetic effects of habitat fragmentation on the ecologically specialized Florida sand skink (*Plestiodon reynoldsi*). *Conservation Genetics*, **10**, 1281-1297.

Roberge J-M, Angelstam P (2004) Usefulness of the Umbrella Species Concept as a Conservation Tool. *Conservation Biology*, **18**, 76-85.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Shama LNS, Kubow KB, Jokela J, Robinson CT (2011). Bottlenecks drive temporal and spatial genetic changes in alpine caddisfly metapopulations. *BMC Evolutionary Biology*, **11**, 278.

Suarez AV, Bolger DT, Case TJ (1998) The effects of fragmentation and invasion on the native ant community in coastal Southern California. *Ecology*, **79**, 2041-2056.

Suarez AV, Richmond JQ, Case TJ (2000) Prey selection horned lizards following the invasion of Argentine ants in southern California. *Ecological Applications*, **10**, 711-725.

Suarez AV, Holway DA, Case TJ (2001) Predicting patterns spread in biological invasions dominated by jump dispersal: lessons from Argentine ants. *Proceedings of the National Academy of Sciences USA*, **98**, 1095-1100.

Sumner J, Jessop T, Paetkau D, Moritz C (2004) Limited effect of anthropogenic fragmentation on molecular diversity in a rain forest skink *Gnypetoscincus queenslandiae*. *Molecular Ecology*, **13**, 259-269.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Thompson RC, Wright AN, Shaffer HB (2016) California Amphibian and Reptile Species of Special Concern. University of California Press, Berkeley USA

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

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{st} do not necessarily require large sample sizes when using many SNP markers. *PloS One*, **7**, e42649. 9. doi:10.1371/journal.pone.0042649

TABLE 1. Sampling locations and numbers of samples per locality/dataset. Latitude and longitude data represent the center of the sampling points at a given site (WGS84; individual sampling points are available upon request).

Sites	County	Latitude	Longitude	Total RAD	Total mtDNA
Agua Chinon	Orange	33.6997	-117.6830	7	6
Agua Tibia Mtns.	Riverside	33.4470	-116.9866	—	2
Aliso Canyon	Los Angeles	34.4504	-118.1526	—	1
Anza	Riverside	33.4708	-116.7232	—	2
Anza Borrego Desert	Imperial	33.4626	-116.6310	—	1
Beauty Mountain	Riverside	33.4668	-116.6967	—	3
Camp Pendleton	San Diego	33.3978	-117.4478	4	2
Cherry Valley	Riverside	33.9907	-116.9840	—	1
Chino Hills	Orange	33.8855	-117.7296	8	7
Covington Flats	San Bernardino	34.0345	-116.3362	5	7
El Monte Valley	San Diego	32.8684	-116.8933	3	3
Elliott Reserve 1995	San Diego	32.8917	-117.0935	8	4
Elliott Reserve 2006	San Diego	32.8911	-117.0960	6	3
Elliott Reserve 2010-12	San Diego	32.8915	-117.0933	8	5
Garner Valley	Riverside	33.6586	-116.6437	—	7
Gavilan Mountains	Riverside	33.6489	-117.2650	—	1
Hollenbeck Canyon	San Diego	32.6779	-116.8192	5	4
Lake Perris SRA	Riverside	33.8699	-117.1700	1	1
Lake Skinner	Riverside	33.5923	-117.0376	8	9
Limestone Canyon	Orange	33.7347	-117.6884	8	5
Little Cedar Ridge 1995	San Diego	32.6264	-116.8646	8	2
Little Cedar Ridge 2011	San Diego	32.6238	-116.8632	9	4
El Segundo Dunes	Los Angeles	32.9186	-117.2142	2	2
Motte Rimrock Reserve	Riverside	33.8101	-117.2568	4	4

Murrieta	Riverside	33.5653	-117.2062	—	2
North Hills	Riverside	33.7012	-117.0167	6	5
Potrero	Riverside	33.8401	-116.8860	—	2
Rancho Jamul	San Diego	32.6753	-116.8496	8	3
Rancho La Costa	San Diego	33.0961	-117.2244	3	1
Rawson Canyon	Riverside	33.6301	-117.0044	9	6
San Diego NWR 1995-96	San Diego	32.7226	-116.9489	8	4
San Diego NWR 2008-10	San Diego	32.7244	-116.9152	5	2
San Mateo Camp Pendleton	San Diego	33.4182	-117.5470	—	1
Santa Margarita ER	San Diego	33.4388	-117.1800	—	1
San Jacinto River	Riverside	33.7696	-116.9200	—	1
Santa Rosa Hills	Riverside	33.6946	-116.8980	—	1
Santa Ysabel ER	San Diego	33.1236	-116.6884	8	5
Silverwood Lake SRA	San Bernardino	34.2998	-117.3384	7	5
Spring Canyon	San Diego	32.5601	-117.0010	1	1
Tenaja Corridor	San Diego	33.5057	-117.3497	9	8
Tijuana Estuary	San Diego	32.5467	-117.1248	8	7
Torrey Pines 2	San Diego	32.9408	-117.2494	8	4
Tule Peak Road	Riverside	33.4712	-116.7380	—	1
Weir Canyon	San Diego	33.8381	-117.7228	5	5
San Pasqual Valley 1995-98	San Diego	33.0954	-116.9801	8	8
San Pasqual Valley 2008-11	San Diego	33.0963	-116.9790	5	5

Table 2. Results of the AMOVA for sites with temporal sampling. The numbers in parentheses represents the number of temporal groups.

Site	$n$loci	Source of Variation	Nested in	%var	F-stat	F-value	SD	c.i.2.5%	c.i.97.5%	P-value
Elliot Reserve	1567	Among individuals	--	0.801	F_{it}	0.199	0.008	0.183	0.215	--
		Among individuals	Population	0.199	F_{is}	0.199	0.008	0.183	0.215	0.001
		Among temp. groups (3)	--	0.000	F_{st}	0.000	0.003	-0.006	0.005	0.513
San Diego NWR	634	Within Individual	--	0.690	F_{it}	0.310	0.016	0.278	0.342	--
		Among Individual	Population	0.254	F_{is}	0.269	0.016	0.236	0.301	0.001
		Among temp. groups (2)	--	0.056	F_{st}	0.056	0.012	0.033	0.080	0.001*
San Pasqual Valley	3007	Within Individual	--	0.831	F_{it}	0.169	0.008	0.153	0.184	--
		Among Individual	Population	0.166	F_{is}	0.167	0.008	0.151	0.182	0.001
		Among temp. groups (2)	--	0.002	F_{st}	0.002	0.003	-0.004	0.009	0.321
Little Cedar Ridge	952	Within Individual	--	0.806	F_{it}	0.194	0.012	0.170	0.217	--
		Among Individual	Population	0.176	F_{is}	0.179	0.012	0.155	0.203	0.001
		Among temp. groups (2)	--	0.018	F_{st}	0.018	0.005	0.009	0.027	0.015*

FIGURE 1. Sampling sites overlaid on conserved lands in southern California (map layer available at http://www.sangis.org/download/available.html). Blue dots indicate sites included in both the mtDNA and SNP datasets; yellow dots indicate mtDNA only. The Aliso Canyon site is not pictured on the map, but occurs in the northern San Gabriel Mountains near the city of Lancaster. Dashed lines approximate the major genetic breaks discussed in the text; yellow opaque polygons indicate sampling gaps in the RAD dataset mentioned in '*Next steps*' in the Discussion.

FIGURE 2. Approximation of the number of clusters based on the ΔK statistic land the curve of the $\ln(\Pr(X \mid K)$ values for datasets with (A) 10% and (B) 20% of samples allowed to have missing data in final locus.

FIGURE 3. Inferred cluster membership proportions from STRUCTURE analyses ($K = 6\text{-}7$ [top and bottom for each panel]) based on datasets with (A) 10% and (B) 20% missing data in a final RAD locus. Populations are oriented latitudinally from left to right, with the northernmost samples on the left side of the plot and the southernmost on the right side. The ΔK statistic favors $K = 6$ for (A) and $K = 7$ for (B); results are shown for both to account for the general uncertainty in estimating an optimal K. Colored squares at the bottom provide a numerical identifier for each cluster and are used only to increase the clarity of text.

FIGURE 4. Mitochondrial gene tree (top) and cluster membership assignments (bottom) for $K = 2$ for SNPs (10% MinCov). Black dots on the tree nodes indicate branch support values of $\geq$ 0.95 posterior probability. Blue arrows highlight San Pasqual Valley haplotypes belonging to both the Southern CA and Northern Baja CA lineages. Colored bars above the assignment plot indicate mtDNA clade affinities of the different sampling sites matched against the SNP clusters. The SNP data suggest a much broader zone of admixture (dark bar above the assignment plot) compared to mtDNA data, where overlap is limited to San Pasqual Valley.
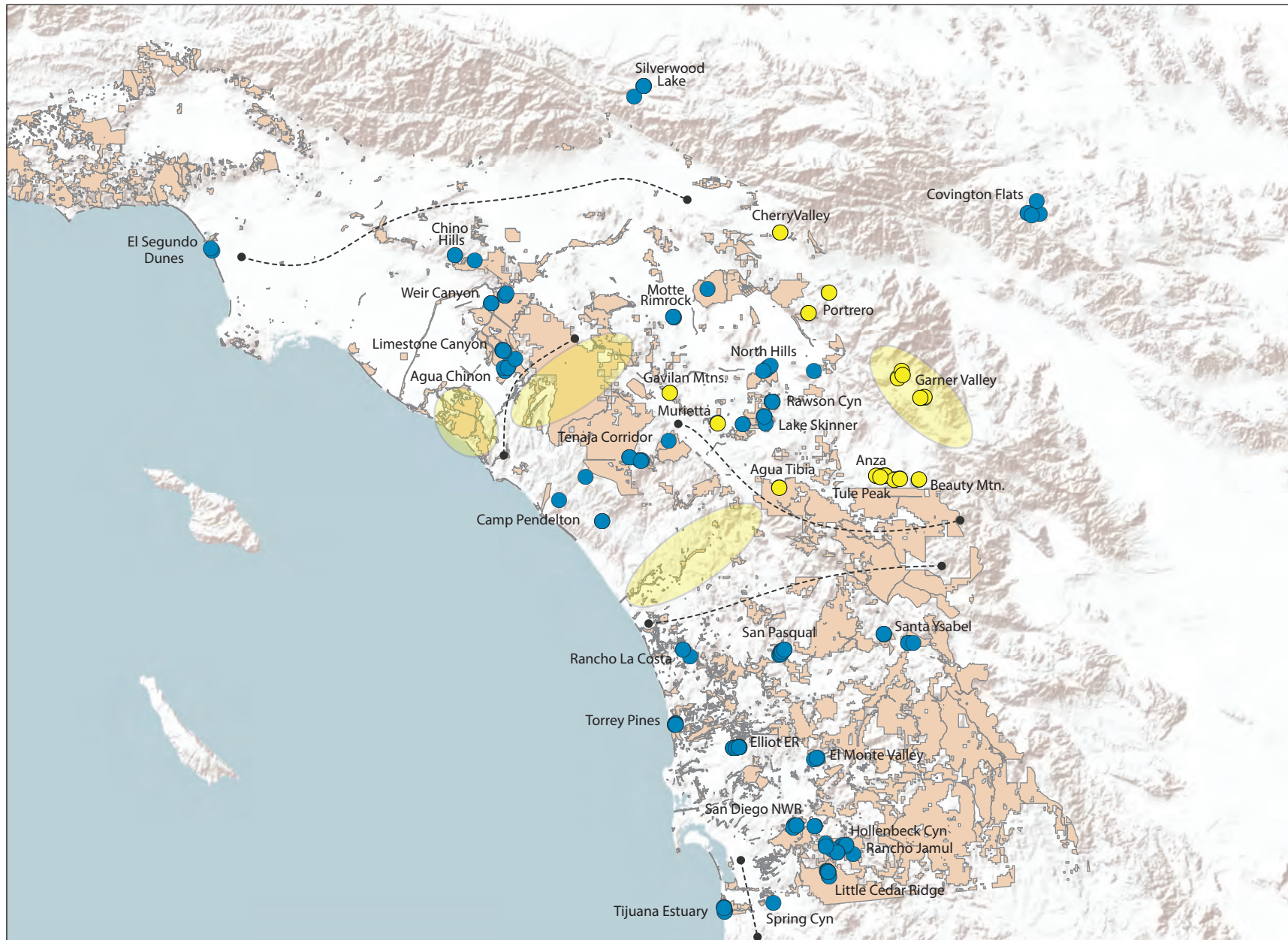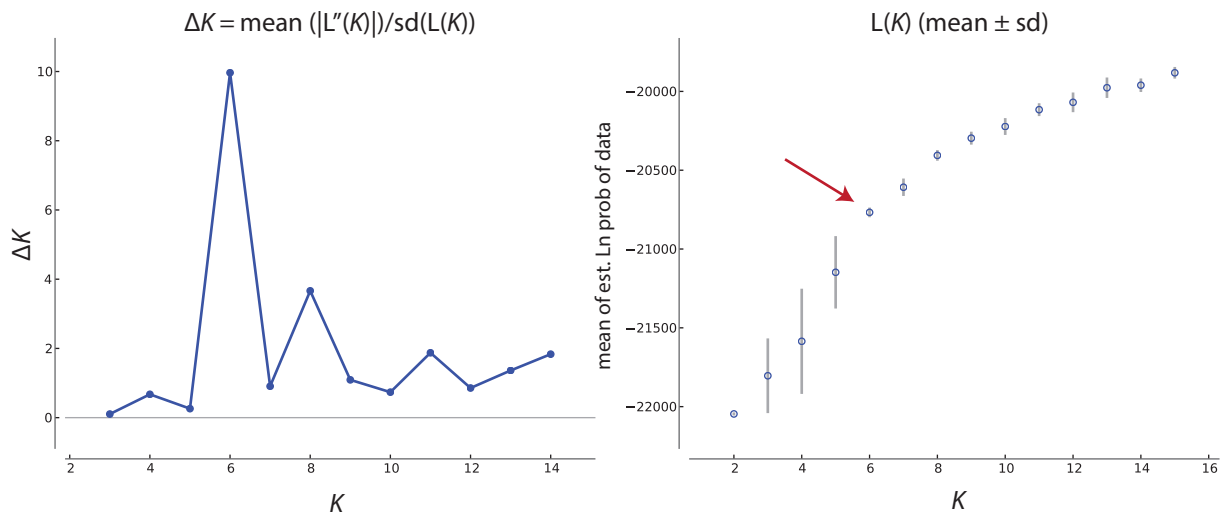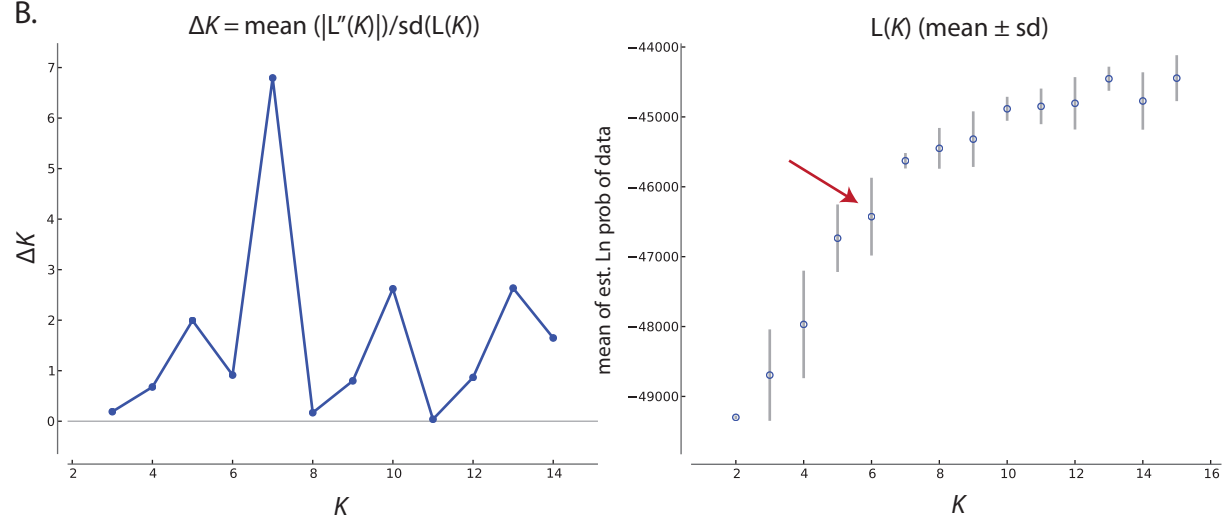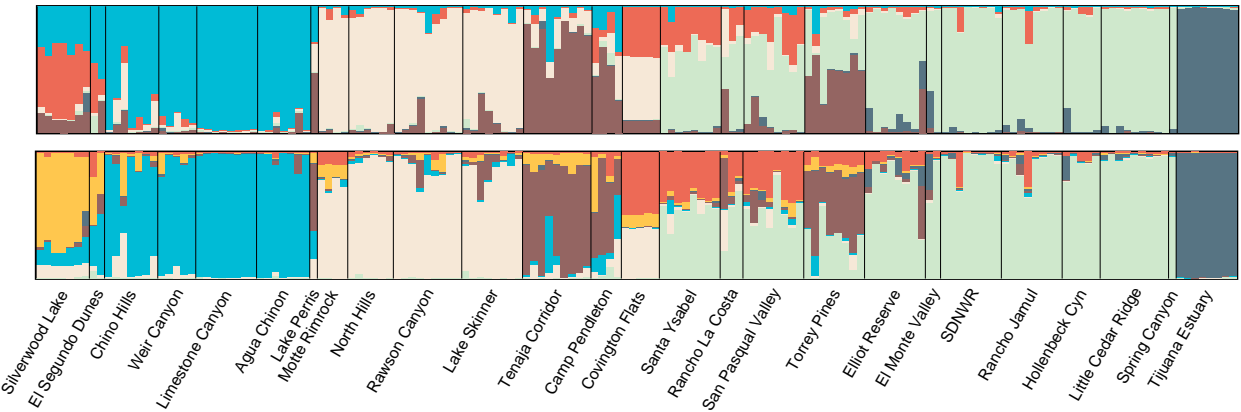
Figure 1.

Figure 2

A.

$\Delta K = mean\ (|L''(K)|)/sd(L(K))$

L(K) (mean ± sd)

B.

$\Delta K = mean\ (|L''(K)|)/sd(L(K))$

L(K) (mean ± sd)

Figure 3

Figure 4