Quantitative assessment of population genetic structure and historical phylogeography of the Alameda Whipsnake *Masticophis lateralis euryxanthus*

Jonathan Q. Richmond[1], Dustin A. Wood, Cindy Hoang, and Amy G. Vandergast

US Geological Survey
Western Ecological Research Center
San Diego Field Station
4165 Spruance Rd. Suite 200
San Diego, CA 92101-0812.

[1] Author to whom correspondence should be addressed (jrichmond@usgs.gov; 619-225-6434).

Table of Contents

	Page
Glossary of terms	1
1. Summary	4
1.1 Background	4
1.2 Major findings	4
1.3 Conclusions	5
2. Introduction	5
3. Material and Methods	7
3.1 Tissue sample acquisition	7
3.2 mtDNA sequence data collection	9
3.3 STR data collection	9
3.4 Organization of analytical methods	10
3.5 Phylogeography, genetic diversity and historical demography based on mtDNA	10
3.5.1 Range-wide phylogeography of M. lateralis	10
3.5.2 Genetic diversity and differentiation: mtDNA	11
3.5.3 Estimates of divergence times	12
3.5.4 Inferences of historical and recent changes in demography	12
3.6 Population genetic analysis of East Bay whipsnakes using STR allele frequencies	14
3.6.1 A brief background on Bayesian assignment methodology	15
3.6.2 Tuning spatial priors	16
3.6.3 Inferring the number of clusters K and cluster assignment	18
3.6.4 Genetic diversity and differentiation: STR loci	19
3.7 Isolation-by-distance analyses	20
4. Results	20
4.1 Analysis of historical phylogeography and demography based on mtDNA	20
4.1.1 Range-wide phylogeography of M. lateralis	20
4.1.2 Genetic diversity and differentiation: STR loci	21
4.1.3 Estimates of divergence times	23
4.1.4 Inferences of historical and recent changes in demography	24
4.2 Analysis of population dynamics based on STR allele frequencies	27
4.2.1 Estimates of the number of clusters	27
4.2.2 Best fit model for estimating admixture coefficients	27
4.2.3 Cluster assignment	29
4.2.4 Genetic diversity and differentiation: STR loci	31
4.3 Isolation-by-distance analyses	34
5. Discussion	34
5.1 Historical biogeography, clade structuring and divergence times; implications for conservation management	36
5.1.1 Do mtDNA clades have implications for taxonomy?	38
5.2 Implications of the Alameda whipsnake phenotype	39
5.3 Population dynamics of East Bay whipsnake populations; patterns and process	40
5.3.1 Genetic diversity status of East Bay populations	42
5.4 Utilizing genetic data in defining *Mle* managements units	43
5.5 Additional research for building a comprehensive *Mle* management plan	46
5.5.1 Best estimates of southern and northern genetic boundaries for Mle	46
5.5.2 Genetic admixture of intermediately spaced samples	47

5.5.3 Roadways as barriers to migration and gene exchange 47
5.5.4 Subspecies phenotypic boundaries 47
6. Acknowledgments 48
7. Literature Cited 48
8. Figures and Tables

Figures
Figure 1. Map of East Bay sampling localities. 8
Figure 2. Voronoi tessellation of sampling sites in the East Bay. 17
Figure 3. Distribution and phylogeography of *M. lateralis*. 22
Figure 4. Bayesian relaxed-clock estimates of divergence times. 24
Figure 5. Bayesian skyline plots for the three major mtDNA clades in California. 25
Figure 6. Mitochondrial DNA haplotype network with a map of corresponding sampling sites. 26
Figure 7. DIC and lnP(D|K) curves used to estimate K_{MAX} in TESS and STRUCTURE. 28
Figure 8. Voronoi tessellation indicating no-admixture cluster assignments. 30
Figure 9. Estimated cluster assignment probabilities for East Bay whipsnake populations. 30
Figure 10. Neighbor-joining tree based on STR theta values. 33
Figure 11. Reduced major axis regression of pairwise genetic similarity values against geographic distance. 35
Figure 12. Conservation priority assignments 45

Tables
Table 1. STR genetic diversity estimates for East Bay sampling sites. 7
Table 2. Best-fit nucleotide substitution models for the mtDNA data partitions. 11
Table 3. Measures of polymorphism and tests of demographic expansion. 21
Table 4. Average sequence divergence within and between the major mtDNA clades. 23
Table 5. Model comparisons using the Deviance Information Criterion for Bayesian assignment tests. 29
Table 6. Comparison of R_{ST} and F_{ST} values by population using Brookfield adjusted versus non-adjusted datasets. 32
Table 7. Results of the Mantel and partial mantel analysis. 34

Appendices (separate document)
8.1 Voucher data for specimens used in the mtDNA analyses
8.2 Map of mtDNA sampling points with corresponding voucher identifications
8.3 Descriptions of STR motifs, multiplex primer sets, and primer sequences
8.4 Allelic patterns across sampling sites
8.5 Nested clade inferences.
8.6 Genetic differentiation at individual STR loci
8.7 Pairwise theta estimates for mtDNA clades
8.8 Genetic distances used in the isolation-by-distance analyses
8.8.1 STR pairwise genetic distances
8.8.2 Mitochondrial DNA pairwise distance

Glossary of Terms

(Terms in the glossary are shown in small caps the first time they appear in the main body of the text, not including the Section 1 Summary.)

Admixture – The sharing of alleles among individuals in separate populations through migration or shared ancestry.

Allele – Variants of a gene at a specific location on a chromosome. In diploid organisms, individuals can have at most two alleles per gene, one inherited from the mother and one from the father.

Assortative mating – Mating of individuals having more traits (e.g. genotypes or phenotypes) in common than is likely with random mating.

Bayesian assignment modeling – A statistical method in which individuals are probabilistically assigned to one of *k* genetic clusters within a dataset, where *k* is either specified or inferred from the data.

Bayesian skyline plot – A Bayesian inference method in which past population dynamics (e.g. increase or decrease in effective population size) are estimated through time from a sample of gene sequences.

Clade – A monophyletic group composed of a single ancestor and all its descendants.

Effective population size (N_e) – The average number of individuals in a population that actually contribute genes to succeeding generations. This number is generally lower than the census population size.

Evolutionary (phyletic) lineage – A series of species or clades arranged in sequence from ancestor to descendent, with each descendent species or clade having evolved from the one which immediately preceding it.

Genomic DNA – The full complement of DNA contained in the genome of an organism.

Gene – Each gene is a linear segment of a DNA molecule that includes a specific sequence of paired bases that are arranged on chromosomes. Each gene is responsible for a single inherited property, characteristic, or function of the organism.

Genotype – A genetic profile indicating the particular alleles present at one or more loci within an organism.

Genetic divergence – A process in which two or more populations accumulate independent genetic changes (DNA sequence mutations) through time as a result of reduced gene flow or complete reproductive isolation for some period of time. Several measures used to estimate genetic divergence in this study include: Wright's fixation index (F_{ST}), Weir & Cockerham's index Θ, R_{ST}, Nei and Chesser's index (G_{ST}).

Genetic diversity – The existing genetic variation within a population or species. Various measures are used to estimate genetic diversity such as determining the number of polymorphic sites across a specified region of DNA sequence or determining the number of heterozygous individuals in a population. Several genetic diversity measures used in this study include: Polymorphic sites (s), nucleotide diversity (π), haplotype diversity (h), heterozygosity (H_E), and allelic richness (A_r).

Haplotype – A haplotype is a series of nucleotides located on the same chromosome that are inherited together as a single unit. The term 'haplotype' is a contraction of the phrase haploid genotype; it is usually applied to individual mitochondrial DNA sequence variants, since the genome is haploid (maternally inherited).

Homozygote – An individual having two identical alleles of a particular gene.

Heterozygote – An individual having two different alleles of a particular gene.

Inbreeding – The mating of closely related individuals, which tends to increase homozygosity and therefore increases the appearance of recessive traits.

Markov Chain Monte Carlo (MCMC) – MCMC is a computational tool that is used to generate simulations from a probability distribution. It is widely used in Bayesian statistics to estimate unknown parameter values and their 95% credible intervals, which represent the degree of uncertainty in the estimate.

Migration – The dispersal of an individual from one population to another.

Mitochondrial DNA (mtDNA) – The genetic material of the mitochondria, an organelle within the cytoplasm that generates energy for the cell. Mitochondria are maternally inherited (i.e. passed from mother to offspring); therefore, only the maternal lineage is reflected in mtDNA. In contrast, most other DNA is found within the cell nucleus and is both maternally and paternally inherited.

Phylogeography – The study of the historical processes that may have been responsible for the contemporary geographic distributions of individuals and genetic lineages. This is accomplished by considering the geographic distribution of individuals in light of the patterns associated with a gene genealogy.

Polymerase chain reaction (PCR) – A technique used for amplifying DNA in a thermal cycler. The process can generate thousands to millions of copies of the original, targeted gene sequence, which can then be visualized, sized and sequenced with fluorescent markers in an automated sequencer.

Population structure – A group of individuals within a certain area that are more closely related to one another than those randomly selected from the general population. Physical barriers to migration and limited dispersal can cause genetic structuring of populations.

Phenotype – Any observable characteristic or trait of an organism (e.g. color pattern, development, behavior, etc.). Phenotypes are a result of the expression of certain genes, environmental factors, and interactions between the two.

Pliocene – A geologic time period that extended from roughly 5.3 to 2.6 million years ago.

Pleistocene – A geologic time period beginning about 1.6 million years ago and ending 10,000 years ago when glaciers covered much of the northern hemisphere (last major ice age).

Range expansion

Demographic – An increase in the effective size of a population or clade.
Geographic– An increase in the geographic area occupied by a population or clade.

Restriction enzyme – A protein that recognizes a specific, short sequence of DNA and cuts the DNA at that point. Different restriction enzymes recognize and cut at different sites along the sequence.

Short tandem repeat loci (STR) – STR's (microsatellites) are short, repeated sequences of non-coding DNA that are found throughout the genome. The repeated sequence is often simple, consisting of two, three or four nucleotides (di-, tri-, and tetranucleotide repeats respectively). Individuals can vary in the number of repeat units per microsatellite allele; this variability constitutes the data used in population genetic analyses. Microsatellites evolve rapidly because mutations (reflected by differences in the number of repeat units) typically have no effect on the organism, making them especially useful for understanding patterns of contemporary evolution (i.e. migration and gene flow).

The most recent common ancestor (TMRCA) – The most recent individual from which all individuals in a group are directly descended.

Wahlund effect – An increase in heterozygosity observed when distinct subpopulations are analyzed jointly, or when they have recently hybridized. Whenever subpopulations vary in gene frequency, the population as a whole will show a Wahlund effect.

1. Summary

1.1 Background

The Alameda whipsnake *Masticophis lateralis euryxanthus* has been the subject of conservation concern in California for over three decades. Management efforts to date have focused mainly on the snake's ecology and preserving San Francisco Bay area populations that fit the description of the *M. l. euryxanthus* phenotype (Riemer 1954); however, these efforts have reached a point where continued management would benefit from knowledge on genetic variation within and among these populations. In this study, we use short tandem repeat loci (STR) and mitochondrial DNA (mtDNA) sequences to examine the population dynamics of *M. l. euryxanthus*, and frame this work within the historical, phylogeographic context of the parent species *M. lateralis*. We use both phylogenetic and population genetic analyses to examine patterns of variation at multiple geographic scales and over different evolutionary time frames.

1.2 Major findings

- We found no evidence that populations with the *M. l. euryxanthus* phenotype form an exclusive mitochondrial clade; rather, *M. l. euryxanthus* haplotypes fall within a geographically restricted and well-supported clade that began to diversify in the late Pleistocene (~0.10– 0.21 million years ago [Ma]), and separated from members of its sister clade during a time period that coincides with the formation of the San Francisco Bay (~0.51 Ma). This same clade, which we refer to as the 'Central CA' clade, also includes populations that do not have the *M. l. euryxanthus* phenotype.
- A single haplotype predominates within the Central CA clade, and its constituent populations show evidence of a demographic and geographic range expansion that began approximately 0.25 Ma ago. The southern boundary of this clade extends substantially further south into the Diablo Range than the known distribution of *M. l. euryxanthus.*
- Levels of genetic divergence among the major mtDNA clades within *M. lateralis* in California exceed that of separate, currently recognized *Masticophis* species in the Cape Region of Baja California. Under a phylogenetic species concept, it would be possible to recognize these clades as separate species.
- A strong signal of genetic isolation-by-distance was detected in both the STR and mtDNA data, and spatially explicit Bayesian assignment models based on the STR allele frequencies unambiguously detected four main clusters within East Bay area populations. These clusters largely assign to critical habitat units in Alameda and Contra Costa counties.
- Divergence among clusters is moderate to high, and Bayesian inference of recent migration rates suggests that the sampled populations have a low proportion of immigrants.
- Populations comprising several clusters in west Alameda and Contra Costa Counties show evidence of unidirectional gene exchange; genes from one cluster, formed by centrally located populations at the Stonebrae/Bailey Ranch site, are apparently moving to neighboring populations to the immediate north and south more frequently than they are moving in.
- There is no strong evidence of recent reductions in effective population size at any of the sampling sites in the East Bay.
- Greater phylogenetic and population genetic distinctiveness of individuals sampled near the Alameda-Santa Clara County line is concordant with a loss in the distinctiveness of the *M. l. euryxanthus* phenotype over the same general area, suggesting that this variation may reflect secondary contact between historically differentiated groups.

1.3 Conclusions

Our results indicate substantial, unrecognized phyletic diversity within *M. lateralis*, with three well-supported mtDNA clades occurring in California and two occurring in Baja California, Mexico. We found no evidence that *M. l. euryxanthus* populations form an exclusive mtDNA group; however, this is not an unexpected result given that neighboring populations belonging to separate subspecies do not form reproductively isolated units, and gene flow has the potential to obscure their phylogenetic history in areas of secondary contact. Spatial patterns of mtDNA haplotypes and STR alleles indicate that the current management units adequately reflect population genetic structuring for *M. l. euryxanthus* in the Bay area, but that individuals appear to be moving out of certain parts of western Alameda County more frequently than they are moving in. We discuss how the spatial dispersion of our sampling and the geography of critical habitat units may be influencing our results, and outline a set of criteria for identifying groups of populations that might be considered high-priority candidates for future management. Our criteria reconcile information from a variety of data, and emphasize a need to not only preserve the snake's ecology, but also the microevolutionary processes that ultimately gave rise the existing genetic and phenotypic diversity within the species. We close with descriptions of future research projects that could add to a comprehensive model for the management of this species.

2. Introduction

The Alameda whipsnake (*Masticophis lateralis euryxanthus*; hereafter *Mle*) is a threatened subspecies limited primarily to coastal sage scrub and chaparral habitat in Contra Costa, and parts of Alameda and Santa Clara Counties, CA. It is one of two currently recognized subspecies of *Masticophis lateralis*, the other of which occurs throughout much of state and extends into Baja California, Mexico (*M. lateralis lateralis*; hereafter *Mll*). The subspecies was described by Riemer (1954) on the basis of color hue and pattern, most notably the orange-rufous suffusion on the anterior light-colored portions of the body (which are cream or pale yellow in other parts of the species range), increased width of the dorsolateral striping relative to *Mll*, and the amount of black spotting on the underside of the head and neck (little spotting in *Mle* and increased spotting in *Mll*). The snake is an important diurnal predator within scrub and chaparral communities, feeding mainly on lizards, small rodents, birds, frogs, salamanders, and other snakes (e. g. Stebbins 2003; Swaim 1994); however, habitat loss and fragmentation, fire suppression, livestock grazing practices, and increased predation from non-native species are likely contributing to its ongoing decline (71 FR-58175).

Because of the snake's restricted distribution and its role as a top predator within chaparral and scrub habitat, *Mle* has been the subject of conservation concern since the mid-1980s (50 FR 37958). In 1994, the US Fish and Wildlife Service published a proposed rule in the Federal Register (59 FR 5377) to list *Mle* as a threatened species, with a final listing rule published in 1997 (62 FR 64306). In 2006, presence/absence data and a series of habitat quality factors (e.g. soil type, vegetation type, vegetation mosaic, and amount of urban infrastructure) were then used to designate six critical habitat areas, totaling approximately 62,659 hectares within the known range of the subspecies (Fig. 1; 71 FR 58175). The intent of these units was to mitigate extinction by preserving high-quality habitat and facilitating movement and dispersal among populations; however, each unit is effectively isolated from the others by urban matrix, and the biological effects of this fragmentation are currently not well-understood.

Many of the issues surrounding the snake's management involve a limited understanding of how *Mle* populations are structured across the landscape and the degree to which *Mle* is

distinctive from neighboring populations of *Mll*. Although 'pure' *Mle* PHENOTYPES are quite distinctive, no quantitative, peer-reviewed studies have ever been published on the phenotypic characters used by Riemer (1954) to distinguish *Mle* from *Mll*, and the original description for *Mle* was based on only six specimens. It is also not uncommon for '*Mle* characters' to appear in whipsnakes that are far removed from Alameda and Contra Costa Counties, and populations currently recognized as *Mle* occasionally consist of individuals that do not exhibit the full suite of putatively diagnostic characters. To add to this ambiguity, a comprehensive analysis of phenotypic variation throughout the remaining parts of the species' range has never been undertaken, and the geographic boundaries of the two subspecies are nebulous. Whipsnakes with intermediate phenotypes (i.e. having character states for both *Mll* and *Mle*; Swaim 1998, unpub. data) have been recovered from areas just south and east of the San Francisco Bay near the Santa Clara/Alameda County line, suggesting that variation along the inner Coast Ranges in California may be clinal (Jennings 1983). Finally, the degree of GENE exchange, GENETIC DIVERGENCE or ADMIXTURE among *Mle* populations has never been examined, and it is unknown whether shifts in color hue and pattern at the putative subspecies boundary coincide with any historical or modern-day disruptions in gene flow.

Despite these unknowns regarding the snake's biology and taxonomy, there is a precedence to target distinctive EVOLUTIONARY LINEAGES for conservation and genetic data constitute a key piece of evidence in identifying such lineages. The rationale behind this approach is based on two factors: (1) unique, historically isolated lineages cannot be recovered if they go extinct, and (2) among-population genetic variability will be maximized by conserving the most phylogenetically divergent lineages (Lesica and Allendorf 1995; Templeton 1986). The latter point emphasizes adaptive potential; by preserving divergent lineages occupying a range of habitats, the species as a whole may be better able to adapt to changing environmental conditions because of the greater breadth of available genetic variation (Meffe and Carroll 1994). Both factors emphasize a need to conserve microevolutionary processes in the overall management of a species – it is ultimately these processes that generate phenotypic variants through time (Frankel 1974; Moritz 1999; Templeton 1986). As no formal phylogenetic or population genetic studies have ever been conducted on *Mle*, it is unknown whether this subspecies corresponds to a distinctive evolutionary lineage, or what factors may be contributing to or restricting its divergence from geographically adjacent populations of *Mll.*

In this study, we use SHORT TANDEM REPEAT LOCI (STR; also known as microsatellites) and MITOCHONDRIAL DNA (mtDNA) sequences to address gaps in the phylogeographic and population genetic knowledge of *Mle*. Our goals were to capture the deeper history of the species writ large to identify the origins and patterns of diversification of central coast populations, as well as to characterize the more recent history of whipsnakes in Alameda and Contra Costa Counties. In particular, we aimed to assess whether the current management units adequately reflect the distribution of genetic diversity within *Mle* populations east of the San Francisco Bay (referred to hereafter as 'East Bay'). To meet these goals, we structured our field sampling and data analyses at multiple geographic scales, beginning with a range-wide assessment of the PHYLOGEOGRAPHY of the species, to a study of within-CLADE population structuring across the northern extension of the inner Coast Ranges in central California, and finally to a series of analyses restricted to populations occurring within the recovery units in the East Bay. We consider this work as a first step toward comprehensively assessing *Mle* population genetics, and use our results as a guide for developing further studies that can be used in conjunction with other forms of data (i.e. morphological and ecological) to develop future management plans.

3. Material and Methods

3.1. Tissue sample acquisition

We obtained tissue samples from monitoring surveys conducted by Swaim Biological Inc. (Livermore, CA), museum collections, professional colleagues and our own field efforts. Samples from Swaim Biological Inc. and our own collecting were obtained using non-lethal techniques (funnel traps and by hand) and consisted of small tail clips preserved in 95% ethanol. Tissues from established *Mle* populations were collected by Swaim Biological Inc. under the authority of Federal Recovery Permit 81537 and a Memorandum of Understanding with the California Department of Fish and Game. The general funnel trap design and use are provided in Swaim (1994). We also harvested tissue from road-kill specimens whenever we encountered them, as road mortality is prevalent in areas where roads bisect preferred habitat. We extracted GENOMIC DNA using standard proteinase K extractions or a Qiagen DNeasy Blood & Tissue Kit.

We used both STR ALLELE frequencies and mtDNA sequence data to study *Mle* population structuring for three reasons. First, the *Mle* distribution is highly restricted and populations were substantially more interconnected in the recent past; thus we assumed that rapidly evolving markers would be needed to resolve any existing POPULATION STRUCTURE. Second, given that the spatial patterning of mtDNA versus STR alleles are influenced in different ways by demography, individual movement tendencies, modes of inheritance[1], and different rates of evolution, we were more likely to understand the interplay between historical abiotic factors and gene flow in shaping contemporary patterns of GENETIC DIVERSITY if we explored data from both the nuclear and mitochondrial genomes. This is because mtDNA are often better at providing information on the deeper history of the species, whereas STR allele frequencies tend to provide more information about recent history (i.e. contemporary MIGRATION and gene flow). Finally, if separate types of data produced consistent results, we could be more confident in our ability to draw reliable conclusions.

For the STR analysis, we focused our sampling on populations spanning the six critical habitat units in the East Bay (Fig. 1). Whipsnakes are difficult to capture by hand because of their wariness, aggressive biting behavior, speed, and agility. They are also active during warm parts of the day, which further exaggerates these behaviors; thus, we relied mainly on trapline surveys conducted by Swaim Biological Inc. to obtain adequate sample sizes for statistical analyses. This resulted in a relatively uniform sampling throughout Alameda and Contra Costa Counties, and allowed us to sample up to 30 individuals per locality (although field sampling did not meet this threshold for all sites). Our targeted sample size was based on three main factors: first, 30+ individuals would allow for robust statistical comparisons among populations and recovery units; second, it would provide for more accurate estimates of genetic diversity and gene flow indices; and third, it would ensure that our results were robust to the effects of any possible allele scoring errors. In total, we obtained STR GENOTYPES from 12 locations, nine of which are well within the range of *Mle* (Swaim 1994; USFWS 2006), two of which are in possible subspecies transition zones,

Table 1. STR genetic diversity estimates for sampling sites in the East Bay. Notations are as follows: the mean sample size *n* over all loci; critical habitat unit label *U*; the proportion of polymorphic loci *P*; the mean no. of observed polymorphic alleles per locus nA_p; the expected heterozygosity H_E; the observed heterozygosity H_O; and an estimate of the fixation index *f*.

Population	*n*	*U*	*P*	nA_p	H_E	H_O	*f*
Los Vaqueros	32.0	4	0.88	4.86	0.44	0.46	0.01
Mt. Diablo	31.8	4	0.88	5.14	0.46	0.43	0.18
Black Diamond	3.0	4	0.63	2.40	0.37	0.40	-0.10
Alhambra	10.8	1	0.75	2.92	0.38	0.32	0.19
Allen Property	35.8	1	0.88	4.07	0.46	0.46	-0.04
Tilden Park	4.0	1	0.94	2.58	0.40	0.33	0.15
Claremont	5.0	6	0.88	2.92	0.42	0.45	0.06
Stonebrae	32.0	3	0.94	4.87	0.47	0.40	0.17
Bailey Ranch	31.0	3	0.88	4.93	0.41	0.34	0.15
Ohlone	25.9	5	0.94	5.73	0.49	0.45	0.08
Del Valle	3.0	5b	0.81	2.46	0.42	0.39	0.10
S300	27.0	5a	0.81	4.08	0.36	0.31	-0.09

[1] In most multicellular organisms, mtDNA is maternally inherited.

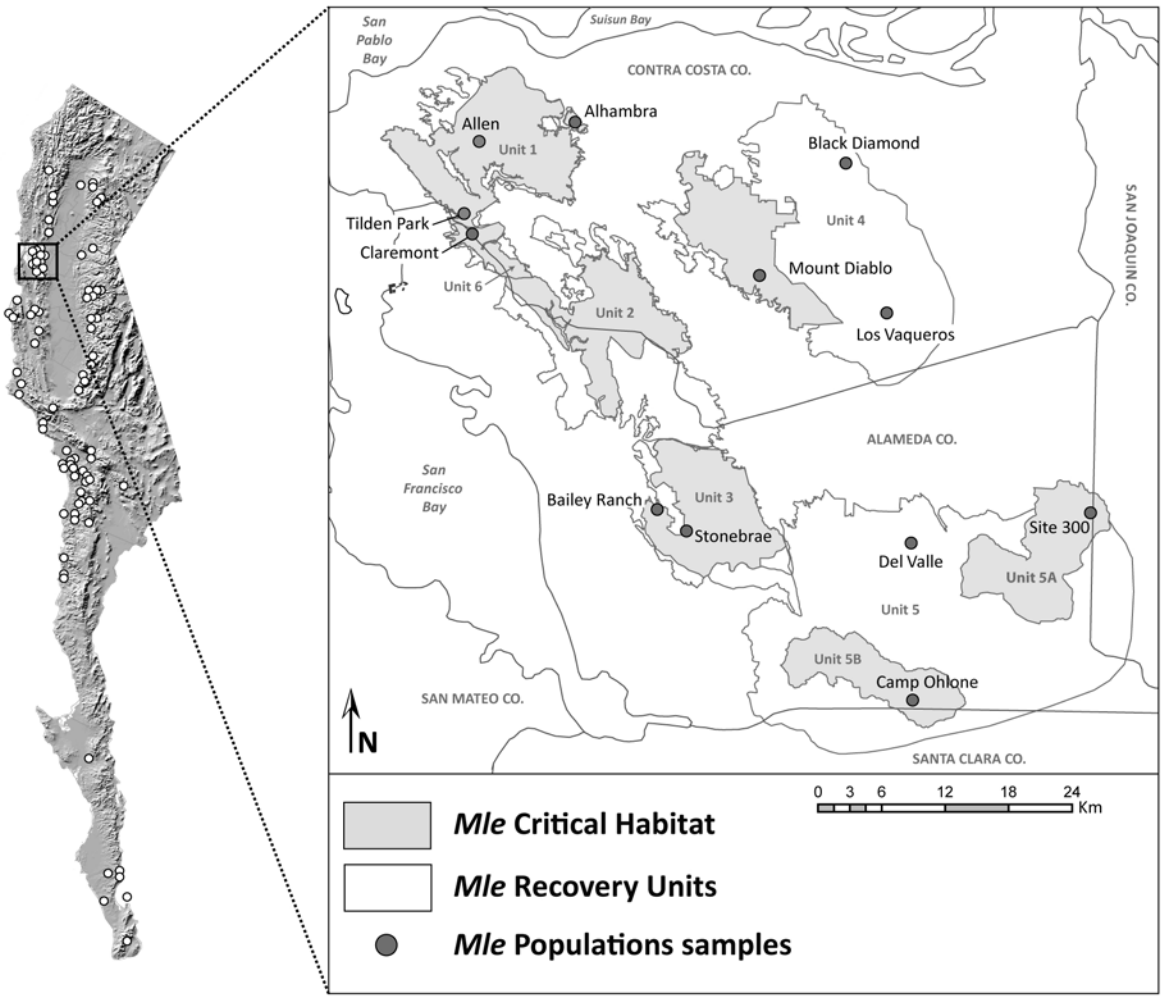


Figure 1. Map of critical habitat and East Bay recovery units for *M. l. euryxanthus*. White bullets on the map of California indicate sampling locations for the mtDNA phylogenetic analysis. Locality data and individual identifiers are provided in Appendix 8.1 and 8.2, respectively.

and one of which is *Mll*. Sample site labels used here are derived from their project names (Swaim Biological Inc.) or study locations (e.g. regional, state, or watershed names, and residential developments). Table 1 provides a summary of the sample sizes for each location.

For the mtDNA analysis, we sequenced 3-5 individuals from each population in which we obtained STR data (in most cases, the same individuals that were genotyped), and from single individuals throughout the geographic range of the species. It was possible to sample fewer individuals per location for mtDNA because within-population variation is limited; most individuals from a single locality have either identical mtDNA HAPLOTYPES, or the haplotypes differ by only one or two substitutions; thus, single individuals are often representative of the mtDNA diversity for a given population. Several *Masticophis* species were used as outgroup taxa, including *M. aurigulus, M. barbouri, M. taeniatus, M. flagellum, M. coluber, M. mentovarious, M. bilineatus,* and *M. schottii*. Two species of *Salvadora* were used to root the tree (Nagy et al. 2004). Locality information and sample sizes for the individuals used in the mtDNA analyses is provided in Appendix 8.1, and a map with mtDNA sampling points and specimen identification is provided in Appendix 8.2.

3.2. mtDNA sequence data collection

For each individual, we amplified the mitochondrial NADH dehydrogenase 4 (ND4) protein coding gene and three flanking tRNAs (tRNA-*his*, tRNA-*ser*, and tRNA-*leu*). This region of the mitochondrial genome is commonly used in reptile phylogeographic studies and is informative at spatial and temporal scales that are relevant to this study (Feldman and Spicer 2006; Mitrovich 2006; Richmond and Reeder 2002; Spinks and Shaffer 2005). We used roughly 50−100 ng of total DNA as template for POLYMERASE CHAIN REACTION (PCR) in a final volume of 12.5 µl containing 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 0.6 µM of each primer, 10X PCR buffer (Qiagen), 5X Q-solution (Qiagen), and 1 unit of DNA TAQ Polymerase. Forward and reverse primers for gene amplification, respectively, consisted of a primer we developed specifically for *M. lateralis*, ND4ML-F (GCAACGACTTTCTAAATAACCTAA) and LEU (Arevalo et al. 1994). PCR cycling conditions consisted of 95°C for 5 min, 35 cycles of 94°C for 45 sec, 52°C for 45 sec, 72°C for 75 sec, followed by a 72°C final extension for 7 min. We purified PCR products using an UltraClean™ PCR Clean-up Kit (Mo BIO Laboratories, Inc.) and directly sequenced on an ABI 3100S automated capillary system. We edited the sequences using Sequencher™ 4.6 and aligned the nucleotide sites by eye.

3.3. STR data collection

We used Genetic Identification Services (Chatsworth, California, USA) to develop STR libraries that were specific to *Mle*. Methods for library construction, enrichment and screening followed Jones *et al*. 2002. Briefly, genomic DNA from a single *Mle* individual was partially restricted with seven blunt-end RESTRICTION ENZYMES. Fragments between 350-700 nucleotides were ligated to double-stranded adaptor sequences, allowing for enrichment of four STR library repeat motifs (CA, AAC, ATG, and TAGA). Positive clones were sequenced using ABI PRISM *Taq* dye-terminator cycle sequencing and an ABI377 automated sequencer (Applied Biosystems, Foster City, CA).

Eighty-six positive clones were recovered from the four libraries. Of these 86, 12 contained di-repeat STRs, 26 contained tri-repeats, and 16 contained tetra-repeats. After screening, we selected 16 primer sets to assess genetic diversity and structure. We optimized a multiplex PCR protocol in which each reaction contained three to four primer sets corresponding to the 16 STR loci. Alleles were amplified in 10 µl reactions using a Qiagen Multiplex PCR kit

with HotStarTaq DNA Polymerase and 50-100 ng of DNA template. Cycling conditions were standardized over all loci as follows: 95°C hot start for 15 min, five cycles of 95°C 30 s, 60°C 30s, 72°C 90 s, five cycles of 95°C 30 s, 58°C 30s, 72°C 90 s, and 20 cycles of 95°C 30 s, 56°C 30s, 72°C 90 s, followed by a 72°C final extension for 10 min. PCR products were subjected to Sephadex® (Sigma-Aldrich, St. Louis, MO) purification to remove unincorporated dyes and primer prior to genotyping. Genotyping runs were performed on an ABI 3100S Automated Sequencer (Applied Biosystems, Foster City, CA) using the LIZ internal size standard (LIZ 600bp, Applied Biosystems). Allelic sizes (in nucleotides) were determined by reference to the internal sizing standard in the software GENEMARKER V1.85 (Softgenetics LLC, State College, PA). Descriptions of STR repeat motifs, multiplex primer sets, and primer sequences are provided in Appendix 3.

We used MICRO-CHECKER v2.2.3 to test for possible allele scoring errors due to stuttering, null alleles, or large allele dropout (Van Oösterhout et al. 2004). The program uses a Monte Carlo simulation method to generate expected HOMOZYGOTE and HETEROZYGOTE allele size difference frequencies. The Hardy-Weinberg theory of equilibrium is used to calculate expected allele frequencies and the frequency of any null alleles detected. Loci showing evidence for any of the three scoring errors in three or more populations were considered suspect. We also verified our results in MICRO-CHECKER by testing whether each locus was in Hardy-Weinberg proportions within populations using FSTAT v2.9.3 (Goudet 2001).

Rather than discarding all data from loci assumed to have null alleles, we used a method by Brookfield (1996) to estimate the null allele frequency (r) and adjusted the genotypes at the affected loci accordingly. This method assumes that heterozygote deficiencies are the result of a true null allele and not caused by WAHLUND EFFECTS, INBREEDING, ASSORTATIVE MATING, etc., and is appropriate for cases such as ours in which one or two bands amplified for all individuals. We used the adjusted frequencies in our statistical analyses to more closely approximate the 'true' values (Brookfield 1996), and verified that the adjusted values did not produce significant differences in the estimates genetic differentiation as inferred by F_{ST} (see methods for comparing different pairwise F_{ST} values below). We prefer this approach to eliminating all data from suspect loci because at least some useful information is contained within those markers.

3.4. Organization of phylogeographic and population-level analyses

From this section forward, we structured our analyses according to geographic scale, beginning at the largest scale and working our way toward successively smaller scales. At the largest scale, we used mtDNA to analyze the phylogeographic structure for the species as a whole, allowing us to place *Mle* within the historical context of populations throughout the entire species range. Next, we focused on a smaller area by analyzing mtDNA data from a geographically restricted clade occurring along the inner coast ranges of central California (the Diablo and Gabilan Ranges), which contains all known populations of *Mle*. Finally, at the smallest scale we examined *Mle* population structuring and gene flow within the East Bay recovery units, using both mtDNA and STR data.

3.5. Phylogeography, genetic diversity and historical demography based on mtDNA

3.5.1. Range-wide phylogeographic assessment of M. lateralis

We used MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) to estimate the phylogenetic tree based on the mtDNA sequence data. This software uses a MARKOV CHAIN MONTE CARLO (MCMC) algorithm to approximate the posterior distribution of phylogenetic trees and model parameter estimates, and produces a summary tree with probability estimates for all inferred relationships, based on the sequence data and a specified model of nucleotide evolution. We

modeled nucleotide evolution by assigning separate substitution models to five data partitions, one for each tRNA gene and one for each codon position in the ND4 gene, and identified the best-fit models using the Akaike Information Criterion in jModelTest 0.1.1 (Posada, *in press*). Codon positions were modeled independently because of well-known differences in the rate of evolution among 1st, 2nd, and 3rd positions. The best-fit models are listed in Table 2. All MCMC simulations were run for 20 million generations using the standard convention of three heated Markov chains and one cold chain. Samples were retained every 1,000th generation to reduce autocorrelation among successive draws from the posterior during MCMC sampling. If the log likelihood scores for separate Markov chains converged on similar values and yielded a split standard deviation <0.005, reflecting the fact that the samples from different chains have become increasingly similar, we considered those parameters to have reached stationarity.

We used Tracer 1.5 (Rambaut and Drummond 2005) to assess convergence by examining plots of all parameter values against numbers of iterations of the Markov chains, and to determine whether the MCMC algorithm had achieved reasonable effective sample sizes (i.e. >200). We also assessed convergence by visually inspecting the cumulative posterior probabilities of clades (or splits) using the online program AWTY (Nylander et al. 2008). Branches with posterior probabilities (*Pp*) ≥ 0.95 were considered strongly supported.

Table 2. Character data and best-fit nucleotide substitution models for the mtDNA data partitions.

Data Partition	No. Characters	Variable Characters	Parsimony Informative Characters	Percent Parsimony Informative	Substitution Model
ND4 – 1st position	440	38	19	4.3	HKY+I
ND4 – 2nd position	440	16	7	1.6	HKY
ND4 – 3rd position	440	164	115	26.1	HKY+I+Γ
tRNA*his*	67	5	3	4.5	HKY+I
tRNA*ser*	58	5	1	1.7	K80
tRNA*leu*	54	0	0	0	none

3.5.2. Genetic diversity and differentiation: mtDNA

Genetic diversity – To measure genetic diversity using mtDNA, we calculated and compared the number of polymorphic sites (*s*), nucleotide diversity (π) and haplotype diversity (*h*) among each of the major clades that were recovered in our estimate of the phylogeny (section 4.2.1 provides a detailed description of these lineages). These diversity indices were measured using DNASP v5.0 (Librado and Rozas 2009). To examine the proportion of mtDNA variation explained by different hierarchical groupings of the data, we performed an analysis of molecular variance (AMOVA: Excoffier et al. 1992a) in ARLEQUIN v3.11 (Excoffier et al. 2005). Variation was assessed at three levels in Alameda and Contra Costa County: (a) among recovery units; (b) among populations within recovery units; and (c) within populations. We also examined the partitioning of variation among sampling sites within the full, regional mtDNA clade that contains the known populations of *Mle*.

Genetic differentiation – To estimate levels of differentiation, we calculated the percent mtDNA sequence divergence among all pairwise combinations of the different regional mtDNA clades within *M. lateralis*. We performed these calculations using a best-fit nucleotide substitution model; the model selection process and the different preferred models follow the descriptions in section 3.5. above. We also calculated pairwise Φ_{ST} values, an analog of Wright's fixation index F_{ST} (described in section 3.8.2 below), and used these values to conduct tests of genetic isolation-by-distance in section 4.5.

3.5.3. Estimates of divergence times

To estimate the TIME TO THE MOST RECENT COMMON ANCESTOR (TMRCA) of the major mtDNA lineages within *M. lateralis*, we used the Bayesian relaxed clock method implemented in BEAST v1.5.4 (Drummond et al. 2006; Drummond and Rambaut 2007). Like MrBayes, this program uses an MCMC algorithm to simultaneously estimate tree topology, model parameters, and the timing of divergence events. Estimates of divergence times are based on calibrations from either the fossil record, historical biogeographic events, or known mutation rates. These analyses were restricted to unique *M. lateralis* haplotypes only and several outgroup taxa (see below), and we used the same MCMC sampling procedures as described above. Analyses were conducted using a constant population tree prior, with evolutionary rates across the tree branches assumed to be uncorrelated and log-normally distributed.

Estimating absolute ages from molecular data alone is not possible unless mutation rates for the specific gene regions of interest are known – this is rarely the case; however, one alternative is to calibrate the clock using the fossil record; if good fossils with reliable ages are available, estimating divergence times based on fossil calibrations may be more reliable because of the simple fact that they incorporate information and uncertainty of the known history of the study organism (Burbrink et al. 2008; Ho and Phillips 2009; Lee 1999; Sanders and Lee 2007). Accuracy can be further improved if multiple calibrations spanning different parts of the evolutionary tree are incorporated in to the same analysis. Here, we calibrated the molecular clock using fossil dates for two temporally spaced nodes in the tree. These calibrations are derived from the North American Miocene snake fossil record (Holman 2000) and have been used to estimate divergence times in a previous phylogenetic study involving many of the same species included here (Burbrink et al. 2008). We used the following calibration references according to this previous work: (1) a mean calibration of 19 Ma and a log normal standard deviation of 0.2 was used to root the tree and corresponds to the earliest known fossils of *Salvadora* and *Paracoluber*, the earliest extant representatives of the monophyletic New World Colubrinae, (2) a mean calibration reference of 11 Ma with a lognormal standard deviation of 0.1 was placed at the most recent common ancestor of all *Masticophis* species and corresponds to the earliest fossil appearance of the genus (early to middle Miocene).

The log-normal standard deviation yields a skewed distribution that some have argued is a good representation of the fossil record (Ho 2007; Sanders and Lee 2007); there is a 'hard' minimum bound, meaning that there is very low probability of dates much younger than the oldest fossil, and a 'soft' maximum bound meaning an infinitely long tail of increasingly unlikely older dates. For comparison, we also performed an analysis using a normally distributed standard deviation, which allows divergence estimates to vary symmetrically with soft bounds. Both approaches are useful for incorporating uncertainty into the age estimation process.

3.5.4. Inferences of recent and historical changes in demography

Inferences of demographic change over different temporal scales can provide useful information for evaluating the conservation status of populations. For example, reductions in EFFECTIVE POPULATION SIZE (N_e) and genetic variability in threatened or endangered species typically signal poor survival prospects because they can increase the rate of fixation of mildly deleterious alleles and rates of inbreeding, and therefore reduce adaptive potential and increase the probability of extinction (Cornuet and Luikart 1996; Hedrick and Miller 1992; Lande 1994).

We tested for changes in the demographic history of *M. lateralis* populations at two different time and geographic scales, one for the deeper history of the species over a broad geographic scale using the mtDNA sequence data, and the other for the more recent history of East Bay populations using the STR data.

Historical demography of M. lateralis based on mtDNA data – Because the demographic history of a population leaves a signature in the DNA of its modern representatives, we were able to perform a series of complementary analyses to infer historical changes in effective population size. Site mismatch distributions are a widely used method for this purpose; distributions of the observed number of pairwise differences between haplotypes are compared to theoretical distributions that describe different demographic models, and the fit of the observed data to the theoretical predictions are tested using one of several statistical indices. The distribution is multimodal in samples drawn from populations at demographic equilibrium, whereas it is smoother and unimodal in populations that have passed through DEMOGRAPHIC EXPANSION (Ray et al. 2003; Rogers and Harpending 1992). Here, we compared the number of pairwise nucleotide site differences against a null distribution of constant population size using the R_2 statistic (Ramos-Onsins and Rozas 2002) and Harpending's raggedness index r_g (Harpending 1994), and tested for significant deviations from zero using 10,000 coalescent simulations in DNASP v5.0.

As an additional method of testing for deviations from demographic equilibrium, we also calculated Tajima's *D* (Tajima 1989) and Fu & Li's *D* (Fu and Li 1993). Both statistics are expected to be zero if lineages have been stable over time, negative if lineages have expanded or are under purifying selection, and positive if lineages have declined or under balancing selection. Significant deviation from the null expectation was assessed using the same coalescent simulations as above for R_2 and r_g.

To augment the above analyses using a genealogical approach (i.e. one that incorporates information on the evolutionary tree), we used BAYESIAN SKYLINE PLOTS (BSP) to test for demographic expansion at different time intervals throughout the evolutionary history of *M. lateralis* (Drummond et al. 2005). This method does not require the a priori specification of a demographic model, and instead can be used as a demographic model selection tool itself (Pybus and Rambaut 2002). The Bayesian skyline plot model uses MCMC sampling to estimate a posterior distribution for effective population size (N_e) through 'slices of time', from the tree tips back to the most recent common ancestor of the gene sequences, based on the gene sequences, tree topology, and a specified nucleotide-substitution model. Plots include credibility intervals for the estimated N_e, which provide an indication of both phylogenetic and coalescent uncertainty. For each mtDNA clade in California, we constructed BSPs using the HKY substitution model and an uncorrelated lognormal relaxed clock in BEAST. We used uniform prior distributions for all demographic and nucleotide substitution parameters, and allowed evolutionary rates to vary independently across the tree branches according to a lognormal distribution. We specified 10 grouped coalescent intervals (*m*) for the Northern CA and Southern CA lineages and 7 grouped intervals (*m*) for the Central CA lineage (see results for a description of these inferred lineages). We estimated the posterior distributions of model parameters and genealogies by sampling Markov chains every 100th generation for a total of 20 million generations. The first 20,000 samples were discarded as burn-in.

Last, we used Nested Clade Phylogeographical Analysis (Templeton 1995; Templeton 2004) to test for historical GEOGRAPHIC RANGE EXPANSION within the clade containing *Mle* haplotypes. To implement NCPA, we first estimated an mtDNA haplotype network in the software TCS v1.21 (Clement et al. 2000). TCS estimates the maximum number of mutational steps among haplotypes as a result of single substitutions with a 95% statistical confidence (described in Templeton et al. 1992) and reconstructs a network based on the most parsimonious explanation of the data. The recovered network was input into GEODIS v2.6 (Posada et al. 2000) together with geographic sampling information, and the NCPA method was used to infer population-level processes that best explain the patterns in the observed data. If a significant

departure from simulated randomness was observed in the distance measures, then the null hypothesis of panmixia could be rejected. We then used the results to make population structure inferences following the most recent inference key included with the GEODIS software (geodisKey_28Apr09).

Recent criticisms of NCPA have focused on a potential bias towards false positives (in particular, inferring isolation by distance when there has not been any), reliance on single locus data, and alleged lack of statistical rigor (Knowles 2008; Panchal and Beaumont 2007; Petit 2008). Templeton has provided myriad counter-arguments to these criticisms (Templeton 2008; Templeton 2010), and the method has been validated by positive controls spanning a wide range of species, geographic scales, and sampling designs. Thus, we are of the opinion (as well as others, e.g. Garrick et al. 2009) that NCPA is useful in conjunction with other approaches; we do not rely on the results from any one methodology or analysis, and carefully consider the limitations of a method before drawing conclusions.

Tests for recent changes in demography based on STR alleles – Fragmented populations that suffer reductions in effective size are important to identify for conservation because they are the most susceptible to extinction. If a population has experienced a recent bottleneck, it generally develops a heterozygosity excess at selectively neutral loci (i.e. the observed heterozygosity calculated from a sample of genes is larger than the expected heterozygosity based on the number of alleles found in the sample if the population were at mutation-drift equilibrium; Cornuet and Luikart 1996). This excess persists only for a certain number of generations before a new equilibrium is established, thus the detection of recent bottlenecks can be sensitive to the timing of specimen collection.

We tested for recent changes in effective population size using the software BOTTLENECK 1.2.02 (Cornuet and Luikart 1996). For each population and locus, BOTTLENECK computes the distribution of the heterozygosity expected from the observed number of alleles (k), given the sample size (n) under the assumption of mutation-drift equilibrium. The distribution is generated by simulating the coalescent process of n genes under the two-phased mutation, stepwise mutation, and infinite alleles models. We simulated distributions using three different models because we wanted to ensure that our conclusions were robust with respect to the assumptions of each model. We then used a Wilcoxon sign-rank test to compare the results of the simulations to the observed heterozygosity to determine whether the proportion of loci with a heterozygosity excess is significantly larger than expected at equilibrium (in a population at mutation-drift equilibrium, the probability that a locus shows a heterozygosity excess or deficiency is approximately equal). The Wilcoxon sign-rank test provides relatively good statistical power and can be used with as few as four polymorphic loci and any number of individuals (15-40 individuals and 10-15 polymorphic loci is recommend to achieve high power).

3.6. Population genetic analysis of East Bay whipsnakes using STR allele frequencies

The next series of analyses focused specifically on the genetic structuring and interconnectedness of East Bay whipsnakes occurring within four critical habitat areas (Fig. 1). The data for these analyses consists exclusively of STR allele frequencies from 12 collecting sites within these habitat units, which cover the known range of *Mle*. Our primary goals were to (1) provide general estimates of genetic diversity at each STR locus and within habitat units, (2) to estimate the number of contemporary genetic groups that exist within this area, (3) to provide inferences on the extent of genetic connectivity among populations, and (4) to estimate how genetic diversity is partitioned among contemporaneous populations.

3.6.1. A brief background on Bayesian assignment methodology

BAYESIAN ASSIGNMENT MODELS are powerful tools for estimating the number of naturally occurring genetic clusters within a dataset. The number of programs that implement these models has grown in recent years, with newer versions offering ever-expanded capabilities for handling complex scenarios. It is up to the user to choose which software best meets the needs of a particular research question – for this work, we used the popular and well-vetted STRUCTURE v2.3 (Pritchard et al. 2000) and a more recently developed software called TESS v2.3.1 (Chen et al. 2007) that incorporates spatial data in its overall assessment of population structure.

Although the programs have similar analytical goals, each has strengths that appeal to different aspects of this study. Both are effective at identifying clusters, clines, and genetic isolation by geographic distance (or 'IBD'), all of which can simultaneously affect the spatial genetic structuring of populations. We use the term 'cluster' to refer to a natural genetic group that forms when populations diverge in response to physical or behavioral barriers that interrupt gene flow. Clines reflect spatial transitions in allele frequencies that form when divergent populations exchange genes in secondary contact (Barton and Hewitt 1985) or by adaptive divergence across selective gradients (Berry and Kreitman 1993). Last, IBD patterns develop when local genetic differences accumulate as a consequence of restricted dispersal, resulting in decreased genetic similarity with distance (Malécot 1948; Slatkin 1993). The three phenomena are not mutually exclusive, and one of the many virtues of these programs is their ability to detect clines and clusters at the level of the genome in a single analysis.

To this end, TESS and STRUCTURE implement a variety of models that fall under two main categories, those with genetic admixture and those without. No-admixture models assume that the data are comprised of some unknown number of clusters K, with each cluster defined by a set of allele frequencies at each locus. Individuals are probabilistically assigned to a cluster in a manner that minimizes departure from Hardy-Weinberg and linkage equilibrium, and the programs report the posterior probability that an individual i belongs to a particular cluster K (with prior probability $1/K$). In admixture models, individuals are assumed to have composite genomes resulting from mixed ancestry. The output consists of the posterior mean estimates of these proportions, called admixture coefficients, and are denoted in a matrix (i.e. the Q matrix). Each element of the matrix, q_{ik}, is the proportion of individual i's genome that originated from cluster k.

Both programs implement similar MCMC techniques to estimate posterior probabilities for model parameters (e. g. allele frequencies, latent clusters for each individual or allele, and admixture proportions), but they differ in one key respect; TESS incorporates an individual's geographic coordinates, be it collecting locations, birth sites, etc., as prior information for estimating cluster membership (Durand et al. 2009; François et al. 2006). The model assumes that nearest neighbor individuals are more likely to exchange genes than distant neighbors, a reasonable assumption for organisms with limited dispersal, and uses a neighborhood network in the form of a Voronoi tessellation as a weighting scheme for assigning cluster membership (Fig 2). The network is built on the GPS coordinates of individuals, and is composed of cells that are centered on a sampled individual (neighboring cells share a common edge). In the no-admixture model in TESS, the prior distribution on the set of cluster configurations incorporates an interaction parameter ψ that influences the tendency of individuals to group together. When $\psi = 0$, the model assumes a non-informative spatial prior that corresponds to the no-admixture, uncorrelated allele frequency model in STRUCTURE. Values in the range ψ (0.5, 1.0) are considered typical for K_{MAX} = 2-10, and the tendency for individuals to cluster together strengthens as ψ approaches 1.0 (François et al. 2006). It is up to the user to explore different ψ

values to assess their influence on the results, although TESS V2.3.1 provides an option to update ψ during the course of an MCMC simulation.

The admixture model in TESS extends this algorithm by incorporating geographic covariates in the prior distributions of the admixture coefficients. The coefficients are modeled as a 'hidden regression' consisting of two components, a trend surface plus a spatially autocorrelated residual term (Cressie 1993). In the statistical parlance, the autocorrelated residual term is a conditional auto-regressive (CAR) model, which determines the magnitude of the spatial neighborhood effects, and is widely applied in ecological studies to study species richness (Lichstein et al. 2002). The two components are intended to capture broad and fine-scale spatial effects, respectively – trend surfaces account for clines in any ordinal direction, and autocorrelated residuals account for IBD (Bocquet-Appel and Sokal 1989; Durand et al. 2009). Models can be fitted with or without these terms and assessed for 'strength-of-effect' by examining whether the 95% credible intervals of the estimated coefficients include zero (which indicates that *m* is not distinguishable from zero).

To fit different models, TESS uses a statistical criterion called the Deviance Information Criterion (DIC), which is computed during MCMC runs as the average model deviance plus a penalty term, p_D (Spiegelhalter et al. 2002); p_D accounts for the effective number of parameters in the model, and approximates the gain in fit by increasing model complexity. The fitting process involves the piecemeal addition of different parameters to a given model, and assessing the change in DIC. Models that receive the most support are those with the lowest DIC scores. More specifically, the smallest DIC indicates the model that would best predict a replicate dataset with the same structure as the observed data.

Because *Mle* are terrestrial animals with limited dispersal and are restricted by habitat type, incorporating the spatial covariates into our analyses (or at least examining their effects) were likely to increase the accuracy of our results. Thus, for assignment modeling we focused our efforts on TESS. As the various spatial models in TESS can be pared down to a typical non-spatial model implemented in STRUCTURE, we used STRUCTURE mainly to check for consistency in our results across platforms, comparing the least complex models first. Similar results using non-spatial models would ensure that we were not promulgating errors from 'the ground up', as we compared model fit by progressing from simpler to more complex models that included the spatial covariates mentioned above. In addition, by beginning with non-spatial models we could better evaluate how spatial priors were influencing our results, if at all.

3.6.2. Tuning of spatial priors Before estimating the number of clusters and individual admixture coefficients, we utilized features of the TESS software that allowed us to account for physical barriers that likely influence whipsnake population structure. One feature enables the user to incorporate 'dummy points' into the neighborhood network – these points represent sites where individuals cannot be sampled, such as urban areas, bodies of water, high mountains, etc (Fig 2). Their purpose is to disrupt the neighborhood network such that formerly connected cells within the Voronoi tessellation no longer share edges, and therefore individuals within those cells have less of a tendency to cluster (in the same way that a physical barrier can restrict gene exchange). The dummies are treated as extra individuals in the spatial network, but the genetic data are coded as missing. We included 10 dummy points in major urban corridors that bisect the East Bay Recovery Units along Interstate 580 (east/west) and 680 (north/south), as whipsnakes are highly unlikely to survive or traverse these urban areas (Fig. 1). Inclusion of only 10 dummy points was conservative - there are almost certainly more sites within the study area where snakes do not occur; however, we used the fewest number of dummy points that would allow us

to modify the spatial network in a biologically meaningful way without introducing potential artifacts in to the analysis.

As a further adjustment to the spatial data, we adjusted the *X*/*Y* coordinates for sampling locations to reflect more of the available area that snakes likely occupy within a given study site, and to account for a lack of information on the birth site of individuals (TESS assumes that the spatial coordinates represent the birth sites, which are rarely known for wild populations). To address these issues, we used the prescribed range of coordinates from pitfall trap lines to generate a random pair of lat/long coordinates for each individual within a sampling site. Using the minimum and maximum values for each lat/long, plus a standard deviation, we randomly sampled points from a normal distribution to recreate the spatial dataset. The standard deviation (SD) was used to manipulate the spread of the sampled points; we used an SD of 0.01, which produced a set of coordinates that were well within the home range size for these snakes and a simulated neighborhood network that largely mirrored the geography of the original sampling (Fig. 2 vs. Fig. 1, respectively). Arguably, we could have sampled random points from a broader area, as the available habitat encompassing trap lines was substantial; however, we again opted for a conservative approach that accounted for biological realism without introducing noise in to the analysis.

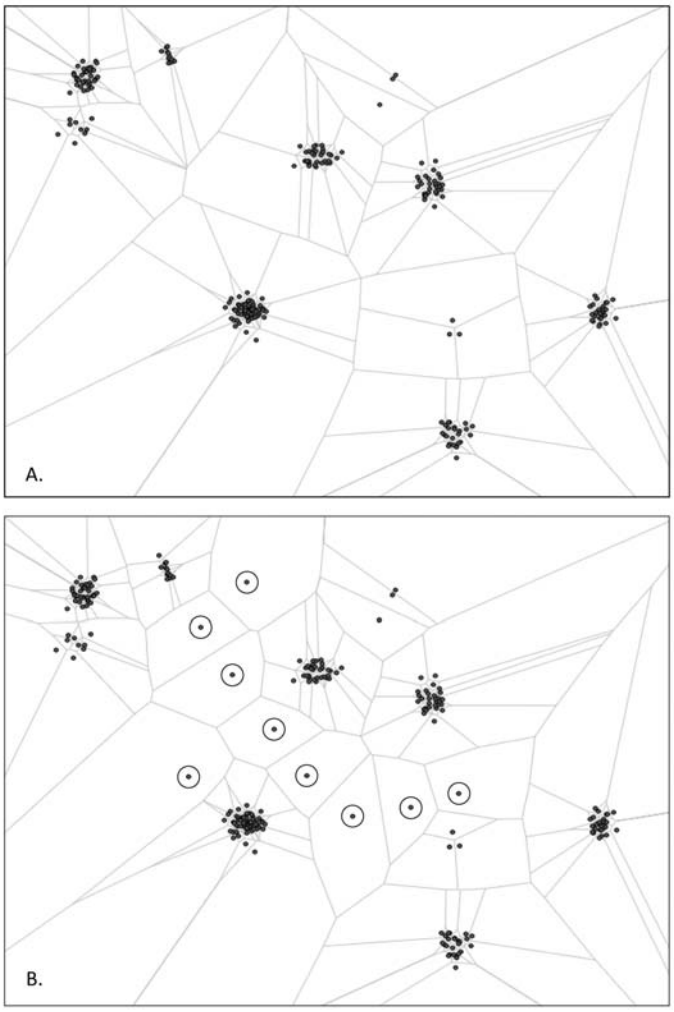


Figure 2. Examples of Voronoi tesselation plots with (A) and without (B) dummy points. Red bullets indicate actual *Mle* sampling sites, whereas dummy points (circled) mimic physical barriers where individuals cannot be sampled (e.g. urban areas, waterways, high mountains, etc.). Polygons surrounding the sampling sites are 'Dirichlet tiles'; the entire configuration forms the neighborhood network, and two sites are considered neighbors if their tiles share a common edge.

It is important to note that these modifications affect only the spatial priors and not the genetic data itself. If the main signal in the full dataset (genetic + spatial data) is derived from the STR allele frequencies, tweaking the spatial priors in this way should have minimal effect on the outcome of the analyses. If minor adjustments on the priors do cause drastic changes in the results, it is likely that the information content of the allelic data is low. One of the major strengths of the Bayesian approach implemented in TESS is the ability to evaluate the affects of spatial priors in a relatively straightforward manner so that the user can determine their magnitude of effect on the results. Ideally, one hopes to obtain parameter estimates that are robust with the respect to the choice of the prior.

3.6.3. Inferring **K** *and cluster assignment*

Using the spatial network and the multilocus genotypes, we then inferred the number of clusters among the 12 main sampling sites (Fig. 1). We ran models with and without admixture, and checked our results for consistency in both TESS and STRUCTURE. To estimate K in TESS, we performed 25 separate runs for K_{MAX} 2 – 10 (50,000 steps; burn-in 40000) and calculated the DIC for each run (Durand et al. 2009). Due to stochastic variance among replicate MCMC simulations, we averaged the DIC values across the 10 best scoring runs and used this value as the representative DIC for a given K_{MAX}. We then plotted the average DIC against K_{MAX}, and selected the value of K_{MAX} that corresponded to the inflection of the DIC curve. This approach is based on the fact that a lower DIC score can always be achieved by increasing the number of partitions in the data, up until each individual is treated as a single cluster K; however, our goal is to find the lowest K_{MAX} that best explains the data to avoid over-fitting the model and exaggerating the amount of structure that is present in the data. We used the same approach for selecting K_{MAX} in STRUCTURE; however STRUCTURE uses $\ln P(D|K)$, the logarithm of the probability of the data, instead of the DIC as a penalized measure of fit. If our estimate of K was unambiguous within an analysis and in agreement across analyses within a particular category (i.e. no-admixture or admixture), we performed all further MCMC simulations by specifying K at this value.

In general, we placed greater emphasis on results from the admixture analyses because mixed ancestry is a common feature of real populations, and models that do not allow admixture are less likely to detect such patterns in the data. Furthermore, it is recommended that the no-admixture models be used mainly to approximate an upper bound on K_{MAX} (TESS v2.3 reference manual).

Next, we used the DIC to test the fit of models that incorporated varying degrees of spatial information. These ranged from non-informative spatial prior (i.e. $\psi = 0.0$), to a trend only prior, and finally to the full model (trend surface plus CAR prior). We explored two surfaces for the trend only prior; trend = 0 (translated, this means that whatever is causing an observation in one place also causes similar observations in other places) and trend = 1.0 (individuals at a given location directly influence the characteristics of nearby locations). Once an appropriate trend surface was established for a given K, we then explored whether the inclusion of the CAR prior further improved model fit. For this component, we weighted the spatial network according to the average great circle distance between collecting sites following Durand et al. 2009.

The final step of the analysis was to perform MCMC simulations at a specified K_{MAX} using the preferred model. We performed 100 replicate searches and retained the 10 runs with the lowest DICs. The MCMC algorithm was run for 50,000 iterations with the first 40,000 samples removed as burn-in. Because the cluster labels are arbitrarily assigned during each replicate simulation, independent analyses of the same data can result in different assignment outcomes. For example, snakes from the Los Vaqueros and Mt. Diablo populations may cluster together in 100% of the total number of MCMC simulations, thereby having the same membership coefficient estimates across runs, but the cluster to which they are assigned may differ among simulations. Thus, to summarize the data from replicate analyses, we used the software CLUMPP v1.1.2 to search for optimal alignments of the estimated membership coefficient matrices (Jakobsson and Rosenberg 2007). The program outputs a permuted version of the same data so that the cluster labels of all replicates match as closely as possible. We then used the program DISTRUCT v1.1 to generate graphical displays of the aligned matrices (Rosenberg 2004).

3.6.4. Genetic diversity and differentiation: STR loci

Genetic diversity – We measured two indices of genetic diversity within populations for the STR data; number of alleles per locus (or allelic richness A_R) and expected heterozygosity (H_E, or genetic diversity). We estimated the mean number of alleles per locus for each population using a rarefaction procedure as implemented in HP-RARE 1.0 (Kalinowski 2005), which corrects for uneven sample sizes across populations. We used FSTAT v2.9.3 (Goudet 2001) to compare allelic richness, observed heterozygosity (H_O), expected heterozygosity (H_E, or gene diversity), and the level of inbreeding F_{IS} among our 12 sampling sites. We also tested for linkage disequilibrium among loci across all populations using a permutation test in FSTAT v2.9.3 (10,000 randomizations).

To visualize differences in allelic patterns across sampling sites in the East Bay, we used GenAlEx v6.4 (Peakall and Smouse 2006) to construct a histogram comparing various genetic diversity indices. This provided a convenient way to qualitatively screen populations for anomalous variation, and if necessary we could then explore the causes of that variation. We also performed an AMOVA in ARLEQUIN to assess the proportion of allelic variation at the same hierarchical levels as described above for the mtDNA.

Genetic differentiation – To estimate levels of genetic differentiation among populations based on allele frequency data, we constructed two different versions of the dataset: one grouped the populations according to the cluster assignments estimated from the Bayesian admixture models, and the other treated all sampling sites as separate populations.

We calculated pairwise F_{ST} and R_{ST} estimates for both datasets in ARLEQUIN using both the Brookfield transformed and non-transformed datasets. F_{ST} is the proportion of the total genetic variance contained in a subpopulation (the $_S$ subscript) relative to the total genetic variance (the $_T$ subscript), with values ranging from 0 to 1. High F_{ST} (i.e. close to 1.0) implies a considerable degree of differentiation among populations whereas low F_{ST} indicates limited differentiation. This index assumes an 'infinite alleles' model because it treats all alleles as being an equal number of mutational steps apart, and all heterozygotes are considered equally distant. The F_{ST} estimator Θ calculated in ARLEQUIN is based on Weir & Cockerham's approach (1984), and represents the proportion of genetic diversity due to differences in allele frequencies among populations. R_{ST} is an alternative measure of the fixation index that instead depends on a stepwise mutation model; alleles that are further apart in size (greater variance in state) are weighed more heavily in the estimate of population structure since they are less likely to be recently identical by decent. In contrast, alleles with more similar states are more likely to be recently identical by decent, and thus make a smaller contribution to the estimate of population subdivisioning.

For comparative measures of differentiation at each locus across populations, we calculated G_{STest} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005), and D_{est} values (Jost 2008) in the software SMOGD (Crawford 2009), as these indices can more accurately account for differences in allelic diversity than traditional F_{ST} and G_{ST} measures in cases where populations are fixed for different alleles (Jost 2008). For example, in a scenario where one population is fixed at allele A, but 10 others are fixed for allele B, G_{ST} will over-estimate a population-wide diversity of 1.0 while D_{est} will report a population diversity of 0.55.

To better visualize population differentiation, we used the software GDA (http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php) to output a distance matrix based on the pairwise estimate of Θ for all 12 sampling sites ($d = -\ln(1 - \Theta)$) in which we obtained genotypic data. We then used the matrix to generate UPGMA (Unweighted Pair Group Method using Arithmetic averages) and NJ (Neighbor-Joining) trees. The UPGMA method builds a tree by first finding the two populations with the smallest distance and treating them as a

unit. The matrix is then recomputed with the distances to the pair being replaced by the average of the two distances to the separate populations. This results in an array that is smaller by one, and the process is repeated until all populations are joined in clusters. Because UPGMA uses mathematical averaging of distances, all branches are assumed to evolve at a constant rate and the root is drawn on the longest branch in the tree. The NJ method does not average the branch lengths at each step so the branches from the same terminal node do not have to be the same length. This allows for uneven rates of evolution across branches. This algorithm clusters populations such that the total tree length is minimized. The tree begins as a star, and in the first step, the two most similar populations are joined and then treated as a single unit. The process is repeated such that the total branch length is minimized, and the inferred tree is unrooted.

3.7. Isolation by distance analyses

To verify and expand on IBD inferences from the Bayesian assignment tests, we performed pairwise population *F* statistics on both the STR and mtDNA data. We used the 12 main sampling sites for estimating pairwise genetic distances from both the STR loci and mtDNA haplotypes (Fig. 1), and added the three additional sites that we sampled exclusively for mtDNA (see Appendix 8.2). We calculated the centroid of each set of clustered collection points and measured Euclidean geographic distances among them in ARCGIS 9.3 (ESRI). The average distance among collection points within clusters was 0.438 km (range 0 - 4.781 km) while the average distance among clusters was 34.92 km (range 2.126 – 67.547 km).

For STR data, we calculated Θ (Wier and Cockerham 1984), an estimator of F_{ST}, between all pairs of populations in the program IBDWS (Jensen et al. 2005). We compared pairwise matrices of geographic distance and linearized Θ (Rousset 1997) using Mantel tests for matrix correlation (Mantel 1967), with significance assessed by 10,000 randomizations of the genetic distance matrix. For mtDNA sequences, we calculated Φ_{ST} in ARLEQUIN v.3.5, an analog of F_{ST} that incorporates haplotype frequency and sequence divergence (Excoffier et al. 1992b). In each analysis, we examined untransformed and log transformed axes, and report the combination that exhibited the highest correlation coefficient.

Alameda whipsnake habitat is fragmented by roads and other urban development, and it is possible that the presence of roads may be affecting patterns of migration and genetic variation. In particular, the management units are separated by three major freeway corridors; Interstate (I)-580, I-680 and California State Highway 24. We used partial Mantel tests to assess the effect of highway presence on population pairwise genetic differentiation, after controlling for geographic distance (Legendre and Legendre 1988). A binary fragmentation matrix was created with values of 1 for population pairs separated by major highways and 0 for pairs not separated. All IBD analyses were performed in IBDWS 2.5 (Jensen et al. 2005) and IBD 1.53 (Bohonak 2002).

4. Results

4.1. Analysis of historical phylogeography and demography based on mtDNA

4.1.1. Range-wide phylogeographic assessment of M. lateralis

Our PCR primers amplified 1320 bp of the ND4 gene, 67 bp of the tRNA-*his*, 57 bp of the tRNA-*ser*, and 55 bp of the tRNA-*leu* (total = 1499 bp). The final data set consisted of 155 *M. lateralis* and 19 outgroup taxa. Based on the collection locations and our current understanding of the subspecies boundary, up to 51 of the *M. lateralis* sampled for mtDNA may represent Alameda whipsnakes (depending on whether populations with intermediate phenotypes are classified as *Mle*). Figure 3 shows the Bayesian estimate of the phylogenetic tree for all mtDNA

haplotypes (50% majority consensus tree from the posterior distribution), alongside a color-coded map that displays the sampling locations. For the remainder of the paper, we refer to clades by their regional names as shown in Figure 3.

Bayesian inference of the phylogenetic tree provided strong support for six regionally structured clades (posterior probabilities of 0.99 – 1.0). These clades extend from the Cape to mid-peninsula Baja California, and from the Peninsular Range of northern Baja up through the Sierra Nevada and Coast Ranges of California (Fig. 3). The most southerly clade, hereafter 'Southern Baja', is disjunct from all remaining clades to the north, and consists of isolated populations that stretch from the mid-peninsula region near San Ignacio south to the Isthmus of La Paz in Baja California Sur. Members of this clade are more closely related to other species of *Masticophis* in the Cape region (*M. aurigulus* and *M. barbouri*) than to other lineages of *M. lateralis*, a relationship that indicates that *M. lateralis* is a paraphyletic species.

Whipsnake populations are absent from areas north of San Ignacio up to the Peninsular Range of Northern Baja California, where populations belonging to a second well-supported clade (Northern Baja) are continuously distributed up into extreme southern San Diego County in California. Here, the Northern Baja clade contacts a third clade we refer to as 'Southern CA', which stretches northward through the Transverse and Coast Ranges along the California coast up to Monterey Bay. Just south of the Monterey Bay area, a few populations belonging to this clade also exist on the opposite side of the Salinas Valley in the southern extension of the Diablo Range.

A fourth widespread clade, referred to as 'Northern CA, extends from the western foothills of the southern Sierra Nevada in Tulare County northward, around the northern rim of the Central Valley, and back south down to the San Francisco Bay Area. The Northern CA clade is the sister taxon to a substantially smaller clade we refer to as 'Central CA'. The water barrier created by San Pablo and Suisun Bay delineates a clear contemporary geographic boundary between these two clades, but their close relatedness suggests a fairly recent historical connection. Haplotypes within the Central CA clade belong to populations that are phenotypically recognized as both *Mle* and *Mll.* Members of this Central clade extend south from the Bay Area well into the Diablo Range in San Benito County, and come fairly close to contacting members of the Southern CA clade. We did not detect any overlap in the distribution of haplotypes belonging to these clades; however, this is likely an artifact of our sampling.

Table 3. Diversity and demographic measures for mtDNA sequences (10,000 coalescent simulations; bold indicates significance at $P<0.05$). Notations are as follows: s = no. of polymorphic sites; π = nucleotide diversity; h = haplotype diversity; r_g = raggedness index of Harpending 1994; R_2 = summary statistic of Ramos-Onsins and Rozas 2002. Modality refers to the shape of the mismatch distributions

Lineage	*n*	*No. of Haplotypes*	*s*	*π*	*h*	*Modality*	*Tajima's* D	*Fu and Li's* D	r_g	R_2
Northern CA	28	17	26	0.003	0.93	unimodal	**-0.081**	**-0.067**	**0.083** (0.00)	**0.123** (0.04)
Central CA	59	14	15	0.001	0.70	unimodal	**-0.043**	**-0.044**	**0.182** (0.02)	**0.106** (0.06)
Southern CA	42	27	48	0.004	0.94	unimodal	**-0.089**	**-0.064**	**0.068** (0.06)	**0.110** (0.00)
North Baja	9	7	19	0.004	0.94	unimodal	-0.067	**-0.072**	0.152 (0.11)	**0.176** (0.07)
South Baja	5	5	17	0.009	1.00	multimodal	**-0.037**	**-0.043**	0.285 (0.71)	**0.232** (0.04)

4.1.2. Genetic diversity and differentiation: mtDNA

Genetic diversity – Table 2 provides summary of the number of nucleotide sites by partition, as well as the number of variable and parsimony informative sites. Not surprisingly, most of the phylogenetic signal in the data came from 3rd codon positions in the protein coding portion of the

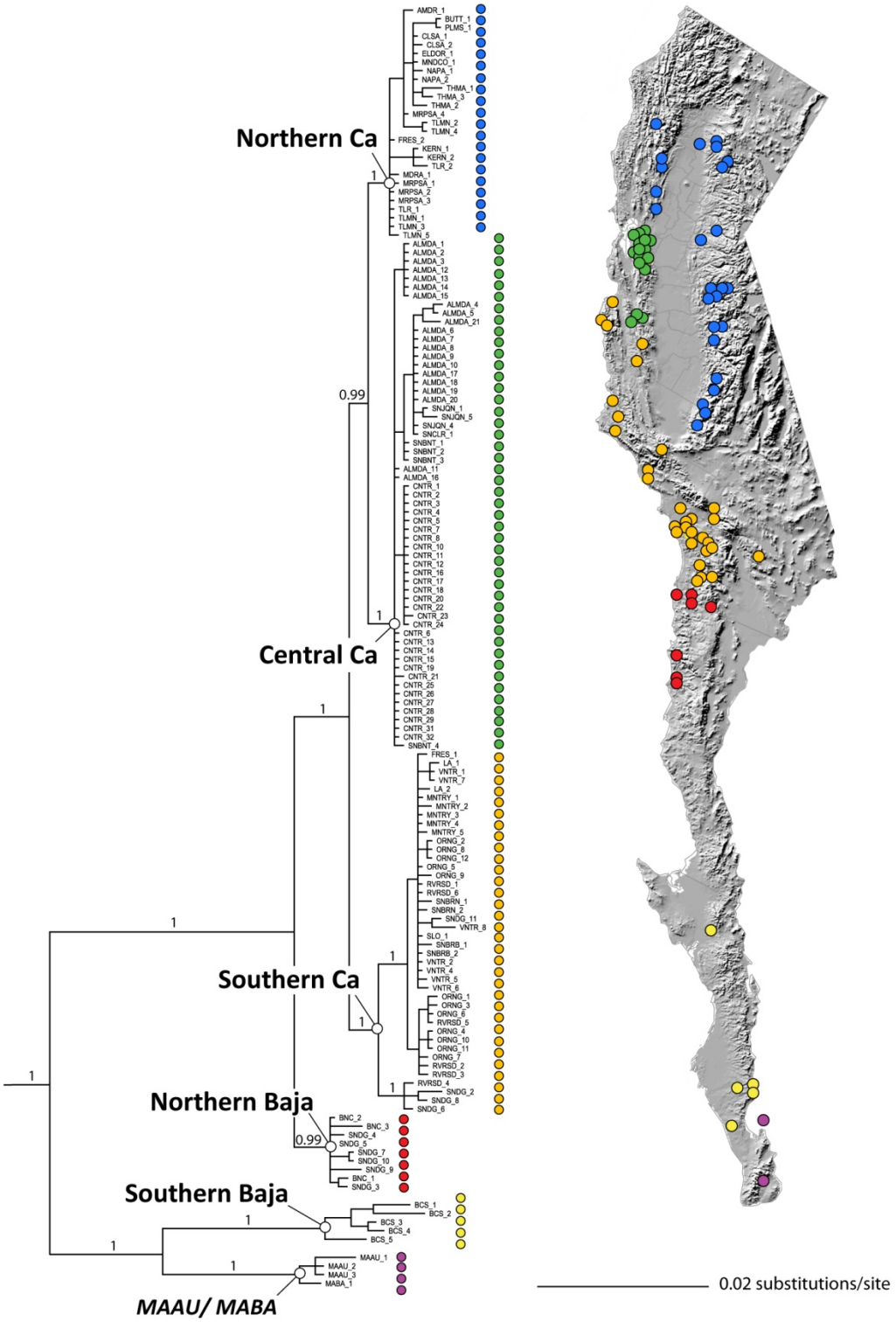


Figure 3. Bayesian estimate of the phylogenetic tree for *M. lateralis* based on mtDNA sequence data. The location and identity of individual haplotypes are provided in Appendix 8.2. Colors highlight the major mitochondrial lineages within the species. Numbers over the tree branches indicate posterior probablilities for the inferred relationships. *MAAU* = *Masticophis aurigulus*; *MABA* = *Masticophis barbouri.* Figure adapted from Leaché et al 2009.

sequence. Genetic diversity estimates as measured by polymorphic sites (*s*), nucleotide diversity (π) and haplotype diversity (*h*) reveals relatively similar variation among three of the five mtDNA lineages (Table 3). The exceptions were the Central CA and Southern Baja lineages, which exhibit the lowest and highest diversity estimates, respectively.

The first AMOVA, which included all populations within the Central CA clade, showed that 77.1% of the total mtDNA variation was explained by differences among the sampling sites ($P < 0.001$), with the remaining 22.9% explained by variation within sites ($P < 0.001$). When considering only the critical habitat units in Alameda and Contra Costa County in the second AMOVA, the majority of the variation (68.1%) was explained by differences among the units ($P < 0.001$), 23.1% was accounted for by variation within a sampling site ($P < 0.001$), and the remaining 8.8% was explained by differences among sampling sites within units ($P < 0.001$).

Genetic differentiation – The average percent sequence divergence of the basal split within *M. lateralis* (between the Southern Baja clade and all other lineages combined) is 8.6%. Among all five of the major mtDNA clades, the highest level of divergence is between the Southern Baja and Central Ca lineages at 8.5%, and lowest between the Northern Ca and Central Ca lineages at 0.9% (Table 4). Average sequence divergence within each of the five lineages ranges from 0.1% – 0.7%; the lowest levels of divergence were recovered for the Central Ca lineage, which contains the known populations of *Mle*, and the highest were recovered in Southern Baja.

We also report pairwise Φ_{ST} estimates for each of the regional mtDNA clades and for all pairwise combinations of sample sites within the Central CA clade in Appendices 8.7 and 8.8, respectively. The pairwise Φ_{ST} estimates for all Central CA sites were used to generate the reduced major axis regressions and Mantel tests for the IBD analyses in section 4.3 below. We call attention to the fact that divergence estimates based on percent sequence differences and Φ_{ST} between the Central and Northern CA sister clades within *M. lateralis* exceed those between two currently recognized species in the Cape Region of Baja California, *M. aurigulus* and *M. barbouri*. We discuss the implications of this result in detail in Section 5.1.1.

Table 4. Average sequence divergence within (diagonal) and between the major mtDNA clades (below diagonal).

Clades	*MAAU/ MABA*	*Southern Baja*	*Northern Baja*	*Southern CA*	*Northern CA*	*Central CA*
MAAU/MABA	*0.004*					
Southern Baja	0.040	*0.007*				
Northern Baja	0.064	0.066	*0.003*			
Southern CA	0.073	0.073	0.020	*0.003*		
Northern CA	0.067	0.071	0.017	0.014	*0.003*	
Central CA	0.082	0.085	0.019	0.015	0.009	*0.001*

4.1.3. Estimates of divergence times

Our 'time to most recent common ancestry' estimates (TMRCA) indicate that diversification of the major clades within *M. lateralis* occurred in the PLIOCENE to late PLEISTOCENE (Fig 4.). The stem origin for the full tree is estimated at approximately 3.79 Ma, and is centered along the mid-peninsular region of Baja California. Following this initial divergence, populations of *M. lateralis* in southern Baja California diverged from *M. aurigulus* and *M. barbouri* (~2.26 Ma), the former of which is restricted to the Sierra La Laguna of the Cape region and separated from *M. lateralis* by the Isthmus of La Paz, and the latter of which is endemic to the Gulf Islands of Espíritu Santo and Partida Sur. More recently, Northern Baja populations diverged from those in California at roughly 1.01 Ma, followed by further divergence of the three California lineages in the late Pleistocene. The Southern CA lineage split from the common ancestor of the Northern and Central CA groups ~0.70 Ma, and the Northern and Central CA lineages separated from each other at approximately 0.51 Ma. The 95% credible intervals for these calibrations, which represent uncertainty in the estimated values, are shown in Fig 4. Our results using the two

different prior distributions for the mean calibrations (lognormal versus normal) produced highly overlapping 95% credible intervals for the TMRCA estimates, indicating that our results were robust with respect to the choice of this prior distribution.

4.1.4. Inferences of historical and recent changes in demography

Historical demography – Tests of historical demographic expansion reveal consistent signatures of increased population size in the three California lineages (Table 3). The results of all of the non-genealogical tests (r_g and R_2 of mismatch distributions, Tajima's *D*, and Fu and Li's *D*) significantly reject the null hypothesis of demographic equilibrium, and site mismatch distributions were all consistent with the signature of an expanding population (i.e. unimodal). Only two of five demographic statistics in the Northern Baja lineage and 3/5 in the Southern Baja lineage were significant, suggesting marginal support for rejecting a null hypothesis of a constant population for these Baja California lineages (Table 3). However our results for the two Baja clades may be influenced by small sample sizes.

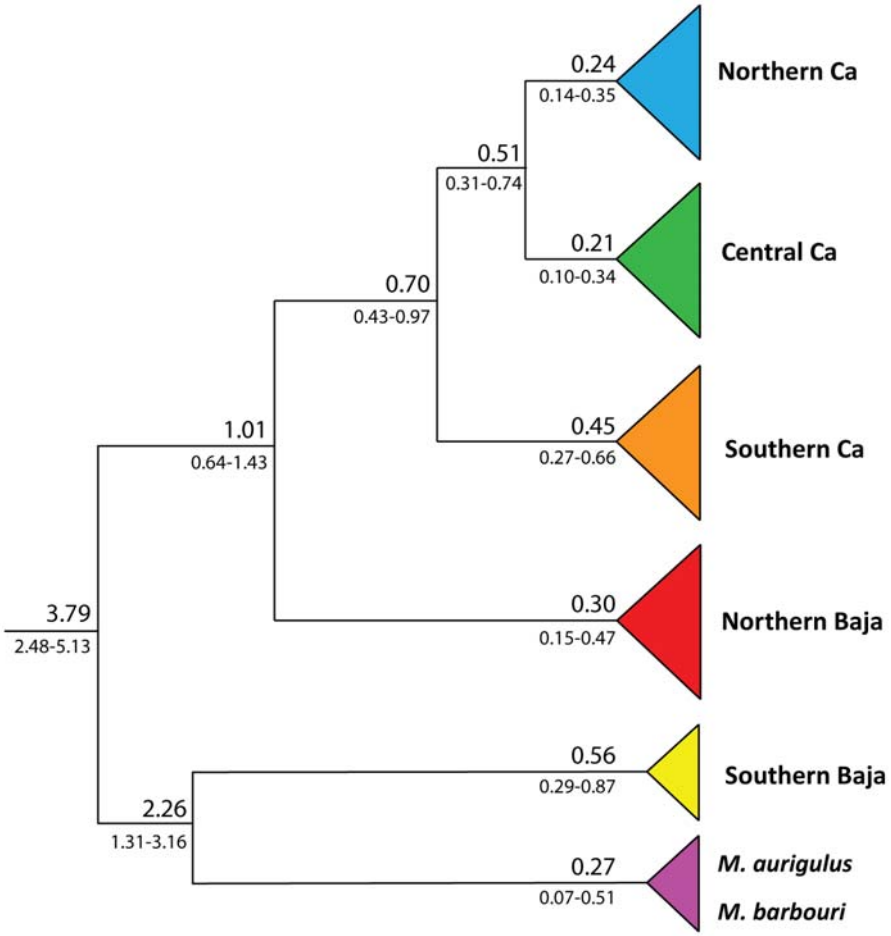


Figure 4. Bayesian relaxed-clock estimates of divergence times for the major mtDNA lineages of *M. lateralis*. Numbers above the branches represent time (millions of years before present); numbers below the branches are the 95% highest posterior densities, and represent the uncertainty of the age estimates. Color codes for the major lineages match those in Figure 3.

Bayesian skyline plots (BSPs) of the three California clades corroborate the results of the more traditional, coalescent-based methods described above; namely, that population sizes have expanded for all three clades. The advantage of the BSPs, however, is that the shape of the population growth curves provides information about the magnitude of change and the dating estimates provide a temporal reference over which the expansion occurred (Fig. 5). Both the Southern and Central clades show relatively stable effective population sizes up until about 50- and 25,000 years ago, respectively, at which time population sizes began to increase. In contrast the Northern CA clade shows a more gradual expansion extending as far back as 200,000 years. The expansion within Central CA clade is clearly the youngest, and the magnitude of change in this lineage is apparently smaller than either the Northern or Southern clades.

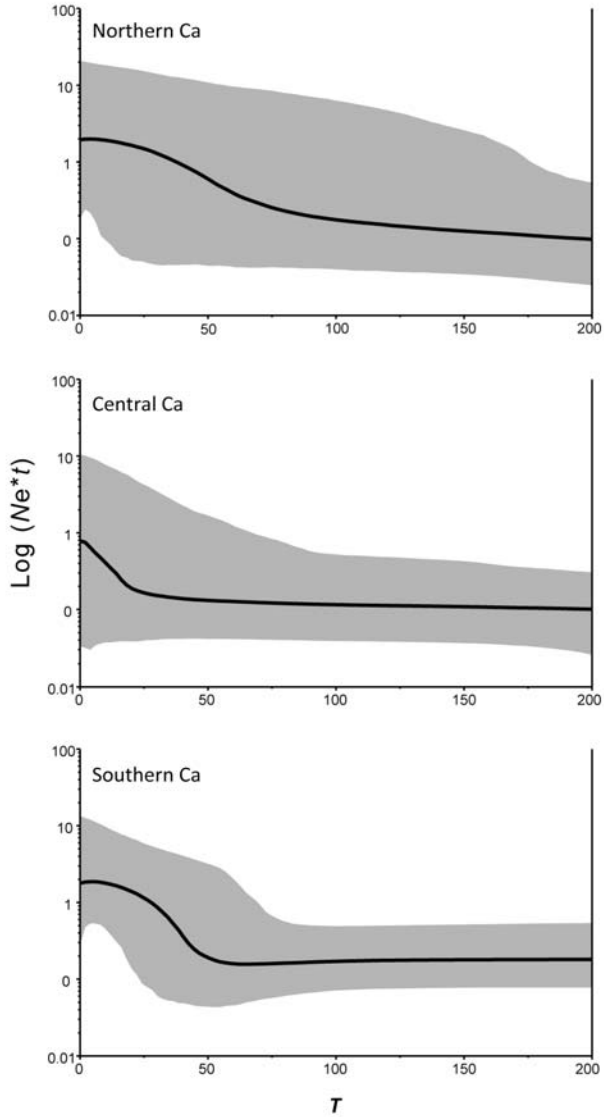


Figure 5. Bayesian skyline plots for the three major mtDNA lineages in California. Time *T* on the x-axis is reported in thousands of years, with the present time at the left. The blue-grey areas represent the credible intervals surrounding the median estimates for the log of the population size (*Ne***t*).

In addition to inferring temporal changes in demography, we used NCPA to test whether populations in the Central CA clade show evidence of geographic range expansion. The first step of this analysis was to construct a haplotype network; all mtDNA haplotypes were connected in a single most parsimonious network that consisted of 17 mutational steps (95% parsimony connection limit: Fig. 6). Haplotype 3 is inferred as the most ancestral haplotype within in the Central Ca network because it represents a transitional node connecting Northern CA haplotypes to all Central CA haplotypes (results not shown). It is also one of the two the most common haplotypes in the Central CA clade, the other being Haplotype 1. Hierarchical nesting procedures reduced six 1-step haplotype clades into two main clades; clade 2-1 extends south from Contra Costa County well in to San Benito County in the Diablo Range, while clade 2-2 is restricted to a substantially smaller area near the Alameda/San Joaquin County border, and south into Santa Clara County (Fig 6).

Three of the eight nesting clades yielded unambiguous NCPA inferences that are consistent with contiguous range expansion (Appendix 8.5). In nested-clade 1-1, the expansion likely occurred from north-to-south, given that the ancestral haplotype (i.e. haplotype 3) is

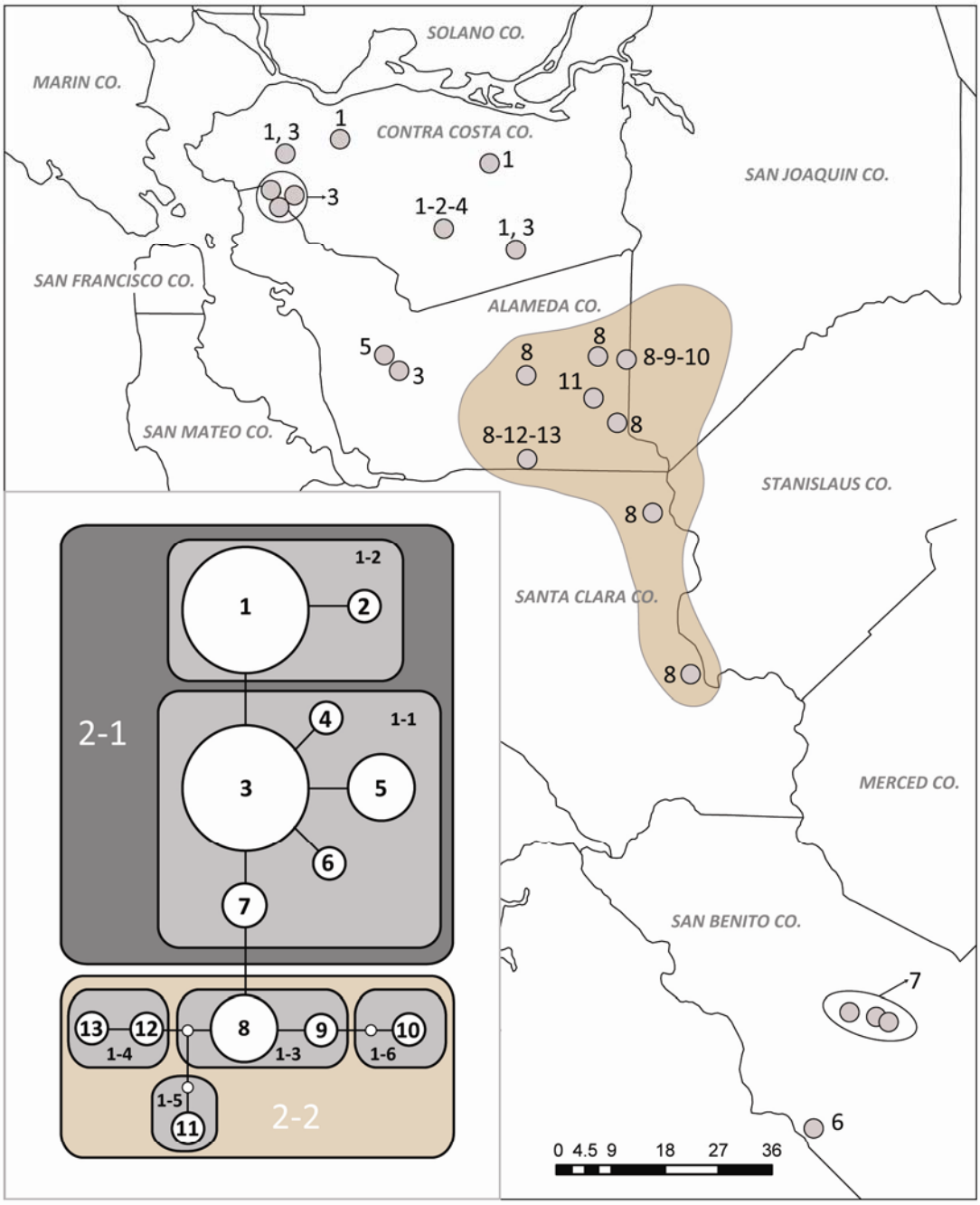


Figure 6. Mitochondrial DNA haplotype network overlain on a map of sampling points. Each branch in the network indicates one mutational step (bullets indicate inferred mutations). The size of the circles in the network is proportional to the frequency that the haplotype was detected in the sample. Colored boxes indicate the arrangement of nested clades as inferred by NCPA. Multiple numbers next to a sampling site indicates that >1 mtDNA haplotype was detected at that site.

limited to Contra Costa and Alameda County to the north, and haplotypes only one mutational step away are found as far south as the Panoche Pass area in San Benito County (Fig 6.). The apparent absence of haplotype 3 in Santa Clara and San Benito Counties may be an artifact of our lower sampling density in this area relative to Contra Costa and Alameda Counties. Following this historical range expansion, genetic connectivity among populations appears to have been restricted by geographic distance, an inference obtained from the remaining higher level clades: nested-clade 2-1, and the total cladogram.

Recent demography – We found no evidence of recent decreases in effective population sizes at any of the sampling locations in the East Bay. To ensure that our results were not sensitive to the way we grouped populations, we performed an initial set of analyses that treated the 12 primary collection sites/populations separately, and a second set of analyses that grouped populations according to their respective recovery units (Fig. 1). Neither approach showed evidence of a heterozygosity excess that was significantly larger than expected at mutation-drift equilibrium.

4.2. Analysis of population dynamics based on STR allele frequencies

4.2.1. Estimates of the number of clusters

Plateaus for the DIC and lnP(D|K) curves for no-admixture models in TESS and STRUCTURE both suggest that four to six clusters exist among the East Bay sampling localities (Fig. 7). The lnP(D|K) curve was slightly more ambiguous as the plateau was less definitive, perhaps tending more towards six groups. As, expected, the inclusion of a spatially informative prior in TESS increased the sensitivity to detecting more clusters; when $\psi = 0.6$, the curve plateaued at $K_{MAX} = 5$, whereas without the spatial prior ($\psi = 0.0$) the curve plateaued at $K_{MAX} = 4$. Because the DIC supported the model that included spatial effects when we varied K_{MAX} from 4 to 10, our results using a no-admixture model suggested that five clusters exist in the dataset. We note that with no-admixture models that impose spatially explicit priors, the DIC can sometimes support a K_{MAX} that overshoots the true number of clusters K; thus, it is recommended that no-admixture models be used mainly to infer an upper limit on the possible number of clusters (TESS V2.3 reference manual).

For admixture models, both the DIC and lnP(D|K) curves had conspicuous inflection points at $K_{MAX} = 4$ (Fig. 7A). In TESS, the curves generated from analyses with out and without spatial interactions (i.e. $\psi = 0.0$ versus 0.6) were largely overlapping, indicating that local spatial effects did not substantially improve model fit. Figure 7A compares the DIC curves of admixture models that assumed a linear trend surface, with a clear plateau at $K_{MAX} = 4$.[2] Thus, based on the more realistic admixture models, we were confident that four main clusters existed within the dataset. We conducted all further model testing by constraining $K_{MAX} = 4$, but ran a single no-admixture analyses (100 replicates, 50,000 sweeps/40,000 burn-in) with $K_{MAX} = 5$ for comparison.

4.2.2. Best fit models for estimating admixture coefficients

Next, we tested the fit of eight models, each representing different ways of capturing spatial information, by varying the strength of the hyper-parameter ψ, fitting a trend surface, and including conditional autoregressive residuals. There was a clear jump in DIC scores between models that fit a constant trend surface (models 1-3) versus a linear trend surface (models 4-7;

[2] We observed the same pattern using a constant trend surface, but did not show the results for graphical clarity.

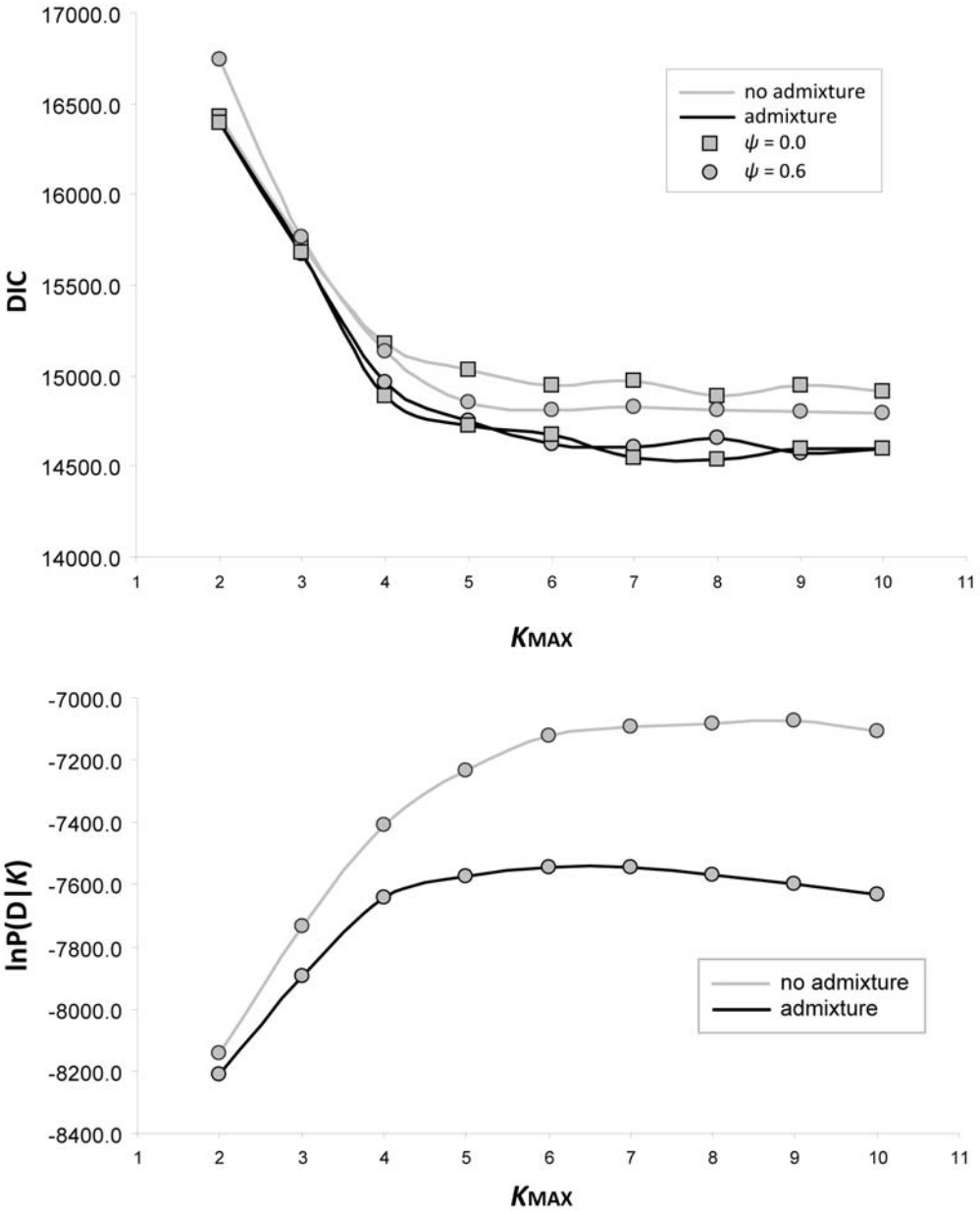


Figure 7. DIC and lnP(D|K) curves used to estimate *K*MAX in TESS and STRUCTURE.

Table 5), indicating the presence of ordinal trends in the data when considering characteristics of all sampling points on average. Models with increased spatial interaction ψ (ψ range = 0.0 – 0.8) tended to fit about the same as those without spatial interactions (models 4-6; Table 5), indicating that local connections do not help in explaining genetic structure. This also emphasizes that our results were primarily being informed by 'signal' in the genetic data, and not by the spatial data. Last, the full model with a linear trend surface plus a CAR prior had a very similar DIC score to the model without the CAR prior (model 7 vs. 5, respectively), and only when ψ and its variance were allowed to update during the MCMC simulation did the DIC drop substantially. Based on these results, we ran two analyses with 100 replicates each (50,000 sweeps, 80% burn-in), one implementing model 4 and one implementing model 8, to estimate admixture coefficients for $K_{MAX} = 4$.

Table 5. Model comparisons using the Deviance Information Criterion (DIC). The smallest DIC indicates the model that would best predict a replicate dataset with the same structure as the observed data*. Notation is as follows: trend = trend surface; CAR = conditional autoregressive model; ψ = parameter that influences the strength of spatial interactions; σ_2 = variance of ψ. We limited our analyses to constant and linear trend surfaces, and explored models with and without CAR priors (yes/no). In model 8, both ψ and σ_2 were updated during the MCMC simulations.

Model	Mean DIC	SD	trend	CAR	ψ	σ^2
1	15145.14	1.23	constant	no	0.0	1.0
2	15128.59	0.89	constant	no	0.6	1.0
3	15128.51	2.73	constant	yes	0.6	1.0
4	14884.77	0.97	linear	no	0.0	1.0
5	14901.71	0.88	linear	no	0.6	1.0
6	14892.21	10.26	linear	no	0.8	1.0
7	14903.57	1.90	linear	yes	0.6	1.0
8	14840.27*	14.51	linear	yes	update	update

4.2.3. Cluster assignments

We show the cluster assignments from the no-admixture analysis in the form of a Voronoi tessellation, where assignments are denoted as different colors (Fig. 8). We present results for $K_{MAX} = 5$ because the DIC favored five clusters for the no-admixture model with spatial effects. Each cluster largely corresponds to one of the designated critical habitat units in the East Bay, with two exceptions. One is for the Claremont sample at the northern edge of Unit 6, which clusters with all populations immediately to the north in Unit 1; the second is the S300 population, which clusters separately from the Camp Ohlone and Del Valle population across Unit 5. The separation of S300 from Camp Ohlone/Del Valle in the STR data mimics the result from our NCPA analyses, where nested clade 2-2 represents a distinctive mtDNA lineage in the general vicinity of the S300 population (Fig. 6).

Cluster assignments based on admixture models produced similar results to the no-admixture model, where each cluster largely corresponded to a critical habitat unit (with the same caveat for the Claremont population as above; Fig. 9); however, instead of S300 being assigned to its own cluster, the admixture model reveals some allele sharing between the S300 population and the Camp Ohlone and Del Valle populations to the west. Nonetheless, the S300 samples exhibited a greater degree of allelic exclusivity from Unit 3 when compared to the Camp Ohlone or Del Valle samples, consistent with the prediction of the no-admixture model. When we added the CAR prior to the model and allowed ψ and its variance to update during the MCMC simulation, we detected greater genetic admixture between the Camp Ohlone-Del Valle samples and those in Unit 3 (Stonebrae and Bailey Ranch). Interestingly, we detected little mixing between S300 and samples in Unit 4 immediately to the north/northwest, even under the most spatially explicit conditions. This result suggests that majority of gene exchange in Unit 5 is

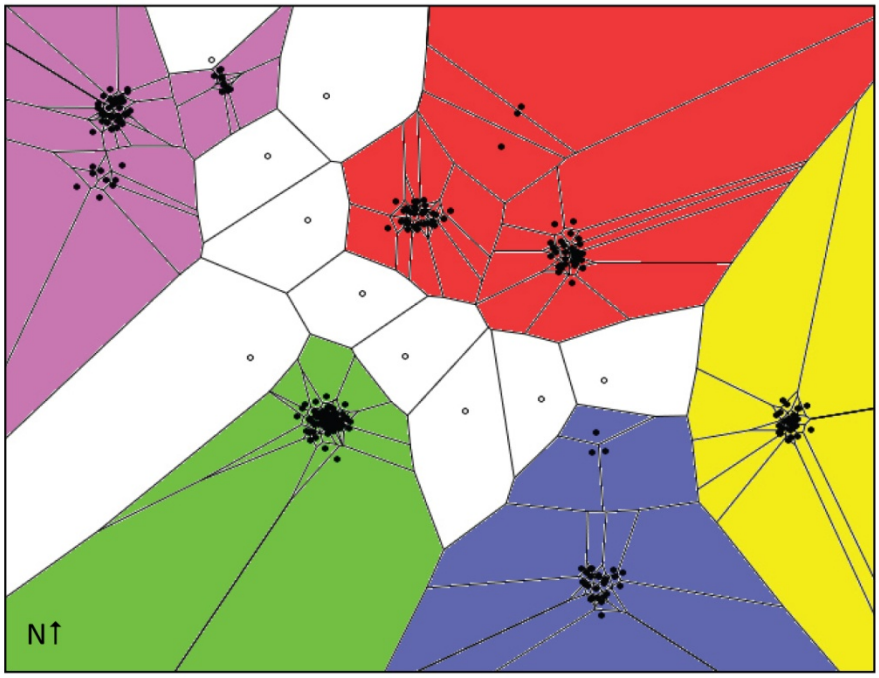


Figure 8. Voronoi tesselation with colors indicating cluster assignments based on the no-admixture model in TESS (*K* = 5). Dark bullets indicate actual sampling points; white cells/open bullets indicate dummy points, which fall along major urban corridors along interstate 580 and 680.

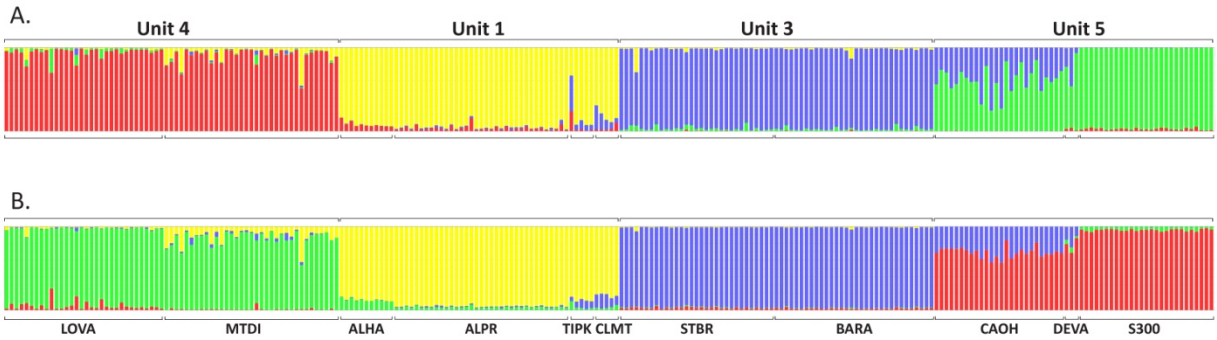


Figure 9. Estimated assignment probabilities of individuals to different geographic clusters *K*. Labels and brackets above the histograms correspond to critical habitat units; labels below correspond to East Bay populations. Each individual in the dataset is represented by a vertical line broken into *K* color segments, with lengths proportional to each of the *K* inferred clusters. A.) Estimated cluster assignments without spatial priors. B.) Cluster assignments with the full CAR model. Results with and without spatial priors are consistent, indicating that the analysis was primarily informed by the allelic data, and not the spatial data (cluster colors do not correspond between runs because they are randomly assigned at the start of each MCMC search).

occurring across an east/west axis, and that the composite ancestry of the Camp Ohlone-Del Valle samples is a reflection either recent historical or ongoing gene exchange.

In general, individuals showed minimal signs of co-ancestry from different critical habitat units, as indicated by the assignment probabilities in Figure 9. Colors within each vertical bar of the histogram represent the proportion of an individual's genome that is assigned to a particular cluster; in this set of populations, individuals are fairly exclusive to a given cluster (i.e. color), indicating a lack of recent migration with gene flow (at least at the distances separating the sampling sites – see Section 5.3 for further discussion). Nonetheless, estimates of the admixture coefficients reveal patterns that are expected under a model of genetic isolation-by-distance. For example, the more northerly population in Unit 4, Mt. Diablo, exhibits a far greater degree of admixture with more geographically proximate populations from Unit 1, than does Los Vaqueros (Fig. 9). Similarly, the two most southern populations in Unit 1, Tilden Park and Claremont, show greater admixture with Unit 3 than do either the Allen Property and Alhambra samples. Finally, the Camp Ohlone and Del Valle populations, which fall in between the Unit 3 and S300, show a high degree of admixture from both sites.

We also note that genes appear to be moving out of Unit 3 more than they appear to be moving in, despite the sandwiched location of Unit 3 populations between Units 1 and 5. For example, Claremont and Tilden Park (U-3) show some influx of genes from Stonebrae and Bailey Ranch (U-5), but there is little to no evidence that genes have moved in the opposite way. Likewise, Unit 3 genes have apparently made their way in to the Unit 5 cluster, but there is no evidence of gene movement in the reverse direction.

4.2.4 Genetic diversity and differentiation: STR loci

Genetic diversity – We recovered 136 alleles from 16 loci, ranging from one to 24 with an average of 8.50 ± 5.28 alleles per locus. A single locus B10 was monomorphic in all Bay Area populations, and was therefore removed from our analyses. Mean number of observed polymorphic alleles across all loci (in populations with >10 individuals) ranged from 2.92 in the Alhambra population to 5.73 at Camp Ohlone (Table 1). When we grouped samples into their respective habitat units[3] and used rarefaction to estimate average allelic richness A_R and private allelic richness A_{PR}, we found that A_R ranged from 4.69 in the Unit 1 group to 6.29 in the Unit 5 group. Private allelic richness showed a similar pattern, with Unit 1 having the lowest A_{PR} at 0.19 and Unit 5 having the highest at 1.52. Of the 30 total private alleles detected in the dataset, 54% were recovered from Unit 5 populations (37% from Camp Ohlone and 17% from S300). In fact, the A_{PR} estimates between Units 1 and Unit 5 were significantly different from each other (Mann-Whitney U test, Z = -3.301, P = 0.001), but neither were significantly different from Units 3 or 4.

The observed (H_O) and expected (H_E) heterozygosity for all loci across populations varied from 0.31 – 0.46 and 0.36 – 0.49, respectively (Table 1). H_O and H_E varied much more dramatically when considering each locus individually, ranging from 0.10 – 0.80 and 0.20 – 0.92. MICROCHECKER detected three loci had greater than expected numbers of homozygotes for certain allele size classes in three or more populations; D11 showed heterozygote deficiencies in 9 out of 12 populations, while B107 and A101 showed deficiencies in 3 and 4 populations, respectively. Based on the results of exact tests (10,000 randomizations), all three loci also deviated significantly from Hardy-Weinberg expectations in three or more populations. Because at least one allele amplified in every individual for each of these loci, our findings are consistent

[3] For rarefaction, we lumped the Claremont samples into Unit 1 because of its location near the border of Units 1 and 6, and because we genotyped only five from this site. Claremont samples also clustered with Unit 1 samples in the Bayesian assignment tests.

with the segregation of a 'real' null allele. Thus, we adjusted the genotypes at these loci according to Brookfield's method. In general, this led to scoring one allele as the observed state and the other as missing data for a proportion of 'homozygotes'. We found no evidence for scoring error due to stuttering or large allele dropout at any loci, or any evidence of linkage disequilibrium among loci.

We provide a visualization of various allelic diversity indices for the 12 East Bay sampling sites in Appendix 8.4. The AMOVA for the STR dataset (Brookfield adjusted) mirrored that for the mtDNA; the majority of the allelic variation in the dataset was explained by differences among all populations (78.2%; σ^2_c= 3.4862, F_{ST} = 0.2180, $P \leq 0.001$), 13.5% was accounted for by difference among units (σ^2_a= 0.6042, F_{CT} = 0.1355, $P \leq 0.001$), and the remainder was explained by differences among populations within units (σ^2_b = 0.3679, F_{SC} = 0.0955, $P \leq 0.001$). Nearly identical results were recovered for the non-transformed STR dataset.

Genetic differentiation – Pairwise F_{ST} estimates for nearly all population pairs (Table 6), showed significant differentiation after Bonferroni correction for both the Brookfield adjusted and the non-adjusted datasets. Estimates from the non-adjusted dataset ranged from 0.0247 between Los Vaqueros and Mt. Diablo to 0.3846 between S300 and Claremont. The same pairwise comparisons also had the lowest and highest F_{ST} values (F_{ST} = 0.0185 and 0.4538, respectively) when basing the measurements on the Brookfield adjusted dataset. Because we used a randomization procedure in ARLEQUIN to test for significant differences in the estimated values, some comparisons involving the smallest sample sizes did not produce statistically significant results despite having values that were similar to those in other comparisons that were significant (e.g. comparisons between DEVA [Del Valle] and CLMT [Claremont] or TIPK [Tilden Park]).

Table 6. Comparison of R_{ST} (above diagonal) and F_{ST} (below diagonal) values by population using Brookfield adjusted versus non-adjusted datasets. Bold values indicate significant differences after Bonferroni correction (α = 0.005). Arlequin uses a permutation test to evaluate significance, thus comparisons involving small sizes may not show statistically significant results despite have values that 'appear' significant (e.g. comparisons between DEVA and CLMT or TIPK).

Non-adjusted	LOVA	MTDI	BLDI	ALPR	ALHA	CLMT	TIPK	BARA	STBR	S300	CAOH	DEVA
LOVA	–	**0.0416**	0.0869	**0.2822**	**0.3481**	**0.1829**	**0.1491**	**0.1306**	**0.0883**	**0.1366**	**0.1222**	0.1475
MTDI	**0.0247**	–	-0.0015	**0.2546**	**0.3659**	**0.1856**	**0.1529**	0.0864	**0.1092**	**0.1814**	**0.0743**	0.2218
BLDI	**0.1267**	**0.0878**	–	0.1273	**0.3230**	0.1366	0.0492	0.0385	0.1123	**0.3145**	0.0760	0.3029
ALPR	**0.2000**	**0.1807**	**0.1819**	–	**0.1172**	**0.3275**	0.0505	**0.2313**	**0.2605**	**0.3648**	**0.2430**	**0.4415**
ALHA	**0.2795**	**0.2672**	**0.3160**	**0.1670**	–	**0.4590**	0.1940	**0.2908**	**0.2488**	**0.4373**	**0.2749**	0.4635
CLMT	**0.2238**	**0.2089**	0.2549	**0.1511**	**0.2661**	–	0.0676	0.0735	0.1374	**0.3491**	**0.2000**	0.3454
TIPK	**0.2180**	**0.2065**	0.2259	**0.0996**	**0.1446**	0.0416	–	0.0986	0.1061	**0.3054**	0.1388	0.2718
BARA	**0.1549**	**0.1584**	**0.2563**	**0.2550**	**0.3008**	**0.2738**	**0.2460**	–	**0.1139**	**0.1012**	0.0633	0.1914
STBR	**0.1265**	**0.1299**	**0.2123**	**0.2153**	**0.2549**	**0.2326**	**0.2068**	**0.0241**	–	**0.1558**	**0.0768**	0.0703
S300	**0.2589**	**0.2661**	**0.3643**	**0.3259**	**0.3348**	**0.3846**	**0.3204**	**0.1671**	**0.1816**	–	**0.1188**	**0.2908**
CAOH	**0.2227**	**0.2323**	**0.2386**	**0.2751**	**0.2992**	**0.3019**	**0.2449**	**0.1244**	**0.1309**	**0.1644**	–	0.1309
DEVA	**0.2202**	**0.2326**	0.2907	**0.2657**	**0.3737**	0.2722	0.2442	**0.1501**	**0.1272**	**0.2185**	**0.0885**	–

Brookfield adjusted	LOVA	MTDI	BLDI	ALPR	ALHA	CLMT	TIPK	BARA	STBR	S300	CAOH	DEVA
LOVA	–	0.0064	-0.0002	**0.1972**	**0.1797**	**0.2117**	0.0995	**0.1315**	**0.0920**	**0.1349**	**0.0875**	0.1308
MTDI	**0.0185**	–	-0.0448	**0.1722**	**0.1884**	**0.2129**	0.1105	**0.0848**	**0.1113**	**0.1045**	**0.0586**	0.2305
BLDI	**0.1240**	**0.0962**	–	0.1443	0.2380	0.1583	0.0939	-0.0025	0.1114	**0.2170**	0.0624	0.3060
ALPR	**0.1884**	**0.1713**	**0.2116**	–	**0.1115**	**0.3557**	0.0762	**0.1508**	**0.2514**	**0.2267**	**0.1677**	**0.3627**
ALHA	**0.2246**	**0.2218**	**0.2977**	**0.1398**	–	**0.3924**	0.1334	**0.1597**	**0.1699**	**0.1218**	**0.1124**	0.1693
CLMT	**0.2496**	**0.2354**	0.2910	**0.1380**	**0.2117**	–	0.0688	0.0685	**0.1603**	**0.4406**	**0.2256**	0.3819
TIPK	**0.2275**	**0.2201**	0.2908	**0.1053**	0.1071	0.0503	–	0.0553	0.1038	**0.2450**	0.1152	0.2176
BARA	**0.1505**	**0.1555**	**0.2688**	**0.2288**	**0.2620**	**0.3160**	**0.2726**	–	**0.1185**	-0.0034	0.0524	0.1939
STBR	**0.1384**	**0.1457**	**0.2541**	**0.2154**	**0.2503**	**0.2894**	**0.2538**	0.0104	–	**0.1186**	**0.0695**	0.0394
S300	**0.2590**	**0.2831**	**0.4257**	**0.2992**	**0.3172**	**0.4538**	**0.3685**	**0.1714**	**0.1925**	–	0.0468	**0.3537**
CAOH	**0.1787**	**0.2025**	**0.2522**	**0.2337**	**0.2491**	**0.3045**	**0.2680**	**0.0682**	**0.0802**	**0.1249**	–	0.1534

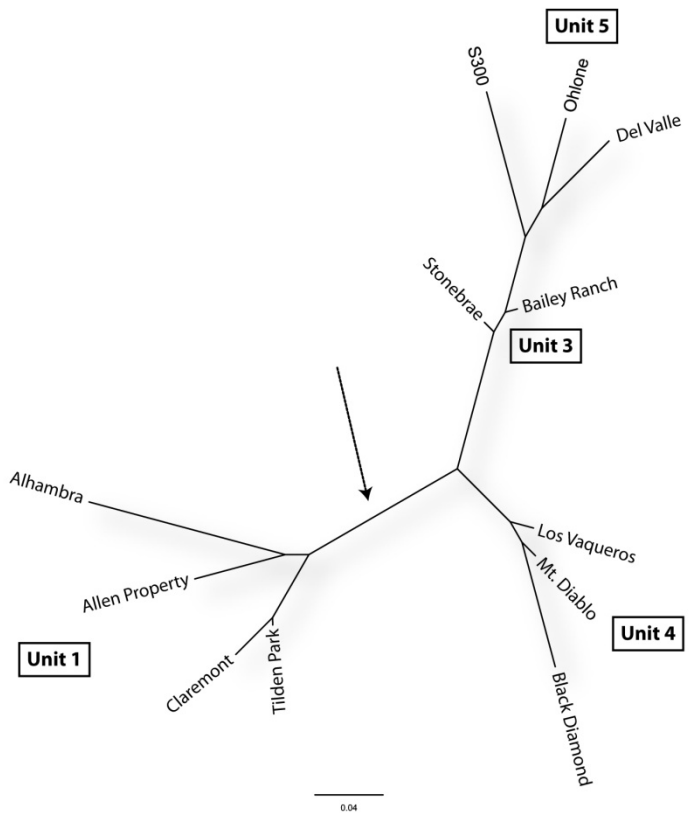


Figure 10. Neighbor-joining tree based on pairwise distance estimates (d = -ln(1- Θ)) between populations. The neighbor-joining method generates an unrooted tree and does not assume constant rates of change along each branch. The UPGMA tree (not shown) has the root placed along the longest branch in the tree (denoted by the arrow).

Pairwise population comparisons based on R_{ST} values were largely consistent with F_{ST} values, but with fewer examples of significant differentiation among certain population pairs (Table 6). As expected, most of the pairs with non-significant differences occur in close geographic proximity, consistent with the pattern of genetic isolation by distance detected in other analyses. Measures of differentiation generated in SMOGD for each locus showed widely varying D_{est} values, ranging from 0.0392 at the A7 locus to 0.6338 at B105. This indicates that our results for genetic diversity and overall population structuring are derived from a series of neutrally-evolving markers with varying rates of evolution.

As a final method for examining population differentiation, we computed a distance matrix based on pairwise estimates of Θ for the 12 sampling sites (d = -ln(1 – Θ), and used it to construct UPGMA and NJ trees (Fig. 10). These results are consistent with four main groupings among the 12 sites, similar to the Bayesian assignment tests and the IBD analyses. We show the NJ tree in Figure 10 because we cannot definitively place the root based on the populations for which we obtained genotypic data (i.e. all are closely related). However, the UPGMA method places the root on the longest branch on the tree, which connects populations from Unit 1 to all other populations in the dataset.

4.3. Isolation by distance analyses

For STR loci, pairwise Θ/1–Θ ranged from 0.0166 to 0.5456 (Appendix 8.5.1). Θ/1–Θ was most strongly correlated with untransformed geographic distances. Genetic differentiation was significantly correlated with geographic distance (Fig. 11). Genetic differentiation was also significantly positively correlated with highway presence; however, this effect was not significant after accounting for geographic distance in the partial mantel analysis (Table 7), suggesting that there is no discernable effect of highway fragmentation.

For mtDNA sequences, pairwise Φ_{ST} among haplotypes ranged from -0.1321 to 1.0000 (Appendix 8.5.2). As with the STR data, genetic differentiation was strongly positively correlated with geographic distance and highway presence. When examining all sampled populations in the Central clade, highway fragmentation was significantly and positively correlated with genetic differentiation, even after accounting for genetic distance; however, when the southernmost collection locations (Henry Coe Regional Wilderness and Panoche Pass) were excluded from the analysis, this relationship was only marginally significant (Table 7).

Table 7. Results of the Mantel and Partial Mantel tests. We report the Pearson correlation coefficient *r* and the associated *P*-values based on 10,000 randomizations of the genetic distance matrices. STR = short tandem repeat; mtDNA' = Central CA clade excluding the Henry Coe and Panoche Hills samples; mtDNA = all sampled populations forming the Central CA clade

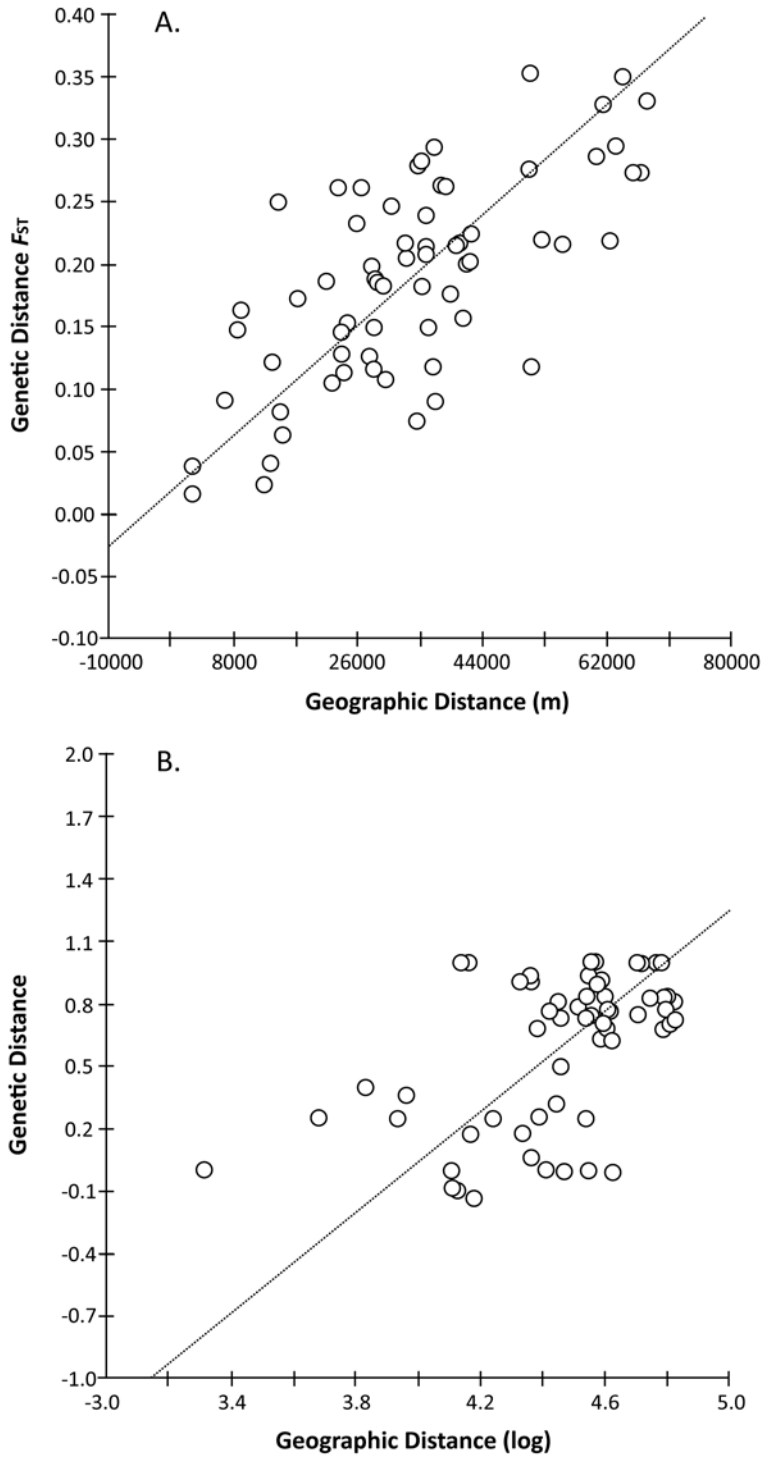
Mantel and Partial Mantel Tests	STR		mtDNA'		mtDNA	
	r	*P*-value	*r*	*P*-value	*r*	*P*-value
Genetics and geographic distance	0.7104	≤0.0001	0.5535	≤0.0006	0.5282	≤0.0002
Genetics and highway presence	0.4935	≤0.0005	0.5546	≤0.0009	0.5260	≤0.0004
Genetics and geographic distance, controlling for highway presence	0.5784	≤0.0005	0.2245	≤0.0533	0.3908	≤0.0011
Genetics and highway presence, controlling for geographic distance	0.0137	≤0.4676	0.2282	≤0.0482	0.3873	≤0.0028

5. Discussion

Our results underscore the complexity of genetic structuring in populations that have historically diverged from others, such as those forming the different mtDNA clades within *M. lateralis*, but have potentially come back in to secondary contact. Such deep genetic structuring within species may indicate either regional restrictions on gene flow or ancient lineage separation that still reveals a signal in the molecular data but does not reflect recent population dynamics. Regardless of the cause for this structuring, identifying divergent populations likely harbor important genetic variation that can ultimately contribute to the long-term sustenance of a metapopulation. It is of course a desirable conservation strategy to identify where that variation exists, with the notion in mind that genetic uniqueness below the level of species can be just as important as that at the level of species when managing populations on local geographic scales.

Although our study did not support *Mle* as an exclusive genetic group, this is hardly a surprising result if gene flow is impeding divergence at putative subspecies boundaries in central coastal California. In fact, perfect overlap between in the patterns of phenotypic and genetic variation would suggest that *Mle* and *Mll* are better candidates for species, but even formally recognized species often fail to meet this standard (Jockusch and Wake 2002; Templeton 1998). Subspecies by their very nature are not reproductively isolated units. Although a group of populations may have diverged to the point of having some detectable phenotypic and genetic differences, these populations may not have achieved sufficient reproductive isolation from neighboring populations to preclude gene flow in secondary contact. Such genetic admixture between divergent populations can and often does create ‘fuzzy’ subspecies borders, and

Figure 11. Reduced major axis regression of pairwise genetic similarity values against geographic distance. Genetic distances are based on STR genotypes (A) and mtDNA haplotypes (B); genetic distance measures are described in the text



emphasizes the fact that subspecies are not simple black and white entities. It also emphasizes that managing subspecies necessarily requires the consideration of how evolutionary processes shape phenotypic and genetic variation.

An even more general issue surrounding conservation management centers on whether phenotypic subspecies are appropriate targets for conservation. This controversy reflects the finding that phenotypic subspecies often fail to correspond to distinctive evolutionary groups as identified by molecular phylogenetic studies (Ball Jr. and Avise 1992; Burbrink et al. 2000; Wood et al. 2008b). As a result, some researchers argue that the focus on subspecies as a management unit has unwittingly turned attention away from clades that harbor distinctive genetic variation, yet are unrecognized on the basis of phenotype (Burbrink et al. 2000; Harris and Froufe 2005; Zink 2004); however, discordance between subspecies and distinct phylogenetic lineages is by no means a rule, as numerous molecular studies have demonstrated support for traditional subspecies boundaries in many taxa (Gavin et al. 1999; Mulcahy 2008; Phillimore and Owens 2006 and references therein). Ultimately, the discordance between phenotypic subspecies and the formation of distinctive lineages rests on whether sufficient time has elapsed for coalescence to have taken place (a historical artifact), and whether sufficient pre- or post-zygotic isolation fails to prevent even small amounts of gene flow from reducing the rate of divergence (a contemporaneous artifact).

5.1. Historical biogeography, clade structuring and divergence times; implications for conservation management

We used a range-wide phylogeographic analysis of the parent species *M. lateralis* to provide a historical context for how East Bay whipsnake populations came to be where they are today, and where their progenitors likely originated from. One of the most important results from this part of the study is that the phyletic diversity within *M. lateralis* is higher than the current taxonomic classification suggests. Based on a limited assessment of color pattern variation (e.g. Riemer 1954), only two forms are presumed to occur in California, *M. l. euryxanthus* and *M. l. lateralis*; however, our mtDNA data reveals three cryptic clades that exist within *Mll* populations alone, indicating that color variation underestimates the genetic diversity within the species. Additionally, none of the recovered clades exclusively correspond to populations with the *Mle* phenotype; rather, *Mle* falls within a more inclusive clade that extends over a large stretch of the inner Coast Ranges of California.

The apparent lack of geographic overlap among these mtDNA clades may be due to the dispersion of our sampling, and more collecting within the putative contact zones (which were unknown beforehand) could potentially reveal sympatry in the distribution of haplotypes. We also cannot rule out the possibility that gene flow is absent between these clades. Sharp mtDNA clade boundaries such as those recovered in this study are common in other taxa, with many revealing similar phylogeographic breaks in the same parts of California (Feldman and Spicer 2006; Kuchta and Tan 2006; Richmond and Reeder 2002; Shaffer et al. 2004; Spinks and Shaffer 2005). Repetitive observations of the same phylogeographic breaks suggest common historical barriers to gene flow, and the exclusivity of different mtDNA clades by region indicates that sufficient time has lapsed for the lineage sorting process to result in monophyletic groups of haplotypes. The biological and taxonomic significance of such clades varies from species to species, but it is generally not warranted to propose new taxonomic descriptions on the basis of mtDNA data alone (see Section 5.1.1 for further discussion of taxonomy and mtDNA phylogeography of *M. lateralis*).

Based on tree topology, branch lengths, and the timing of divergence events, it is clear that the oldest *M. lateralis* haplotypes are from the southern part of the species range – the

branches leading to the tips of the Southern Baja clade are clearly the longest tips in tree, and it is this clade that causes *M. lateralis* to be paraphyletic. This pattern is not uncommon in intraspecific phylogeographic studies based on mtDNA. In fact, the coalescence process dictates that the earliest stages of speciation are characterized by a polyphyletic assortment of tip lineages (or a 'star phylogeny') that slowly progress toward a paraphyletic assemblage over time, and then ultimately to a monophylic group as the lineage sorting process goes to completion (Templeton 1998). Our results indicate that the two nominal species *M. aurigulus* and *M. barbouri* evolved relatively recently from within lineages of *M. lateralis*, or as 'buds' of the phylogenetic tree. Both are striped whipsnakes that look very similar to *M. lateralis*, and it is likely that intermittent historical isolation events in the Cape Region of Baja California played a role in their mtDNA exclusivity and their minor phenotypic differences (Garrick et al. 2009; Grismer 2002; Riddle et al. 2000). In fact many reptile taxa have distinctive pattern classes and/or sister species across this same general area in the Isthmus of La Paz (Grismer 2002).

The mid-peninsula hiatus between the Southern and Northern Baja clades is also a well-documented and recurrent biogeographic pattern across many taxa (Garrick et al. 2009; Leaché et al. 2007; Riddle et al. 2000), although the causes for this gap are not clear. Marine incursions, volcanism, and ecological gradients have all be cited as possible causes, and comparative molecular dating among several mammal and reptilian taxa suggests that patterns of genetic diversification across this area cannot be attributed to a single vicariant event (Leaché et al. 2007 and references therein). Our TMRCA estimate of ~3.8 ma between the Southern and Northern Baja clades is consistent with the timing of a major divergence event in other reptiles with similar distributions.

The split between the Northern Baja and California lineages is enigmatic given that there are no known historical or contemporary barriers near the US/Mexico border, but again the same pattern shows up in host of other reptilian taxa with overlapping distributions (Leaché et al. 2009; Mitrovich 2006; Wood et al. 2008a). Populations within the Southern CA clade evidently progressed northward along the major montane corridors, passing through the Transverse Ranges northward into the central Coast Ranges, but stopping short of Monterey Bay in an area that was formerly inundated by the San Joaquin Embayment. This embayment served as the main drainage from the San Joaquin Valley to the Pacific Ocean as far back as ~2.5 Ma, well before the formation of the San Francisco Bay (Dupré et al. 1991; Hall Jr. 2002; Jacobs et al. 2004). Similar to *M. lateralis,* the signature of this barrier is evident in the phylogeography of other reptile and amphibian taxa (Feldman and Spicer 2006; Richmond and Reeder 2002; Shaffer et al. 2004).

The Northern CA clade, which extends along the western foothills of the Sierra Nevada, around the northern end of the Central Valley, and back down to the San Francisco Bay Area, is the sister taxon to a monophyletic Central CA lineage, which contains haplotypes belonging to both *Mle* and *Mll* populations. This result provides unambiguous evidence that whipsnakes invaded the present day San Francisco Bay area from the north, at a time when the Bay itself did not constitute a significant water barrier. Indeed, our TMRCA estimate of the Northern and Central CA split (mean = 0.51 Ma, 95% highest posterior density = 0.31 – 0.74) coincides with geologic inferences of a marine connection between the Central Valley and the Pacific Ocean via the San Francisco Bay at approximately 0.6 Ma (Dupré et al. 1991). That same configuration remains unchanged to the present day.

The monophyly and strong statistical support for the Central CA clade also provides good evidence that its constituent populations were genetically isolated from both the Northern and Southern CA clades as recently as the late Pleistocene, providing an opportunity for divergence to occur without the influence of gene flow from divergent populations. It is possible that during

this time of isolation, the phenotype now recognized as *Mle* began to evolve. The relatively small contemporary geographic distribution, low haplotypic diversity and evidence for a demographic and geographic range expansion, all suggest that the Central CA lineage was likely bottlenecked at its inception, increasing the prospects for genetic drift to accelerate the sorting of this lineage into a monophyletic group.

5.1.1: Do the mtDNA clades have implications for taxonomy?

The statistically well-supported and regionally distinctive mtDNA clades within *M. lateralis* naturally raise the question of how these data can or should be used to inform taxonomy. In fact, many studies have used mtDNA to identify taxonomic boundaries, with some showing congruence between phenotypic subspecies and molecular phylogenetic data (see reviews in Phillimore and Owens 2006). In this study, we found that *Mle* does not form a distinctive mtDNA group, but rather falls within a more inclusive clade that also includes populations of *Mll*. Thus, there may be some inclination to recognize the whole Central CA clade, as well as other regional clades as subspecies, instead of the subset of the populations within the clade that have the *Mle* phenotype. Thus, there may be some inclination to recognize the whole Central CA clade, as well as other regional clades within *M. lateralis* as subspecies, instead of the subset of the populations within the clade that have the *Mle* phenotype. However, this approach presents complications of its own for the following reasons:

1. Subspecies by traditional definition are based on phenotype, not genotype (following Darwin 1896; Mayr 1943); molecular data can be used to validate or refute subspecies taxonomy as an indicator of evolutionary lineages, but whether mtDNA variation should be used as a sole criterion to validate classical subspecies is questionable and controversial (see point 2).
2. The use of mtDNA haplotypes as a single line of evidence for recognizing new taxa can be misleading for several reasons (for a detailed review see Ballard and Whitlock 2004):
    a. Sex-biased dispersal: If males and females have different dispersal abilities and behaviors, then nuclear and mtDNA divergence is expected to be incongruent. If males are the more dispersive sex, for example, then we would expect greater signals of genetic differentiation in the mtDNA than in nuclear genetic markers.
    b. Different effective population size: Because the mtDNA genome is a single-strand of DNA (a haploid genome) with maternal inheritance, its effective population size is roughly ¼ that of nuclear markers. Thus genetic drift and fixation of new alleles should be faster in mtDNA, again leading to a greater signal of structure in mtDNA versus nuclear markers.
    c. Selection: Although the mtDNA genome consists of many functional and non-coding gene regions, it is replicated and inherited as a single unit. Therefore, even if the sequenced gene region appears to be evolving neutrally, selection at other sites can potentially pull the rest along. A strong selective sweep has the identical genetic signature of a demographic expansion. These can only readily be distinguished when additional, unlinked, genetic markers are also examined.
    d. Stochasticity in the genetic sorting process: Any single gene genealogy may vary substantially from the actual species history due to random sampling over the coalescent process (Edwards and Beerli 2000).

We noted previously that divergence estimates between the Central and Northern CA clades exceed those between currently recognized species in southern Baja California. Using a phylogenetic species concept (Cracraft 1983), one could argue that each of these clades is a

cryptic species within *M. lateralis*. However, this argument depends on whether one agrees with a phylogenetic concept of species, which in itself is problematic and not universally accepted since it easily leads to taxonomic inflation (Avise 2004; Hudson and Coyne 2002; Isaac et al. 2004), and whether *M. aurigulus* and *M. barbouri* deserve species-level recognition. The two differ subtly in phenotype (and both differ only slightly from *M. lateralis*) and ND4 sequences are currently less than 0.5% divergent. In fact, one of the main reasons for recognizing them as distinct species is because *M. barbouri* occurs on an island in the Gulf of California and will presumably continue to diverge from *M. aurigulus* into the future (Grismer 1999).

This same rationale cannot hold true for divergent mainland populations that do not have obvious physical barriers to gene flow and are not reproductively isolated. Furthermore, there are no rules as to how much genetic or phenotypic divergence it takes to be considered a good species. We draw the comparison between levels of divergence within *M. lateralis* and other *Masticophis* 'species' simply to make the point that the differences among mtDNA clades in California is substantial – our comparison is intended to add a qualitative measure as to how different the Northern and Central CA clades really are. If one uses a phylogenetic species concept to identify island and mainland *Masticophis* species in Mexico, they would almost certainly recognize the Northern, Central and Southern clades as good species as well, given that they are distinctive phylogenetic groups. We do not share either view, but rather recognize that the divergence between the Northern and Central lineages is on par with other distinctive regional lineages in the species. In fact, applying a phylogenetic species concept in this case would rid the taxonomy of *Mle* altogether, since species would be considered the smallest, irreducible monophyletic groups (i.e. any grouping within species do not warrant taxonomic standing, so subspecies are not recognized; Cracaft 1983).

In sum, we recognize the Central CA clade as a well-supported group of mtDNA haplotypes that share common ancestry and that some of those haplotypes belong to snakes that are phenotypically *Mle*. *While mtDNA haplotypes provide useful information about the history of populations within this species, we do not advocate their use as a single line of evidence for drawing new taxonomic boundaries* – clade level boundaries are likely an artifact of historical isolation in this case (due to concordant phylogeographic breaks found in the same areas in other species), and whether patterns of differentiation found in the mtDNA are reflective of the species genome as a whole is unclear. Additional studies focusing on nuclear gene exchange across the mtDNA contact zones would provide a better picture of the degree to which historical population subdivisioning is responsible for the observed patterns in the mtDNA phylogeny, as allelic data from nuclear loci would be better able to detect the effects of recent migration with gene flow.

5.2. Implications of the Alameda whipsnake phenotype

It is unclear why the *Mle* phenotype is so restricted geographically, or more generally, why the stereotypical whipsnake phenotype has diverged in Alameda and Contra Costa Counties and not in other places with similar or greater levels of phylogenetic divergence. At least some of the phenotypic distinctiveness in *Mle* may be due to local, directional natural selection (either contemporary or historical), namely the orange-reddish suffusion over the lighter colored parts of the body, as a number of other taxa have evolved similar coloration in the East Bay (Riemer 1954). Such convergent evolution is often attributed to selection, as the random process of genetic drift is not likely to result in recurrent evolution of similar phenotypes (Endler 1986; Schluter 2000). We also cannot rule out the possibility that stabilizing selection is acting on the more stereotypical *Mll* phenotype throughout most of the range, and that for some reason this selection is relaxed in the Bay Area, allowing those populations to diverge. Finally, some or all of the *Mle* characters may have no adaptive significance at all, and the phenotype may be the

result of genetic drift acting on a small and isolated ancestral population in the East Bay.

We lean towards the former hypothesis of local, directional selection for the following reasons. There is clear evidence of both demographic and geographic range expansion over the past ~25,000 years for the Central CA clade, and the spatial distribution of mtDNA haplotypes indicates that the expansion proceeded from north-to-south. *The* Mle *phenotype likely evolved after whipsnakes moved into the East Bay and became isolated from the Northern CA clade, and continued to remain isolated from the Southern CA clade.* Based on the north-to-south expansion of haplotypes, one might expect the *Mle* phenotype to extend further south along the inner Coast Ranges than it actually does. However, introgression of the *Mle* mtDNA genome in areas of secondary contact with *Mll* may have obscured this expected pattern. In other words, the mtDNA of populations with the *Mle* phenotype has introgressed southward, but the phenotype itself has not moved along in step. We loosely suggest that two interplaying factors may be responsible for this pattern: (1) directional selection historically drove the evolution of the *Mle* phenotype in the East Bay area (or that stabilizing selection was relaxed in this area), promoting divergence from the more stereotypical *Mll* phenotype; and (2) there is currently insufficient reproductive isolation to preclude gene exchange with *Mll* populations occurring further south, and that secondary contact with gene flow is disrupting the spread of the *Mle* phenotype (while allowing mtDNA to move freely) out of its area of origin.

Discordance between mtDNA and phenotypic subspecies boundaries is common, and it is because of this recurrent finding that some researchers believe that subspecies should be disregarded in conservation and invalidated as a taxonomic unit (Harris and Froufe 2005; Ramey et al. 2005; Zink et al. 2000). However, as stated in section 5.1.1, there are several reasons why mtDNA genetic structure may be different from that found within the nuclear genome, and hence the phenotype (for a detailed review see Ballard and Whitlock 2004). We also do not share the view that subspecies should be wholly disregarded in conservation; if the suite of characters that make up the *Mle* phenotype have evolved in response to selective pressures, then the phenotype may represent a legacy of possible adaptive evolution, and in conjunction may be a useful marker for identifying conservation units.

5.3. Population dynamics of East Bay whipsnake populations; patterns and process

Our phylogeographic analyses provided an important, historical context for the population-level inquiries that specifically targeted whipsnakes in Alameda and Contra Costa Counties. With knowledge of the deeper evolutionary history of the species in hand, we can then use the population-level analyses to better understand how more contemporaneous evolutionary processes, namely secondary contact with gene flow, have rewound the changes that previously built up when the Central CA clade was isolated from other *Mll* clades to the north and south.

This study, the first to conduct a genetic analysis of *Mle*, relied on concentrated sampling within designated critical habitat areas to obtain estimates of overall population structuring, demography and intrapopulation genetic diversity. We were limited to this sampling approach because the bulk of our tissues from Alameda and Contra Costa Counties were collected during monitoring surveys by Swaim Biological Inc. over the past 10 years, versus implementation of a specific sampling design for this study. Nonetheless, this material and sampling scheme enabled us to view the genetic landscape over much of the East Bay and pin down critical information about the population genetics of the subspecies. Our results also highlight specific areas where future genetic studies should perhaps take place, such as those representing transitional areas between the major clusters identified in this study. The detection of actual clines would provide useful information about the effective distance at which snakes (and therefore genes) move between core habitat areas, and where the most important areas of gene exchange exist. In turn,

managers may want to focus their efforts on conserving the habitat in these particular places.

Our inferences of overall genetic structuring in the East Bay are likely a reflection of not only true biological artifacts of the snake, namely dispersal ability, but also of sampling artifacts. Bayesian assignment models revealed substantial genetic partitioning by critical habitat units, with few individuals showing evidence of mixed ancestry, and fitting certain models with and without spatially structured priors (i.e. $\psi = 0.6$ versus 0.0, respectively) showed that short distance, local interactions did not help in explaining genetic structure[4]. In fact, we detected little evidence for clinal transitions in allele frequencies or extensive co-ancestry within individuals, even though a strong overall pattern of genetic isolation by distance was recovered in both the STR and mtDNA datasets. Because of limited admixture among clusters and a high estimated proportion of non-immigrants within our sampling sites, it is possible that the dispersion of our sampling exceeded the distance at which snake movement between sites would result in strong, identifiable admixture. Even in the Camp Ohlone and Del Valle populations, which displayed increased admixture relative to other populations in the recovery units, the sampling was so geographically clustered that there was minimal evidence of a true cline in the assignment plots. We suspect that as we fill in sampling gaps, allelic clines will begin to surface between the major clusters we have identified here.

We draw attention to our results that show mixed ancestry of individuals belonging to the Camp Ohlone and Del Valle populations; this strongly suggests that there is genetic linkage between Units 3 and 5 (Fig. 9). It is notable that there is so little admixture between whipsnakes in Unit 5 (particularly those from Del Valle and S300) and those in Unit 4 directly to the north; rather, gene flow appears to occur from east-to-west far more frequently than north-to-south within Unit 5 (Fig. 1). In fact, both the STR and mtDNA data indicate that there is unique variation in the southeastern corner of Alameda County (and into San Joaquin and Santa Clara Counties), and that these populations have diverged to a greater extent than elsewhere in the East Bay. This result is seen in the mtDNA NCPA analysis (clade 2-2; Fig. 6), the STR cluster assignment plots (Figs. 8-9), and in the number of private alleles.

The discontinuity of gene exchange between Units 4 and 5 is likely explained by the lack of high-quality chaparral/scrub habitat in the Livermore Valley. Currently, there is approximately a 10.0 mile gap separating the edges of suitable habitat within the respective recovery units, and the closest known *Mle* occurrences are between the Mt. Diablo area in Unit 4 and Sunol-Cedar Mountain area in Unit 5 (B. Solvesky USFWS, pers. comm.). The lack of any chaparral or scrub vegetation within the Livermore Valley, combined with the absence of gene flow between the recovery units, suggests that the Livermore Valley may constitute a longstanding historical barrier, rather than a more recent barrier created by human activity. This disconnect is further exacerbated by the 129 km stretch of Interstate 580 (I-580), which is a heavily traveled spur route of I-80 that connects the San Francisco Bay area with I5 in the Central Valley, and bisects Units 4 and 5; however, because of its young age (built 1963) the road would have to impose a severe movement barrier for its effects to surface in the genetics of these populations over such a short span of evolutionary time (see section 5.5.3 for additional discussion about road effects).

The genetic admixture that we did detect suggests that genes are moving out of certain areas, but not back into them to the same extent. The best example is the clustering assignment for populations in Unit 3, which are largely genetically exclusive; however, Unit 3 genes are

[4] This does not mean that such spatial interactions do not exist in the wild; it simply means that these interactions were not detected to the point where they were informative for the clustering algorithm, given our current data and sampling regimen.

apparently making their way north and south, as evidenced in the admixture plot for Tilden/Claremont and Camp Ohlone/Del Valle, respectively. More specifically, the predominant colors identifying Units 1 and 5 in the admixture plot are not observed in Unit 3, whereas the identifying color for Unit 3 is clearly visible in the most geographically proximate populations residing in Units 1 (Tilden/Claremont) and 5 (Camp Ohlone/Del Valle). It is interesting that snakes are apparently moving out of this part of Alameda County but not in, because corridor areas provide opportunities for dispersal in either direction. To add to this pattern of asymmetric migration, phenotypic traits are ostensibly transitional between *Mle* and *Mll* in this same general area (i.e. between Units 3 and 5; Swaim 1998), with a decreased frequency of *Mle* traits to the southeast. Taken together, this 'breakdown' of phenotypic traits, evidence of asymmetric migration out of Unit 3, divergent mtDNA haplotypes within Unit 5, and the complete lack of allele sharing between Units 4 and 5, provides evidence of secondary contact between two formerly isolated groups within *M. lateralis*.

5.3.1 Genetic diversity status of East Bay populations

Although there is clear genetic structuring among the 12 East Bay sampling sites, we did not detect significant differences in allelic richness, gene diversity, or levels of inbreeding among the critical habitat units. In fact, the only significant difference we detected, aside from differences in allele frequencies and cluster assignment among sites, were in the estimates of private allelic richness between the northern populations in Unit 1 and the southern populations in Unit 5. Based on the increased frequency of private alleles in the southern part of our sampling area, combined with a less distinctive *Mle* phenotype over the same general region, it is quite possible that our sampling reached an interface with a historically distinctive cluster that extends further south in Santa Clara County, as described in the previous section. This could be interpreted as a subtle genetic break within Unit 5, and provides some evidence that the phenotypic boundaries of the two subspecies align with the boundaries identified by our neutrally evolving genetic markers.

While our results indicate substantial genetic structuring by habitat unit, we found no evidence for recent reduction in the effective sizes of *Mle* populations. At face value, this suggests that recent habitat loss and fragmentation has not had any real, deleterious genetic effects on these populations. However, it is important to consider the time lag that nearly always takes place between the ecological and genetic effects of anthropogenic disturbance and the onset of the disturbance itself (with the possible exception of all individuals in a single population dying off in one fell-swoop, which itself is unlikely). Such inertia is dictated by several key factors, including population size, generation time, and severity of the disturbance (Wiens 1997; Wright 1978; Young and Clarke 2000). In general, small populations with short generation times will show more immediate effects than larger populations with longer generation times because genetic variability decreases at a constant rate of $1(2/N_e)$ per generation in a closed population (Wright 1978). In other reptile species, it has been demonstrated that 50 or more years can pass before even mild genetic effects are detectable in genome-wide, neutrally evolving genetic markers (Driscoll and Hardy 2005; Richmond et al. 2008; Sumner et al. 2004), but again this lag period is determined by a number of factors that are intrinsic to particular organisms. We mention this here as a cautionary note because researchers often evaluate the genetic status of populations based on data that represent a snapshot in time, while ignoring inertia in the effects of habitat disturbance on ecological and genetic processes.

5.4. Utilizing genetic data in defining *Mle* managements units

One of the goals of this study is to offer recommendations for appropriate management units based on the genetic data. In a recent review focused on defining subspecies and distinct population segments for conservation, Haig et al (2006) recommend that at least two separate criteria be examined in assessing the discreteness of any particular 'population' or 'unit' in relation to the remainder of the species (e.g. molecular, morphological, ecological, behavioral). In keeping with this theme, we present three management options below, discuss their strengths and weaknesses, and highlight the approach that we believe represents a sound strategy for conservation. Our suggestions are based on the data in hand and will no doubt require modification as more information is gathered. However, our main findings are unlikely to change drastically as new genetic information is added to dataset, and the data in hand provide a solid framework on which to base upcoming management decisions in 2011. Our approach below reconciles the information gained from all forms of data, capitalizes on existing knowledge and previous work conducted on *Mle*, and emphasizes the preservation of genetically and phenotypically distinctive populations. Most importantly, it targets the critical units in need of conservation, a basic tenet of the U.S. Endangered Species Act (1973).

Option A – Continue with the current strategy of protecting all populations that possess the *Mle* phenotype and retain the current management units. ***Strengths***: (1) A conceptually straightforward management plan is already in place, and that is, snakes possessing the 'Riemer characters' (1954) belong to a unique subset of populations that should be protected; (2) the *Mle* phenotype might be adaptive, and therefore by preserving populations that possess the phenotype, we may be protecting individuals that are most likely to survive. ***Weaknesses***: (1) The subspecies boundaries are difficult to define without further study – our present knowledge of the boundaries rests on the intuition of a few seasoned investigators and unpublished data (e.g. Swaim 1998; manuscripts in prep by Swaim Biological Inc.), but not on any peer-reviewed study that has quantitatively evaluated the limits of the *Mle* distribution. Furthermore, it is unclear how to taxonomically treat snakes with intermediate phenotypes, which are common in certain areas; (2) it is unknown if the *Mle* phenotype is adaptive, as no tests of fitness or performance have ever been conducted; (3) the sole reliance on phenotype as the management unit ignores potential genetic diversity that may be critical for the long-term survival of the species.

Option B – Consider the Central CA mtDNA lineage as the management unit, and focus conservation efforts on all populations included within this lineage. ***Strengths***: (1) It is straightforward to sequence the ND4 gene and identify the mtDNA lineage membership of any whipsnake; (2) possibly captures a larger subset of the total evolutionary diversity within the species than Option A; (3) protects populations within the Central CA lineage that are already recognized as *Mle*. ***Weaknesses***: (1) Relies on a single marker type to identify management units, and for reasons discussed in Section 5.1.1, this could potentially create unreliable flagships for conservation; (2) ignores useful information on phenotype and obfuscates the traditional definition of *Mle*; (3) the rationale of identifying an mtDNA lineage as a 'subspecies' would require that all other mtDNA lineages be recognized as subspecies without any firm knowledge of the phenotypic variation within and among those clades; (4) treats all populations within the clade as 'equal', even though certain populations may harbor more genetic diversity than others and are therefore more appropriate targets for conservation; (5) inflates taxonomy in a potentially non-productive way.

Option C – Consider the management unit as the most geographically cohesive set of populations that forms a distinctive or closely related STR clusters, contains mtDNA haplotypes with the fewest number of steps removed from the common ancestral haplotype for the Central CA clade, and possess the *Mle* phenotype. ***Strengths***: (1) This approach does not rely on any single form of data, but instead incorporates information from the both mitochondrial and nuclear genomes, as well as morphology (and ideally ecology as well); (2) delimits a management unit that captures evolutionary diversity while preserving the genetic legacy of the species; (3) combines traditional and contemporary approaches to management, where traditional use of the phenotype as an indicator of the management unit is incorporated with a conservation strategy that seeks to preserve evolutionary processes. ***Weaknesses*** (1) Possibly more labor intensive and more costly, but only in the initial stages of dataset development; (2) requires a greater integration of datasets and investigator expertise.

Option C is an amalgam of Options A and B, but also incorporates information from the nuclear genome. By targeting populations with distinctive mtDNA and nuclear genetic variation and to some degree the *Mle* phenotype, we can logically assign hierarchical conservation priorities to different whipsnake populations within the Central CA clade, which can be viewed as the most inclusive management unit. Under Option C, priorities could be arranged according to the following strategy:

Priority 1: Priority 1 status would be assigned to populations occurring within critical habitat units 1, 3, and 4, and presumably units 2 and 6 (Fig. 9). Bayesian assignment tests show that STR alleles among these units form distinctive but closely related clusters, and populations within them show a high degree of allelic admixture. Mitochondrial haplotypes are exclusive to clade 2-1 in the haplotype network (Fig. 6), which falls within the more inclusive Central CA clade. In the phylogenetic tree, these populations are represented by the following individuals: ALMDA1-3, 11-16; CNTR1-8, 10-29, 31-32 (Fig. 12); note that the haplotypes belonging to these individuals are more removed from the tree tips than those associated with Priority 2 populations below, and therefore more closely resemble the common ancestral haplotype for the Central CA clade. By targeting the ancestral 'type' for conservation, our presumption is that we would be preserving the original variation that ultimately gave rise to the diversity that exists today. The phenotype of these snakes is stereotypical *Mle*.

Priority 2: Priority 2 status would be assigned to Central CA populations in which at least two types of data point to a common, distinctive group that lies in the putative intergrade area between *Mll* and *Mle*. These populations are located in the eastern section of Unit 5 in Alameda and Santa Clara Counties, where STR alleles form a distinctive cluster and mtDNA haplotypes have diverged more from the common ancestor of the Central CA clade than other haplotypes within the group (Fig. 12). There is also a higher prevalence of private STR alleles in this area, which may indicate that we sampled in to the northern margin of a new genetic cluster. Snakes in this region have relatively few *Mle* characters as described by Reimer (1954) and tend toward the *Mll* phenotype. In the phylogenetic tree, individuals belonging to these populations are identified by the following labels: ALMDA4-10, 17-21; SNJQN1, 4-5; SNCLR1-4 (Fig. 12). These same populations are represented by clade 2-2 in the haplotype network (Fig. 6). Because the sequences differ by relatively few substitutions, the network provides a clearer representation of Priority 2 populations, and it is possible to see which one-step mutational differences are most similar.

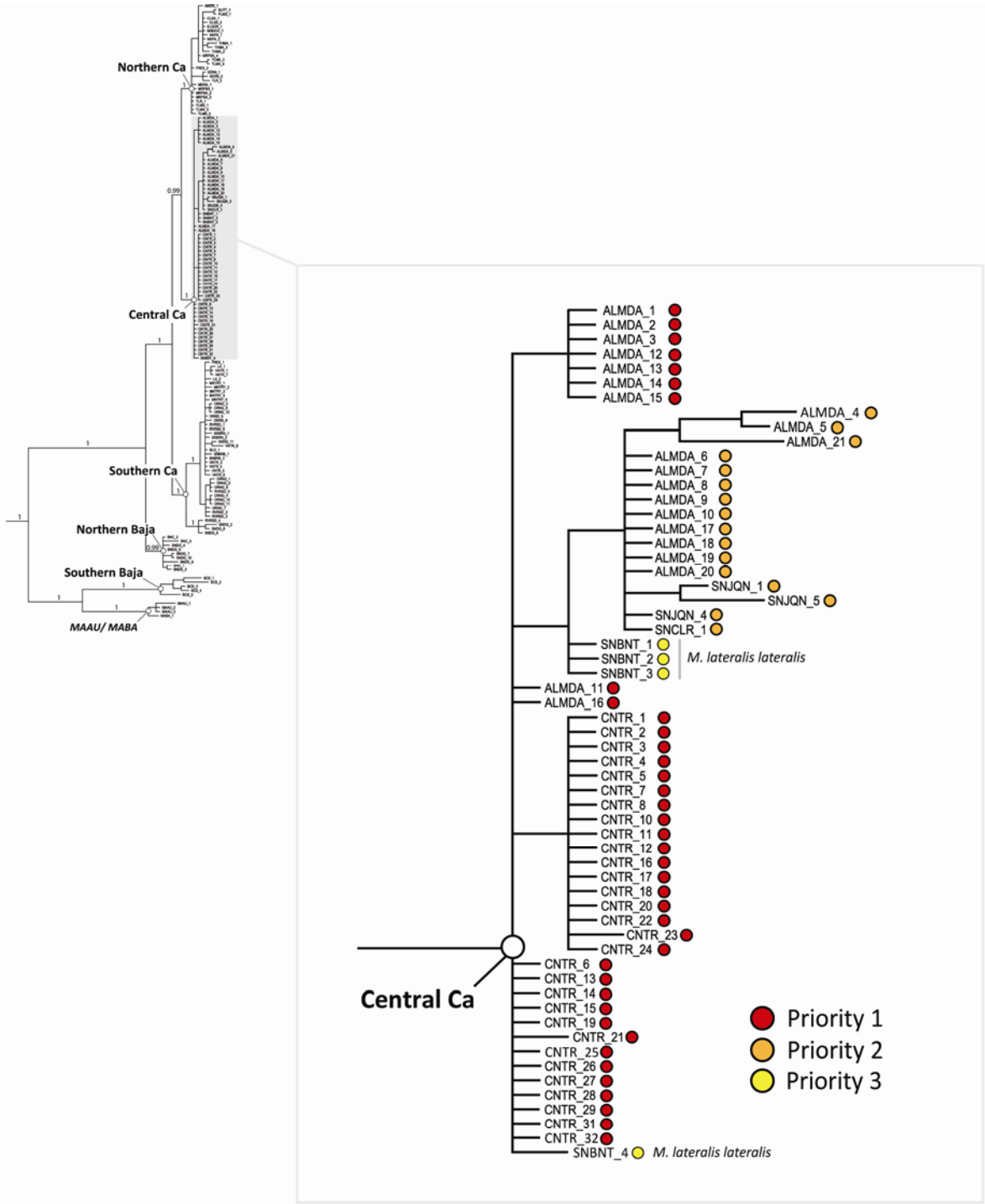


Figure 12. Central CA clade pruned from the full mtDNA phylogeny, shown in the upper left. Colors represent different conservation priorities assigned to individuals (i.e. locations), following the text in Section 5.4. Priority 3 populations are phenotypically *M. lateralis lateralis*, whereas all others are *M. lateralis euryxanthus.*

Priority 3: Priority 3 status would be assigned to populations with the *Mll* phenotype found well south of the Alameda-Santa Clara County border (haplotypes 6 & 7, Fig. 6; SNBNT1-4, Fig. 12). Although these populations are still within the Central CA clade, they contain a mixture of ancestral and clade 2-2 haplotypes. Based on our demographic analyses, this pattern is probably reflective of historical range expansion and introgression of East Bay area mitochondria. We do not have STR data for these populations, so we are unable to assess whether they assign to the same cluster as individuals collected from Unit 5 at this time. We are currently gathering this data, and some populations may need to be reassigned as Priority 2 in light of new information.

Because Priority 1 sites contain the most concordant genetic and phenotypic variation, emphasis should be placed on maintaining connectivity between these sites and others in Alameda, San Joaquin, and Santa Clara Counties. Such a strategy would assist in preserving the microevolutionary processes identified in this region (i.e. probably secondary contact with gene flow). In addition, our recommended framework does not include a threat assessment for these populations, namely habitat quality. Although most of the distribution of clade 2-1 (Fig.6) falls within currently designated critical habitat and recovery units, a reassessment of habitat quality in this area would be warranted if the genetic aspects of our proposal are implemented. We stress that the genetic data constitute one piece of the puzzle needed to successfully manage these distinctive population segments; the findings presented here should be used in conjunction with other ecological, behavioral, and demographic data to arrive at the most effective ways to do this.

5.5. Additional research for building a comprehensive *Mle* management plan

Our work is a first-attempt at characterizing the population genetics of whipsnake populations in the East Bay area, and our field sampling was constrained by available material gathered from previous monitoring surveys. Because of the spatial characteristics of our sampling over the total area, new questions have emerged as a result of this work. Below, we outline some of these new research questions and highlight several older, outstanding issues for which further investigation would provide useful information for future management of this snake.

5.5.1: What are the best estimates of southern and northern genetic boundaries for Mle?

We focused our STR data collection and analyses on East Bay Area populations identified as phenotypically 'pure' *Mle*, 'transitional' *Mle*/*Mll*, and at least one that is *Mll,* and restricted our field efforts to areas south of Suisun Bay. This allowed us to examine population structuring across a predominantly north/south transect along the northern most reach of the Diablo Range, and into an area where the two subspecies are presumed to intermingle; however, our ability to detect a more definitive border, if one exists, would improve with further sampling in eastern Alameda, Santa Clara and San Benito Counties to the south (i.e. 30 individuals per site with at least some interspersed sampling), and allow us to tease out whether Unit 5 populations truly represent a transitional area. In a true contact zone in which gene flow between divergent population assemblages is restricted either now or in the recent past, we would expect the admixture coefficients for the Bayesian assignment tests to have a sigmoidal shape across space, similar to theoretical predictions for allele frequency curves in a hybrid zone (Barton and Hewitt 1985). A steep curve would suggest a hard boundary with very limited gene flow, whereas a gradual curve would indicate gene flow with isolation by distance.

In addition, population dynamics remain poorly understood at the extreme northern edge of the *Mle* range in Contra Costa, Napa, and Solano Counties. Our data suggest that there is a sharp genetic boundary in southern Napa County, presumably due to water barriers created by

the San Joaquin and Sacramento Rivers and Suisun and San Pablo Bay; however, we sampled only two individuals in southern Napa County, and we have yet to recover sympatric Northern and Central clade haplotypes (if they exist). Furthermore, the Central/Northern clade boundary is identified by mtDNA haplotypes only, providing an incomplete picture of genetic structuring in this region. Further studies would allow us to assess the degree to which water barriers created by the San Joaquin and Sacramento Rivers and Suisun and San Pablo Bay have influenced the biogeographic history of *Mle* populations in northern Contra Costa County. For example, we could test whether genetic variation in the northern part of the subspecies range is more abruptly transitional than to south, as might be expected given the water barrier, which in turn would provide an indication of how rapidly shared ancestry is lost at the level of the genome in these snakes.

5.5.2: Are individual snakes more genetically admixed in the 'in-between' areas where we currently lack samples?

Our current sampling, although extensive within four core areas, is patchy over the entire *Mle* distribution. Thus, with increased population sampling in areas that fall in between these core areas, it would be possible to test whether individuals are more genetically admixed over smaller spatial scales. The dispersion of our samples and the strong pattern of IBD in the current data set no doubt contribute to the appearance of genetically divergent groups among Recovery Units, but it is quite possible that we have sampled beyond the distance at which migration can reveal itself in the admixture coefficients. If detected, increased admixture in linkage areas would suggest that more gene exchange is occurring between habitat units than can be appreciated under the current sampling. If strong population structuring is still apparent even after filling in these gaps, managers would need to reassess the factors that may be driving this divergence (i.e. insufficient corridors between units), or whether population sizes within units are sufficient to withstand the potentially adverse effects of genetic drift.

5.5.3: Do roadways constitute significant barriers to migration and gene exchange?

Roads are have been suggested to contribute to *Mle* population declines in recent decades, yet this presumption is not based on any rigorous quantification of how roads or vehicular traffic might be affecting genetic connectivity. For both the mtDNA and STR datasets in this study, our results vary in a direction that suggests that roads do in fact influence population structuring in these snakes, but the distances between sampling sites prevented us from definitively disentangling the effects of roads versus genetic isolation-by-distance. We also cannot rule out the possibility that roads exist in areas that may have constituted natural barriers to begin with; however, roads would likely make such areas even less 'porous' than they were initially, and more intensive, systematic sampling closer to major road ways would allow us to more accurately test the degree to which they alter migration and gene flow (as has been shown recently in other protected snakes; Clark et al. 2010). For example, it would be possible for us to correlate genetic differentiation between sites with various road statistics, such as number of vehicles per unit time, width and length of the road, age of the road, and whether there are significant medians or underpasses.

5.5.4: Where are the phenotypic boundaries between* Mle *and* Mll*?

No quantitative analysis of the phenotypic subspecies boundaries has ever been published, and ideas about where these boundaries exist are largely intuitive and stem from unpublished reports or manuscripts. Reimer (1954) highlights eight characters that purportedly distinguish *Mle* from the more typical *Mll*; we propose using the current information on genetic and phenotypic

boundaries to guide a detailed, transect study of these eight characters in southern Alameda County and into Santa Clara County. Swaim Biological Inc. has collected meristic character data for snakes occurring in Contra Costa and Alameda Counties, and also from museum specimens sampled throughout California (manuscript in prep). However, data for one of the most potentially informative characters, color hue and degree of color suffusion over the total body area, are lacking. Robertson and Robertson (2008) developed a method for quantifying attributes of color hue and pattern in amphibians that can be easily translated to the whipsnake system. It involves the use of digital photos of the animal on a standardized color background, and uses imaging software to produce continuously-valued character data that can be analyzed in a statistical framework (Robertson and Robertson 2008). If sufficient sample sizes could be obtained in the vicinity of the Alameda/Santa Clara County border, it would be possible to perform statistical tests of correlated spatial patterns between phenotype and STR/mtDNA data.

The discovery of previously unrecognized phylogenetic diversity within *M. lateralis* also suggests that further investigation of phenotypic variation throughout the range of *Mll* is warranted. We know from personal field experience that undescribed variation exists in localized areas, and it would now be possible to test whether that variation corresponds to particular genetic subgroups within the three main California clades.

6. Acknowledgments

Funding for this project was provided by the USFWS Science Support Program and from the USFWS's Central Valley Project Improvement Act Habitat Restoration Program. This work would have not been possible without the field assistance and expertise of Karen Swaim and members of Swaim Biological Inc., who collected the bulk of the tissue samples for the genetic work on East Bay populations. For additional tissue collections, we thank Michael Bernard and the University of California Natural Reserve System's Quail Ridge Reserve, Brad Hollingsworth and the San Diego Museum of Natural History, Paul Johnson, Jimmy McGuire and the Museum of Vertebrate Zoology, Milan Mitrovich, Robert Murphy and the Royal Museum of Ontario, Greg Pauly, Jens Vindum and the California Academy of Sciences, Michael Westphal, and Kevin Wiseman. The report benefited greatly from comments by Kelly Barr, Jeffery Markert, Megan Lahti, MaryEllen Mueller, Josh Hull, Ben Solvesky, and Karen Swaim. Views expressed in this project summary are the authors' and do not necessarily reflect the view of the U.S. Geological Survey.

7. Literature Cited

Arevalo, E., S. K. Davis, and J. W. Sites, Jr. 1994. Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. Systematic Biology 43:387-418.

Avise, J. C. 2004. Molecular markers, natural history, and evolution. 2nd edition. Sinauer, Sunderland MA.

Ball Jr., R. M., and J. C. Avise. 1992. Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. Auk 109:626-636.

Ballard, J. W. O. and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744.

Barton, N. H., and G. Hewitt. 1985. Analysis of hybrid zones. Annual Review of Ecology and Systematics 16:113-148.

Berry, A., and M. Kreitman. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. Genetics 134:869-893.

Bocquet-Appel, J. P., and R. R. Sokal. 1989. Spatial autocorrelation analysis of trend residuals in biological data. Systematic Zoology 38:333-341.

Bohonak, A. J. 2002. IBD (Isolation by distance): A program for analysis of isolation by distance. Journal of Heredity 93:153-154.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5:453-455.

Burbrink, F. T., F. Fontanella, R. A. Pyron, T. J. Guiher, and C. Jimenez. 2008. Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes: Colubridae: *Coluber constrictor*). Molecular Phylogenetics and Evolution 47:274-288.

Burbrink, F. T., R. Lawson, and J. B. Slowinski. 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. Evolution 54:2107-2118.

Chen, C., E. Durand, F. Forbes, and O. Francois. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes 7:747-756.

Clark, R. W., W. S. Brown, R. Stechert, and K. R. Zamudio. 2010. Roads, interrupted dispersal, and genetic diversity in Timber Rattlesnakes. Conservation Biology 24:1059-1069.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 10:1657-1660.

Cornuet J.M. and Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

Cracraft, J. 1983. Species concepts and speciation analyses. Current Ornithology 1: 159-187.

Crawford, N. G. 2009. SMOGD: software for the measurement of genetic diversity. Molecular Ecology Resources 10:556-557.

Cressie, N. A. C. 1993, Statistics for spatial data. Wiley, New York.

Darwin, C. 1896. The variation of animals and plants under domestication. D. Appleton, New York.

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148:1269-1284.

Driscoll, D. A., and C. M. Hardy. 2005. Dispersal and phylogeography of the agamid lizard *Amphibolurus nobbi* in fragmented and continuous habitat. Molecular Ecology 14:1613-1629.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:699-710.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular Biology and Evolution 22:1185-1192.

Dupré, W. R., R. B. Morrison, H. E. Clifton, K. R. Lajoie, and D. J. Ponti. 1991. Quaternary geology of the Pacific margin, Pages 141-214 *in* R. B. Morrison, ed. Quaternary Nonglacial Geology; Conterminous US. Geological Society of America, Boulder, CO.

Durand, E., F. Jay, O. E. Gaggiotti, and O. François. 2009. Spatial inference of admixture proportions and secondary contact zones. Molecular Biology and Evolution 26:1963-1973.

Edwards, S. V. and P. Beerli. 2002. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.

Endler, J. A. 1986, Natural selection in the wild. Princeton University Press, Princeton.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992a. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial-DNA restriction data. Genetics 131:479-491.

—. 1992b. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Feldman, C. R., and G. S. Spicer. 2006. Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. Molecular Ecology 15:2201-2222.

François, O., S. Ancelet, and G. Guillot. 2006. Bayesian clustering using hidden Markov random fields in spatial population genetics. Genetics 174:805-816.

Frankel, O. H. 1974. Genetic conservation: our evolutionary responsibility. Genetics 78:53-65.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Gavin, T. A., P. W. Sherman, E. Yensen, and B. May. 1999. Population genetic structure of the northern Idaho ground squirrel (*Spermophilus brunneus brunneus*). Journal of Mammalogy 80: 156-168.

Garrick, R. C., J. D. Nason, C. A. Meadows, and R. J. Dyer. 2009. Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. Molecular Ecology 18:1916-1931.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversity and fixation indices (version 2.9.3) available from www.unil.ch/izea/softwares/fstat.html.

Grismer, L. L. 1999. An evolutionary classification of reptiles on islands in the Gulf of California, México. Herpetologica 55:446-469.

Grismer, L. L. 2002, Amphibians and Reptiles of Baja California, including its Pacific Islands and the Islands in the Sea of Cortes. University of California Press, Berkeley.

Haig, S. M., E. A. Beever, S. M. Chambers, H. M. Draheim, B. D. Dugger, S. Dunham, E. Elliott-Smith, J. B. Fontaine, D. C. Kesler, B. J. Knaus, I. F. Lopes, P. Loschl, T. D. Mullins, and L. M. Sheffield. 2006. Taxonomic considerations in listing subspecies under the US Endangered Species Act. Conservation Biology 20:1584-1594.

Hall Jr., C. A. 2002. Nearshore marine paleoclimatic regions, increasing zoogeographic provinciality, molluscan extinctions, and paleoshorelines, California: Late Oligocene (27 Ma) to late Pliocene (2.5 Ma). Geological Society of America Special Papers 357:1-489.

Harpending, H. 1994. Signature of ancient population growth in a low resolution mitochondrial mismatch distribution. Human Biology 66:131-137.

Harris, J. D. and E. Froufe. 2005. Taxonomic inflation: species concept or historical geopolitical bias? Trends in Ecology and Evolution 20:6-8.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633-1638.

Hedrick, R. W. and P. S. Miller. 1992. Conservation genetics; techniques and fundamentals. Ecological Applications 2:30-46.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38:409-414.
Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367-380.
Holman, J. A. 2000, Fossil Snakes of North America: Origin, Evolution, Distribution, Paleoecology. Indiana University Press, Indianapolis.
Hudson, R.R. and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557-1565.
Huelsenbeck, J. P., and F. Ronquist. 2001. Bayesian inference of phylogenetic trees. Bioinformatics 17:1754-1755.
Isaac, N. J., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology and Evolution 19:464-469.
Jacobs, D. K., T. A. Haney, and K. D. Louie. 2004. Genes, diversity, and geologic processes on the Pacific coast. Annual Review of Earth and Planetary Sciences 32:601-652.
Jakobsson, M., and N. A. Rosenberg. 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, M. R. 1983. *Masticophis lateralis* (Hallowel), Striped racer. Catalogue of American Amphibians and Reptiles. Society for the Study of Amphibians and Reptiles.
Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service. BMC Genetics 6:v.3.16 http://ibdws.sdsu.edu/
Jockusch, E. L., and D. B. Wake. 2002. Falling apart and merging: diversification of slender salamanders (Plethodontidae: *Batrachoseps*) in the American West. Biological Journal of the Linnean Society 76:361-391.
Jost, L. 2008. Gst and its relatives do not measure differentiation. Molecular Ecology 17:4015-4026.
Kalinowski, S. T. 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5:187-189.
Knowles, L. L. 2008. Why does a method that fails continue to be used? Evolution 62:2713-2717.
Kuchta, S. R, and A. Tan. 2006 Lineage diversification on an evolving landscape: phylogeography of the California newt *Taricha torosa* (Caudata: Salamandridae).
Lande, R. 1994. Risk of population extinction from fixation of new deleterious mutations. Evolution 48: 1460-1469.
Leaché, A. D., S. C. Crews, and M. J. Hickerson. 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biology Letters 3:646-650.
Leaché, A. D., M. S. Koo, C. L. Spencer, T. J. Papenfuss, R. N. Fisher, and J. A. McGuire. 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). Proceedings of the National Academy of Sciences 106:12418-12423.
Lee, M. S. Y. 1999. Molecular clock calibrations and Metazoan divergence dates. Journal of Molecular Evolution 49:385-391.
Legendre, P., and L. Legendre. 1988, Numerical Ecology. Elsevier, New York.
Lesica, P., and F. W. Allendorf. 1995. When are peripheral populations valuable for conservation? Conservation Biology 9:753-760.
Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 73:445-463.
Luikart, G. and J. M. Cornuet. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology 12(1):228-237.
Malécot, G. 1948, Les Mathématiques de l'Hérédité Masson and Cie, Paris.
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.
Mayr, E. Animal species and evolution. Belknap Press, Cambridge MA.
Meffe, G. K., and C. R. Carroll. 1994, Principles of Conservation Biology. Sunderland and Associates, Sunderland, MA.
Mitrovich, M. J. 2006. A case study in conservation science: The spatial ecology and evolutionary history of the coachwhip snake (*Masticophis flagellum*) and striped Racer (*M. lateralis*). Unpublished Ph.D. dissertation, University of California at Davis and San Diego State University.
Moritz, C. 1999. Conservation units and translocations: Strategies for conserving evolutionary processes. Hereditas 130:217-228.
Mulcahy, D. G. 2008. Phylogeography and species boundaries of the western North American Nightsnake (*Hypsiglena torquata*): Revisiting the subspecies concept. Molecular Phylogenetics and Evolution 46:1095-1115.
Nagy, Z. T., R. Lawson, U. Joger, and M. Wink. 2004. Molecular systematics of racers, whipsnakes and relatives (Reptilia: Colubridae) using mitochondrial and nuclear markers. Journal of Zoological Systematics and Evolutionary Research 42:223-233.
Nei, M., and R. K. Chesser. 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47:253-259.
Nylander, J. A. A., J. C. Wilgenbusch, D. L. Warren, and D. L. Swofford. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581-583.
Panchal, M., and M. A. Beaumont. 2007. The automation and evaluation of nested clade phylogeographic analysis. Evolution 61:1466-1480.
Peakall, R., and P. E. Smouse. 2006. GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.
Petit, R. J. 2008. The coup de grâce for the nested clade phylogeographic analysis? Molecular Ecology 17:516-518.
Phillimore, A. B., and I. P. F. Owens. 2006. Are subspecies useful in evolutionary and conservation biology? Proceedings of the Royal Society B: Biological Sciences 273:1049-1053.
Posada, D., K. A. Crandall, and A. R. Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Molecular Ecology 9:487-488.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.
Pybus, O. G., and A. Rambaut. 2002. GENIE: estimating demographic history from molecular phylogenies. Bioinformatics 18:1404-1405.
Rambaut, A., and A. J. Drummond. 2005, Tracer - v.1.3. http://evolve.zoo.ox.ac.uk/software.html?id=tracer.
Ramos-Onsins, S. E., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution 19:2092-2100.

Ramey, R. R. H., H-P Liu, C. W. Epps, M. Carpenter, and J. D. Wehausen. 2005. Genetic relatedness of the Preble's meadow jumping mouse (*Zapus budsonius preblei*) to nearby subspecies of *Z. budsonius* as inferred from variation in cranial morphology, mitochondrial DNA and microsatellite DNA: implications for taxonomy and classification. Animal Conservation 8: 329-346.

Ray, N., M. Currant, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. Molecular Biology and Evolution 20:76-86.

Richmond, J. Q., and T. W. Reeder. 2002. Evidence for parallel ecological speciation in scincid lizards of the *Eumeces skiltonianus* species group (Squamata: Scincidae). Evolution 56:1498-1513.

Richmond, J. Q., D. T. Reid, K. G. Ashton, and K. R. Zamudio. 2008. Delayed genetic effects of habitat fragmentation on the ecologically specialized Florida sand skink (*Plestiodon reynoldsi*). Conservation Genetics 10:1281-1297.

Riddle, B. R., D. J. Hafner, L. F. Alexander, and J. R. Jaeger. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. Proceedings of the National Academy of Sciences of the United States of America 97:14438-14443.

Riemer, W. J. 1954. A new subspecies of the snake *Masticophis lateralis*. Copeia 1954:45-48.

Robertson, J. M., and A. D. Robertson. 2008. Spatial and temporal patterns of phenotypic variation in a Neotropical frog. Journal of Biogeography 35:830-843.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9:552-559.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137-138.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145:1219-1228.

Sanders, K. L., and M. S. Y. Lee. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biology Letters 3:275-279.

Schluter, D. 2000, The Ecology of Adaptive Radiation. Oxford University Press, Inc., New York.

Shaffer, H. B., G. B. Pauly, J. C. Oliver, and P. C. Trenham. 2004. The molecular phylogenetics of endangerment: cryptic variation and historical phylogeography of the California tiger salamander, *Ambystoma californiense*. Molecular Ecology 13:3033-3049.

Slatkin, M. 1993. Isolation by distance in equilibrium and non equilibrium populations. Evolution 47:264-279.

Spiegelhalter, D. J., N. G. Best, and B. P. Carlin. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B 64:1-34.

Spinks, P. Q., and H. B. Shaffer. 2005. Range-wide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. Molecular Ecology 14:2047-2064.

Stebbins, R. C. 2003, Field Guide to Western Reptiles and Amphibians. Houghton Mifflin Company, New York.

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

Swaim, K. E. 1994. Aspects of the Ecology of the Alameda Whipsnake (*Masticophis lateralis euryxanthus*); unpublished Master's thesis, California State University Hayward.

Swaim, K. E. 1998. Results of a live trapping survey for the Alameda whipsnake (*Masticophis lateralis euryxanthus*) at the Site 300 facilities of the Lawrence Livermore National Laboratory. Prepared for the Deputy Direction, Site 300. 18pp+ appendices (unpubl).

Swofford, D. L. 1991.PAUP: Phylogenetic analysis using parsimony, v3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Templeton, A. 1986. Coadaptation and outbreeding depression *in* M. E. Soulé, ed. Conservation Biology. The Science of Scarcity and Diversity Sinauer Associates, Sunderland, MA.

Templeton, A. R. 1995. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*. Genetics 140:767-782.

—. 1998. Species and Speciation: Geography, Population Structure, Ecology, and Gene Trees *in* D. J. Howard, and S. H. Berlocher, eds. Endless Forms, Species and Speciation. Oxford University Press, Oxford.

—. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Molecular Ecology 13:789-809.

—. 2008. Nested clade analysis: an extensively validated method for strong phylogeographic inference. Molecular Ecology 17:1877-1880.

—. 2010. Coalescent-based, maximum likelihood inference in phylogeography. Molecular Ecology 19:431-435.

U.S. Fish and Wildlife Service. 1985. Review of Vertebrate Wildlife; Notice of Review. Federal Registrar 50:37958-37967.

U.S. Fish and Wildlife Service. 1994. Proposed Endangered Status for the Callippe Silverspot Butterfly, Behren's Silverspot Butterfly, and the Alameda Whipsnake From Northern and Central California. Federal Registrar 59:5377-5384.

U.S. Fish and Wildlife Service. 1997. Determination of Endangered Status for the Callippe Silverspot Butterfly of Behren's Silverspot Butterfly and Threatened Status for the Alameda Whipsnake. Federal Registrar 62:64306-64320.

U.S. Fish and Wildlife Service. 2006. Endangered and threatened wildlife and plants; designation of critical habitat for the Alameda whipsnake. Federal Registrar 71:58176-58230.

Van Oösterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

Wiens, J. A. 1997. Metapopulation dynamics and landscape ecology, Pages 32-60 *in* I. A. Hanski, and M. E. Gilpin, eds. Metapopulation Biology. Ecology, Genetics and Evolution. Academic Press, San Diego.

Wier, B. S., and C. C. Cockerham. 1984. Estimating *F* statistics for the analysis of population structure. Evolution 38:1358-1370.

Wood, D. A., R. N. Fisher, and T. W. Reeder. 2008a. Novel patterns of historical isolation, dispersal, and secondary contact across Baja California in the Rosy Boa (*Lichanura trivirgata*). Molecular Phylogenetics and Evolution 46:484-502.

Wood, D. A., J. Meik, A. Holycross, R. Fisher, and A. Vandergast. 2008b. Molecular and phenotypic diversity in *Chionactis occipitalis* (Western Shovel-nosed Snake), with emphasis on the status of *C. o. klauberi* (Tucson Shovel-nosed Snake). Conservation Genetics 9:1489-1507.

Wright, S. 1978, Evolution and the Genetics of Populations. University of Chicago Press, Chicago.

Young, A. C., and G. M. Clarke. 2000, Genetics, Demography and Viability of Fragmented Populations. University Press, Cambridge.

Zink, R. M. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. Proceedings of the Royal Society B: Biological Sciences 271:561-564.

Zink, R. M., G. F. Barrowclough, J. L. Atwood, and R. C. Blackwell-Rago. 2000. Genetics, taxonomy, and conservation of the threatened California gnatcatcher. Conservation Biology 14:1394-1405.

Appendix 8.1. Locality data and voucher/tissue numbers for *Masticophis lateralis* and outgroup taxa used in the study. The mtDNA site code corresponds to the tissues used in the mtDNA study (abbreviations refer to the county in which the samples were collected). Taxonomic abbreviations are as follows: MALAEU = *Masticophis lateralis euryxanthus*, MALALA = *M. l. lateralis*; MALA∗ = snakes of intermediate status (MALAEU x MALALA) based on published descriptions (Reimer 1954; Jennings 1983); MAAU = *M. aurigulus*; MABA = *M. barbouri*; MABI = *M. bilineatus*; MACO = *M. coluber*; MAFL = *M. flagellum*; MAME = *M. mentovarius*; MATA = *M. taeniatus*; MASC = *M. schotti*; SAHE = *Salvadora hexalepis*; and SAMX = *S. mexicana*. Tissues obtained from museum collections are listed by standard museum abbreviations: CAS, California Academy of Sciences; MVZ, Museum of Vertebrate Zoology at Berkeley; and ROM, Royal Ontario Museum. Non-standard abbreviations and collector abbreviations are as follows: SD Field, San Diego Natural History Museum; HBS, H. Bradley Shaffer; JQR, Jonathan Q. Richmond; RNF, Robert N. Fisher; all other tissues were obtained from field surveys conducted by Swaim Biological, Inc and the San Diego Field Station, US Geological Survey. Latitude and longitude coordinates are given in the WGS84 datum surface.

Voucher/ Tissue Number	mtDNA Site code	Taxon	Country	State	County	Locality	Latitude	Longitude
25850	ALMDA_1	MALAEU	USA	California	Alameda	Bailey Ranch	37.643300	-122.019000
25851	ALMDA_2	MALAEU	USA	California	Alameda	Bailey Ranch	37.643300	-122.019000
26858	ALMDA_3	MALAEU	USA	California	Alameda	Bailey Ranch	37.644780	-122.020620
26909		MALAEU	USA	California	Alameda	Bailey Ranch	37.643340	-122.018850
27392		MALAEU	USA	California	Alameda	Bailey Ranch	37.644990	-122.019680
L5		MALAEU	USA	California	Alameda	Bailey Ranch	37.645670	-122.020670
L6		MALAEU	USA	California	Alameda	Bailey Ranch	37.647790	-122.018940
L7		MALAEU	USA	California	Alameda	Bailey Ranch	37.646100	-122.019940
L8		MALAEU	USA	California	Alameda	Bailey Ranch	37.649950	-122.017330
L9		MALAEU	USA	California	Alameda	Bailey Ranch	37.647820	-122.019130
L13		MALAEU	USA	California	Alameda	Bailey Ranch	37.647410	-122.018050
L14		MALAEU	USA	California	Alameda	Bailey Ranch	37.643660	-122.019060
L15		MALAEU	USA	California	Alameda	Bailey Ranch	37.647440	-122.018140
L16		MALAEU	USA	California	Alameda	Bailey Ranch	37.647220	-122.017510
L17		MALAEU	USA	California	Alameda	Bailey Ranch	37.649560	-122.017100
L18		MALAEU	USA	California	Alameda	Bailey Ranch	37.647190	-122.017810
R5		MALAEU	USA	California	Alameda	Bailey Ranch	37.646120	-122.019960
R6		MALAEU	USA	California	Alameda	Bailey Ranch	37.646100	-122.019950
R7		MALAEU	USA	California	Alameda	Bailey Ranch	37.644130	-122.019300

Appendix 8.1. Tissue identification and locality data.

R8		MALAEU	USA	California	Alameda	Bailey Ranch	37.647450	-122.018140
R9		MALAEU	USA	California	Alameda	Bailey Ranch	37.645910	-122.020410
R12		MALAEU	USA	California	Alameda	Bailey Ranch	37.646090	-122.019940
R13		MALAEU	USA	California	Alameda	Bailey Ranch	37.647420	-122.017960
R14		MALAEU	USA	California	Alameda	Bailey Ranch	37.650950	-122.017080
R15		MALAEU	USA	California	Alameda	Bailey Ranch	37.648970	-122.018270
R16		MALAEU	USA	California	Alameda	Bailey Ranch	37.647430	-122.018100
R17		MALAEU	USA	California	Alameda	Bailey Ranch	37.647440	-122.018130
R18		MALAEU	USA	California	Alameda	Bailey Ranch	37.649840	-122.017250
KS4		MALAEU	USA	California	Alameda	Bailey Ranch	37.642730	-122.019020
KS5		MALAEU	USA	California	Alameda	Bailey Ranch	37.646110	-122.019940
KS6		MALAEU	USA	California	Alameda	Bailey Ranch	37.646140	-122.019960
Ohlone01	ALMDA_4	MALA*	USA	California	Alameda	Camp Ohlone	37.487390	-121.743580
Ohlone02	ALMDA_5	MALA*	USA	California	Alameda	Camp Ohlone	37.487280	-121.743440
Ohlone03	ALMDA_6	MALA*	USA	California	Alameda	Camp Ohlone	37.486860	-121.743180
Ohlone04		MALA*	USA	California	Alameda	Camp Ohlone	37.486760	-121.743030
Ohlone05		MALA*	USA	California	Alameda	Camp Ohlone	37.486610	-121.742600
Ohlone06		MALA*	USA	California	Alameda	Camp Ohlone	37.486290	-121.742330
Ohlone07		MALA*	USA	California	Alameda	Camp Ohlone	37.486290	-121.742330
Ohlone08	ALMDA_7	MALA*	USA	California	Alameda	Camp Ohlone	37.486220	-121.742200
Ohlone09		MALA*	USA	California	Alameda	Camp Ohlone	37.487280	-121.743440
Ohlone10		MALA*	USA	California	Alameda	Camp Ohlone	37.487230	-121.743920
Ohlone11		MALA*	USA	California	Alameda	Camp Ohlone	37.485790	-121.743160
Ohlone12		MALA*	USA	California	Alameda	Camp Ohlone	37.485790	-121.743160
Ohlone13		MALA*	USA	California	Alameda	Camp Ohlone	37.488680	-121.745240
Ohlone14		MALA*	USA	California	Alameda	Camp Ohlone	37.487230	-121.743920
Ohlone15		MALA*	USA	California	Alameda	Camp Ohlone	37.486030	-121.741860
Ohlone16		MALA*	USA	California	Alameda	Camp Ohlone	37.486230	-121.753060
Ohlone17		MALA*	USA	California	Alameda	Camp Ohlone	37.486760	-121.743030
Ohlone18		MALA*	USA	California	Alameda	Camp Ohlone	37.486530	-121.743500
Ohlone19		MALA*	USA	California	Alameda	Camp Ohlone	37.487390	-121.743580
Ohlone20		MALA*	USA	California	Alameda	Camp Ohlone	37.487390	-121.743580
Ohlone21		MALA*	USA	California	Alameda	Camp Ohlone	37.486800	-121.743100
Ohlone22		MALA*	USA	California	Alameda	Camp Ohlone	37.487230	-121.743920
Ohlone23		MALA*	USA	California	Alameda	Camp Ohlone	37.487390	-121.743580

Appendix 8.1. Tissue identification and locality data.

Ohlone24		MALA∗	USA	California	Alameda	Camp Ohlone	37.486030	-121.741860
Ohlone25		MALA∗	USA	California	Alameda	Camp Ohlone	37.489500	-121.745900
KS3		MALA∗	USA	California	Alameda	Camp Ohlone	37.487390	-121.743580
DelValle02	ALMDA_8	MALA∗	USA	California	Alameda	Del Valle	37.619530	-121.746810
DelValle03	ALMDA_9	MALA∗	USA	California	Alameda	Del Valle	37.619640	-121.746700
DelValle01	ALMDA_10	MALA∗	USA	California	Alameda	Del Valle	37.619530	-121.746810
36427	ALMDA_11	MALAEU	USA	California	Alameda	Stonebrae	37.626480	-121.988810
36428	ALMDA_12	MALAEU	USA	California	Alameda	Stonebrae	37.635710	-121.998070
36605	ALMDA_13	MALAEU	USA	California	Alameda	Stonebrae	37.622010	-121.995470
36694	ALMDA_14	MALAEU	USA	California	Alameda	Stonebrae	37.650240	-121.994840
39017	ALMDA_15	MALAEU	USA	California	Alameda	Stonebrae	37.637170	-122.002290
36631		MALAEU	USA	California	Alameda	Stonebrae	37.630100	-121.999970
36687		MALAEU	USA	California	Alameda	Stonebrae	37.635490	-121.997580
36688		MALAEU	USA	California	Alameda	Stonebrae	37.636070	-121.999320
36693		MALAEU	USA	California	Alameda	Stonebrae	37.650240	-121.994840
36705		MALAEU	USA	California	Alameda	Stonebrae	37.636150	-122.000320
36753		MALAEU	USA	California	Alameda	Stonebrae	37.636150	-122.000320
36827		MALAEU	USA	California	Alameda	Stonebrae	37.626000	-121.990040
36828		MALAEU	USA	California	Alameda	Stonebrae	37.663660	-122.009410
36831		MALAEU	USA	California	Alameda	Stonebrae	37.649830	-121.995480
36851		MALAEU	USA	California	Alameda	Stonebrae	37.650290	-121.994740
36861		MALAEU	USA	California	Alameda	Stonebrae	37.660880	-122.000810
36880		MALAEU	USA	California	Alameda	Stonebrae	37.660500	-122.001810
36883		MALAEU	USA	California	Alameda	Stonebrae	37.660760	-122.001570
39020		MALAEU	USA	California	Alameda	Stonebrae	37.637170	-122.002290
39021		MALAEU	USA	California	Alameda	Stonebrae	37.650140	-121.994730
39242		MALAEU	USA	California	Alameda	Stonebrae	37.635890	-122.001150
39281		MALAEU	USA	California	Alameda	Stonebrae	37.635890	-122.001150
39286		MALAEU	USA	California	Alameda	Stonebrae	37.626670	-121.990550
39362		MALAEU	USA	California	Alameda	Stonebrae	37.629300	-121.991760
39363		MALAEU	USA	California	Alameda	Stonebrae	37.626580	-121.988650
39368		MALAEU	USA	California	Alameda	Stonebrae	37.627030	-121.991010
39383		MALAEU	USA	California	Alameda	Stonebrae	37.622010	-121.995470
39395		MALAEU	USA	California	Alameda	Stonebrae	37.635840	-122.000500
43381		MALAEU	USA	California	Alameda	Stonebrae	37.626410	-121.990080

Appendix 8.1. Tissue identification and locality data.

43526		MALAEU	USA	California	Alameda	Stonebrae	37.650130	-121.994610
43532		MALAEU	USA	California	Alameda	Stonebrae	37.626730	-121.990320
43858		MALAEU	USA	California	Alameda	Stonebrae	37.646860	-122.001010
CAS 227730	ALMDA_16	MALAEU	USA	California	Alameda	Berkeley, Wildcat Canyon Rd, 0.5 mi NNW of Shasta Rd	37.894167	-122.252028
CAS 214877	ALMDA_17	MALA∗	USA	California	Alameda	Mines Rd, 5.75 mile marker	37.550858	-121.570597
MVZ 230735	ALMDA_18	MALA∗	USA	California	Alameda	Tesla Rd, 2.9 road mi W of Alameda County line	37.648750	-121.599400
MVZ 230734	ALMDA_19	MALA∗	USA	California	Alameda	Tesla Rd, 3.3 road mi W of Alameda County line	37.651060	-121.602870
MVZ 230736	ALMDA_20	MALA∗	USA	California	Alameda	Tesla Rd, 4.6 road mi W of Alameda County line	37.648360	-121.618030
MVZ 164944	ALMDA_21	MALA∗	USA	California	Alameda	Mines Rd, 9.5 mi S Tesla Rd, Livermore	37.587150	-121.617490
HBS 17056	AMDR_1	MALALA	USA	California	Amador	14880 Tyler Rd, near Fiddletown	38.508749	-120.744763
CAS 233408	BUTT_1	MALALA	USA	California	Butte	North Fork Feather River, Cresta Reach	39.848840	-121.393303
CAS 209526	CLSA_1	MALALA	USA	California	Colusa	South Fork Stony Ck, ca 100 m upstream of Forest Rd	39.358533	-122.654256
CAS 212745	CLSA_2	MALALA	USA	California	Colusa	Little Sullivan Creek	39.293861	-122.614722
MALA5	CNTR_1	MALAEU	USA	California	Contra Costa	Alhambra	37.974920	-122.116460
MALA6	CNTR_2	MALAEU	USA	California	Contra Costa	Alhambra	37.646060	-122.019890
MALA9	CNTR_3	MALAEU	USA	California	Contra Costa	Alhambra	37.974920	-122.116460
KS8	CNTR_4	MALAEU	USA	California	Contra Costa	Alhambra	37.969720	-122.110840
KS40	CNTR_5	MALAEU	USA	California	Contra Costa	Alhambra	37.976910	-122.121190
MALA8		MALAEU	USA	California	Contra Costa	Alhambra	37.974920	-122.116460
MALA13b		MALAEU	USA	California	Contra Costa	Alhambra	37.974920	-122.116460
KS9		MALAEU	USA	California	Contra Costa	Alhambra	37.969720	-122.110840
KS41		MALAEU	USA	California	Contra Costa	Alhambra	37.972720	-122.113680
KS42		MALAEU	USA	California	Contra Costa	Alhambra	37.978290	-122.120010
KS43		MALAEU	USA	California	Contra Costa	Alhambra	37.972850	-122.113280
31476	CNTR_6	MALAEU	USA	California	Contra Costa	Allen Property	37.952970	-122.214850
32042	CNTR_7	MALAEU	USA	California	Contra Costa	Allen Property	37.949490	-122.217120
32043	CNTR_8	MALAEU	USA	California	Contra Costa	Allen Property	37.949350	-122.217020
32765	CNTR_9	MALAEU	USA	California	Contra Costa	Allen Property	37.950140	-122.217780
39856	CNTR_10	MALAEU	USA	California	Contra Costa	Black Diamond	37.941500	-121.824400
39865	CNTR_11	MALAEU	USA	California	Contra Costa	Black Diamond	37.941260	-121.824500

Appendix 8.1. Tissue identification and locality data.

KS34	CNTR_12	MALAEU	USA	California	Contra Costa	Black Diamond	37.941450	-121.824350
12757	CNTR_13	MALAEU	USA	California	Contra Costa	Claremont	37.874740	-122.225060
23280	CNTR_14	MALAEU	USA	California	Contra Costa	Claremont	37.874230	-122.225850
26043	CNTR_15	MALAEU	USA	California	Contra Costa	Claremont	37.874830	-122.225090
26254		MALAEU	USA	California	Contra Costa	Claremont	37.874290	-122.224790
27524		MALAEU	USA	California	Contra Costa	Claremont	37.874290	-122.224790
14115	CNTR_16	MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810290	-121.769390
14139	CNTR_17	MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814350	-121.777190
14368	CNTR_18	MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810350	-121.769480
20353	CNTR_19	MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811910	-121.772490
33755	CNTR_20	MALAEU	USA	California	Contra Costa	Los Vaqueros	37.809870	-121.768410
14402		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.813640	-121.775900
14527		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810080	-121.769190
14555		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811810	-121.771940
14749		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811470	-121.770350
14878		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.809570	-121.768040
15221		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811980	-121.771100
16028		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.809490	-121.767960
16306		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.813790	-121.776140
16756		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810970	-121.771660
17912		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811910	-121.772490
19442		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814800	-121.776710
20153		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814450	-121.776220
20367		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810610	-121.771010
20400		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814400	-121.776610
23911		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.813650	-121.775540
24199		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814760	-121.775070
24356		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.808850	-121.765750
24357		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.809240	-121.766190
26082		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814560	-121.775570
28126		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814050	-121.775120
33440		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.811660	-121.772300
33497		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814650	-121.776150
34396		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.814660	-121.775710
35036		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.812100	-121.770990

Appendix 8.1. Tissue identification and locality data.

35164		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810540	-121.770480
39130		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810080	-121.769190
39216		MALAEU	USA	California	Contra Costa	Los Vaqueros	37.810080	-121.769190
13167	CNTR_21	MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838430	-121.909340
13196	CNTR_22	MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844790	-121.915060
13199	CNTR_23	MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844340	-121.913200
13449	CNTR_24	MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844560	-121.916150
13234		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844690	-121.916150
13235		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844340	-121.913200
13241		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.847110	-121.904240
13244		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.847640	-121.901600
13303		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.847910	-121.901550
13350		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.835790	-121.901370
13374		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.847690	-121.930320
13378		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.843530	-121.910070
13422		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837400	-121.904440
13460		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837180	-121.906490
13587		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.846550	-121.900620
13615		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.842080	-121.922390
13925		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.843320	-121.911580
13977		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.846750	-121.904120
23867		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844530	-121.910090
23910		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838080	-121.908430
23954		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837260	-121.906670
24065		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.844130	-121.927010
24073		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838000	-121.909680
24081		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837260	-121.906670
25542		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837380	-121.904510
25595		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838400	-121.909240
25629		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.845870	-121.932260
25777		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838430	-121.909340
26602		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.842010	-121.922870
26623		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.837180	-121.906490
26800		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.835790	-121.901370
28093		MALAEU	USA	California	Contra Costa	Mt. Diablo	37.838530	-121.909340

Appendix 8.1. Tissue identification and locality data.

36874	CNTR_25	MALAEU	USA	California	Contra Costa	Allen Property	37.950360	-122.216440
36889	CNTR_26	MALAEU	USA	California	Contra Costa	Allen Property	37.954340	-122.217680
36892	CNTR_27	MALAEU	USA	California	Contra Costa	Allen Property	37.954200	-122.217600
31476		MALAEU	USA	California	Contra Costa	Allen Property	37.952970	-122.214850
32029		MALAEU	USA	California	Contra Costa	Allen Property	37.953810	-122.218410
32042		MALAEU	USA	California	Contra Costa	Allen Property	37.949490	-122.217120
32043		MALAEU	USA	California	Contra Costa	Allen Property	37.949350	-122.217020
32759		MALAEU	USA	California	Contra Costa	Allen Property	37.953670	-122.217010
32765		MALAEU	USA	California	Contra Costa	Allen Property	37.950140	-122.217780
32766		MALAEU	USA	California	Contra Costa	Allen Property	37.950140	-122.217780
32780		MALAEU	USA	California	Contra Costa	Allen Property	37.953670	-122.217010
36893		MALAEU	USA	California	Contra Costa	Allen Property	37.953530	-122.216890
36894		MALAEU	USA	California	Contra Costa	Allen Property	37.950360	-122.216440
36895		MALAEU	USA	California	Contra Costa	Allen Property	37.950490	-122.216700
36896		MALAEU	USA	California	Contra Costa	Allen Property	37.954050	-122.217780
36900		MALAEU	USA	California	Contra Costa	Allen Property	37.954340	-122.217680
36906		MALAEU	USA	California	Contra Costa	Allen Property	37.954340	-122.217680
36929		MALAEU	USA	California	Contra Costa	Allen Property	37.949250	-122.216280
36935		MALAEU	USA	California	Contra Costa	Allen Property	37.950440	-122.216510
36936		MALAEU	USA	California	Contra Costa	Allen Property	37.953380	-122.216800
36948		MALAEU	USA	California	Contra Costa	Allen Property	37.954050	-122.217780
36959		MALAEU	USA	California	Contra Costa	Allen Property	37.953760	-122.218210
36960		MALAEU	USA	California	Contra Costa	Allen Property	37.954160	-122.217940
36961		MALAEU	USA	California	Contra Costa	Allen Property	37.950480	-122.216190
38451		MALAEU	USA	California	Contra Costa	Allen Property	37.953740	-122.216830
38452		MALAEU	USA	California	Contra Costa	Allen Property	37.954090	-122.218120
38453		MALAEU	USA	California	Contra Costa	Allen Property	37.951490	-122.218800
38455		MALAEU	USA	California	Contra Costa	Allen Property	37.954300	-122.217480
38464		MALAEU	USA	California	Contra Costa	Allen Property	37.953530	-122.216890
43029		MALAEU	USA	California	Contra Costa	Allen Property	37.953670	-122.217010
43052		MALAEU	USA	California	Contra Costa	Allen Property	37.949490	-122.217120
43060		MALAEU	USA	California	Contra Costa	Allen Property	37.954050	-122.217780
43152		MALAEU	USA	California	Contra Costa	Allen Property	37.954050	-122.217780
53915		MALAEU	USA	California	Contra Costa	Allen Property	37.950140	-122.217780
55136		MALAEU	USA	California	Contra Costa	Allen Property	37.954200	-122.217600

Appendix 8.1. Tissue identification and locality data.

MALA1		MALAEU	USA	California	Contra Costa	Allen Property	37.949350	-122.217020
36932	CNTR_28	MALAEU	USA	California	Contra Costa	Tilden Park	37.891250	-122.233100
36933	CNTR_29	MALAEU	USA	California	Contra Costa	Tilden Park	37.891070	-122.233040
36953	CNTR_30	MALAEU	USA	California	Contra Costa	Tilden Park	37.893360	-122.234310
38454	CNTR_31	MALAEU	USA	California	Contra Costa	Tilden Park	37.893670	-122.234920
CAS 201051	CNTR_32	MALAEU	USA	California	Contra Costa	El Toyonal Rd at intersection of Camino del Cielo	37.883233	-122.205583
TildenMala65	CNTR_33	MALAEU	USA	California	Contra Costa	Tilden Park	37.893670	-122.234920
TildenMala66	CNTR_34	MALAEU	USA	California	Contra Costa	Tilden Park	37.893670	-122.234920
CAS 234680	ELDOR_1	MALALA	USA	California	El Dorado	Poho Ridge, confluence of South Fork American River & Silver Creek	38.794080	-120.589393
JQR 859	ELDOR_2	MALALA	USA	California	El Dorado	Kyburz	38.768520	-120.317120
JQR 860	ELDOR_3	MALALA	USA	California	El Dorado	Kyburz	38.768520	-120.317120
JQR 861	ELDOR_4	MALALA	USA	California	El Dorado	Kyburz	38.768520	-120.317120
RNF 7863	FRES_1	MALALA	USA	California	Fresno	0.6mi E of San Benito/Fresno County lines on Los Gatos Rd	36.291100	-120.679090
CAS 213085	FRES_2	MALALA	USA	California	Fresno	Forest Rd 10S43, 1.2 mi E of Peterson Mill Rd & Forest Rd	37.029530	-119.293330
CAS 220925	KERN_1	MALALA	USA	California	Kern	Cedar Creek	35.722944	-118.608056
JQR 652	KERN_2	MALALA	USA	California	Kern	Breckenridge Mountain	35.476550	-118.571030
MALA434_2A	LA_1	MALALA	USA	California	Los Angeles	Hennessey		
PUE001	LA_2	MALALA	USA	California	Los Angeles	Puente Hills	33.975154	-117.947965
CAS 212948	MDRA_1	MALALA	USA	California	Madera	Rock Creek, upstream of Forest Rd 7S47A	37.275230	-119.340910
CAS 208884	MRPSA_1	MALALA	USA	California	Mariposa	South Fork Merced River, Hite Cove Trail	37.640190	-119.875680
MVZ 241099	MRPSA_2	MALALA	USA	California	Mariposa	ca 0.25 mi S along Foresta Rd	37.710925	-119.734545
MVZ 245803	MRPSA_3	MALALA	USA	California	Mariposa	Big Oak Flat Rd, 0.4 mi W of Foresta Rd	37.715570	-119.738210
MVZ 245804	MRPSA_4	MALALA	USA	California	Mariposa	Big Oak Flat Rd, 1 mi E of Crane Flat	37.748080	-119.780350
CAS 213088	MNDCO_1	MALALA	USA	California	Mendocino	SE end of Williams Valley, 28201 Mendocino Pass Rd	39.831694	-123.176778
MVZ 232836	MNTRY_1	MALALA	USA	California	Monterey	Hwy 1 at Grimes Canyon, 2 mi SE Nepenthe	36.208500	-121.734100
MVZ 229167	MNTRY_2	MALALA	USA	California	Monterey	Tassajara Rd, 14.5 mi S junction with Carmel Valley Rd	36.209847	-121.554467

Appendix 8.1. Tissue identification and locality data.

HBS 38979	MNTRY_3	MALALA	USA	California	Monterey	Fort Ord	36.624200	-121.780801
JQR833	MNTRY_4	MALALA	USA	California	Monterey	S side of Vineyard Canyon Road, ca 5 mi NW of Parkfield	35.920467	-120.523617
JQR834	MNTRY_5	MALALA	USA	California	Monterey	S side of Vineyard Canyon Road, ca 5 mi NW of Parkfield	35.920467	-120.523617
GP26APR10	NAPA_1	MALALA	USA	California	Napa	Along Knoxville-Berryessa Rd, ca 6.2 mi S of Lake-Napa County Line	38.807790	-122.322430
HBS 16262	NAPA_2	MALALA	USA	California	Napa	Lake Berryessa	38.491161	-122.126164
AGC001	ORNG_1	MALALA	USA	California	Orange	Agua Chinon	33.700262	-117.692142
AWC001	ORNG_2	MALALA	USA	California	Orange	Aliso-Wood Canyon	33.552860	-117.744775
LIM001	ORNG_3	MALALA	USA	California	Orange	Limestone	33.728334	-117.671135
LIM002	ORNG_4	MALALA	USA	California	Orange	Limestone	33.728334	-117.671135
ORA001	ORNG_5	MALALA	USA	California	Orange	Orange Hills	33.798488	-117.793988
PET001	ORNG_6	MALALA	USA	California	Orange	Peter's Canyon	33.767418	-117.765298
CHI001	ORNG_7	MALALA	USA	California	Orange	Chino Hills	33.898787	-117.735700
SJH001	ORNG_8	MALALA	USA	California	Orange	San Joaquin Hills	33.584401	-117.792301
RAT001	ORNG_9	MALALA	USA	California	Orange	Rattlesnake Reserve	33.732596	-117.732902
STA001	ORNG_10	MALALA	USA	California	Orange	Starr Ranch	33.612789	-117.550626
WEI001	ORNG_11	MALALA	USA	California	Orange	Weir Canyon	33.824307	-117.739370
EDI001	ORNG_12	MALALA	USA	California	Orange	Edison Easement	33.668590	-117.639680
CAS 233409	PLMS_1	MALALA	USA	California	Plumas	Hwy 70 at Rodger's Flat	39.963361	-121.276833
JOS001	RVRSD_1	MALALA	USA	California	Riverside	Joshua Tree National Park	33.938183	-115.986886
NOR001	RVRSD_2	MALALA	USA	California	Riverside	North Hills	33.703287	-117.013271
SKI001	RVRSD_3	MALALA	USA	California	Riverside	Lake Skinner	33.585804	-117.034575
RAW001	RVRSD_4	MALALA	USA	California	Riverside	Rawson Canyon	33.629229	-117.009483
MOT001	RVRSD_5	MALALA	USA	California	Riverside	Motte Rimrock Reserve	33.808841	-117.255888
TEN001	RVRSD_6	MALALA	USA	California	Riverside	Tenaja Corridor	33.509318	-117.366034
RNF 7860	SNBNT_1	MALALA	USA	California	San Benito	Panoche Rd, 16mi W of Little Panoche Rd	36.660040	-121.112810
RNF 7861	SNBNT_2	MALALA	USA	California	San Benito	Panoche Rd, 16mi E of Hwy 25	36.655000	-121.061870
RNF 7862	SNBNT_3	MALALA	USA	California	San Benito	Panoche Rd, 17mi E of Hwy 25	36.648200	-121.039200
JQR675	SNBNT_4	MALALA	USA	California	San Benito	Pinnacles National Monument, Bear Gulch Headquarters	36.483318	-121.177638
SIL001	SNBRN_1	MALALA	USA	California	San Bernardino	Silverwood Lake	34.286220	-117.330191
CAS 210354	SNBRN_2	MALALA	USA	California	San Bernardino	State Route 330, 4 mi N of Redlands, Cook Creek Fire Station	34.135817	-117.212264

Appendix 8.1. Tissue identification and locality data.

SD Field 2546	SNDG_2	MALALA	USA	California	San Diego	Barona, 3.9 mi NE of, on Wildcat Canyon Rd	32.978900	-116.824310
MAR001	SNDG_3	MALALA	USA	California	San Diego	Marron Valley	32.594590	-116.766790
TJE001	SNDG_4	MALALA	USA	California	San Diego	Tijuana Estuary	32.546033	-117.120117
JQR830	SNDG_5	MALALA	USA	California	San Diego	Boulevard, 0.2mi S of I-8 on the E side of Ribbonwood Rd	32.673620	-116.291197
JQR836	SNDG_6	MALALA	USA	California	San Diego	S2, S of San Felipe	33.197630	-116.596520
DRC7-27-15	SNDG_7	MALALA	USA	California	San Diego	Rancho Jamul Ecological Reserve	32.687600	-116.855000
DRC5-12-7	SNDG_8	MALALA	USA	California	San Diego	Rancho Jamul Ecological Reserve	32.683000	-116.818360
KAF8-10-38	SNDG_9	MALALA	USA	California	San Diego	Rancho Jamul Ecological Reserve	32.689500	-116.818000
SC15-54-88	SNDG_10	MALALA	USA	California	San Diego	Rancho Jamul Ecological Reserve	32.685450	-116.854280
SDField 2114	SNDG_11	MALALA	USA	California	San Diego	Hellhole Canyon Open Space Preserve	33.224279	-116.939791
20436	SNJQN_1	MALA*	USA	California	San Joaquin	Site 300	37.645290	-121.550740
20441		MALA*	USA	California	San Joaquin	Site 300	37.648200	-121.554480
20467		MALA*	USA	California	San Joaquin	Site 300	37.644110	-121.552180
20523	SNJQN_4	MALA*	USA	California	San Joaquin	Site 300	37.649690	-121.555020
20865	SNJQN_5	MALA*	USA	California	San Joaquin	Site 300	37.644240	-121.552120
20484		MALA*	USA	California	San Joaquin	Site 300	37.648470	-121.554900
20588		MALA*	USA	California	San Joaquin	Site 300	37.644670	-121.551820
20672		MALA*	USA	California	San Joaquin	Site 300	37.649480	-121.554960
20714		MALA*	USA	California	San Joaquin	Site 300	37.649400	-121.553460
20780		MALA*	USA	California	San Joaquin	Site 300	37.644110	-121.552180
20821		MALA*	USA	California	San Joaquin	Site 300	37.645340	-121.551560
20845		MALA*	USA	California	San Joaquin	Site 300	37.649630	-121.553060
20847		MALA*	USA	California	San Joaquin	Site 300	37.644670	-121.551820
20859		MALA*	USA	California	San Joaquin	Site 300	37.649610	-121.554410
20869		MALA*	USA	California	San Joaquin	Site 300	37.649440	-121.553310
21034		MALA*	USA	California	San Joaquin	Site 300	37.649520	-121.554820
21082		MALA*	USA	California	San Joaquin	Site 300	37.649480	-121.554960
21094		MALA*	USA	California	San Joaquin	Site 300	37.648470	-121.554900
21095		MALA*	USA	California	San Joaquin	Site 300	37.648200	-121.554480
21142		MALA*	USA	California	San Joaquin	Site 300	37.648480	-121.554490
21304		MALA*	USA	California	San Joaquin	Site 300	37.649560	-121.553150
21777		MALA*	USA	California	San Joaquin	Site 300	37.644880	-121.550130
30683		MALA*	USA	California	San Joaquin	Site 300	37.648340	-121.554480

Appendix 8.1. Tissue identification and locality data.

31462		MALA*	USA	California	San Joaquin	Site 300	37.648200	-121.554480
31699		MALA*	USA	California	San Joaquin	Site 300	37.648430	-121.554660
32188		MALA*	USA	California	San Joaquin	Site 300	37.645200	-121.550560
32448		MALA*	USA	California	San Joaquin	Site 300	37.646050	-121.552320
CAS 208508	SLO_1	MALALA	USA	California	San Luis Obispo	HWY 1, near Calander Rd, Pismo Dunes SVRA	35.054950	-120.594880
JQR674	SNBRB_1	MALALA	USA	California	Santa Barbara	Foxen Canyon Rd, 0.3mi S of the jct of Foxen Canyon Rd & Orcutt Garet Rd	34.878310	-120.311950
09DAW2P3	SNBRB_2	MALALA	USA	California	Santa Barbara	Rancho Aakatada, off Hwy 101, ca 1.8 mi S of Santa Ynez River crossing	34.574610	-120.192930
MVZ 243356	SNCLR_1	MALALA	USA	California	Santa Clara	ca 2 mi N on San Antonio Valley Rd from jct with Del Puerto Rd	37.414715	-121.497816
HCMALA1	SNCLR_2	MALALA	USA	California	Santa Clara	Henry Coe State Park	37.169305	-121.422761
HCMALA2	SNCLR_3	MALALA	USA	California	Santa Clara	Henry Coe State Park	37.169305	-121.422761
HCMALA3	SNCLR_4	MALALA	USA	California	Santa Clara	Henry Coe State Park	37.169305	-121.422761
JQR 846	THMA_1	MALALA	USA	California	Tehama	Dye Creek Preserve, near Molinos	39.989700	-122.049680
CAS 199521	THMA_2	MALALA	USA	California	Tehama	1.5 km downstream from Black Rock Campground	40.183390	-121.719300
CAS 227000	THMA_3	MALALA	USA	California	Tehama	Deer Creek, from crossing of Forest Rd 28N29 to confluence of Beaver Creek	40.081433	-121.693439
CAS 220886	TLR_1	MALALA	USA	California	Tulare	North Fork Middle Fork Tule River	36.188333	-118.668611
CAS 224762	TLR_2	MALALA	USA	California	Tulare	Greenhorn Mtns, 2 mi W of California Hot Springs	35.872031	-118.704361
MVZ 241100	TLMN_1	MALALA	USA	California	Tuolomne	ca 1.2 mi E of Harden Rd on Hwy 120	37.817799	-119.911246
MVZ 249944	TLMN_2	MALALA	USA	California	Tuolomne	Rancheria Falls campground, Hetch Hetchy	37.956290	-119.715300
MVZ 249942	TLMN_3	MALALA	USA	California	Tuolomne	Hetch Hetchy Rd	37.935620	-119.793950
MVZ 249943	TLMN_4	MALALA	USA	California	Tuolomne	Hetch Hetchy Rd	37.921450	-119.804200
MVZ 241101	TLMN_5	MALALA	USA	California	Tuolomne	1.7 mi W of Merced Grove Trail head, Big Oak Flat Rd	37.775920	-119.861270
JQR 676	VNTR_1	MALALA	USA	California	Ventura	918 Sudario Ct, Camarillo, CA 93010	34.248763	-119.063429
MALA0037	VNTR_2	MALALA	USA	California	Ventura	North Mugu	34.101500	-119.035480
MALA0046	VNTR_4	MALALA	USA	California	Ventura	North Mugu	34.101500	-119.035480

Appendix 8.1. Tissue identification and locality data.

MALA_11_51	VNTR_5	MALALA	USA	California	Ventura	Simi		
MALA0033	VNTR_6	MALALA	USA	California	Ventura	White Oak		
JQR812	VNTR_7	MALALA	USA	California	Ventura	Camarillo Estates	34.252598	-119.047670
JQR 845	VNTR_8	MALALA	USA	California	Ventura	ca 8 mi west of Frazier Park on the south slope of Mt Pinos	34.769723	-119.047565
SDField 0731	BCN_1	MALA	Mexico	Baja California Norte		San Vincente, 21.3 km S of	31.158700	-116.140360
SDField 0753	BCN_2	MALA	Mexico	Baja California Norte		La Grulla, 39.5 km marker S of Ensenada	31.630360	-116.463380
SDMALA3	BCN_3	MALA	Mexico	Baja California Norte		22 km S of San Vicente	31.152579	-116.136988
MVZ 182251	BCS_1	MALA	Mexico	Baja California Sur		8.9 mi E San Ignacio by Mexico Hwy 1	27.325556	-113.016111
RWM 925	BCS_2	MALA	Mexico	Baja California Sur		San Pedro La Presa, 4.5 mi E	24.867297	-111.051722
SDField 0458	BCS_3	MALA	Mexico	Baja California Sur		Los Dolores	25.066843	-110.850153
SDField 0525	BCS_4	MALA	Mexico	Baja California Sur		3 km NNW of San Evaristo base-camp	24.933342	-110.733406
SDField 0601	BCS_5	MALA	Mexico	Baja California Sur		Estacion Microondas el Cardon; 6 km west of	24.152750	-110.831911

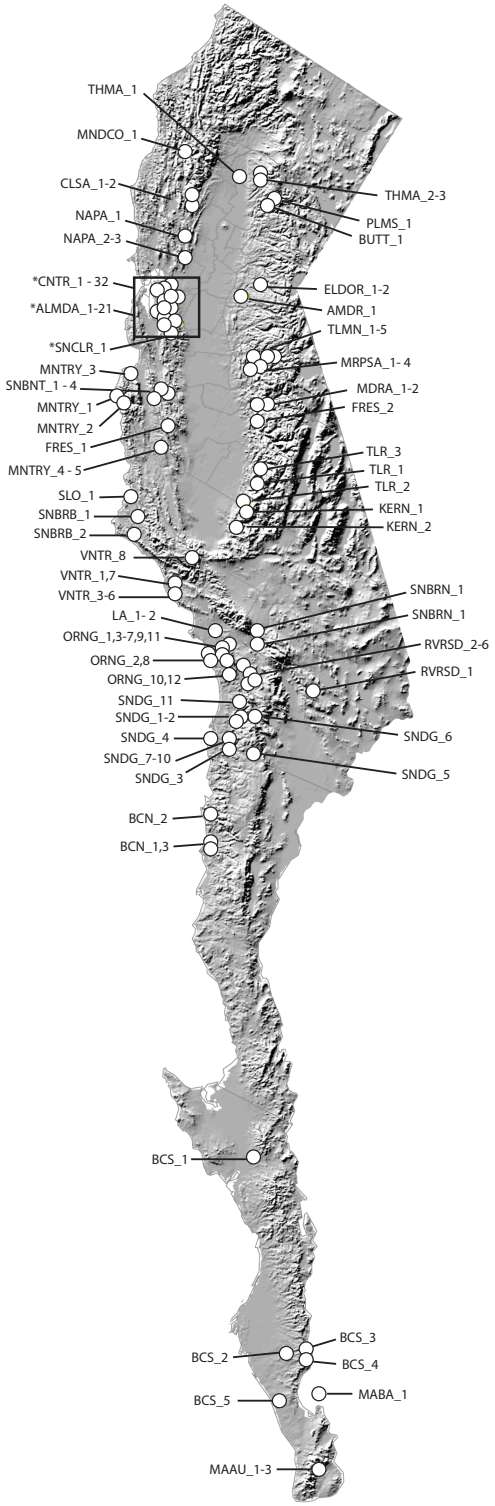
Appendix 8.1. Tissue identification and locality data.

Outgroup taxa

Voucher/ Tissue Number	mtDNA Site code	Taxon	Country	State	County	Locality	Latitude	Longitude
MVZ 234614	MAFL_1	MAFL	USA	California	San Bernardino	near Kelso Dune	34.902430	-115.745640
SDMALA1	MAFL_2	MAFL	Mexico	Baja California Norte		29 km S of San Vicente	31.109032	-116.159777
RLS 2565	MABA_1	MABA	Mexico	Baja California Sur		Isla Espiritu Santo	24.492104	-110.371825
MVZ 236399	MAAU_1	MAAU	Mexico	Baja California Sur		Agua Caliente Dam 2.4 mi W of Agua Caliente	23.440000	-109.808333
JEC 889	MAAU_2	MAAU	Mexico	Baja California Sur		Agua Caliente Dam 2.4 mi W of Agua Caliente	23.440000	-109.808333
JEC 890	MAAU_3	MAAU	Mexico	Baja California Sur		Agua Caliente Dam 2.4 mi W of Agua Caliente	23.440000	-109.808333
RNF 6724	MACO_1	MACO	USA	California	San Diego	Santa Ysabel Ecological Reserve		
ROM 28973	MACO_2	MACO	USA	California				
MVZ 204113	MAME_1	MAME	Costa Rica	Prov Guanacaste		14.5 km W Canas on Hwy 1	10.483330	-85.216670
MVZ 207361	MAME_2	MAME	Costa Rica	Prov. Alajuela		La Argentina de Grecia	10.033330	-84.350000
GAA 83-112	MATA_5	MATA	Mexico	Coahuila		Mohovano		
GAA 83-103	MATA_4	MATA	Mexico	Coahuila		Mohovano		
ROM 15326	MATA_6	MATA	Mexico	Durango		Ceballos, 26.6 mi S, base of microondas Conejo, Sierra de las Banderas		
GAA 83-97	MATA_3	MATA	Mexico	Nuevo Leon		Doctor Arroyo, 34.8 km N on Hwy 61		
GAA 83-68	MATA_2	MATA	Mexico	Nuevo Leon		Linares, 23 km W on Hwy 58		
CAS 206499	MATA_1	MATA	USA	California	Mono	Hot Creek	37.653760	-118.82411

Appendix 8.1. Tissue identification and locality data.

MVZ 175911	MABI_1	MABI	USA	Arizona	Cochise	Paradise Rd, 2.6 mi from jct with Portal Rd	31.932200	-109.185500
ROM 14965	MABI_2	MABI						
MVZ 233302	MASC_1	MASC	Mexico	Queretaro		0.5-1.0 mi E San Pablo Junction		
GENBANK	AF138748	SAHE	USA	California	San Bernardino	Camp Rock Rd, jct of Upper Johnson & Oard valleys		
GENBANK	AY487075	SAMX	Mexico	Michoacan		17 km S of Quatro Caminos		



Appendix 8.2. Samples used for the mtDNA analysis. Labels match the phylogenetic tree presented in Figure 3. Labels with asterisks correspond to samples with the *M. l. euryxanthus* phenotype.

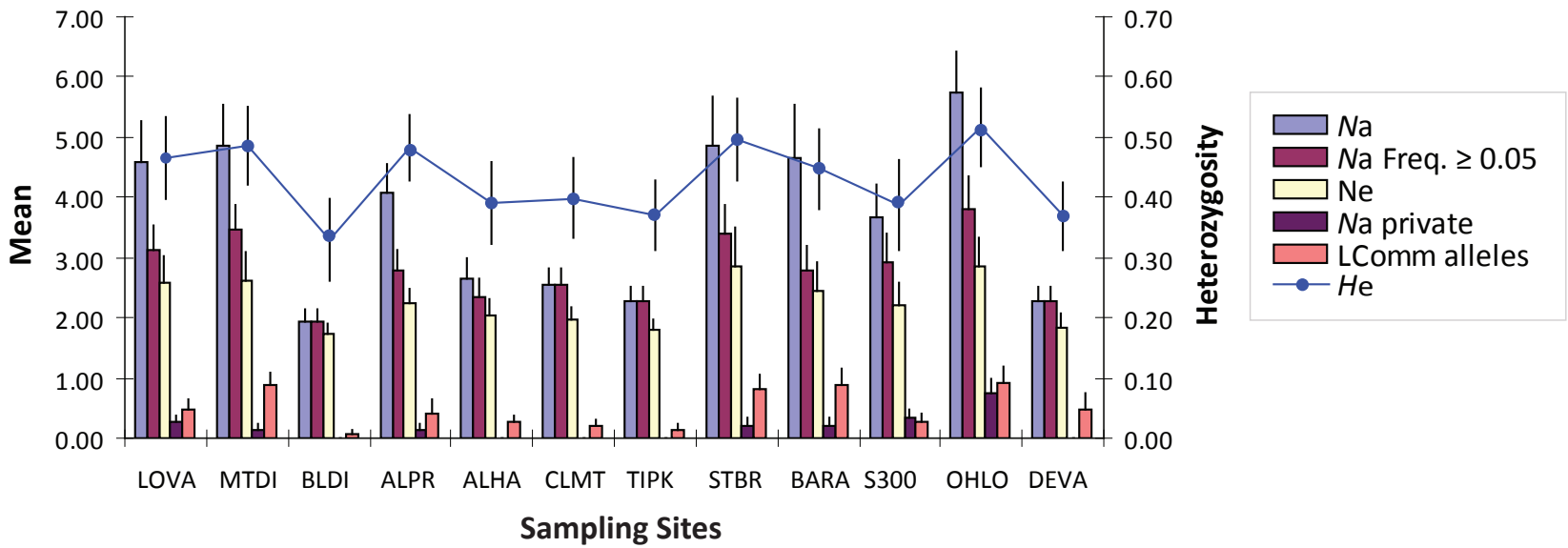
Appendix 8.3. Characteristics of 16 STR loci for *Masticophis lateralis euryxanthus.* Repeat motifs and PCR fragment lengths are based on cloned alleles from a single individual. The 'Type' category refers to whether the motif had a consistent set of repeat units for a given stretch of sequence (i.e. perfect), or contained stray nucleotide insertions at some location within the STR (i.e. imperfect). Forward primers are listed first.

Primer	***Length (nt)***	***Motif***	**N***a*	***Type***	***Primer sequences (5'-3')***	**H*o***	**H*s***	**H*t***
*Mle*A7	119	CA	7	perfect	AAAAGAGAATGGCTGAGGTATC AAGCAATAGTCAAGACTCCATG	0.381	0.359	0.397
*Mle*A114	281	CA	7	perfect	ATTCCCGTATCAAGAGATGG AGCACAAGCCAGAAGGAC	0.496	0.532	0.700
*Mle*B10	153	GTT	1	perfect	GCTAAGCAAGGCAGGAGTG GCAAAGTCAATGTGGAAAGC	0.000	0.000	0.000
*Mle*B107	137	GTT	5	perfect	CCCAGTTTTGTAGCAGATGC AGGCAAGAAGTTTGGTTGG	0.167	0.287	0.371
*Mle*C10	292	ATC	9	perfect	TTAGGGACTCTAATGGTTTGAG CACCCAGTCAGCTTTTGTAC	0.482	0.511	0.712
*Mle*D11	177	TAGA	9	perfect	TAATGGTTGACCTATGTCAGC CATTCTGCAAGCTACTGAGAG	0.420	0.699	0.792
*Mle*A109	115	AC	7	imperfect	GCTCTGTCCTGGTATTGATG GGGTTGCTCCTAAACTGAAC	0.281	0.319	0.453
*Mle*D122	259	TAGA	24	perfect	CACAGCAGATGTAACCATGTC AGCACTGTTTCCCAATATGAC	0.745	0.809	0.912
*Mle*D5	192	TCCA	16	perfect	AGCCATCTGTCTGTCCCTCT CGGAAGGATAGAAGGCTGAG	0.598	0.546	0.634
*Mle*A101	157	TG, GA	10	perfect	GAATCTGAGGCACCTTCTGAG CTCTTGAAATGGAGCCAACAC	0.328	0.368	0.467
*Mle*B105	180	GTT	8	perfect	CTCCAGGTTGCTTTTCTCCTT GTCGGGTTTGAATGTGATCATG	0.342	0.346	0.701

Appendix 8.3 Characteristics of STR loci.

*Mle*C110	209	ATC	7	perfect	CAATTTCTTCAGGATCATCACC GCAAAATATGGACAGTAACC	0.254	0.204	0.261
*Mle*C116	189	CAT	4	perfect	AACTCATCATTTGGTTCTCCC TAGATATGAAGCCCTAGGAAG	0.234	0.274	0.367
*Mle*A4	217	CA	4	perfect	CAAACAGGCATCAGGATAAT TTCCCTCAGTCCCACTTAC	0.427	0.440	0.533
*Mle*C108	282	TCA	9	imperfect	TCTTTGGTGAATCCCACTTA CAACCTGTCGGTATCACTTC	0.557	0.555	0.662
*Mle*C115	140	ATC	9	perfect	TGAGGTGAATGTTGCATACAG GGGCATGTAATAGAGAGTGACC	0.592	0.575	0.637

Appendix 8.4. Allelic patterns across sampling sites in the East Bay area. Figure notations are as follows: *N*a = no. of alleles; *N*a Freq. ≥ 0.05 = no. of different alleles with a frequency ≥ 5%; Ne = no. of effective alleles; *N*a private = no. of private alleles; LComm alleles = no. of locally common alleles (Freq. ≥ 5%) found in 25% or fewer populations; *He* = expected heterozygosity. Note that BLDI, ALHA, CLMT, TIPK, and DEV have samples size of 11 or fewer individuals.



Appendix 8.5. Demographic inferences for nested clades with significant phylogeographic structure, as determined by χ^2 contingency tests. The 'Chain of inference' and 'Inferred process', which have been the main targets of NCPA controversy in recent years (discussed in Section 3.6.3 of the project summary), are from the demographic inference keys presented in Templeton (1995) and (2004).

Significant nested clades	***χ2 permutation***		
	P-Value	*Chain of inference*	*Inferred process*
Clade 1-1	<0.0001	1-2-11-12-No	Contiguous Range Expansion
Clade 2-1	<0.0001	1-2-3-4-No	Restricted gene flow w/ Isolation by distance
Total cladogram	<0.0001	1-2-3-4-No	Restricted gene flow w/ Isolation by distance

Appendix 8.6. Genetic diversity estimates for individual STR loci as calculated in SMOGD. We report the mean estimates and 95% upper-lower confidence limits for each index. H_{Sest} = nearly unbiased estimator of within-subpopulation heterozygosity (Nei and Chesser 1983); H_{Test} = nearly unbiased estimator of total-subpopulation heterozygosity (Nei and Chesser 1983); G_{STest} = nearly unbiased estimator of relative differentiation (Nei and Chesser 1983); G'_{STest} = standardized measure of genetic differentiation (Hedrick 2005); $D_{_est}$ = estimator of actual differentiation (Jost 2008). * indicates loci that showed evidence of null alleles.

Locus ID	H_{S_est}	H_{T_est}	G_{ST_est}	G'_{ST_est}	$D_{_est}$
A7	0.3282	0.3480	0.0568	0.0938	0.0392
	0.2816–0.3699	0.2974–0.3954	0.0404–0.0918	0.0652–0.1525	0.0245–0.0686
A114	0.5016	0.6910	0.2741	0.6419	0.5067
	0.4699–0.5253	0.6653–0.7145	0.2419–0.3210	0.5856–0.7078	0.4508–0.5734
B107*	0.2572	0.3391	0.2416	0.3531	0.1470
	0.2176–0.2922	0.2912–0.3801	0.1665–0.3442	0.2485–0.4792	0.0939–0.2151
C108	0.6890	0.7345	0.0620	0.2451	0.1952
	0.6528–0.7124	0.7092–0.7563	0.0441–0.0947	0.1842–0.3489	0.1456–0.2815
C10	0.5767	0.6953	0.1705	0.4802	0.3734
	0.5316–0.6081	0.6633–0.7217	0.1382–0.2163	0.4103–0.5637	0.3115–0.4464
D11*	0.7469	0.7999	0.0662	0.3267	0.2790
	0.7093–0.7609	0.7796–0.8139	0.0506–0.1028	0.2587–0.4480	0.2181–0.3856
A109	0.3131	0.4279	0.2682	0.4312	0.2228
	0.2669–0.3499	0.3911–0.4620	0.1941–0.3668	0.3251–0.5518	0.1612–0.2928
D122	0.8860	0.9169	0.0337	0.3831	0.3616
	0.8672–0.8876	0.9077–0.9210	0.0326–0.0499	0.3608–0.4928	0.3387–0.4677
D5	0.6330	0.6655	0.0488	0.1610	0.1180
	0.5883–0.6642	0.6224–0.7057	0.0402–0.0725	0.1264–0.2390	0.0892–0.1809
A101*	0.5006	0.5521	0.0934	0.2181	0.1376
	0.4560–0.5377	0.5087–0.5953	0.0697–0.1373	0.1621–0.3078	0.0973–0.2038
A4	0.5179	0.5736	0.0970	0.2359	0.1539
	0.4833–0.5414	0.5520–0.5933	0.0668–0.1455	0.1682–0.3312	0.1087–0.2191
C115	0.5876	0.6095	0.0359	0.1041	0.0708
	0.5507–0.6131	0.5763–0.6374	0.0220–0.0650	0.0659–0.1789	0.0437–0.1238
B105	0.4144	0.6927	0.4018	0.7809	0.6338
	0.3773–0.4426	0.6754–0.7067	0.3600–0.4559	0.7348–0.8279	0.5850–0.6849
C110	0.1915	0.1992	0.0386	0.0508	0.0127
	0.1534–0.2281	0.1597–0.2400	0.0248–0.0662	0.0322–0.0891	0.0071–0.0243
C116	0.2815	0.3690	0.2371	0.3609	0.1623
	0.2451–0.3170	0.3298–0.4069	0.1656–0.3162	0.2585–0.4649	0.1079–0.2194

Appendix 8.7. Pairwise population Θ_{ST} values for the major mtDNA clades within *M. lateralis* (*$P < 0.001$). Note that the nominal species *Masticophis aurigulus* and *M. barbouri* are substantially less divergent from one another than any of the main clades within *M. lateralis.*

Clades	***MAAU***	***MABA***	**Southern Baja**	**Northern Baja**	**Southern CA**	**Northern CA**	**Central CA**
***MAAU* Baja**	–						
***MABA* Baja**	0.3839	–					
Southern Baja	0.8575	0.8051	–				
Northern Baja	0.9301	0.8998	0.9125*	–			
Southern CA	0.9358*	0.9166	0.9250*	0.8106*	–		
Northern CA	0.9386*	0.9193	0.9349*	0.8106*	0.7665*	–	
Central CA	0.9725*	0.9652	0.9699*	0.9061*	0.8310*	0.7653*	–

Appendix 8.8.1. STR distances; pairwise θ/1-θ (above diagonal), and F_{ST} (below diagonal).

	STBR	DEVA	ALPR	BARA	CLMO	BLDI	ALHA	LOVA	CAOH	MTDI	TIPK	S300
STBR	–	0.1177	0.2750	0.0166	0.2765	0.0992	0.3522	0.1437	0.1335	0.1472	0.2234	0.2147
DEVA	0.1053	–	0.2777	0.1255	0.3800	0.2628	0.5456	0.2277	0.0883	0.2316	0.2836	0.2088
ALPR	0.2157	0.2173	–	0.3240	0.1739	0.0812	0.1943	0.2504	0.3732	0.2232	0.1011	0.4924
BARA	0.0163	0.1115	0.2447	–	0.3273	0.1329	0.4153	0.1753	0.1201	0.1724	0.2579	0.1859
CLMO	0.2166	0.2753	0.1481	0.2466	–	0.3130	0.3318	0.2792	0.4041	0.2459	0.0392	0.5406
BLDI	0.0902	0.2081	0.0751	0.1173	0.2384	–	0.3048	0.0680	0.1343	0.0414	0.1747	0.2742
ALHA	0.2604	0.3530	0.1627	0.2934	0.2492	0.2336	–	0.3921	0.4169	0.3523	0.1385	0.4865
LOVA	0.1257	0.1855	0.2003	0.1491	0.2183	0.0637	0.2817	–	0.2741	0.0227	0.2530	0.3530
CAOH	0.1177	0.0811	0.2718	0.1072	0.2878	0.1184	0.2942	0.2151	–	0.2894	0.2820	0.1805
MTDI	0.1283	0.1880	0.1825	0.1471	0.1974	0.0397	0.2605	0.0222	0.2245	–	0.2259	0.3543
TIPK	0.1826	0.2209	0.0918	0.2050	0.0377	0.1487	0.1216	0.2019	0.2200	0.1843	–	0.3757
S300	0.1768	0.1727	0.3299	0.1568	0.3509	0.2152	0.3273	0.2609	0.1529	0.2616	0.2731	–

Appendix 8.8.2. Mitochondrial DNA distances (Θ_{ST})

	ALPR	TESL	CLMO	BLDI	ALHA	DEVA	STBR	CAOH	MTDI	PANO	TIPK	HECO	S300	LOVA
ALPR	–													
TESL	0.8335	–												
CLMO	0.2500	1.0000	–											
BLDI	0.2500	1.0000	1.0000	–										
ALHA	0.3617	1.0000	1.0000	0.0000	–									
DEVA	0.8335	0.0000	1.0000	1.0000	1.0000	–								
STBR	0.7077	0.9334	0.7876	0.8985	0.9163	0.9334	–							
CAOH	0.7096	0.1802	0.6711	0.7470	0.8119	0.1802	0.8152	–						
MTDI	-0.0086	0.7295	0.3205	-0.0912	0.0622	0.7295	0.6812	0.6302	–					
PANO	0.7274	1.0000	1.0000	1.0000	1.0000	1.0000	0.8985	0.5304	0.6130	–				
TIPK	0.4000	1.0000	0.0000	1.0000	1.0000	1.0000	0.8362	0.7789	0.5032	1.0000	–			
HECO	0.8335	0.0000	1.0000	1.0000	1.0000	0.0000	0.9334	0.1802	0.7295	1.0000	1.0000	–		

Appendix 8.8 Genetic distances used in the isolation-by-distance analyses

S300	0.7277	0.2501	0.7003	0.7696	0.8398	0.2501	0.8374	0.2626	0.6276	0.5716	0.8156	0.2501	–	
LOVA	-0.0073	0.9061	0.6809	-0.1321	0.0000	0.9061	0.8167	0.7461	-0.0856	0.8548	0.7736	0.9061	0.7665	–