When Does Gene Flow Stop? A Mechanistic Approach to the Formation of Phylogeographic Breaks in Nature

by

Iris Holmes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Ecology and Evolutionary Biology
in the University of Michigan
2020

Doctoral Committee:

Assistant Professor Alison Davis Rabosky, Chair
Research Professor Liliana Cortés Ortiz
Professor Patrick Schloss
Associate Professor Stephen Smith

Iris A. Holmes

iholmes@umich.edu

ORCID iD: 0000-0001-6150-6150

Dedication

I dedicate this thesis to Michael Grundler, who is always there.

Acknowledgements

The research in this dissertation was supported by funding from the University of Michigan, including the Department of Ecology and Evolutionary Biology, the Museum of Zoology, and the Rackham Graduate School. It was also supported by grants from the Bureau of Land Management, and the STEPS Institute for Innovation in Environmental Research at the University of California. The research in my dissertation was greatly facilitated by the National Science Foundation Graduate Research Fellowship, the Rackham Predoctoral Fellowship, and the Rackham Graduate School Anna Olcott Smith Women in Science Award.

I would like to thank my adviser, Alison Davis Rabosky, for her care and attention in developing both my strengths and weaknesses as a scientist. I would also like to thank the rest of my committee, Patrick Schloss, Stephen Smith, and Liliana Cortez Ortiz, for their help and support in completing my dissertation. In addition, I have had the privilege to work with excellent coauthors on the manuscripts in this dissertation, including Maggie Grundler, William Mautz, Ivan Monagan Jr, and Mike Westphal.

I would like to thank the communities in the University of Michigan Museum of Zoology, and the UMMZ Department of Herpetology and Genomic Diversity Laboratory in particular, as well as the University of Michigan Department of Ecology and Evolutionary Biology. I am particularly indebted to the Davis Rabosky and Rabosky labs for their help and support of all kinds.

Table of Contents

List of Tables

List of Figures

Abstract

I present a dissertation that examines the stochastic and deterministic predictors of population genetic demes between populations of organisms in nature. Gene flow patterns can increase the demographic viability of a population by increasing genetic diversity, and therefore reducing inbreeding depression and improving the population's ability to adapt to changes in their biotic or abiotic environment. Conversely, gene flow patterns can negatively impact demographic health by flooding locally adapted phenotypes and reducing the population's overall fitness, causing outbreeding depression. On the species level, disruptions in gene flow allow species to accumulate a variety of adaptations, and therefore provides the underlying variation necessary for the first step in diversification. In my dissertation, I use a variety of empirical and theoretical approaches to examine the mechanisms behind reductions in gene flow between natural populations. I propose that there are three common contributors to a breakdown in gene flow between populations: 1) the existence of an abiotic barrier that reduces migration by imposing intractable physiological costs on organisms that try to cross it, 2) ecological or behavioral properties of a population that reduce dispersal, such as reluctance to cross open areas that the species is still physically able to move through or 3) failure of dispersing individuals to survive in a new environment long enough to breed, possibly due to interspecific interactions, including predation and parasitism. The mechanisms behind these barriers are not mutually exclusive. Further, signatures of low gene flow between natural populations can reflect both deterministic barriers and stochastic patterns of movement of alleles across a landscape. Untangling all of this complexity requires a variety of theoretical and empirical approaches. I present empirical methods to identify geographic barriers that reduce gene flow between populations, and to identify cases in which genetic deme boundaries are not related to

abiotic or biotic barriers but instead reflect stochastic patterns of migration. For this work I use night lizards from the genus *Xantusia*, including *Xantusia vigilis* and *Xantusia riversiana.* I use the discordance between SNP data derived from a ddRAD sequencing and microsatellite markers to better understand the demographic circumstances that result in the maintenance of phylogeographic breaks in the absence of a clear barrier to dispersal. I use an individual-based stochastic simulation inspired by natural polymorphic systems including *Sonora* snakes, *Oophaga* dart frogs, and *Heliconius* butterflies to identify mechanisms behind reduction gene flow due to predation. I propose a pipeline to use existing next-generation sequence data to integrate parasite biogeographic patterns into examinations of the mechanisms behind host population genetic patterns. My work on the complex and contingent nature of barriers to gene flow suggests that a hypothesis-testing framework, in which a suite of potential mechanisms are sequentially ruled out by the available evidence, might be the most productive approach to understanding patterns of genetic diversity in nature. Such an approach would be useful in conservation genetics, particularly in planning to maintain corridors for dispersal. Further, I suggest that improved understanding of the physiological tolerance of species of interest, and improved understanding of their biological context, can improve our predictions about the types of scenarios that lead to reductions in gene flow and ultimately set populations on the path to diversification.

Chapter 1 Introduction

What processes allow an allele to spread across geographic space? What processes prevent that movement from occurring? If all species consisted of individuals that had identical ecological and behavioral traits, and if biological communities were homogeneous throughout the world, answering this question would be straightforward. Examination of empirical systems show diverse examples of populations maintaining contiguity over challenging terrain, only to stop at seemingly insignificant, or even undetectable, physical barriers (Barber, Palumbi, Erdmann, & Moosa, 2002; Jolly, Jollivet, Gentil, Thiébaut, & Viard, 2005; Leavitt, Bezy, Crandall, & Sites Jr, 2007). Given these patterns, barriers are best viewed as combinations of abiotic environment, biotic interactions, and the characteristics of the individual organisms in the populations that cease to expand (Irwin, 2002; Marshall, Monro, Bode, Keough, & Swearer, 2010; Zamudio, Bell, & Mason, 2016).

Gene flow, the process of an allele moving from one location to another, requires organisms to both move across space and leave viable offspring in their new location (Slatkin, 1985). This process can have both positive and negative demographic consequences for the recipient population (Frankham, 2015; Frankham et al., 2011). On the positive side, migration can be a major source for genetic diversity within populations (Kronenberger et al., 2017; Luijten, Kery, Oostermeijer, & Den Nijs, 2002). Genetic diversity is the raw material for adaptation within a population (Eizaguirre & Baltazar-Soares, 2014), allowing populations to face changing abiotic or biotic circumstances. Adaptive potential can determine long-term population viability, and can influence the species' ability to compete with other organisms in its community (Antonovics, 1976). Ultimately, species with high connectivity among populations may have a net survival advantage in their local community compared to those that are more isolated.

Conversely, differentiation between populations can allow evolutionary novelties to persist when they might be swamped in a more-connected population (Keller, Kollmann, & Edwards, 2000). High levels of gene flow from other populations can swamp locally adapted alleles, reducing the overall fitness of the population. At the scale of an entire species, lower gene flow can therefore increase net genetic and adaptive diversity (Lynch, 1991). Hypothetically, lower gene flow could therefore lead to higher total adaptive potential for a species.

To better understand the mechanisms behind population differentiation, I focus on areas of species' ranges with low or nonexistent migration that reveal the conditions under which the processes that facilitate gene flow break down. I propose that there are three common contributors to a breakdown in gene flow within a species: 1) the existence of an abiotic barrier that reduces migration by imposing intractable physiological costs on organisms that try to cross it, 2) ecological or behavioral properties of a population that reduce dispersal, such as behavioral reluctance to cross open areas that the species is still physically able to move through (Laurance, Stouffer, & Laurance, 2004) or 3) failure of dispersing individuals to survive in a new environment long enough to breed, possibly due to interspecific interactions, such as predation on mimic butterfly morphs move into an area in which they do not match local toxic models (Kapan, 2001). Any one of these factors alone might be significant enough to disrupt gene flow, or a break might be a result of two or more factors working in combination. Developing a working, generalizable model of the factors that restrict gene flow therefore requires incorporating the role of contingency and multiple causal mechanisms.

For almost as long as population genetics has been a recognizable field, scientists have been working on developing theoretical frameworks to handle the contextual nature of gene flow between populations (Haldane, 1924; Mayr, 1954; Slatkin, 1987). Researchers throughout the 20th and 21st centuries have seen the benefits of using ecological principles to understand the processes that lead to reduction in gene flow between populations. In 1947, RA Fisher and EB Ford integrated the nascent fields ecology and genetics to quantify the spread of an allele causing dark pigmentation in a population of moths (Fisher & Ford, 1947). Prior to the discovery of DNA, Fisher and Ford relied on breeding

experiments to understand inheritance patterns of visible traits in their focal species. With the data and analysis methods available at the time, Fisher and Ford were unable to determine the precise adaptive mechanisms behind the allele's spread, but they were confident that examining both natural history and genetics would be a fruitful path for the future.

By the 1970's, allozyme studies began to allow geneticists to identify allele frequency differences between natural populations. Doing so allowed researchers to connect allele-frequency traits, such as low heterozygosity, to demographic trends in natural populations. In 1976, Janis Antonovics synthesized the effort to connect within-population demographic trends to the function of species within their ecological communities. He formalized a set of rules for drawing a direct line from population genetic processes to ecological characteristics of natural populations, drawing on the ways he and others used these ideas in the literature (Antonovics, 1976). Specifically, he identified natural selection as both a genetic process and an ecological one, and asserted that it should be analyzed in both modes at once. Ecological roles create selection pressure, and a population's response to that pressure can change its ecological niche.

As direct sequencing of DNA became widely available through the 1980s and 1990s, researchers were able to track truly neutral traits in populations through time. John Avise coined the term 'phylogeography' to describe the process of placing mitochondrial gene trees onto maps to determine how intraspecific patterns of genetic similarity mapped onto geographic space (J C Avise et al., 1987). Taking inspiration from useful ecological generalizations such as Haldane's rule, Avise proposed a set of phylogeographic hypotheses describing processes that govern the development of phylogeographic patterns. One of his hypotheses proposed that most genetic demes, or spatially contiguous populations in which all members are more related to each other than they are to any other member of their species, generally arise from extrinsic barriers to gene flow. While such barriers are common in natural systems, this emphasis combined with the ease of measurement of potential geographical barriers may have guided the field toward a more absolute and less contextual view of the restriction of gene flow than is warranted by the available evidence.

In the genomic era, we have more tools than ever to identify population histories and understand the biotic communities in which they occur. At the same time, we have come to realize that not all genes flow across the landscape in the same way (Latta & Mitton, 1997; Teeter et al., 2007). The earliest examples of the differing patterns of movement between different types of genetic marker, called marker-type discordance, came from comparisons between phylogeographic signals in mitochondrial and nuclear DNA. In species with sex-biased dispersal patterns, the maternal-only inheritance of mitochondrial DNA can result in discordance in geographical signal between mitochondrial and nuclear DNA (Melnick & Hoelzer, 1992). As methods for examining different types of genetic markers proliferated, so too did instances of discordance in signal between marker types (Battey & Klicka, 2017; Fontenot, Makowsky, & Chippindale, 2011; Lemaire, Versini, & Bonhomme, 2005; Toews & Brelsford, 2012). These observations strengthen the inference that barriers to gene flow are context-dependent, as they show that even the variation in selection regimes on different genotypic markers can result in strongly contrasting deme membership within a single individual.

This dissertation includes two chapters that examine empirical predictors of a population's migration ability across physical barriers. I show that by integrating information from within-population demographic history and between-population gene flow, it is possible to improve our understanding of whether a geographic feature is an absolute barrier that permanently impedes all gene flow on time scales relevant to population genetics, or a contextual barrier that could allow gene flow if biotic or climatic conditions change. I examine one case in which geographic barriers have a strong predictive effect on gene flow patterns, and another case in which populations that have strong signals of differentiation in the absence of observable physical barriers. My second two chapters deal with the ecological context of barriers. I use a simulation study to understand how morph-specific predation can shape genetic diversity within a population and gene flow between populations, and I present a method for identifying parasite DNA in genome-scale datasets. My parasite pipeline can be applied to questions about the ways parasite faunas shape patterns of migration in their host species.

Together, my dissertation covers three contributors to reduced gene flow: abiotic barriers, ecological conditions that result in reduced migration, and community interactions that prevent migrants from successfully establishing in a new population. Following the efforts of many researchers interested in both patterns of genetic diversity across landscapes and the eco-evolutionary processes that generated them, my dissertation seeks to maximize the efficacy of the data available to me in understanding the natural processes behind gene flow. As in all such studies, the processes I investigate took place over many generations. Thus, they are not directly observable. Here, I address that challenge by using newly available tools, including genetic markers with differing mutation rates, simulation studies based on recently-identified genetic architectures from natural species, and newly-available geological data that contextualizes the histories of my focal species. Used together, these tools advance the state of knowledge of the context behind the regulation of gene flow between natural populations.

Chapter 2 Historical Environment is Reflected in Modern Population Genetics and Biogeography of an Island Endemic Lizard (*Xantusia riversiana reticulata*)[1]

2.1 Abstract

The restricted distribution and isolation of island endemics often produces unique genetic and phenotypic diversity of conservation interest to management agencies. However, these isolated species, especially those with sensitive life history traits, are at high risk for the adverse effects of genetic drift and habitat degradation by non-native wildlife. Here, we study the population genetic diversity, structure, and stability of a classic "island giant" (*Xantusia riversiana*, the Island Night Lizard) on San Clemente Island, California following the removal of feral goats. Using DNA microsatellites, we found that this population is reasonably genetically robust despite historical grazing, with similar effective population sizes and genetic diversity metrics across all sampling locations irrespective of habitat type and degree of degradation. However, we also found strong site-specific patterns of genetic variation and low genetic diversity compared to mainland congeners, warranting continued special management as an island endemic. We identify both high and low elevation areas that remain valuable repositories of genetic diversity and provide a case study for other low-dispersal coastal organisms in the face of future climate change.

1 **Iris Holmes**, William Mautz, Alison Davis Rabosky. Historical environment is reflected in modern population genetics and biogeography of an island endemic lizard (*Xantusia riversiana reticulata*) (PLoS ONE, Vol. 11(11), Article e1063738, November 2016)

2.2 Introduction

Islands, due to their isolation, often support suites of highly endemic species and contain some of the most threatened habitats in the world (Courchamp, Chapuis, & Pascal, 2003; N. Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000; Rumeu, Afonso, Fernández-Palacios, & Nogales, 2014). Islands face similar threats to mainland habitats, such as habitat degradation and global climate change (Paolucci, MacIsaac, & Ricciardi, 2013; Powell & Lenton, 2013). However, island endemics are also sensitive to loss of genetic diversity and stochastic population fluctuations caused by the small size and isolation of their habitat (Grueber, Wallis, & Jamieson, 2013; Ringsby, Sæther, Jensen, & Engen, 2006). Insular species can be particularly vulnerable to invasive species, through predation, competition, or habitat destruction (Coblentz, 1978; Tabak, Poncet, Passfield, & Martinez del Rio, 2014). Despite these challenges, islands also offer the opportunity to completely eradicate invasive species, which is not often feasible on continental scales (Kappes & Jones, 2014; Monks, Monks, & Towns, 2014; Spatz et al., 2014). Improving conservation techniques for island endemics is a goal of global importance, as island species often represent unique phenotypic or genetic diversity that is of high priority to wildlife managers. Maintaining genetic connectivity and genetic diversity are central concerns in conservation genetics (Culver, Hedrick, Murphy, O'Brien, & Hornocker, 2008; Frankham, 2015; Hedrick & Fredrickson, 2010). Species with relatively low vagility present a particular challenge because genes may not move freely despite having no discernible barriers to gene flow (Lusini et al., 2014; Mouret et al., 2011). In such species, loss of a small area of habitat can represent a substantial reduction in the overall genetic diversity and evolutionary potential of the population, if that habitat hosts a genotype that has not diffused through the population (Lesica & Allendorf, 1995). As such, thorough survey of the available genetic diversity is a critical step in conservation planning for the maintenance of that diversity. Moreover, concentrations of genetic diversity can form in sections of habitat that were formerly connected to large populations that have recently contracted, as simulation studies indicate that many generations must pass at low population size before genetic diversity is lost (Garza & Williamson, 2001). Identifying these

hotspots may be facilitated by both genetic approaches and knowledge of the historical extent of the species.

2.2.1 San Clemente Island biogeography

San Clemente Island, like the other seven California Channel Islands, formed as the Farallon plate subducted under the North American plate during the Miocene (Merifield, Lamar, & Stout, 1971; Stock & Lee, 1994; Ward & Valensise, 1996). At that time, San Clemente Island may have been connected to Baja California, far to the south of its current position (Crouch, 1979; Noonan et al., 2013). After the Miocene, the subduction zone moved away from southern California, and the counterclockwise rotation of the Pacific Plate led to the formation of a field of roughly parallel faults with northwestward movement, one of which continues to push San Clemente island north, west, and upward (Hauksson, Kanamori, Stock, Cormier, & Legg, 2014; Stock & Lee, 1994). The tectonic uplift resulted in the carving of twenty sea terraces currently visible on the west side of the island (Muhs, 1983). The oldest terrace, currently 300 meters above sea level, was carved 1.25 mya, while the youngest, five meters above sea level, was carved 90,000 years ago (Muhs, 1983). At the time of the last glacial maximum (LGM), global sea levels were 120 meters lower than current levels (Porcasi, Porcasi, & O'Neill, 1999). The land exposed by lowered sea levels increased the area of San Clemente Island to 1.7 times its current size, expanding the total land area available to the island organisms. Most of the exposed land was on the western side of the island. By 2000 YA, the global sea level had stabilized close to its current position (Porcasi et al., 1999). This large, geologically recent habitat reduction may have affected the current population genetic patterns of terrestrial species on the island.

2.2.2 *Xantusia riversiana* and its phylogenetic relationships

Xantusia is a genus of secretive lizards that occurs in the western US and Mexico (Noonan et al., 2013). *Xantusia riversiana* is the largest member of the genus at 80–110 mm snout-vent length and is a classic example of island gigantism (Mautz, 1987). A true island endemic, it only occurs on three of the California Channel Islands (asterisks in Fig 2.1a): San Clemente Island, Santa Barbara Island, and San

Nicolas Island (where it is considered a separate subspecies, *Xantusia riversiana riversiana*). Night lizards tend to exhibit high site fidelity, with adults spending their 20–30 year life spans within small territories. Over six months, 45% of marked *X. riversiana* on San Clemente Island were found at their original capture locations, while the others dispersed an average of 3 m (Mautz, 1987). As young *Xantusia* rarely disperse far from their natal site (an average of 4.2 meters in Xantusia vigilis; Alison R. Davis, 2012), their population genetic distribution reflects demographic events both ancient and modern (Noonan et al., 2013).

Xantusia riversiana may be the sister species to all other members of the genus, from which it diverged approximately 14 to 16 MYA (Noonan et al., 2013; but see Vicario, Caccone, & Gauthier, 2003) and (Sinclair, Bezy, Bolles, & Sites, 2004) for alternative tree topologies and divergence time estimates). At that time, the flora of Baja California was primarily tropical deciduous forest, with evergreen broadleaf forest present in riparian areas (Martínez-Cabrera, Cevallos-Ferriz, & Poole, 2006). Fossil tree species from areas in Baja California near the putative attachment point of Miocene San Clemente Island most closely resemble extant tropical dry forests on the west coast of Mexico, which is within the current range of *Lepidophyma*, the sister genus to *Xantusia* (Noonan et al., 2013). If this phylogenetic hypothesis is correct, the ancestral *Xantusia* had split from *Lepidophyma* and occurred in Baja California 20 million years ago. San Clemente Island's separation from the mainland may have divided the ancestral *X. riversiana* from the rest of the *Xantusia* lineage, which subsequently speciated on the mainland. Alternatively, the islands could have been colonized by dispersal over water after they separated from the mainland. Both San Nicolas and Santa Barbara islands were completely submerged at times during the Miocene, while San Clemente Island remained above water, making it likely that the San Clemente population is ancestral to the populations on other islands (Hauksson et al., 2014; Ward & Valensise, 1996).

2.2.3 Human impacts on San Clemente Island

San Clemente Island's vegetation was severely damaged by invasive goats between their introduction in 1875 (Keegan, Coblentz, & Winchell 1994) and removal in 1991 (Wylie, 2012). In some areas, goat herbivory catalyzed a change from the endemic marine sage scrub to grassland that is largely composed of invasive Mediterranean grasses. These disturbances are expected to have had negative conservation consequences for *X. riversiana* by removing prime habitat that supports the highest density of lizards (Mautz, 1993), but to date there has been no investigation of genetic population structure across habitat types to help inform management decisions. The native flora has been rebounding in the absence of the goats (Keegan, Coblentz, & Winchell 1994).

In this paper, we assess the population genetics and historical biogeography of the San Clemente Island Night Lizard, *Xantusia riversiana reticulata*, a species endemic to the Channel Islands of California, USA (Fig S2.1 and Fig 2.1a). First, we use DNA microsatellite markers augmented by genome-scale SNP data to estimate population genetic diversity, structure, connectivity, and stability in order to identify populations of particular conservation genetic importance. Second, we use the island's history to examine the relative genetic impacts of two habitat disruptions: 1) the drastic reduction in size of the total island due to post-Pleistocene sea level rise, which occurred around 15,000 years ago, and 2) habitat degradation due to introduced goat herbivory, which began 150 years ago. Together, these analyses provide important insight into past demographic history, contextualize future management priorities and strategies, and offer a case study for comparison to other low-dispersal island organisms that have had significant habitat disturbance due to invasive species.

2.3 Methods

2.3.1 Ethics Statement

The U.S. Department of the Nave supported field collection and granted access to San Clemente Island, and the U.S. Fish and Wildlife Service issued scientific collecting permits to WJM. All methods

were approved by the Chancellor's Animal Research Committee (protocol #Sine0002-1) at the University of California, Santa Cruz.

2.3.2 Tissue collection and site descriptions

We collected 530 tissue samples from 12 sites across San Clemente Island between 2005–2007 (Fig 2.1b; see Table 2.1 for site-specific sample sizes).We captured lizards through a combination of rock turn surveys and a previously established grid of pitfall traps (Mautz, 1993). At each capture, we measured the mass, snout-vent length (SVL), tail condition (broken, regenerated, intact), and determined the lizard's sex by shining a light through the lizard's tail base to visualize hemipenes in juvenile males (Alison R. Davis & Leavitt, 2007) and by assessing hemipenal bulging to identify adult males. Because population monitoring was ongoing, some captured individuals had already been toe clipped during previous population sampling, in which case the lizard's clip combination and status as a recapture were noted. We toe-clipped each newly captured individual with a unique combination for future identification. We took a 0.5–1 cm piece of tail tissue (stored in 95% ethanol) from every lizard for genetic analyses. We did not use analgesia, as toe-clipping with sharp scissors does significantly not increase lizard stress-hormone levels and the increased manipulation necessary to use analgesia could potentially further stress the lizards (Langkilde & Shine, 2006). We immediately released all lizards at their exact location of capture and recorded the latitude and longitude coordinates of each location using Magellan® eXplorist® 300 GPS units. We classified habitat by the dominant vegetation type present at each collection site across the island and following the Holland Code classification system (Holland, 1986; Fig. S2.2). The Maritime Succulent Scrub (MSS) habitat was split into two subcategories based on whether the structurally dominant plant species was boxthorn (*Lycium sp.*; MSS-Ly, N = 5 sites) or cactus (*Opuntia* or *Cylidropuntia sp.*; MSS-Op, N = 4 sites). The grassland habitats (N = 3 sites) were any location with more than 75% grass (*Stipa sp.*) cover, even if the landscapes were a composite that included some minor component of MSS or other habitat (Fig 2.1b).

2.3.3 Primer design and locus amplification

To identify variable microsatellite loci with reliable amplification within *X. riversiana*, we a) screened primer sets designed for microsatellites in the congener *X. vigilis* (A. Davis, Corl, Surget-Groba, & Sinervo, 2011) and b) created a de novo enriched library of microsatellite loci using DNA extracted from four *X. riversiana* from San Clemente Island. DNA was extracted from tail tips stored in ethanol using Qiagen DNEasy kits, and four individuals were pooled for library construction. We then screened this genomic DNA for di- and tetranucleotide (CA and AAAG) repeats following the protocol in (Gow, Johansson, Surget-Groba, & Thorpe, 2006). From this screen, we sequenced 96 prospective clones and identified 60 unique microsatellite motifs, including the motifs screened for and motifs we opportunistically identified in the cloned sequences. (GenBank accession numbers KT696132-KT696166).We designed primers using PRIMER3 (Rozen & Skaletsky, 1999; Table S2.1). Six of these presumptive loci, as well as two of the loci from X. vigilis (N = 8 total loci), amplified reliably. We simultaneously amplified these loci with Qiagen Multiplex PCR Kits according to manufacturer's instructions (final reaction volume was 10µl), with at least 5 loci amplifying in 516 of the samples (Dryad repository DOI: doi:10.5061/dryad.6c7p5).

Since we had samples from multiple years in several sites, we ensured that each genotype was unique. We found no duplicate genotypes in our sample. We tested for differentiation between year-cohorts in the populations that had been sampled in multiple years (Table S2.2). We used a per-locus exact G test implemented in Genepop with 100 batches of 100 iterations each, with 1000 dememorization steps (Raymond & Rousset, 1995; Rousset, 2008). Loci were largely undifferentiated between years, with the exception of three populations (LA, SC, and ST) that had less than ten individuals sampled in 2005, and HN and HS, which showed differentiation in one and four loci, respectively. With this caveat in mind, we combined genotypes across years for our remaining analyses. To assess the quality of our markers, we calculated the number of alleles per locus, the observed and expected heterozygosity, polymorphism information content, and random match probability for each locus (Hameed, Ommer,

Murad, & Mohammed, 2015); Table S2.1). We use the last two metrics as a proxy for the reliability of our dataset for distinguishing population-level patterns of differentiation.

2.3.4 Next generation sequencing

In a brief exploration of potential marker skew, we genotyped three *X. riversiana*, one each from high, medium, and low elevations, using a double-digest RADseq approach followed by sequencing on an Illumina HiSeq platform (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Using the program pyRAD (Eaton, 2014), we clustered the resulting sequences at 85% similarity across the three individuals, and retained the sequences that had at least one polymorphism, resulting in 3136 informative loci. For comparative purposes, we include seven individuals from the congeneric *Xantusia vigilis* that were sampled from populations in mainland central California (data not shown).

2.3.5 Locus linkage and disequilibrium

To test for sex linkage, we reduced the data to the 130 individuals for which sex was positively known (N = 83 females, 47 males).We performed a chi-square analysis on the allele frequencies at each locus, and found no significant sex linkage. Three loci (XrivB1, XrivG2, and XrivY3) were homozygous for all individuals whose sex had been recorded and were not included in this analysis.

To test for deviations from Hardy-Weinberg equilibrium, we performed an exact test in ARLEQUIN (Excoffier & Lischer, 2010) using a 100,000 step Markov Chain with 1,000 burn in steps (Guo & Thompson, 1992). The test returns a P-value that indicates the two-tailed likelihood of observed heterozygosity in each locus and population. We then used MICROCHECKER (Van Oosterhout, Weetman, & Hutchinson, 2006) to identify null alleles and their frequency, large allele dropout, and stutter by simulating 1000 randomizations to find the expected numbers of heterozygotes in each sampling location, assuming Hardy-Weinberg equilibrium. We set the confidence interval to 95%, Bonferroni corrected. To test for linkage disequilibrium, we built contingency tables of observed allele frequencies at each pair of loci in ARLEQUIN. We then permuted the genotypes and used a 1000 step

Markov chain to explore the contingency table space for each pair and obtain the probabilities of the observed contingency tables.

2.3.6 Genetic diversity, gene flow, and population structure

To estimate structure among collection locations across the island, we calculated average genetic diversity and location-specific FST and FIS in ARLEQUIN. We identified private alleles using GENALEX (Peakall & Smouse, 2012) for each collection location. We calculated an indirect measure of gene flow using the formula FST = 1/(4Nm+1), rearranged to solve for Nm, the number of migrants per time step (Wright, 1931).We used BAYESASS to find an alternate measure for asymmetric migration rates between pairs of populations (Wilson & Rannala, 2003).

As a further test of population structure, we used the program TESS (Durand, Jay, Gaggiotti, & Francois, 2009) to identify genetic demes and the proportion of each individual's genome that belongs to those demes. For any individuals for which we did not have unique collection locations, we generated unique coordinates sampled from a 50 by 50 meter square centered on the coordinates for the entire collection site. For individuals caught in the same pitfall trap, we perturbed the collection location of all but one individual by one meter in a randomly chosen cardinal direction.

We computed pairwise geographic distances using the Euclidean option due to many of our samples being in close proximity. We began our runs with the no-admixture model with the spatial interaction parameter set to zero to mimic the Structure algorithm. Visual assessment of this analysis indicated that the optimal number of clusters was around 5, so we concentrated on Kmax values between 3 and 7.We then implemented the CAR admixture model. DIC values for short runs of this model indicated that the optimal Kmax was between 3 and 5.We refined this estimate by looking at the proportion of the available genomes that were assigned to each cluster. For Kmax of 4, one cluster had less than 5 percent of the available genomes assigned to it, indicating that it was a dummy cluster that had not completely emptied after 40,000 iterations. The same was true for two clusters for Kmax of 5. Thus, we concentrated our analysis on the runs with a Kmax of 3.We repeated the analysis with several

variations of subsamples versions of the two populations for which we had more than 60 individuals, and we found the results to be very robust to all permutations.

2.3.7 Effective population size

Island populations can show much lower effective population sizes relative to census size (Grueber et al., 2013; Ringsby et al., 2006). To assess the difference between effective and census size estimates, we used coalescent simulations in LAMARC to determine effective population size at each collection location (Kuhner, 2006). For locations with more than 15 individuals sampled in a year, we randomly subsampled two sets of 10–15 genotypes. We calculated the parameter theta ($4N_e\mu$) five times from each data set and averaged those calculations. Finally, we generated the mean of the means of the two subsamples of individuals. We find a relative measure of population size by standardizing our theta values by our lowest reconstructed size. We used the standard microsatellite mutation rate of 0.001 mutations per individual per generation to calculated Ne for each collection location (Sun et al., 2012).We compared the results to demographic capture rate (lizard/trap/day) data (Mautz, 2007) for the locations at which those data are available, including an estimation of capture rates at rock turn sites calculated from an average rate of two rocks flipped and replaced per minute, which includes lizard handling time (lizards/hours sampled/daily flip rate; Table 2.1).

To calculate the effective population size for all night lizards on San Clemente, we used the 'pegas' package (Paradis, 2010) in R ver. 2.12.1 (R Core Development Team 2010) to find the value for theta calculated across all individuals and all loci. We calculated three values for theta (Haasl & Payseur, 2010; Kimmel et al., 1998), and used the formula $\theta = 4N_e\mu$ to calculate effective population size. We used a value of 0.001 for the microsatellite mutation rate (Sun et al., 2012) and averaged across loci to find the mean and standard deviation for the effective population size.

We used the library 'pegas' in R to calculate theta based on heterozygosity for each SNP from our ddRAD dataset (Paradis, 2010).We calculated Ne using the standard equation $\theta = 4N_e\mu$.We set μ to the standard estimate for sequence mutation rate, 2.5×10^{-8} (Nachman & Crowell, 2000). We contrasted

this estimate with two mainland effective population size estimates. The first was from *Xantusia vigilis* from the Panoche Hills in the central part of mainland California. We also used *X. vigilis* from the Cuyama Hills. We repeated the analysis with one *X. vigilis* drawn from Pinnacles National Park, the Panoche Hills, and the Cuyama Hills.

2.3.8 Population bottlenecks

There are have been two major reductions in habitat availability in the recent history of San Clemente. The first was the forty percent reduction in size the island suffered as sea levels rose after the LGM, which would have primarily effected the low-lying night lizard populations. The second was the century-long habitat degradation by goats, which was more severe on the higher elevation areas. To determine the relative impacts of these two events, we used the Garza-Williamson (GW) index to identify any reductions in effective population size in each location (Garza & Williamson, 2001). We calculated the statistic for each locus in ARLEQUIN, and report the mean and standard deviation of those calculations. The GW index is vulnerable to inaccurate assumptions about the mutation model. As we do not know the mutation rate of our loci, we use a comparative approach in our analysis. We regard populations with relatively low GW indices as having experienced greater relative declines than those with higher indices. If a population is bottlenecked and then recovers, simulations show that the GW index will return to near its pre-bottleneck levels within 400 generations (Peery et al., 2012). Thus, any detected perturbations should indicate a fairly recent demographic event. If vegetation loss has caused a major bottleneck in grass-living populations, then this statistic would be lower in these populations than those living in Maritime Succulent Scrub (MSS) habitats.

2.3.9 Spatial genetic structure

We used a discriminant analysis of principle components implemented in the R package ‘adegenet’ to assess spatial genetic structures (Jombart, 2008). We calculated the coordinates of each individual on the first two PC axes, and from these found the centroid for each collection location. We used the ‘raster’ (Hijmans, 2015) package in R to extract the elevations of our collection locations from

the DEM, and assigned each collection location a habitat type based on observation during collection (Fig 2.1B).We performed an ANOVA on the elevations in R with the TESS deme identity of each location as a three level factor. We investigated the correlation between deme identity and potential explanatory variables by using a Generalized Additive Model (GAM) implemented in the R package 'gam' (Hastie, 2015). For each deme, we performed a GAM with just elevation, just habitat, and elevation + habitat as the explanatory variables. We calculated relative AIC values for each model calculated for each deme. The model was unable to definitively distinguish between elevation and elevation + habitat for the medium and high demes, and selected elevation as the best explanatory variable for the low deme. We therefore used elevation to extrapolate the probable deme identity of populations below the current water line. We used a bathymetry raster publicly available from the NOAA National Geophysical Data Center (National Geophysical Data Center, 2012), and used the latitude, longitude and elevation information for each raster cell to provide the independent variables for the model. The dependent variable was the percentage of identity in the TESS-identified demes at each sampling location. We did a separate GAM for each of the demes.

In order to assess the correlation between genetic distance and geographic features, we used Mantel tests implemented in IBDWS (Jensen, Bohonak, & Kelley, 2005). To generate our pairwise geographic resistance matrices, we used the program CIRCUITSCAPE (McRae, Shah, & Mohapatra, 2013). The rows and columns of the matrices were randomized and the correlation statistic was recalculated for each randomization and compared to the original statistic to generate a p-value for the significance of the correlation. We used the pairwise FST matrix from ARLEQUIN for genetic distance. CIRCUITSCAPE models a habitat raster as grid nodes, each with a specified resistance. The model calculates the total resistance between each pair of population locations.

We obtained a 7.5 minute digital elevation model (DEM) map from webGIS (U.S. Geological Survey, 2009d, 2009a, 2009c, 2009b, 2009e, 2009f).We specified habitat rasters by calculating slope from a digital elevation model raster using the 'raster' package in R (Hijmans, 2015). We calculated pairwise resistances between collection locations for the slope raster, and for a raster in which every cell

was set to the mean value, such that it had the same total resistance but did not include any slope information. This raster served as a test for strict isolation by distance, rather than isolation by resistance.

2.4 Results

2.4.1 Locus behavior, genetic diversity, and population structure

None of the five loci with variation among adults were sex linked (p-values between 0.115 and 0.630). The number of alleles ranged from four to 32, with a total of 93 unique alleles. Observed heterozygosity was less than expected heterozygosity for all loci except XrivR1. Random match probability, roughly a measure of the probability of any two individuals in the population having the same multilocus genotype, was 4.37×10^{-6}.We calculated polymorphic information content and random match probability for each locus (Table S2.1). We found an excess of homozygotes in three populations for locus B1, two populations for G1, and one population each for G2 and Y3, but no other deviations were detected. Two sites had one or more loci that were out of Hardy-Weinberg equilibrium, and every locus was out of equilibrium for at least one collection site (Table S2.3). However, there were no consistent trends in the lack of equilibrium across loci and populations, so we retained all loci for our further analysis. We found no, or very minimal, evidence of any linkage disequilibrium in most sites. However, one site on the low-lying western side of the island (HN) had three pairs of loci in linkage disequilibrium (Table S2.4).

Across all collection sites, the average proportion of polymorphic loci was 6.33/8 loci, with the minimum (two sites) having five polymorphic loci, and the maximum (two sites) having eight polymorphic loci (Table S2.5). The mean number of alleles per site was 6.17 ± 5.05 (Table S2.6). Average gene diversity (P) was not significantly different for any sampling site, with all locations showing a value near 0.4 (± 0.2). FIS values were also between -0.11 and 0.012, while location-specific FST values were between 0.098 and 0.1. The majority of the population-specific private alleles occurred in HN and HS (Table S2.6). Pairwise FST values ranged from 0.1575 between BO and LA, and 0.0047 between HN and HS (Table S2.8). We found the highest levels of migration among the populations in the

central third of the island, ranging from 15–53 migrants per generation (Fig 2.1c). All other pairwise populations exchange less than ten migrants per generation (Table S2.7). These levels of migration must be interpreted in context of the effective population sizes of the populations. BAYESASS showed the highest migration values from ES to EP, from LA and HN to HS, from SC and TE to ST, and WI to WS (Table S2.9). All of these pairs of populations are geographically close. The most distant pair is LA to HS, but this result is mirrored by the TESS analysis (Fig 2.2a).

The TESS analysis recovered three regional demes (Fig 2.2a). One deme is made up of two collection locations at the southern end of the island combined with the northernmost collection location, which is 23 km from these southern localities. A second deme covers the intervening central populations, while a third is concentrated around the lowest elevation sites of HN and HS, with considerable representation in the LA and TE sites. For convenience, we use the terms low, medium, and high to describe the demes, based on the relative elevations of the majority of the populations.

We independently estimated the deme affiliation of each collection location using Principle Components Analysis. We again recovered three groups, mainly differentiated along PC1, with membership generally concordant with the TESS analysis: the southern sites of BO and SH, the low elevation HN and HS sites, and then the remaining sites. The main difference between the two analyses is that WI does not group with BO and SH. PC2 loosely differentiates LA, WI and WS, although their convex hulls overlap other populations (Fig 2.2b).

2.4.2 Effective population size and genetic bottlenecks

Standardized population sizes ranged from 1 to 2.2, with generally larger theta values in the HS, HN, and EP collection locations. Using a mutation rate of 0.001 mutations per individual per generation, we calculated estimates for effective population size for each collection location, which ranged from ~2300–5100 individuals per site (Table 2.1).We also found that these site-specific effective population sizes were not correlated with capture rates ($F_{1,10} = 1.47$, $P = 0.254$), even when rock turn sites were excluded ($F_{1,8} = 0.64$, $P = 0.447$). For the entire island, we found estimates of effective population size to

range from 42,083 (±79,060) to 62,109 (±106,308) individuals, depending on the method used to calculate theta. Two sites on the western MSS-rich side of the island (HS and ES) show relatively low Garza-Williamson index compared to the other sites, indicating that they have undergone proportionally greater population reductions in recent history than the other sampled locations (Fig 2.2c). Their upper standard error bars fall below the 0.68 cutoff that Garza and Williamson found to indicate a history of a bottleneck in empirical and simulated populations. The ddRAD dataset had 1046 polymorphic loci across one individual each from the high elevation, middle elevation, and low elevation demes, and gave us an effective population size of 3,178,936.We combined sets of three *Xantusia vigilis* from a total of seven samples taken across central California. The effective population sizes ranged from 3,235,742 to 3,836,402 depending on the grouping we used.

2.4.3 Spatial genetic structure

We found that the first two principle component axes isolated the "high" demes from the "low" and "medium" demes (Fig 2.2b). Both the isolation by slope (P = 0.0049) and the isolation by distance (P = 0.0094) rasters showed a significant relationship with genetic distance. Reduced Major Axis regression showed a tighter correlation and higher slope (slope = 0.7402, R2 = 0.218) for the isolation by slope than the isolation by distance test (slope = 0.4402, R2 = 0.183). The output slope raster had values between 0 and 1.0037, with a mean value of 0.203 (Fig 2.1c shows slopes). Overall, we found a better, although still weak, correlation between genetic deme and elevation (t = -2.108, df = 10, P = 0.061) than between deme and habitat ($\chi 2$ = 3.45, df = 4, P = 0.486). The effect of elevation was strongest on the mid-elevation deme (F1,10 = 5.001, P = 0.049), although not significant for the high (F1,10 = 2.05, P = 0.182) or low demes (F1,10 = 2.07, P = 0.181). The GAM models for the high deme showed relative AIC's of 0.13, 0.01, and one for the only elevation, only habitat, and habitat + elevation models, respectively. The mid-elevation deme showed relative AIC scores of 0.25, 0.07, and one for the three models, and the high-elevation deme showed relative scores of one, 0.29, and 0.14 for the three models. Using the GAM models, the projected deme affiliation across the island suggests that most of the island's area has lizards from the high

elevation deme, with the low elevation (and high genetic diversity) deme restricted to only a small fraction of the total above-water island area (Fig 2.2d).

2.5 Discussion

We found no evidence of recent bottlenecks in any of the populations, regardless of the degree of degradation in the local habitat. We found evidence of an older bottleneck detected by the M-ratio tests in two low-elevation populations. Goats were introduced 150 years ago, which, assuming a lizard generation time of at least 5 years (estimated lifespan of 13+ years; (Porcasi et al., 1999)), represents no more than 30 generations for *X. riversiana* (Martínez-Cabrera et al., 2006). Bottlenecks severe enough to results in measurable reduction in population genetic diversity generally must continue for at least 50 generations, given a reasonable population size (Peery et al., 2012). Our results indicate that the eradication of the goats occurred quickly enough after their introduction that genetic diversity was maintained in even the most heavily impacted populations. This finding suggests that ongoing management for re-vegetation should have positive impacts on the island night lizard. It also indicates that other island herbivore removal projects could be useful applications of conservation effort, since similar species may also have maintained genetically healthy populations despite long-term habitat degradation. However, our relatively low number of microsatellite loci and low allelic richness of these loci mean that our results should be interpreted with caution. Further genetic work should be conducted on the night lizards to confirm our recommendations.

2.5.1 Genetic diversity of *Xantusia riversiana*

We found that pairwise genetic distance between collection locations increased with the pairwise geographic distance and the degree of slope between two populations, which could account for the roughly elevationally stratified demes. In practice, it indicates that low elevation populations should not be counted on to provide natural demographic or genetic “rescues” of higher elevation populations in this low-vagility species (Noël, Machon, & Robert, 2013). Currently, the majority of high quality MSS habitat on San Clemente Island is at low elevations, and higher elevations are still degraded from the effects of

past herbivory by introduced grazers, although these habitats have continued to recover post-removal of goats. The high elevation locations we sampled had lower capture rates than the low elevation populations, indicating potential demographic fragility in those areas, despite their lack of evidence for genetic bottlenecks (Table 2.1). The low elevation deme also contained genetic variation of conservation priority, with a relatively high proportion of private alleles and limited distribution across the island. Some low elevation populations also had a signature of a past genetic bottleneck, indicating that the deme had undergone a large and likely long-term reduction in population size in the past. This finding is consistent with the conjecture that night lizards belonging to the low deme once occupied the area that was exposed in the Pleistocene, and the deme underwent dramatic reduction as the area became submerged.

An indirect comparison of these genetic variation levels to the mainland congener *X. vigilis* (see Langkilde & Shine, 2006) suggests *X. riversiana* on San Clemente Island has about half the variation of one population of *X. vigilis* across multiple metrics (proportion of polymorphic loci, observed heterozygosity, number of unique alleles per locus). Reduced genetic variation is a classic characteristic of island fauna, and special care should be taken to preserve the diversity that does exists across the island. However, we suspect that our overall estimate of island-wide effective population size (mean N_e = 50,000) may be significantly lower than the true Ne, especially considering the estimated demographic census size (N) of approximately 21 million individuals (Sun et al., 2012). Analysis of RADSeq loci using one individual from each of the three demes on San Clemente Island yielded a Ne of 3.18 million, which is far closer to the expected Ne in a lizard system given the large size estimated for N. Further study of this species should be undertaken to better understand the true absolute and effective population sizes. The differences in effective population size estimates between the sequence data and microsatellite data may have several causes. We may not have enough microsatellite loci to adequately describe the effective population size, while the large number of loci in our RADseq data more accurately capture actual effective population size. Identifying the primary causes of the differences in estimates between the

microsatellites and sequence data will take future research on the reaction of different genetic marker types to various demographic scenarios.

As *Xantusia riversiana* also inhabits two other Channel Islands (Fig 2.1a), the greatest outstanding questions about their genetics are 1) what is the genetic diversity on the other two islands, and 2) how long ago did the populations split? The best available genetic evidence (allozyme and karyotyping data for many individuals (Bezy, Gorman, Adest, & Kim, 1980) and Sanger sequencing data for a few samples from each island (Noonan & Comeault, 2009)) suggests that these three island populations are each quite distinct and have been separated without gene flow for at least 500,000 years, with the greatest standing variation in both allelic diversity and color pattern of the three populations being found on San Nicolas Island. A thorough investigation and comparison of genetic diversity across islands will be essential to management of the species as an integrated unit.

2.5.2 Effects of previous conservation action: goat removal

The conservation concern that prompted *Xantusia riversiana*'s listing as federally threatened was habitat destruction due to invasive herbivores, farming, and invasive plants (Bezy et al., 1980). The goats have since been eradicated, and native flora is rebounding under the management of the U.S. Navy through targeted re-vegetation efforts, natural recovery, and invasive species control (Keegan, Coblentz, & Winchell 1994; Martínez-Cabrera et al., 2006). We found no greater evidence of genetic bottlenecking in areas where the vegetation was heavily degraded due to past goat herbivory than in areas in which it remained more pristine. Our results indicate that island populations may be genetically robust to some types of multidecade habitat degradation. In populations in which this is the case, invasive species removal can be a worthwhile conservation measure.

2.5.3 Future conservation focus

Since the removal of feral goats, investment into native plant restoration, feral cat management, erosion control, and the creation of Island Night Lizard Management Areas (INLMA) across key MSS habitat have helped ensure population stability of *Xantusia* on San Clemente Island. This study adds key

information about the main dispersal corridors across the middle of the island and demonstrating the importance of elevation in predicting genetic variation across sea terraces, which should help with future land management decisions.

Climate change models indicate that, in the absence of a worldwide reduction in carbon output by 2050, we should be prepared for sea level rise of at least 0.3 meters by 2100 (Horton, Rahmstorf, Engelhart, & Kemp, 2014). Although the population currently shows signs of demographic health, the sea-level HS/HN sites are at more risk than the other areas. The HS/HN sites also contain a disproportionate number of private alleles, indicating that they contain important genetic resources for the species. As slopes are barriers to dispersal, and these collection locations are surrounded by sea terraces (Fig 2.1c), the natural dispersal process of *X. riversiana* may be inadequate to offset habitat contraction. For example, sea levels rose continually between 13,500 years ago and 2,000 years ago without homogenizing the genetic differences between these sites and some of their nearest neighbors, suggesting very low up-slope dispersal. Depending on the trajectory of sea level rise, future conservation efforts may need to focus on safeguarding this population. Similar species may follow this pattern, indicating that managers should determine whether their species have clusters of genetic diversity that could be at risk due to sea-level rise.

2.6 Data Archiving

All sequences have been submitted to GenBank, accession numbers KT696132-KT696166. Microsatellite genotypes are available on Dryad, with DOI doi:10.5061/dryad.6c7p5.

2.7 Acknowledgments

We thank the U.S. Department of the Navy for supporting field collection and for granting access to San Clemente Island. WJM is contracted as a Biological Consultant for San Clemente Island by the U.S. Navy.

2.8 Funding

Funding for molecular work was provided by the STEPS Institute for Innovation in Environmental Research at the University of California, Santa Cruz and by start-up funds from the University of Michigan to ARDR

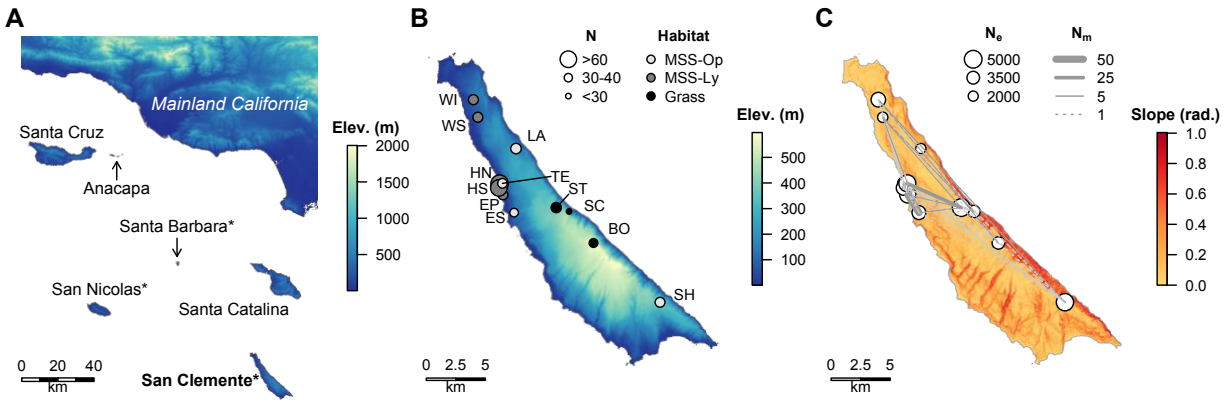


Figure 2.1 - San Clemente Island sampling sites

A) Relief map of the Channel Islands and mainland California. Island names with asterisks are inhabited by *Xantusia riversiana*, with our study island in bold. B) Relief map of San Clemente Island showing collection locations colored by habitat type and scaled by sample size (N). C) Map of San Clemente Island colored by slope steepness and showing dispersal rates among collection locations. Points are scaled by effective population size (N_e) and lines by number of migrants per generation (N_m).

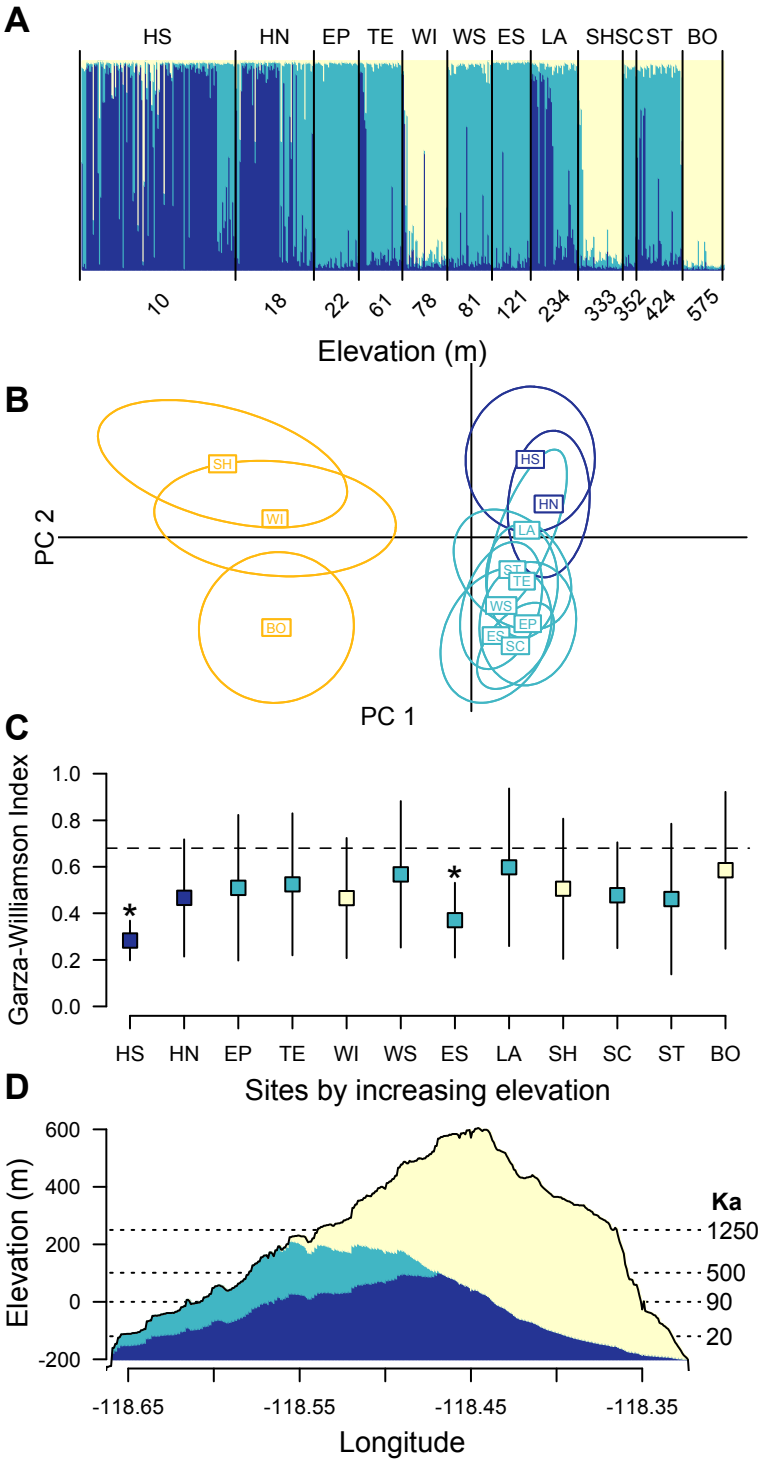


Figure 2.2 - *Xantusia riversiana* genetic demes follow elevation on San Clemente Island.

A) Percentages of deme identity found for each individual by collection location generally correspond to elevation, with a low (dark blue), mid (light blue), and high elevation (cream) deme. B) Discriminant analysis of principle components concordantly recovers three clusters. Locations are color-coded according to the TESS deme that had the highest representation at that location. C) Mean ±1 s.e.m. of the Garza-Williamson Index for each collection location shows two sites with evidence of historical bottleneck (asterisks). D) Elevational profile plot of the highest point on San Clemente Island at each longitude, colored by deme representation from the GAM prediction. Gray shading denotes area currently below sea level, and dashed lines show historical sea levels.

Table 2.1 Sampling sites on San Clemente Island

Site	N	θ_1	θ_2	mean θ	N_e	Lizards/ trap/day[2]
BO	32		12.60	12.60	3150.0	0.538*
EP	37	24.41	16.69	20.55	5137.5	0.267
ES	38	17.57	14.09	15.83	3959.3	0.300
HN	70	19.72	14.09	17.31	4577.0	0.200
HS	139	15.37	15.14	15.26	3814.3	0.224
LA	39	8.44	10.32	9.38	2343.8	0.497
SC	11		12.46	12.46	3115.0	0.067
SH	44	13.57	17.89	15.73	3930.0	0.356
ST	50	18.36	18.08	18.22	4555.0	0.214
TE	35		17.35	17.35	4335.0	0.357
WI	49	12.18	14.93	13.55	3387.5	0.254
WS	37		10.56	10.57	2642.5	0.310*

Number of genotyped individuals (N), theta values, estimated effective population sizes (N_e), and demographic trap capture rates for 12 sampling locations on San Clemente Island.
[1]2005
[2]2007
*rock turn surveys (see Methods)

Figure S 2.1 *Xantusia riversiana* photo.

Xantusia riversiana reticulata, the Island Night Lizard, from San Clemente Island, CA.

Figure S 2.2 San Clemente Island habitats

Major habitat types on San Clemente Island. A) Maritime Succulent Scrub, *Opuntia* phase (MSS-Op). B) Maritime Succulent Scrub, *Lycium* phase (MSS-Ly). C) Grassland (primarily *Stipa* sp.), with secondary invasion of shrubs (*Heteromeles* sp.) post-removal of feral goats.

Table S 2.1 Marker characteristics

Locus	Repeat Motif	Primer Sequence (5′-3′) (F=*dye*-forward, R=reverse)	Range (bp)	N	H_o	H_e	PIC	RMP	GenBank Accession #
XrivB1	CTT	F: *FAM*-TCGCATCCACCTACACAAGC R: GGTTTGGTGTGCTGCCTAGT	125-132	4	0.000	0.019	0.019	0.981	KT833329
XvGLA	AAAG	F: *FAM*-TTGCCTGTCCCAAAAGTCTC R: CCTGACTGGAAGGAGCTCAG	260-346	32	0.888	0.931	0.931	0.009	FJ197164
XrivG2	TG	F: *VIC*-ACACTCTGCTCCCCTTCAGA R: GCCCAAGGTTACCCAGTGAG	130-134	5	0.008	0.025	0.025	0.969	KT696148
XrivG1	GT	F: *VIC*-AAGCTTCGCATCCAGCAGTT R: CCCTTTCATCCGTTGCCAGA	182-220	14	0.655	0.748	0.748	0.116	KT833330
XvCHEL	AAAG	F: *VIC*-ATGTTTTCCTGTCCCAAAGG R: GGCAAGCTATCCTCTGCTTG	250-290	14	0.727	0.750	0.750	0.079	FJ197163
XrivY3	CTT	F: *NED*-AGCTTCGAGCCGATCTTGAG R: AGCAGGAACATCACTCCACG	263-271	7	0.025	0.029	0.029	0.947	KT833332
XrivR1	ACC	F: *PET*-TCCCATGCACAGCAAAAAGC R: CCTTGCCCTCCAAGAAGGTT	142-176	10	0.558	0.485	0.484	0.254	KT833331
XrivR2	CAA	F: *PET*-AAGACAGCCGCCAAATCCTT R: GAATGGTGCGTAGACGGTGA	224-234	7	0.393	0.609	0.609	0.231	KT696134

Characteristics of eight multiplexed microsatellite loci in *Xantusia riversiana reticulata* (*N*=516 individuals). Sequences for 54 additional candidate clones are available through GenBank (KT696132-KT696166). Note abbreviations for number of alleles (N) and observed (H_o), expected (H_e) heterozygosities, polymorphism information content (PIC), and random match probability (RMP). All loci were out of Hardy-Weinberg equilibrium when analyzed island-wide, but not by population (Table S2.3).

Table S 2.2 Year-cohort differentiation

	Xriv B1	Xv GLA	Xriv G2	Xriv G1	Xv CHEL	Xriv Y3	Xriv R1	Xriv R2	Chi_2	P-val
EP	-	0.664	-	0.084	0.910	0.362	0.888	1.000	8.24 (12)	0.766
ES	0.646	0.772	0.646	0.622	0.520	1.000	0.925	0.348	6.786 (16)	0.977
HN	0.201	0.378	0.194	0.001	0.000	-	0.000	0.000	Inf (14)	0.000
HS	0.581	0.151	0.789	0.809	0.163	0582	0.391	0.000	29.83 (16)	0.019
LA	-	0.429	-	0.451	0.339	-	0.000	0.000	Inf (10)	0.000
SC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Inf (16)	0.000
SH	-	0.266	-	0.235	0.850	0.855	0.396	0.612	9.01 (12)	0.702
ST	-	0.262	-	0.001	0.000	1.000	0.024	0.004	49.91 (12)	0.000
WI	-	0.687	1.000	0.393	0.617	-	0.391	0.266	8.11 (12)	0.777

P-values of G tests of genetic differentiation between year-cohorts for populations sampled in multiple years. A dash denotes monomorphic locus-population combinations for which differentiation cannot be assessed. The final two columns show the whole-population Chi_2 statistic, and the whole-population *P*-value.

Table S 2.3 Per locus and population Hardy-Weinberg equilibrium

	Xriv B1	Xv GLA	Xriv G2	Xriv G1	Xv CHEL	Xriv Y3	Xriv R1	Xriv R2
EP	-	0.210	-	0.080	0.381	1.000	1.000	0.381
ES	0.015	0.006	0.017	0.001	0.029	1.000	1.000	0.068
HN	**0.000**	0.004	**0.000**	**0.000**	0.102	-	0.256	0.001
HS	0.005	0.246	0.011	**0.000**	0.006	0.004	**0.000**	**0.000**
LA	-	0.433	-	0.019	0.591	-	0.50932	0.287
SC	-	0.505	-	0.061	0.929	1.000	1.000	1.000
SH	-	0.110	-	0.208	0.911	1.000	1.000	0.001
ST	-	0.118	-	0.701	0.784	1.000	0.952	0.002
TE	-	0.172	-	0.984	0.743	1.000	1.000	0.107
WI	-	0.619	1.000	0.217	0.195	1.000	0.657	0.008
WS	-	0.505	-	0.007	0.203	1.000	0.586	0.130

P-values for tests of deviation from Hardy-Weinberg equilibrium (HWE) per locus (columns) and collection site (rows). A dash denotes monomorphic locus-population combinations for which HWE cannot be assessed. Bold lettering denotes significant deviation after Bonferroni correction (*N*=6).

Table S 2.4 Loci in linkage disequilibrium per site

Site	Locus 1	Locus 2	*P*-value
BO	XvCHEL	XrivR2	0.033
EP	XrivG1	XrivR2	0.029
ES	XvGLA	XrivG1	0.016
ES	XrivB1	XrivG2	0.033
ES	XrivB1	XrivY3	0.034
ES	XrivG2	XrivY3	0.035
HN	**XrivB1**	**XrivG2**	**0.000**
HN	**XvCHEL**	**XrivR2**	**0.000**
HN	**XrivR1**	**XrivR2**	**0.000**
HN	XrivG1	XrivR1	0.002
HN	XvGLA	XrivR1	0.005
HN	XvCHEL	XrivR1	0.010
HN	XrivG1	XrivR2	0.021
HN	XrivG1	vCHEL	0.024
HS	**XrivR1**	**XrivR2**	**0.000**
HS	XvCHEL	XrivR1	0.007
HS	XrivB1	XrivY3	0.007
HS	XrivY3	XrivR2	0.008
HS	XrivB1	XrivG2	0.009
HS	XrivG2	XrivY3	0.009
HS	XrivB1	XrivR2	0.009
HS	XvGLA	XrivG2	0.012
HS	XrivG1	XrivR1	0.014
HS	XvCHEL	XrivR2	0.014
HS	XrivY3	XrivR1	0.015

HS	XrivB1	XrivR1	0.016
HS	XrivG2	XrivR2	0.017
HS	XrivG2	XrivR1	0.031
HS	XvGLA	XrivG1	0.048
LA	**XrivR1**	**XrivR2**	**0.000**
LA	XrivG1	XrivR2	0.021
LA	XrivG1	XvCHEL	0.022
SH	XvGLA	XvCHEL	0.017
SH	XvGLA	XrivY3	0.023
SH	XrivG1	XrivR2	0.037
ST	**XrivR1**	**XrivR2**	**0.000**
ST	XvCHEL	XrivR1	0.035
TE	**XvGLA**	**XvCHEL**	**0.000**
TE	XrivR1	Xriv_2	0.002
WI	XvCHEL	XrivR2	0.024
WS	XrivG1	XrivR2	0.030

Linkage disequilibrium between pairs of loci by collection site. Bold lettering denotes significant linkage after Bonferroni correction (*N*=7)

Table S 2.5 Per locus and population allele counts

	BO	EP	ES	HN	HS	LA	SC	SH	ST	TE	WI	WS	mean	s.d.
XrivB1	1	1	2	2	2	1	1	1	1	1	1	1	*1.25*	*0.43*
XvGLA	14	16	15	18	22	16	8	16	21	23	17	11	*16.42*	*4.15*
XrivG2	1	1	2	2	4	1	1	1	1	1	2	1	*1.50*	*0.87*
XrivG1	7	12	8	10	10	7	3	6	9	8	9	7	*8.00*	*2.20*
XvCHEL	9	9	9	12	12	11	8	13	9	10	10	11	*10.25*	*1.48*
XrivY3	1	2	2	1	2	1	2	3	2	2	2	2	*1.83*	*0.55*
XrivR1	5	5	7	6	10	6	4	6	6	7	7	5	*6.17*	*1.46*
XrivR2	4	4	3	4	6	2	2	6	4	5	4	3	*3.92*	*1.26*
mean	*5.25*	*6.25*	*6*	*6.875*	*8.5*	*5.625*	*3.625*	*6.5*	*6.625*	*7.125*	*6.5*	*5.125*	*6.17*	*1.16*
s.d.	*4.323*	*5.19*	*4.359*	*5.6*	*6.225*	*5.195*	*2.69*	*5.074*	*6.224*	*6.772*	*5.074*	*3.887*	*5.05*	*1.08*
No. gene copies	64	72	62	126	248	76	22	72	74	70	72	74	*86.00*	*53.41*
No. total loci	8	8	8	8	8	8	8	8	8	8	8	8	*8.00*	*0.00*
No. usable loci	8	8	8	8	8	8	8	8	8	8	8	8	*8.00*	*0.00*
No. poly loci	5	6	8	7	8	5	6	6	6	6	7	6	*6.33*	*0.94*

Allele counts and summary statistics for each locus (rows) by collection site (columns).

Table S 2.6 Private alleles (N=12) by collection location

Site	Locus	Allele	Frequency
EP	GLA	350	0.014
ES	XrivB1	132	0.032
HN	XrivB1	125	0.048
HN	XrivB1	114	0.008
HS	GLA	287	0.004
HS	XrivG2	120	0.004
HS	XrivG2	128	0.008
HS	XrivR1	137	0.008
HS	XrivR2	237	0.008
SH	GLA	291	0.014
ST	GLA	346	0.014
TE	GLA	344	0.014

Sites in bold (10 of 12) are all within 5km of each other on the low-lying western side of the island characterized by high quality MSS habitat (see Fig 2.1b).

Table S 2.7 Pairwise numbers of migrants

	BO	EP	ES	HN	HS	LA	SC	SH	ST	TE	WI	WS
BO	0.000											
EP	1.795	0.000										
ES	2.042	**52.829**	0.000									
HN	1.515	3.509	3.025	0.000								
HS	1.337	2.273	2.044	**23.835**	0.000							
LA	1.549	4.137	3.552	**21.358**	**10.739**	0.000						
SC	1.355	4.869	4.143	1.773	1.271	2.124	0.000					
SH	3.315	1.900	2.067	1.398	1.287	1.505	1.438	0.000				
ST	2.337	**7.127**	**7.957**	**5.814**	4.049	**8.695**	4.595	2.113	0.000			
TE	1.617	**7.528**	**6.197**	**6.252**	4.703	**14.883**	3.729	1.820	**26.690**	0.000		
WI	**8.001**	2.311	2.345	2.352	2.258	2.467	1.761	**8.650**	3.432	2.591	0.000	
WS	1.689	4.518	4.361	2.395	1.893	3.646	**9.006**	1.635	**6.094**	**6.739**	2.301	0.000

Number of migrants (N_m) per generation between pairs of collection sites. Migration rates higher than 5 are in bold.

Table S 2.8 Pairwise F_{ST} between all pairs of populations

	BO	EP	ES	HN	HS	LA	SC	SH	ST	TE	WI	WS
BO	0											
EP	0.139	0										
ES	0.1223	0.0570	0									
HN	0.1091	0.0658	0.0047	0								
HS	0.1416	0.0116	0.0665	0.0763	0							
LA	0.1575	0.0228	0.0991	0.1090	0.0104	0						
SC	0.1558	0.1053	0.0488	0.0569	0.1236	0.1643	0					
SH	0.0701	0.1424	0.1163	0.1079	0.1517	0.1627	0.1481	0				
ST	0.0967	0.0280	0.0339	0.0305	0.0412	0.0582	0.0516	0.1058	0			
TE	0.1339	0.0165	0.0321	0.0388	0.0385	0.0505	0.0628	0.1208	0.0093	0		
WI	0.0303	0.0920	0.0976	0.0963	0.0961	0.0997	0.1243	0.0281	0.0679	0.0880	0	
WS	0.1290	0.0641	0.0524	0.0542	0.0945	0.117	0.027	0.132	0.0394	0.0358	0.098	0

Table S 2.9 Pairwise BayesAss measurements between all pairs of populations

	BO	EP	ES	HN	HS	LA	SC	SH	ST	TE	WI	WS
BO	0.8305	0.0148	0.0071	0.0115	0.012	0.0082	0.0075	0.0155	0.0382	0.0094	0.013	0.0322
EP	0.0146	0.8404	0.0068	0.0176	0.0169	0.0073	0.0071	0.0239	0.0247	0.0119	0.0102	0.0186
ES	0.0159	0.1951	0.6747	0.0098	0.0136	0.0096	0.0098	0.0123	0.0262	0.0097	0.0092	0.014
HN	0.022	0.0365	0.0047	0.7141	0.1549	0.0056	0.0046	0.0093	0.0246	0.0061	0.0057	0.0119
HS	0.0036	0.0076	0.0035	0.0125	0.9094	0.0032	0.0033	0.0125	0.0243	0.0049	0.0053	0.0099
LA	0.0241	0.0128	0.0069	0.0149	0.1518	0.6735	0.0067	0.0157	0.0191	0.0075	0.0074	0.0598
SC	0.0192	0.0241	0.0146	0.0148	0.0188	0.0147	0.6813	0.0207	0.1032	0.0158	0.0225	0.0503
SH	0.0125	0.0746	0.0071	0.0221	0.0308	0.0079	0.0071	0.7474	0.0285	0.0108	0.0122	0.039
ST	0.0169	0.0198	0.0069	0.0292	0.064	0.0075	0.0064	0.019	0.7831	0.0102	0.0177	0.0194
TE	0.0087	0.0149	0.0074	0.0168	0.0663	0.0076	0.0068	0.0136	0.0927	0.6749	0.0093	0.081
WI	0.0173	0.0285	0.0068	0.013	0.0387	0.0073	0.0071	0.0221	0.0426	0.0083	0.6771	0.1313
WS	0.0627	0.0118	0.0068	0.009	0.013	0.0076	0.0068	0.0325	0.0191	0.0084	0.0141	0.8081

Chapter 3 Parsing Variance by Marker Type: Testing Biogeographic Hypotheses and Differential Contribution of Historical Processes to Population Structure in a Desert Lizard

3.1 Abstract

A fundamental goal of population genetic studies is to identify historical biogeographic patterns and understand the processes that generate them. However, localized demographic events can skew population genetic inference. Assessing populations with multiple types of genetic markers, each with unique mutation rates and responses to changes in population size, can help to identify potentially confounding population-specific demographic processes. Here, we compared population structure and connectivity inferred from DNA microsatellites and RAD loci among 17 populations of an arid-specialist lizard (*Xantusia vigilis*) in Central California to test among historical processes structuring population-level genetic diversity. We found that both marker types yielded generally concordant insights into genetic population structure including a major phylogenetic break maintained between two populations separated by less than 10 kilometers, suggesting that either marker type could be used to understand generalized demographic patterns across the region for management purposes. However, we also found that populations exhibited varying levels of marker discordance in heterozygosity and allelic privacy that we could use to resolve the effects of differential population histories across the landscape and test among otherwise indistinguishable biogeographic hypotheses. Our results suggest that comparisons of within-population diversity across marker types provide powerful opportunities for leveraging marker discordance for analytical benefit rather than viewing it as a nuisance, particularly for understanding the creation and maintenance of contact zones among clades.

3.2 Introduction

Population-level processes are critical for structuring species-level genetic diversity across both space and time. Especially in the context of complex landscapes, comparative tests across multiple species offer important insight into the differential effects of historical and demographic processes on spatial signatures of genetic diversity (John C. Avise, 2009; Knowles, 2009). From these studies, one of the biggest unresolved questions remains the mechanistic origin and maintenance of sharp phylogeographic breaks across relatively small spatial scales, especially in the absence of corresponding geographic barriers or evidence of other causes of dispersal limitation (B. Charlesworth, Charlesworth, & Barton, 2003; Irwin, 2002). What constitutes a barrier to dispersal, and by what mechanisms do barriers create and maintain these patterns in natural populations living in complex - or simple - habitats? Why do these potential geographic barriers seem to affect different species or even populations within a single species in different ways, even when the underlying biology of those species suggests similar dispersal capabilities and habitat usage (E. A. Myers et al., 2019)?

One powerful way to test among competing processes that generate standing genetic diversity is to compare marker types with different a) mutational properties and b) response to shared demographic history (Fischer et al., 2017; Miller et al., 2014). Theory predicts that two marker types will have different patterns of change in diversity in response to the same demographic events (Fischer et al., 2017; Miller et al., 2014). DNA microsatellites of small tandem repeats tend to overestimate within-population diversity due to their rapid mutation rates, large numbers of alleles, and the tendency for researchers to select highly polymorphic loci (Putman & Carbone, 2014; Queirós et al., 2015). Microsatellites have many more alleles per locus than do sequence polymorphisms (SNPs). As a result, demographic events such as moderate-strength bottlenecks that purge rare alleles should disproportionately impact microsatellites (Garza & Williamson, 2001). RAD (Restriction Site Associated DNA) markers based on SNPs are likely to be in slower-mutating sections of the genome, because finding homologous RAD sites across many individuals and even species relies on their two- to four-base pair long enzyme cut sites

being maintained (Lowry et al., 2017). Since microsatellites have a higher mutation rate than RAD loci, private microsatellite alleles should emerge in an isolated population before private RAD alleles. Accounting for effective population size, we should be able to group timing of population isolation into three separate phases: 1) no private alleles, 2) relative excess of private microsatellite alleles, and 3) private alleles for both marker types. Using these three bins, we can tease apart demographic events that otherwise produce indistinguishable genetic signatures (Fig. 3.1), even when both marker types reveal similar phylogeographic patterns in natural populations (DeFaveri, Viitaniemi, Leder, & Merilä, 2013; Gärke et al., 2012). Overall, leveraging variation in marker response to historical processes remains a powerful, but underutilized, approach to population and evolutionary genetic analyses.

Marker comparisons specifically within geographically complex and biodiverse regions create the strongest tests of both biogeographic hypotheses and the relative contribution of different historical processes to generating spatial patterns of standing diversity. California, with its complex geological structure due to tectonic activity, has long been considered an engine for generating biogeographic diversity, including high endemism, species richness, and complexity of population structure (Gottscho, 2016; Lancaster & Kay, 2013). For Central California, the most important abiotic factors affecting species distribution and population structure are 1) topographic structure and history, particularly the uplift of the north-south Sierra Nevada and Southern Coast Ranges and the tectonic-induced rotation of the east-west Transverse Ranges (see Fig. 3.1) and 2) rainfall gradients across this topography, especially the replicated rain shadow effects along eastern-facing mountain slopes (Hughes, Hall, & Fovell, 2009). The disjunct arid habitats in the Central Californian rain shadows have high conservation importance and represent the northernmost distribution of many Californian desert species (Hill, 2003). These factors all combine to make Central California biodiversity an ideal system for marker comparison studies.

In this study, we used an arid specialist lizard species (the Desert Night Lizard, *Xantusia vigilis*) to a) test among competing biogeographic hypotheses about the historical distribution and connectivity of populations in Central California's xeric ecozones and b) assess contemporary genetic diversity and gene flow to inform population management practices. Then, we compared marker-specific patterns of

diversity and allelic evolution to test among two explanations for an unexpected combination of long-distance gene flow and a deep genetic break across short geographic distance (<10km). Together, these tests help us understand how geography, habitat, and history interact to control barriers to migration among populations.

3.2.1 Biogeographic hypotheses

Xantusia vigilis is a very small (adult mass = 1.5g), secretive lizard commonly found throughout arid regions of the southwestern US and the Baja peninsula of Mexico (Stebbins, 2003). Presumably due to limited dispersal rates and distances and low frequencies of inbreeding (A. Davis et al., 2011), this species maintains genetic signatures of historical processes over long periods of time and boundaries between demes tend to be well-maintained (Sinclair, Bezy, Bolles, Camarillo, et al., 2004). However, this species also shows genetic evidence of long-distance dispersal events, creating unexpected and dramatic patterns of connectivity among non-neighboring populations (Leavitt, Bezy, Crandall, & Sites Jr, 2007). *Xantusia vigilis* is considered a habitat specialist intimately tied to plant or rock cover objects, and several authors have suggested the importance of these specialized habitat associations in predicting historical distribution and resolving unexpected patterns in population connectivity (Noonan et al., 2013; Sinclair, Bezy, Bolles, Camarillo, et al., 2004). These factors have also contributed to the presence of highly fragmented and disjunct populations across the northern range of *X. vigilis*, with the most extreme example being two isolated populations of *X. vigilis* about 150 miles northwest of the main range in the Joshua trees (*Yucca brevifolia*) of the Mojave Desert: one population is found in isolated outcrops of Chapparal Yucca (*Hesperoyucca whipplei*) in the Panoche Hills and another in Gray Pine (*Pinus sabiniana*) within Pinnacles National Monument (Fig. 3.2).

California has the highest diversity in the world of unique, and deeply divergent, lineages of *Xantusia*. Although the membership within several distinct clades of *X. vigilis sensu stricto* (also referred to as “Clade A” *X. vigilis* in (Sinclair, Bezy, Bolles, Camarillo, et al., 2004)) across Central California has been well supported with phylogenetic work using both mitochondrial and nuclear loci (Leavitt et al.,

2007; Noonan et al., 2013), these previous studies have not resolved relationships *among* these clades, suggesting rapid range expansion and diversification. These relationships inform an important outstanding question in the historical biogeography of the system: the directionality and timing of range expansion along Central Californian dispersal corridors. The two competing hypotheses about expansion from ancestral populations generally fall into the categories of "North-to-South" or "South-to-North," and they have differing implications for both the drivers of expansion and conservation importance of the disjunct northern populations (Morafka & Banta, 1973). In the North-to-South scenario, the populations along the Coast Range derive from an ancestral population near the northern range limit in the Pinnacles or Panoche area, with little to no input from the populations in the Mojave Desert to the southwest. In this case, the main direction of population expansion and gene flow is from the northern populations to the southern populations, and the Pinnacles/Panoche populations would have important conservation value as repositories of high diversity ancestral genetic variation. In the South-to-North scenario, the ancestral population centered in the main species range of the Mojave Desert, and the northernmost Pinnacles/Panoche populations are simply the most recent outpost of post-glacial range expansion along newly-created habitat. Both North-to-South and South-to-North scenarios entail demographic processes of expansion, isolation, and population size changes that should be reflected in the contemporary genetic diversity and distribution alleles across the landscape.

3.2.2 Predictions of marker variance

Heterozygosity of the two marker types can depart from co-linearity in several specific scenarios (Fig. 2.1a). Due to their relatively quicker mutation rate, microsatellites should have higher allelic richness when populations have gone through an acute reduction in size, but have since rebounded demographically (Hoelzel, 1999; Martínez-Cruz, Godoy, & Negro, 2004). Alternatively, the population might have recently received migrants. Since microsatellites mutate more quickly than sequence data, these new migrants are more likely to introduce new microsatellite alleles than SNP alleles. RAD data should have higher relative heterozygosity in very recently established populations. Rare alleles are

disproportionately likely to drop out during founder events (Garza & Williamson, 2001). Since microsatellites can have more alleles per locus than sequence data, founder events could reduce microsatellite heterozygosity more sharply than RAD heterozygosity. Low variance in both marker types can be produced by prolonged isolation at low population size, or by a recent and severe founder event.

The relative abundance of private alleles in the populations can distinguish between the scenarios posed above, and provide information on the length of time during which a population has been isolated (Fig. 2.1b). The longer a population has been isolated, the more likely it is to have private alleles (Harpak, Bhaskar, & Pritchard, 2016). However, due to the difference in mutation rates between the marker types, populations should go through four distinct phases following a demographic event that reduces population size. First, neither marker type will have private alleles. In the second phase, new microsatellite alleles will emerge, but new RAD alleles will still be absent. In the third phase, both marker types will show many private alleles. In a theoretically possible fourth phase, the rapid mutation rate of microsatellites will lead to plesiomorphic identical alleles in other populations, rendering former private microsatellite alleles no longer detectable as private. The combination of two marker types will allow us to identify and date demographic events to a greater degree of precision than either of our markers independently.

Using a combination of phylogeographic and demographic analyses, we reconstructed historical patterns of population structure and connectivity in *Xantusia vigilis* across a complex geological and ecological landscape. In doing so, we leveraged the two marker types to differentiate historical migration patterns across a range of demographic scenarios. By assessing historical biogeographic drivers structuring genetic marker discordance across populations, we provided a clear mechanism for testing among otherwise indistinguishable hypotheses that is useful for other systems with similarly intractable population histories.

3.3 Methods

3.3.1 Field collection and tissue acquisition

We collected tissue samples in the field from 354 *Xantusia vigilis* between 2007 and 2014, which we augmented with five museum samples mainly from the Northeastern Mojave Desert (Table S3.1). We sampled across 17 populations in Central California, loosely clustered into four regions (Pinnacles National Monument, Panoche Hills, northwestern Transverse Ranges, and the Antelope Valley) along a north-south latitudinal gradient in the Central Coast range and Western Mojave (Fig. 3.2). For most sites, we captured lizards by lifting, rolling, or opening decaying yucca trunks and rosettes (see also A. Davis et al., 2011). At the Pinnacles site, we captured lizards by first moving fallen *Pinus* logs onto a white sheet and prying off flakes of bark by hand, as well as flipping associated flakes of talus underneath the logs. After capture, we took tail clip samples that were stored in 95% ethanol and frozen until analysis. For outgroup comparisons, we included tissues from *Xantusia wigginsi.*, *X. extorris*, *X. gilberti*, *X. sherbrookei*, and *X. riversiana*, representing five additional species in the final analyses.

3.3.2 Habitat classifications

We collected samples from three major habitat types (Fig. 3.2). North of the Transverse Ranges, *Hesperoyucca whipplei* is distributed into discrete patches on sandstone formations, where it is the dominant forb, therefore all of the collection sites were straightforwardly characterized as *Hesperoyucca* mesohabitat. In the disjunct population at Pinnacles National Monument, the habitat has been defined as grey pine (*Pinus sabiniana*)-blue oak (*Quercus douglasii*) woodland (Sawyer & Keeler-Wolf, 2009) and lizards also occurred in grey pine-manzanita (*Arctostaphylos sp.*) associations at higher elevations. In the Transverse Range collection sites, lizards were also collected in *H. whipplei*, but here the yucca tended to be interspersed with woody shrubs such as *Ephedra californica* and *Artemisia tridentata* as well as at least two tree species, *Juniperus californica* and *Pinus monophylla*. Our Mojave sites were dominated by

creosote bush, *Laria tridentata,* and lizards were found mostly under Joshua trees (*Yucca brevifolia)*, Mojave yucca (*Y. schidigera*), and Banana yucca (*Y. baccata*).

3.3.3 DNA extraction and Microsatellite Genotyping

We extracted DNA from tissue samples using a standard Qiagen Blood and Tissue spin column kit or a Chelex-based protocol and amplified 8 microsatellite loci using a Qiagen Multiplex PCR kit under standard amplification conditions (see A. Davis et al., 2011 for primer information). We visualized fluorescently-labelled products on ABI 3170XL machines at the University of California, Berkeley and on ABI 3730XL machines at University of Michigan, and we scored genotypes in GeneMapper v4.0 using allele panels created from previous analysis of 1,140 *X. vigilis* from the Antelope Valley population (A. Davis et al., 2011; A. R. Davis, 2012; Davis Rabosky, Corl, Liwanag, Surget-Groba, & Sinervo, 2012) . We discarded any individual from further analysis that we could not confidently genotype at five or more loci. We checked our microsatellite data for null alleles using the R package 'PopGenReport' (Adamack & Gruber, 2014) and Hardy-Weinberg equilibrium using the R package 'pegas' (Paradis 2010). One locus was dropped from further analysis due to a high frequency of null alleles.

3.3.4 Next-Generation Sequencing and Data Processing

We performed double digest Restriction site Associated DNA (ddRAD) sequencing on a subset of individuals ($N = 104$ *X. vigilis*, plus 10 outgroup samples) following the protocol developed by (Peterson et al., 2012). We restricted total genomic DNA using the enzymes EcoR1 and Msp1 and then used a QIAquick gel extraction kit to size select fragments between 100 and 200 base pairs. We used 24 unique barcodes and four unique indices (following Peterson et al., 2012) to individually mark genomic DNA from 96 individuals per multiplexed lane. We sequenced individuals across three runs on an Illumina HiSeq 2500 at the University of Michigan Sequencing Core with 200 base pair paired-end reads.

We processed our RAD fragments using a pipeline developed by (Singhal et al., 2017). We removed low-quality and short reads using TRIMMOMATIC v0.35 (Bolger, Lohse, & Usadel, 2014), then assembled reads within each individual using RAINBOW 2.0.4 (Chong, Ruan, & Wu, 2012). We

identified homologous loci by clustering at 97% similarity in VSEARCH v.1.1.0 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016), and assembled a pseudogenome from the consensus sequences for each homologous fragment. We aligned loci for each individual using BWA 0.7.12 (Heng Li & Durbin, 2010) and GATK (McKenna et al., 2010). We called SNPs using SAMTOOLS 1.2 (H. Li et al., 2009). We retained fragments with coverage depths of 10x or higher.

We identified SNPs and indels in each RAD fragment. For downstream analyses, we focused on SNPs only. To do so, we used a custom R script to insert the appropriate SNP call for each individual into the consensus sequence. We screened all fragments for deviations from Hardy-Weinberg equilibrium in each population using the function 'hw.test' in the R package 'pegas' (Paradis, 2010). We removed all fragments that had one or more populations out of Hardy-Weinberg equilibrium at a significance value of 0.05. Following screening, retained 1706 fragments, with 9260 SNPs. We removed individuals that were had more than 90% missing data. After removing these individuals, we retained 82 *X. vigilis* individuals and nine outgroup individuals.

3.3.5 Locus diversity and characteristics

We calculated within-population and within-region heterozygosity using the R package diveRsity for microsatellites (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013). We used a custom R script to calculate the proportion of private alleles per locus and individual for each population and region. For RAD data, we used a custom R script that calculated the average proportion of heterozygote calls per individual for the variable locations on each DNA fragment within populations and regions. To find private alleles, we considered each fragment as a whole, and used a custom R script to determine whether the allelic state across all SNPs within a fragment was unique to a given population or region.

3.3.6 Phylogeography

We used the program RAxML to infer a tree from our RAD samples (Stamatakis, 2014) and reconstruct the phylogenetic relationships and major genetic splits within *Xantusia vigilis*. We used the R

package ‘phylotools’ to create a supermatrix and a gene partition file from the RAD fragments (J. Zhang, 2017). We used GTRCAT model with 100 bootstraps, retaining one *Xantusia wigginsi* sample designated as an outgroup for display. To assess the relationship between phylogeny and geography, we mapped each sample from its collection point to its location on the phylogenetic tree using phytools (Revell, 2012).

We used Structure to identify population-level deme groupings and find evidence of admixture in our RAD and microsatellite data (Pritchard, Stephens, & Donnelly, 2000). For the RAD Structure input, we selected one SNP per fragment. Some of our fragments varied only in the outgroups, so our Structure file contained 1529 SNPs across 82 *X. vigilis* individuals. We ran each marker type at K values of 3 and 4. For microsatellites, we performed 1,000,000 steps and 1,000,000 burnin steps. For the RAD data, we did 100,000 steps with 100,000 burnin steps. We performed 10 runs at each K value for each marker types, and selected the output file with the highest posterior probability.

For comparison with our Structure results, we mapped the size range between our largest and smallest microsatellite alleles in each population. This visual analysis allowed us to understand the underlying allelic characteristics driving the Structure groupings. We also found the average pairwise number of dissimilarities between each pair of individuals in our RAD dataset and used these values to calculate the average pairwise dissimilarity within and between populations. This analysis allowed us to contextualize the Structure results in relation to the underlying variance characteristics of the data.

3.3.7 Historical movement corridors

We used the program Estimated Effective Migration Surfaces (eems) to explore the relative rates of effective migration between our focal populations in both our microsatellite and RAD datasets (Petkova, Novembre, & Stephens, 2015). We created a convex hull around our collection location, and laid a network of 1000 points across the polygon it formed. Eems used this grid to find effective migration rates by adjusting the expected rates of migration between pairs of populations until the

posterior genetic dissimilarities match the observed dissimilarities. Dissimilarities were calculated using Euclidean genetic distances between pairs of individuals.

Finally, we assessed direction of migration between the major population clusters found in the RAxML tree. In this analysis, we followed the philosophy of rangeExpansion package (Peter & Slatkin, 2013) which relies on the observation that one-time, directional population movements often increase the proportion of rare alleles in the population (Hallatschek, Hersen, Ramanathan, & Nelson, 2007; Klopfstein, Currat, & Excoffier, 2006). Since rare alleles are more likely to be derived than ancestral, the expected outcome of a range expansion event is to increase the proportion of derived alleles in the recipient population (Slatkin & Excoffier, 2012). However, since some of our populations have been separated for many generations, they are often fixed at many loci. To best use the available data, we created a simple pairwise 'directionality' metric between populations. For each RAD locus, we found the putatively ancestral allelic state using a designated outgroup animal, our *Xantusia wigginsi* sample. If an allele from the *X. vigilis* sample matched the outgroup allele, we considered that allele to be ancestral. We then found all derived alleles from the *X. vigilis* individuals in the sample. In cases in which one population had only derived alleles while a second population contained both ancestral and derived alleles, we considered ancestral+derived allele population to be the source population. For each pair of populations, we selected three individuals from per population. If a fragment was not represented in at least three individuals in both populations, as well as our outgroup individual, we removed it from the analysis of that pair of populations. From this rarefied dataset, we counted the number of fragments that pointed to each population in the pair being the 'source' relative to the other population. To compare across population pairs, we divided the number of fragments that showed directionality by the total number of fragments that had more than one allele in the focal pair of populations. We followed the groupings of populations in our RAxML tree, first finding directionality between the major north and south geographical areas, then between the regional demes nested within each area.

3.4 Results

3.4.1 Locus characteristics, diversity, and private alleles

The observed frequency of null alleles in the microsatellite loci retained for analysis ranged from -0.0391 to -0.0033 per locus. No locus was significantly out of Hardy-Weinberg equilibrium in more than two populations. We found per-population heterozygosity levels ranging from 0.071 in Quatal Canyon to 0.817 in Antelope Valley in microsatellites. Private allele frequency per population range from 0 for several populations to 0.444 in the Pinnacles population, while private allele frequency per region ranged from 0.067 in the North Transverse region to 0.443 in the Mojave region per locus and individual in microsatellites (Table 3.1). We found that many microsatellite loci were region-specific rather than population-specific, so we concentrate on patterns of regional privacy going forward.

For RAD data, the per-population mean within-individual frequency heterozygote calls per SNP ranged from 0.006 at Curry Mountain to 0.088 at Antelope Valley. We found that the regional per-fragment frequency of private alleles ranged from 0.373 in the South Transverse region to 0.721 in the Mojave region. At the population level, the frequency of private alleles ranged from 0.056 in Dry Canyon to 0.617 in Antelope Valley (see Table 3.1 for full results).

3.4.2 Phylogeography

We recovered two major clades of *Xantusia vigilis* across Central California: one in the Mojave, and another in the Central Valley area (Fig. 3.2). Each geographic cluster forms a monophyletic clade with high bootstrap support, while within-clade genetic variation and phylogenetic structure are minimal. The west Mojave *X. vigilis* are nested within the east Mojave samples. Also nested in that clade, sister to the west Mojave samples, are individuals from three populations in the southern Transverse Ranges: Quatal Canyon, Dry Canyon, and Apache Canyon. The Central Valley clade contains individuals from the Transverse Ranges from Ballinger Canyon and Cuyama Valley. Slightly farther north are the Caliente

Ridge samples, which also cluster in this clade. The northern Central Valley samples are split between local Panoche and Pinnacles clades, with Curry Mountain clustering with Pinnacles.

The Structure results largely recapitulate the RAxML tree (Fig. 3.3a). At K=3, RAD results cluster Panoche and Pinnacles, with the southern Transverse Range and Mojave populations forming their own deme. At K=4, the Pinnacles populations split from Panoche in the RAD data. At both the K=3 and K=4, the two populations closest to the North Transverse/South Transverse split show a small but consistent amount of introgression from the opposite deme. The microsatellite results split Panoche and Pinnacles, with the Transverse Ranges North populations clustering with Pinnacles. The microsatellite data split some Ciervo samples from the rest of the Panoche populations at K=4. Overall, microsatellites show much greater admixture than the RAD data.

Microsatellite allele ranges show that both the Northern and Southern Transverse Ranges have both become fixed for a single allele at some loci, but those loci differ between the two demes (Fig. 3.3b). For most loci, the Mojave and Panoche populations have the largest size ranges, with other populations intermediate between them. The RAD pairwise difference heat map shows lower pairwise distances between populations within a deme than those between, demonstrating that the Structure demes reflect raw genetic dissimilarities in the data (Fig. 3.3c). Together, the RAxML, Structure, and locus characteristic analyses point to a strong biogeographic break between northern and southern groups that occurs in the Cuyama Valley in the Transverse Ranges.

3.4.3 Historical movement corridors

Our microsatellite eems runs show high effective migration rates within the Panoche populations and around the Mojave populations. The Pinnacles, Curry Mountain, and southern Transverse Ranges populations show relatively lower effective migration rates. The RAD dataset shows broadly similar outcomes, but with an indication of higher effective migration rates between the Pinnacles and Curry Mountain populations. This analysis confirms that our two locus types recover similar historical

migration corridors, and provide support for the phylogeographic and population genetic results that separate the North and South Transverse populations (Fig. 3.4a).

Our RAD-allele based directionality analysis shows that the southern populations (Mojave and South Transverse) are a source of migrant alleles to the northern populations (Panoche, Pinnacles, and North Transverse). Within the southern populations, the Mojave is a source for the South Transverse populations. In the northern populations, Panoche is a source for Pinnacles and the North Transverse populations (Fig. 3.4b). The number of polymorphic fragments between pairs of populations ranged from 796 (Mojave/South Transverse pair) to 590 (North Transverse/Pinnacles). DNA fragments with clear directionality ranged from 138 showing the Mojave as ancestral to South Transverse to 25 placing Pinnacles ancestral to Panoche.

3.4.4 Locus comparison

Our comparison between RAD and microsatellite heterozygosity showed populations occupying each of our four quadrants (Fig. 3.5a), which we set using the mean values for heterozygosity for both marker types. Antelope Valley had the highest heterozygosity for both markers, while the South Transverse populations had low values for both markers. Pinnacles and North Transverse populations showed variable levels of deviation from our regression line. Possibly reflecting the admixture found in our Structure analysis (Fig. 3.3a), the Quatal Canyon (South Transverse group) and Ballinger Canyon (North Transverse group) populations had higher RAD heterozygosity levels than the other populations in their respective regional demes. The Panoche populations all clustered in the lower right quadrant, which we previously designated as indicating post-bottleneck rebound. No other regional deme showed a consistent pattern of clustering within a single quadrant.

Private allele discordance indicates that the Panoche and South Transverse populations have experienced a recent bottleneck, while private allele patterns show the Pinnacles population is recovering from an older bottleneck (Table 3.1, Fig. 3.5b). The Antelope Valley populations have been large and

stable, while the North Transverse has experienced demographic events that inflated RAD private allele rate over microsatellite private allele rate.

3.5 Discussion

In this paper, we leveraged patterns of variation between two types genetic markers across populations of the desert night lizard, *Xantusia vigilis,* to test among biogeographic hypotheses that could not be resolved using conventional approaches. Within the complex geological and climatic history of California, we paradoxically found that 1) the biggest phylogeographic break appears without any clear geographic barriers to gene flow, despite that 2) the species showed relatively large amounts of movement along corridors that do appear to have substantial barriers. By testing the historical drivers structuring each type of genetic marker discordance found across populations in both heterozygosity and allele privacy, our methods for leveraging genetic marker type discordance to our analytical advantage can be applied across other systems with similarly intractable population histories.

Despite the inherent biases associated with both of our marker types (Arnold, Corbett-Detig, Hartl, & Bomblies, 2013; Putman & Carbone, 2014), they recover highly concordant results for both phylogeographic structure (Fig. 3.2-3.3) and within-population diversity (Fig. 3.5). Both of our paired analyses (Fig. 3.3-3.4) generally agree on population groupings and on the strong differentiation between the north and south Transverse Range populations. We also found several areas of disagreement between marker types, for which the nature, magnitude, and directionality of discordance were informative for reconstructing both population histories and understanding the creation and maintenance of contact zones among clades. Our study area contains many complex biogeographic patterns, which have been influenced by geological and climatic history. To demonstrate how our analytic approaches helped to discriminate among demographic hypotheses and how these methods could be applied to organisms with similar ecologies or geological histories, we discuss four inferences below: one wholistic bioegeographic perspective and three regional ‘case studies’ within our broader study area.

3.5.1 Tests of biogeographic hypotheses: Expansions and boundaries

One of the unresolved biogeographic questions we tested is whether the northern range-limit populations of *Xantusia vigilis* represent a historical refugium for the species (North-to-South hypothesis), or whether they are a recent offshoot of the main, Mojave Desert population (South-to-North hypothesis; (Morafka & Banta, 1973)). We found support for the South-to-North hypothesis, although this event most likely occurred many generations in the past. Our phylogeographic tree showed reciprocal monophyly between the two regions, rather than showing one nested in the other (Fig. 3.2). To determine directionality, we instead relied on the evolutionary history of individual RAD fragments (Fig. 3.4b). Our results indicate that the Mojave populations are the source of the northern populations, but that the expansion event happened so long ago that the Central Valley *X. vigilis* represent a valuable and unique genetic resource within the broader species. This result is consistent with previous mitochondrial trees that showed a quick expansion of the A-clade *X. vigilis* approximately 1.5 mya (Leavitt et al., 2007).

Our study also resolved the geographic location a major phylogenetic break in the desert night lizard, *Xantusia vigilis,* to a single small valley in California's Transverse Ranges (Fig. 3.2). *Xantusia vigilis* has a 'Mojave' clade that extends into the southern Transverse ranges, and a set of 'northern' clades in the mountains west of California's Central Valley that meet in the Cuyama Valley in the Transverse Ranges (Fig. 3.2, Fig. 3.3). In this broad pattern, *X. vigilis* are similar to many other California species or species groups that show biogeographic breaks in the Transverse Ranges (Chatzimanolis & Caterino, 2007; Gottscho, 2016). However, the break we detect falls in between the edges of Transverse Range-specific biogeographic units detected by comparative phylogeography (Chatzimanolis & Caterino, 2007). The phylogeographic patterns used to identify the Transverse Range biogeographic regions may be old enough that they were formed during the Transverse Range uplift, which occurred between five and three million years ago (Nicholson, Sorlien, Atwater, Crowell, & Luyendyk, 1994). In contrast, *X. vigilis* likely entered the area 1.5 million years ago, dispersing over the Transverse Ranges rather than being divided by their uplift (Leavitt et al., 2007). These populations show a secondary subdivision between the

northern Transverse range populations and the Panoche-Pinnacles clade, which could correspond to the glacial-lake barrier shown in other Central Valley lizards (Richmond et al., 2017).

The precise paleoclimatic events that facilitated expansion by the A-clade *X. vigilis* are likely to remain elusive. However, looking at modern vegetation could provide some clues. Both marker types show lower genomic diversity in populations that shelter under *Hesperoyucca whipplei* relative to those that use other sheltering habitat (Fig. 3.2, Table 3.1, Fig. 3.5). We hypothesize that this pattern reflects the isolation and small size of *Hesperoyucca* patches in our study area. *Yucca* stands in the Mojave and pine stands in the Pinnacles area seem to provide better opportunities for dispersal and gene flow. From this evidence, we might hypothesize that a region-wide vegetation more similar to either modern Mojave or Pinnacles could have facilitated the rapid expansion of *X. vigilis* into their current range. In addition, our observed correlations between habitat type and genomic diversity indicate that habitat destruction in *Hesperoyucca* areas will have a proportionally higher regional genomic diversity consequence for *X. vigilis* that a similar amount of habitat destruction in other regions.

3.5.2 Transverse Ranges populations: A phylogeographic museum

The Cuyama Valley, a small area in the northern Transverse Ranges, holds populations from the two major phylogeographic lineages of California A-clade *X. vigilis* (Fig. 3.2 and 3.3). Our directionality analysis shows a strong signal of the Mojave population being a source of migrants for the South Transverse population, and Panoche being a source of migrants for the North Transverse populations (Fig. 3.4b). However, both the South and North Transverse populations are monophyletic rather than nested within their source populations, indicating that they have been in place for many generations. Our Structure results do show a small amount of introgression between the demes in the two most geographically adjacent populations, Quatal and Ballinger (Fig. 3.3a). All individuals in both populations show a small amount of assignment to their opposite deme, indicating that the admixture event is not recent. Given their apparent long residency and geographic proximity, why do we observe so little admixture between these two regional demes?

Given the available evidence, we hypothesize that at least one of the major regional lineages arrived in the Cuyama Valley just prior to some paleoclimatic event that dramatically reduced migration rates to their low modern levels. Given our observed differences in genetic diversity between cover-vegetation types, a reasonable scenario could involve a rapid drying event that replaced contiguous tree cover with more xeric vegetation, with fragmented stands of appropriate cover plants interspersed in an inhospitable matrix. This scenario would fit with patterns of quick expansion followed by long-term stasis from other locations within our study areas, and from other work on the species (Leavitt et al., 2007).

Today, Cuyama Valley *X. vigilis* populations are isolated from conspecific populations. Paleoclimatic reconstruction shows that Cuyama had a relatively wet climate during the dry periods of the last glacial maximum and the Younger Dryas (DeLong, Minor, & Arnold, 2007). As such, it may function as a phylogeographic museum, preserving historical patterns of reticulate population identity formerly common to the broader area, rather than an example of conditions that create absolute barriers to expansion in *X. vigilis*. Worldwide, many other areas served as climatically-stable refugia during the last glacial maximum, including areas of the Amazon (Bonaccorso, Koch, & Peterson, 2006) , the Eastern Afromontane Biodiversity Hotspot (Demos, Kerbis Peterhans, Agwanda, & Hickerson, 2014), and southern Australia (Byrne, 2008). These locations may also preserve a disproportionate amount of phylogeographic diversity, particularly if conditions have since changed to reduce migration rates of the species concerned. If this mechanism is widespread, high lineage diversity of a variety of low-dispersal organisms, such snails, plants that spread mostly through vegetative mechanisms, or tropical-forest understory specialist birds, might be preserved in historical stable habitat patches. As in the case of the Cuyama Valley *X. vigilis*, these phylogeographic hot spots may not be readily distinguishable from surrounding habitat in the modern day.

3.5.3 Panoche Hills: Recovery from old bottleneck results in microsatellite - RAD data conflict

Movement of the San Andreas fault near the Central Valley may help explain another unexpected pattern of deme affiliation. Our Panoche and Curry Mountain samples are on the eastern (stationary) side of the fault, while the Pinnacles population is on the western (moving) side. In the time lag since the putative rapid expansion of *X. vigilis* 1.5 mya, the fault has moved approximately 70 km (Argus & Gordon, 2001, Greg Middleton pers. comm.). The Pinnacles populations group with Panoche in the RAD marker Structure results, but with North Transverse in the microsatellite results (Fig. 3.3a). Our directionality analysis shows Panoche as a source of migrants to both Pinnacles and North Transverse populations (Fig. 3.4b). Prior to the relocation of Pinnacles due to San Andreas movement, Pinnacles and North Transverse were closer together, and both were south of Panoche. A secondary southward migration, first to Pinnacles and then to the North Transverse area, could account for our observed RAD data patterns. A further discordance between marker types occured within the Panoche population. The microsatellite Structure results detect a fourth deme in a subset of the Ciervo Hills samples, which is not reflected in the RAD data (Fig. 3.3a).

Insights about the demographic histories of our populations from discordance in allelic diversity between our marker types can resolve these observations. The Panoche populations, clustered in the lower right quadrant in Fig. 3.5a, all show a signature of an old bottleneck followed by a rebound (Fig. 3.1, 3.5). The rebound after such a bottleneck could account for the local demes within the Panoche populations in the microsatellite Structure results (Fig. 3.3a), which could have emerged due to changes in relative allele frequency during the rebound process. Evidence from geology and marker comparisons supports the following scenario: at some time in the past, migrants from Panoche moved south, first to the former location of the Pinnacles populations, then to the North Transverse areas. Subsequently, the Panoche region experienced a bottleneck and population rebound, which altered the microsatellite allelic profiles of the Panoche populations, but did not strongly impact their RAD data profile. Under this scenario,

Pinnacles and North Transverse share a regional ancestral microsatellite allelic signature. This interpretation is strengthened by the observation that the microsatellite signatures in Panoche proper show significant admixture with the Pinnacles/North Transverse deme, while the lower-diversity Griswold, Ciervo, and Tumey Hills populations carry the signatures of the location-specific demes (Fig. 3.2, Fig. 3.3a).

Similar mechanisms likely act worldwide, particularly in areas that experienced suboptimal climate conditions in recent paleoclimatic history. Panoche's location in a rain shadow may be particularly relevant here. Such rain shadow habitats might be both optimal for survival of dry-adapted organisms during cold and wet climatic conditions, but vulnerable to bottlenecking due to drought when global conditions change. Rain shadow-driven arid areas of global biological importance include but are not limited to the Atacama Desert on the western coast of South America (Rech et al., 2010), the Eastern Arc mountains in Tanzania and Kenya (Burgess et al., 2007; Lovett, 1996), and the Central Asian high plateau north of the Himalayas (Tewari & Kapoor, 2013).

3.5.4 Curry Mountain: Tectonic drift separates populations that retain strong co-ancestry

Another seemingly paradoxical result we found that may be explained by San Andreas fault movement was the relationship between the samples taken from the main areas of Pinnacles and Panoche versus Curry Mountain, which is approximately 70 km south of both of the larger populations (Fig. 3.2). The Curry Mountain samples cluster with the Pinnacles lizards in both marker types, despite Panoche being geographically closer and more similar in habitat. However, at the time of the expansion of the A-clade *X. vigilis*, Pinnacles would have been geographically closer to Curry Mountain, explaining the ongoing genetic similarity between the populations. This type of strike-slip faulting displacement occurs elsewhere throughout the world. The rotating Pacific plate, which drives the movement of the San Andreas fault, also causes faulting throughout the Pacific rim. Locations along the North American west

coast (Brothers et al., 2020), in Japan (Hosoi et al., 2019), and in New Zealand (Michailos, Warren-Smith, Savage, & Townend, 2020) could experience similar displacement.

3.6 Conclusions

Our phylogeographic results show that *Xantusia vigilis* has the ability to maintain biogeographic breaks between two closely adjacent demes over large timescales. In an apparently contradictory pattern, they also show close population co-ancestry over large geographic distances. When combined with previous work on this species, we hypothesize that the biogeographic history of *X. vigilis* is largely composed of long periods of population stasis, with little dispersal of any kind. On rare but significant occasions, there has been substantial long-distance gene flow in which individuals successfully established new populations and introgressed with existing populations. For similar scenarios of punctuated dispersal regimes across long time periods, our work demonstrates the utility of comparing phylogeographic signal in marker types with different mutational properties to successfully resolve complex histories of migration and demographic change. We propose that our approach is applicable to organisms in similarly tectonically active and paleoclimatically complex habitats worldwide.

3.7 Data Archiving

Raw fastq files are available in the University of Michigan's Deep Blue archive, at https://doi.org/10.7302/mep1-5124.

3.8 Acknowledgments

We thank California Department of Fish and Game, the U.S. Fish and Wildlife Service, and Pinnacles National Monument for scientific collection permits. We thank many people for assistance in field collection of samples: Ammon Corl, Dean Leavitt, Heather Mostman Liwanag, William Mautz, Theodore J. Papenfuss, Amy Patten, and Barry Sinervo. We thank the Natural History Museum of Los Angeles County, Monte L. Bean Museum at Brigham Young University, and the Museum of Vertebrate Zoology

at the University of California, Berkeley for loaning tissue samples. We also thank Paul Johnson for helpful comments in the preparation of this manuscript.

3.9 Funding

This study was supported by startup funds from the University of Michigan to ARDR and from the U.S. Bureau of Land Management to MFW and ARDR.

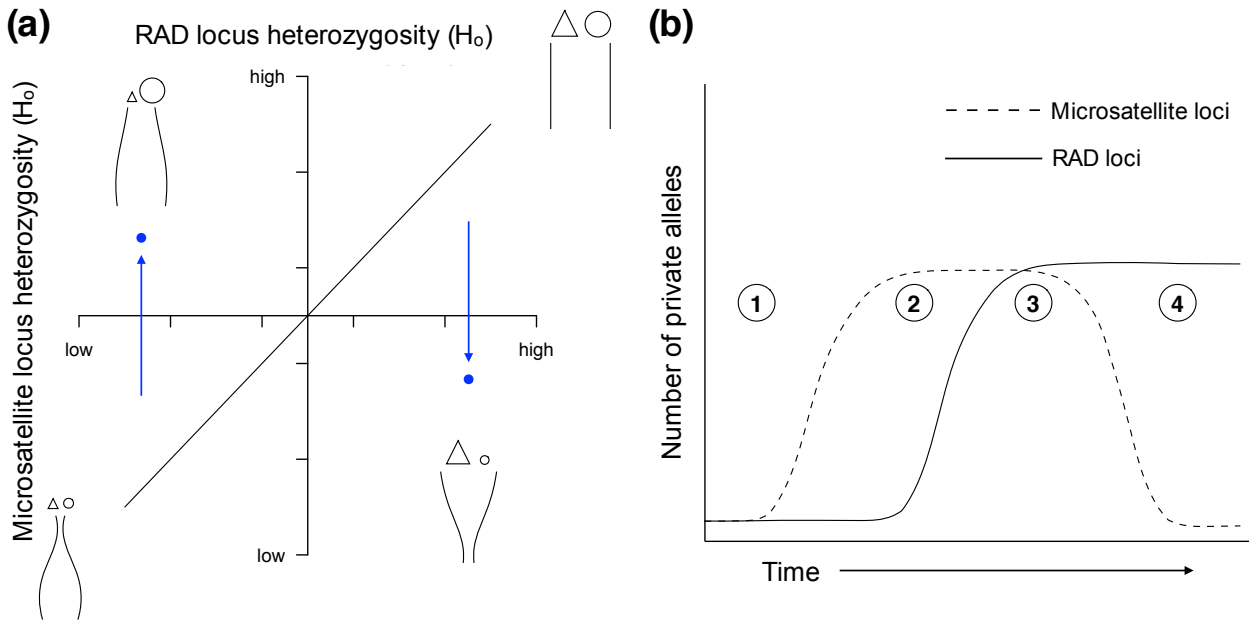


Figure 3.1 Genetic marker variation supporting different historical scenarios

(a) The concordance (or lack thereof) between values of observed heterozygosity (H_o) across RAD and microsatellite loci reflects different population histories (population size through time shown by curved lines in each quadrant, with time increasing towards the top). Most population histories will yield generally concordant levels of heterozygosity (blue points) across both marker types (both high or both low, as shown by the relative sizes of triangles for RAD loci and circles for microsatellites). The slope of the regression line does not need to be 1, but it is predicted to be positive. Discordant heterozygosity values between the two marker types in which one marker is much higher than the other (lower right and upper left quadrants) suggest timing and strength of historical bottlenecks or founder events that will differentially affect these marker types. (b) The predictions from heterozygosity can be integrated with numbers and identity of private alleles across marker types to create a framework for testing among competing biogeographic hypotheses across populations (numbers in circles). Variation in correspondence between these values over time is due to the differential mutational and saturation properties of the two marker types. Recently isolated populations should have few private alleles at either marker type (1). Higher microsatellite mutation rates will generate microsatellite private alleles before

RAD private alleles (2). Eventually RAD alleles will occur (3), and finally random mutations in other populations could result in plesiomorphy, rendering microsatellite private alleles undetectable.

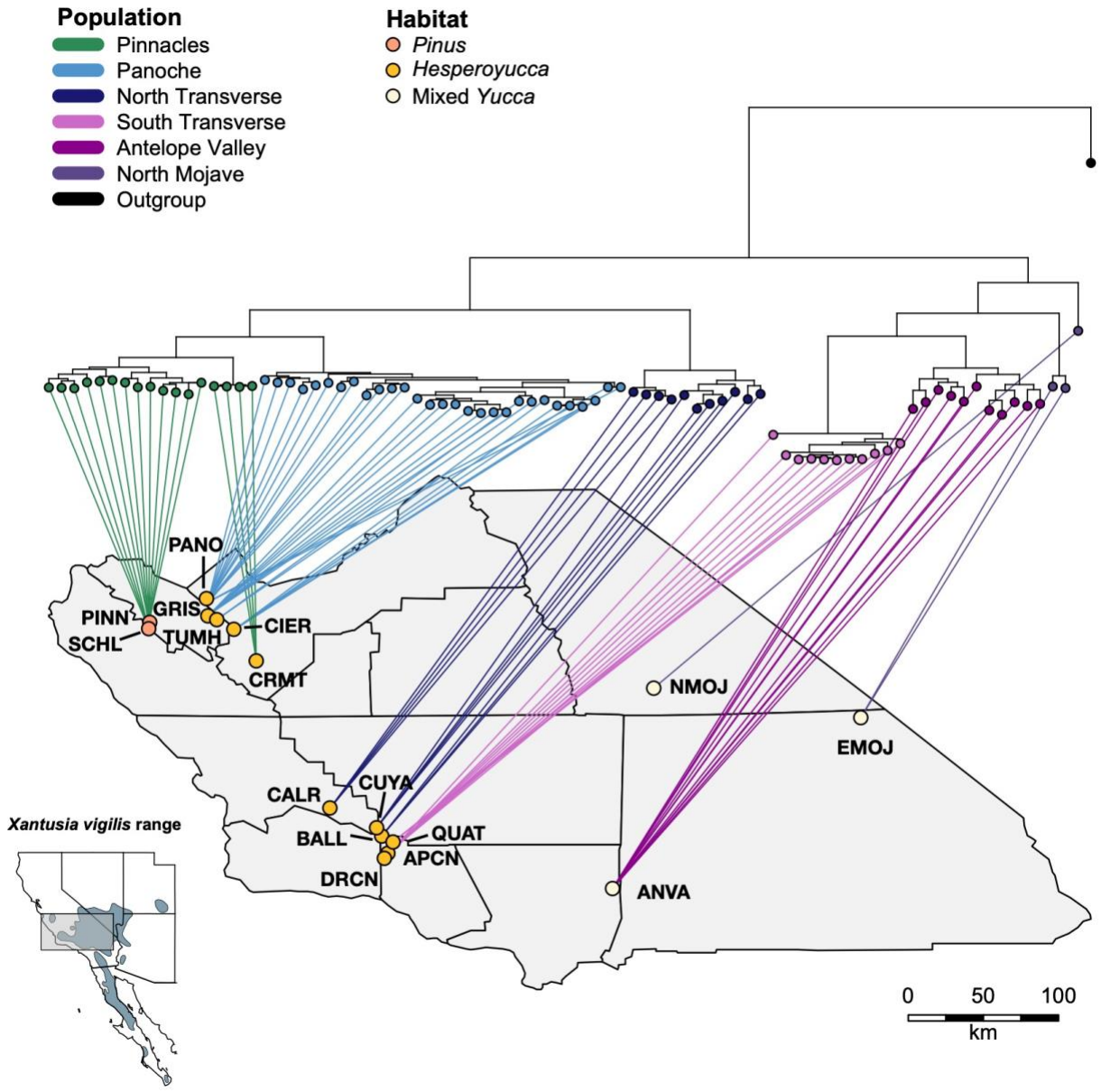


Figure 3.2 Collection locations, habitat, and phylogeographic relationships of seventeen *Xantusia vigilis* populations

Lizards in the northern part of the range shelter under the monocot shrub *Hesperoyucca whippeli* or under bark of fallen Gray Pine (*Pinus sabiana*) logs, unlike the mixed *Yucca* (*brevifolia*, *baccata*, *schidigera*) sheltering sites found in the Mojave desert. Sites names are indicated by codes in Table 1. The full geographic range of *X. vigilis* range and the portion of the range sampled here (gray box) is indicated in

the inset map. Note that the two main phylogenetic clades meet across a short geographic distance in the Transverse Ranges, demonstrating a biogeographic break that does not follow habitat breaks. The northern and southern clades are reciprocally monophyletic, so no directionality of north-south colonization can be inferred from the tree.

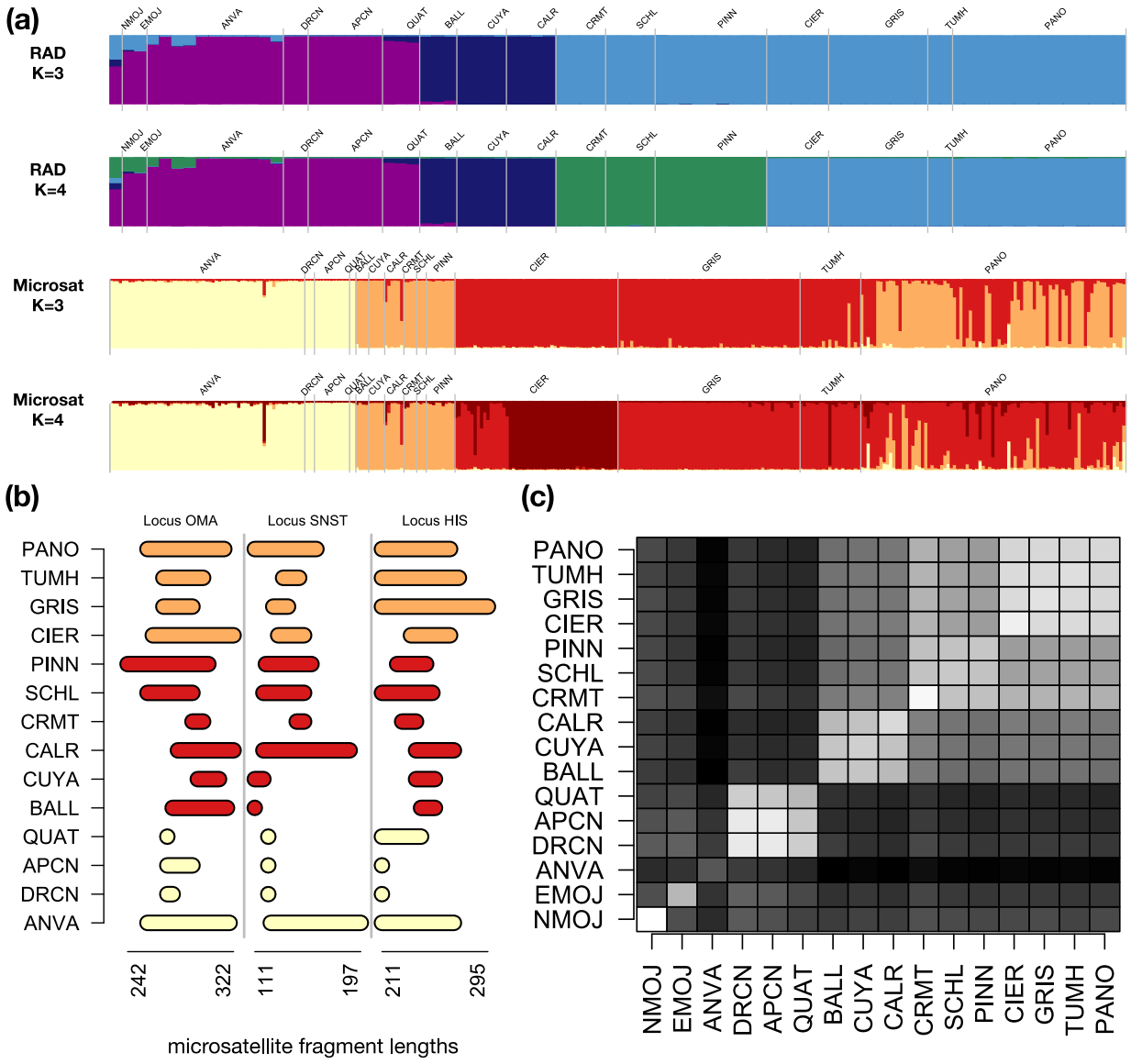


Figure 3.3 Structure demes and marker-specific patterns of diversity

(a) RAD SNP K=3 and K=4 followed by microsatellite results K=3 and K=4. Structure separates the southern from the northern Transverse populations at both K values for both marker types. Microsatellites group the northern Transverse populations with Pinnacles, while the SNP data assigns them to their own deme. Both marker types group the southern Transverse populations with the Mojave samples. The Mojave SNP runs show introgression from the Panoche/Pinnacles region that is absent in the southern

Transverse samples. Microsatellite K=4 shows a unique deme within the Ciervo population. (b) Range between the largest and smallest allele in each population for three representative loci. Populations are colored by their K=3 structure deme. North and south Transverse populations are observably different in allele size range, while allele size range seems to shift more smoothly between the north Transverse, Pinnacles, and Panoche regions. (c) Average pairwise number of SNP differences between each pair of individuals in each population. Note that this measure recapitulates the Structure demes. It also shows the relative similarity between the Pinnacles and Panoche groups, which accounts for their occupying the same deme with K=3. The three South Transverse populations are the most distinct from any other regional deme (dark bands), which the two Pinnacles populations are the most similar to every other group (lighter bands).

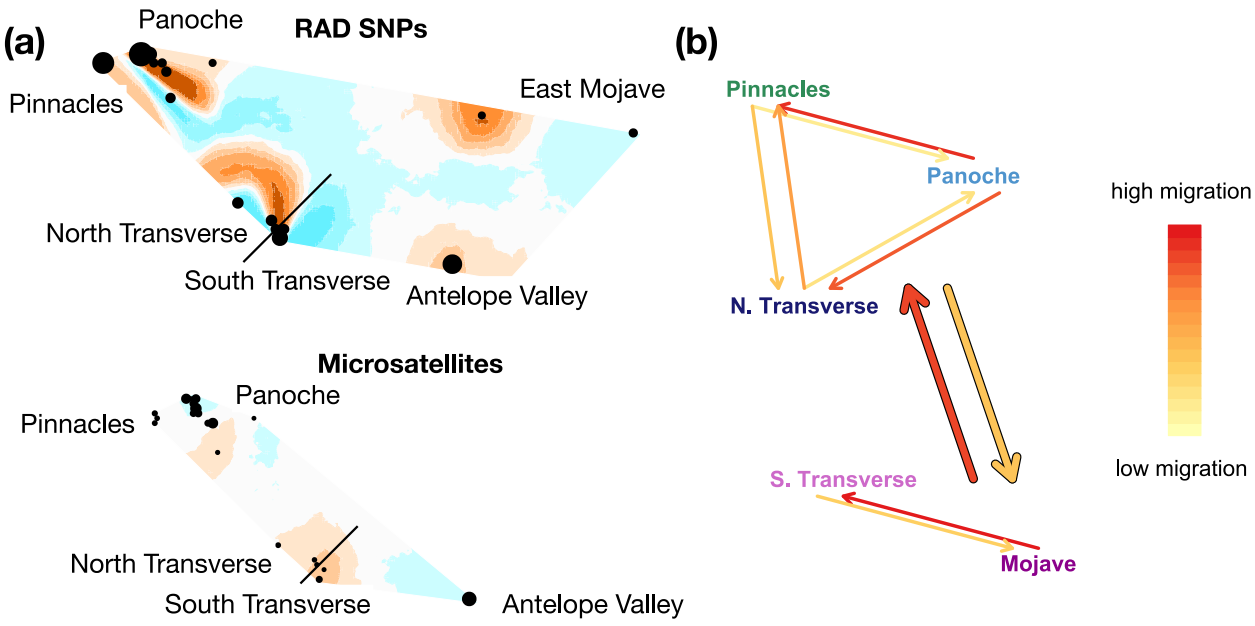


Figure 3.4 Patterns of regional connectivity by marker type

(a) Estimated effective migration rates for RAD SNPs (top) and microsatellites (bottom). In both cases, the Transverse Ranges are an area of low effective migration, while the Panoche and Mojave regions are areas of high effective migration. (b) Historical signatures of expansion between the major clades recovered in our RAxML analysis using RAD SNPs. We follow the nested structure of the phylogeographic tree, first testing directionality between the northern and southern groups, and then between major divisions within the groups. Our analysis shows expansion from south to north (thick arrows). Within the two major groups, expansion from the Mojave populations to the South Transverse populations, and from the Panoche to Pinnacles and North Transverse populations.

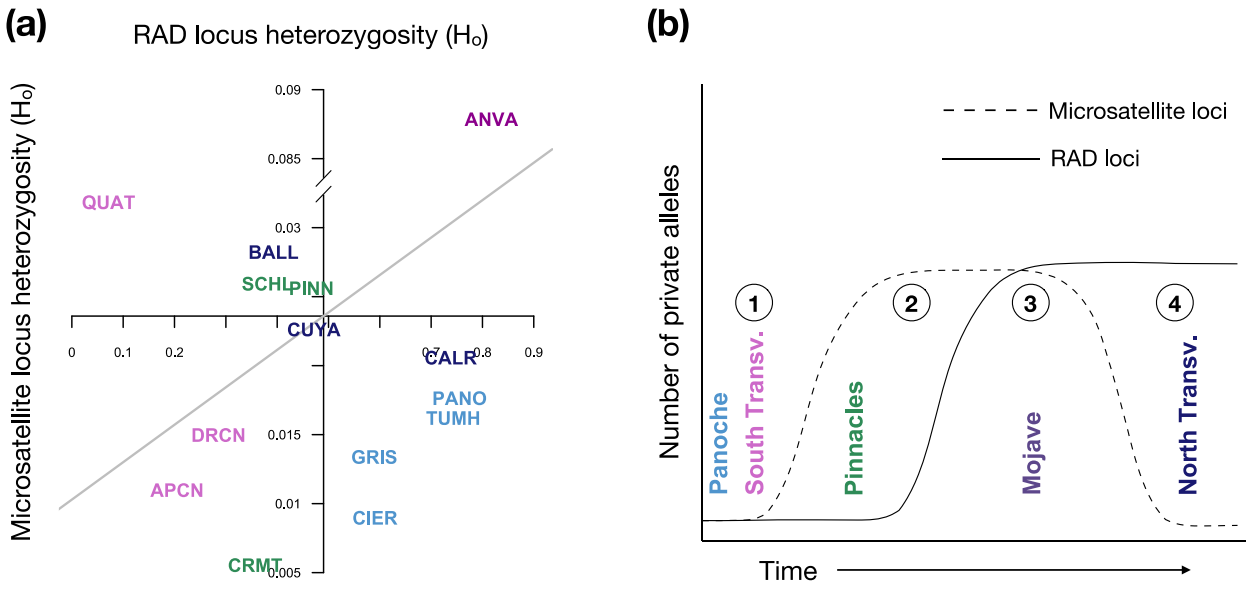


Figure 3.5 Demographic events shown by marker discordance in private allele rates and heterozygosity

(a) Heterozygosity discordance shows two North Transverse populations in the upper left quadrant, showing signs of more recent bottlenecks, while Panoche populations occur in the lower right quadrant, indicating an older bottleneck followed by a population rebound. Axes are placed at mean heterozygosity values for each marker type, and the grey line represents the results from a linear regression of RAD heterozygosity against microsatellite heterozygosity. (b) Discordance in private allele rates at our two marker types place the populations on a time-since-isolation axis. The North Transverse and Pinnacles populations show signatures of long-term isolation.

Table 3.1 Coordinates, sample sizes, and locus diversity characteristics

Population	Code	Latitude	Longitude	RAD N	micro satellite N	RAD Ho	micro satellite Ho	RAD private alleles	microsatellite private alleles
Panoche Region				29	210	0.0149	0.631	0.3588	0.0571
Panoche	PANO	36.6607	-120.75175	14	83	0.0176	0.7553	0.2894	0.0482
Griswold	GRIS	36.5313	-120.74115	8	57	0.0134	0.5892	0.1994	0.0526
Tumey	TUMH	36.503	-120.67454	2	19	0.0162	0.7431	0.1218	0.0526
Ciervo	CIER	36.4308	-120.54703	5	51	0.009	0.5906	0.1083	0
Pinnacles Region				17	16	0.021	0.5231	0.4271	0.25
Pinnacles	PINN	36.4829	-121.17629	9	9	0.0256	0.4667	0.341	0.4444
South Chalone	SCHL	36.4355	-121.18401	4	3	0.0259	0.381	0.2863	0
Curry Mountain	CRMT	36.1948	-120.37915	4	4	0.0055	0.3571	0.0772	0
North Transverse Region				11	15	0.0234	0.55	0.4288	0.0667
Caliente Ridge	CALR	35.094	-119.828	4	6	0.0206	0.7381	0.2109	0

Cuyama	CUYA	34.9459	-119.47849	4	5	0.0226	0.4714	0.2166	0
Ballinger	BALL	34.8836	-119.43983	3	4	0.0282	0.3929	0.2524	0.25
South Transverse Region				11	16	0.0174	0.1053	0.373	0.125
Quatal	QUAT	34.8392	-119.35552	3	2	0.0318	0.0714	0.2593	0
Apache Canyon	APCN	34.7551	-119.39395	6	11	0.011	0.2049	0.1097	0.0909
Dry Canyon	DRCN	34.7172	-119.42048	2	3	0.015	0.2857	0.0564	0
Mojave Region				14	61	0.0765	0.7738	0.721	0.4426
east Mojave	EMOJ	35.7699	-115.85529	2	-	0.0307	-	0.2732	-
northeast Mojave	NMOJ	35.9903	-117.40484	1	-	0.0424	-	0.2353	-
Antelope Valley	ANVA	34.4911	-117.71298	11	61	0.0879	0.8173	0.6166	0.4426

Locus diversity characteristics for regional demes (as defined by RAxML analysis) appear in the grey bars.

Table S 3.1 Sample names, coordinates, sampling populations and regions, and the marker type(s) genotyped for each individual

Sample	Longitude	Latitude	Population	Code	Region	RAD genotype	microsatellite genotype
31BFF	-120.806667	36.65267	Panoche	PANO	Panoche	yes	no
6Af	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	no
A1M	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	no
AC1f	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	no
AC3	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	no
AC8	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	no
Apache11	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	yes	yes
AQ1302	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
AQ1309	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
AQ1322	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes

AQ1323	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
BALX2F	-119.43984	34.88359	Ballinger	BALL	North Transverse	yes	yes
BALX3	-119.4398	34.88358	Ballinger	BALL	North Transverse	yes	yes
BALX4	-119.43984	34.88359	Ballinger	BALL	North Transverse	yes	yes
CALR2	-119.82372	35.09582	Caliente Ridge	CALR	North Transverse	yes	yes
CALR3	-119.82372	35.09582	Caliente Ridge	CALR	North Transverse	yes	yes
CALR7	-119.82372	35.09582	Caliente Ridge	CALR	North Transverse	yes	yes
CALX2_13_11	-119.83228	35.09214	Caliente Ridge	CALR	North Transverse	yes	no
CIER2_14	-120.526937	36.43247	Ciervo	CIER	Panoche	yes	yes
CIER2_5	-120.526937	36.43247	Ciervo	CIER	Panoche	yes	yes
CIER2_7	-120.526937	36.43247	Ciervo	CIER	Panoche	yes	yes
CIER3_2	-120.57716	36.42823	Ciervo	CIER	Panoche	yes	yes
CIER3_9	-120.57716	36.42823	Ciervo	CIER	Panoche	yes	yes
CurryXN1	-120.47121	36.10288	Curry Mountain	CRMT	Pinnacles	yes	yes
CurryXN2	-120.47117	36.10241	Curry Mountain	CRMT	Pinnacles	yes	yes

CurryXN3	-120.47192	36.10232	Curry Mountain	CRMT	Pinnacles	yes	yes
CurryXN4	-120.10229	36.47166	Curry Mountain	CRMT	Pinnacles	yes	yes
CUYX1	-119.47857	34.94585	Cuyama	CUYA	North Transverse	yes	yes
CUYX3	-119.47854	34.94588	Cuyama	CUYA	North Transverse	yes	yes
CUYX5	-119.47845	34.94594	Cuyama	CUYA	North Transverse	yes	yes
CYX4F	-119.47853	34.94591	Cuyama	CUYA	North Transverse	yes	yes
DCX2	-119.420483	34.71718	Dry Canyon	DRCN	South Transverse	yes	yes
GRIS2_F_1	-120.726266	36.54471	Griswold	GRIS	Panoche	yes	yes
GRIS2_U_4	-120.726266	36.54471	Griswold	GRIS	Panoche	yes	yes
Gris3_13	-120.76881	36.50372	Griswold	GRIS	Panoche	yes	yes
Gris3_14	-120.76881	36.50372	Griswold	GRIS	Panoche	yes	yes
Gris3_22	-120.76881	36.50372	Griswold	GRIS	Panoche	yes	yes
Gris4_15	-120.728383	36.54543	Griswold	GRIS	Panoche	yes	yes
Gris4_17	-120.728383	36.54543	Griswold	GRIS	Panoche	yes	yes
Gris4_5	-120.728383	36.54543	Griswold	GRIS	Panoche	yes	yes
MTD1	-120.7265	36.63967	Panoche	PANO	Panoche	yes	yes
MTD2	-120.7265	36.63967	Panoche	PANO	Panoche	yes	yes

MVZ232669	-119.420483	34.71718	Dry Canyon	DRCN	South Transverse	yes	yes
MVZ232719	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	yes	yes
MVZ232720	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	yes	yes
N2_N4_2	-120.74113	36.69145	Panoche	PANO	Panoche	yes	yes
N4_10_9A	-120.741133	36.69145	Panoche	PANO	Panoche	yes	yes
N4_8	-120.741133	36.69145	Panoche	PANO	Panoche	yes	yes
PINNXBG1	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	yes	yes
PINNXFC1	-121.164	36.47395	Pinnacles	PINN	Pinnacles	yes	yes
PINNXFC2	-121.163517	36.4738	Pinnacles	PINN	Pinnacles	yes	yes
PINNXFC3	-121.163833	36.47422	Pinnacles	PINN	Pinnacles	yes	no
PINNXFC5	-121.16385	36.47383	Pinnacles	PINN	Pinnacles	yes	yes
PINNXOP4	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	yes	yes
PINNXSC11	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	yes	yes
PINNXSC13	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	yes	yes
PINNXSC2	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	yes	yes
PINNXSC9	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	yes	no
PINNXWS1	-121.182267	36.50825	Pinnacles	PINN	Pinnacles	yes	yes
QTX1M	-119.35902	34.83679	Quatal	QUAT	South Transverse	yes	yes

QTX4F	-119.35471	34.83979	Quatal	QUAT	South Transverse	yes	no
QTX5F	-119.35463	34.83976	Quatal	QUAT	South Transverse	yes	no
RT2	-120.750438	36.722	Panoche	PANO	Panoche	yes	no
RTF8	-120.750438	36.722	Panoche	PANO	Panoche	yes	yes
SE4_14	-120.740856	36.63511	Panoche	PANO	Panoche	yes	yes
SE4_6	-120.740856	36.63511	Panoche	PANO	Panoche	yes	yes
SE4_7	-120.740856	36.63511	Panoche	PANO	Panoche	yes	no
SE4_9	-120.740856	36.63511	Panoche	PANO	Panoche	yes	no
SW17	-120.806667	36.65267	Panoche	PANO	Panoche	yes	yes
SW18	-120.806667	36.65267	Panoche	PANO	Panoche	yes	yes
TC1003	-115.85587	35.76888	east Mojave	EMOJ	Mojave	yes	no
TC1006	-115.85471	35.7709	east Mojave	EMOJ	Mojave	yes	no
TC2081	-117.40484	35.99031	northeast Mojave	NMOJ	Mojave	yes	no
TUM1_15	-120.674541	36.50301	Tumey	TUMH	Panoche	yes	yes
TUM1_21	-120.674541	36.50301	Tumey	TUMH	Panoche	yes	yes
WW1492	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
WW1494	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
WW1496	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes

WW1497	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
WW1510	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
WW1545	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
WW1567	-117.713	34.49113	Antelope Pass	ANVA	Mojave	yes	yes
APAC10	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC12	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC2	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC3	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC4	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC7	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
APAC9	-119.39395	34.75505	Apache Canyon	APCN	South Transverse	no	yes
CALR1	-119.83228	35.09214	Caliente Ridge	CALR	North Transverse	no	yes

CALR4	-119.83228	35.09214	Caliente Ridge	CALR	North Transverse	no	yes
CALX2_13_4	-119.83228	35.09214	Caliente Ridge	CALR	North Transverse	no	yes
CALX2_15	-119.83228	35.09214	Caliente Ridge	CALR	North Transverse	no	yes
CIERa1	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa10	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa11	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa14	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa15	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa16	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa17	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa18	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa19	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa2	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa3	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa4	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa5	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa6	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa7	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa8	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERa9	-120.53444	36.43439	Ciervo	CIER	Panoche	no	yes
CIERb1	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb10	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes

CIERb11	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb12	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb13	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb15	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb16	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb17	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb18	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb19	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb2	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb20	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb3	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb4	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb6	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb8	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERb9	-120.526937	36.43247	Ciervo	CIER	Panoche	no	yes
CIERc1	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIER3_11	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc2	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc4	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc5	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc6	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc7	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERc8	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERca1	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERca3	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes

CIERca4	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERca5	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
CIERd1	-120.57716	36.42823	Ciervo	CIER	Panoche	no	yes
GRISa1	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa10	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa2	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa3	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa4	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa5	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa6	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa7	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa8	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISa9	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb10	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb11	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb12	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb13	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb3	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb4	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb5	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb6	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb7	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb8	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISb9	-120.726266	36.54471	Griswold	GRIS	Panoche	no	yes
GRISc10	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes

GRISc11	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc12	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc16	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc17	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc18	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc19	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc20	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc21	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc6	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc7	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc8	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISc9	-120.76881	36.50372	Griswold	GRIS	Panoche	no	yes
GRISd10	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd11	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd14	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd16	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd18	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd19	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd2	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd20	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd21	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd22	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd23	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd25	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd4	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes

GRISd6	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd7	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
GRISd8	-120.728383	36.54543	Griswold	GRIS	Panoche	no	yes
MTD3	-120.7265	36.63967	Panoche	PANO	Panoche	no	yes
MTD4	-120.7265	36.63967	Panoche	PANO	Panoche	no	yes
MTD5	-120.7265	36.63967	Panoche	PANO	Panoche	no	yes
NORF1	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF11	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF12	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF13	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF14	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF15	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF16	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF17	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF18	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF2	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF20	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF21	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF22	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF3	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF4	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF5	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF6	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF7	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORF9	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes

NORO1	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORO2	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORO3	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORO4	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORO5	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
NORO6	-120.74113	36.69145	Panoche	PANO	Panoche	no	yes
SEF1	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF10	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF11	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF12	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF13	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF15	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF16	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF17	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF18	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF19	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF3	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF4	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SEF5	-120.740856	36.63511	Panoche	PANO	Panoche	no	yes
SW1	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW10	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW11	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW12	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW13	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW14	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes

SW15	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW16	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW19	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW20	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW21	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW22	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW24	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW25	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW26	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW27	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW28	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW29	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW3	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW30	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW32	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW4	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW5	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW6	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW7	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW8	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
SW9	-120.806667	36.65267	Panoche	PANO	Panoche	no	yes
PANW1	-120.808158	36.65115	Panoche	PANO	Panoche	no	yes
PANW2	-120.808158	36.65115	Panoche	PANO	Panoche	no	yes
PANW3	-120.808158	36.65115	Panoche	PANO	Panoche	no	yes
PANW4	-120.808158	36.65115	Panoche	PANO	Panoche	no	yes

TUMa1	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa10	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa11	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa12	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa13	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa14	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa16	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa17	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa18	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa19	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa2	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa20	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa22	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa23	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa5	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa6	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa7	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
TUMa9	-120.674541	36.50301	Tumey	TUMH	Panoche	no	yes
BALX5F	-119.43984	34.88359	Ballinger	BALL	North Transverse	no	yes
DCX1	-119.420483	34.71718	Dry Canyon	DRCN	South Transverse	no	yes
FC6	-121.163833	36.47422	Pinnacles	PINN	Pinnacles	no	yes
NW1	-121.163516	36.4747	Pinnacles	PINN	Pinnacles	no	yes
OP1	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	no	yes

OP3	-121.182633	36.48035	Pinnacles	PINN	Pinnacles	no	yes
RT4	-120.750438	36.722	Panoche	PANO	Panoche	no	yes
RTF2	-120.750438	36.722	Panoche	PANO	Panoche	no	yes
RTU4	-120.750438	36.722	Panoche	PANO	Panoche	no	yes
PINN1	-121.163516	36.4747	Pinnacles	PINN	Pinnacles	no	yes
SC1	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	no	yes
SC3	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	no	yes
SC4	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	no	yes
SC12	-121.18401	36.43547	South Chalone	SCHL	Pinnacles	no	yes
WS2	-121.182267	36.50825	Pinnacles	PINN	Pinnacles	no	yes
WS3	-121.182267	36.50825	Pinnacles	PINN	Pinnacles	no	yes
CYX6F	-119.47852	34.94599	Cuyama	CUYA	North Transverse	no	yes
QUAX3	-119.35463	34.83976	Quatal	QUAT	South Transverse	no	yes
WW1003	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1086	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1160	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes

WW1168	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1205	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1241	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1246	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1388	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1449	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1459	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1477	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1485	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1538	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1540	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1583	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes

WW1620	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW1629	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW302	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW344	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW408	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW554	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW587b	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW691	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW872a	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
WW954	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1076	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1089	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes

AQ1101	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1140	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1151	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1165	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1243	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1251	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1258	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1259	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1273	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1276	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1299	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ1301	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes

AQ301	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ312	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ317	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ338a	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ381	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ705	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ741	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ759	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ764	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ795	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
AQ808	-117.713	34.49113	Antelope Pass	ANVA	Mojave	no	yes
XrES43			Xantusia riversiana	OUTG	Outgroup	yes	no

XrHN257			Xantusia riversiana	OUTG	Outgroup	yes	no
Xawi			Xantusia wigginsi	OUTG	Outgroup	yes	no
Xe004			Xantusia extorris	OUTG	Outgroup	yes	no
Xe005			Xantusia extorris	OUTG	Outgroup	yes	no
Xg27			Xantusia gracilis	OUTG	Outgroup	yes	no
Xg9			Xantusia gracilis	OUTG	Outgroup	yes	no
Xs13			Xantusia sierrae	OUTG	Outgroup	yes	no
Xs17			Xantusia sierrae	OUTG	Outgroup	yes	no

Chapter 4 Predator Perspective Drives Geographic Variation in Frequency-Dependent Polymorphism[2]

4.1 Abstract

Color polymorphism in natural populations can manifest as a striking patchwork of phenotypes in space, with neighboring populations characterized by dramatic differences in morph composition. These geographic mosaics can be challenging to explain in the absence of localized selection because they are unlikely to result from simple isolation-by-distance or clinal variation in selective regimes. To identify processes that can lead to the formation of geographic mosaics, we developed a simulation-based model to explore the influence of predator perspective, selection, migration, and genetic linkage of color loci on allele frequencies in polymorphic populations over space and time. Using simulated populations inspired by the biology of *Heliconius* longwing butterflies, *Cepaea* land snails, *Oophaga* poison frogs, and Sonora ground snakes, we found that the relative sizes of predator and prey home ranges can produce large differences in morph composition between neighboring populations under both positive and negative frequency-dependent selection. We also demonstrated the importance of the interaction of predator perspective with the type of frequency dependence and localized directional selection across migration and selection intensities. Our results show that regional-scale predation can promote the formation of phenotypic mosaics in prey species, without the need to invoke spatial variation in selective regimes. We suggest that predator behavior can play an important and underappreciated role in the formation and maintenance of geographic mosaics in polymorphic species.

[2] **Iris Holmes** and Maggie Rose Grundler, Alison Davis Rabosky. Predator perspective drives geographic variation in frequency-dependent polymorphism. (American Naturalist, Vol. 190, no. 4, pages E78-E94, October 2017)

4.2 Introduction

Intraspecific color polymorphism is an iconic evolutionary phenomenon in which multiple distinct color phenotypes occur within the same population. Color polymorphism has been widely used to study the origins and maintenance of phenotypic variation because the long-term persistence of multiple sympatric morphs within a species suggests the operation of evolutionary processes other than isolation by distance and genetic drift (Clegg & Durbin, 2000; Cox & Davis Rabosky, 2013; Endler, 1980; Mallet & Joron, 1999; Roff, 1996; Sandoval, 1994). Understanding the origin and maintenance of polymorphism within single species may also inform our understanding of lineage diversification because polymorphism can serve as a precursor to speciation (Corl, Davis, Kuchta, & Sinervo, 2010; Hugall & Stuart-Fox, 2012; McLean & Stuart-Fox, 2014; West-Eberhard, 1986).

Despite extensive work on the maintenance of color polymorphism in single populations, some broad-scale spatial patterns of polymorphism remain challenging to explain in empirical systems. The spatial distribution of morphs in most polymorphic systems can be classified into three major spatial arrangements (Fig. 4.1). In the ubiquitous type (Fig. 4.1A), all morphs are present across all populations and in approximately equal frequencies, suggesting that the polymorphism is maintained by a common density-dependent regulatory mechanism (Gosden, Stoks, & Svensson, 2011; E. I. Svensson & Abbott, 2005). In the clinal type (Fig. 4.1B), a gradual change in the extrinsic environment is paralleled by the increase of one morph relative to another (Hegna, Nokelainen, Hegna, & Mappes, 2013). Two common mechanisms generating clinal systems are the differential adaptation of discrete morphs to environmental factors that form a gradient across geographic space or secondary contact between divergent populations (McLean & Stuart-Fox, 2014). In the mosaic type (Fig. 4.1C), polymorphic species display rampant variation in the presence and absence of morphs that is not well predicted by geography or underlying habitat characteristics, with adjacent populations showing very different morph complements despite their proximity and similar habitats (e.g., the state of Arizona in Fig. 4.1C; McLean & Stuart-Fox, 2014).

While the first two types of geographic variation in polymorphism can be explained by relatively straightforward mechanisms, geographic mosaics are the most complex and least understood spatial arrangement of phenotypic diversity (but see Thompson, 2005 for a discussion of the production of mosaics by coevolutionary arms races as a separate phenomenon). Mosaics occur in many polymorphic systems, yet the mechanisms behind their origin remain largely untested (Bonansea & Vaira, 2012; Cox & Davis Rabosky, 2013; Greenwood, 1974; Langham, 2004; Mochida, 2009). Although many studies have tested for the relative roles of neutral and selective processes, most studies that find nonneutral mosaic polymorphism usually invoke the interaction between multiple selective processes or spatial variability in selection pressure to explain their results (L M Cook, 1967; Cox & Davis Rabosky, 2013; Franks & Oxford, 2009; Gigord, Macnair, & Smithson, 2001; Kikuchi & Sherratt, 2015; Langham, 2004; Erik I. Svensson, Abbott, & Härdling, 2005).

However, whether it is necessary to rely on complex selective scenarios to explain this pattern of spatial variation is unclear. Here, we ask whether a single mechanism can simultaneously produce both stable polymorphism within a population and a geographic mosaic across populations within a species. In other words, can geographic mosaics be generated by the same selective mechanism that maintains polymorphism, or do they necessarily represent spatial variation in those pressures?

4.2.1 Frequency-Dependent Selection across Space

One well-supported mechanism for the maintenance of within-population polymorphism is frequency-dependent selection (FDS) by a predator (Bonansea & Vaira, 2012; Greenwood, 1974). Predator-generated negative FDS (NFDS) occurs when the predators disproportionately focus their efforts on more common morphs in a population, allowing rare phenotypes to increase in frequency over time. This mechanism can maintain multiple morphs in a population indefinitely (Huang, Haubold, Hauert, & Traulsen, 2012) and can be caused by dietary wariness (a combination of neophobia and dietary conservatism) exhibited by predators foraging in polymorphic populations (Franks & Oxford, 2009; Marples, Kelly, & Thomas, 2005; Marples & Mappes, 2011; McMahon, Conboy, O'Byrne-White,

Thomas, & Marples, 2014; Smith, 1977). We note that wariness does not require the formation of a search image, over which much controversy exists (see Punzalan, Rodd, & Hughes, 2005 for a review). Positive FDS (PFDS) occurs when predators disproportionately consume rare morphs and often is a result of aposematism (warning coloration associated with unpalatability) discouraging predation on familiar, common morphs (Langham, 2004; see Chouteau, Arias, & Joron, 2016 for other examples of PFDS). Though expected to result in purifying selection, some species under PFDS display striking levels of polymorphism both within and among populations (Borer, Van Noort, Rahier, & Naisbit, 2010; Langham, 2004; Richards-Zawacki & Cummings, 2011), with variability in predator behavior or in predation across different spatial scales speculated as a potential mechanism for the maintenance of geographic diversity (Langham, 2004; Marples & Mappes, 2011; Noonan & Comeault, 2009; Sherratt, 2006).

Although many studies suggest that spatially invariant FDS alone is not enough to explain the mosaic distribution of morph frequency across space (requiring at least the addition of spatial variation in selection or predator behavior; see Aubier & Sherratt, 2015; Gompert, Willmott, & Elias, 2011; McLean & Stuart-Fox, 2014), other research has suggested that simply varying the spatial neighborhood over which FDS operates can lead to disparate outcomes in prey distribution or phenotype as long as dispersal and/or prey encounter rates are varied (Endler & Rojas, 2009; Houston, Stevens, & Cuthill, 2007; Molofsky, Bever, & Antonovics, 2001, Molofsky, Bever, Antonovics, & Newman 2002). Similarly, field experiments involving polymorphic prey species have suggested that variation in morph frequencies among populations may be influenced by whether predators are restricted to foraging on single prey populations (Comeault & Noonan, 2011; Mallet et al., 1990; Noonan & Comeault, 2009). These studies suggest that discordance between the ranges of predators and prey can produce variation in the outcome of a single selective process, without requiring variation in selective intensity over space.

However, while studies such as these indicate the importance of considering the relative home ranges of predator and prey, they do not directly compare the effects of predator foraging at different spatial scales on morph frequency changes within and among polymorphic prey populations to other forces that can oppose the generation of geographic mosaics. Explicit study of these effects in a single

coherent framework is critical to explain geographic polymorphism, as many polymorphic populations that are subject to NFDS or PFDS by wide-ranging predators differ in various other important ecological or genetic components (see Table 4.1). Only a systematic comparison of the processes commonly observed to affect morph distribution can answer our motivating question: Can a single, uniformly applied, selective force both maintain polymorphism and create a mosaic in empirical systems?

To achieve this framework, we consider both complex mixtures of selection types and the effects of genetic architecture that are found in naturally occurring polymorphic mosaics. While some polymorphism results from multiple unlinked genes or loci, other lineages have evolved polymorphism under the control of a supergene of tightly linked loci (D. Charlesworth, 2016; Schwander, Libbrecht, & Keller, 2014). Because the genetic architecture of a polymorphism may constrain how color and pattern traits are able to behave under selection, the degree of genetic linkage could be critical in understanding the causes and consequences of geographic variation in polymorphism (McLean & Stuart-Fox, 2014; Sinervo & Calsbeek, 2006).

In this article, we explore how the interaction among relative predator range size, type, and strength of FDS and the genetic linkage between color loci affects the phenotypic similarity of polymorphic populations. We base our simulations on empirical knowledge of four polymorphic species that exhibit geographic mosaics and are subject to either NFDS or PFDS (Table 4.1): (1) *Sonora* ground snakes (unlinked loci under NFDS; Cox & Davis Rabosky, 2013; Davis Rabosky, Cox, & Rabosky, 2016), (2) *Oophaga* poison frogs (unlinked loci under PFDS; Richards-Zawacki & Cummings, 2011; Richards-Zawacki, Wang, & Summers, 2012), (3) *Cepaea* land snails (linked loci under NFDS; Ożgo, 2011), and (4) *Heliconius numata* longwing butterflies (linked loci under PFDS; Brown & Benson, 1974). Each of these systems is preyed upon by avian predators that might forage either locally or regionally (Table 4.1) and therefore may exert significant influence on prey phenotypic diversity simply by means of the spatial scale of foraging behavior. Because the two systems that display genetic linkage also experience locally unequal fitness among morphs (e.g., directional selection) due to habitat matching (*Cepaea*) and mimicry (*Heliconius*), we also test the effect of additional components of selection by

comparing these results to two theoretical systems with linked loci under pure FDS (no directional selection), for a total of six systems. By structuring our model around natural systems in a systematic and biologically applicable approach, we explicitly explore the effects of the spatial context of predator behavior on geographic patterns of phenotypic diversity in comparison to empirical observations.

4.3 Methods

In our model, we follow two polymorphic prey populations. They are preyed upon by a predator guild that can either view one population at a time (local perspective) or both populations simultaneously (regional perspective). We allow populations to exchange migrants at rates of 0– 10% of the population per generation, and we vary the proportion of the total mortality in the prey population that is due to predator-induced frequency dependence from 0 to 1. For each unique combination of migration and strength of selection parameter values, the populations follow a sequential set of steps (Fig. 4.2) through migration, breeding, frequency-dependent selection, and density-dependent mortality to a carrying capacity of 500 individuals for 250 generations, at which point the populations have been at a stable equilibrium of morph frequency for at least 50 generations. We then calculate the level of differentiation between the two populations in order to assess the geographic mosaicism of the resulting polymorphism. For each empirical or theoretical system, we run 10 iterations of each model to assess variability among outcomes. The model is written in the R programming language (R Core Team 2016), and the source code is provided as a supplementary data file (supplemental code, available online).

4.3.1 Initializing Starting Populations

We begin by simulating two populations of a prey species under the same starting conditions with alleles at equal frequencies and in Hardy-Weinberg equilibrium (step 1 in Fig. 4.2). Each individual has a phenotype, coded for by two loci with two randomly generated alleles each. The alleles show simple dominance, with recessive indicated by 0 and dominant by 1, and together code for four phenotypes. The loci can be unlinked or linked into a supergene with 2.5% recombination between them. We begin the

first generation by setting the starting frequency of each dominant allele to 0.275, which results in equal morph frequency, and then stochastically generating 200 individuals.

4.3.2 Random Migration

The populations exchange migrants (step 2 in Fig. 4.2), with between 0 and 10% of each population simultaneously moving to the opposite population. The migrants are chosen randomly with respect to their phenotype and genotype. We run simulations with rates of migration increasing by steps of 1% from 0 to 10%. We chose the upper threshold level of migration to exceed the levels of migration found in the empirical literature between populations in natural mosaics but still remain within a realistic range (L. M. Cook, 1998; Twomey et al., 2013). We vary migration rates to understand how the homogenizing effects of gene flow oppose forces generating mosaicism.

4.3.3 Random Breeding

Once migrants are exchanged, individuals can breed only with other individuals in their current population (step 3 in Fig. 4.2). We select 50% of each population for breeding in each generation and assign individuals into pairs randomly with respect to morph. We create all possible gametes from both parents, randomly draw four from each parent's pool, and then combine them to create four offspring genotypes. For the unlinked models, there are four possible gametes from each parent. For the linked models, we create only the two possible gametes from each parent and then recombine them with 2.5% probability according to empirical estimates (Joron et al., 2006; Richards et al., 2013). Finally, the parental generation and the offspring are combined into a single population (no age structuring).

4.3.4 Calculating Total Mortality

We use a standard logistic equation of the form

eq.1

$$N + \left[qN\left(1 - \frac{N}{N^*}\right)\right]$$

with a set carrying capacity N* and a growth rate q to determine how many total individuals should experience mortality in each post breeding population given a carrying capacity set to 500 (step 4 in Fig. 4.2). Because the populations begin at 200 individuals, mortality does not affect the populations until three generations into the simulation, such that there is an initial population growth phase.

4.3.5 Morph-Specific Mortality

To model frequency dependence, we use (van Leeuwen, Brännström, Jansen, Dieckmann, & Rossberg, 2013) elegant closed-form solution describing the functional response of a polymorphic prey population to morph-specific predation (their eq. [1], shown below; step 5 in Fig. 4.2). The term "functional response," in this context, refers to the rate at which prey of morph i are consumed by a predator as a function of the relative frequency of morph i in the population. The equation uses four constants to find fi, or the percentage of the total predator-related mortality that is accounted for by mortality of morph i: the predator's base attack rate on individuals of morph i (c_i), the likelihood of switching from morph i to j (s_{ij}), the rate at which predators transition from handling morph i back to the searching state (T_{ij}), and the abundances of each morph at time t 2 1 (N_i).

eq.2

$$f_i = \frac{c_i N_i \sum_{k=1}^{n} s_{ik} c_k N_k}{\sum_{k=1}^{n} c_k N_k \left(1 + \sum_{k=1}^{n} s_{kj} T_{kj} c_j N_j\right)}$$

The equation replaces an individual-based stochastic model describing the behavior of foraging predators and is more computationally efficient than modeling predators individually. While the equation contains necessary simplifying assumptions, it nonetheless produces a similar functional response to those observed in nature and approximates the behavior of avian predator guilds (van Leeuwen et al., 2013).

We summarize van Leeuwen et al.'s equation here, starting by briefly describing the stochastic Markov model from which they derived the closed-form solution and the assumptions they make to do so. In the stochastic model, predators are either in a state of searching for prey or handling prey. The model uses constant values that describe the transition rates of the predator from one state to another. The mean time that the predators spend in a given state is the inverse of the transition rate between states, and the

time that each individual spends in a state is an exponential random variable. Once prey has been attacked, predators switch to a handling state and transition back to a searching state at rate T_{ij}. This rate depends on the prey type being handled and the prey type the predator previously handled. If all s_{ii} are greater than s_{ij} (the diagonal elements of the prey-type similarity matrix are greater than any other element), the predator will be more likely to attack many of the same prey type in a row than to switch between prey types.

In order to derive the closed-form solution, van Leeuwen et al. assumed that predators eat a large number of prey in their lifetime, and the predators are as likely to switch from consuming morph i to consuming morph j as they are to switch from j to i, a property they refer to as inversion indifference. Van Leeuwen et al. derive the closed-form solution (eq. [2]) from their semi-Markovian model in their appendix 2. In using this equation in our simulation, we make one additional assumption that all morphs have equal transition rates out of handling ($T_{ij} = T_{ii} = T_{ji}$ for all morphs). We arbitrarily set T to 1 so that the term drops out of the functional response equation. We make this assumption because the morphs of our prey species are identical except in coloration, while the model was originally derived to accommodate more than one prey species that might vary greatly in handling time. We therefore specify that our morphs differ in search time but are identical in handling time. We induce negative frequency-dependent behavior by setting all s_{ii} 10 times greater than s_{ij} (1 vs. 0.1) and positive frequency dependence by setting all s_{ii} 10 times less than s_{ij} (0.1 vs. 1). Additionally, we set base attack rates (c) to 0.5 for each morph, except for the locally unequal fitness systems (e.g., *Heliconius* and *Cepaea*), for which we reduce c for one morph in each population (Table S4.1). We use the van Leeuwen et al. equation to calculate f_i for each morph. Under local predator perspective, morph specific f_i is calculated within each population separately, while populations are pooled by morph under regional predator perspective. We then take the fi values for each morph and normalize them to sum to 1 so that we can find the percentage of the FDS mortality that is contributed by each morph. We vary the percentage of morph-specific mortality (relative to random density-dependent mortality) from 0 to 100%, increasing by increments of 5% (step 5 in Fig. 4.2). Varying this parameter allows us to understand how large a

proportion of the prey population must be available to FDS for frequency-dependent effects to become the dominant force in shaping morph distribution. This variation reflects empirical populations where only a proportion of the predator population exhibits dietary wariness or in which prey have limited surface activity (Franks & Oxford, 2009; Marples & Mappes, 2011; Richards et al., 2013; Smith, 1977; Thomas, Bartlett, Marples, Kelly, & Cuthill, 2004). To calculate the total number of mortalities due to FDS, we multiply the total mortality value from step 4 by the availability parameter. We then multiply this value by the vector of the four normalized mortality frequencies to get the total number of individuals of each morph that will be removed by FDS. We remove these individuals randomly with respect to the underlying genotype that creates a given morph.

4.3.6 Random Mortality

When the percentage of mortality due to FDS is less than 100, we remove individuals randomly with respect to morph until the population hits the carrying capacity calculated in step 4 (step 6 in Fig. 4.2). This step creates density-dependent mortality that keeps the population size from increasing beyond the first few generations. After this step, the populations then cycle back to the migration step to repeat the loop of functions until 250 generations are completed and a stable equilibrium is reached.

4.3.7 Evaluating Population Outcomes

To summarize polymorphism levels within and between populations and assess mosaicism, we used a modified Shannon- Wiener diversity index based on (Aubier & Sherratt, 2015) and implemented using the R package vegan (Oksanen et al., 2018; step 7 in Fig. 4.2). Aubier and Sherratt (2015) calculate the Shannon index for their regional population (a matrix of local populations) and then subtract the Shannon indexes calculated for each local population to find a relative local Shannon index. We slightly modify their approach to get a single summary statistic for our two-population model. Following Aubier and Sherratt, we first calculate the Shannon index value (H) for each population ($\alpha 1$ and $\alpha 2$, respectively) and the regional pool (γ) as

eq.3

$$H_{a/g} = -\sum_{i=1}^{n} p_i \ln p_i$$

where n is the number of observed morphs and pi is the frequency of morph i in that population. For all models, we calculate morph frequencies by first averaging genotypes across the final 50 generations of each unique parameter set and then compiling morphs from the averaged genotypes into population vectors. For the regional (γ) pool, we simply sum the two population vectors before performing the Shannon calculation. To then create a single summary statistic, we find the index for both populations and subtract the mean of those two values from the regional Shannon index as

eq.4

$$H = H_g - \frac{\left(H_{a1} + H_{a2}\right)}{2}$$

This index differentiates between populations that have equal morph frequencies and those that have a skewed distribution with one or more rare morphs. However, it does not strongly differentiate between populations fixed for a morph and those with one common morph and one rare morph. Additionally, it gives widely varying values for ubiquitous polymorphism arrangements (Fig. 4.1A), from near zero to near maximum depending on relative morph frequencies, making mosaicism difficult to interpret from this metric alone. Because the presence and absence of morphs says more about the evolutionary potential of the population than the frequency of the morphs and is more directly indicative of geographic mosaicism, we further alter the Shannon index such that we code a morph as either present (1) or absent (0) in a local population and calculate the Shannon index for this reduced-complexity vector. We then weight the regional H_γ by 1.5 to provide unique values to every unique population outcome of polymorphism. We report results from both the frequency-based H (Fig. S4.1) and this weighted presence/absence H (main text), and in both cases, we average H values from the 10 independent iterations and tally the number of unique outcomes across iterations (most frequent outcome over the final 50 generations of each unique parameter set). Data underlying figures 3–5 are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5gb05.

4.3.8 Model Predictions

We develop predictions for our simulations based on empirical observations of polymorphic species that display geographic mosaicism. We are interested in the processes that generate variation within and between natural populations. We select three factors to incorporate in our simulation based on their appearances across many polymorphic mosaic systems: (1) We predict that having predators foraging regionally compared to locally will promote maintenance of morphs in the regional population. Our empirical mosaic examples often have birds as their major predators, with home ranges that are likely to encompass multiple prey populations and thus promote regional scale foraging. (2) We predict that linked loci will promote regional mosaicism because we see supergenes in polymorphic mosaic systems that are maintained in the face of purifying selection regimes (e.g., PFDS and directional selection in *Cepaea* and *Heliconius* populations; (Brown & Benson, 1974; J. S. Jones, Leith, & Rawlings, 1977)). And (3) we predict that our results will differ qualitatively depending on whether selection is positive or negative frequency dependent. There are empirical examples of both types of selection producing mosaics, but negative frequency dependence is more likely to maintain multiple morphs in a single local population than positive frequency dependence.

4.4 Results

Overall, we find that both predator perspective and type of FDS had a stronger effect than the genetic architecture of the color loci on the mosaicism of polymorphism across populations (Fig. 4.3 at a single set of parameter values for migration and selection; Fig. 4.4 across all parameter values). We find that migration opposes the formation of mosaics, while even low levels of FDS can drive mosaicism as long as migration is low. Particularly for the cases of pure frequency dependence, regionally foraging predators drive greater dissimilarity between neighboring populations (e.g., more mosaicism) than locally foraging predators, especially at moderate to high levels of FDS. However, the specific morph composition outcomes vary greatly by system, as described below.

4.4.1 Negative Frequency-Dependent Selection: Unlinked Loci, Pure Frequency Dependence

We find that the role of predator perspective is particularly strong during simple negative FDS on unlinked loci, as represented by Sonora snakes (see Table 4.1; Figs. 4.3A, 4.4A). Local predator perspective results in a ubiquitous polymorphism (e.g., type I in Fig. 4.1A) regardless of the level of migration or the proportion of the mortality that is due to frequency-dependent selection. All four morphs are easily maintained in both populations by NFDS, with all morphs remaining at roughly equal frequency in every generation (Fig. 4.3A). In the regional predator perspective, however, we see a mosaic system established when one allele in one population stochastically begins to decline in frequency (Fig. 4.3A), thereby reducing the frequency of the two morphs that have that allele. Eventually, this allele will be lost in one local population, leaving two morphs present. In the other population, the two morphs with the opposite allele at that locus will be favored by regional NFDS and eventually fixed. Once one locus is fixed for opposite alleles in the two populations, the other locus must remain polymorphic in both populations, as loss of another morph from the regional pool is strongly opposed by NFDS. This process results in a two-by-two morph regional mosaic (Fig. 4.3A, red-colored high values corresponding to outcome "2,2,4" in fig. 4), except for situations where migration is high and a low proportion of total mortality is frequency dependent, which lead to a ubiquitous morph arrangement (Fig. 4.4A, midrange values shown in light yellow corresponding to outcome "4,4,4"). At low levels of migration, NFDS needs to make up only 10%– 15% of total mortality to result in a strong mosaic. Additionally, regionally foraging predators drive greater variability in outcomes of morph distribution among iterations (as shown by more gray than blue colors in Fig. 4.5C, top left).

4.4.2 Negative Frequency-Dependent Selection: Linked Loci, Pure Frequency Dependence

The pure NFDS with linkage theoretical framework shows very similar behavior across the parameter space (Fig. 4.4C) to the unlinked model above. Genetic linkage leads to a more gradual switch between the ubiquitous and mosaic arrangements than in the unlinked Sonora case (Fig. 4.3C) and some

additional variability in the 3-D surface of population outcomes (Fig. 4.4C). These results are likely due to the greater role stochasticity can play in morph presence when alleles are linked in parental genotypes at population inception (note greater variation in morph frequency and cycling amplitude in Fig. 4.3C cf. Fig. 4.3A). In both cases, the regional predator view leads to a mosaic arrangement at very low percentages of NFDS relative to total mortality. As in the unlinked simulation, regionally foraging predators drive greater variability in outcomes of morph distribution among iterations (Fig. 4.5C, middle left).

4.4.3 Negative Frequency-Dependent Selection: Linked Loci, Frequency Dependence Plus Local Selection for Morphs

Predator perspective has less effect in this system than in the simpler NFDS systems. In systems with linked color loci and directional selection with NFDS (e.g., *Cepaea* land snails; see Table 4.1), we see three common diversity outcomes across broad regions of parameter space (Fig. 4.4E). First, the favored morph can fix in each population (Fig. 4.3E). This pattern is most common when migration rate is low and the proportion of mortality due to FDS is large (corresponding to outcome "1,1,2" in Fig. 4.4). The local model contains a greater area of parameter space in which this outcome occurs. Alternatively, the same morph can fix in both populations (corresponding to outcome "1,1,1" in Fig. 4.4). This outcome is opposed by NFDS and is therefore most common when the proportion of mortality due to FDS is small. Last, at low values of migration, populations can maintain polymorphism. Unlike the nondirectional models, the morphs do not stay in equal frequency in either the local or regional populations. Instead, the favored morphs (either recessive/recessive or dominant/dominant) are most abundant where they are locally selected, with a few individuals of the recessive/ dominant morphs occurring in both populations (Fig. 4.3E). This pattern also results in a great number of potential morph arrangement outcomes (Fig. 4.5C, bottom left), especially across the high diversity ridge seen under locally foraging predators in Figure 4.4E. The local model contains a set of outcomes not reached by the regional model, all of high diversity (H) values.

4.4.4 Positive Frequency-Dependent Selection: Unlinked Loci, Pure Frequency Dependence

In systems with unlinked color loci and no directional selection (e.g., *Oophaga* frogs; see table 1), PFDS leads to the same morph dominating both populations ("1,1,1" outcomes) under localized predator perspective at nearly all regions of parameter space (Fig. 4.4B), except for low levels of migration or selection, which can produce "1,1,2" outcomes (Figs. 4.3B, 4.4B). This kind of purifying selection is traditionally predicted for systems under PFDS. With no migration, the populations can fix for different morphs that stochastically rise to higher numbers in the first few generations of the simulation (Fig. 4.3B). Averaging these two most common outcomes (1 and 2 in Fig. 4.5A; top two panels of Fig. 4.5B) produces the light blue ridge in Figure 4.4B; however, only regional predation allows populations to remain polymorphic with a subset of the four morphs at relatively low migration and intermediate to high levels of selection (Fig. 4.4B), although we do not see this outcome when migration is absent (Fig. 4.3B). As in the unlinked and linked NFDS models, regionally foraging predators drive greater variability in morph distribution outcomes among iterations across most parameter space (Fig. 4.5C, top right).

4.4.5 Positive Frequency-Dependent Selection: Linked Loci, Pure Frequency Dependence

As in the results for NFDS, the theoretical framework of simple PFDS on linked loci with no additional components of selection (Figs. 4.3D, 4.4D) looks similar to those from the simulation of PFDS on unlinked loci (Figs. 4.3B, 4.4B). Linkage allows more of the parameter space to fix for a single morph under both predator perspectives. We hypothesize that linked alleles are stochastically fixed for a single locus more quickly than unlinked alleles. Again, regionally foraging predators drive greater variability in outcomes of polymorphism among iterations across most parameter space (Fig. 4.5C, middle right).

4.4.6 Positive Frequency-Dependent Selection: Linked Loci, Pure Frequency Dependence Plus Local Selection for Morphs

In the simulation of PFDS on linked loci with localized directional selection (e.g., *Heliconius* longwing butterflies with different local models; see table 1), low migration and high selection lead to the fixation of a different morph in each population under both localized and regional predator perspective ("1,1,2" outcomes; Figs. 4.3F, 4.4F). Individuals with the dominant/dominant phenotype (favored by directional selection in one population) will always leave offspring that share the parental morph, while the recessive/recessive phenotype (favored in the other population) will have only same morph offspring after mating with another recessive/recessive parent. Thus, the favored morph in each population quickly climbs to numerical dominance in our local predator perspective simulation, where the effects of PFDS can reinforce the effects of directional selection. This outcome is because PFDS in one population is not complicated by the rising frequency (due to directional selection) of the favored morph in the other population or high rates of migration of the dominant/dominant morph.

As migration increases, higher levels of selection are required for directional selection to produce the "1,1,2" outcome. At low to intermediate values of selection and high levels of migration, PFDS can oppose directional selection: migration gradually introduces dominant/dominant morphs from population 2 into population 1,where they mate with recessive/recessive morphs and produce dominant/dominant offspring. PFDS then favors the dominant/dominant morph despite local selection against the morph in population 1. The dominant/dominant morph, therefore, eventually goes to fixation in both populations (blue regions of Fig. 4.4F). Under regional predator perspective, there is a small area of parameter space at low levels of migration and selection that produces higher diversity values and more polymorphism through the unique counterbalancing of directional and frequency-dependent selection. However, similar to the *Cepaea* system, results are not substantially different under regional predator perspective, and variability in outcomes is distributed equally between the two perspectives (Fig. 4.5C). Overall, however, these more complex mixtures of frequency-dependent and directional selection as described in systems

like *Cepaea* snails and *Heliconius* butterflies produce the highest levels of population mosaicism (largest orange areas, Fig. 4.4).

4.5 Discussion

Our investigation of the effects of predator perspective demonstrates that across a broad area of parameter space, two adjacent populations under identical selection regimes can exhibit distinctly different dynamics in morph frequency, depending on whether the predator guild forages locally or on a regional scale encompassing both populations. However, we also found that the presence or absence of directional selection can fundamentally alter the outcomes of frequency dependent selection and the response of phenotypic diversity to predation at different spatial scales. We demonstrate that both negative and positive frequency dependence can result in mosaic systems, although additional components of selection may alter within-population polymorphism and make prey populations less sensitive to changes in predator perspective. The proportion of migrants in the population per generation and the proportion of the total mortality that is frequency dependent also determine whether the populations show a mosaic, fixation for a single morph in both populations, or a ubiquitous morph arrangement.

The results from this model provide an important explanation for empirical observations of geographic mosaics in polymorphic species. (Cox & Davis Rabosky, 2013) suggest that striking differences in morph frequency between adjacent populations of the Sonora ground snake are due to a combination of negative FDS and spatial and temporal heterogeneity in the type, strength, or direction of selection. However, there is no explicit mechanism explaining why selection should vary in intensity or direction, especially over small spatial scales. Using our model to combine aspects of selection and spatial scale—negative frequency dependence and large predator range size—we provide evidence that a single selective mechanism can produce dramatic differences in morph composition between geographically close populations without invoking any spatial variation in selection strength or type. We find a qualitatively similar result for systems under simple PFDS as in *Oophaga* poison frogs, demonstrating

that a regional predator perspective can oppose the purifying effects of PFDS and create a regional mosaic in which the two populations demonstrate differences in morph composition. We hypothesize that the maintenance of diversity is due to a balance of regional PFDS and local population dynamics, whereby local morph frequencies can buffer morphs that are regionally rare against the effects of PFDS. This result can explain polymorphic mosaics of poison frogs such as those demonstrated by *Oophaga pumilio* in the Bocas del Toro archipelago (Richards-Zawacki & Cummings, 2011), where some polymorphic populations exhibiting strong mosaicism exist in such close geographic proximity that avian predators are likely to forage among them (Table 4.1).

For polymorphic systems under FDS that demonstrate complex patterns of population-specific selection, we find that local directional selection operating alongside FDS (negative or positive) can significantly affect outcomes of morph distribution. Both *Cepaea* and *Heliconius* simulations, which are under localized selection for crypsis and mimicry, respectively, displayed similar levels of geographic mosaicism under both local and regional predator perspective. Directional selection tends to oppose both positive and negative FDS to drive fixation for the locally fit morph, even under a regional predator perspective, thereby reducing the total number of morphs across a region compared to populations not under directional selection. As a result, although NFDS is able to maintain within-population polymorphism, it is with fewer morphs than are observed in the pure frequency-dependence models. Effects of directional selection are enhanced under local predator perspective at low levels of migration, driving high levels of mosaicism; in contrast, regional predator perspective slightly depresses mosaicism in this area of parameter space by allowing regional-scale NFDS to counter the effects of directional selection that would otherwise drive locally favored morphs to fixation. This result could support the highly variable levels of polymorphism found in natural populations of *Cepaea* (L. M. Cook, 2008).

Conversely, our results suggest that systems like *Heliconius* butterflies are less qualitatively influenced by predator perspective than species subject to NFDS. Therefore, a regional predator perspective is unlikely to account for within-population polymorphism in populations with linked polymorphism under both PFDS and localized directional selection (e.g., *Heliconius numata* populations

under selection for mimicry). However, because predators with a range encompassing multiple *Heliconius* prey populations will drive fixation for opposite alleles between the two simulated populations, our results also suggest that predators need not be segregated among the mimicry rings on which they forage in order to maintain a spatially polymorphic system, as has been previously suggested for *Heliconius erato-melpomene* mimicry complexes (Mallet et al., 1990). We note that most other *Heliconius* species have not evolved supergene architecture, and these systems display regional rather than local polymorphism (Huber et al., 2015). As such, future work could examine directional selection and FDS in the absence of linkage and variation in the strength of directional selection.

Because our simulations examining NFDS and PFDS with linked loci and equal morph fitness do not look very different from their unlinked counterparts, it is possible that genetic linkage does not exert a significant influence on the formation of geographic mosaics under FDS. However, this result may also be due to how we modeled genetic linkage through the random fixation of the initial parental allele combinations and by enforcing all morphs to be present at equal frequencies at population inception. With many supergenes, the rare recombinant phenotypes are often nonrandom and unlikely to be present at high frequency, especially with respect to supergenes in mimicry systems. At least for Batesian mimics, there can be strong selection for either competent mimics with both color components or cryptic individuals with neither, rather than conspicuous but nonmimetic recombinants displaying one of the colors but not both (D. Charlesworth, 2016). One difference we do note in the PFDS simulations is that more areas of parameter space reach fixation for a single morph in the linked models compared to unlinked, as fixation with linkage requires only one loss of an allele as opposed to two. A more nuanced exploration of how genetic linkage can impact the spatial variation in polymorphism would help clarify these effects.

By presenting side-by-side simulations that change only whether the predator guild forages locally or regionally, our results show that spatially varying selection pressures do not need to be invoked to explain dramatic variation in morph presence in adjacent populations. The maintenance of both local and regional phenotypic diversity by a regionally foraging predator occurs in the majority of our

simulations, despite differences in other ecological and genetic processes known to affect morph distribution. Given the greater range sizes of many predators relative to their prey (due to larger body size and/or differences in mobility), this mechanism is likely to pertain to many natural systems in addition to the case studies presented here. However, we also show that directional selection combined with FDS can produce geographic mosaicism regardless of predator perspective. By using a systematic and cohesive process to reveal the ecological and genetic pathways through which a single selective process can and cannot produce patterns in polymorphism, our model provides a new framework for understanding the different ways in which intraspecific diversity is maintained within different taxa.

4.6 Data Archiving

Data underlying figures 3–5 are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5gb05

4.7 Acknowledgments

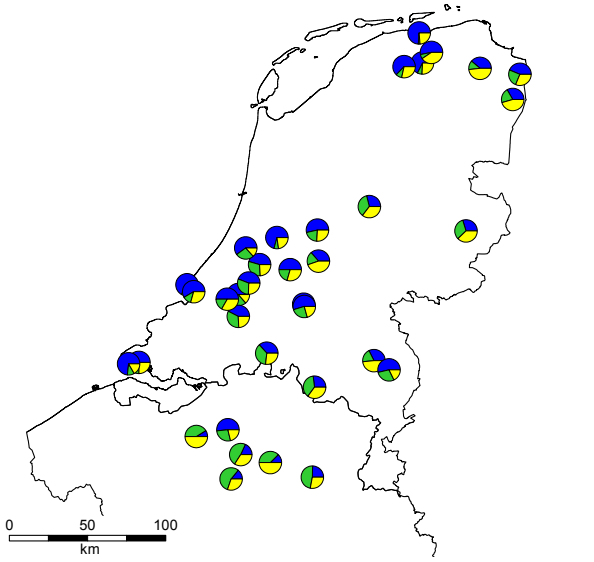
We thank D. Rabosky and the other members of the Rabosky lab at the University of Michigan and five anonymous reviewers for their valuable comments that improved the manuscript. We also thank J. P. Lawrence for use of the photos of *Oophaga pumilio* in figure 2 and the Advanced Research Computing center at the University of Michigan for providing high-performance computing resources to run the simulations.

4.8 Funding

This study was supported by startup funds from the University of Michigan to ARDR.

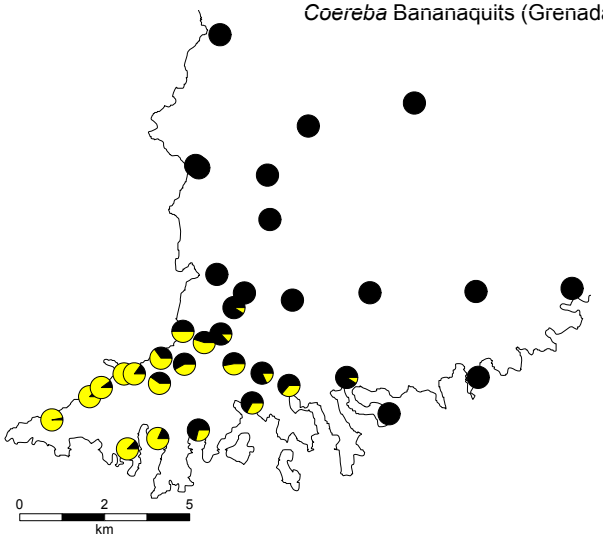
A

Type I: Ubiquitous
Ischnura damselflies (Belgium/Netherlands)



B

Type II: Clinal
Coereba Bananaquits (Grenada)



C

Type III: Mosaic
Sonora snakes (USA/Mexico)

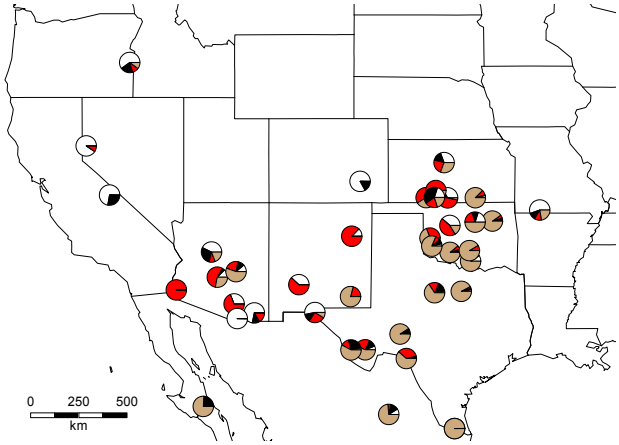


Figure 4.1 Major classes of geographic variation in color polymorphism

A) Type I variation describes the ubiquitous presence of all color morphs in nearly all populations, like the androchome (blue wedges), infuscans (green wedges), and infuscans-obsoleta (yellow wedges) morphs of female damselflies (*Ischnura elegans*) in Europe (data from Gosden et al., 2011; Hammers & Hans Van Gossum, 2008). B) Type II variation describes a geographic cline in frequencies of morphs with fixation at either end, like the yellow-breasted (yellow wedges) and melanistic (black wedges) morphs of bananaquits (*Coereba flaveola*) on the Caribbean island of Grenada (data from Maccoll & Stevenson, 2003; Wunderle, 1981). C) Type III variation describes a geographic mosaic of polymorphism, in which neighboring populations can have drastically different morph complements, and there is no geographic pattern to morph loss or fixation, like the mimetic (white wedges), black-banded (black wedges), red-striped (red wedges), and uniform brown (tan wedges) morphs of ground snakes (*Sonora semiannulata*) in North America (data from Cox & Davis Rabosky, 2013). These classes are not necessarily mutually exclusive but are hypothesized to represent differences in selective pressures, with type III being the most difficult to explain theoretically.

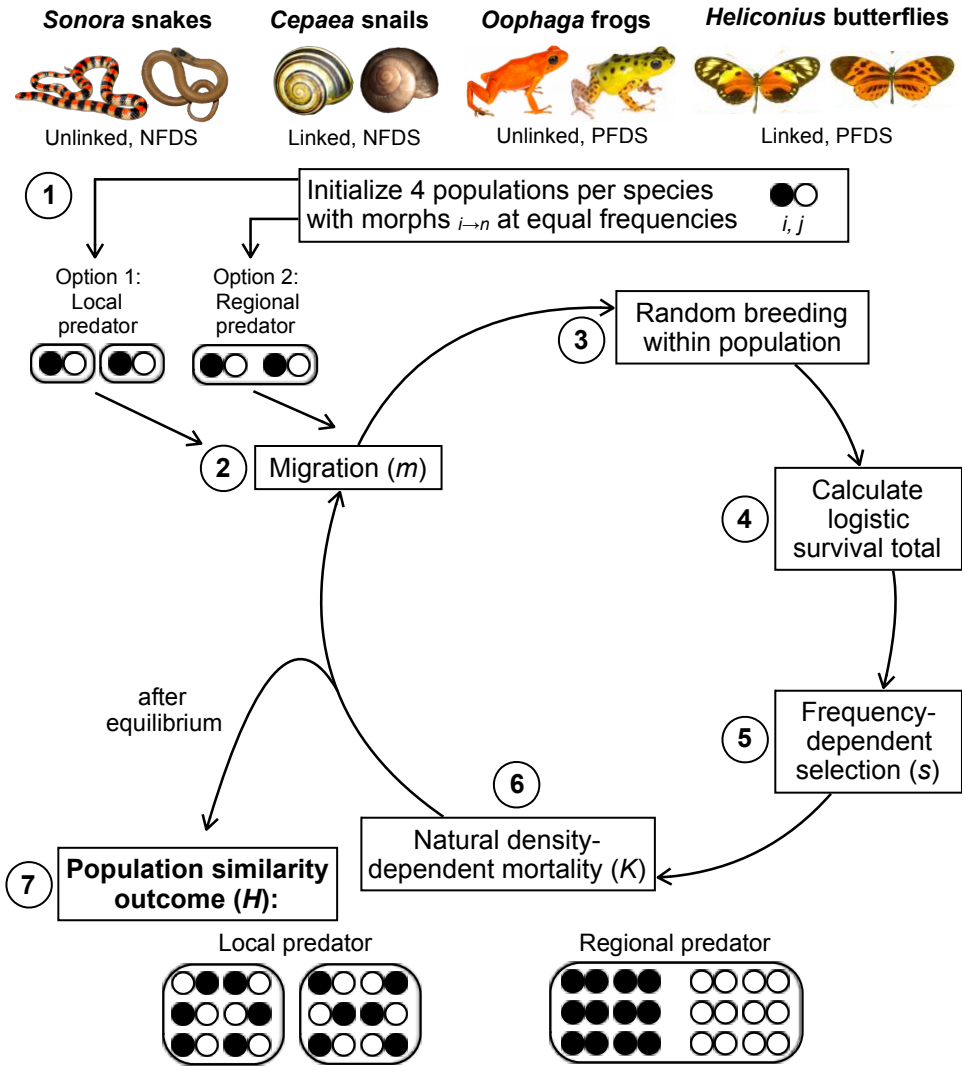


Figure 4.2 Schematic representation of the simulation model

For each modeled system (four empirically derived systems shown along the top, plus two theoretically possible systems), two polymorphic populations experience migration followed by random breeding and mortality by frequency-dependent selection (FDS) by a predator or by density dependence. In separate paths of the model, prey are subject to a locally (L) or a regionally (R) foraging predator. Resulting populations repeat steps 2–5 for 250 generations, leading to differential outcomes under L versus R predation in the level of similarity between the populations (modified Shannon-Wiener diversity index: H; see "Methods").

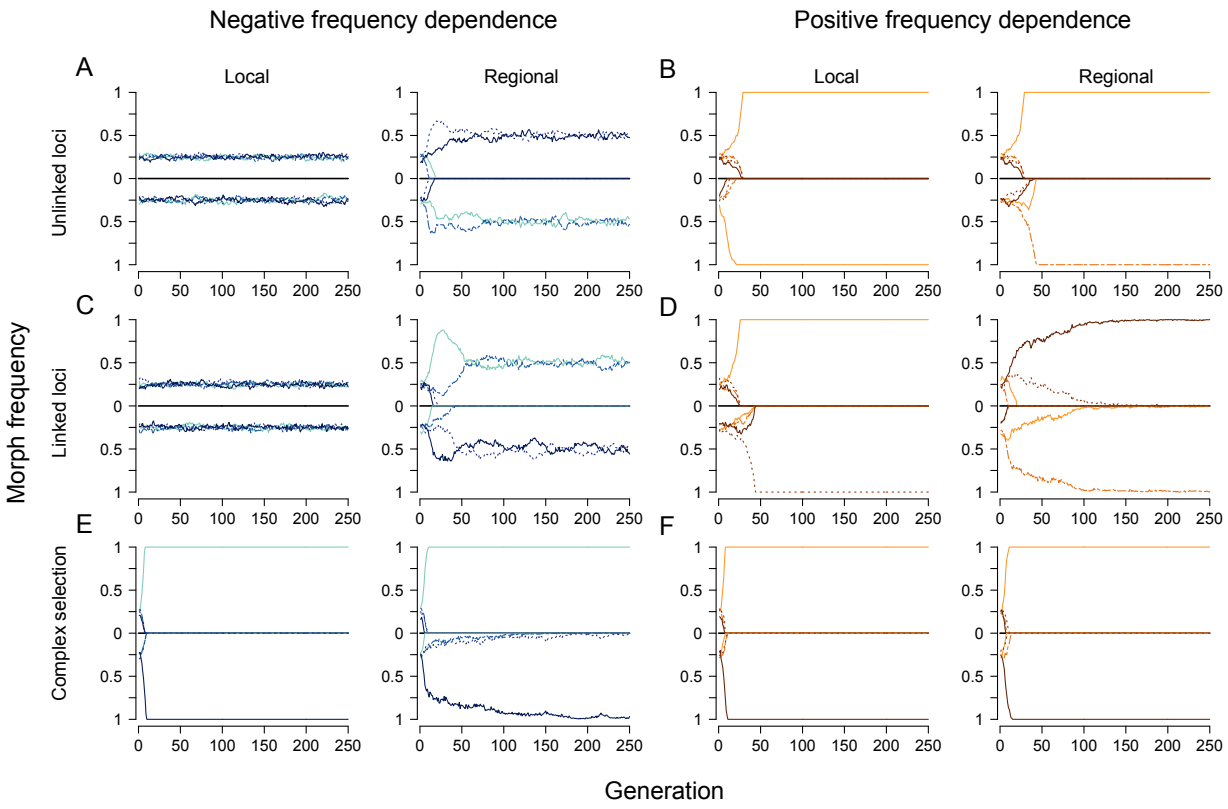


Figure 4.3 Morph frequencies through time

under different selection regimes and scenarios of genetic linkage between color loci, with population 1 on top of and population 2 below the reference line at 0 in each plot. All plots begin at t_0, with all four morphs at equal frequencies and with migration rate (m) at 0, and proportion of selection due to frequency dependence (s) at 0.35. The number and identity of the color morphs at equilibrium are used to calculate an index of population similarity. In systems under pure frequency-dependent selection like *Sonora* snakes (A), *Oophaga* frogs (B), and our theoretical systems (C, D), both the type of frequency dependence (negative vs. positive) and predator perspective (local vs. regional) have stronger effects than genetic linkage on the number and identity of morphs in each population. However, more complex mixtures of frequency-dependent and directional selection as described in systems like *Cepaea* snails (E) and *Heliconius* butterflies (F) produce similar strong mosaicism irrespective of predator perspective at

these parameter values. In general, regionally foraging predators can create more dissimilarity among populations in number and identity of morphs (e.g., mosaicism) than locally foraging predators.

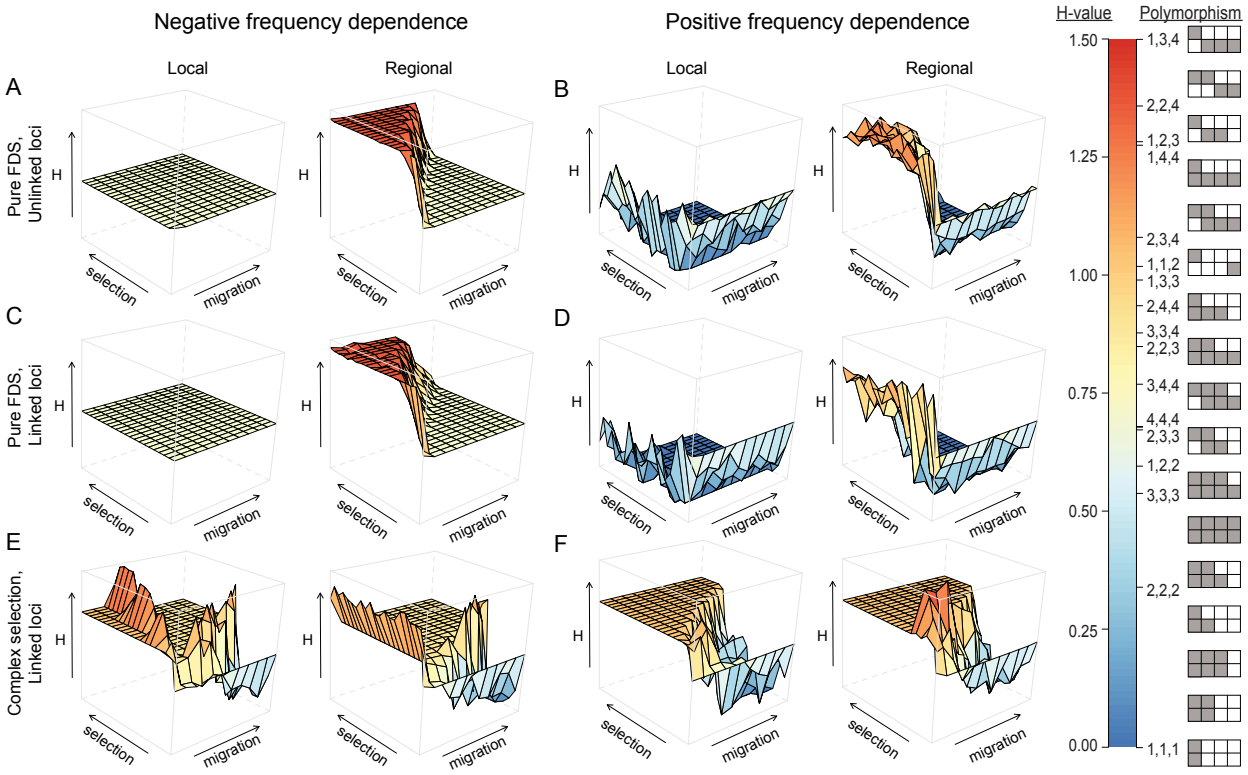


Figure 4.4 Variation in the similarity of morph composition of populations

Variation in the similarity of populations (modified Shannon-Wiener diversity index, H) is driven by type of selection, predator perspective, migration (from 0 to 0.1), and proportion of selection due to frequency dependence (from 0 to 1). Higher values (warmer colors) of the diversity index correspond to greater dissimilarity (more mosaicism) among populations. Each polymorphism outcome (see key at right) that corresponds to a particular H value is represented as both a numerical code (e.g., "1,3,4," in which the left value is the number of morphs in population 1, the middle value is the number of morphs in population 2, and the right value is the number of unique morphs when viewing both populations simultaneously) and graphically (the top four boxes representing the possible morphs in population 1 and the bottom four boxes representing population 2). In systems under pure frequency dependent selection (FDS) like *Sonora* snakes (A), *Oophaga* frogs (B), and our theoretical systems (C, D), regionally foraging predators create higher levels of mosaicism than locally foraging predators, while the genetic linkage of the coloration loci

has minimal effect. However, directional selection in more complex systems like *Cepaea* snails (E) and *Heliconius* butterflies (F) tends to dominate the effects of FDS and predator perspective, creating less variation in outcomes between local and regional models in favor of fixation for a locally selected morph. In all cases, the values of both migration and selection alter the population similarity outcome.

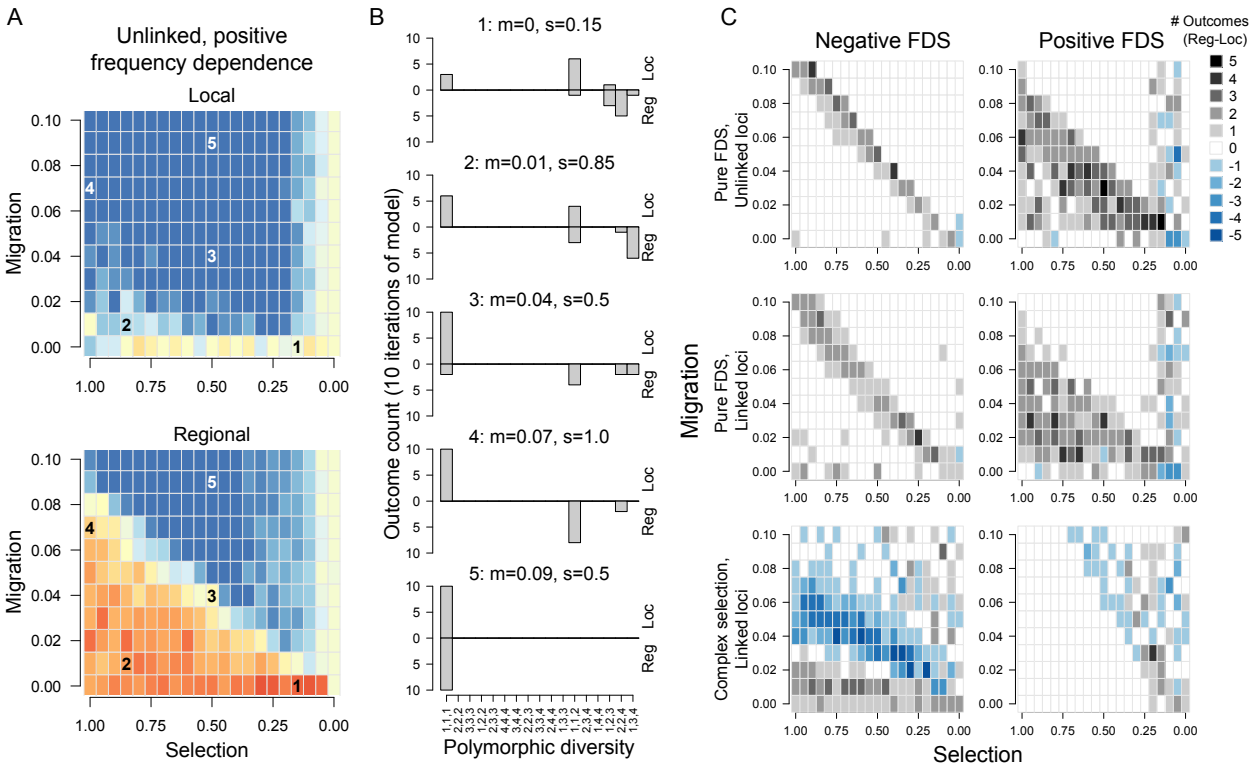


Figure 4.5 Regionally foraging predators tend to increase the number of distinct population outcomes across simple systems

(A) As illustrated by the unlinked, positive frequency dependent selection (PFDS) system of *Oophaga*, the flattened 2-D version of figure 4B (using the same color scale) can be used to explore outcomes at different combinations of selection and migration (numbered 1–5, corresponding to panel B) across the 10 simulation iterations for each system. (B) Not only do regional predators promote higher diversity values (further to the right on the X-axis), they also promote more variability in distinct outcomes across multiple areas of parameter space. (C) When this difference in outcome number (regional 2 local) is explored across all systems and parameter combinations, systems under pure frequency-dependent selection (top four panels) show a greater number of distinct polymorphism outcomes under regionally foraging (more gray coloration) than locally foraging predators (blue coloration), especially for PFDS.

For the systems with complex selection (*Cepaea* [bottom left] and *Heliconius* [right]), there is great heterogeneity in outcomes decoupled from the relative range size of the predator, suggesting that mosaics should be common under both predation scenarios.

Table 4.1 Empirical information supporting model design for each polymorphic system represented in the simulations

Type FDS	Linkage	Representative Study System	Genetic Architecture	Additional Selective Pressures	Main Predator Guild	Predator Territory Size/Observed Foraging Distance	Observed Distance Between Prey Populations	Results Under Localized Predation	Results Under Regional Predation
Negative	Unlinked	*Sonora* ground snakes	Two unlinked loci, simple Mendelian inheritance[1,2,7,8,9]	Batesian mimicry balanced by crypsis[7,10]	Avian (free-ranging, migratory raptor species)[3,4,5,7]	>100km[6]	<100km[7,9]	Ubiquitous polymorphism	Geographic mosaic
Negative	Linked	*Cepaea* land snails	"Supergene"; recombination < 5%[12,13,15,16,18]	Crypsis[13,14,17]	Avian (Song thrush, *Turdus philomelos*)[14,17]	~1200m[11]	~200m[15]	Ubiquitous polymorphism	Geographic mosaic
Positive	Unlinked	*Oophaga* dendrobatid frogs	Likely 1-3 genes[19,23,24]	---	Avian (Amazonian species)[20,21,22]	0.16 sq. km[20]	100-1000m[21,22]	Regional monomorphism or localized fixation	Geographic mosaic
Positive	Linked	*Heliconius* longwing butterflies	"Supergene"; recombination nearly absent[26,30]	Müllerian mimicry[25,29,30]	Avian (Rufous-tailed Jacamar, *Galbula ruficauda*)[27,28,29]	~500-1000m[28]	~30m[25]	Regional monomorphism	Regional monomorphism

Sources of genetic and empirical data: *Sonora:* 1= H. Bechtel & Bechtel, 1962; 2 = H. B. Bechtel, 1978; 3= Kochert & Kucera, 1979; 4= Greene & McDiarmid, 1981; 5= Steenhof, 1983; 6= Steenhof, Kochert, & Moritsch, 1984; 7= Cox & Davis Rabosky, 2013; 8= Cox & Chippindale, 2014; 9= Davis Rabosky, Cox, & Rabosky, 2016, 10= Davis Rabosky, Cox, Rabosky, et al., 2016. *Cepaea*: 11= Davies & Snow, 1965; 12= L. M. Cook, 1967; 13= Greenwood, 1974; 14= J. S. Jones et al., 1977; 15= Bantock & Ratsey, 1980; 16= L. M. Cook, 1998; 17= L. M. Cook, 2013; 18= Richards et al., 2013. *Oophaga:* 19= Summers, Cronin, & Kennedy, 2004; 20= Noonan & Comeault, 2009; 21= Comeault & Noonan, 2011; 22= Richards-Zawacki & Cummings, 2011; 23= Richards-Zawacki et al., 2012; 24= Vestergaard, Twomey, Larsen, Summers, & Nielsen, 2015. *Heliconius:* 25= Turner, 1971; 26= Brown & Benson, 1974; 27= Chai, 1986; 28= Mallet et al., 1990; 29= Langham, 2004; 30=Joron et al., 2006.

Table S 4.1 Reference table for simulation parameters and variables in the model

Parameter	Definition	Default value
s	Strength of frequency-dependent selection (percent total mortality of each prey population due to frequency-dependent predation)	{0, 0.05, 0.10, … 1}
m	Migration between populations	{0, 0.01, 0.02, … 0.10}
K	Carrying capacity of each population	500
Variable		
f_i	Frequency-dependent mortality of morph i (percent of s that is due to frequency-dependent predation on morph i)	Depends on the remaining variables. See equation 2.
c_i	Base attack rate on individuals of morph i (in the absence of FDS, the percentage of time that a morph will be attacked upon each encounter)	In absence of localized directional selection, vector of values for each morph: [0.5, 0.5, 0.5, 0.5] With localized directional selection: Habitat/mimicry model 1: [0.01, 0.83, 0.83, 0.83] Habitat/mimicry model 2: [0.83, 0.83, 0.83, 0.01]

s_{ij} (= s_{ji})	Similarity between morphs *i* and *j* (lower similarity describes proportionally rarer switches of the predator guild from consuming morph *i* to consuming morph *j*)	Matrix of values comparing each morph: For NFDS: $\begin{bmatrix} 1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1 \end{bmatrix}$ For PFDS: $\begin{bmatrix} 0.1 & 1 & 1 & 1 \\ 1 & 0.1 & 1 & 1 \\ 1 & 1 & 0.1 & 1 \\ 1 & 1 & 1 & 0.1 \end{bmatrix}$
N_i	Frequency of morph *i* at time t-1	Depends on results from reproduction and predation (f_i). (Reproduction at generation 1 establishes approximately equal frequencies of all morphs.)

Chapter 5 Natural History Bycatch: a Pipeline for Identifying Metagenomic Sequences in RADseq Data[3]

5.1 Abstract

Reduced representation genomic datasets are increasingly becoming available from a variety of organisms. These datasets do not target specific genes, and so may contain sequences from parasites and other organisms present in the target tissue sample. In this paper, we demonstrate that (1) RADseq datasets can be used for exploratory analysis of tissue-specific metagenomes, and (2) tissue collections house complete metagenomic communities, which can be investigated and quantified by a variety of techniques. We present an exploratory method for mining metagenomic "bycatch" sequences from a range of host tissue types. We use a combination of the pyRAD assembly pipeline, NCBI's blastn software, and custom R scripts to isolate metagenomic sequences from RADseq type datasets. When we focus on sequences that align with existing references in NCBI's GenBank, we find that between three and five percent of identifiable double-digest restriction site associated DNA (ddRAD) sequences from host tissue samples are from phyla to contain known blood parasites. In addition to tissue samples, we examine ddRAD sequences from metagenomic DNA extracted snake and lizard hind-gut samples. We find that the sequences recovered from these samples match with expected bacterial and eukaryotic gut microbiome phyla. Our results suggest that (1) museum tissue banks originally collected for host DNA archiving are also preserving valuable parasite and microbiome communities, (2) that publicly available RADseq datasets may include metagenomic sequences that could be explored, and (3) that restriction site

[3] Iris Holmes and Alison Davis Rabosky. Natural history bycatch: a pipeline for identifying metagenomic sequences in RADseq data. (PeerJ, Vol. 6, e4662, April 2018)

approaches are a useful exploratory technique to identify microbiome lineages that could be missed by primer-based approaches.

5.2 Introduction

Next generation sequencing techniques have dramatically increased our understanding of the phylogenetic diversity of microbial communities, both in the environment and as metagenomic communities within multicellular hosts (Tremblay et al., 2015). Sequencing allows investigations of microbial communities without expensive, time-consuming, and sometimes unreliable culturing techniques (Browne et al., 2016). Metagenomic approaches also allow investigators to assess the relative abundances and activity levels of microbes as they occur in nature (Kozich, Westcott, Baxter, Highlander, & Schloss, 2013; Schloss et al., 2009). New techniques for assessing microbial communities are continually being developed and refined. One area of concern for methods development is binding bias in primer sites that could result in some metagenomic taxa being overlooked in sequencing-based surveys (Clooney et al., 2016), if primer-based approaches are the first and only method of analysis for that community.

The most common approach to sequencing metagenomes requires researchers to amplify a pre-determined barcode primer that can bind to all target taxa. In this paper, we consider all mutualistic, commensal, and parasitic or pathogenic organisms to be part of the host's metagenome. Relationships between hosts and the microbes and larger parasites that live in their tissues are often complex and context-dependent, so we prefer the most general term possible. One of the central problems in designing primers for metagenome-scale analysis is deciding which taxa should be considered during primer design (Dollive et al., 2013). User-friendly bioinformatics techniques and full mitochondrial and nuclear genomes of many taxa of interest have made it easier to design primers for metabarcoding techniques (Riaz et al., 2011). However, primers designed to fit known taxa in a community could completely miss unknown taxa that are present and potentially of interest to the investigator. Even within specifically-

targeted higher taxa, primers often preferentially amplify some taxa and bind poorly to others, thereby potentially altering downstream community-scale analyses (Clooney et al., 2016; Tremblay et al., 2015).

Exploratory techniques that avoid primer-binding bias can help to identify target taxa for primer design, avoiding these problems in later barcoding analysis. Here, we demonstrate that enzyme-based reduced representation library approaches primarily used for host genomic analyses often co-amplify metagenomic DNA along with target host sequences. Our contribution is to provide a pipeline based on widely used analysis platforms, and present proof-of-concept examples on a range of sequence types. Double-digest restriction site associated DNA sequencing (ddRAD) approaches are relatively cheap compared to other non-primer-based next generation sequencing methods (Peterson et al., 2012). We present a protocol for identifying metagenomic DNA incidentally amplified during ddRAD and other short-read sequencing of multicellular host tissues. Previous work has shown that metagenomic sequences can appear in full genome assemblies, indicating that they may also be present in RADseq data (Orosz, 2015). We work across a range of host tissue types, including those commonly preserved as archival DNA in museum collections. We demonstrate that tissue samples collected primarily for genetic work on hosts can now be used to look at blood and tissue metagenome taxa, underlining the importance of long-term tissue preservation in publicly available collections. Additionally, our pipeline will allow researchers designing host-associated metabarcoding projects to survey publicly available datasets in order to refine their set of target taxa.

5.3 Methods

5.3.1 Sample collection and preservation

We sequenced tissues from multiple sources. We collected tail tissue samples (skin, muscle, and cartilage) in the field from two species of horned lizards*, Phrynosoma modestum* (six individuals) and *Phrynosoma cornutum* (nine individuals). We collected the samples using heat-sterilized scissors to avoid contamination, and stored them in RNALater in the field. These samples were collected in southwestern New Mexico in the summer of 2015 (permit number 3,606 to M. Grundler, University of Michigan

IACUC protocol number PRO00006234). Samples were stored at ambient temperature in the field (for up to a month) and at -20 C after being returned to the lab. We also sequenced DNA from night lizard (genus *Xantusia*) liver tissues (Natural History Museum of Los Angeles County, accession numbers TC1002, TC1003, TC1006, and RLB5221). Collection protocol is not available for the museum samples. The first three samples are Xantusia vigilis liver samples collected in 2012, and the last is a *Xantusia riversiana* liver collected in 1972. We also collected cloacal swabs (sterile rayon swabs, MW113) from two ribbon snakes, *Thamnophis sauritus*, and one water snake, *Nerodia sipedon*, in southeastern Michigan in the fall of 2015 (collected under a Michigan Scientific Collecting Permit 9-16-2015 to I. Holmes). We prevented the swab from coming into contact with the environment, the sampler, or the skin of the animals. We avoided contact with the skin around the cloaca by gently applying pressure to the ventral surface of the animal just anterior to the vent. This pressure slightly everted the cloaca, exposing the mucous membrane and allowing us to insert the swab cleanly. We removed the swab and placed it in a sterile 2 mL vial, then broke the shank so that the cap could be put on. We handled the shank of the swab only above the portion that will be preserved in the vial. The samples were transferred to -20 C storage within hours of collection. We also sequenced samples from whole digestive tracts of two *Sceloporus jarrovi* preserved in 95% ethanol and stored at -80 C (permit number SP673841 to Robert M. Cox). To acquire the samples, we dissected out the total lower intestines. We filled a sterile pipette tip with 100 mL of distilled water. We inserted the pipette tip into the intestine section, and depressed the plunger to force the water through the intestine. We collected the wash in a sterile 1.5 mL vial. The samples were sequenced in the fall of 2015.

5.3.2 Laboratory protocol

We extracted total genomic and metagenomic DNA using Qiagen blood and tissue kits with a 12-hour incubation with proteinase-K prior to the spin column extraction. We used a double-digest RADSeq approach (Peterson et al., 2012), with the enzymes EcoR1 and Msp1 from New England Biolabs. We ligated barcoded Illumina adapters to the sticky ends left by the enzyme cuts, and used a PCR to attach

barcoded Illumina primers to double barcode the sequence. We size-selected fragments with genomic inserts between 200 and 300 bp using a Pippin Prep cassette. We performed 125 bp paired-end sequencing the fragments on an Illumina HiSeq platform with V2500 reagents at the University of Michigan Sequencing Core.

5.3.3 Publicly available sequence analysis

We downloaded three double digest RADseq datasets from NCBI's Short Read Archive (SRR1947260 to SRR1947262). They are from the coral snake *Micrurus fulvius* (Streicher et al., 2016). Details of preservation are reported in the original paper. The authors report that samples were liver, heart, shed skin, or scales, and were preserved in ethanol or stored at -80 F. Details of storage on a per-sample basis were not available. Samples were restricted using the enzymes Sbf1 and Sau3A1, and paired-end sequenced on an Illumina HiSeq 2500 platform (Streicher et al., 2016).

5.3.4 Sequence preparation

We demultiplexed the sequences using pyRAD, removed the low-quality sequences, and clustered reads within samples to 97% identity (pyRAD steps 1–3) (Eaton, 2014). We chose this clustering threshold because many microbial ecologists use a 3% difference in sequences to identify operational taxonomic units. We used the resulting fasta file of clustered sequences for each individual (the pyRAD *.edit file) for all further analyses. Any combination of sequence quality control and clustering programs can be used for this step, for example FastQC or Trimmomatic for filtering (Andrews, 2010; Bolger et al., 2014), or vsearch for clustering (Rognes et al., 2016). For the *Phrynosoma* and *Xantusia* samples, we continued the pyRAD pipeline to cluster reads across individuals, and used the resulting "*.loci" file for further analyses. In the pyRAD *.loci files, the sequences for each locus are listed in a group, with the individual that provided the sequence identified in the name of that sequence. A standard line break string separates the sequences for each locus. We used a custom R script to take the first sequence for each locus and combine them into a fasta file to be passed to our analysis pipeline.

5.3.5 Investigating metagenomic sequences

We use NCBI's discontinuous megablast algorithm to compare all sequences from the *.edit and *.loci files to reference sequences in the online NCBI nucleotide database (Camacho et al., 2009). We use the R package taxize to find the genus and species of each sequence (Chamberlain & Szocs, 2013; Chamberlain et al., 2016). We discard results that aligned to more than one kingdom or phylum with greater than 80% identity. To assess how the threshold for similarity affected the number of sequences that can be identified to phylum, we imposed percent similarity thresholds to the closest matching sequence of 70%, 80%, 85%, 90%, 95%, and 97%. To assess the distribution of parasite sequences across hosts, we screened sequences that clustered across individuals of two horned lizard species, *P. cornutum* and *P. modestum*. We performed a similar analysis on samples from two *Xantusia* species. Finally, we built rarefaction curves using the R package vegan to identify the depth of sampling necessary to identify all genera of metagenomic DNA present in the sample (Oksanen et al., 2018). For each genus for which we have tissue metagenomes (*Phrynosoma*, *Xantusia*, and *Micrurus*) we created a community matrix in which samples are rows and columns are the the number of Chordata sequences, and the number of sequences in each genus of blood parasite. We set a 90% identity match for this analysis.

5.4 Results

We identify sequences that match with reference flat worms (Platyhelminthes), round worms (Nematoda), and Apicomplexans (the phylum that contains malarial parasites) from the majority of tissue samples we sequenced (Table 5.1). When sequences are examined at a 97% similarity to reference sequence threshold, we find that from an average of 1,252,549 (s. e. +/- 1,080,872) sequences per sample, 466 (s. e. +/- 301) are identifiable host sequences, 40 (s. e. +/- 70) are from platyhelminths, and 20 (s. e. +/- 34) are from nematodes. On average, we identify 3.2 (s. e. +/- 1.7) unique playthelminth taxa and 2.9 (s. e. +/- 1.8) unique nematode taxa per individual. The large majority of sequences do not have any significant match in the BLAST database at the 97% similarity threshold (Table 5.1). We present two examples in which we alter the threshold for similarity of a sequence to its top hit in GenBank from 70%,

80%, 85%, 90%, 95% to 97% (Fig. 5.1). Increasing the similarity threshold causes the number of sequences matched to each phylum decrease, but generally not to go to zero. Results from gut samples show more and greater diversity of metagenomic taxa relative to sequences from muscle tissue.

We screen a dataset of three desert night lizards (*X. vigilis*) and one island night lizard (*X. riversiana*). Nineteen unique metagenomic sequences have two or more representatives in the final assembly (Fig. 5.2), out of 81,966 total sequences. One hundred and ninety-seven of the sequences in that assembly align with Chordata reference sequences with 97% similarity. Fifteen sequences out of the final assembly align to Platyhelminthes sequences with 97% similarity. Fourteen of these match with *Protopolystoma xenopodis* and one with *Diphyllobothrium latum*. Four nematode sequences align with the species *Nippostrongylus brasiliensis*, *Strongyloides stercoralis*, *Soboliphyme baturini*, and *Elaeophora elaphi* at 97% similarity.

We also investigate the efficacy of the ddRAD approach in surveying the diversity of the hindgut microbiome. We find that the approach reliably returns sequences from the three most common phyla of gut bacteria in reptiles (Fig. 5.3): Proteobacteria, Firmicutes, and Bacteroidetes (Colston & Jackson, 2016). We also retrieve sequences from Platyhelminthes, Nematoda, and Apicomplexa. All three phyla are known gut community members (de Chambrier & de Chambrier, 2010; Molnár, Ostoros, Dunams-Morel, & Rosenthal, 2012, Peichoto et al., 2016). In addition to the phyla common to all four samples, we find taxa specific to individual hosts. These include the bacterial phyla Actinobacteria, Chloroflexi, Tenericutes, Planctomycetes, Cyanobacteria, Synergistetes, Deinococcus-Thermus, Armatimonadetes, Thermotogae, Verrucomicrobia, Ignavibacteriae, Spirochaetes, Fibrobacteres, Acidobacteria, Fusobacteria, and the Archaea phylum Euryarchaeota. We also find the fungal taxa Ascomycota, Basidiomycota, and Entomophthoromycota. The first two samples also contain sequences that align with Cnidaria. These are likely Myxozoans, a branch of cnidarians that parasitize vertebrate guts (Foox & Siddall, 2015).

Our rarefaction curves show that most tissue datasets need to have at least 40,000 identified sequences to capture metagenomic communities (Fig. 4). Some of our samples (notably corn8, mod7, and

TC1003) fall far short of that threshold, while others are closer to it but still likely to have undetected metagenomic information. Raw data files use in this paper are available at https://doi.org/10.6084/m9.figshare.5593522.v1.

5.5 Discussion

We find that metagenomic sequences can be identified from a range of tissue types with variable preservation histories (Table 5.1). Our screened tissues included liver preserved in ethanol from a museum collection, lizard tail-tip tissues preserved in RNALater, lizard guts preserved in ethanol, and cloacal swabs preserved in ethanol. All preservation and sample types yield metagenomic sequences, indicating that most or all of the tissues currently preserved for DNA extraction in museum collections worldwide (Yeates, Zwick, & Mikheyev, 2016) are also repositories of metagenomic information. These repositories of metagenomic sequences can be analyzed using a range of approaches, including the RADseq exploratory techniques we present here and more conventional amplicon-based metagenomic profiling. Our exploratory approach shows that the majority of sequences generated by RADseq for our host tissues are not identifiably similar to any publicly available reference sequence. Less than 1% of the sequences that cluster across individual hosts are 97% similar to any NCBI GenBank sequence. Of those that do hit our similarity threshold, the vast majority are Chordate (host) sequences, when tissue samples are the source of DNA. Other DNA sources, such as cloacal swabs and intestinal rinses, have different taxonomic profiles. However, a number of sequences amplified from lizard tail and liver tissue align with phyla known to occur in the blood and tissue metagenome: Platyhelminthes (blood flukes and relatives), Nematoda (round worms), and Apicomplexa (malaria parasites and relatives). More that 50% of the metagenomic sequences identified in Phrynosoma were found in both species present at the site, indicating that the parasites are common within and between closely related host species.

The double digest RADseq approach worked across a range of sample types, including standard tissue samples commonly used for host genetic analysis (liver and muscle), and less conventional sources, such as cloacal swabs and rinses from preserved digestive tract. Any metagenomic source that can

produce the necessary 100–200 ng of DNA for the ddRAD protocol should be tractable for this type of analysis (Peterson et al., 2012). All tissue types produced large numbers of sequences that could not be matched to publicly available references at any similarity threshold (Fig. 5.1). As percent identity threshold levels increased, the number of sequences that matched to the host dropped quickly. There was no threshold that completely excluded sequences that matched common parasite or microbiome phyla, indicating that they were present in the extracted DNA with high confidence.

5.5.1 Limitations and caveats

The major limitation of our approach is that it relies on public databases to determine the taxonomic identity of sequences. However, public databases do not accurately reflect the diversity of metagenomic taxa. For example, we recover relatively few Archaea sequences from our hind-gut samples. We hypothesize that this reflects the relative lack of Archaea genetic sequences in GenBank to compare against, rather than an absence of Archaea from our samples. However, the number of publicly available, taxonomically identified reference sequences is quickly increasing, so this source of bias should be reduced in the future. Second, our identifications are based on randomly restricted DNA samples, rather than widely-accepted barcode sequences. These sequences can't be corrected for copy number variation, and we have little to no ability to determine whether two different sequences represent different individuals or whether they are two separate samples of the genome of a single individual. Due to the inherent stochasticity of the ddRAD approach, this method should not be used to quantify the relative or absolute abundance of metagenomic communities. Finally, the sequences in this paper are from an Illumina HiSeq platform. Negative controls are not recommended on this platform, as running one leads to uneven ratios in barcode sequences, which can damage the sequencing quality for the entire run. The lack of negative controls is one reason that this approach should be considered exploratory, rather than as a method for quantifying microbial load in specimens. Well-designed primer sets can account for all of these problems, and should be used to answer questions about relative abundances of metagenomic lineages or community structure.

We note some unexpected taxonomic identifications in our sequences. Specifically, we note that some of the sequences align with highest percent identity with arthropod sequences, or with Streptophyta. These sequences may represent sequencing error that alters a highly conserved host or metagenomic sequence to erroneously align more closely with a non-target sequence. Alternatively, sequences could be labelled incorrectly in the NCBI database, or they could be contamination in genome assemblies (Orosz, 2015). Regardless of the source, this taxonomic error indicates that our method should be used for exploratory purposes only.

5.5.2 Ecology of metagenomic sequences

All of the putative parasite taxa that we can identify with 97% certainty are known parasites of vertebrate hosts. *Protopolystoma xenopodis* is known from African clawed frogs, in which it attaches to the kidney and feeds on blood, thereby potentially releasing its own DNA into the host's bloodstream (Theunissen, Tiedt, & Du Preez, 2014). *Diphyllobothrium latum* is known from the digestive tracts of a range of vertebrates, including mammals and fish (Schurer et al., 2016; Wicht, Gustinelli, Fioravanti, Invernizzi, & Peduzzi, 2009). *Nippostrongylus brasilensis* and *Strongyloides stercoralis* are nematodes known from mammals. Their lifecycle begins with free-living juveniles that find a host and bore into the bloodstream through the skin. The juveniles migrate to the lungs, where they develop into adults before entering the digestive tract to breed (Haley, 1961; Koutz & Groves, 1953). *Soboliphyme baturini* is known from mammals, and infects the stomach (Zarnke, Whitman, Flynn, & Hoef, 2004). *Elaeophora elaphi* occurs in red deer, where it lives in the portal vein near the heart (Carrasco et al., 1995).While the sequences we detected are probably not the same species as their closest match, they should be closely related, and are likely to have similar life histories. All of the life histories here indicate that parasite DNA could plausibly be shed into the bloodstream.

We found many families of bacteria that are known from vertebrate guts, and some that have not previously been recorded. Actinobacteria, Acidobacteria, Cyanobacteria, Fusobacteria, Spirochaetes, Synergistetes, and Tenericutes have been reported from wild snake hindguts (Colston, Noonan, &

Jackson, 2015). Plancomycetes and Verrucomicrobia have been found in the guts of wild apes (Yildirim et al., 2010). Euryarchaeota, Deinococcus-Therums, Thermatogae, and Fibrobacteres have been found in dog gut microbiomes (Swanson et al., 2011). Chloroflexi has been recorded from human guts (Campbell et al., 2014), and Synergistetes has been recorded from the gut of young calves (R. W. Li, Connor, Li, Baldwin VI, & Sparks, 2012). To the best of our knowledge, the bacterial phyla Ignavibacteriae and Armatimonadetes, and the fungal phylum Entomophthoromycota have not previously been reported from vertebrate hind microbiomes. A species in Entomophthoromycota has been found in a cyst in the esophagus of a rat snake, *Elaphe obsoleta* (Dwyer et al., 2006). Other Entomophthoromycota are pathogens of invertebrates (Gryganskyi et al., 2012), indicating that they are capable of invading multicellular hosts. Ignavibacteriae is a sister phylum to Bacteroidetes and Chlorobi, both known from gut microbiomes (Podosokorskaya et al., 2013). The phylum has been sequenced from wastewater, indicating that it can survive in organic waste (Meng, Li, Wang, Ma, & Zhang, 2015). Armatimonadetes is primarily a soil phylum, but also participates in plant rhizobial communities (Tanaka et al., 2012), and has been found in mosquito salivary glands (Sharma et al., 2014) and decomposing swine manure (Tuan, Chang, Yu, & Huang, 2014). The three new phyla are all reasonable candidates for the gut microbiome, as they are known to occur within multicellular host tissues. However, caution should be exercised because both the known and novel taxa we identified from the gut samples can also be found in environmental samples. Further study, using carefully selected barcode loci, should be undertaken before these taxa are considered an established part of the gut microbiome.

5.6 Conclusion

Our results demonstrate the value of long-term storage of a variety of tissue types in publicly available collections. Techniques that have not yet been developed at the time of tissue collection may later become available, rendering the samples and their metadata (geographic locality, time of year collected, and other ecological data) highly relevant. Similarly, current publicly available short read

datasets may include as yet unrecognized metagenomic sequences. Investigators designing amplicon-based approaches to microbial communities in specific host tissue types could mine

5.7 Data Archiving

Data are available on the Figshare website, DOI https://doi.org/10.6084/m9.figshare.5593522.v1. Supplemental scripts can be found online at http://dx.doi.org/ 10.7717/peerj.4662#supplemental-information.

5.8 Acknowledgments

We thank Christian L. Cox and Robert M. Cox for providing samples, Michael Grundler for assistance in the field, and the Natural History Museum of Los Angeles County for providing tissues. Laboratory protocols were performed at the University of Michigan Genomic Diversity Laboratory.

5.9 Funding

Funding was provided by a startup grant from the University of Michigan to Alison R. Davis Rabosky and University of Michigan Department of Ecology and Evolutionary Biology and University of Michigan Museum of Zoology graduate student grants to Iris Holmes.

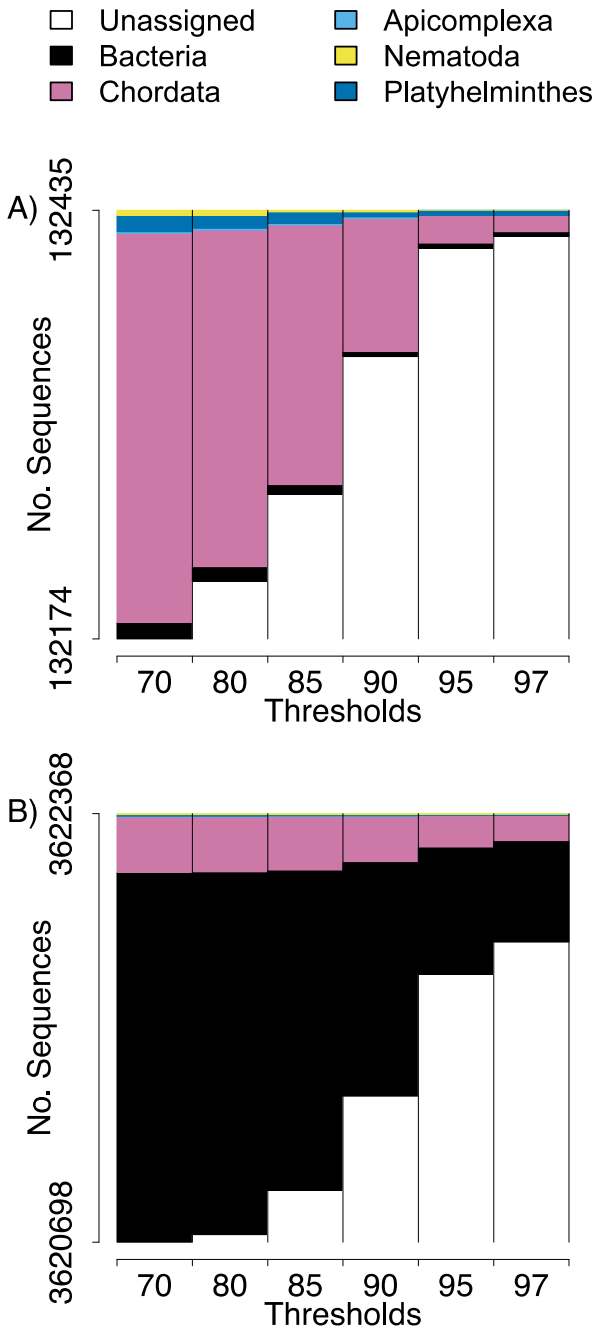


Figure 5.1 Numbers of parasite bycatch sequences at increasing thresholds of similarity to reference sequences.

Changes in the relative number of sequences that aligned to an identifiable phylum in the NCBI database at six thresholds for percent similarity between the sequence and its closest GenBank hit. (A) Is from Texas horned lizard, *Phrynosoma cornutum*, tail tissue, (B) is from a ribbon snake, *Thamnophis sauritius*, cloacal swab.

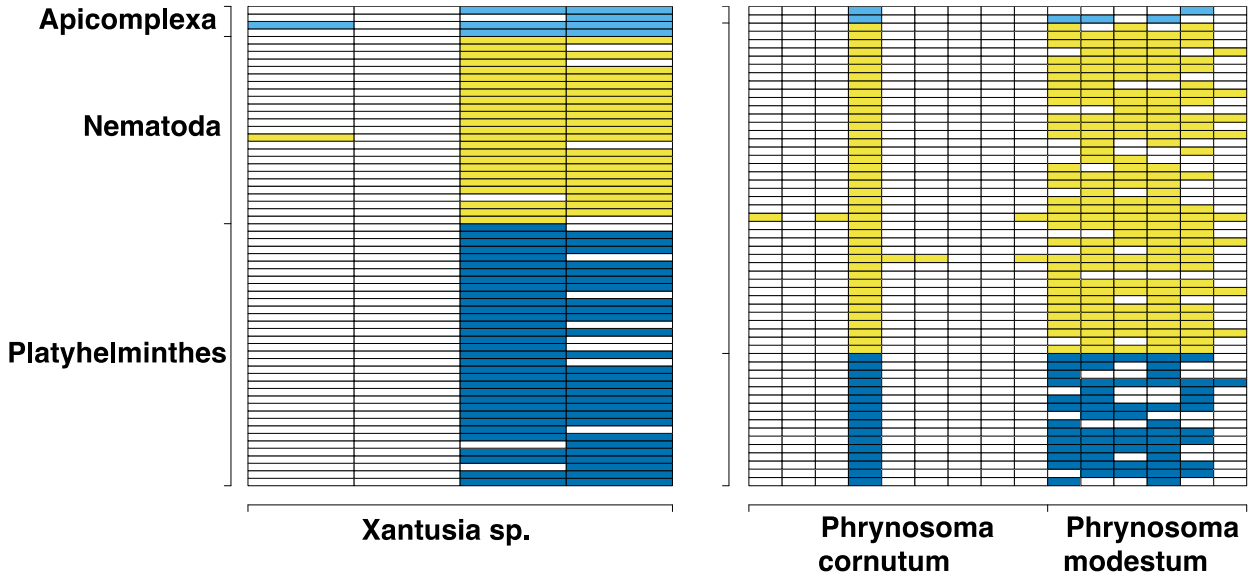


Figure 5.2 Identifiable metagenomic sequences between congeneric hosts

Data from each host is shown in columns; rows indicate distinct, identifiable metagenomic sequences. White sequences are absent from a given host, sequences with color were detected in a given host. For comparison purposes, we assign different colors to each parasite phylum.

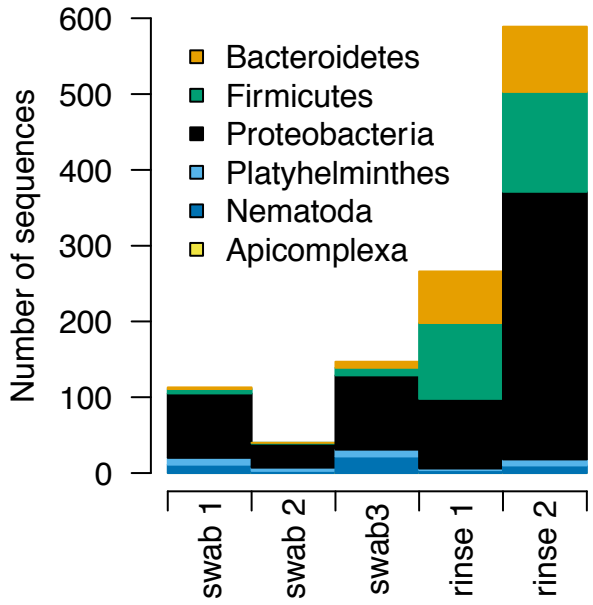


Figure 5.3 Numbers of gut microbiome phyla from two collections techniques

RADseq approaches amplify a range of identifiable bacterial and eukaryotic sequences from intestinal rinse (*Sceloporus jarrovi*) and cloacal swab (*Thamnophis sauritus* and *Nerodia sipedon*) sequences.

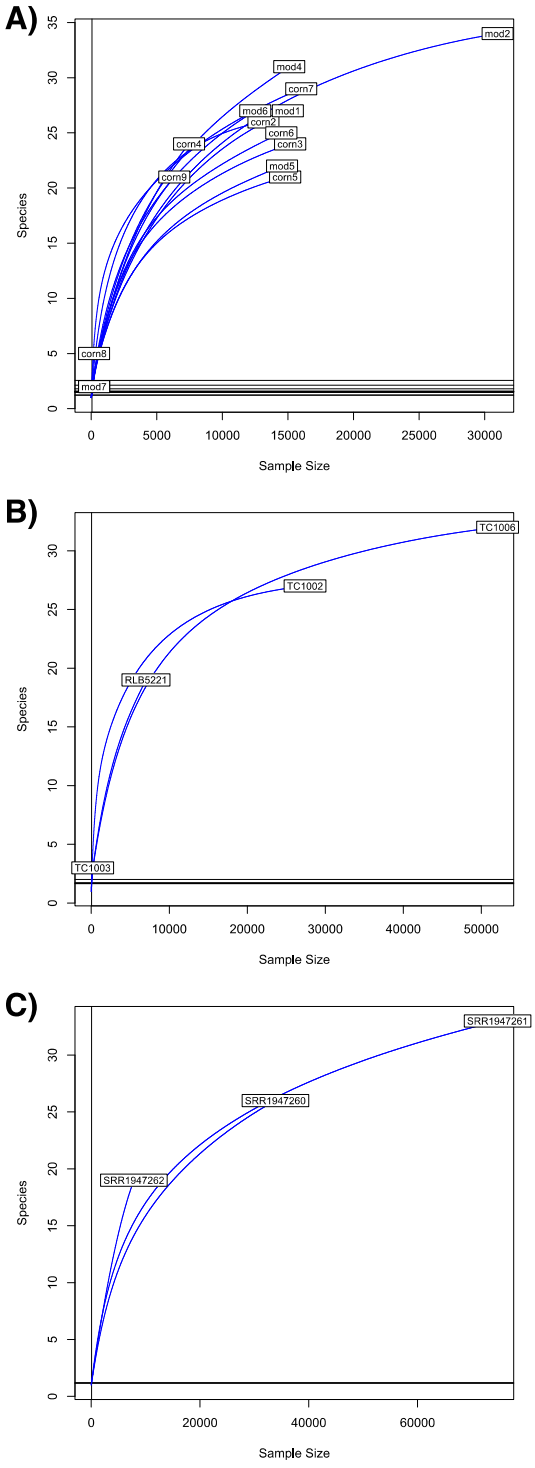


Figure 5.4 Rarefaction curves for Chordata and blood parasite metagenomic sequences

Rarefaction curves for genera of blood parasites and host sequences from the three genera of hosts: *Phrynosoma* (A), *Xantusia* (B), and *Micrurus* (C).

Table 5.1 Sequences aligned with 97% similarity to host or parasite templates

Sample	Total	Chordata	Platyhelminthes	Nematoda	Apicomplexa	species	sample type
corn1	379196	12	1	0	0	*Phrynosoma cornutum*	tail tissue
corn2	138897	533	61	49	0	*Phrynosoma cornutum*	tail tissue
corn3	846104	564	29	5	0	*Phrynosoma cornutum*	tail tissue
corn4	1918882	292	9	1	0	*Phrynosoma cornutum*	tail tissue
corn5	2591254	501	10	1	0	*Phrynosoma cornutum*	tail tissue
corn6	550893	543	31	7	0	*Phrynosoma cornutum*	tail tissue
corn7	57273	633	313	138	0	*Phrynosoma cornutum*	tail tissue
corn8	132434	10	0	3	0	*Phrynosoma cornutum*	tail tissue
corn9	927740	213	13	8	0	*Phrynosoma cornutum*	tail tissue
mod1	2052740	603	23	4	0	*Phyrnosoma modestum*	tail tissue
mod2	2423509	1363	46	11	0	*Phyrnosoma modestum*	tail tissue
mod4	3070606	564	12	4	0	*Phyrnosoma modestum*	tail tissue
mod5	1018036	622	28	11	0	*Phyrnosoma modestum*	tail tissue
mod6	298447	554	28	22	0	*Phyrnosoma modestum*	tail tissue
mod7	111733	10	1	0	0	*Phyrnosoma modestum*	tail tissue
RLB5221	342173	478	9	12	0	*Xantusia riversiana*	liver tissue
TC1002	91094	759	212	231	1	*Xantusia vigilis*	liver tissue

TC1003	2154606	6	1	0	0	*Xantusia vigilis*	liver tissue
TC1006	836745	1360	73	30	0	*Xantusia vigilis*	liver tissue
SRR1947260	350330	2966	1	3	0	*Micrurus fulvius*	Streicher et al. 2016
SRR1947261	296572	7343	5	2	0	*Micrurus fulvius*	Streicher et al. 2016
SRR1947262	323453	694	3	2	0	*Micrurus fulvius*	Streicher et al. 2016
ribbon1	1174345	3594	9	11	2	*Thamnophis sauritus*	cloacal swab
ribbon2	260079	1148	4	3	2	*Thamnophis sauritus*	cloacal swab
water	1772451	6762	9	22	2	*Nerodia sipedon*	cloacal swab
Sc0055	3622367	98	2	4	1	*Sceloporus jarrovi*	dissected gut
Sc0100	2715018	577	8	9	0	*Sceloporus jarrovi*	dissected gut

Preserved host muscle tissue also preserves genetic material from three major taxa of parasites, Platyhelminthes, Nematoda, and Apicomplexa. Only sequences that had 97% or greater similarity to a sequence in GenBank are included. Total number of sequences for each sample included for reference.

Chapter 6 Conclusion

In this dissertation, I examine three classes of mechanisms that reduce gene flow between natural populations. Taken together, my work points to the importance of understanding areas of low in gene flow rates in natural populations in their abiotic and biotic context. In addition, boundaries between genetically distinct populations should be understood as the result of a mix of stochastic and deterministic processes. My work will provide a framework for researchers who identify phylogeographic boundaries or patterns of allelic diversity across landscapes that are not well explained by either observable abiotic boundaries or purely neutral processes. In doing so, my dissertation will advance the project of understanding how diversity is generated in natural populations, as isolation is a key first step in populations evolving phenotypic innovations.

The first class of barriers I examine is comprised of discrete, abiotic features that impose strong physiological costs on the organisms attempting to disperse across them. Both in this dissertation and in the larger literature, this class of barriers is most often identified by mapping phylogeographic demes onto their habitat, and noting the geographic features that consistently align with breaks between demes. This practice can be traced back to the paper that first coined the term 'phylogeography' (J C Avise et al., 1987). While both the observational approach and more quantitative approaches such as cost-path modeling (Wang, Savage, & Bradley Shaffer, 2009) have greatly increased our understanding of how organisms move through their environment, I propose that organism's physiology and performance be incorporated directly into our understanding of abiotic barriers. Recent work has identified adaptation toward a 'dispersal phenotype' at the moving front of species expanding their ranges (Hudson, McCurry, Lundgren, McHenry, & Shine, 2016). If individual organisms' phenotypes can predict movement across

the landscape, they should be similarly able to predict inability to move. By more explicitly incorporating physiology into studies of phylogeographic patterns, the field will be able to predict barriers to gene flow on a landscape scale.

A second category of barrier occurs through biotic interactions that prevent or reduce dispersal. These barriers arise when organisms are physiologically capable of dispersing to a new population, but dispersal is reduced either through direct ecological interactions or through behavioral adaptations. For example, a population of organisms might experience high predation when in open habitats. The predation itself could restrict gene flow between populations that are not connected by adequate cover. Over evolutionary time, the surviving members of the population will be those organisms with a strong reluctance to leave cover. Then, even if predation pressure changes, behavioral syndromes could maintain isolation between populations (Laurance et al., 2004). Phylogeography contextualized by the focal species' biotic and abiotic environment can provide circumstantial evidence for this type of barrier. If the demes align with a given habitat feature in the presence of one or more other species, but do not when those species are absent, there is evidence that this second category of mechanism may be in play. A true test differentiating the first type of barrier from the second would include both information from interspecific interactions, and physiological and performance information from the species of interest.

Gene flow between populations can also be reduced even when individuals can successfully disperse between locations. True gene flow requires the dispersing individuals to successfully leave offspring in their new population. In some situations, for example in ring species, dispersing animals cannot or will not breed with the conspecifics in their new location (Irwin, 2002; Moritz, Schneider, & Wake, 1992). I exclude these situations from the third category of barriers to gene flow, because they involve processes of incipient speciation. Instead, I focus on interspecific interactions that prevent a dispersing individual from surviving long enough to breed in their new location. The classic example of this type of barrier comes from the well-studied *Heliconius* butterfly sytem. Gene flow within species but between mimicry complexes is very low, because dispersing butterflies that do not match the local morph are quickly eaten by predators (Kapan, 2001). There are some similar examples of parasite faunas that

exclude dispersing hosts from populations in which the parasite is absent (Q.-G. Zhang & Buckling, 2016). A large number of locally coadapted complexes of host and parasite could significantly reduce gene flow between populations across both host's and parasite's range. Identifying the third type of barrier in wild populations requires first showing that migration can occur, and then demonstrating that migrants fail to leave offspring in their new location. Determining the mechanisms behind high migrant mortality will require community-scale identification of locally coadapted predatory, parasitic, and competitive relationships.

Finally, all categories of barrier I discuss in this dissertation are highly contextual. Interspecific interactions that reduce gene flow, and even abiotic barriers to dispersal, can change with climatic shifts or other alterations in habitat. The abundance of examples of populations fragmented in the Pleistocene that experience secondary contact during the Holocene attest to the contextual nature of abiotic boundaries (Arbogast, Browne, & Weigl, 2001; Canestrelli, Cimmaruta, & Nascetti, 2008; K. L. Jones, Krapu, Brandt, & Ashley, 2005). Similarly, a population might in the future evolve the ability to navigate the biotic or abiotic feature that restricts gene flow in the present.

At any given time, a range of deterministic and stochastic processes shape the observable patterns of gene flow in natural populations. Identifying the causal agents in reduction of gene flow between populations might be best approached in a hypothesis-testing framework, in which potential mechanisms are sequentially ruled out rather than identified. The hypothesis-testing framework could be particularly useful in instances in which multiple interacting factors are reducing gene flow. None of the barrier categories I identify are mutually exclusive, a situation that lends itself to hypothesis testing.

In my dissertation, I discuss a range of mechanisms, both stochastic and deterministic, that result in patterns of deme boundaries on the landscape. I propose that incorporating data on the genetic and physiological characteristics of the focal organisms can improve our understanding of the mechanisms behind reduction in gene flow. In addition, I propose that incorporating information from the biotic communities in which the focal species occurs can help to identify cases in which gene flow is reduced due to biotic interactions. Finally, using a hypothesis testing framework to sequentially exclude possible

mechanisms for reduced gene flow, rather that identifying them, could help to incorporate the role of both stochasticity and complex, contingent mechanisms creating in phylogeographic pattern.

Bibliography

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antonovics, J. (1976). The Input from Population Genetics: "The New Ecological Genetics." *Systematic Botany*, *1*(3), 233. doi: 10.2307/2418718

Arbogast, B. S., Browne, R. A., & Weigl, P. D. (2001). Evolutionary genetics and Pleistocene biogeography of North American tree squirrels (*Tamiasciurus*). *Journal of Mammalogy*, *82*(2), 18.

Argus, D. F., & Gordon, R. G. (2001). Present tectonic motion across the Coast Ranges and San Andreas fault system in central California. *Geological Society of America Bulletin*, 13.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, *22*(11), 3179–3190. doi: 10.1111/mec.12276

Aubier, T. G., & Sherratt, T. N. (2015). Diversity in Müllerian mimicry: The optimal predator sampling strategy explains both local and regional polymorphism in prey. *Evolution*, *69*(11), 2831–2845. doi: 10.1111/evo.12790

Avise, J C, Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., … Saunders, N. C. (1987). Intraspecific phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, *18*, 489–522.

Avise, John C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15. doi: 10.1111/j.1365-2699.2008.02032.x

Bantock, C. R., & Ratsey, M. (1980). Natural selection in experimental populations of the land snail *Cepaea nemoralis* (L.). *Heredity*, *44*(1), 37–54. doi: 10.1038/hdy.1980.3

Barber, P., Palumbi, S., Erdmann, M., & Moosa, M. (2002). Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, *11*, 659–674.

Battey, C. J., & Klicka, J. (2017). Cryptic speciation and gene flow in a migratory songbird Species Complex: Insights from the Red-Eyed Vireo (*Vireo olivaceus*). *Molecular Phylogenetics and Evolution*, *113*, 67–75. doi: 10.1016/j.ympev.2017.05.006

Bechtel, H. B. (1978). Color and Pattern in Snakes (Reptilia, Serpentes). *Journal of Herpetology*, *12*(4), 521. doi: 10.2307/1563357

Bechtel, H., & Bechtel, E. (1962). Heredity of albinism in the corn snake, *Elaphe guttata guttata*, demonstrated in captive breedings. *Copeia*, 436–437.

Bezy, R., Gorman, G., Adest, G., & Kim, Y. (1980). Divergence in the island night lizard *Xantusia riversiana* (Sauria: Xantusiidae). In *The California Islands: Proceedings of a Multidisciplinary Symposium* (pp. 565–583). Santa Barbara, CA: Haagen Printing.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bonaccorso, E., Koch, I., & Peterson, A. T. (2006). Pleistocene fragmentation of Amazon species' ranges. *Diversity and Distributions*, *12*(2), 157–164. doi: 10.1111/j.1366-9516.2005.00212.x

Bonansea, M. I., & Vaira, M. (2012). Geographic and intrapopulational variation in colour and patterns of an aposematic toad, *Melanophryniscus rubriventris* (Amphibia, Anura, Bufonidae). *Amphibia-Reptilia*, *33*(1), 11–24. doi: 10.1163/156853811X619754

Borer, M., Van Noort, T., Rahier, M., & Naisbit, R. E. (2010). Positive frequency-dependent selection of warning color in alpine leaf beetles. *Evolution*, *64*(12), 3629–3633. doi: 10.1111/j.1558-5646.2010.01137.x

Brothers, D. S., Miller, N. C., Barrie, J. V., Haeussler, P. J., Greene, H. G., Andrews, B. D., … Dartnell, P. (2020). Plate boundary localization, slip-rates and rupture segmentation of the Queen Charlotte Fault based on submarine tectonic geomorphology. *Earth and Planetary Science Letters*, *530*, 115882. doi: 10.1016/j.epsl.2019.115882

Brown, K. S., & Benson, W. W. (1974). Adaptive Polymorphism Associated with Multiple Mullerian Mimicry in *Heliconius numata* (Lepid. Nymph.). *Biotropica*, *6*(4), 205. doi: 10.2307/2989666

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., … Lawley, T. D. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature*, *533*(7604), 543–546. doi: 10.1038/nature17645

Burgess, N. D., Butynski, T. M., Cordeiro, N. J., Doggart, N. H., Fjeldså, J., Howell, K. M., … Stuart, S. N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, *134*(2), 209–231. doi: 10.1016/j.biocon.2006.08.015

Byrne, M. (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, *27*(27–28), 2576–2585. doi: 10.1016/j.quascirev.2008.08.032

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. doi: 10.1186/1471-2105-10-421

Campbell, A. G., Schwientek, P., Vishnivetskaya, T., Woyke, T., Levy, S., Beall, C. J., … Podar, M. (2014). Diversity and genomic insights into the uncultured *Chloroflexi* from the human microbiota: Uncultured human-associated *Chloroflexi*. *Environmental Microbiology*, *16*(9), 2635–2643. doi: 10.1111/1462-2920.12461

Canestrelli, D., Cimmaruta, R., & Nascetti, G. (2008). Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica* - insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, *17*(17), 3856–3872. doi: 10.1111/j.1365-294X.2008.03870.x

Carrasco, L., Fierro, Y., Sanchezcastillejo, J., Bautista, M., Gomezvillamandos, J., & Sierra, M. (1995). Elaeophorosis in red deer caused by *Elaeophora elaphi* - lesions of natural disease. *Veterniary Pathology*, *32*(3), 250–257.

Chai, P. (1986). Field observations and feeding experiments on the responses of rufous-tailed jacamars (*Galbula ruficauda*) to free-flying butterflies in a tropical rainforest. *Biological Journal of the Linnean Society*, *29*(3), 161–189. doi: 10.1111/j.1095-8312.1986.tb01772.x

Chamberlain, S., & Szocs, E. (2013). taxize - taxonomic search and retrieval in R. *F1000Research*, *2*(191). Retrieved from http://f1000research.com/articles/2-191/v2

Chamberlain, S., Szocs, E., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., … O'Donnell, J. (2016). taxize: Taxonomic information from around the web (Version 0.7.8). Retrieved from https://github.com/ropensci/taxize

Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The Effects of Genetic and Geographic Structure on Neutral Variation. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 99–125. doi: 10.1146/annurev.ecolsys.34.011802.132359

Charlesworth, D. (2016). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications*, *9*(1), 74–90. doi: 10.1111/eva.12291

Chatzimanolis, S., & Caterino, M. S. (2007). Toward a Better Understanding of the "Transverse Range Break": Lineage Diversification in Southern California. *Evolution*, *61*(9), 2127–2141. doi: doi: 10.1111 /j.1558-5646

Chong, Z., Ruan, J., & Wu, C.-I. (2012). Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, *28*(21), 2732–2737. doi: 10.1093/bioinformatics/bts482

Chouteau, M., Arias, M., & Joron, M. (2016). Warning signals are under positive frequency-dependent selection in nature. *Proceedings of the National Academy of Sciences*, *113*(8), 2164–2169. doi: 10.1073/pnas.1519216113

Clegg, M. T., & Durbin, M. L. (2000). Flower color variation: A model for the experimental study of evolution. *Proceedings of the National Academy of Sciences*, *97*(13), 7016–7023. doi: 10.1073/pnas.97.13.7016

Clooney, A. G., Fouhy, F., Sleator, R. D., O'Driscoll, A., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. *PLoS One*, *11*(2), e0148028.

Coblentz, B. E. (1978). The effects of feral goats (*Capra hircus*) on island ecosystems. *Biological Conservation*, *13*(4), 279–286. doi: 10.1016/0006-3207(78)90038-1

Colston, T. J., & Jackson, C. R. (2016). Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Molecular Ecology*, *25*(16), 3776–3800. doi: 10.1111/mec.13730

Colston, T. J., Noonan, B. P., & Jackson, C. R. (2015). Phylogenetic Analysis of Bacterial Communities in Different Regions of the Gastrointestinal Tract of *Agkistrodon piscivorus*, the Cottonmouth Snake. *PLOS ONE*, *10*(6), e0128793. doi: 10.1371/journal.pone.0128793

Comeault, A. A., & Noonan, B. P. (2011). Spatial variation in the fitness of divergent aposematic phenotypes of the poison frog, *Dendrobates tinctorius*: Spatial variation in fitness. *Journal of Evolutionary Biology*, *24*(6), 1374–1379. doi: 10.1111/j.1420-9101.2011.02258.x

Cook, L M. (1967). The genetics of *Cepaea nemoralis*. *Heredity*, *22*(3), 397–410. doi: 10.1038/hdy.1967.49

Cook, L. M. (1998). A two–stage model for *Cepaea* polymorphism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1375), 1577–1593. doi: 10.1098/rstb.1998.0311

Cook, L. M. (2008). Variation with habitat in *Cepaea nemoralis*: the Cain & Sheppard diagram. *Journal of Molluscan Studies*, *74*(3), 239–243. doi: 10.1093/mollus/eyn011

Cook, L. M. (2013). Selection and disequilibrium in *Cepaea nemoralis*. *Biological Journal of the Linnean Society*, *108*(3), 484–493. doi: 10.1111/j.1095-8312.2012.02027.x

Corl, A., Davis, A. R., Kuchta, S. R., & Sinervo, B. (2010). Selective loss of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation.

Proceedings of the National Academy of Sciences, *107*(9), 4254–4259. doi: 10.1073/pnas.0909480107

Courchamp, F., Chapuis, J.-L., & Pascal, M. (2003). Mammal invaders on islands: impact, control and control impact. *Biological Reviews*, *78*(3), 347–383. doi: 10.1017/S1464793102006061

Cox, C. L., & Chippindale, P. T. (2014). Patterns of genetic diversity in the polymorphic ground snake (*Sonora semiannulata*). *Genetica*, *142*(4), 361–370. doi: 10.1007/s10709-014-9780-7

Cox, C. L., & Davis Rabosky, A. R. (2013). Spatial and Temporal Drivers of Phenotypic Diversity in Polymorphic Snakes. *The American Naturalist*, *182*(2), E40–E57. doi: 10.1086/670988

Crouch, J. K. (1979). Neogene tectonic evolution of the California Continental Borderland and western Transverse Ranges. *Geological Society of America Bulletin Part 1*, *90*, 338–345.

Culver, M., Hedrick, P. W., Murphy, K., O'Brien, S., & Hornocker, M. G. (2008). Estimation of the bottleneck size in Florida panthers. *Animal Conservation*, *11*(2), 104–110. doi: 10.1111/j.1469-1795.2007.00154.x

Davies, P. W., & Snow, D. W. (1965). Territory and food of the Song Thrush. *British Birds*, *58*(5), 161–175.

Davis, A., Corl, A., Surget-Groba, Y., & Sinervo, B. (2011). Convergent evolution of kin-based sociality in a lizard. *Proceedings of the Royal Society B*, *278*, 1508–1514.

Davis, A. R. (2012). Kin presence drives philopatry and social aggregation in juvenile Desert Night Lizards (*Xantusia vigilis*). *Behavioral Ecology*, *23*(1), 18–24. doi: 10.1093/beheco/arr144

Davis, Alison R., & Leavitt, D. H. (2007). Candlelight vigilis: A noninvasive method for sexing small, sexually monomorphic lizards. *Herpetological Review*, *38*(4), 402–404.

Davis Rabosky, A. R., Corl, A., Liwanag, H. E. M., Surget-Groba, Y., & Sinervo, B. (2012). Direct Fitness Correlates and Thermal Consequences of Facultative Aggregation in a Desert Lizard. *PLoS ONE*, *7*(7), e40866. doi: 10.1371/journal.pone.0040866

Davis Rabosky, A. R., Cox, C. L., & Rabosky, D. L. (2016). Unlinked Mendelian inheritance of red and black pigmentation in snakes: Implications for Batesian mimicry. *Evolution*, *70*(4), 944–953. doi: 10.1111/evo.12902

Davis Rabosky, A. R., Cox, C. L., Rabosky, D. L., Title, P. O., Holmes, I. A., Feldman, A., & McGuire, J. A. (2016). Coral snakes predict the evolution of mimicry across New World snakes. *Nature Communications*, *7*(1). doi: 10.1038/ncomms11484

de Chambrier, S., & de Chambrier, A. (2010). Two new genera and two new species of proteocephalidean tapeworms (*Eucestoda*) from reptiles and amphibians in Australia. *Folia Parasitologica*, *57*(4), 263.

DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources*, *13*(3), 377–392. doi: 10.1111/1755-0998.12071

DeLong, S. B., Minor, S. A., & Arnold, L. J. (2007). Late Quaternary alluviation and offset along the eastern Big Pine fault, southern California. *Geomorphology*, *90*(1–2), 1–10. doi: 10.1016/j.geomorph.2007.01.018

Demos, T. C., Kerbis Peterhans, J. C., Agwanda, B., & Hickerson, M. J. (2014). Uncovering cryptic diversity and refugial persistence among small mammal lineages across the

Eastern Afromontane biodiversity hotspot. *Molecular Phylogenetics and Evolution*, *71*, 41–54. doi: 10.1016/j.ympev.2013.10.014

Dollive, S., Chen, Y.-Y., Grunberg, S., Bittinger, K., Hoffmann, C., Vandivier, L., … Bushman, F. D. (2013). Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment. *PLoS ONE*, *8*(8), e71806. doi: 10.1371/journal.pone.0071806

Durand, E., Jay, F., Gaggiotti, O. E., & Francois, O. (2009). Spatial Inference of Admixture Proportions and Secondary Contact Zones. *Molecular Biology and Evolution*, *26*(9), 1963–1973. doi: 10.1093/molbev/msp106

Dwyer, J., Burwell, B., Humber, R., Mcleod, C., Fleetwood, M., & Johnson, T. (2006). *Schizangiella serpentis* infection in a Virginia ratsnake (*Elaphe obsoleta*). *Veterniary Pathology*, *43*(5), 819.

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. doi: 10.1093/bioinformatics/btu121

Eizaguirre, C., & Baltazar-Soares, M. (2014). Evolutionary conservation-evaluating the adaptive potential of species. *Evolutionary Applications*, *7*(9), 963–967. doi: 10.1111/eva.12227

Endler, J. A. (1980). Natural selection on color patterns in *Poecilia reticulata*. *Evolution*, *34*(1), 76–91. doi: 10.1111/j.1558-5646.1980.tb04790.x

Endler, J. A., & Rojas, B. (2009). The Spatial Pattern of Natural Selection When Selection Depends on Experience. *The American Naturalist*, *173*(3), E62–E78. doi: 10.1086/596528

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., … Widmer, A. (2017). Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. *BMC Genomics*, *18*(1). doi: 10.1186/s12864-016-3459-7

Fisher, R. A., & Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula L*. *Heredity*, *1*(2), 143–174. doi: 10.1038/hdy.1947.11

Fontenot, B. E., Makowsky, R., & Chippindale, P. T. (2011). Nuclear–mitochondrial discordance and gene flow in a recent radiation of toads. *Molecular Phylogenetics and Evolution*, *59*(1), 66–80. doi: 10.1016/j.ympev.2010.12.018

Foox, J., & Siddall, M. E. (2015). The Road To Cnidaria: History of Phylogeny of the Myxozoa. *Journal of Parasitology*, *101*(3), 269–274. doi: 10.1645/14-671.1

Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610–2618. doi: 10.1111/mec.13139

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the Probability of Outbreeding Depression: Predicting Outbreeding Depression. *Conservation Biology*, *25*(3), 465–475. doi: 10.1111/j.1523-1739.2011.01662.x

Franks, D. W., & Oxford, G. S. (2009). The evolution of exuberant visible polymorphisms. *Evolution*, *63*(10), 2697–2706. doi: 10.1111/j.1558-5646.2009.00748.x

Gärke, C., Ytournel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., & Simianer, H. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken

populations: Population differentiation: SNPs vs SSRs. *Animal Genetics*, *43*(4), 419–428. doi: 10.1111/j.1365-2052.2011.02284.x

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, *10*(2), 305–318. doi: 10.1046/j.1365-294x.2001.01190.x

Gigord, L. D. B., Macnair, M. R., & Smithson, A. (2001). Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina (L.)* Soo. *Proceedings of the National Academy of Sciences*, *98*(11), 6253–6255. doi: 10.1073/pnas.111162598

Gompert, Z., Willmott, K., & Elias, M. (2011). Heterogeneity in predator micro-habitat use and the maintenance of Müllerian mimetic diversity. *Journal of Theoretical Biology*, *281*(1), 39–46. doi: 10.1016/j.jtbi.2011.04.024

Gosden, T. P., Stoks, R., & Svensson, E. I. (2011). Range limits, large-scale biogeographic variation, and localized evolutionary dynamics in a polymorphic damselfly. *Biological Journal of the Linnean Society*, *102*(4), 775–785. doi: 10.1111/j.1095-8312.2011.01619.x

Gottscho, A. D. (2016). Zoogeography of the San Andreas Fault system: Great Pacific Fracture Zones correspond with spatially concordant phylogeographic boundaries in western North America: Zoogeography of the San Andreas Fault system. *Biological Reviews*, *91*(1), 235–254. doi: 10.1111/brv.12167

Gow, J. L., Johansson, H., Surget-Groba, Y., & Thorpe, R. S. (2006). Ten polymorphic tetranucleotide microsatellite markers isolated from the *Anolis roquet* series of Caribbean lizards. *Molecular Ecology Notes*, *6*(3), 873–876. doi: 10.1111/j.1471-8286.2006.01382.x

Greene, H. W., & McDiarmid, R. W. (1981). Coral Snake Mimicry: Does It Occur? *Science*, *213*(45), 1207–1212.

Greenwood, J. J. D. (1974). Visual and other selection in *Cepaea*: A further example. *Heredity*, *33*(1), 17–31. doi: 10.1038/hdy.1974.61

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2013). Genetic drift outweighs natural selection at toll-like receptor (*TLR*) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology*, *22*(17), 4470–4482. doi: 10.1111/mec.12404

Gryganskyi, A. P., Humber, R. A., Smith, M. E., Miadlikovska, J., Steven Wu, Voigt, K., … Vilgalys, R. (2012). Molecular phylogeny of the Entomophthoromycota. *Molecular Phylogenetics and Evolution*, *65*(2), 682–694. doi: 10.1016/j.ympev.2012.07.026

Guo, S. W., & Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, *48*(2), 361. doi: 10.2307/2532296

Haasl, R. J., & Payseur, B. A. (2010). The Number of Alleles at a Microsatellite Defines the Allele Frequency Spectrum and Facilitates Fast Accurate Estimation of. *Molecular Biology and Evolution*, *27*(12), 2702–2715. doi: 10.1093/molbev/msq164

Haldane, J. (1924). A mathematical theory of natural and artificial selection: Part IV. Isolation. *Transactions of the Cambridge Philosophical Society*, *26*, 220–230.

Haley, A. J. (1961). Biology of the Rat Nematode *Nippostrongylus brasiliensis* (Travassos, 1914). I. Systematics, Hosts and Geographic Distribution. *The Journal of Parasitology*, *47*(5), 727. doi: 10.2307/3275460

Hallatschek, O., Hersen, P., Ramanathan, S., & Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, *104*(50), 19926–19930. doi: 10.1073/pnas.0710150104

Hameed, I. H., Ommer, A. J., Murad, A. F., & Mohammed, G. J. (2015). Allele frequency data of 21 autosomal short tandem repeat loci in Mesan and Basra provinces in South Iraq. *Egyptian Journal of Forensic Sciences*, *5*(4), 150–156. doi: 10.1016/j.ejfs.2014.10.003

Hammers, M., & Hans Van Gossum. (2008). Variation in female morph frequencies and mating frequencies: random, frequency-dependent harassment or male mimicry? *Animal Behaviour*, *76*(4), 1403–1410. doi: 10.1016/j.anbehav.2008.06.021

Harpak, A., Bhaskar, A., & Pritchard, J. K. (2016). Mutation Rate Variation is a Primary Determinant of the Distribution of Allele Frequencies in Humans. *PLOS Genetics*, *12*(12), e1006489. doi: 10.1371/journal.pgen.1006489

Hastie, T. (2015). gam: Generalized Additive Models (Version 1.12). Retrieved from http://CRAN.R-project.org/ package=gam

Hauksson, E., Kanamori, H., Stock, J., Cormier, M.-H., & Legg, M. (2014). Active Pacific North America Plate boundary tectonics as evidenced by seismicity in the oceanic lithosphere offshore Baja California, Mexico. *Geophysical Journal International*, *196*(3), 1619–1630. doi: 10.1093/gji/ggt467

Hedrick, P. W., & Fredrickson, R. (2010). Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics*, *11*(2), 615–626. doi: 10.1007/s10592-009-9999-5

Hegna, R. H., Nokelainen, O., Hegna, J. R., & Mappes, J. (2013). To quiver or to shiver: increased melanization benefits thermoregulation, but reduces warning signal efficacy in the wood tiger moth. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755), 20122812. doi: 10.1098/rspb.2012.2812

Hijmans, R. (2015). raster: Geographic data analysis and modeling (Version 2.5-2). Retrieved from http://CRAN.R-project. org/package=raster

Hill, K. (2003). Measure of Biodiversity: Richness, Rarity, and Endemism (Reptiles). In *Atlas of the Biodiversity of California* (pp. 30–31). California Department of Fish and Game.

Hoelzel, A. R. (1999). Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society*, *68*(1–2), 23–39. doi: 10.1111/j.1095-8312.1999.tb01156.x

Holland, R. F. (1986). *Preliminary Descriptions of the Terrestrial Natural Communities of California* [The Resources Agency Non-game Heritage Program]. Sacaramento CA: Department of Fish and Game, State of California.

Horton, B. P., Rahmstorf, S., Engelhart, S. E., & Kemp, A. C. (2014). Expert assessment of sea-level rise by AD 2100 and AD 2300. *Quaternary Science Reviews*, *84*, 1–6. doi: 10.1016/j.quascirev.2013.11.002

Hosoi, J., Danhara, T., Iwano, H., Matsubara, N., Amano, K., & Hirata, T. (2019). Development of the Tanakura strike-slip basin in Japan during the opening of the Sea of Japan: Constraints from zircon U–Pb and fission-track ages. *Journal of Asian Earth Sciences*, 104157. doi: 10.1016/j.jseaes.2019.104157

Houston, A. I., Stevens, M., & Cuthill, I. C. (2007). Animal camouflage: compromise or specialize in a 2 patch-type environment? *Behavioral Ecology*, *18*(4), 769–775. doi: 10.1093/beheco/arm039

Huang, W., Haubold, B., Hauert, C., & Traulsen, A. (2012). Emergence of stable polymorphisms driven by evolutionary games between mutants. *Nature Communications*, *3*(1), 919. doi: 10.1038/ncomms1930

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., … Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, *114*(5), 515–524. doi: 10.1038/hdy.2015.22

Hudson, C. M., McCurry, M. R., Lundgren, P., McHenry, C. R., & Shine, R. (2016). Constructing an Invasion Machine: The Rapid Evolution of a Dispersal-Enhancing Phenotype During the Cane Toad Invasion of Australia. *PLOS ONE*, *11*(9), e0156950. doi: 10.1371/journal.pone.0156950

Hugall, A. F., & Stuart-Fox, D. (2012). Accelerated speciation in colour-polymorphic birds. *Nature*, *485*(7400), 631–634. doi: 10.1038/nature11050

Hughes, M., Hall, A., & Fovell, R. G. (2009). Blocking in Areas of Complex Topography, and Its Influence on Rainfall Distribution. *Journal of the Atmospheric Sciences*, *66*(2), 508–518. doi: 10.1175/2008JAS2689.1

Irwin, D. E. (2002). Phylogeographic Breaks without Geographic Barriers to Gene Flow. *Evolution*, *56*(12), 2383–2394.

Jensen, J. L., Bohonak, A. J., & Kelley, S. T. (2005). Isolation by distance, web service. *BMC Genetics*, *6*(13). doi: 10.1186/1471-2156-6-13

Jolly, M. T., Jollivet, D., Gentil, F., Thiébaut, E., & Viard, F. (2005). Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the North coast of France. *Heredity*, *94*(1), 23–32. doi: 10.1038/sj.hdy.6800543

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Jones, J. S., Leith, B. H., & Rawlings, P. (1977). Polymorphism in *Cepaea*: A Problem with Too Many Solutions? *Annual Review of Ecology and Systematics*, *8*(1), 109–143. doi: 10.1146/annurev.es.08.110177.000545

Jones, K. L., Krapu, G. L., Brandt, D. A., & Ashley, M. V. (2005). Population genetic structure in migratory sandhill cranes and the role of Pleistocene glaciations. *Molecular Ecology*, *14*(9), 2645–2657. doi: 10.1111/j.1365-294X.2005.02622.x

Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., … Jiggins, C. D. (2006a). A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *PLoS Biology*, *4*(10), e303. doi: 10.1371/journal.pbio.0040303

Kapan, D. D. (2001). Three-butterfly system provides a field test of Muellerian mimicry. *Nature*, *409*, 338–340.

Kappes, P. J., & Jones, H. P. (2014). Integrating seabird restoration and mammal eradication programs on islands to maximize conservation gains. *Biodiversity and Conservation*, *23*(2), 503–509. doi: 10.1007/s10531-013-0608-z

Keegan, D. R., Coblentz, B. E., & Winchell, C. S. (1994). Feral Goat Eradication on San Clemente Island, California. *Wildlife Society Bulletin*, *22*(1), 56–61.

Keenan, K., McGinnity, P., Cross, T., Crozier, W., & Prodöhl, P. (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. doi: 10.1111/2041-210X.12067

Keller, M., Kollmann, J., & Edwards, P. J. (2000). Genetic introgression from distant provenances reduces fitness in local weed populations. *Journal of Applied Ecology*, *37*, 647–659.

Kikuchi, D. W., & Sherratt, T. N. (2015). Costs of Learning and the Evolution of Mimetic Signals. *The American Naturalist*, *186*(3), 321–332. doi: 10.1086/682371

Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S., & Jorde, L. B. (1998). Signatures of Population Expansion in Microsatellite Repeat Data. *Genetics*, *148*, 1921–1930.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution*, *23*(3), 482–490. doi: 10.1093/molbev/msj057

Knowles, L. L. (2009). Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 593–612.

Kochert, M., & Kucera, T. (1979). *Snake River Birds of Prey special research report* [USDI Report]. Boise, ID: BLM.

Koutz, E., & Groves, H. (1953). *Strongyloides stercoralis* from a dog in Ohio. *Journal of the American Veterinary Medical Association*, *122*(912), 211–213.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-Index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. doi: 10.1128/AEM.01043-13

Kronenberger, J. A., Funk, W. C., Smith, J. W., Fitzpatrick, S. W., Angeloni, L. M., Broder, E. D., & Ruell, E. W. (2017). Testing the demographic effects of divergent immigrants on small populations of Trinidadian guppies. *Animal Conservation*, *20*(1), 3–11. doi: 10.1111/acv.12286

Kuhner, M. K. (2006). LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, *22*(6), 768–770. doi: 10.1093/bioinformatics/btk051

Lancaster, L. T., & Kay, K. M. (2013). Origin and diversification of the California flora: re-examining hypotheses with molecular phylogenies. *Evolution*, *67*(4), 1041–1054. doi: 10.1111/evo.12016

Langham, G. M. (2004). Specialized Avian Predators Repeatedly Attack Novel Color Morphs of *Heliconius* Butterflies. *Evolution*, *58*(12), 2783–2787.

Langkilde, T., & Shine, R. (2006). How much stress do researchers inflict on their study animals? A case study using a scincid lizard, *Eulamprus heatwolei*. *Journal of Experimental Biology*, *209*(6), 1035–1043.

Latta, R. G., & Mitton, J. B. (1997). A Comparison of Population Differentiation Across Four Classes of Gene Marker in Limber Pine (*Pinus flexilis* James). *Genetics*, *146*, 1153–1163.

Laurance, S. G. W., Stouffer, P. C., & Laurance, W. F. (2004). Effects of Road Clearings on Movement Patterns of Understory Rainforest Birds in Central Amazonia. *Conservation Biology*, *18*(4), 1099–1109. doi: 10.1111/j.1523-1739.2004.00268.x

Leavitt, D. H., Bezy, R. L., Crandall, K. A., & Sites Jr, J. W. (2007a). Multi-locus DNA sequence data reveal a history of deep cryptic vicariance and habitat-driven convergence in the desert night lizard *Xantusia vigilis* species complex (Squamata: Xantusiidae). *Molecular Ecology*, *16*(21), 4455–4481. doi: 10.1111/j.1365-294X.2007.03496.x

Lemaire, C., Versini, J.-J., & Bonhomme, F. (2005). Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology*, *18*(1), 70–80. doi: 10.1111/j.1420-9101.2004.00828.x

Lesica, P., & Allendorf, F. W. (1995). When Are Peripheral Populations Valuable for Conservation? *Conservation Biology*, *9*(4), 753–760. doi: 10.1046/j.1523-1739.1995.09040753.x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Heng, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. doi: 10.1093/bioinformatics/btp698

Li, R. W., Connor, E. E., Li, C., Baldwin VI, R. L., & Sparks, M. E. (2012). Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools: Metagenomics and the rumen microbiota. *Environmental Microbiology*, *14*(1), 129–139. doi: 10.1111/j.1462-2920.2011.02543.x

Lovett, J. C. (1996). Elevational and latitudinal changes in tree associations and diversity in the Eastern Arc mountains of Tanzania. *Journal of Tropical Ecology*, *12*(5), 629–650. doi: 10.1017/S0266467400009846

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. doi: 10.1111/1755-0998.12635

Luijten, S. H., Kery, M., Oostermeijer, J. G. B., & Den Nijs, H. (J.) C. M. (2002). Demographic consequences of inbreeding and outbreeding in *Arnica montana*: a field experiment. *Journal of Ecology*, *90*(4), 593–603. doi: 10.1046/j.1365-2745.2002.00703.x

Lusini, I., Velichkov, I., Pollegioni, P., Chiocchini, F., Hinkov, G., Zlatanov, T., … Mattioni, C. (2014). Estimating the genetic diversity and spatial structure of Bulgarian *Castanea sativa* populations by SSRs: implications for conservation. *Conservation Genetics*, *15*(2), 283–293. doi: 10.1007/s10592-013-0537-0

Lynch, M. (1991). The genetic interpretation of inbreeding and outbreeding depression. *Evolution*, *45*(3), 622–629. doi: 10.1111/j.1558-5646.1991.tb04333.x

MacColl, A. D. C., & Stevenson, I. R. (2003). Stasis in the morph ratio cline in the Bananaquit on Grenada, West Indies. *The Condor*, *105*, 821–825.

Mallet, J., Barton, N., Lamas, G. M., Santisteban, J. C., Muedas, M. M., & Eeley, H. (1990). Estimates of Selection and Gene Flow From Measures of Cline Width and Linkage Disequilibrium in *Heliconius* Hybrid Zones. *Genetics*, *124*, 921–936.

Mallet, J., & Joron, M. (1999). Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation. *Annual Review of Ecology and Systematics*, *30*(1), 201–233. doi: 10.1146/annurev.ecolsys.30.1.201

Marples, N. M., Kelly, D. J., & Thomas, R. J. (2005). The evolution of warning coloration is not paradoxical. *Evolution*, *59*(5), 933–940. doi: 10.1111/j.0014-3820.2005.tb01032.x

Marples, N. M., & Mappes, J. (2011). Can the dietary conservatism of predators compensate for positive frequency dependent selection against rare, conspicuous prey? *Evolutionary Ecology*, *25*(4), 737–749. doi: 10.1007/s10682-010-9434-x

Marshall, D. J., Monro, K., Bode, M., Keough, M. J., & Swearer, S. (2010). Phenotype–environment mismatches reduce connectivity in the sea. *Ecology Letters*, *13*(1), 128–140. doi: 10.1111/j.1461-0248.2009.01408.x

Martínez-Cabrera, Hugo. I., Cevallos-Ferriz, S. R. S., & Poole, I. (2006). Fossil woods from early Miocene sediments of the El Cien Formation, Baja California Sur, Mexico. *Review of Palaeobotany and Palynology*, *138*(3–4), 141–163. doi: 10.1016/j.revpalbo.2006.01.001

Martínez-Cruz, B., Godoy, J. A., & Negro, J. J. (2004). Population genetics after fragmentation: the case of the endangered Spanish imperial eagle (*Aquila adalberti*). *Molecular Ecology*, *13*(8), 2243–2255. doi: 10.1111/j.1365-294X.2004.02220.x

Mautz, W. (1987). Ecology and energetics of the Island Night Lizard, *Xantusia riversiana*, on San Clemente Island. *American Zoologist*, *27*(4), A147–A147.

Mautz, W. (1993). Ecology and energetics of the island night lizard, *Xantusia riversiana*, on San Clemente Island, California. In *Third California Islands Symposium: Recent Advances in Research on the California Islands* (pp. 417–442). Santa Barbara, CA: Santa Barbara Museum of Natural History.

Mautz, W. (2007). *Island night lizard surveys on San Clemente Island*. Commander Navy Region Southwest.

Mayr, E. (1954). Change of genetic environment and evolution. In *Evolution as a process* (pp. 157–180). London: Allen & Unwin.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi: 10.1101/gr.107524.110

McLean, C. A., & Stuart-Fox, D. (2014). Geographic variation in animal colour polymorphisms and its role in speciation: Geographic variation in polymorphism. *Biological Reviews*, *89*(4), 860–873. doi: 10.1111/brv.12083

McMahon, K., Conboy, A., O'Byrne-White, E., Thomas, R. J., & Marples, N. M. (2014). Dietary wariness influences the response of foraging birds to competitors. *Animal Behaviour*, *89*, 63–69. doi: 10.1016/j.anbehav.2013.12.025

McRae, B., Shah, V., & Mohapatra, T. (2013). *Circuitscape 4 User Guide*. Retrieved from http://www.circuitscape.org

Melnick, D. J., & Hoelzer, G. A. (1992). Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *International Journal of Primatology*, *13*(4), 379–393. doi: 10.1007/BF02547824

Meng, L.-W., Li, X., Wang, K., Ma, K.-L., & Zhang, J. (2015). Influence of the amoxicillin concentration on organics removal and microbial community structure in an anaerobic EGSB reactor treating with antibiotic wastewater. *Chemical Engineering Journal*, *274*, 94–101. doi: 10.1016/j.cej.2015.03.065

Merifield, P. M., Lamar, D. L., & Stout, M. L. (1971). Geology of Central San Clemente Island, California. *Geological Society of America Bulletin*, *82*(7), 1989. doi: 10.1130/0016-7606(1971)82[1989:GOCSCI]2.0.CO;2

Michailos, K., Warren-Smith, E., Savage, M. K., & Townend, J. (2020). Detailed spatiotemporal analysis of the tectonic stress regime near the central Alpine Fault, New Zealand. *Tectonophysics*, *775*, 228205. doi: 10.1016/j.tecto.2019.228205

Miller, J. M., Malenfant, R. M., David, P., Davis, C. S., Poissant, J., Hogg, J. T., … Coltman, D. W. (2014). Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity*, *112*(3), 240–247.

Mochida, K. (2009). A parallel geographical mosaic of morphological and behavioural aposematic traits of the newt, *Cynops pyrrhogaster* (Urodela: Salamandridae). *Biological Journal of the Linnean Society*, *97*(3), 613–622. doi: 10.1111/j.1095-8312.2008.01182.x

Molnár, K., Ostoros, G., Dunams-Morel, D., & Rosenthal, B. M. (2012). *Eimeria* that infect fish are diverse and are related to, but distinct from, those that infect terrestrial vertebrates. *Infection, Genetics and Evolution*, *12*(8), 1810–1815. doi: 10.1016/j.meegid.2012.06.017

Molofsky, J., Bever, J. D., & Antonovics, J. (2001). Coexistence under positive frequency dependence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1464), 273–277. doi: 10.1098/rspb.2000.1355

Molofsky, J., J. D. Bever, J. Antonovics, and T. J. Newman. (2002). Negative frequency dependence and the importance of spatial scale. *Ecology* 83:21–27.

Monks, J. M., Monks, A., & Towns, D. R. (2014). Correlated recovery of five lizard populations following eradication of invasive mammals. *Biological Invasions*, *16*(1), 167–175. doi: 10.1007/s10530-013-0511-2

Morafka, D. J., & Banta, B. H. (1973). The Distribution and Microhabitat of *Xantusia vigilis* (Reptilia: Lacertilia) in the Pinnacles National Monument, San Benito and Monterey Counties, California. *Journal of Herpetology*, *7*(2), 97. doi: 10.2307/1563207

Moritz, C., Schneider, C. J., & Wake, D. B. (1992). Evolutionary Relationships Within the *Ensatina eschscholtzii* Complex Confirm the Ring Species Interpretation. *SYSTEMATIC BIOLOGY*, *41*(3), 273–291.

Mouret, V., Guillaumet, A., Cheylan, M., Pottier, G., Ferchaud, A.-L., & Crochet, P.-A. (2011). The legacy of ice ages in mountain species: post-glacial colonization of mountain tops rather than current range fragmentation determines mitochondrial genetic diversity in an endemic Pyrenean rock lizard: Post-glacial history of a Pyrenean mountain lizard. *Journal of Biogeography*, *38*(9), 1717–1731. doi: 10.1111/j.1365-2699.2011.02514.x

Muhs, D. R. (1983). Quaternary Sea-Level Events on Northern San Clemente Island, California. *Quaternary Research*, *20*(3), 322–341. doi: 10.1016/0033-5894(83)90016-9

Myers, E. A., Xue, A. T., Gehara, M., Cox, C. L., Davis Rabosky, A. R., Lemos-Espinal, J., … Burbrink, F. T. (2019). Environmental heterogeneity and not vicariant biogeographic barriers generate community-wide population structure in desert-adapted snakes. *Molecular Ecology*, *28*(20), 4535–4548. doi: 10.1111/mec.15182

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. doi: 10.1038/35002501

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics*, *156*, 297–304.

National Geophysical Data Center. (2012). *U.S. Coastal Relief Model Southern California vers. 2*. Retrieved from http://www.ngdc.noaa.gov/mgg/fliers/04mgg01.html

Nicholson, C., Sorlien, C. C., Atwater, T., Crowell, J. C., & Luyendyk, B. P. (1994). Microplate capture, rotation of the western Transverse Ranges, and initiation of the San Andreas transform as a low-angle fault system. *Geology*, *22*, 491–495.

Noël, F., Machon, N., & Robert, A. (2013). Integrating demographic and genetic effects of connections on the viability of an endangered plant in a highly fragmented habitat. *Biological Conservation*, *158*, 167–174. doi: 10.1016/j.biocon.2012.07.029

Noonan, B. P., & Comeault, A. A. (2009). The role of predator selection on polymorphic aposematic poison frogs. *Biology Letters*, *5*(1), 51–54. doi: 10.1098/rsbl.2008.0586

Noonan, B. P., Pramuk, J. B., Bezy, R. L., Sinclair, E. A., Queiroz, K. de, & Sites, J. W. (2013a). Phylogenetic relationships within the lizard clade Xantusiidae: Using trees and divergence times to address evolutionary questions at multiple levels. *Molecular Phylogenetics and Evolution*, *69*(1), 109–122. doi: 10.1016/j.ympev.2013.05.017

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2018). vegan: Community Ecology Package (Version R package version 2.5-2). Retrieved from https://CRAN.R-project.org/package=vegan

Orosz, F. (2015). Two recently sequenced vertebrate genomes are contaminated with apicomplexan species of the Sarcocystidae family. *International Journal for Parasitology*, *45*(13), 871–878.

Ożgo, M. (2011). Rapid evolution in unstable habitats: a success story of the polymorphic land snail *Cepaea nemoralis* (Gastropoda: Pulmonata). *Biological Journal of the Linnean Society*, *102*(2), 251–262. doi: 10.1111/j.1095-8312.2010.01585.x

Paolucci, E. M., MacIsaac, H. J., & Ricciardi, A. (2013). Origin matters: alien consumers inflict greater damage on prey populations than do native consumers. *Diversity and Distributions*, *19*(8), 988–995. doi: 10.1111/ddi.12073

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, *28*(19), 2537–2539. doi: 10.1093/bioinformatics/bts460

Peery, M. Z., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Bëer, E., Robinson, S., … Palsbøll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology*, *21*(14), 3403–3418. doi: 10.1111/j.1365-294X.2012.05635.x

Peichoto ME, Sanchez MN, Lopez A, Salas M, Rivero MR, Teibler P, Toledo GDM, Tavares FL. (2016). First report of parasitism by *Hexametra boddaertii* (Nematoda: Ascaridae) in *Oxyrhopus guibei* (Serpentes: Colubridae). *Veterinary Parasitology*, 224:60–64. doi: 10.1016/j.vetpar.2016.05.017.

Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, *67*(11), 3274–3289. doi: doi: 10.1 111/evo. 12202

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. doi: 10.1371/journal.pone.0037135

Petkova, D., Novembre, J., & Stephens, M. (2015). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. doi: 10.1038/ng.3464

Podosokorskaya, O. A., Kadnikov, V. V., Gavrilov, S. N., Mardanov, A. V., Merkel, A. Y., Karnachuk, O. V., … Kublanov, I. V. (2013). Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class *Ignavibacteria* , and a proposal of a novel bacterial phylum

Ignavibacteriae: *Melioribacter roseus* gen. nov., sp. nov. and *Ignavibacteriae*. *Environmental Microbiology*, *15*(6), 1759–1771. doi: 10.1111/1462-2920.12067

Porcasi, P., Porcasi, J. F., & O'Neill, C. (1999). Early Holocene Coastlines of the California Bight: The Channel Islands as First Visited by Humans. *Pacific Coast Archaeological Society Quarterly*, *35*(2 and 3).

Powell, T. W. R., & Lenton, T. M. (2013). Scenarios for future biodiversity loss due to multiple drivers reveal conflict between mitigating climate change and preserving biodiversity. *Environmental Research Letters*, *8*(2), 025024. doi: 10.1088/1748-9326/8/2/025024

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Punzalan, D., Rodd, F. H., & Hughes, K. A. (2005). Perceptual Processes and the Maintenance of Polymorphism Through Frequency-dependent Predation. *Evolutionary Ecology*, *19*(3), 303–320. doi: 10.1007/s10682-005-2777-z

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, n/a-n/a. doi: 10.1002/ece3.1305

Queirós, J., Godinho, R., Lopes, S., Gortazar, C., de la Fuente, J., & Alves, P. C. (2015). Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. *Molecular Ecology Resources*, *15*(4), 747–760. doi: 10.1111/1755-0998.12349

R Core Team. (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Raymond, M., & Rousset, F. (1995). GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*, *86*(3), 248–249. doi: 10.1093/oxfordjournals.jhered.a111573

Rech, J. A., Currie, B. S., Shullenberger, E. D., Dunagan, S. P., Jordan, T. E., Blanco, N., … Houston, J. (2010). Evidence for the development of the Andean rain shadow from a Neogene isotopic record in the Atacama Desert, Chile. *Earth and Planetary Science Letters*, *292*(3–4), 371–382. doi: 10.1016/j.epsl.2010.02.004

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods in Ecology and Evolution*, *3*(2), 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21), e145–e145. doi: 10.1093/nar/gkr732

Richards, P. M., Liu, M. M., Lowe, N., Davey, J. W., Blaxter, M. L., & Davison, A. (2013). RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, *22*(11), 3077–3089. doi: 10.1111/mec.12262

Richards-Zawacki, C. L., & Cummings, M. E. (2011). Intraspecific reproductive character displacement in a polymorphic poison dart frog, *Dendrobates pumilio*. *Evolution*, *65*(1), 259–267. doi: 10.1111/j.1558-5646.2010.01124.x

Richards-Zawacki, C. L., Wang, I. J., & Summers, K. (2012). Mate choice and the genetic basis for colour variation in a polymorphic dart frog: inferences from a wild pedigree. *Molecular Ecology*, *21*(15), 3879–3892. doi: 10.1111/j.1365-294X.2012.05644.x

Richmond, J. Q., Wood, D. A., Westphal, M. F., Vandergast, A. G., Leaché, A. D., Saslaw, L. R., … Fisher, R. N. (2017). Persistence of historical population structure in an endangered species despite near-complete biome conversion in California's San Joaquin Desert. *Molecular Ecology*, *26*(14), 3618–3635. doi: 10.1111/mec.14125

Ringsby, T. H., Sæther, B.-E., Jensen, H., & Engen, S. (2006). Demographic Characteristics of Extinction in a Small, Insular Population of House Sparrows in Northern Norway. *Conservation Biology*, *20*(6), 1761–1767. doi: 10.1111/j.1523-1739.2006.00568.x

Roff, D. A. (1996). The Evolution of Threshold Traits in Animals. *The Quarterly Review of Biology*, *71*(1), 3–35. doi: 10.1086/419266

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. doi: 10.7717/peerj.2584

Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Rozen, S., & Skaletsky, H. (1999). Primer3 on the WWW for General Users and for Biologist Programmers. In *Methods in Molecular Biology*: *Vol. 132*. *Bioinformatics Methods and Protocols* (pp. 365–386). Springer.

Rumeu, B., Afonso, V., Fernández-Palacios, J. M., & Nogales, M. (2014). Diversity, distribution and conservation status of island conifers: a global review. *Diversity and Distributions*, *20*(3), 272–283. doi: 10.1111/ddi.12163

Sandoval, C. P. (1994). The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. *Evolution*, *48*(6), 1866–1879. doi: 10.1111/j.1558-5646.1994.tb02220.x

Sawyer, J. O., & Keeler-Wolf, T. (2009). *A Manual of California Vegetation* (2nd ed.). , Sacramento, CA: California Native Plant Society.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. doi: 10.1128/AEM.01541-09

Schurer, J. M., Pawlik, M., Huber, A., Elkin, B., Cluff, H. D., Pongracz, J. D., … Jenkins, E. J. (2016). Intestinal parasites of gray wolves (*Canis lupus*) in northern and western Canada. *Canadian Journal of Zoology*, *94*(9), 643–650. doi: 10.1139/cjz-2016-0017

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and Complex Phenotypes. *Current Biology*, *24*(7), R288–R294. doi: 10.1016/j.cub.2014.01.056

Sharma, P., Sharma, S., Maurya, R. K., De, T. D., Thomas, T., Lata, S., … Dixit, R. (2014). Salivary glands harbor more diverse microbial communities than gut in *Anopheles culicifacies*. *Parasites & Vectors*, *7*(1), 235.

Sherratt, T. N. (2006). Spatial mosaic formation through frequency-dependent selection in Müllerian mimicry complexes. *Journal of Theoretical Biology*, *240*(2), 165–174. doi: 10.1016/j.jtbi.2005.09.017

Sinclair, E. A., Bezy, R. L., Bolles, K., Camarillo, J. L. R., Crandall, K. A., & Sites Jr, J. W. (2004). Testing Species Boundaries in an Ancient Species Complex with Deep Phylogeographic History: Genus *Xantusia* (Squamata: Xantusiidae). *The American Naturalist*, *164*(3), 396–413.

Sinervo, B., & Calsbeek, R. (2006). The Developmental, Physiological, Neural, and Genetical Causes and Consequences of Frequency-Dependent Selection in the Wild. *Annual Review*

of Ecology, Evolution, and Systematics, *37*(1), 581–610. doi: 10.1146/annurev.ecolsys.37.091305.110128

Singhal, S., Huang, H., Title, P. O., Donnellan, S. C., Holmes, I., & Rabosky, D. L. (2017). Genetic diversity is largely unpredictable but scales with museum occurrences in a species-rich clade of Australian lizards. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854), 20162588. doi: 10.1098/rspb.2016.2588

Slatkin, M. (1985). Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, *16*(1985), 393–430.

Slatkin, M. (1987). Gene Flow and the Geographic Structure of Natural Populations. *Science*, *236*(4803), 787–792.

Slatkin, M., & Excoffier, L. (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics*, *191*(1), 171–181. doi: 10.1534/genetics.112.139022

Smith, S. M. (1977). Coral-snake pattern recognition and stimulus generalisation by naive great kiskadees (Aves: Tyrannidae). *Nature*, *265*(5594), 535–536. doi: 10.1038/265535a0

Spatz, D. R., Newton, K. M., Heinz, R., Tershy, B., Holmes, N. D., Butchart, S. H. M., & Croll, D. A. (2014). The Biogeography of Globally Threatened Seabirds and Island Conservation Opportunities: Seabird Conservation Opportunities. *Conservation Biology*, *28*(5), 1282–1290. doi: 10.1111/cobi.12279

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Stebbins, R. (2003). *A field guide to western reptiles and amphibians.* (3rd ed.). New York: Houghton Mifflin Publishing Company.

Steenhof, K. (1983). Prey weights for computing percent biomass in raptor diets. *Raptor Research*, *17*(1), 15–27.

Steenhof, K., Kochert, M., & Moritsch, M. (1984). Dispersal and migration of southwestern Idaho raptors. *Journal of Field Ornithology*, *55*, 357–368.

Stock, J. M., & Lee, J. (1994). Do microplates in subduction zones leave a geological record? *Tectonics*, *13*(6), 1472–1487. doi: 10.1029/94TC01808

Streicher, J. W., McEntee, J. P., Drzich, L. C., Card, D. C., Schield, D. R., Smart, U., … Castoe, T. A. (2016). Genetic surfing, not allopatric divergence, explains spatial sorting of mitochondrial haplotypes in venomous coralsnakes. *Evolution*, *70*(7), 1435–1449. doi: 10.1111/evo.12967

Summers, K., Cronin, T. W., & Kennedy, T. (2004). Cross-Breeding of Distinct Color Morphs of the Strawberry Poison Frog (*Dendrobates pumilio*) from the Bocas del Toro Archipelago, Panama. *Journal of Herpetology*, *38*(1), 1–8. doi: 10.1670/51-03A

Sun, J. X., Helgason, A., Masson, G., Ebenesersdóttir, S. S., Li, H., Mallick, S., … Stefansson, K. (2012). A direct characterization of human mutation based on microsatellites. *Nature Genetics*, *44*(10), 1161–1165. doi: 10.1038/ng.2398

Svensson, E. I., & Abbott, J. (2005). Evolutionary dynamics and population biology of a polymorphic insect: Polymorphism and evolutionary dynamics. *Journal of Evolutionary Biology*, *18*(6), 1503–1514. doi: 10.1111/j.1420-9101.2005.00946.x

Svensson, Erik I., Abbott, J., & Härdling, R. (2005). Female Polymorphism, Frequency Dependence, and Rapid Evolutionary Dynamics in Natural Populations. *The American Naturalist*, *165*(5), 567–576. doi: 10.1086/429278

Swanson, K., Dowd, S., Suchodolski, J., Middelbos, I., Vester, B., Barry, K., … Fahey Jr, G. C. (2011). Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *The ISME Journal*, *5*, 639–649.

Tabak, M. A., Poncet, S., Passfield, K., & Martinez del Rio, C. (2014). Invasive species and land bird diversity on remote South Atlantic islands. *Biological Invasions*, *16*(2), 341–352. doi: 10.1007/s10530-013-0524-x

Tanaka, Y., Tamaki, H., Matsuzawa, H., Nigaya, M., Mori, K., & Kamagata, Y. (2012). Microbial Community Analysis in the Roots of Aquatic Plants and Isolation of Novel Microbes Including an Organism of the Candidate Phylum OP10. *Microbes and Environments*, *27*(2), 149–157. doi: 10.1264/jsme2.ME11288

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., … Tucker, P. K. (2007). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, *18*(1), 67–76. doi: 10.1101/gr.6757907

Tewari, V. P., & Kapoor, K. S. (2013). Western Himalayan Cold Deserts: Biodiversity, Eco-Restoration, Ecological Concerns and Securities. *Annals of Arid Zone*, *52*(3 & 4), 225–232.

Theunissen, M., Tiedt, L., & Du Preez, L. H. (2014). The morphology and attachment of *Protopolystoma xenopodis* (Monogenea: Polystomatidae) infecting the African clawed frog *Xenopus laevis*. *Parasite*, *21*, 20. doi: 10.1051/parasite/2014020

Thomas, R. J., Bartlett, L. A., Marples, N. M., Kelly, D. J., & Cuthill, I. C. (2004). Prey selection by wild birds can allow novel and conspicuous colour morphs to spread in prey populations. *Oikos*, *106*(2), 285–294. doi: 10.1111/j.0030-1299.2004.13089.x

Thompson, J. N. (2005). Coevolution: The Geographic Mosaic of Coevolutionary Arms Races. *Current Biology*, *15*(24), R992–R994. doi: 10.1016/j.cub.2005.11.046

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. doi: 10.1111/j.1365-294X.2012.05664.x

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., … Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, *6*. doi: 10.3389/fmicb.2015.00771

Tuan, N. N., Chang, Y.-C., Yu, C.-P., & Huang, S.-L. (2014). Multiple approaches to characterize the microbial community in a thermophilic anaerobic digester running on swine manure: A case study. *Microbiological Research*, *169*(9–10), 717–724. doi: 10.1016/j.micres.2014.02.003

Turner, J. R. G. (1971). Experiments on the Demography of Tropical Butterflies. II. Longevity and Home-Range Behaviour in *Heliconius erato*. *Biotropica*, *3*(1), 21. doi: 10.2307/2989703

Twomey, E., Yeager, J., Brown, J. L., Morales, V., Cummings, M., & Summers, K. (2013). Phenotypic and Genetic Divergence among Poison Frog Populations in a Mimetic Radiation. *PLoS ONE*, *8*(2), e55443. doi: 10.1371/journal.pone.0055443

U.S. Geological Survey. (2009a). *San Clemente Island Central, CA*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

U.S. Geological Survey. (2009b). *San Clemente Island Central OE E,*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

U.S. Geological Survey. (2009c). *San Clemente Island Central OE S, CA*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

U.S. Geological Survey. (2009d). *San Clemente Island North*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

U.S. Geological Survey. (2009e). *San Clemente Island South, CA*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

U.S. Geological Survey. (2009f). *San Clemente Island South OE E, CA*. Retrieved from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html

van Leeuwen, E., Brännström, Å., Jansen, V. A. A., Dieckmann, U., & Rossberg, A. G. (2013). A generalized functional response for predators that switch between multiple prey species. *Journal of Theoretical Biology*, *328*, 89–98. doi: 10.1016/j.jtbi.2013.02.003

Van Oosterhout, C., Weetman, D., & Hutchinson, W. F. (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes*, *6*(1), 255–256. doi: 10.1111/j.1471-8286.2005.01082.x

Vestergaard, J. S., Twomey, E., Larsen, R., Summers, K., & Nielsen, R. (2015). Number of genes controlling a quantitative trait in a hybrid zone of the aposematic frog *Ranitomeya imitator*. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20141950. doi: 10.1098/rspb.2014.1950

Vicario, S., Caccone, A., & Gauthier, J. (2003). Xantusiid "night" lizards: a puzzling phylogenetic problem revisited using likelihood-based Bayesian methods on mtDNA sequences. *Molecular Phylogenetics and Evolution*, *26*(2), 243–261. doi: 10.1016/S1055-7903(02)00313-5

Wang, I. J., Savage, W. K., & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Molecular Ecology*, *18*(7), 1365–1374. doi: 10.1111/j.1365-294X.2009.04122.x

Ward, S. N., & Valensise, G. (1996). Progressive growth of San Clemente Island, California, by blind thrust faulting: implications for fault slip partitioning in the California Continental Borderland. *Geophysical Journal International*, *126*(3), 712–734. doi: 10.1111/j.1365-246X.1996.tb04699.x

West-Eberhard, M. J. (1986). Alternative adaptations, speciation, and phylogeny (A Review). *Proceedings of the National Academy of Sciences*, *83*(5), 1388–1392. doi: 10.1073/pnas.83.5.1388

Wicht, B., Gustinelli, A., Fioravanti, M., Invernizzi, S., & Peduzzi, R. (2009). Prevalence of the broad tapeworm *Diphyllobothrium latum* in perch (*Perca fluviatilis*) and analysis of abiotic factors influencing its occurrence in Lake Lario (Como, Italy). *Bulletin of the European Association of Fish Pathologists*, *29*(2), 58–65.

Wilson, G. A., & Rannala, B. (2003). Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, *163*, 1177–1191.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Wunderle, J. M. (1981). Movements of Adult and Juvenile Bananaquits within a Morph-Ratio Cline. *The Auk*, *98*, 571–577.

Wylie, J. (2012). *Vegetation change on San Clemente Island following removal of feral herbivores*. San Diego State University.

Yeates, D. K., Zwick, A., & Mikheyev, A. S. (2016). Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science*, *18*, 83–88. doi: 10.1016/j.cois.2016.09.009

Yildirim, S., Yeoman, C. J., Sipos, M., Torralba, M., Wilson, B. A., Goldberg, T. L., … Nelson, K. E. (2010). Characterization of the Fecal Microbiome from Non-Human Wild Primates Reveals Species Specific Microbial Communities. *PLoS ONE*, *5*(11), e13963. doi: 10.1371/journal.pone.0013963

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, *113*(29), 8041–8048. doi: 10.1073/pnas.1602237113

Zarnke, R., Whitman, J., Flynn, R., & Hoef, J. (2004). Prevalence of *Soboliphyme baturini* in marten (*Martes americana*) populations from three regions of Alaska, 1990-1998. *Journal of Wildlife Disease*, *40*(3), 452–455.

Zhang, J. (2017). phylotools: Phylogenetic Tools for Eco-Phylogenetics (Version R package version 0.2.2). Retrieved from https://CRAN.R-project.org/package=phylotools

Zhang, Q.-G., & Buckling, A. (2016). Migration highways and migration barriers created by host-parasite interactions. *Ecology Letters*, *19*(12), 1479–1485. doi: 10.1111/ele.12700