Title

Species Conservation in the Big Data Era: Leveraging Genomic and Community Science Datasets for Conservation Management

Author

Curti, Joseph Nikko

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Species Conservation in the Big Data Era: Leveraging Genomic and Community Science Datasets for Conservation Management

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

by

Joseph Nikko Curti

2024

ABSTRACT OF THE DISSERTATION

Species Conservation in the Big Data Era: Leveraging Genomic and Community Science Datasets for Conservation Management

by

Joseph Nikko Curti

Doctor of Philosophy in Biology

University of California, Los Angeles, 2024

Professor Brad Shaffer, Chair

Many vertebrate species across the planet are experiencing major declines due to rapid expansion of urban areas, habitat destruction for animal agriculture, and human-caused climate change, among other drivers. Recently, the establishment of public data repositories for species occurrences and genetic sequence data have begun to resolve some of the former data limitations that inhibited conservation biologists from acting on declines due to lack of fundamental knowledge for many species. Conservation biologists are therefore primed to now make immense strides in addressing species declines by applying these publicly available datasets to species conservation worldwide. Here I describe two primary approaches to species conservation in the era of big publicly available data using genomic sequencing data and community science occurrence records. Specifically, in Chapter 1, I demonstrate how whole genome datasets can be used to evaluate the barrier effect of roadways on wildlife movement and gene flow in a North American ground bird, the California quail (*Callipepla californica*). I show that compared to other factors including differences in environment or habitat suitability, the presence of roads is the most important factor shaping quail gene flow in Southern California. In Chapter 2, which is now

published in *Journal of Heredity*, we use PacBio HiFi long reads and Omni-C chromatin-proximity sequencing technology to generate one of the most complete de novo genome assemblies for an abundant and widespread North American bat species, the Yuma myotis bat (*Myotis yumanensis*). In Chapter 3, I leverage the novel genomic resource generated in Chapter 2 to summarize genome-wide diversity, historical demography, and range-wide phylogenetics of Yuma myotis to evaluate current subspecies designations and establish genomically-informed management units. Through this work, I found that genomic datasets are generally discordant with existing subspecies designations, revealing two primary genomic groups of Yuma myotis across North America. Additionally, I found that populations of Yuma myotis have high genome-wide diversity and high estimates of contemporary effective population sizes across most populations assessed, which presents a positive conservation outlook for the species. In Chapter 4, which is now published in *PLOS ONE*, we demonstrate how another big data source – community science data from iNaturalist – can be used to evaluate urban affinities of Southern California native vertebrate taxa. Taken as a whole, this body of work demonstrates how big data sets can be applied to conservation on multiple spatial scales – from guiding local biodiversity initiatives for the City of Los Angeles, to suggesting range-wide, continental-scale management units for species conservation.

The dissertation of Joseph Nikko Curti is approved.

Rachel Blakey

Kirk E. Lohmueller

Seth P. Riley

Morgan W. Tingley

Brad Shaffer, Committee Chair

University of California, Los Angeles

2024

DEDICATION PAGE

This dissertation is dedicated to my former advisor Bob Wayne, for his support and mentorship, and most of all, to my partner Anthony Cantatore, for his love, encouragement, and support throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

This dissertation represents the collective work and support of many people. I would like to first acknowledge my committee members, Dr. Rachel V. Blakey, Dr. Kirk E. Lohmueller, Dr. Seth P. Riley, Dr. Morgan W. Tingley, and Dr. H. Bradley Shaffer, for their guidance and feedback that greatly improved this manuscript. I would like to especially thank my advisor Dr. H. Bradley Shaffer for integrating me into his lab group, spending countless hours discussing my research, and for always encouraging me to apply myself and be a better scientist. I also thank members past and present of the Wayne, Shaffer, and Lohmueller labs for your continued feedback on my research over these past six years and for making working at the office enjoyable. Finally, I would like to acknowledge my family and friends who encouraged and supported me throughout my dissertation. Specifically, I would like to acknowledge my parents, Melynda and Edward Curti, who sacrificed immeasurably to get me where I am today. Also, to my cousins Lisa and Don Schmude, thank you for being a home away from home ever since I started my undergraduate studies at UCLA in 2016 – you mean so much to me. To my friends Michelle and David Barton, Natalie Hamilton, Kenzie Harris, Anthony Opocensky, Tim Myer, Kelsi Nelson, Brenna Gibbs, Nurit Katz, Chris Kyriazis, Conner Philson, Peerawat Prasatcharoen, Mike Arellano, Whitney Tsai, and many others, thank you for always being there for me, making me laugh, and encouraging me. To Mary Beth Bourne: I could not have made it through this process if it were not for you, your friendship and love mean everything to me. And most of all, I thank my partner Anthony Cantatore, who has been my best friend and confidant throughout this process.

Chapter One is work of Joseph N. Curti, Phred M. Benham, Rauri C. K. Bowie, Katy S. Delaney, Ryan J. Harrigan, Kirk E. Lohmueller, Zachary. G. MacDonald, Seth P. D. Riley, Robert K. Wayne, H. Bradley Shaffer. JNC primarily carried out genomic analyses, with assistance from RJH and ZGM. Further, sample collection was primarily conducted by JNC, with assistance from PMB, RCKB, and KSD. The manuscript was written by JNC, with input from RJH, ZGM, and all committee members. We would first like to posthumously acknowledge RKW, who invaluably contributed to early stages of this work. We would also like to acknowledge Pedro Chavarria for initial training on quail trapping methods.

We would also like to thank three undergraduate students – Ashley Soto, Rebecca Garcia and Claire Ramirez – for their assistance in the field. For assistance in permitting and landowner permissions, we would like to thank the UCLA Animal Research Committee, Danielle Lefer of California State Parks, staff at the National Park Service of the Santa Monica Mountains, and Paul Edelman of Mountains Recreation and Conservation Authority. For guidance in statistical analysis, we thank Morgan Tingley and Kirk Lohmueller. For assistance with sequencing, we acknowledge QB3 Genomics, UC Berkeley, Berkeley, CA, RRID:SCR_022170. Finally, this project was supported by funding from several groups including Sea and Sage Audubon Society, Pasadena Audubon Society, San Fernando Valley Audubon Society, Santa Monica Bay Audubon Society, the American Pheasant and Waterfowl Society, the La Kretz Center for Conservation Science, and the National Park Service Research Learning Center.

Chapter Two is a reprint of the article:

Joseph N Curti, Devaughn Fraser, Merly Escalona, Colin W Fairbairn, Samuel Sacco, Ruta Sahasrabudhe, Oanh Nguyen, William Seligmann, Peter H Sudmant, Erin Toffelmier, Juan Manuel Vazquez, Robert Wayne, H Bradley Shaffer, Michael R Buchalski (2004) A genome assembly of the Yuma myotis bat, *Myotis yumanensis*. *Journal of Heredity*, *115*(1), 139–148.

https://doi.org/10.1093/jhered/esad053

This article is reprinted by permission from Oxford University Press. The work for Chapter Two was under the direction of article coauthors Michael R. Buchalski and Bradley Shaffer.

Chapter Three is work of Joseph N. Curti, Natalie M. Hamilton, Alicia Kubicki, Robert K. Wayne, H. Bradley Shaffer, Michael R. Buchalski. JNC primarily carried out the genomic analyses, with assistance from NMH. Sample collection and processing was primarily conducted by JNC, with assistance from AK and MJV. The manuscript was written by JNC, with input from NMH, AK, MRB, and all committee members. We would first like to posthumously acknowledge RKW, who invaluably contributed to early stages of this work. We would also like to thank UCLA undergraduate students Claire Ramirez and Jake Jacobsen for their assistance in sample collection. We thank Kristen Ahrens, Ryan Peat, and Scott Osborn from the California Department of Fish for help with sample collection and permitting.

From California State Parks, we would like to thank Danielle Lefer, Nita Barve, and Jessica Vannatta for their assistance in the field and with permitting. For assistance with sample curation, library preparation, and sequencing, we would like to thank Daniel Oliveira, Andrew Tully, Erin Toffelmier, Courtney Miller, and Erik Enbody from the California Conservation Genomics Project. For access to historic MYYU samples we would like to acknowledge Drew Stokes, Vryce Hough, and Philip Unitt from the San Diego Museum of Natural History, Kayce Bell and Shannen Robson from the Los Angeles Museum of Natural History, Kyndall Hildebrandt and Link Olson from the University of Alaska Museum of the North, Caleb Phillips and Heath Garner from the Texas Tech Natural Science Research Laboratory, Loren Ammerman from the Angelo State University Natural History Collections, Chris Conroy from the UC Berkeley Museum of Vertebrate Zoology, and Celia López-González from the Instituto Politécnico Nacional, CIIDIR Unidad Durango. We would also like to thank the following people for contributing samples, assisting with field sampling, or providing advice for field collection or analytical methods: Emily Hamblen, Makenzie Henk, Patricia Brown, Bill Rainey, Jill Carpenter, Erika Noel, Dave Johnston, Abigail Tobin, Bethany Shulze, Austin Waag, Juan Manuel Vazquez, and Winifred Frick. Finally, this work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Chapter Four is a reprint of the article:

Joseph N Curti, Michelle Barton, Rhay G Flores, Maren Lechner, Alison Lipman, Graham A Montgomery, Albert Y Park, Kirstin Rochel, and Morgan W Tingley (2024) Using unstructured crowd-sourced data to evaluate urban tolerance of terrestrial native animal species within a California Mega-City. *PLoS ONE 19*(5): e0295476. https://doi.org/10.1371/journal.pone.0295476

 The work for Chapter Four was under the direction of article coauthor Morgan W. Tingley.

VITA

2016	B.S. in Environmental Science with minors in Conservation Biology & Musicology
2016 – 2019	Grants & Science Administrator, Mountains Restoration Trust, CA, USA

PUBLICATIONS

- Hoffman, S., Benson, L., Philson, C., Chock, R. Y., Curti, J. N., Flores-Negrón, C., Grether, G. F. (2024). Reduction in mutualistic ant aggressive behavior upon sugar supplementation. *Biotropica*: e13390. https://doi.org/10.1111/btp.13390
- Curti, J. N., Barton, M., Flores, R. G., Lechner, M., Lipman, A. J., Montgomery, G. A., Park, A. Y., Rochel, K., Tingley, M. W. (2024) Using Unstructured Crowd-sourced Data to Evaluate Urban Tolerance of Terrestrial Native Species within a California Mega-City. *PLoS ONE*, *19*(5): e0295476. https://doi.org/10.1371/journal.pone.0295476
- Curti, J. N., Fraser, D., Escalona, M., Faibairn, C. W., Sacco, S., Sahasrabudhe, R., Nguyen, O., Seligmann, W., Sudmant, P. H., Toffelmier, E., Vazquez, J. M., Wayne, R., Shaffer, H. B., and Buchalski, M. R. (2023) A genome assembly of the Yuma myotis bat, *Myotis yumanensis*. *Journal of Heredity*, *115*(1), 139–148. https://doi.org/10.1093/jhered/esad053
- Curti, J. N., Fergus, C. E., and Palma-Dow, A. A. D. (2021). State of the ART: Using artificial refuge traps to control invasive crayfish in southern California streams. *Freshwater Science*, *40*(3), 000-000. https://doi.org/10.1086/716185
- De Palma-Dow, A. A., Curti, J. N., & Fergus, C. E. (2020). It's a Trap! An evaluation of different passive trap types to effectively catch and control the invasive red swamp crayfish (*Procambarus clarkii*) in streams of the Santa Monica Mountains. *Management of Biological Invasions*, *11*(1), 44. https://doi.org/10.3391/mbi.2020.11.1.04

- Sastre, N., Francino, O., Curti, J.N., Armenta, T.C., Fraser, D.L., Kelly, R.M., Hunt, E., Silbermayr, K., Zewe, C., Sánchez, A. and Ferrer, L. (2016). Detection, prevalence and phylogenetic relationships of Demodex spp and further skin prostigmata mites (Acari, Arachnida) in wild and domestic mammals. *PLoS One*, *11*(11), e0165765. https://doi.org/10.1371/journal.pone.0165765

AWARDS

2023	Graduate Diversity, Equity, and Inclusion Award, Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA
2023	Graduate Teaching Award, Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA
2022	Karl F. Koopman Award, 50th annual North American Society for Bat Research conference, Austin, TX

CHAPTER 1: **Why didn't the quail cross the road? Using whole genome data to evaluate roadways as barriers to gene flow in a North American ground-dwelling bird**

ABSTRACT

Roadway infrastructure is a dominating feature of landscapes across the globe and the ecological impacts of these roadways are leading to declines in many natural populations. To date, research on the genetic impacts of roadways on bird populations has produced mixed results, with some studies demonstrating that roads correlate with observed levels of genetic differentiation while other studies have failed to find an effect. Here, we use a high coverage whole genome dataset to quantify the impacts of roadway infrastructure on the population genetic differentiation of California quail (*Callipepla californica*), a ground-dwelling and ground-nesting bird species. We non-lethally sampled wild quail from populations located on either side of major and minor roadways in Southern California, a region with the densest road network in the United States. Using a combination of Reciprocal Causal Modeling, Principal Component Analysis, and Estimated Effective Migration Surface analyses, we demonstrate that roadways are the most important factor shaping patterns of genomic differentiation in southern California quail populations. We also show that quail populations sampled alongside the two busiest roadways traversing their southern California habitat have higher levels of inbreeding than other populations within the region resulting from close sibling matings within the last 10 generations. Findings from this study will help guide conservation management efforts for this species, including the construction of the Wallis Annenberg Wildlife Crossing over the 101 Freeway and other efforts to improve wildlife connectivity.

INTRODUCTION

Habitat loss and fragmentation pose significant threats to global biodiversity (Wilson et al. 2016). One of the major contributors to global habitat loss and fragmentation is through the construction of roadways (Jaeger et al. 2005; Jaeger et al. 2007; Madadi et al. 2017). Roadway infrastructure is extensive and increasing, with approximately 3,097,279 km of roadways in the United States alone, and this number

is projected to increase 15% by 2050 (Meijer et al. 2018). Roadways can impact wildlife in several ways, including both direct effects such as mortality due to vehicle strikes (Summers et al. 2011) and acting as barriers to wildlife movement and gene flow (Forman and Alexander 1998), as well as indirect effects such as noise (McClure et al. 2013) and light pollution (Blackwell et al. 2015). These indirect factors also effectively extend the ecological impact of roadways – termed the "road effect zone" – and this has been estimated to cover approximately 19% of the continental United States (Forman and Alexander 1998).

Many bird species across the world are in a state of decline and a primary cause for declines are habitat loss and degradation, including that due to roads (Lees et al. 2022). For example, Cooke et al. (2020) found that at least 47 species of birds in Great Britain, or 53% of all species assessed, decreased in abundance as roadway exposure increased. The impact of roads on avian communities can be attributed to both direct and indirect factors associated with roads (Kociolek et al. 2011), for example road mortalities of bird species (Loss et al. 2014), the fragmentation sensitivity of a given species due to their association with shrub cover (Bolger et al. 2001), or the impacts of artificial light at night (La Sorte et al. 2022) and noise pollution (Parris and Schneider 2009) from vehicles. Not all bird species are equally impacted by roads, however, impacts appear greatest for birds within more wooded environments (Kroeger et al. 2022), species with smaller population sizes, smaller body size, and that are migratory (Cooke et al. 2020), and birds with non-carnivorous diets (de Jonge et al. 2022).

Roadway infrastructure can also leave a detectable signature on the genomic structure of populations by impeding gene flow across the landscape. To date, research aimed at understanding the impacts of roadway infrastructure on gene flow and genetic diversity has largely focused on non-volant terrestrial mammals (Fusco et al. 2021) and herpetofauna (Beninde et al. 2016; McCartney-Melstad et al. 2018; Schmidt and Garroway 2021). For these groups, measures of anthropogenic impacts including human population density and road density can be strong predictors of genetic diversity (Schmidt et al. 2020; Habrich et al. 2021), and in some cases explain up to 50% of all genetic differentiation between populations (Remon et al. 2022; but see Beninde et al. 2016). Perhaps unsurprisingly, the genetic impact of roads seems to be negatively correlated with dispersal ability, as volant mammals such as bats showed

little genetic differentiation across urban environments compared to exurban sites, likely due to their increased vagility and species-specific associations with urban environments (Richardson et al. 2021). Little research has analyzed the potential impacts of roadway infrastructure on bird genetic diversity and gene flow, but due to their increased vagility it is possible they would respond in a similar way as bats. This is supported by some studies that have shown no consistent effect of roadway infrastructure on genetic diversity and differentiation across seven resident North American bird species (Schmidt et al. 2020). However, not all birds are equally mobile, and the degree to which avian populations are impacted by roadways may also depend on species-specific natural history, dispersal ability and life history traits. For example, Roy and Gregory (2016) found that composite metrics of human disturbance that include roadway infrastructure (i.e., human footprint index [Venter et al. 2016]) significantly explained genetic differentiation in Greater Prairie-chickens (*Tympanuchus cupido*), a ground-dwelling bird species, and Delaney et al. (2010) reported high levels of genetic differentiation related to urban habitat fragmentation and a large roadway for a habitat-specialist bird in Southern California. These conflicting results indicate that more research is needed to fully understand the impact of roadways on structuring levels of genetic differentiation in avian communities.

To properly attribute putative impacts of roadway infrastructure to observed patterns of landscape-level genetic differentiation, it is necessary to first compare support for competing hypotheses. Generally, the null expectation for varying levels of genetic differentiation across a landscape is isolation by distance (IBD), defined as a positive relationship between genetic and geographic distance (Wright 1943; Slatkin et al. 1993). Genetic differentiation can also be related to the resistance of the intervening landscape to the organism as it moves between habitat patches (isolation by resistance, IBR; McRae 2006; Zeller et al. 2012), and local barriers to gene flow reflecting selection against maladapted migrants (isolation by environment, IBE; Wange and Summers 2010; Wang and Bradburd 2014, Sexton et al. 2014). Determining the specific factors limiting gene flow is often a critical consideration for wildlife managers seeking to establish the boundaries of management units based on population connectivity (Oyler-McCance et al. 2016; Cassin-Sackett et al. 2019).

In the Santa Monica Mountains and Simi Hills ("SMM") north of Los Angeles (Los Angeles and Ventura Counties, California, USA), the effect of roadways on gene flow has been studied in mule deer (*Odocoileus hemionus*; Pease et al. 2009; Fraser et al. 2019), bobcats (*Lynx rufus*) and coyotes (*Canis latrans*; Riley et al. 2006), mountain lions (*Puma concolor*; Riley et al. 2014) and three species of lizard (*Uta stansburiana*, *Plestiodon skiltonianus*, *Sceloporus occidentalis*) and wrentits (*Chamaea fasciata*; Delaney et al. 2010; Thomassen et al. 2018). These studies indicate that even wide-ranging species, including the wrentit, a small, non-migratory bird, experience reduced gene flow and population fragmentation from urbanization and roads, and that this form of cryptic, but intense habitat fragmentation may be a ubiquitous feature of urban landscapes. This pattern of reduced gene flow as a result of roads may well be due to the magnitude of Los Angeles roadway infrastructure development, which is the densest road network in the United States with more than 39,000 kilometers of roads (Fraser and Chester 2015).

To further explore the relationship between roadway infrastructure, genetic diversity and gene flow in Southern California birds, we chose the California quail (*Callipepla californica*) as a model organism for studying these impacts. California quail are a common, ground-foraging species across Southern California (Leopold 1985). Their range is often dissected by large roadways and urban land features, and they are generally thought to be limited dispersers, with estimates of median dispersal distance of 633 meters (Rushing et al. 2022) and maximum dispersal between 2.03 km (Rushing et al. 2022) and 8 km (Ahlborn and Johnson 2024). To date, there has been no attempt to explicitly quantify the impact of roadways on California quail, despite substantial evidence of enigmatic declines in the species (Brennan 1994) and evidence from closely related quail species indicating increased local extinction risk when habitat patches become too small (i.e., from encroaching roadways; Fies et al. 2002). Further, several studies have identified California quail as a "fragmentation sensitive" species (Bolger et al. 2001; Crooks et al. 2003), suggesting that roads may impact their population genetic connectivity.

Using high coverage whole genome resequencing data from 61 individual quail, we set out to understand how roadway infrastructure influenced observed patterns of genetic differentiation and

diversity. Specifically, we had the following aims: 1) evaluate genetic population structure of California quail across the SMM and specifically across roadways of differing traffic volume, 2) estimate the impact of different landscape features such as roadway presence and traffic volume on quail migration and connectivity, and 3) describe patterns of genome-wide heterozygosity and homozygosity and evaluate any possible relationships with the presence of roadway infrastructure. We predicted that patterns of population structure would emerge around large roadways with high traffic volume such as the US-101 and I-405 freeways and that populations separated by smaller roadways would not form genetically distinct populations. We also predicted that models would identify major roadways such as the US-101 and I-405 freeways as significant barriers to migration and that roadway infrastructure would explain most of the variation in our genetic dataset.

Finally, this project was motivated by ongoing construction of the Wallis Annenberg Wildlife Crossing over the 101 freeway in Southern California. This wildlife crossing will be a first of its kind, spanning one of busiest freeways in the nation with an estimated peak daily traffic of 171,000 vehicles in 2022 (California Department of Transportation 2022). As connectivity of wildlife in light of increasing levels of urbanization continues to be a major issue across the region and largely for many cities across the world, this study will provide timely insights into the genomic implications of this decreased connectivity. Further, with only a handful of studies on the impacts of roadways on southern California native wildlife species, and none on the impacts to primarily ground-dwelling bird species, our California quail dataset provides a critical baseline for future studies on the purported benefits of wildlife crossing structures.

METHODS

Sampling, DNA Extraction, and Sequencing

Blood samples were collected from 47 wild California quail across the Santa Monica Mountains and Simi Hills, Los Angeles and Ventura, CA, USA (Figure 1-1; approximate centroid: 34.068589N, 118.770905W) between July 2020 - April 2022. Quail were non-lethally captured using baited funnel

traps modified from Smith et al. (1981) set approximately 30 minutes before sunrise until 12PM or when quail were captured. Captured quail were removed from traps and immediately placed in an individual cotton drawstring bag until they were processed. Processing including aging of birds based on plumage patterns and molt, as well as sexing and the collection of morphometric data including tarsus, bill, and wing lengths. Holding time never exceeded 50 minutes, and all trapping was approved by the UCLA Institutional Animal Care and Use Committee (IACUC Protocol #: ARC-2020-034) and the California Department of Fish and Wildlife (Scientific Collecting Permit S-201150005-20147-001). A sample of 30 - 70 µl of blood was collected via brachial vein puncture (Owen 2011) using a 25 gauge needle and a capillary tube and stored in avian blood buffer (Seutin et al. 1991, 1:30 blood to buffer). DNA was extracted from blood using a DNeasy Blood and Tissue Kit (Catalog #: 69504, QIAGEN Sciences, Maryland, USA), quantified using a Qubit™ fluorometer (Thermo Fisher Scientific, Maryland, USA) and a 2100 Bioanalyzer (Agilent Technologies, California, USA), and sent for library construction and sequencing on Illumina NovaSeq 6000 at either the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley or Novogene sequencing facility, depending on the sample.

We also sampled two specimens from the Los Angeles County Natural History Museum (Catalog #: LACM 107541, LACM 112287) and two from the Western Foundation of Vertebrate Zoology (Catalog #: WFVZ 52698, WFVZ 53206). Approximately 25 grams of tissue was sampled from dried toe pads, DNA was phenol-chloroform extracted following Tsai et al. (2018), and sent for library construction and sequencing on Illumina NovaSeq 6000 at the Daicel Arbor Biosciences sequencing facility. Ten additional quail genomes provided by study collaborators were included in this study. These samples were collected from quail populations in Northern California and Southern California outside of the SMM and were used to provide broader landscape-level context for genomic data observed in the SMM (Benham et al. in prep).

Data Processing and Alignment

We processed all raw sequencing reads using pipelines from Kyriazis et al. (2023) (https://github.com/ckyriazis/moose_WGS_project/) and Nigenda-Morales et al. (2023) (https://github.com/snigenda/Fin_whale_Population_Genomics/tree/V1.0), following the Genome Analysis Toolkit (GATK) Best Practices (DePristo et al. 2011; Van der Auwera et al. 2013). Samples with multi-lane sequencing reads were concatenated prior to alignment. Raw 150 bp paired end reads were assessed for quality using FastQC v. 0.12.1 (Andrews 2010) and adapter sequences were trimmed using picard *MarkIlluminaAdapters* v. 3.1.1. (https://broadinstitute.github.io/picard/). Adapter-free sequencing reads were then aligned to the California quail reference genome (GCA_023055505.1 [bCalCai1.0.p]; Benham et al. 2023) using BWA-MEM (Li 2013) and alignment quality was assessed using qualimap v. 2.3 (Okonechnikov et al. 2015).

Sex Chromosome Identification

Given that sex chromosomes spend unequal times in males and females, the effects of selection and genetic drift differ from autosomal chromosomes (Johnson and Lachance 2013). Thus, we excluded sex chromosomes from all analyses. We identified the quail Z chromosome by performing a genome-to-genome long-read alignment between the California quail reference genome and the chicken (*Gallus gallus*) reference genome (GCA_016700215.2 [bGalGal1]) in minimap2 v. 2.24 (Li 2018) allowing for up to 5% sequence divergence (-asm5). The resulting .sam file was converted to a .bed file using BEDOPS v. 2.4.41 *sam2bed* (Neph et al. 2012), and the output file was visually inspected for matches between scaffolds in the quail reference genome and the identified scaffold for the Z chromosome in the chicken reference genome (CM035082.1). We did not perform a search for the W chromosome because the reference genome for the California quail was generated using tissues derived from the homogametic sex (a ZZ male).

Genotype Calling and Filtering

Joint genotyping of invariant and variant sites across the quail autosomal genome was performed using GATK v. 3.8. PCR duplicates were removed from aligned .bam files using picard *MarkDuplicates*, genotypes were called for individual samples using GATK *HaplotypeCaller* which included filtering to remove low quality bases (--min-base-quality-score = 20), Z chromosome scaffolds were removed, and joint genotyping was performed using GATK *GenotypeGVCF* on all autosomal scaffolds >1Mb in length for a total genome length of 951,641,657 Bp (87.68% of total genome size of 1,085,331,016 bp). We do not have a known database of variants for California quail, therefore we did not conduct Base Quality Score Recalibration. Instead, hard filtering was performed.

We filtered all autosomal SNPs using a series of depth, missingness, and quality filters modified from GATK filtering recommendations, retaining only high quality biallelic SNPs and monomorphic sites. Prior to hard filtering, we ran GATK *SelectVariants* to select SNPs from our unfiltered .vcf file, and GATK *VariantsToTable* to output INFO field variant annotations. We then visualized the distributions of these annotations in R studio v. 4.2.2 (R Studio Team 2020) to determine appropriate hard filtering thresholds. We also filtered for depth across all samples and on the individual sample basis. For all sample depth filters, we filtered for sites with depth less than 6x and greater than the 99th percentile of depth values across all samples. We then filtered depth on the individual level by subsetting our unfiltered .vcf file to the individual-level using VCFtools v. 0.1.16 (Danecek et al. 2011) and calculating the 99th percentile of depth for each individual sample using a custom python script. We then filtered our dataset using GATK *VariantFiltration* and a custom python script to remove all SNPs that fell outside of these all sample and individual sample depth filters. We also applied the following GATK hard filtering parameters: QUAL < 30, QD < 4.0, FS > 12.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, SOR > 3.0. Finally, we masked all repetitive regions across the autosomal quail genome identified by using a combination of RepeatModeler v. 2 (Flynn et al. 2020) and 'RepeatMasker' v. 4.1.2 (Smit et al. 2015).

Relatedness and Population Structure

To disentangle family structure from population structure, we subsetted our filtered .vcf file to only contain individuals with below second-degree kinship, or the equivalent of retaining samples that were at most first cousins (Kinship Coefficient ≦ 0.0884), and used this set of individuals in all subsequent analyses. We further refined our SNP set by pruning for linkage disequilibrium (LD) using an r^2 threshold of 0.2 and filtered for minor allele frequency (MAF) < 0.01 using *snpgdsLDpruning* in SNPRelate v.1.32.0 (Zheng et al. 2012), and this SNP set was used for all population structure analyses unless otherwise directed by the package manual.

To address our first study aim of evaluating genetic population structure across the SMM, we ran ADMIXTURE v. 1.3.0 (Alexander et al. 2009) on all individuals after filtering for LD and MAF. We first converted our filtered and pruned .vcf file to .ped file using PLINK v. 1.90b6.21 (Purcell et al. 2007), and then ran ADMIXTURE for K = 1 through 10, with 10 iterations each. We determined the best fitting model using cross validation (CV) analysis by averaging CV scores for all 10 runs of a given K value and selecting the model with the lowest CV score. To further investigate population structure, we conducted a principal component analysis (PCA) in SNPRelate using the *snpgdsPCA* function.

To evaluate genetic differentiation across quail populations in the SMM only, we calculated pairwise F_{ST} using a version of Hudson F_{ST} (Hudson et al. 1992; Bhatia et al. 2013) implemented in PLINK v. 2.00a5.12. We transformed F_{ST} to (F_{ST}/1-F_{ST}) (Slatkin 1993) and visualized the resulting values of pairwise genetic difference using a symmetrical pairwise difference matrix using the *df_to_pw_mat* function in the graph4lg v. 1.8.0 package in R. We also calculated values of pairwise Euclidean distance between sampling points in graph4lg using the *mat_geo_dist* function. We then log transformed these geographic distance values and tested for IBD using Mantel tests in the R package vegan v. 2.6-4 (Oksanen et al. 2013).

Migration Between Sampling Locations

To address our second objective of estimating the impact of different landscape features such as roadway presence and traffic volume on quail migration and connectivity, we first assessed if there were differences in quail migration patterns across the SMM using by modeling the effective migration surface using EEMS v. 0.0.0.900 (https://github.com/dipetkov/eems/tree/master; Petkova et al. 2016). As input into EEMS we used a pairwise average genetic dissimilarity matrix generated using the *bed2diffs* function, a habitat perimeter file generated online (https://www.birdtheme.org/useful/v3tool.html), and a .txt file of our sampling locations. Following the package documentation, we ran the function *runeems-snps* to calculate our migration surface using parameter settings for proposal variances that resulted in accepted proposals no less than 10% and no more than 40% of the time. We ran separate models for 100, 200 and 300 demes, and each model was run three times with different starting parameters (i.e., a different random seed for each MCMC chain) for a total of 5,000,000 iterations with a 2,500,000 burn in period. We assessed MCMC chain convergence via a posterior probability trace plot and visualized EEMS outputs using the reemsplot2 and ggplot2 packages in R (https://github.com/dipetkov/reemsplots2/tree/master).

Estimating The Impact of Roads on Patterns of Genetic Differentiation

To further address our second aim, and to specifically identify what environmental features of the landscape might impact quail migration and connectivity, we performed reciprocal causal modeling on spatial data sets of habitat suitability, landscape and environmental variables, as well layers of road presence and traffic volume.

Habitat suitability modeling

To test whether landscape features such as habitat suitability and roadway infrastructure shape patterns of genetic differentiation in quail populations, or if the null expectation of IBD is the predominant force shaping genetic differentiation in our samples, we generated spatial datasets of select

landscape features. To capture variation in habitat quality, we generated a predicted habitat suitability surface using MaxEnt v.3.2.3 (Phillips et al., 2006) implemented in the R package dismo v. 1.3-9 (Hijmans et al., 2024). Following MacDonald et al. (2022), we defined the study extent as a minimum convex polygon encompassing all sequenced individuals buffered by 50 km (>50% the maximum Euclidean distance between samples), ensuring that extent edges did not interfere with downstream connectivity analyses. Input presence data included the capture locations of all quail samples as well as presence data from iNaturalist (iNat, accessed 2024 August 13; research-grade observations, coordinate uncertainty ≤ 300 meters). After thinning to one presence per 300x300 m raster cell (see below for raster details), 304 occurrences remained for model training. Predictor variables were generated according to MacDonald et al. (*In Review*) and included the Enhanced Vegetation Index (EVI; a measure of photosynthetic activity), terrain ruggedness index (TRI), the presence of and distance to surface water, land cover classification based on the National Land Cover Database (Dewitz 2021), percent imperviousness, VIIRS artificial light at night (Elvidge et al. 2021), and a number of environmental variables generated using ClimateNA v. 7.50 software (Wang et al. 2016). Environmental variables included mean temperature in summer and winter, total precipitation in summer and winter, the difference in mean temperature between coldest and warmest months (continentality), and heat load (MacDonald et al. In Review). All spatial data layers were generated in or resampled to a 300x300m spatial resolution in an equal-area projection optimized for California (NAD 1983 California [Teale] Albers Equal Area Conic; EPSG:3310) . For a complete list of data sources for these layers, see Table S1-1.

We ran MaxEnt using 10,000 random background points to characterize the distribution of predictor variables within the study extent (Beninde et al. 2023). All feature classes were made available and the regularization parameter was set to 1.0 (Phillips 2006). A predicted habitat suitability surface was then generated using MaxEnt's "clog-log" output (*predict* function the the R package raster v. 3.6-26 (Hijmans 2023). Prediction accuracy was assessed using using area under the receiver operating characteristic curve (AUC_{ROC}) (Phillips et al. 2017). Additional information on generation of predictor layers and MaxEnt modeling can be found in the Supplemental Methods 1-I.

Spatial layers of roads and traffic volume

To test for the impacts of roadway infrastructure on quail genetic differentiation, we generated a continuous spatial layer of roadways in southern California by constructing two spatial layers of roads. First, we downloaded all roadway data for the study extent from the US Census Bureau TIGER/line dataset (https://www.census.gov/cgi-bin/geo/shapefiles/index.php), filtered to include only primary and secondary roadways (MTFCC values of S1100 and S1200), as other road types were minimally present within the study area. We then converted these linear features into a raster (equal in resolution and projection to rasters generated above) using the R package terra v. 1.7-29 (Hijmans 2024), assigning all road pixels a value of one and all non-road pixels zero. Second, we used the same TIGER/line dataset and combined this with statewide data on traffic volume (California Department of Transportation 2022) to create a raster (~82.5 m^2 resolution), bound by the polygons identified as primary and secondary roadways, of annual average daily traffic (AADT), a measure of the average daily traffic volume derived from the annual number of vehicles that pass over a given roadway. Additional information the making of the traffic volume layer can be found in the Supplemental Methods 1-I.

Geographic, Ecological, and Environmental Distances

For geographic distances between our sampling points, we calculated euclidean distances using the *spDists* function in the R package sp v. 2.1-2 (Pebesma and Bivand 2005). For landscape layers of habitat suitability and roadways, we calculated both least cost paths and resistance distances between sampling points. Least cost path distances often have had more power than euclidean distance alone in explaining genetic differentiation (Storfer et al. 2007; Wang et al. 2009). Alternatively, resistance distances use circuit theory to calculate an integrated measure of distance between many random walk paths between two points (McRae and Beier 2007). The efficacy of each of these two measures of resistance distance to predict gene flow between populations can depend on species-specific traits (Spear et al. 2015); therefore we present both measures. Least cost path and resistance distance were calculated using the R package gDistance v. 1.6.4 (van Etten 2017), using resistance surfaces created from the

inverse of our habitat suitability model as well as our layers of road presence and traffic volume. Finally, we also calculated environmental distances for summer and winter temperature and precipitation. Environmental distances were calculated as the absolute difference in environmental variable values between sampling locations (Wang et al. 2013).

Genetic Distances

Choice of genetic distance metric can impact landscape genomics results (Beninde et al. 2024). We calculated both Euclidean genetic distance and Nei's genetic distance. Euclidean genetic distance is defined as the sum of the squared difference of genotypes at a given locus and makes no assumptions about the underlying processes that generate genetic variation (Georges et al. 2023). Nei's genetic distance is another common distance metric based on difference in allele frequencies across genomic loci (Nei 1972) and increases as divergence between populations increases due to mutation, selection, and genetic drift. We calculated Euclidean genetic distance using the *dist* function in the R package adegenet v. 2.1.10 (Jombart 2008) and Nei's D using the *stamppNeisD* function in the R package StAMPP v. 1.6.3 (Pembleton et al. 2013). After calculating these two genetic distances, we assessed their degree of correlation in R using the *cor.test* function and found that these two measures were nearly perfectly correlated (Pearson's $R(1033) = 0.996$, $p < 0.001$). Therefore, we only present results for Euclidean genetic distance given its ease of interpretability and presence in the literature (Shirk et al. 2017; MacDonald et al. 2020; MacDonald et al. 2022).

Reciprocal Causal Modeling

We used reciprocal causal modeling (RCM) to identify which factors (i.e., landscape resistance, environmental, or geographic) best explained observed patterns of genetic differentiation and gene flow. RCM has been shown to have high statistical power in identifying causal models that influence patterns of genetic differentiation between highly correlated variables (Shirk et al. 2012) and has shown to effectively resolve landscape genomic relationships in both simulation (Shirk et al. 2018) and natural

(MacDonald et al. 2020; MacDonald et al. 2022) settings. We performed partial Mantel tests between genetic distances and pairs of landscape/resistance, environmental, and geographic factors with 999 permutations in the R package vegan. This method calculates the Mantel R coefficient between genetic distance and a focal variable ($R_{PM\text{-}A}$), conditioned on a second alternative variable (Cushman et al. 2013). Focal and alternative variables are then reversed, and Mantel correlation coefficients are re-calculated ($R_{PM\text{-}B}$), and whatever focal variable receives a higher mantel R coefficient is determined to be the variable with the highest support. We then calculated an index of relative support for each comparison of the focal and alternative variables by taking the difference, where ($R_{PM\text{-}A} > R_{PM\text{-}B}$) indicates that the focal variable has more support and ($R_{PM\text{-}A} < R_{PM\text{-}B}$) indicates that the alternate variable has more support. Finally, these relative support values are averaged across all focal variables to generate a single relative support value, and we visualized these pairwise comparisons with a heatmap.

Genetic Diversity and Runs of Homozygosity

To address our third study aim of describing patterns of genome-wide heterozygosity and homozygosity and evaluating any possible relationships with the presence of roadway infrastructure, we first calculated autosomal heterozygosity, or the total number of heterozygous sites divided by the total number of variant and invariant sites (i.e., per base pair heterozygosity), within 10 Mb sliding windows across the genome. Following Robinson et al. (2021), we filtered out windows where the number of called sites was less than half the length of the sliding window (i.e., 5Mb) to control for the potential effect of low genotyping rate along certain scaffolds on heterozygosity calculations.

We also calculated the proportion of the genome covered in runs of homozygosity (F_{ROH}) using BCFtools *RoH* v. 1.9 (Narasimhan et al. 2016), as this has been shown to be a robust measure of inbreeding when compared to other methods (Caballero et al. 2020). We calculated F_{ROH} by summing the total length of runs of homozygosity (ROH) and dividing by the total length of the scaffolds used in the analysis (951,641,657 bp). We visualized the results in binned size categories (0.1 - 1Mb, 1 - 10Mb, and 10 - 100Mb) of ROHs using a custom R script (Kyriazis et al. 2023).

We also estimated the relative age of ROH segments following Stoffel et al. 2021. Based on the chicken genome, we assumed that 2.7 centimorgans (cM) approximately equals 1 Mb (Robinson et al. 2021; Groenen et al. 2024) To calculate time to the most recent common ancestor (T_{MRCA}) of ROHs, we used the formula: ((100/2*g) cM / (2.7 cM/Mb)). We binned values in non-overlapping windows of 2 generations ago (g), 2 - 4g, 4 - 8g, 8 - 16g, 16 - 32g, and greater than 32g.

To test for significant differences in per base pair heterozygosity and the F_{ROH} between sampling locations across the state, we performed a non-parametric Kruskal-Wallis rank sum test in the R package 'stats' (R Core Team 2023) and with post-hoc Dunn's tests using the R package rstatix (Kassambara 2023). To test for significant differences in the amount of ROHs within binned T_{MRCA} groups and between roads, we used bootstrapping to test the null hypothesis that there is no difference in median number of summed ROHs within an age group between quail on either sides of the US101 and I405 freeways, and all other quail sampling locations that are not separated by these major roadways (Calmettes et al. 2013). Briefly, we conjoined summed lengths of ROHs by age class from both sampling locations into a single set, and resampled with replacement from this set. This was repeated 10,000 times, and the 5% and 95% cutoffs of the 10,000 computed differences were used as the cutoffs to evaluate statistical significance ($\alpha = 0.05$).

RESULTS

Genomic Data

We achieved high per-individual average genomic sequence coverage (before filtering mean = 24.8, range = 15.0 - 47.5; after filtering mean = 16.4, range = 11.5 - 34.6). Missingness per sample was generally low (average = 0.026, range = 0.0041 - 0.097), but one sample that was derived from preserved museum tissues had high average missingness (x = 0.69) and was subsequently removed from analysis. After filtering, we retained a total of 940,401,236 nucleotides containing 25,199,223 SNPs.

Relatedness, Population Structure, and Genetic Differentiation

We detected several samples with first-degree (kinship coefficient = 0.177 - 0.354) and second-degree (kinship coefficient = 0.0884 - 0.177) relationships (Figure S1-1). For samples with second-degree relationships and above, we randomly removed individuals from the dataset so that only one of the related individuals remained for all analyses that assumed that input samples were unrelated to each other (e.g., analyses focused on detecting population structure).

ADMIXTURE analysis failed to detect population structure across sampling locations in California or within the SMM (Figure S1-3). Cross validation results indicated that the top supported model has K = 1 population clusters and this was consistent across hierarchical runs of ADMIXTURE that excluded Northern and Southern California samples (Figure S1-4). Principal component analysis, which can be more sensitive, revealed large-scale population structure between sampling locations, with the first principal component axis (PC1) separating Northern California, SMM, and Southern California samples (Figure 1-2A). Within the SMM cluster (Figure 1-2B), PC1 separated the eastern SMM samples (I-405) from the western samples (US-101, N-23, Newbury Park, Kanan Road), while PC2 separated samples from either side of the I-405 freeway in the eastern SMM. This pattern was generally consistent with more and less explanatory principal components, although at higher principal component axes, the cluster of samples from east of I-405 became more dispersed and clustered more closely with samples from Topanga Cyn (Figure S1-2).

Estimates of genetic differentiation (F_{ST}) from quail across different sampling locations were generally low (mean = 0.034, range = -0.0082 to 0.071; Figure S1-5), although some sampling sites such as South US101, west of Malibu Cyn., Montenido, and Stunt Rd. had moderate ($x > 0.07$) levels of pairwise genetic differentiation. There was a weak positive relationship between Euclidean genetic distance and geographic distance, providing little support for a model of isolation by distance (mantel R = 0.068, p = 0.33; Figure S1-6), suggesting that something other than geographic distance is contributing to levels of genetic differentiation.

Estimated migration surface

We estimated effective migration of quail across the SMM in EEMS. For each iteration of the total number of demes, all three MCMC chains converged based on visualization of the posterior trace plots (Figure S1-7). The migration surface for 100, 200 and 300 modeled nDemes all indicated low effective migration for samples on either side of the I405 freeway and Malibu Canyon Rd., and to a lesser extent the US101 freeway. Further, the migration surface for 100 and 200 modeled nDemes indicated an increase in effective migration along Topanga Canyon and the two eastmost sites north of the US101 (Figure 1-3).

Habitat Suitability Modeling

We used MaxEnt to generate an ecological niche model for Santa Monica Mountains quail. The resulting model performed well in accurately discriminating habitat suitability (AUC = 0.88; Figure S1-8). The relative contributions (RC) and permutational important (PI) for the top five variables are as follows: Terrain Ruggedness Index (RC = 39.8, PI = 29.7), winter temperature (RC = 13.2, PI = 9.2), night light (RC = 11.6, PI = 15.7), percent impervious surface (RC = 10.7, PI = 19.7), and EVI (RC = 6.5, PI = 9.2; Table S1-1). We used the inverse of this model to parameterize a resistance surface for Reciprocal Causal Modeling.

Reciprocal Causal Modeling

We assessed support for geographic, environmental, and resistance variables in explaining observed patterns of genetic differentiation in our data using reciprocal causal modeling, and summarized these results in a heatmap (Figure 1-4). The heatmap displays the focal variable of the partial mantel test along the y-axis (rows) and the alternative variable along the x-axis (columns). Each cell represents a partial mantel test between the focal variable (row) and the matrix of genetic differentiation while partialling out the effect of the alternative variable (column). Variables with higher support are visualized with warmer colors. Across all variables, the variable with the highest correlation with genetic

differentiation was road resistance distance, followed by least cost paths distance of roads, least cost paths distance of the habitat suitability model, Euclidean distance between sampling locations, and the resistance distance of the habitat suitability model (Table S1-2). Environmental variables were the least supported correlates in RCM analysis and all had negative relative support values with the exception of winter temperature.

Genetic Diversity and Runs of Homozygosity

Across California, California quail genome-wide diversity was high (average = 5.01×10^{-3} het/bp, range = 4.25×10^{-3} to 5.29×10^{-3} het/bp). Quail significantly differed in levels of diversity between sampling locations ($\chi^2(2) = 10.76$, $p = 0.0046$; Figure 1-5A), with the highest overall genomic diversity in quail sampled from the SMM (average = 5.05×10^{-3}, range = 4.60×10^{-3} to 5.29×10^{-3}), followed by Southern California samples (average = 4.91×10^{-3}, range = 4.83×10^{-3} to 5.03×10^{-3}), and Northern California samples (average = 4.43×10^{-3}, range = 4.25×10^{-3} to 4.71×10^{-3}). This represents a significant increase in overall heterozygosity of about 12.3% between populations of quail in the SMM compared to quail sampled in Northern California ($z = 2.92$, adj. $p = 0.0106$).

Levels of inbreeding across all sampled quail populations were low, with an average 2.66% of the autosomal genome in ROHs ≥ 100 Kb (F_{ROH} range = 0.87 to 9.00%). Average values of F_{ROH} were nearly twice as large in northern California (average = 4.24%, range = 2.62 to 5.11%) compared to southern California (average = 2.62%, range = 1.28 to 3.35%) and the SMM (average = 2.58%, range = 0.872 to 9.00%), although this difference was not significant (Figure 1-5B). The fraction of the genome coverage in ROHs ≥ 1 Mb was generally low (1.57%, range = 0.159 to 7.77%), with the highest levels in northern California (average = 2.06%, range = 0.995 to 2.96%) and much lower values in the SMM (average = 1.55%, range = 0.159 to 7.77%) and southern California (average = 1.40% , range = 0.592 to 2.00%). Two samples from west of the I-405 and north of the US-101 in the SMM had ROHs ≥ 10 Mb, totalling in 10.21 Mb and 28.34 Mb in size (Figure 1-5C). Given these high values, we investigated the timing of inbreeding, comparing samples adjacent to the US101 and I405 freeways to all other roadways studied in

the SMM. We found an increase in recent inbreeding between 2 - 4 and 4 - 8 generations ago at sites near the US-101 and I-405 when compared to all other roadways in the SMM (Figure 1-5D), although these comparisons were not statistically significant (Figure S1-9). Further, we found an increase in older inbreeding events at all other roadways when compared to the US-101 and I-405 freeways including between 8 - 16, 16 - 32, and greater than 32 generations ago, although only the most ancient inbreeding events were statistically significant ($p = 0.03$).

DISCUSSION

Understanding the factors that influence genetic diversity, at local and regional scales, is critical to landscape ecology and conservation biology. Although we often think of birds as vagile organisms that can readily seek out suitable habitat and thus escape most of the human-caused impacts on the landscape, we also know that some species are still affected by roads and urbanization, especially in xeric western North America (Benítez-López et al. 2010; Aronson et al. 2014; Seress and Liker 2015). California quail, an important game species, is a strong, but generally reluctant flier that has never been genetically evaluated as its California habitat has become increasingly fragmented and urbanized. We found that major roadways that dissect the SMM left a detectable signature on California quail genomic architecture, and reciprocal causal modeling demonstrated that road presence is the most important factor shaping quail population structure and genetic differentiation. Fortunately, overall levels of genome-wide diversity were high for California quail, although levels of inbreeding increased in birds living proximal to the largest roads in our study. To our knowledge, our study is the first genomic dataset published for this species and one of the few attempts to determine the genomic consequences of roadway infrastructure on birds. These results indicate that quail tend not to cross major freeways, and provide a valuable baseline of differentiation related to connectivity in the region, including before the widely anticipated Annenberg Wildlife Overpass traversing US-101 in the SMM is completed.

Roads as barriers to gene flow

Results from the Principal Components Analysis, Reciprocal Causal Modeling, and Estimated Effective Migration Surface analysis provide strong support for isolation by resistance acting as the prevailing force structuring observed levels of genomic differentiation in SMM California quail populations. Specifically, both least cost paths and resistance distances generated using the road presence resistance surface had the highest relative support values when compared to other distance matrices for habitat suitability and environmental factors. These findings were further substantiated by clustering of samples from either side of the US-101 and I-405 in the PCA analysis and evidence of decreased migration near these same roadways in the EEMS analysis. Relative model support values for the presence of roads were higher than those generated using the road traffic volume resistance surface, indicating that road surface avoidance behavior is more important than traffic avoidance behavior in structuring quail populations within the study area, and this result adds to the growing information that bird abundances are affected by distance to roadway infrastructure but not the intensity of traffic (Benítez-López et al. 2010). We notably did not test for the hypothesis that quail avoid roads due to the increased noise levels (Jaeger et al. 2005). Road noise is an important factor shaping bird communities (McClure et al. 2013) and future studies should include measures of the "soundscape" to better understand its effect.

In the context of other landscape genomics research in birds, factors that best explain patterns of genetic differentiation vary greatly depending on the taxa being studied. For example, in some studies the most important factors for shaping genetic differentiation in birds include linear structures such as rivers (Wenzel et al. 2016), manmade features such as reservoirs (Cros et al. 2020), the amount of managed agricultural lands and exposed soils (Malpica and González 2024), as well as landscape resistance due to elevation (Sonsthagen et al. 2022; Jiao et al. 2024; Kimmitt et al. 2024) and vegetation cover (Row et al. 2018; Cheek et al. 2022; Luna et al. 2023). However, support for IBE (Manthey and Moyley 2015; Provost et al. 2022; Wang et al. 2022) and IBD (Rodríguez-Bardía et al. 2022; Spurgin et al 2024) are also common in avian systems. To our knowledge, this is the first study to find roads as the principal

factor structuring genetic differentiation in a bird species after ruling out competing hypotheses of IBD and IBE.

Roads and genetic diversity

Only a few studies have evaluated the relationship between roadway infrastructure and levels of standing genetic variation in birds. Schmidt et al. (2020) used microsatellite data for 129 bird species across the United States and Canada and found no significant association between levels of genetic diversity and measures of urbanization that included the presence of roadways. This study attributed a lack of a relationship between urbanization and genetic diversity to the high vagility of bird species, which they suggested can confer some resilience to the genetic impacts of habitat fragmentation. However, research from the same region that we studied found that wrentits, a habitat specialist bird species, exhibited significant levels of genetic differentiation associated with intense fragmentation including across a major roadway (Delaney et al. 2010; Thomassen et al. 2017). We found that California quail harbor high levels of genome-wide diversity (mean per base pair heterozygosity = 5.01×10^{-3}) throughout much of their range, presumably reflecting the large populations that were historically present in the state (Leopold 1985), but that there were no significant differences in heterozygosity within the SMM relating to sampling location or what side of the road quail were sampled from. Further, unlike wrentits, we did not find significant genomic differentiation across the US-101 or between any of our sampling locations (mean $F_{ST} = 0.034$), indicating that the standing genetic variation that we observed is largely shared across sampling sites. Relative to genomic diversity in other bird species, these values are higher than average (Ellegren 2013; Brüniche-Olsen et al. 2021; Mathur and DeWoody 2021), and higher than in other species of quail (northern bobwhite quail [*Colinus virginianus*], per base pair heterozygosity = 4.17×10^{-3}, Oldeschulte et al. 2017; scaled quail [*Callipepla squamata*], per base pair heterozygosity = 2.48×10^{-3}, Oldeschulte et al. 2017, and Montezuma quail [*Cyrtonyx montezumae*], per base pair heterozygosity = 1.2×10^{-3}, Mathur and DeWoody 2021). Overall, the high levels of standing genetic variation observed here is a positive finding for the species, as high levels of genetic diversity have been

demonstrated in many systems to be correlated to fitness and adaptive potential (DeWoody et al. 2021). Further, these values provide an important baseline that can be reevaluated in the future in response to changes in census size estimates.

In addition to providing traditional measures of genetic diversity, whole genome sequencing data enables researchers to more effectively estimate levels of inbreeding in wild populations. Inbreeding, which results in an overall increase in homozygosity, can lead to many detrimental impacts to individuals. In birds, this can include congenital defects and decreased embryonic survival (Fu et al. 2019), decreases in juvenile survival rates (Duntsch et al. 2023), decreased hatchling weight (Chen et al. 2016), decreased hatching and fledging success (Suzulkin et al. 2007; Chen et al. 2016), and decreases in lifetime fitness (Harrisson et al. 2019). To our knowledge, no studies have tried to link levels of inbreeding in a wild bird population with the presence of roadway infrastructure. We found that all quail populations in the SMM harbored on average 2.66% of their genomes in long tracts of homozygous alleles greater than 100 Kb in length, and that some quail sampled ~1 mile from the US-101 and I-405 freeways had large ROHs greater than 10Mb that had an estimated time to the most recent common ancestor of 2- 4 generations ago. ROHs of this length indicate recent inbreeding, which can lead to decreased fitness due to inbreeding depression (Robinson et al. 2023). While we did not sample quail to explicitly test how distance from roadways impacts inbreeding, there is at least some evidence that quail near major roadways such as the US-101 and I-405 freeways have increased levels of non-random mating between close kin. This information is important for wildlife managers in the region, as inbreeding is a process that removes genetic diversity from populations over time, and this future loss in diversity could impact quail populations fitness and survival in the future.

Compared to other avian species, these levels of inbreeding are relatively low. For example, values of F_{ROH} in SMM quail are similar to values for recovering populations of Seychelles paradise flycatcher (*Terpsiphone corvina*; $F_{ROH} \sim 0.01\%$; Femerling et al. 2023), Canary Island populations of Berthelot's pipit (*Anthus berthelotii*; $F_{ROH} = 0.8 - 3.90\%$; Martin et al. 2023), and outbred wide-ranging species such as turkey vultures (*Cathartes aura*; $F_{ROH} = 4.24\%$; Robinson et al. 2021), although these

values are also similar to the ‘Alalā (*Corvus hawaiiensis*; F_{ROH} = 5.6%; Sutton et al. 2018) which is now extinct in the wild. However, the values of F_{ROH} that we found in CA quail are far lower than those observed in wild avian species threatened with extinction, which can range from 14.4 - 53% (Fu et al. 2019; Dussex et al. 2021; Duntsch et al. 2023; Sato et al. 2023). Values of observed F_{ROH} in the SMM could be due to a variety of factors including local population declines, overall low levels of movement between populations due to roads and other anthropogenic features on the landscape, and heightened levels of consanguineous matings (Robinson et al. 2018; Kyriazis et al. 2023). Further increases in levels of inbreeding in SMM quail populations could have long lasting impacts on viability of these birds, as populations that were historically large have an increased risk from inbreeding depression due to their increased burden of deleterious variants across the genome (Kyriazis et al. 2021). As genomic resources are made available for the species including an annotation to the existing reference genome, it will be possible to establish whether or not there is spatial variation in the accumulation of deleterious variation including in these populations where large ROHs were observed. Determining exactly how prevalent these recent inbreeding events have become across the landscape will help determine if additional conservation actions are needed and where they should be implemented to prevent increases in inbreeding in the future.

Mitigating the barrier effect of roads

In light of our findings regarding the impact of roadways on quail gene flow, a natural next question is ‘can conservation actions such as wildlife crossings mitigate the barrier effect caused by roadways for California quail’? Unfortunately, it is not a straightforward question to answer. The general lack of information regarding the efficacy of wildlife crossings on gene flow in avian systems makes it unclear whether or not the addition of a wildlife overpass will have its intended impact on quail populations. Research on the efficacy of wildlife crossing structures on lessening the barrier effect of roadways is extremely limited (Denneboom et al. 2021), and we found no research or established guidelines relating to wildlife crossing design that might suggest ways to improve crossings for birds

(Kociolek et al. 2015) let alone quail, which we predict may be more affected by roads than other bird species. In non-avian taxa, a few examples from the literature demonstrate that the addition of wildlife crossing structures can increase gene flow in regions where habitat is bisected by a major roadway. For example, Sawaya et al. (2014) found evidence of bidirectional gene flow and reproduction between formerly isolated bear populations following the construction of multiple wildlife crossing structures. Similarly, Soanes et al. (2018) found that road crossing structures erased the signature of genetic structure for some populations of gliding marsupials as soon as five years after construction. There is at least some evidence that quail and other ground birds use underpass structures (Kociolek et al. 2015; Smith et al. 2015; Caldwell and Klip 2020; Young et al. 2023) and that wildlife overpasses can create crossing opportunities for bird species that otherwise would not cross roads due to their dispersal ability (Pell and Jones 2015). Based on our relatedness analysis alone, we did not find evidence of recent reproduction between quail populations separated by major roadways which may be attributable to the near complete barrier these structures create for quail movement. Therefore, if quail adjacent to the Annenberg Wildlife Crossing are able to locate and utilize the structure, it seems plausible that this wildlife crossing could alleviate some of the constraints to quail migration that roadways in Southern California cause. In order to evaluate the efficacy of the wildlife crossing for quail populations in the SMM, we recommend long-term deployment of passive methods to assess quail use of the crossing such as camera traps or acoustic detectors as well as additional genetic sampling post-construction to evaluate potential increases in gene flow and reproduction. Ultimately, this first assessment of quail genomic health provides invaluable baseline data to understand how conservation actions such as wildlife crossings can benefit this iconic game bird species across the state.

FIGURES

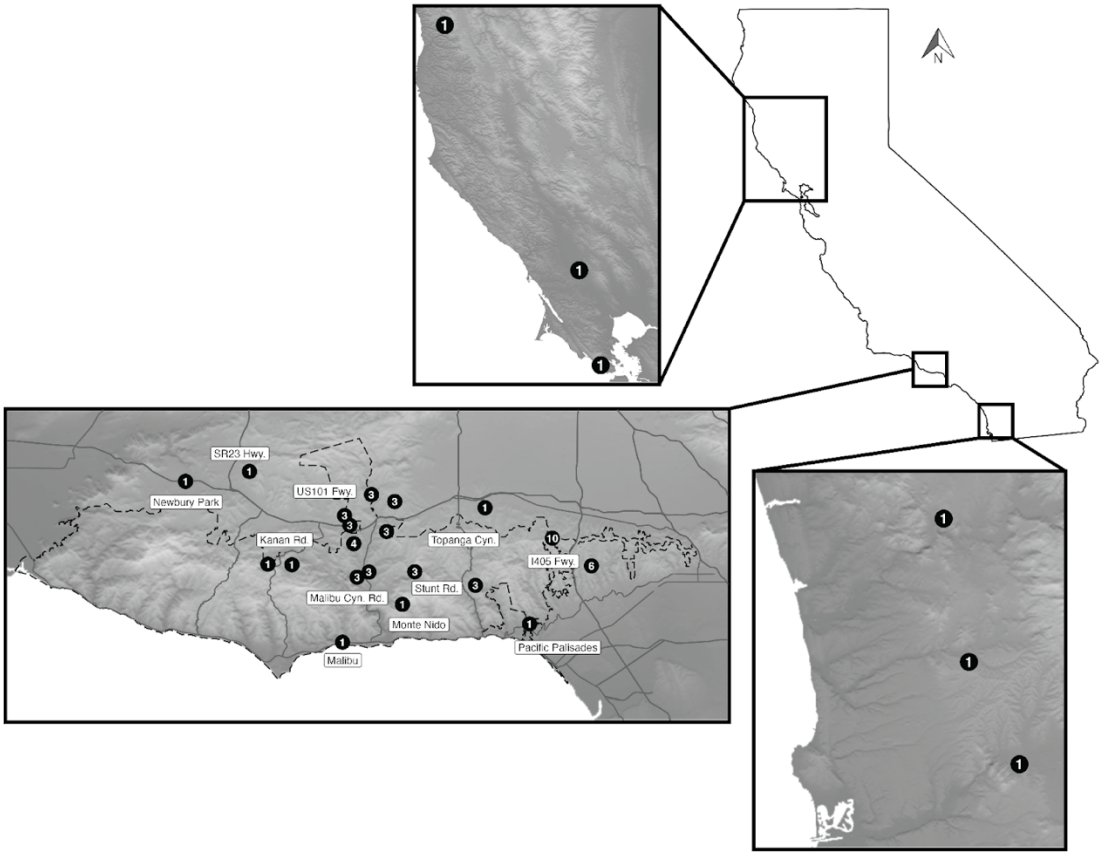

Figure 1-1. Map of the study area. Black circles indicate sample locations and number of samples. Dashed line represents the boundary of Santa Monica Mountains National Recreation Area.

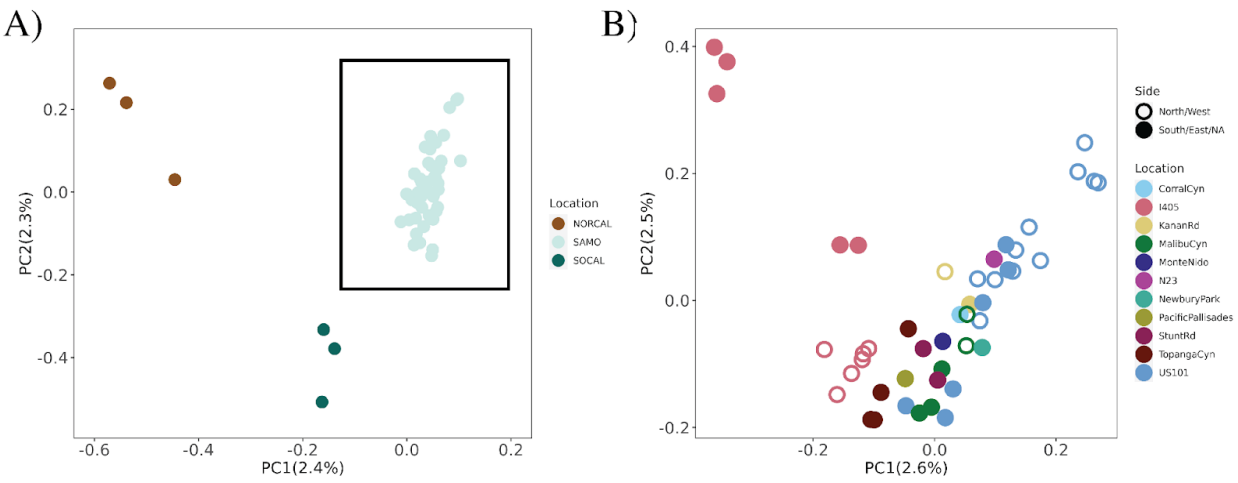

Figure 1-2. A) PCA of 53 samples colored by sampling location, black box indicating samples from panel B, and B) PCA of 46 samples excluding outgroup samples from northern and southern California, colored by sampling location, shape indicates side of the road the sample was collected from if applicable.

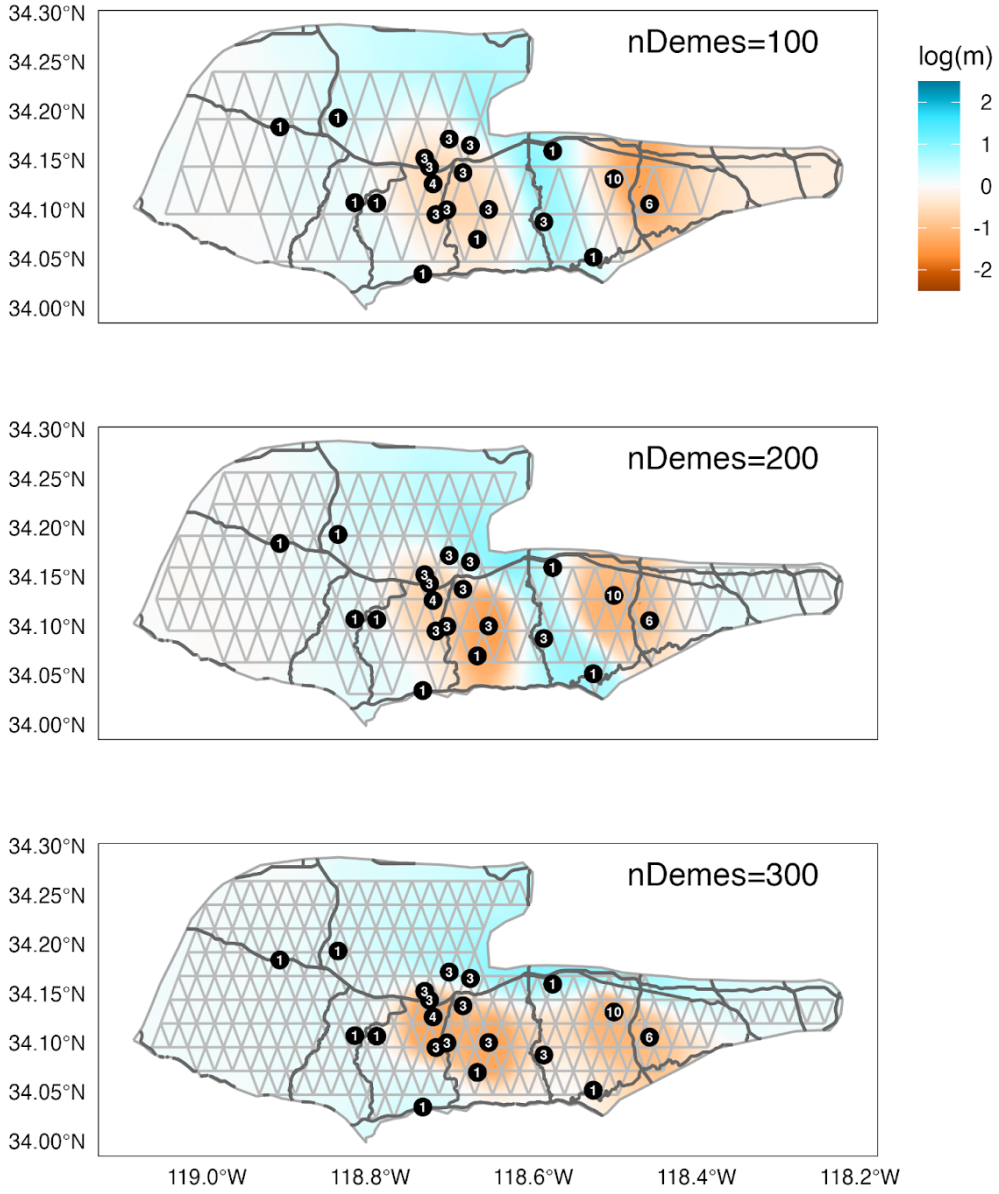

Figure 1-3. Maps of the Santa Monica Mountains area with effective migration surface estimated using EEMS. Warmer colors indicate below average migration and cooler colors indicate above average migration. Black circles with numbers indicate the sampling locations and number of samples.

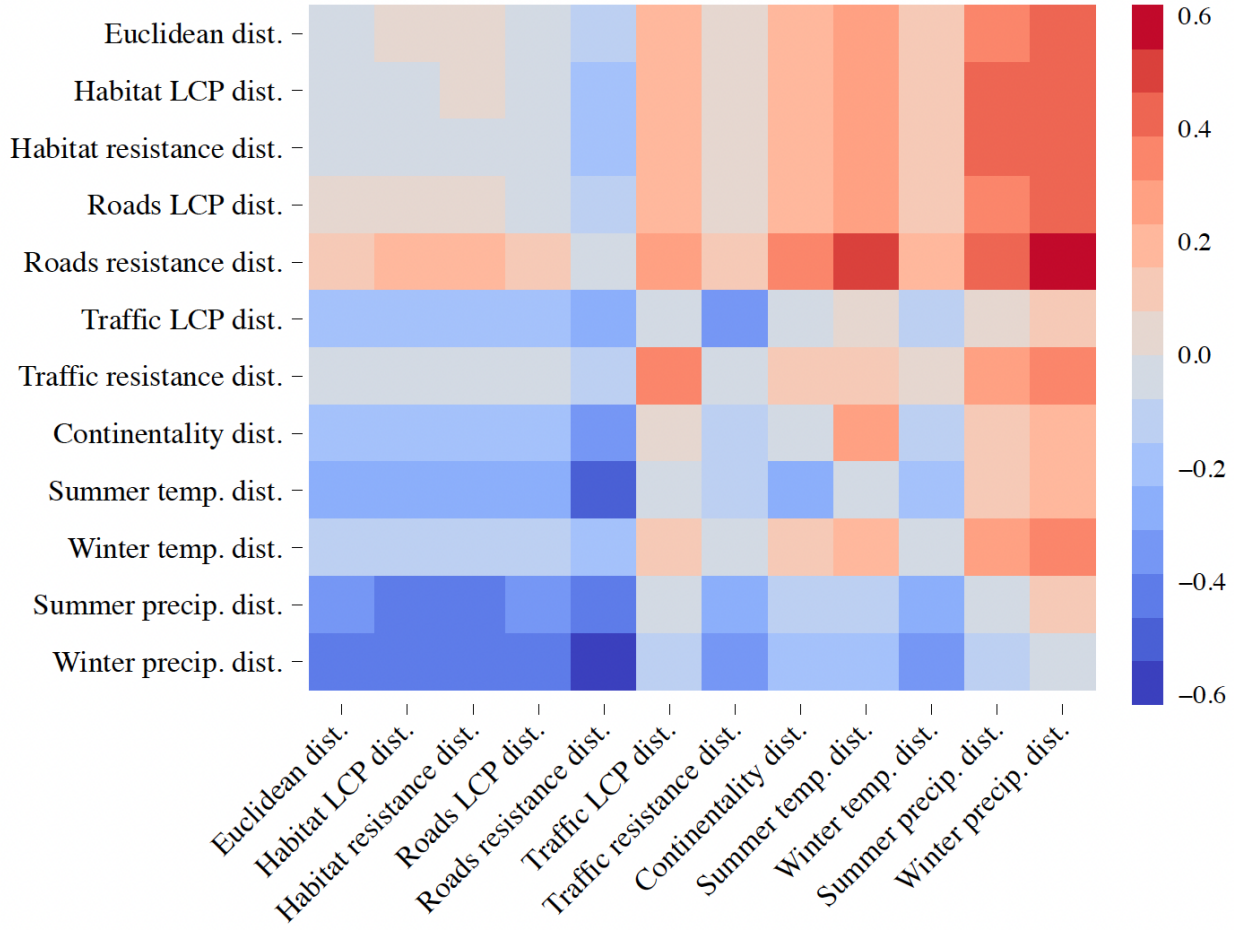

Figure 1-4. Heatmap summarizing Reciprocal Causal Modeling outputs. Rows represent focal variables while columns represent alternate variables. For any given comparison within a row, warmer colors indicate more support for the focal variable in explaining patterns of genetic differentiation while partialling out the effect of the alternate variable. Values in scale correspond to correlation coefficients of partial mantel tests.

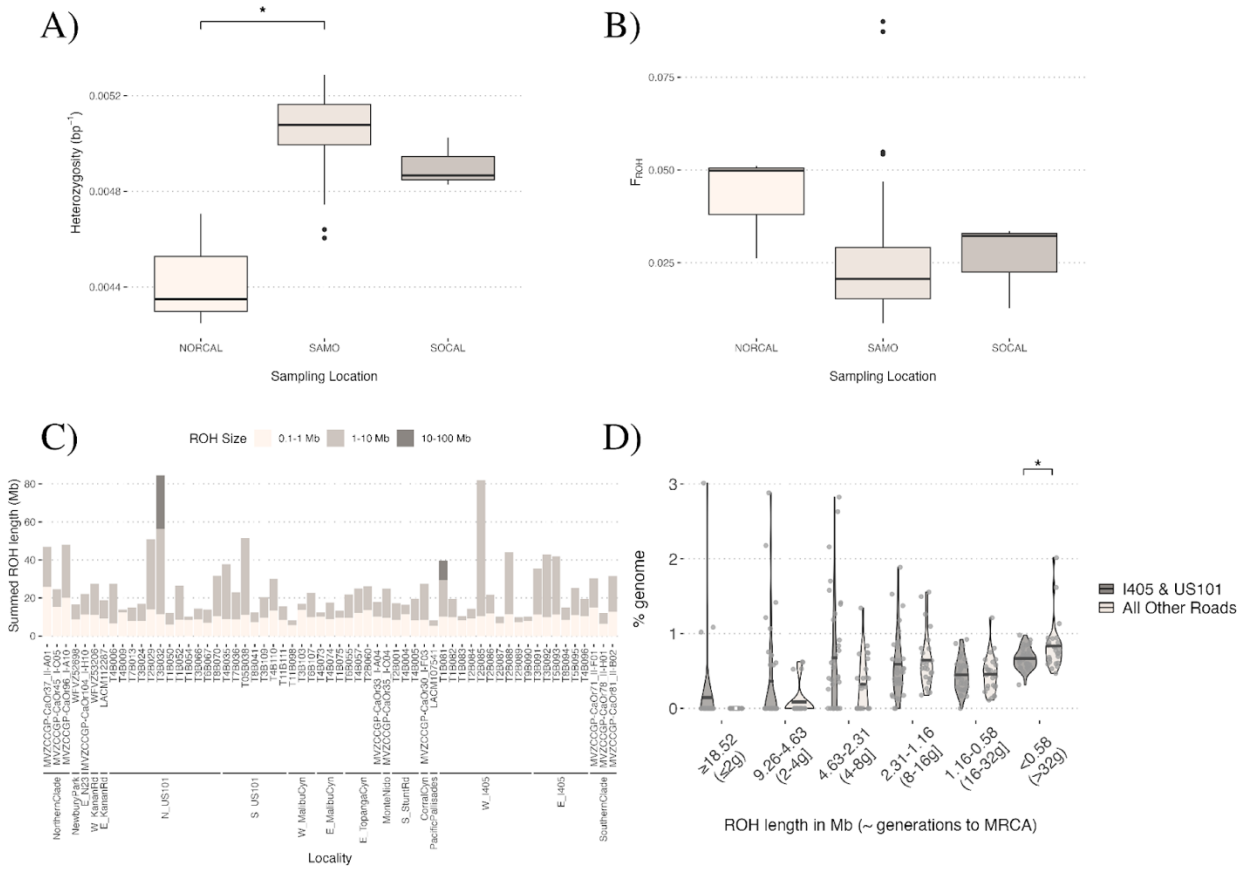

Figure 1-5. Genetic diversity across sampling locations, including A) Genome-wide autosomal heterozygosity, B) proportion of the genome in runs of homozygosity ≥ 100 Kb (F_{ROH}), C) Summed lengths of runs of homozygosity, and D) T_{MRCA} of ROHs for samples located on either side of the US101 and I405 freeways compared to samples from all other roadways.

APPENDIX 1-I: SUPPLEMENTAL METHODS

Predictor variables used in MaxEnt Modeling

Template

As a basis for predictor variables, we first downloaded tiled Digital Elevation Model (DEM) rasters from the USGS 3D Elevation Program (3DEP) at 1 arc-second (https://data.usgs.gov/datacatalog/data/USGS:35f9c4d4-b113-4c8d-8691-47c428c29a5b). We next reprojected and resampled the tiled collection to NAD 1983 California (Teale) Albers Equal Area Conic (EPSG: 3310) at a spatial resolution of 300 x 300 m (all resulting grid cells are square and identical in size—a problem if spatial data are reprojected but not resampled). This projection is optimized for state-wide mapping and area-based calculations, and is recommended by the California Department of Fish and Wildlife. This raster served as the basic template for generating all other spatial data layers (identical origin, resolution, and projection).

Terrain

We generated a terrain ruggedness index, estimated as the mean of absolute differences between the elevation value of each cell and the 8 surrounding cells ("terrain" function, terra R package). We also estimated a heat load index using the R package spatialEco (Evans & Ram, 2021), measuring variation in solar radiation according to the slope and aspect of each cell.

Water

We downloaded surface water vector data from the Hydrosheds database (https://www.hydrosheds.org, accessed Aug 15, 2024). These vector data were buffered by 100 m, intersected with the template raster, and turned into binary raster layers (value of 1 if a raster cell contained water, 0 otherwise). This was performed separately for lotic and lentic data, creating separate raster layers. We also generated a raster measuring the shortest distance from each raster cell's centroid to any surface water.

Land cover

Land cover GIS data was acquired from the Commission for Environmental Cooperation (www.cec.org, accessed Aug 15, 2024), generated using 2020 Landsat satellite imagery. The fine native resolution (30 m) of these data meant no down sampling was required when reprojecting and resampling (bilinear method) to match our template rasters. Fifteen land cover categories are present within the study area, including different types of human land use (e.g., "urban" and "cropland").

Vegetation

We quantified the amount of photosynthetic activity on the landscape using the enhanced vegetation index (EVI) (Jiang et al., 2008). This was generated following MacDonald et al. (In Reivew) using the entire MODIS AQUA (MYD09A1) (Vermote, 2021a) and TERRA (MOD09A1) (Vermote, 2021b) collection. We generated median EVI values from 2000 to 2023 using Google Earth Engine (Gorelick et al., 2017). We chose median values rather than mean because the annual distribution of EVI values are often moderately to highly skewed (Dong et al., 2019)

Urbanization

As an index of variation in urbanization, we downloaded Nighttime Light maps from the Earth Observation Group's Version 4 DMSP-OLS Nighttime Lights Time Series (Elvidge et al., 1997; Baugh et al., 2010; native resolution of 30 arc-seconds). From this dataset, we reprojected and resampled the average of the visible band digital number values to match template rasters. Values ranged from 0-63, with no missing data present within the study extent.

Environmental data

We downloaded data for five environmental variables, generated by MacDonald et al. (In Review) using ClimateNA v7.30 software (Wang et al., 2016). Variables included mean temperature of the warmest quarter, mean temperature of the coldest quarter, difference between mean temperatures of the warmest

and coldest quarters (WorldClim) (hereafter, “continentality”), mean precipitation of the warmest quarter, and mean precipitation of the coldest quarter. We selected these variables to best capture environmental variation while minimizing collinearity ($|r| < 0.75$). These data were generated using an origin, resolution, and projection identical to our template, meaning no processing was required. These layers are approximately 10X finer than WorldClim 2.1 data (Fick & Hijmans, 2017), conferring greater power to resolve heterogeneity in habitat suitability at a fine spatial scale.

Creating continuous layers of traffic volume

To create a continuous layer of traffic volume for use in the RCM analysis, we first overlaid values of “ahead AADT” (estimated in a single direction and useful estimators when road widths are relatively uniform) from the California Department of Transportation (https://gisdata-caltrans.opendata.arcgis.com/datasets/d8833219913c44358f2a9a71bda57f76_0/about) for the year 2022, on to our roads vector layer using QGIS v. 3.20.2-Odense (QGIS 2024). As these data are available as multiple point estimates where traffic volume is recorded, rather than as a continuous layer, we interpolated the traffic volume along the entire road by separating primary and secondary roads and individually performed an inverse distance weighted interpolation across the study using the five nearest neighbor “ahead AADT” points. We then clipped this interpolated surface to the polygon representing primary and secondary roadways (generated by buffering the centerline of the roadway by 50-meters in each direction) and set all background points equal to zero. This surface was then used as a predictor, along with other rasters representing continuous environmental variation, in downstream migration and modeling methods.

APPENDIX 1-II: SUPPLEMENTAL FIGURES

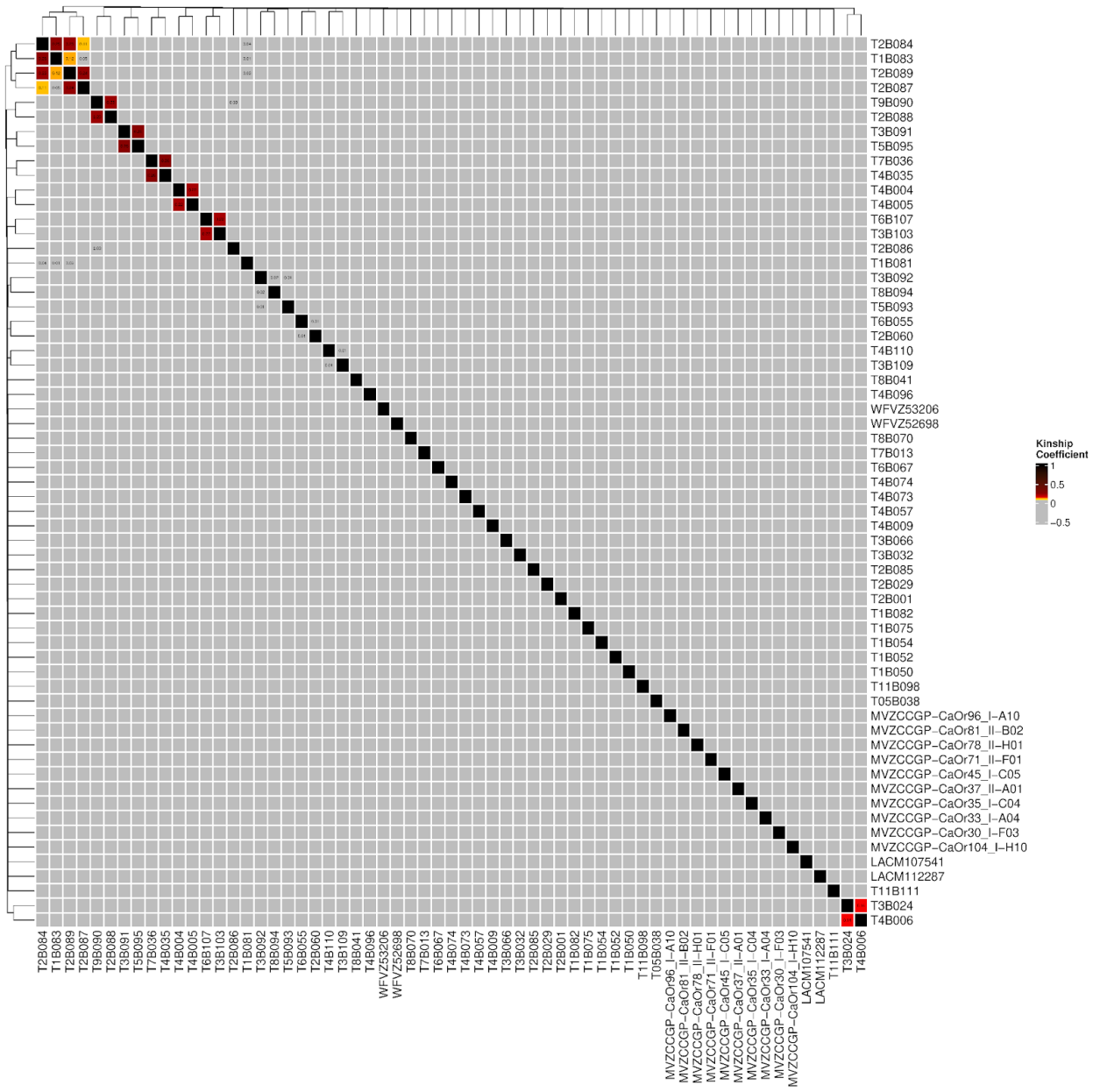

Figure S1-1. Pairwise identity by descent (IBD) matrix of quail samples. Samples in yellow indicate individuals with second-degree relationships and samples in red indicate samples with first-degree relationships. Self-comparisons in dark gray.

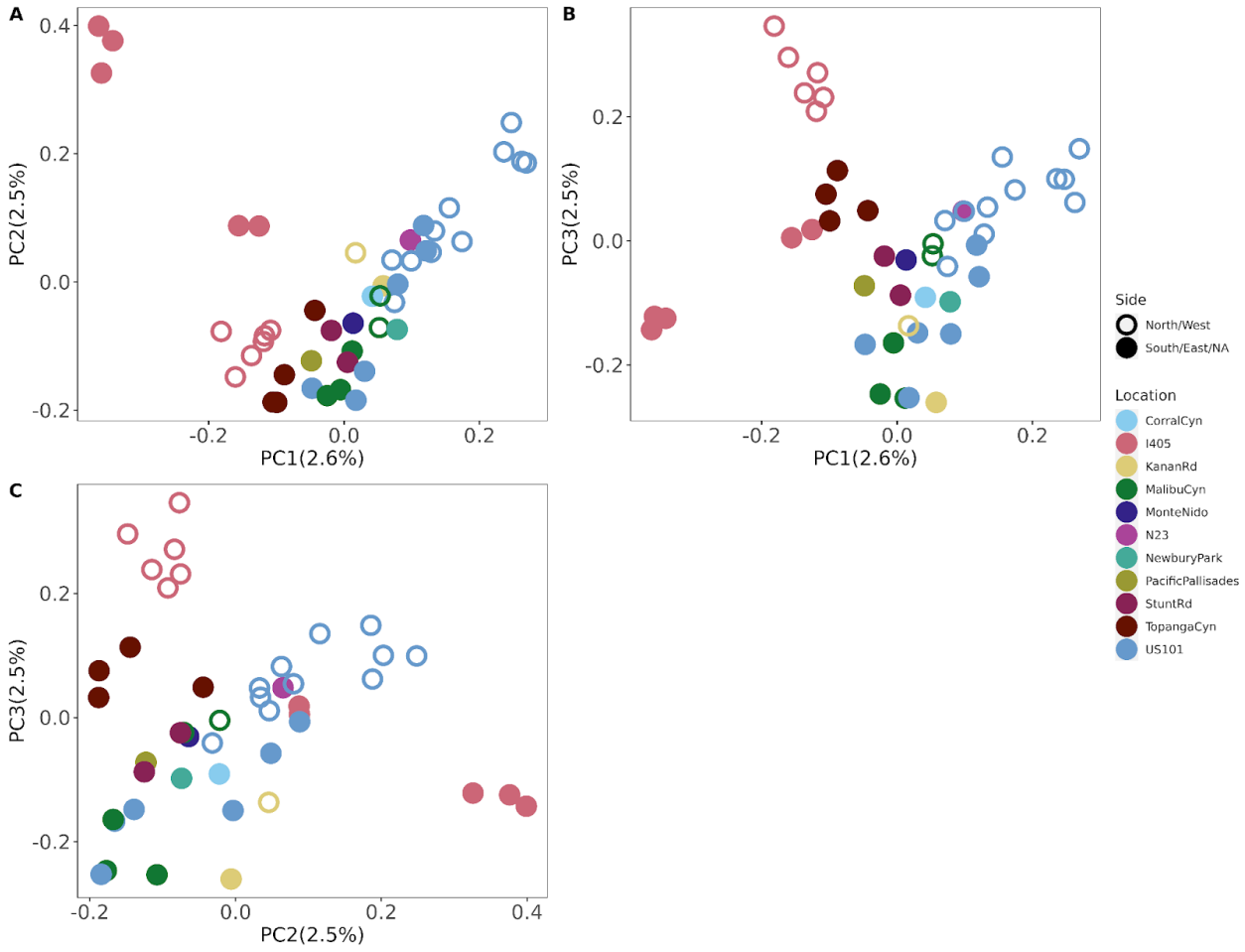

Figure S1-2. Principal component analysis of 46 samples comparing the first three principal components. Points are colored by sampling location, shape indicates side of the road the sample was collected from if applicable.

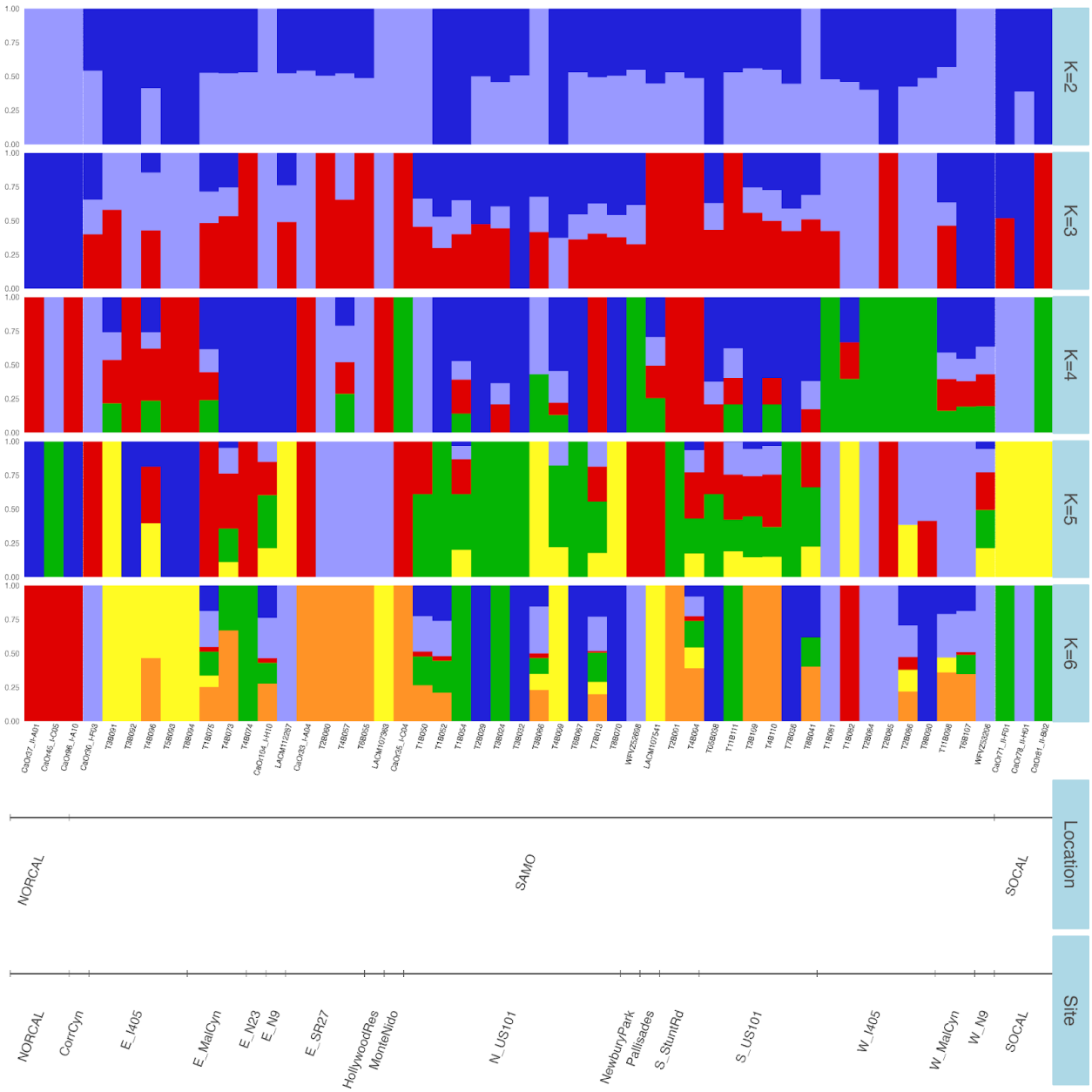

Figure S1-3. Admixture output for k values 1 - 6. Best supported k value = 1.

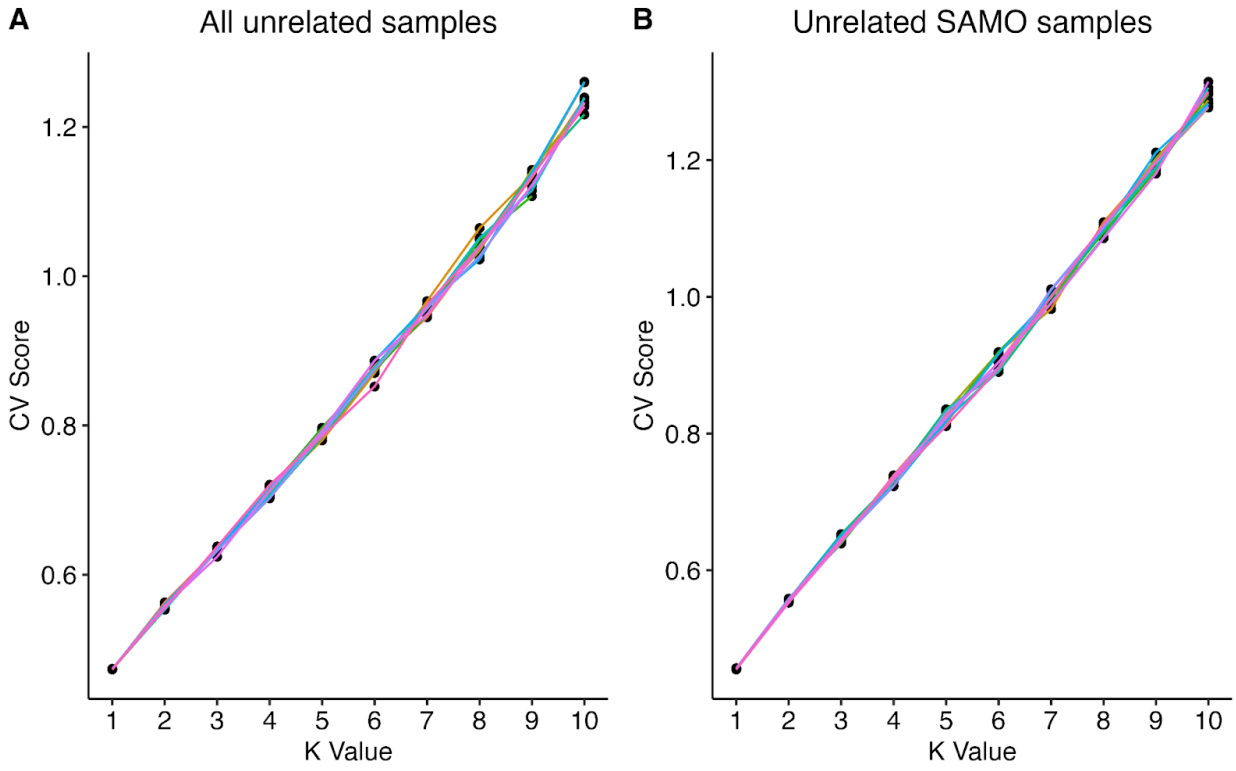

Figure S1-4. Cross validation score for each K value for A) all unrelated quail samples, and B) unrelated quail samples from the Santa Monica Mountains. Lower CV score indicates better model fit.

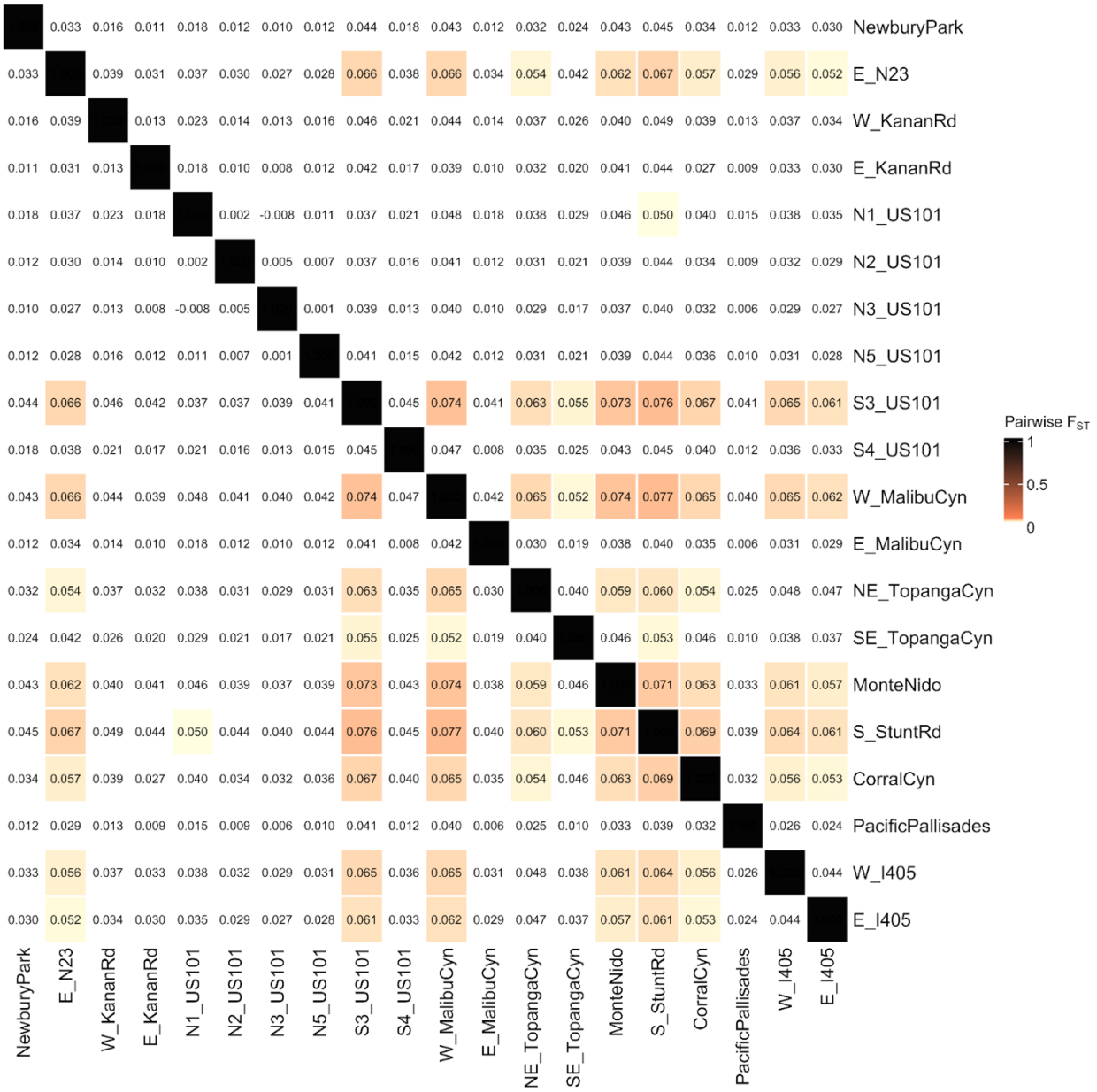

Figure S1-5. Pairwise F_{ST} heatmap of quail samples. Self-comparisons in black along the diagonal.

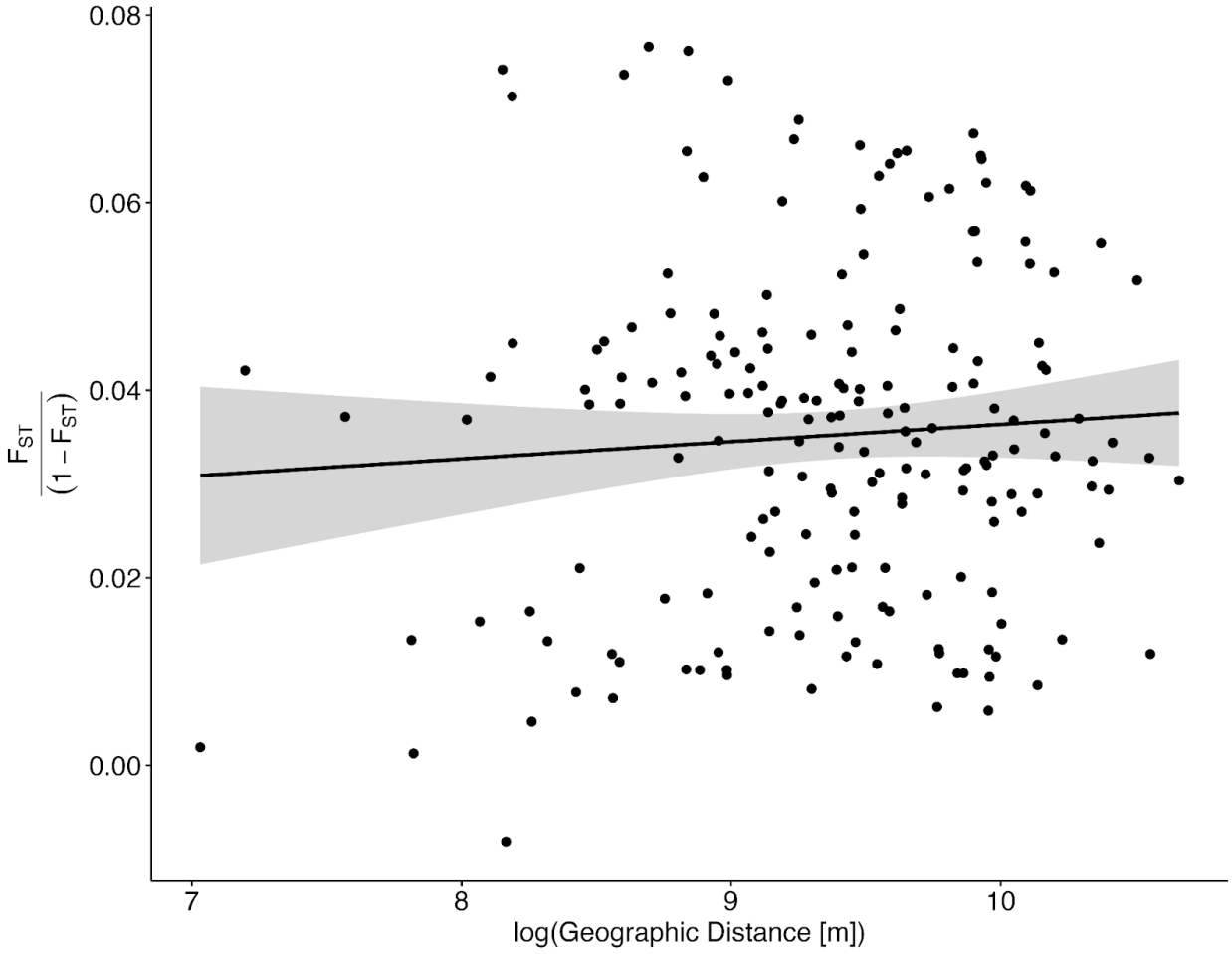

Figure S1-6. Scatter plot with fit line of pairwise linearized F_{ST} values and log transformed geographic distance from 46 samples

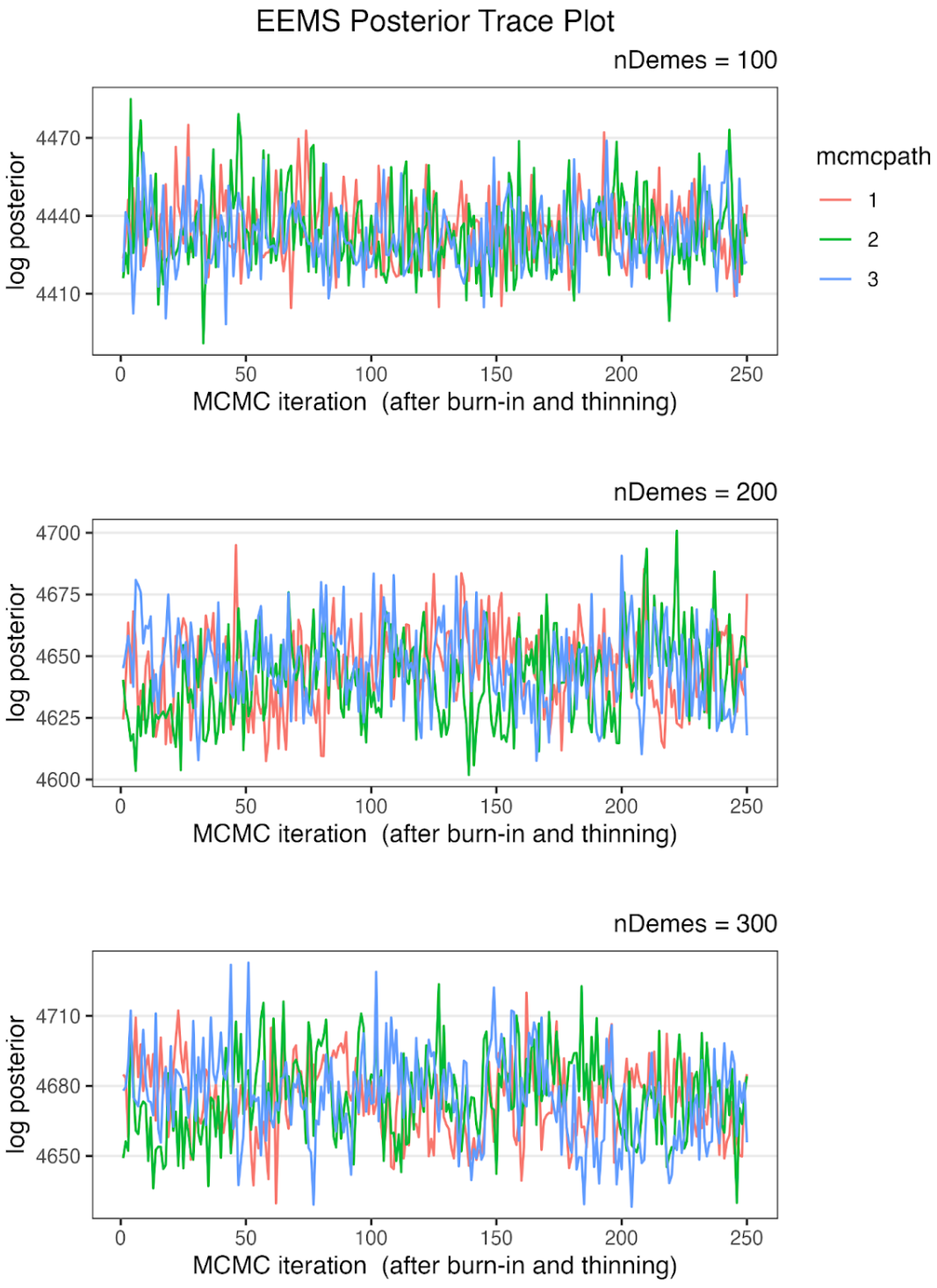

Figure S1-7. Posterior trace plots showing model convergence for three MCMC runs each of three different numbers of nDemes (100 - 300).

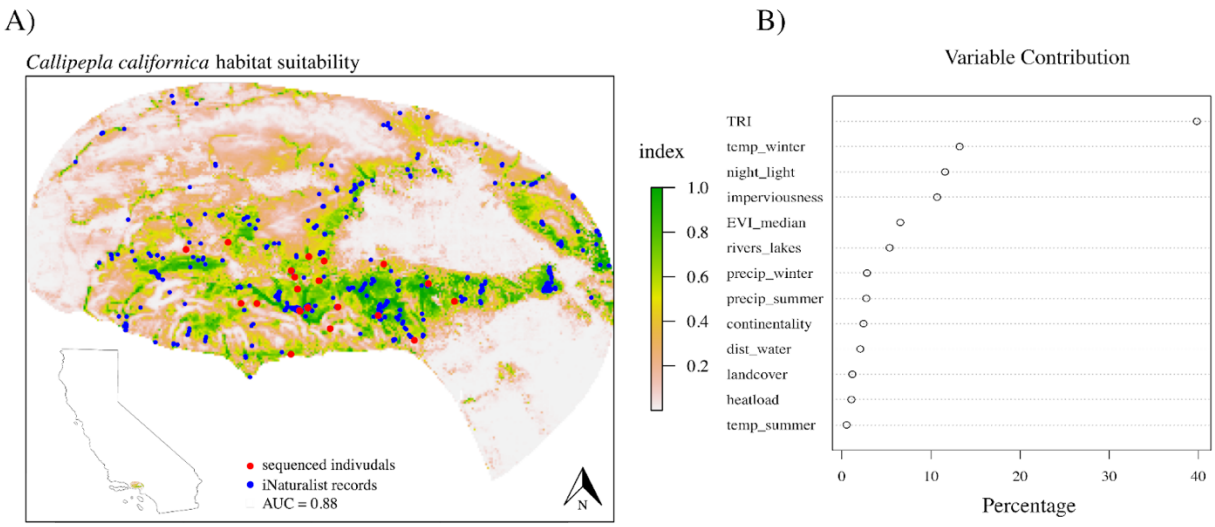

Figure S1-8. Results of ecological niche modeling including A) map of study extent with predictive suitability surface built using locations of quail from this study (red points) and from publicly accessible community science data from iNaturalist (blue points), with more suitable habitat in green and less suitable habitat in red, and B) percentage contribution of each landscape and environmental variable included in the MaxEnt model.

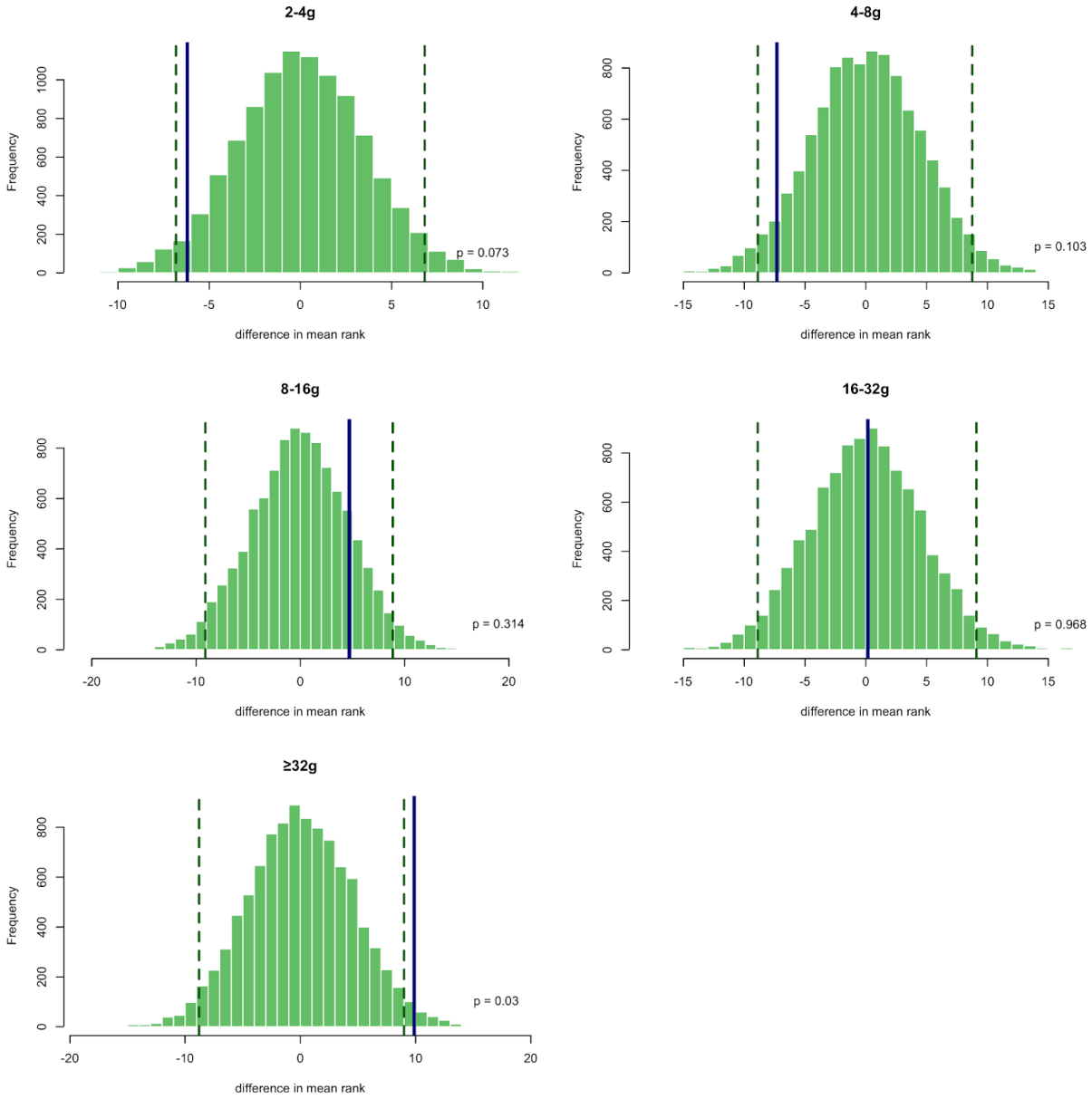

Figure S1-9. Histograms of bootstrapped distribution of mean F_{ROH} lengths by age of autozygous segments. Dotted lines indicate the 5% and 95% percentile and the blue line indicates the observed mean difference between F_{ROH} lengths between samples from the US101 and I405 freeways and all other sampling locations in the Santa Monica Mountains.

APPENDIX 1-III: SUPPLEMENTAL TABLES

Table S1-1. Variables used in the ecological niche model listed in order of their percent contribution and permutational importance.

Variable	Percent Contribution	Permutation Importance
TRI	39.8	29.7
temp_winter	13.2	9.2
night_light	11.6	15.7
imperviousness	10.7	19.7
EVI_median	6.5	9.2
rivers_lakes	5.4	1.6
precip_winter	2.8	1.2
precip_summer	2.7	4.2
continentality	2.4	3
dist_water	2.1	2.7
landcover	1.2	0.5
heatload	1.1	0.1
temp_summer	0.5	3.3

Table S1-2. Distance metrics used in the Reciprocal Causal Modeling analysis and their relative support values.

Distance Measure	Relative Support
roads_resistance_dist	2.987
roads_least_cost_dist	1.478
habitat_least_cost_dist	1.461
euclidean_dist	1.441
habitat_resistance_dist	1.396
traffic_resistance_dist	0.948
temp_winter_dist	0.400
continentality_dist	-0.625
traffic_least_cost_dist	-1.337
temp_summer_dist	-1.946
precip_summer_dist	-2.733
precip_winter_dist	-3.472

REFERENCES

Ahlborn, G., & Johnson, N. (n.d.). *California Quail Life History Account*. California Department of Fish and Wildlife.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* [Computer software]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Aronson, M. F. J., La Sorte, F. A., Nilon, C. H., Katti, M., Goddard, M. A., Lepczyk, C. A., Warren, P. S., Williams, N. S. G., Cilliers, S., Clarkson, B., Dobbs, C., Dolan, R., Hedblom, M., Klotz, S., Kooijmans, J. L., Kühn, I., MacGregor-Fors, I., McDonnell, M., Mörtberg, U., … Winter, M. (2014). A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), 20133330. https://doi.org/10.1098/rspb.2013.3330

Benham, P. M., Cicero, C., Escalona, M., Beraut, E., Marimuthu, M. P. A., Nguyen, O., Nachman, M. W., & Bowie, R. C. K. (2023). A highly contiguous genome assembly for the California quail (*Callipepla californica*). *Journal of Heredity*, *114*(4), 418–427. https://doi.org/10.1093/jhered/esad008

Beninde, J., Delaney, T. W., Gonzalez, G., & Shaffer, H. B. (2023). Harnessing iNaturalist to quantify hotspots of urban biodiversity: The Los Angeles case study. *Frontiers in Ecology and Evolution*, *11*, 983371. https://doi.org/10.3389/fevo.2023.983371

Beninde, J., Feldmeier, S., Werner, M., Peroverde, D., Schulte, U., Hochkirch, A., & Veith, M. (2016). Cityscape genetics: Structural vs. functional connectivity of an urban lizard population. *Molecular Ecology*, *25*(20), 4984–5000. https://doi.org/10.1111/mec.13810

Beninde, J., Wittische, J., & Frantz, A. C. (2024). Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric. *Molecular Ecology Resources*, *24*(1), e13831. https://doi.org/10.1111/1755-0998.13831

Benítez-López, A., Alkemade, R., & Verweij, P. A. (2010). The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. *Biological Conservation*, *143*(6), 1307–1316. https://doi.org/10.1016/j.biocon.2010.02.009

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Research*, *23*(9), 1514–1521. https://doi.org/10.1101/gr.154831.113

Blackwell, B. F., DeVault, T. L., & Seamans, T. W. (2015). Understanding and Mitigating the Negative Effects of Road Lighting on Ecosystems. In R. Van Der Ree, D. J. Smith, & C. Grilo (Eds.), *Handbook of Road Ecology* (1st ed., pp. 143–150). Wiley. https://doi.org/10.1002/9781118568170.ch18

Bolger, D. T., Scott, T. A., & Rotenberry, J. T. (2001). Use of corridor-like landscape structures by bird and small mammal species. *Biological Conservation*, *102*(2), 213–224. https://doi.org/10.1016/S0006-3207(01)00028-3

Brennan, L. A., & Covington, W. (1994). Broad-scale population declines in four species of North American quail: An examination of possible causes. *Sustainable Ecological Systems: Implementing an Ecological Approach to Land Management. USDA Forest Service General Technical Report RM-247*, 44–50.

Brüniche-Olsen, A., Kellner, K. F., Belant, J. L., & DeWoody, J. A. (2021). Life-history traits and habitat availability shape genomic diversity in birds: Implications for conservation. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1961), 20211441. https://doi.org/10.1098/rspb.2021.1441

Caballero, A., Villanueva, B., & Druet, T. (2021). On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evolutionary Applications*, *14*(2), 416–428. https://doi.org/10.1111/eva.13126

Caldwell, M. R., & Klip, J. M. K. (2020). Wildlife Interactions within Highway Underpasses. *The Journal of Wildlife Management*, *84*(2), 227–236. https://doi.org/10.1002/jwmg.21801

California Department of Transportation. (2022). *Traffic Volumes AADT* [Dataset]. https://gis.data.ca.gov/datasets/d8833219913c44358f2a9a71bda57f76_0/explore?locatio=36.988062%2C-119.281935%2C6.20

Calmettes, G., John, S. A., Weiss, J. N., & Ribalet, B. (2013). Hexokinase–mitochondrial interactions regulate glucose metabolism differentially in adult and neonatal cardiac myocytes. *Journal of General Physiology*, *142*(4), 425–436. https://doi.org/10.1085/jgp.201310968

Cassin-Sackett, L., Welch, A. J., Venkatraman, M. X., Callicrate, T. E., & Fleischer, R. C. (2019). The Contribution of Genomics to Bird Conservation. In R. H. S. Kraus (Ed.), *Avian Genomics in Ecology and Evolution* (pp. 295–330). Springer International Publishing. https://doi.org/10.1007/978-3-030-16477-5_10

Cheek, R. G., Forester, B. R., Salerno, P. E., Trumbo, D. R., Langin, K. M., Chen, N., Scott Sillett, T., Morrison, S. A., Ghalambor, C. K., & Chris Funk, W. (2022). Habitat-linked genetic variation supports microgeographic adaptive divergence in an island-endemic bird species. *Molecular Ecology*, *31*(10), 2830–2846. https://doi.org/10.1111/mec.16438

Chen, N., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., & Clark, A. G. (2016). Genomic Consequences of Population Decline in the Endangered Florida Scrub-Jay. *Current Biology*, *26*(21), 2974–2979. https://doi.org/10.1016/j.cub.2016.08.062

Cooke, S. C., Balmford, A., Donald, P. F., Newson, S. E., & Johnston, A. (2020). Roads as a contributor to landscape-scale variation in bird communities. *Nature Communications*, *11*(1), 3125. https://doi.org/10.1038/s41467-020-16899-x

Cros, E., Ng, E. Y. X., Oh, R. R. Y., Tang, Q., Benedick, S., Edwards, D. P., Tomassi, S., Irestedt, M., Ericson, P. G. P., & Rheindt, F. E. (2020). Fine-scale barriers to connectivity across a fragmented South-East Asian landscape in six songbird species. *Evolutionary Applications*, *13*(5), 1026–1036. https://doi.org/10.1111/eva.12918

Cushman, S., Wasserman, T., Landguth, E., & Shirk, A. (2013). Re-Evaluating Causal Modeling with Mantel Tests in Landscape Genetics. *Diversity*, *5*(1), 51–72. https://doi.org/10.3390/d5010051

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Jonge, M. M. J., Gallego-Zamorano, J., Huijbregts, M. A. J., Schipper, A. M., & Benítez-López, A. (2022). The impacts of linear infrastructure on terrestrial vertebrate populations: A trait-based approach. *Global Change Biology*, *28*(24), 7217–7233. https://doi.org/10.1111/gcb.16450

Delaney, K. S., Riley, S. P. D., & Fisher, R. N. (2010). A Rapid, Strong, and Convergent Genetic Response to Urban Habitat Fragmentation in Four Divergent and Widespread Vertebrates. *PLoS ONE*, *5*(9), e12767. https://doi.org/10.1371/journal.pone.0012767

Denneboom, D., Bar-Massada, A., & Shwartz, A. (2021). Factors affecting usage of crossing structures by wildlife – A systematic review and meta-analysis. *Science of The Total Environment*, *777*, 146061. https://doi.org/10.1016/j.scitotenv.2021.146061

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dewitz, J. (2021). *National Land Cover Database (NLCD) 2019 Products*. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54

DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, *30*(17), 4147–4154. https://doi.org/10.1111/mec.16051

Dong, C., MacDonald, G., Okin, G. S., & Gillespie, T. W. (2019). Quantifying drought sensitivity of Mediterranean climate vegetation to recent warming: A case study in Southern California. *Remote Sensing*, *11*(24), 2902. https://doi.org/10.3390/rs11242902

Duntsch, L., Whibley, A., De Villemereuil, P., Brekke, P., Bailey, S., Ewen, J. G., & Santure, A. W. (2023). Genomic signatures of inbreeding depression for a threatened Aotearoa New Zealand passerine. *Molecular Ecology*, *32*(8), 1893–1907. https://doi.org/10.1111/mec.16855

Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., … Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics*, *1*(1), 100002. https://doi.org/10.1016/j.xgen.2021.100002

Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 239–259. https://doi.org/10.1146/annurev-ecolsys-110411-160327

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogrammetric Engineering and Remote Sensing*, *63*(6), 727-734.

Evans, J. S., & Ram, K. (2021). Package 'spatialEco'. R CRAN Project.

Femerling, G., Van Oosterhout, C., Feng, S., Bristol, R. M., Zhang, G., Groombridge, J., P. Gilbert, M. T., & Morales, H. E. (2023). Genetic load and adaptive potential of a recovered avian species that narrowly avoided extinction. *Molecular Biology and Evolution*, *40*(12), msad256.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302-4315. https://doi.org/10.1002/joc.5086

Fies, M. L., Puckett, K. M., & Larson-Brogdon, B. (2002). Breeding Season Movements and Dispersal of Northern Bobwhites in Fragmented Habitats of Virginia. *National Quail Symposium Proceedings*, *5*(1). https://doi.org/10.7290/nqsp050tez

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Forman, R. T. T., & Alexander, L. E. (1998). Roads and Their Major Ecological Effects. *Annual Review of Ecology and Systematics*, *29*(1), 207–231. https://doi.org/10.1146/annurev.ecolsys.29.1.207

Fraser, A., & Chester, M. V. (2016). Environmental and Economic Consequences of Permanent Roadway Infrastructure Commitment: City Road Network Lifecycle Assessment and Los Angeles County. *Journal of Infrastructure Systems*, *22*(1), 04015018. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000271

Fraser, D. L., Ironside, K., Wayne, R. K., & Boydston, E. E. (2019). Connectivity of mule deer (*Odocoileus hemionus*) populations in a highly fragmented urban landscape. *Landscape Ecology*, *34*(5), 1097–1115. https://doi.org/10.1007/s10980-019-00824-9

Fu, C.-Z., Guang, X.-M., Wan, Q.-H., & Fang, S.-G. (2019). Genome Resequencing Reveals Congenital Causes of Embryo and Nestling Death in Crested Ibis (*Nipponia nippon*). *Genome Biology and Evolution*, *11*(8), 2125–2135. https://doi.org/10.1093/gbe/evz149

Fusco, N. A., Carlen, E. J., & Munshi-South, J. (2021). Urban Landscape Genetics: Are Biologists Keeping Up with the Pace of Urbanization? *Current Landscape Ecology Reports*, *6*(2), 35–45. https://doi.org/10.1007/s40823-021-00062-3

Georges, A., Mijangos, L., Kilian, A., Patel, H., Aitkens, M., & Gruber, B. (2023). *Distances and their visualization in studies of spatial-temporal genetic variation using single nucleotide polymorphisms (SNPs)*. https://doi.org/10.1101/2023.03.22.533737

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, *202*, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Groenen, M. A. M., Wahlberg, P., Foglio, M., Cheng, H. H., Megens, H.-J., Crooijmans, R. P. M. A., Besnier, F., Lathrop, M., Muir, W. M., Wong, G. K.-S., Gut, I., & Andersson, L. (2009). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research*, *19*(3), 510–519. https://doi.org/10.1101/gr.086538.108

Habrich, A. K., Lawrence, E. R., & Fraser, D. J. (2021). Varying genetic imprints of road networks and human density in North American mammal populations. *Evolutionary Applications*, *14*(6), 1659–1672. https://doi.org/10.1111/eva.13232

Harrisson, K. A., Magrath, M. J. L., Yen, J. D. L., Pavlova, A., Murray, N., Quin, B., Menkhorst, P., Miller, K. A., Cartwright, K., & Sunnucks, P. (2019). Lifetime Fitness Costs of Inbreeding and Being Inbred in a Critically Endangered Bird. *Current Biology*, *29*(16), 2711-2717.e4. https://doi.org/10.1016/j.cub.2019.06.064

Hijmans, R. J. (2024a). *raster: Geographic Data Analysis and Modeling*. https://rspatial.org/raster

Hijmans, R. J. (2024b). *terra: Spatial Data Analysis*. https://github.com/rspatial/terra

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2024). *dismo: Species Distribution Modeling*. https://github.com/rspatial/dismo

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583–589. https://doi.org/10.1093/genetics/132.2.583

Jaeger, J. A. G., Bowman, J., Brennan, J., Fahrig, L., Bert, D., Bouchard, J., Charbonneau, N., Frank, K., Gruber, B., & Von Toschanowitz, K. T. (2005). Predicting when animal populations are at risk from roads: An interactive model of road avoidance behavior. *Ecological Modelling*, *185*(2–4), 329–348. https://doi.org/10.1016/j.ecolmodel.2004.12.015

Jaeger, J. A. G., Schwarz-von Raumer, H.-G., Esswein, H., Müller, M., & Schmidt-Lüttmann, M. (2007). Time Series of Landscape Fragmentation Caused by Transportation Infrastructure and Urban Development: A Case Study from Baden-Württemberg, Germany. *Ecology and Society*, *12*(1), art22. https://doi.org/10.5751/ES-01983-120122

Jiao, X., Wu, L., Zhang, D., Wang, H., Dong, F., Yang, L., Wang, S., Amano, H. E., Zhang, W., Jia, C., Rheindt, F. E., Lei, F., & Song, G. (2024). Landscape Heterogeneity Explains the Genetic Differentiation of a Forest Bird across the Sino-Himalayan Mountains. *Molecular Biology and Evolution*, *41*(3), msae027. https://doi.org/10.1093/molbev/msae027

Johnson, N. A., & Lachance, J. (2012). The genetics of sex chromosomes: Evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences*, *1256*(1). https://doi.org/10.1111/j.1749-6632.2012.06748.x

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

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. https://CRAN.R-project.org/package=rstatix

Kimmitt, A. A., Pegan, T. M., Jones, A. W., Winker, K., & Winger, B. M. (2024). How Veeries vary: Whole genome sequencing resolves genetic structure in a long-distance migratory bird. *Ornithology*, *141*(2), ukad061. https://doi.org/10.1093/ornithology/ukad061

Kociolek, A., Grilo, C., & Jacobson, S. (2015). Flight Doesn't Solve Everything: Mitigation of Road Impacts on Birds. In R. Van Der Ree, D. J. Smith, & C. Grilo (Eds.), *Handbook of Road Ecology* (1st ed., pp. 281–289). Wiley. https://doi.org/10.1002/9781118568170.ch33

Kociolek, A. V., Clevenger, A. P., St. Clair, C. C., & Proppe, D. S. (2011). Effects of Road Networks on Bird Populations: Effects of Roads on Birds. *Conservation Biology*, *25*(2), 241-249. https://doi.org/10.1111/j.1523-1739.2010.01635.x

Kroeger, S. B., Hanslin, H. M., Lennartsson, T., D'Amico, M., Kollmann, J., Fischer, C., Albertsen, E., & Speed, J. D. M. (2022). Impacts of roads on bird species richness: A meta-analysis considering road types, habitats and feeding guilds. *Science of The Total Environment*, *812*, 151478. https://doi.org/10.1016/j.scitotenv.2021.151478

Kyriazis, C. C., Beichman, A. C., Brzeski, K. E., Hoy, S. R., Peterson, R. O., Vucetich, J. A., Vucetich, L. M., Lohmueller, K. E., & Wayne, R. K. (2023). Genomic underpinnings of population persistence in Isle Royale moose. *Molecular Biology and Evolution*, *40*(2), msad021.

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, *5*(1), 33–47. https://doi.org/10.1002/evl3.209

La Sorte, F. A., Johnston, A., Rodewald, A. D., Fink, D., Farnsworth, A., Van Doren, B. M., Auer, T., & Strimas-Mackey, M. (2022). The role of artificial light at night and road density in predicting the seasonal occurrence of nocturnally migrating birds. *Diversity and Distributions*, *28*(5), 992–1009. https://doi.org/10.1111/ddi.13499

Lees, A. C., Haskell, L., Allinson, T., Bezeng, S. B., Burfield, I. J., Renjifo, L. M., Rosenberg, K. V., Viswanathan, A., & Butchart, S. H. M. (2022). State of the World's Birds. *Annual Review of Environment and Resources*, *47*(1), 231–260. https://doi.org/10.1146/annurev-environ-112420-014642

Leopold, A. S. (1985). *The California quail*. University of California Press.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Loss, S. R., Will, T., & Marra, P. P. (2014). Estimation of bird-vehicle collision mortality on U.S. roads. *The Journal of Wildlife Management*, *78*(5), 763–771. https://doi.org/10.1002/jwmg.721

Luna, L. W., Williams, L. M., Duren, K., Tyl, R., Toews, D. P. L., & Avery, J. D. (2023). Whole genome assessment of a declining game bird reveals cryptic genetic structure and insights for population management. *Molecular Ecology*, *32*(20), 5498–5513. https://doi.org/10.1111/mec.17129

MacDonald, Z. G., Dupuis, J. R., Davis, C. S., Acorn, J. H., Nielsen, S. E., & Sperling, F. A. H. (2020). Gene flow and climate-associated genetic variation in a vagile habitat specialist. *Molecular Ecology*, *29*(20), 3889–3906. https://doi.org/10.1111/mec.15604

MacDonald, Z. G., Snape, K. L., Roe, A. D., & Sperling, F. A. H. (2022). Host association, environment, and geography underlie genomic differentiation in a major forest pest. *Evolutionary Applications*, *15*(11), 1749–1765. https://doi.org/10.1111/eva.13466

Madadi, H., Moradi, H., Soffianian, A., Salmanmahiny, A., Senn, J., & Geneletti, D. (2017). Degradation of natural habitats by roads: Comparing land-take and noise effect zone. *Environmental Impact Assessment Review*, *65*, 147–155. https://doi.org/10.1016/j.eiar.2017.05.003

Malpica, A., & González, C. (2023). Landscape anthropization explains the genetic structure of an endemic Mexican bird (*Thryophilus sinaloa*: Troglodytidae) across the tropical dry forest biodiversity hotspot. *Landscape Ecology*, *38*(12), 3249–3268. https://doi.org/10.1007/s10980-023-01777-w

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. *Molecular Ecology*, *24*(14), 3628–3638. https://doi.org/10.1111/mec.13258

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an island-colonizing bird. *Molecular Ecology*, *32*(8), 1972–1989. https://doi.org/10.1111/mec.16865

Mathur, S., & DeWoody, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications*, *14*(6), 1540–1557. https://doi.org/10.1111/eva.13216

McCartney-Melstad, E., Vu, J. K., & Shaffer, H. B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Molecular Ecology*, *27*(22), 4430–4443. https://doi.org/10.1111/mec.14892

McClure, C. J. W., Ware, H. E., Carlisle, J., Kaltenecker, G., & Barber, J. R. (2013). An experimental investigation into the effects of traffic noise on distributions of birds: Avoiding the phantom road. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1773), 20132290. https://doi.org/10.1098/rspb.2013.2290

McRae, B. H. (2006). Isolation by Resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, *104*(50), 19885–19890. https://doi.org/10.1073/pnas.0706568104

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., & Schipper, A. M. (2018). Global patterns of current and future road infrastructure. *Environmental Research Letters*, *13*(6), 064006. https://doi.org/10.1088/1748-9326/aabd42

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283– 292.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: High-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Nigenda-Morales, S. F., Lin, M., Nuñez-Valencia, P. G., Kyriazis, C. C., Beichman, A. C., Robinson, J. A., Ragsdale, A. P., Urbán R, J., Archer, F. I., Viloria-Gómora, L., & others. (2023). The genomic footprint of whaling and isolation in fin whale populations. *Nature Communications*, *14*(1), 5465.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292– 294. https://doi.org/10.1093/bioinformatics/btv566

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., Wagner, H., & others. (2013). Package 'vegan.' *Community Ecology Package, Version*, *2*(9), 1–295.

Oldeschulte, D. L., Halley, Y. A., Wilson, M. L., Bhattarai, E. K., Brashear, W., Hill, J., Metz, R. P., Johnson, C. D., Rollins, D., Peterson, M. J., Bickhart, D. M., Decker, J. E., Sewell, J. F., & Seabury, C. M. (2017). Annotated Draft Genome Assemblies for the Northern Bobwhite (*Colinus virginianus*) and the Scaled Quail (*Callipepla squamata*) Reveal Disparate Estimates of Modern Genome Diversity and Historic Effective Population Size. *G3: Genes|Genomes|Genetics*, *7*(9), 3047–3058. https://doi.org/10.1534/g3.117.043083

Owen, J. C. (2011). Collecting, processing, and storing avian blood: A review: Avian Blood Collection Techniques. *Journal of Field Ornithology*, *82*(4), 339–354. https://doi.org/10.1111/j.1557-9263.2011.00338.x

Oyler-McCance, S. J., Oh, K. P., Langin, K. M., & Aldridge, C. L. (2016). A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. *The Auk*, *133*(4), 626–648. https://doi.org/10.1642/AUK-16-49.1

Parris, K. M., & Schneider, A. (2009). Impacts of Traffic Noise and Traffic Volume on Birds of Roadside Habitats. *Ecology and Society*, *14*(1), art29. https://doi.org/10.5751/ES-02761-140129

Pease, K. M., Freedman, A. H., Pollinger, J. P., Mccormack, J. E., Buermann, W., Rodzen, J., Banks, J., Meredith, E., Bleich, V. C., Schaefer, R. J., Jones, K., & Wayne, R. K. (2009). Landscape genetics of California mule deer (*Odocoileus hemionus*): The roles of ecological and historical factors in generating differentiation. *Molecular Ecology*, *18*(9), 1848–1862. https://doi.org/10.1111/j.1365-294X.2009.04112.x

Pebesma, E. J., & Bivand, R. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13.

Pell, S., & Jones, D. (2015). Are wildlife overpasses of conservation value for birds? A study in Australian sub-tropical forest, with wider implications. *Biological Conservation*, *184*, 300–309. https://doi.org/10.1016/j.biocon.2015.02.005

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*, 946–952. https://doi.org/10.1111/1755-0998.12129

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Provost, K., Shue, S. Y., Forcellati, M., & Smith, B. T. (2022). The genomic landscapes of desert birds form over multiple time scales. *Molecular Biology and Evolution*, *39*(10), msac200.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Remon, J., Moulherat, S., Cornuau, J. H., Gendron, L., Richard, M., Baguette, M., & Prunier, J. G. (2022). Patterns of gene flow across multiple anthropogenic infrastructures: Insights from a multi-species approach. *Landscape and Urban Planning*, *226*, 104507. https://doi.org/10.1016/j.landurbplan.2022.104507

Richardson, J. L., Michaelides, S., Combs, M., Djan, M., Bisch, L., Barrett, K., Silveira, G., Butler, J., Aye, T. T., Munshi-South, J., DiMatteo, M., Brown, C., & McGreevy, T. J. (2021). Dispersal ability predicts spatial genetic structure in native mammals persisting across an urbanization gradient. *Evolutionary Applications*, *14*(1), 163–177. https://doi.org/10.1111/eva.13133

Riley, S. P. D., Pollinger, J. P., Sauvajot, R. M., York, E. C., Bromley, C., Fuller, T. K., & Wayne, R. K. (2006). FAST-TRACK: A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*, *15*(7), 1733–1741. https://doi.org/10.1111/j.1365-294X.2006.02907.x

Riley, S. P. D., Serieys, L. E. K., Pollinger, J. P., Sikich, J. A., Dalbeck, L., Wayne, R. K., & Ernest, H. B. (2014). Individual Behaviors Dominate the Dynamics of an Urban Mountain Lion Population Isolated by Roads. *Current Biology*, *24*(17), 1989–1994. https://doi.org/10.1016/j.cub.2014.07.029

Robinson, J. A., Bowie, R. C. K., Dudchenko, O., Aiden, E. L., Hendrickson, S. L., Steiner, C. C., Ryder, O. A., Mindell, D. P., & Wall, J. D. (2021). Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology*, *31*(13), 2939-2946.e5. https://doi.org/10.1016/j.cub.2021.04.035

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, *28*(21), 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Robinson, J., Kyriazis, C. C., Yuan, S. C., & Lohmueller, K. E. (2023). Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annual Review of Animal Biosciences*, *11*(1), 93–114. https://doi.org/10.1146/annurev-animal-080522-093311

Rodríguez-Bardía, M., Fuchs, E. J., Barrantes, G., Madrigal-Brenes, R., & Sandoval, L. (2022). Genetic structure in neotropical birds with different tolerance to urbanization. *Scientific Reports*, *12*(1), 6054. https://doi.org/10.1038/s41598-022-09961-9

Row, J. R., Doherty, K. E., Cross, T. B., Schwartz, M. K., Oyler-McCance, S. J., Naugle, D. E., Knick, S. T., & Fedy, B. C. (2018). Quantifying functional connectivity: The role of breeding habitat, abundance, and landscape features on range-wide gene flow in sage-grouse. *Evolutionary Applications*, *11*(8), 1305–1321. https://doi.org/10.1111/eva.12627

Roy, C. L., & Gregory, A. J. (2019). Landscape genetic evaluation of a tallgrass prairie corridor using the Greater Prairie-chicken (*Tympanuchus cupido*). *Landscape Ecology*, *34*(6), 1425–1443. https://doi.org/10.1007/s10980-019-00862-3

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Rushing, G., Conley, J. T., Whitt, J. G., & Reyna, K. S. (2022). Translocating Wild California Valley Quail to Texas: An Evaluation of Survival, Dispersal, Tracking Efficacy, and Roost Preference. *National Quail Symposium Proceedings*, *9*. https://doi.org/10.7290/nqsp09oDPh

Sato, Y., Humble, E., & Ogden, R. (2023). Genomic data reveal strong differentiation and reduced genetic diversity in island golden eagle populations. *Biological Journal of the Linnean Society*, blad172. https://doi.org/10.1093/biolinnean/blad172

Sawaya, M. A., Kalinowski, S. T., & Clevenger, A. P. (2014). Genetic connectivity for two bear species at wildlife crossing structures in Banff National Park. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), 20131705. https://doi.org/10.1098/rspb.2013.1705

Schmidt, C., Domaratzki, M., Kinnunen, R. P., Bowman, J., & Garroway, C. J. (2020). Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proceedings of the Royal Society B*, *287*: 20192497. http://dx.doi.org/10.1098/rspb.2019.2497

Schmidt, C., & Garroway, C. J. (2021). The population genetics of urban and rural amphibians in North America. *Molecular Ecology*, *30*(16), 3918–3929. https://doi.org/10.1111/mec.16005

Seress, G., & Liker, A. (2015). Habitat urbanization and its effects on birds. *Acta Zoologica Academiae Scientiarum Hungaricae*, *61*(4), 373–408. https://doi.org/10.17109/AZH.61.4.373.2015

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, *69*(1), 82–90. https://doi.org/10.1139/z91-013

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

Shirk, A. J., Cushman, S. A., & Landguth, E. L. (2012). Simulating Pattern-Process Relationships to Validate Landscape Genetic Models. *International Journal of Ecology*, *2012*, 1–8. https://doi.org/10.1155/2012/539109

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2018). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, *18*(1), 55–67. https://doi.org/10.1111/1755-0998.12709

Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*, *47*(1), 264–279.

Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0. 2013–2015*. Seattle, USA.

Smith, D. J., Van Der Ree, R., & Rosell, C. (2015). Wildlife Crossing Structures: An Effective Strategy to Restore or Maintain Wildlife Connectivity Across Roads. In R. Van Der Ree, D. J. Smith, & C. Grilo (Eds.), *Handbook of Road Ecology* (1st ed., pp. 172–183). Wiley. https://doi.org/10.1002/9781118568170.ch21

Smith, H. D., Stormer, F., & Godfrey, R. (1981). *A collapsible quail trap*. USDA Forest Service.

Soanes, K., Taylor, A. C., Sunnucks, P., Vesk, P. A., Cesarini, S., & Van Der Ree, R. (2018). Evaluating the success of wildlife crossing structures using genetic approaches and an experimental design: Lessons from a gliding mammal. *Journal of Applied Ecology*, *55*(1), 129–138. https://doi.org/10.1111/1365-2664.12966

Sonsthagen, S. A., Wilson, R. E., & Talbot, S. L. (2022). Species-specific responses to landscape features shaped genomic structure within Alaska galliformes. *Journal of Biogeography*, *49*(2), 261–273. https://doi.org/10.1111/jbi.14294

Spear, S. F., Cushman, S. A., & McRae, B. H. (2015). Resistance Surface Modeling in Landscape Genetics. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Eds.), *Landscape Genetics* (1st ed., pp. 129–148). Wiley. https://doi.org/10.1002/9781118525258.ch08

Spurgin, L. G., Bosse, M., Adriaensen, F., Albayrak, T., Barboutis, C., Belda, E., Bushuev, A., Cecere, J. G., Charmantier, A., Cichon, M., Dingemanse, N. J., Doligez, B., Eeva, T., Erikstad, K. E., Fedorov, V., Griggio, M., Heylen, D., Hille, S., Hinde, C. A., … Slate, J. (2024). The great tit HapMap project: A continental-scale analysis of genomic variation in a songbird. *Molecular Ecology Resources*, *24*(5), e13969. https://doi.org/10.1111/1755-0998.13969

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. P. (2007). Putting the 'landscape' in landscape genetics. *Heredity*, *98*(3), 128–142. https://doi.org/10.1038/sj.hdy.6800917

Summers, P. D., Cunnington, G. M., & Fahrig, L. (2011). Are the negative effects of roads on breeding birds caused by traffic noise? *Journal of Applied Ecology*, *48*(6), 1527–1534. https://doi.org/10.1111/j.1365-2664.2011.02041.x

Sutton, J. T., Helmkampf, M., Steiner, C. C., Bellinger, M. R., Korlach, J., Hall, R., Baybayan, P., Muehling, J., Gu, J., Kingan, S., & others. (2018). A high-quality, long-read de novo genome assembly to aid conservation of Hawaii's last remaining crow species. *Genes*, *9*(8), 393.

Szulkin, M., Garant, D., Mccleery, R. H., & Sheldon, B. C. (2007). Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, *20*(4), 1531– 1543. https://doi.org/10.1111/j.1420-9101.2007.01325.x

Thomassen, H. A., Harrigan, R. J., Semple Delaney, K., Riley, S. P. D., Serieys, L. E. K., Pease, K., Wayne, R. K., & Smith, T. B. (2018). Determining the drivers of population structure in a highly urbanized landscape to inform conservation planning. *Conservation Biology*, *32*(1), 148–158. https://doi.org/10.1111/cobi.12969

Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2019). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, 1755-0998.13077. https://doi.org/10.1111/1755-0998.13077

Van Der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, *43*(1). https://doi.org/10.1002/0471250953.bi1110s43

van Etten, J. (2017). *R* Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, *76*(13). https://doi.org/10.18637/jss.v076.i13

Vermote, E. (2021). MODIS/Aqua Surface Reflectance 8-Day L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes DAAC: Missoula, MT, USA.

Vermote, E. (2021). MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes DAAC: Missoula, MT, USA.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, *16*(2), 175–182. https://doi.org/10.1111/ele.12025

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. https://doi.org/10.1111/j.1365-294X.2009.04122.x

Wang, I. J., & Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, *19*(3), 447–458. https://doi.org/10.1111/j.1365-294X.2009.04465.x

Wang, P., Liu, S., Hu, J., Zhang, J., Wang, Z., Xu, J., Yao, H., Wang, B., Chen, D., Zhang, Z., & Liu, Y. (2022). Disentangling the relative roles of geographical and ecological factors in driving genomic variations of a widely distributed bird across a longitudinal gradient. *Journal of Avian Biology*, *2022*(7), e02979. https://doi.org/10.1111/jav.02979

Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *PLOS ONE*, *11*(6), e0156720. https://doi.org/10.1371/journal.pone.0156720

Wenzel, M. A., Douglas, A., James, M. C., Redpath, S. M., & Piertney, S. B. (2016). The role of parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, *25*(1), 324–341. https://doi.org/10.1111/mec.13473

Wilson, M. C., Chen, X.-Y., Corlett, R. T., Didham, R. K., Ding, P., Holt, R. D., Holyoak, M., Hu, G., Hughes, A. C., Jiang, L., Laurance, W. F., Liu, J., Pimm, S. L., Robinson, S. K., Russo, S. E., Si, X., Wilcove, D. S., Wu, J., & Yu, M. (2016). Habitat fragmentation and biodiversity conservation: Key findings and future challenges. *Landscape Ecology*, *31*(2), 219–227. https://doi.org/10.1007/s10980-015-0312-3

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.

Young, G., King, R., & Allen, B. L. (2023). Where do wildlife cross the road? Experimental evaluation reveals fauna preferences for multiple types of crossing structures. *Global Ecology and Conservation*, *46*, e02570. https://doi.org/10.1016/j.gecco.2023.e02570

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, *27*(6), 777–797. https://doi.org/10.1007/s10980-012-9737-0

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

CHAPTER 2: A genome assembly of the Yuma myotis bat, *Myotis yumanensis*

Journal of Heredity, 2024, **115**, 139–148
https://doi.org/10.1093/jhered/esad053
Advance access publication 15 September 2023
Genome Resources

Genome Resources

A genome assembly of the Yuma myotis bat, *Myotis yumanensis*

Joseph N. Curti[1,*], Devaughn Fraser[2], Merly Escalona[3], Colin W. Fairbairn[4], Samuel Sacco[4], Ruta Sahasrabudhe[5], Oanh Nguyen[5], William Seligmann[4], Peter H. Sudmant[6], Erin Toffelmier[1], Juan Manuel Vazquez[6], Robert Wayne[1,†], H. Bradley Shaffer[1,7] and Michael R. Buchalski[8]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles (UCLA), Los Angeles, CA, United States,
[2]Connecticut Department of Energy and Environmental Protection, Hartford, CT, United States,
[3]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, United States,
[4]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA, United States,
[5]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California, Davis, Davis, CA, United States,
[6]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, United States,
[7]Institute of the Environment and Sustainability, La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles (UCLA), Los Angeles, CA, United States,
[8]Wildlife Genetics Research Unit, Wildlife Health Laboratory, California Department of Fish and Wildlife, Sacramento, CA, United States

[*]Corresponding author: Department of Ecology and Evolutionary Biology, University of California, Los Angeles (UCLA), Los Angeles, CA, United States. Email: jcurti3@g.ucla.edu
[†]Deceased

Corresponding Editor: William Murphy

Abstract

The Yuma myotis bat (*Myotis yumanensis)* is a small vespertilionid bat and one of 52 species of new world *Myotis* bats in the subgenus *Pizonyx*. While *M. yumanensis* populations currently appear relatively stable, it is one of 12 bat species known or suspected to be susceptible to white-nose syndrome, the fungal disease causing declines in bat populations across North America. Only two of these 12 species have genome resources available, which limits the ability of resource managers to use genomic techniques to track the responses of bat populations to white-nose syndrome generally. Here we present the first de novo genome assembly for Yuma myotis, generated as a part of the California Conservation Genomics Project. The *M. yumanensis* genome was generated using a combination of PacBio HiFi long reads and Omni-C chromatin-proximity sequencing technology. This high-quality genome is one of the most complete bat assemblies available, with a contig N50 of 28.03 Mb, scaffold N50 of 99.14 Mb, and BUSCO completeness score of 93.7%. The Yuma myotis genome provides a high-quality resource that will aid in comparative genomic and evolutionary studies, as well as inform conservation management related to white-nose syndrome.

Key words: California Conservation Genomics Project, CCGP, chiroptera, long-read assembly, *Myotis yumanensis*, reference genome

Introduction

Bats (order Chiroptera) are the second-most diverse mammalian order, representing 22% of global mammal diversity (Simmons and Cirranello 2018; Mammal Diversity Database 2022). Despite their global distribution and ecological and economic importance, the conservation status of bats is less well understood than other species of mammals or birds (Frick et al. 2020). In step with data gaps in the global conservation status of bats, genomic resources for bats are also underdeveloped. Since the first reference genome of the little brown bat (*Myotis lucifigus*) was published by the Broad Institute in 2011 (Lindblad-Toh et al. 2011), 50 additional bat reference genomes have been made publicly available, although 37 (74%) of these genomes are highly fragmented, primarily short-read assemblies. Eleven of the 19 currently recognized chiropteran families have at least one reference genome, and most are from species in the families Pteropodidae, Phyllostomidae, and Vespertilionidae, including four in the genus *Myotis*. Given that the genus contains more than 120 globally distributed species, many of which have experienced declines in recent decades, additional genomic resources are sorely needed for the group.

The Yuma myotis bat (hereafter "Yuma bat"; *Myotis yumanensis*; Allen 1864) is one of 47 bat species endemic to North America. The Yuma bat is abundant and widely distributed, occurring as far north as British Columbia, Canada, south throughout most of the western United States, and as far south as Morelos, Mexico (Braun et al. 2015). Yuma

Received July 12, 2023; Accepted September 13, 2023

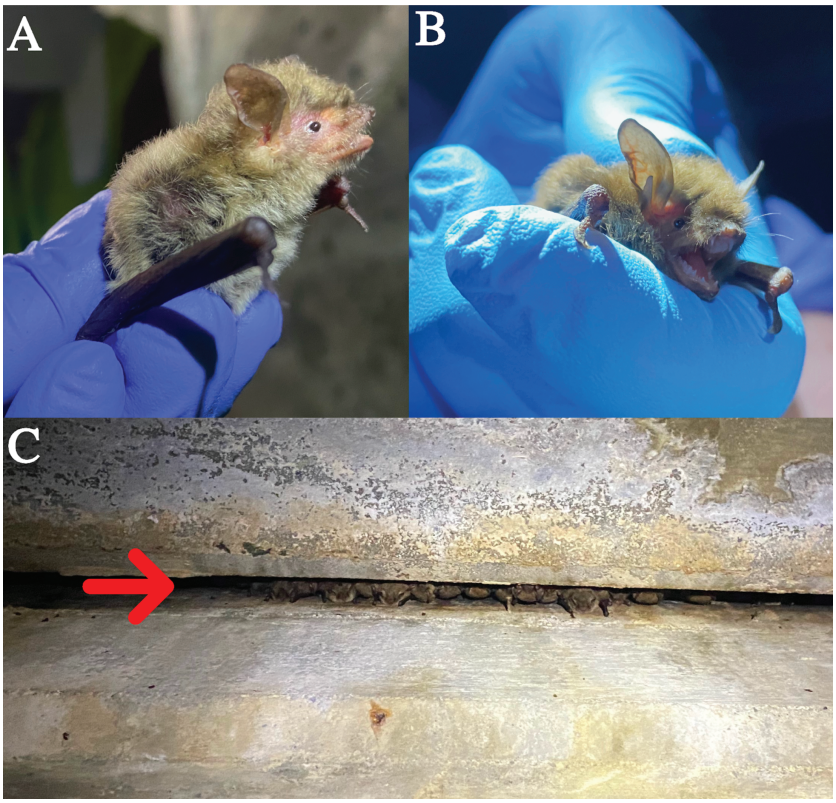

Fig. 1. (A) Profile view and (B) front-on view of Yuma myotis bats (*Myotis yumanensis*). (C) *M. yumanensis* day roost in a longitudinal joint of a bridge in Riverside County, California, USA.

bats are closely associated with riparian habitat for foraging (Brigham et al. 1992; Duff and Morrell 2007) and utilize a variety of natural (Braun et al. 2015) and manmade (Evelyn et al. 2004) roost types (Fig. 1). There are six putative subspecies of Yuma bat including *M. y. lambi, M. y. lutosus, M. y. oxalis, M. y. saturatus, M. y. sociabilis,* and *M. y. yumanensis*, although the extent to which these subspecies are supported as evolutionarily distinct lineages by genomic data is unknown (Braun et al. 2015).

The Yuma bat is also one of 12 bat species in North America with confirmed detection of *Pseudogymnoascus destructans* (Pd), the fungus responsible for white-nose syndrome (WNS). For some species of bats such as the little brown bat, WNS has resulted in more than 90% loss from certain colonies (Frick et al. 2010). Furthermore, although the IUCN considers the Yuma bat stable across its native range (Solari 2019), occupancy models derived from acoustic data indicate a slight decline in summer occupancy over the three-year period of 2016–2019 (Udell et al. 2022). As WNS continues to spread across North America (Duncan 2023), it will be important to monitor common, abundant species such as the Yuma bat to detect and document population declines as they occur.

Genomic data provide an effective, efficient tool to monitor WNS-related mortalities in bat populations, as well as the genes underlying survival. Using whole genome resequencing data, researchers have identified single nucleotide polymorphisms related to torpor and immune function in bat populations that survive WNS (Lilley et al. 2020b; Gignoux-Wolfsohn et al. 2021) and have investigated potential declines in genomic diversity following mass die offs (Lilley et al. 2020b). Genomic studies such as these rely heavily on the availability of high-quality reference genomes (Brandies et al. 2019).

Here, we describe the genome assembly for *M. yumanensis*, generated through the California Conservation Genomics Project (CCGP; Shaffer et al. 2022). One of the primary goals of the CCGP is to generate reference genomes and whole genome resequencing data for a comprehensive set of 153 ecologically and phylogenetically diverse species across California (Shaffer et al. 2022), and the Yuma bat is one of two chiropteran species in the project. Using PacBio HiFi long reads and Omni-C chromatin-proximity sequencing technology, we generated the first assembly for the species. The Yuma bat genome is an invaluable resource for basic research on diversification among *Myotis* species and the evolution of unique traits like echolocation and disease resistance, as well as more applied work on population size, connectivity, and genomic health that will aid in WNS management planning.

Methods

Tissue collection and cell culture

We captured a juvenile male Yuma bat from a maternity colony located in Chester, Plumas County, California. The specimen was collected by California Department of Fish and Wildlife (CDFW) staff under the department's jurisdiction as the trustee for wildlife management in the state of California, CA Fish & Game Code § 1802 (2015). The animal was transported to a CDFW laboratory facility where it was humanely euthanized via a combination of isoflurane and cervical dislocation. The carcass was immediately dissected and tissues were collected for genome sequencing. Several aliquots of kidney, lung, heart, spleen, liver, testes, intestine, skeletal muscle, and brain were washed sequentially in molecular grade water, ethanol, and water again before being flash frozen in liquid nitrogen. One aliquot of each tissue was reserved for generating primary cell cultures. Species identity was confirmed through Sanger sequencing of a fragment of the cytochrome oxidase subunit 1 (COI) mitochondrial gene using the methodology of Walker et al. (2016).

Primary cell cultures from the skin (plagiopatagium and body), heart, brain, cartilage, and eye were grown following Yohe et al. (2019) with modifications. Tissue samples were rinsed serially in baths of DPBS, 70% ethanol, and DPBS, and then stabilized in a cell culture medium consisting of BenchStable DMEM/F12 (Gibco Cat. #A4192002, Thermofisher Scientific Inc., Waltham, MA) supplemented with 20% FBS (Gibco Cat. #26140087), 0.2% Primocin (InvivoGen Cat. #ant-pm-1, San Diego, CA), and 15 mM HEPES (Gibco Cat. #15630080). Tissues were minced in 500 µL of DPBS using surgical scissors, and the tissues were digested overnight in 1 mg/mL Collagenase IV (Stemcell Technologies Cat. #07909, Vancouver, Canada) supplemented with 0.2% Primocin. The dissociated tissues were centrifuged at 500 × *g* for 5 min, and washed twice with DPBS (Gibco Cat. #14190144). Cells were plated in T75 flasks containing cell culture media formulated as described, and grown in a 37 °C incubator with 5% CO_2 atmosphere.

Adherent cells were passaged four days post-collection ("Passage 0") using 0.05% Trypsin-EDTA (Gibco Cat. #25300054). Cells were then counted and replated in high glucose DMEM (Gibco Cat. #10569010) with pyruvate and GlutaMax supplementation, plus 10% FBS and 1% penicillin-streptomycin (Gibco Cat. #10378016). Three T175 flasks were seeded with approximately two million cells each after the first passage to generate triplicates of 10 million cell aliquots for DNA and RNA extraction.

Nucleic acid library preparation

High molecular weight genomic DNA (HMW gDNA) was isolated from cultured cells following a protocol described previously (Jain et al. 2018). Briefly, 10 million cultured skin fibroblast cells were lysed with 2 mL lysis buffer containing 10 mM NaCl, 25 mM EDTA, 0.5% (weight/volume) SDS, and 100 µg/mL Proteinase K overnight at room temperature. The lysate was treated with RNase A for 30 min at 37 °C and cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio Cat. #2302830, Beverly, MA). The HMW gDNA was precipitated by adding 0.4× volume of 5 M ammonium acetate and 3× volume of ice cold ethanol. The pellet was washed with 70% ethanol twice and resuspended in elution buffer (10 mM Tris, pH 8.0). The purity was accessed using NanoDrop spectrophotometer (260/280 = 1.8 and 260/230 = 2.0) and the integrity of the HMW gDNA was verified on a Femto pulse system (Agilent Technologies, Santa Clara, CA).

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences of California [PacBio] Cat. #100938900, Menlo Park, CA) according to the manufacturer's instructions. HMW gDNA was sheared to a target size distribution between 15 and 20 kb. The sheared gDNA was concentrated using 0.45× of AMPure PB beads (PacBio Cat. #100265900) for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 min, ligation of overhang adapter v3 at 20 °C for 60 min and 65 °C for 10 min to inactivate the ligase, then nuclease treated at 37 °C for 1 h. The SMRTbell library was purified and concentrated with 0.45× AMPure PB beads for size selection using the BluePippin/PippinHT system (Sage Science Inc. Cat. #BLF7510/HPE7510, Beverly, MA) to collect fragments greater than 79 kb. The 15–20 kb average HiFi SMRTbell library was sequenced at the University of California, Davis, DNA Technologies Core (Davis, CA) using three SMRT Cell 8M Trays (PacBio Cat. #101389001), Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

The Omni-C library was prepared using a Dovetail Omni-C Kit (Dovetail Genomics Cat. #21005, Scotts Valley, CA) according to the manufacturer's protocol with slight modifications. First, cultured cell pellets (Sample ID: MYYU_CA2020_CCGP) were resuspended in 1× PBS. Then, chromatin was fixed in place in the nucleus, and the fixed chromatin was digested with DNase I and extracted. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed and the DNA was purified from proteins, purified DNA was treated to remove biotin that was not internal to ligated fragments, and a sequencing library was generated using the NEBNext Ultra II (New England Biolabs Inc. Cat. #E7645, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post capture product was split into two replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at the Vincent J. Coates Genomics Sequencing Laboratory (Berkeley, CA) on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) to generate approximately 100 million 2 × 150 bp read pairs per Gb of genome size.

Nuclear genome assembly

We assembled the *M. yumanensis* genome following the CCGP assembly pipeline Version 5.0, as outlined in Table 1, which lists the tools and nondefault parameters used. The pipeline uses PacBio HiFi reads and Omni-C data to produce high quality and highly contiguous genome assemblies. First, we removed the remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022) and generated the initial dual or partially phased diploid assembly (http://lh3.github.io/2021/10/10/introducing-dual-assembly) using HiFiasm (Cheng et al. 2022) on Hi-C mode, with the filtered PacBio HiFi reads and the Omni-C dataset. We then aligned the Omni-C data to both assemblies following the Arima Genomics Mapping Pipeline

Table 1 Assembly pipeline and software used. Software citations are listed in the main text

Assembly	Software and any non-default options	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k=21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (Hi-C Mode, –primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Omni-C scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloser (-mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 0e34c3b
Omni-C contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.0.3
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l mammalia)	5.0.0
	Merqury	2020-01-29
Contamination screening		
Local alignment tool	BLAST+ (-db nt, -outfmt '6 qseqid staxids bitscore std' , -max_target_seqs 1, -max_hsps 1, -evalue 1e-25)	2.1
General contamination screening	BlobToolKit	2.3.3
Mitochondrial assembly		
Mitochondrial genome assembly	MitoHiFi (-r, -p 50, -o 1)	2.2
Comparing available genome assemblies		
Genome contiguity	ggplot2	3.4.1 (R version 4.2.3)
	Custom script (https://github.com/joeycurti3/myyu_joh)	Commit 3f5c8dd
Genome genic completeness	gVolante (-cuttoff length = 1, -sequence type = Genome (nucleotide), -ortholog search pipeline = BUSCO v5, -ortholog set = mammalia	2.0.0

(https://github.com/ArimaGenomics/mapping_pipeline) and scaffolded both assemblies with SALSA (Ghurye et al. 2017, 2019).

Both genome assemblies were manually curated by iteratively generating and analyzing their corresponding Omni-C contact maps. To generate the contact maps we aligned the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Open2C et al. 2023). We generated a multi-resolution Omni-C matrix with cooler (Abdennur and Mirny 2020) and balanced it with hicExplorer (Ramírez et al. 2018). We used HiGlass (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps where we identified misassemblies and misjoins, and finally modified the assemblies using the Rapid Curation pipeline from the Wellcome Trust Sanger Institute, Genome Reference Informatics Team (https://gitlab.com/wtsi-grit/rapid-curation). Some of the remaining gaps (joins generated during scaffolding and curation) were closed using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser). Finally, we checked for contamination using the BlobToolKit Framework (Challis et al. 2020).

Genome assembly assessment

We generated k-mer counts from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl). The k-mer counts were then used in GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) to estimate genome features including genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich et al. 2013). We evaluated genome quality and functional completeness using BUSCO (Manni et al. 2021) with the Mammalia ortholog database (mammalia_odb10) which contains 9,226 genes. Assessment of base level accuracy (QV) and k-mer completeness was

performed using the previously generated meryl database and merqury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in Korlach et al. (2017). Measurements of the size of the phased blocks are based on the size of the contigs generated by HiFiasm on HiC mode. We followed the quality metric nomenclature established by Rhie et al. (2021), with the genome quality code $x{\cdot}y{\cdot}P{\cdot}Q{\cdot}C$, where, $x = \log_{10}$[contig NG50]; $y = \log_{10}$[scaffold NG50]; $P = \log_{10}$ [phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a karyotype of 2n = 44 (Braun et al 2015). Quality metrics for the notation were calculated on the assembly for Haplotype 1.

Mitochondrial genome assembly

We assembled the mitochondrial genome of *M. yumanensis* from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (Allio et al. 2020; Uliano-Silva et al. 2021). The mitochondrial sequence of an existing *M. yumanenis* (NCBI:NC_036319.1; Platt et al. 2018) was used as the starting reference sequence. After completion of the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST+ (Camacho et al. 2009) and filtered out contigs and scaffolds from the nuclear genome with a percentage of sequence identity >99% and size smaller than the mitochondrial assembly sequence.

Comparing available genome assemblies

We queried the National Library of Medicine's National Center for Biotechnology Information (NCBI) on 11 April 2023 for all representative genome assemblies using the taxon id for *Chiroptera* (search term: txid9397[Organism:exp]). For each assembly, we recorded the genomes's global statistics including genome size, scaffold number, scaffold N50, contig number, and contig N50. To compare the contiguity of available genomes, we accessed NCBI full sequence reports for all 50 available bat genomes and plotted the cumulative coverage of the genome by scaffold of a given size (NGx plot) in R (R Core Team 2022), using the package "ggplot2" (Wickham 2016) following scripts from Lin et al. (2022). To compare completeness of available genomes, we downloaded fasta sequences for all 50 available bat genomes on NCBI and we used gVolante (Nishimura et al. 2017, 2019) to run BUSCO using the Mammalian ortholog database (mammalia_odb10).

Results

The Omni-C and PacBio HiFi sequencing libraries generated 120.4 million read pairs and 4.7 million reads, respectively. The latter yielded ~40-fold coverage (N50 read length 16,323 bp; minimum read length 43 bp; mean read length 16,158 bp; maximum read length of 52,146 bp) based on the Genomescope 2.0 genome size estimation of 1.9 Gb. Based on PacBio HiFi reads, we estimated 0.194% sequencing error rate and 0.809% nucleotide heterozygosity rate. The k-mer spectrum based on PacBio HiFi reads show a bimodal distribution with two major peaks at ~38 and ~75-fold coverage, where peaks correspond to homozygous and heterozygous states of a diploid species (Fig. 2A).

The final assembly (mMyoYum1) consists of two partially phased haplotypes that vary slightly in size compared with the estimated value from GenomeScope 2.0 (Fig. 2A), as has been observed in other taxa (see e.g. Pflug et al. 2020). Haplotype 1 consists of 476 scaffolds spanning 1.94 Gb with contig N50 of 28.03 Mb, scaffold N50 of 99.14 Mb, longest contig of 120.09 Mb, and largest scaffold of 240.34 Mb. The Haplotype 2 assembly consists of 250 scaffolds, spanning 2.05 Gb with contig N50 of 26.79 Mb, scaffold N50 of 94.21 Mb, longest contig of 59.72 Mb, and largest scaffold of 216.39 Mb. Assembly statistics are reported in Table 2, and graphical representation for the primary assembly in Fig. 2B.

During manual curation, we generated a total of 12 breaks and 153 joins, with 6 breaks per haplotype, 79 joins for Haplotype 1, and 74 joins were made for Haplotype 2. We were able to close 45 gaps, 19 on Haplotype 1 and 26 on Haplotype 2, and we filtered out 2 contigs (1 per haplotype), corresponding to mitochondrial contamination. No further contigs were removed. The Omni-C contact maps show that both assemblies are highly contiguous (Fig. 2C and 2D). We have deposited both assemblies on NCBI (see Table 2 and Data Availability for details).

Haplotype 1 has a BUSCO completeness score of 93.7% using the Mammalian ortholog database, a per-base quality (QV) of 63.62, a kmer completeness of 89.64, and a frameshift indel QV of 40.98. Haplotype 2 has a BUSCO completeness score of 91.2% using the same ortholog database, a per-base quality (QV) of 63.88, a kmer completeness of 93.97, and a frameshift indel QV of 40.27. The Omni-C contact maps show that both assemblies are highly contiguous with some chromosome-length scaffolds (Fig. 2C and 2D, respectively; see Table 2 and Data availability for details).

The final mitochondrial genome size was 17,366 bp. The base composition of the final assembly version is A = 33.55%, C = 22.93%, G = 13.44%, T = 30.08%, and consists of 22 unique transfer RNAs and 13 protein-coding genes.

Across all available bat genomes, genome contiguity based on scaffold N50 values ranged from 0.0107 to 171.1 Gb ($\bar{x}$ = 29.73). Furthermore, completeness based on BUSCO percentage of complete genes detected ranged from 47.33 to 96.61 ($\bar{x}$ = 85.39). Generally, short-read genome assemblies were less contiguous ($\bar{x}$ = 11.03 Mb) and less complete ($\bar{x}$ = 81.91) than assemblies that used a combination of long and short reads ($\bar{x}$ = 92.44 Mb and $\bar{x}$ = 95.21%, respectively).

Discussion

Here we provide the first genome assembly for the Yuma bat. This genome is highly contiguous and when compared against standards set by the Vertebrate Genome Project (VGP; https://vertebrategenomesproject.org/), this genome exceeds the proposed standards for the VGP2020 category (Rhie et al. 2021), with the exception of the "chromosome status" quality category, since we did not name or match chromosomes. This genomic resource is comparable in its contiguity and completeness to other modern de novo genome assemblies that use a combination of short and long-read technologies, and is one of the most contiguous bat genomes currently available based on scaffold N50 (99.14 Mb for Yuma bat, range of other taxa: 0.0107–171.1 Gb). When compared with the other available genomes for bats in the genus *Myotis*, this genome is the most contiguous based on scaffold N50 (99.14 Mb for Yuma bat, range of other taxa: 3.226–94.45 Mb; Fig.

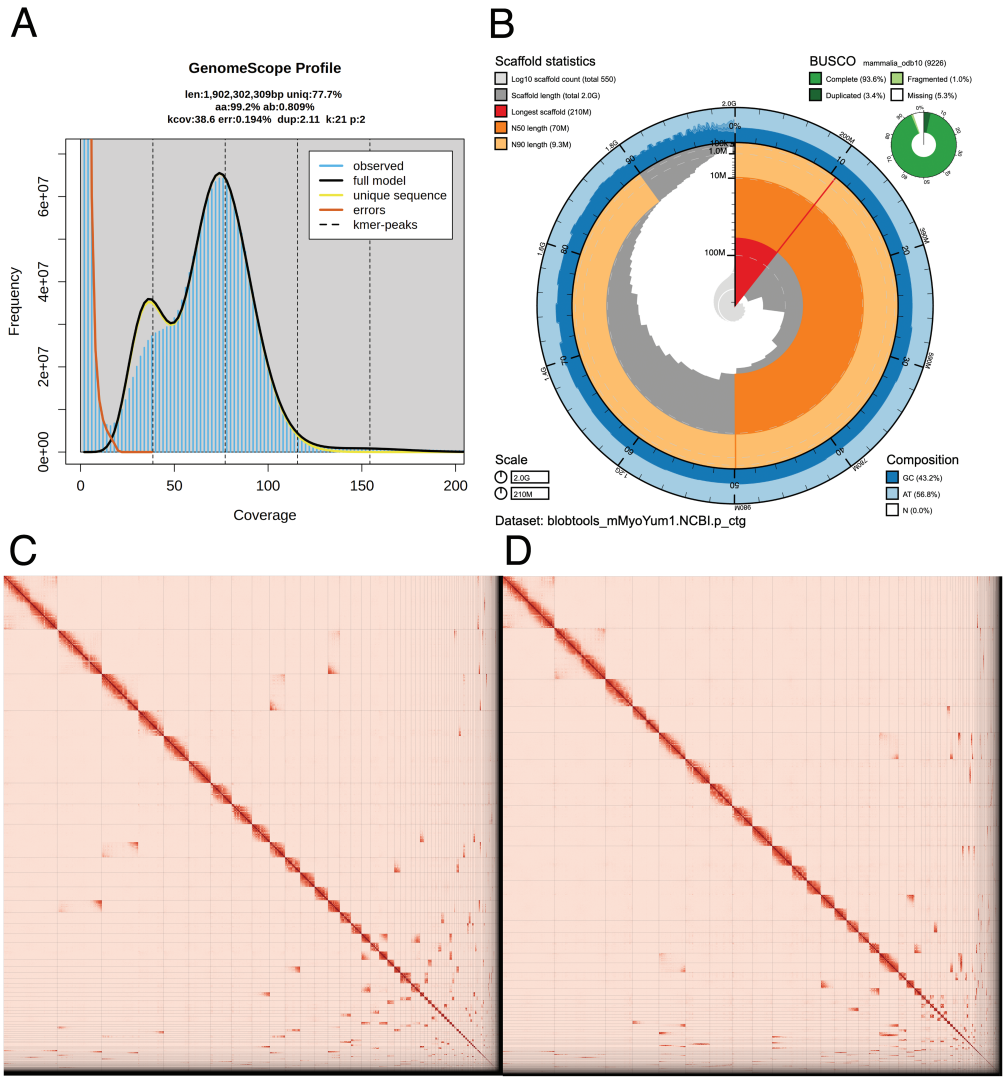

Fig. 2. Visual overview of genome assembly metrics. (A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome. K-mers covered at lower coverage and lower frequency correspond to differences between haplotypes, whereas the higher coverage and higher frequency k-mers correspond to the similarities between haplotypes. (B) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *M. yumanensis* primary assembly (mMyoYum1.0.hap1). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly. The dark versus light blue area around it shows mean, maximum, and minimum GC versus AT content at 0.1% intervals (Challis et al. 2020). (C-D) The Omni-C contact map for the primary (C) and alternate (D) genome assemblies generated with PretextSnapshot. Omni-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two such regions. Scaffolds are separated by black lines, and higher density corresponds to higher levels of fragmentation (See online version for color figure).

Table 2 Sequencing and assembly statistics, and accession numbers

Bio Projects & Vouchers	CCGP NCBI BioProject		PRJNA720569				
	Genera NCBI BioProject		PRJNA765635				
	Species NCBI BioProject		PRJNA777197				
	NCBI BioSample		SAMN30526064				
	Specimen identification		MYYU_CA2020_CCGP				
Genome Sequence	NCBI Genome accessions		**Haplotype 1 (Primary)**			**Haplotype 2 (Alternate)**	
	Assembly accession		JAPQVT000000000			JAPQVU000000000	
	Genome sequences		GCA_028538775.1			GCA_028536395.1	
Sequencing Data	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel II) run: 4.7 M spots, 76.5 G bases, 57 Gb				
		Accession	SRX19740654				
	Omni-C Illumina reads	Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 120.5 M spots, 36.4 G bases, 11.9 Gb				
		Accession	SRX19740655, SRX19740656				
Genome Assembly Quality Metrics	Assembly identifier (Quality code*)		mMyoYum1(7.7.P7.Q63.C96)				
	HiFi Read coverage§		33.26X				
			Haplotype 1			**Haplotype 2**	
	Number of contigs		685			465	
	Contig N50 (bp)		28,025,655			26,795,370	
	Contig NG50§		28,147,841			28,130,932	
	Longest Contigs		120,097,812			597,242,388	
	Number of scaffolds		476			250	
	Scaffold N50		99,144,700			94,205,551	
	Scaffold NG50§		99,144,700			109,018,441	
	Largest scaffold		240,344,003			2,163,927,272	
	Size of final assembly		1,952,479,771			2,050,500,308	
	Phased block NG50§		27,204,636			27,189,810	
	Gaps per Gbp (# Gaps)		107 (209)			104 (215)	
	Indel QV (Frame shift)		40.98297536			40.27268042	
	Base pair QV		63.6294			63.8881	
			Full assembly = 63.76				
	k-mer completeness		89.6446			93.9753	
			Full assembly = 99.442				
	BUSCO completeness (mammalia) n = 9226		**C**	**S**	**D**	**F**	**M**
		H1‡	93.70%	90.20%	3.50%	1.00%	5.30%
		H2‡	95.80%	92.20%	3.60%	1.00%	3.20%
	Organelles		1 complete mitochondrial sequence			CM053173.1	

* Assembly quality code x.y.P.Q.C derived notation, from (Rhie et al. 2021). x = $\log_{10}$ [contig NG50]; y = $\log_{10}$ [scaffold NG50]; P = $\log_{10}$ [phased block NG50]; Q = Phred base accuracy QV (Quality value); C = % genome represented by the first 'n' scaffolds, following a known karyotype for *M. yumanensis* of 2n = 44 (Braun et al 2015). Quality code for all the assembly denoted by Haplotype 1 assembly (mMyoYum1.0.hap1)
§ Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.95 Gb
‡ (H1) Haplotype 1 and (H2) Haplotype 2 assembly values.

3), and the second most complete based on its BUSCO score of 93.7% (range of other taxa: 86.57%–96.18%). Future work could further improve this assembly through additional manual curation of scaffold placement and targeted DNA-FISH to assign scaffolds to true karyotypes (Shakoori 2017). Such work, along with gene annotation using RNA-seq for gene prediction, is planned for future versions of this assembly.

Genomic data are increasingly being applied to investigate the unique traits bats possess, including the ability to act as hosts to many pathogens without succumbing to illness (Chattopadhyay et al. 2020; Moreno Santillán et al. 2021), the physiological basis of unique feeding behaviors like sanguivory (blood feeding; Zepeda Mendoza et al. 2018), and the exceptional longevity of bats relative to their small body size (Foley et al. 2018; Sullivan et al. 2022). While these and other bat genomic studies have the potential to prove useful to human biomedical research as well as our understanding of chiropteran evolution, they are often severely limited by the availability and quality of genomic resources. For example, of the 50 bat reference genomes currently available, 34 (74%) are short-read

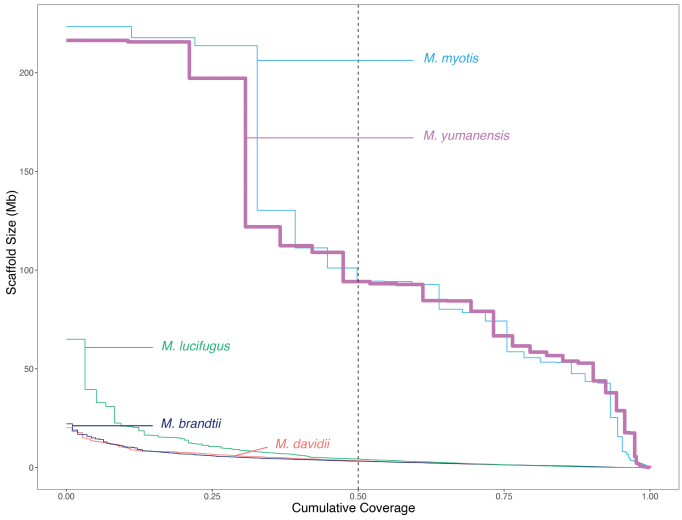

Fig. 3. NGx plot comparing contiguity of available de novo reference genomes for bats in the genus *Myotis*. The plot depicts the fraction of the genome (x-axis) that is covered by scaffolds of a given size in Mb (y-axis). The vertical dashed line depicts the N50 value, or half of the genome. The thick pink line is the genome of *M. yumanensis* presented in this paper.

assemblies with very low overall contiguity and completeness (Supplementary Materials). Our Yuma bat assembly provides a high quality, near-chromosome level resource in support of these research efforts. At the level of California biodiversity, the Yuma bat genome is the first chiropteran reference genome sequenced by the CCGP, filling a major gap in our emerging phylogeny of California biodiversity (Toffelmier et al. 2022). It contributes a new reference genome that will help in resolving outstanding questions on both species delimitation and phylogenetic relationships for the hyperdiverse genus *Myotis*, including the role of hybridization in shaping contemporary genomic architecture (Korstian et al. 2022). The CCGP will also generate 163 resequenced genomes throughout the species' distributional range, including all currently recognized subspecies, and this reference genome will be critical to evaluating the validity of, and relationships among, those taxa.

Genomic resources can also enhance the conservation and management of bat species, both in California (Fiedler et al. 2022) and globally. Two major foci of bat conservation are to better understand the susceptibility of individuals and species to WNS, and predict the spread of the pathogen among North American populations. Currently, only 5 of 20 bat species known to be affected by WNS have available genomic resources, including the reference genome presented here. Increasing genomic resources for these species will facilitate research on impacts of WNS, including the loss of genetic diversity due to population declines (Lilley et al. 2020b) and genomic predictions regarding individual-to-individual spread of the pathogen across landscapes (Lilley et al. 2020a).

In conclusion, we present the first high-quality genomic resource for the Yuma bat, a currently abundant and widespread North American species. This highly contiguous and complete de novo genome assembly will be a valuable resource for studies aimed at understanding the evolution of unique bat traits and will contribute to bat conservation and management planning.

Supplementary material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Acknowledgements

PacBio Sequel II library prep and sequencing were carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data. We also thank Dr. Courtney Miller for assistance with editing early drafts of this manuscript. This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group.

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224], and a White-nose Syndrome Recovery Research Grant [ID: F18AS00119]

awarded by the United States Fish and Wildlife Service to California Department of Fish and Wildlife. JMV was funded by a National Science Foundation Postdoctoral Research Fellowship in Biology [ID: 2109915].

Data availability

Data generated for this study are available under NCBI BioProject PRJNA777197. Raw sequencing data for sample MYYU_CA2020_CCGP (NCBI BioSample SAMN30526064) are deposited in the NCBI Short Read Archive (SRA) under SRX19740654 for PacBio HiFi sequencing data, and SRX19740655 and SRX19740656 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_028538775.1 and GCA_028536395.1; and for genome sequences JAPQVT000000000 and JAPQVU000000000. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly.

References

Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020:36:311–316. https://doi.org/10.1093/bioinformatics/btz540.

Allen H. Monograph of North American bats. *Smithsonian Misc Collect*. 1864:7:184.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020:20:892–905. https://doi.org/10.1111/1755-0998.13160.

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. *Genes*. 2019:10:846. https://doi.org/10.3390/genes10110846.

Braun JK, Yang B, Gonzalez-Perez SB, Mares MA. *Myotis yumanensis* (Chiroptera: Vespertilionidae). *Mamm Species*. 2015:47:1–14. https://doi.org/10.1093/mspecies/sev001.

Brigham RM, Aldridge HDJN, Mackey RL. Variation in habitat use and prey selection by yuma bats, *Myotis yumanensis*. *J Mammal*. 1992:73:640–645. https://doi.org/10.2307/1382036.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinf*. 2009:10:421. https://doi.org/10.1186/1471-2105-10-421.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit – interactive quality assessment of genome assemblies. *G3 Genes Genom Genet*. 2020:10:1361–1374.

Chattopadhyay B, Garg KM, Ray R, Mendenhall IH, Rheindt FE. Novel de Novo genome of *Cynopterus brachyotis* reveals evolutionarily abrupt shifts in gene family composition across fruit bats. *Genome Biol Evol*. 2020:12:259–272. https://doi.org/10.1093/gbe/evaa030.

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022:40:1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

Duff AA, Morrell TE. Predictive occurrence models for bat species in California. *J Wildl Manag*. 2007:71:693–700. https://doi.org/10.2193/2005-692.

Duncan, T. First Colorado bat tests positive for deadly white-nose syndrome. Colorado parks and wildlife; 2023. [accessed 2023 April 24]. https://cpw.state.co.us/aboutus/Pages/News-Release-Details.aspx?NewsID=3797.

Evelyn MJ, Stiles DA, Young RA. Conservation of bats in suburban landscapes: roost selection by *Myotis yumanensis* in a residential area in California. *Biol Conserv*. 2004:115:463–473. https://doi.org/10.1016/S0006-3207(03)00163-0.

Fiedler PL, Erickson B, Esgro M, Gold M, Hull JM, Norris JM, Shapiro B, Westphal M, Toffelmier E, Shaffer HB. Seizing the moment: the opportunity and relevance of the California Conservation Genomics Project to state and federal conservation policy. *J Hered*. 2022:113:589–596.

Foley NM, Hughes GM, Huang Z, Clarke M, Jebb D, Whelan CV, Petit EJ, Touzalin F, Farcy O, Jones G, et al. Growing old, yet staying young: the role of telomeres in bats' exceptional longevity. *Sci Adv*. 2018:4:eaao0926. https://doi.org/10.1126/sciadv.aao0926.

Frick WF, Kingston T, Flanders J. A review of the major threats and challenges to global bat conservation. *Ann N Y Acad Sci*. 2020:1469:5–25. https://doi.org/10.1111/nyas.14045.

Frick WF, Pollock JF, Hicks AC, Langwig KE, Reynolds DS, Turner GG, Butchkoski CM, Kunz TH. An emerging disease causes regional population collapse of a common North American bat species. *Science*. 2010:329:679–682. https://doi.org/10.1126/science.1188594.

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017:18:527. https://doi.org/10.1186/s12864-017-3879-z.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2019:15:e1007273. https://doi.org/10.1371/journal.pcbi.1007273.

Gignoux-Wolfsohn SA, Pinsky ML, Kerwin K, Herzog C, Hall M, Bennett AB, Fefferman NH, Maslo B. Genomic signatures of selection in bats surviving white-nose syndrome. *Mol Ecol*. 2021:30:5643–5657. https://doi.org/10.1111/mec.15813.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013:29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018:36:338–345. https://doi.org/10.1038/nbt.4060.

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018:19:125. https://doi.org/10.1186/s13059-018-1486-1.

Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, Cantin L, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017:6:1–16. https://doi.org/10.1093/gigascience/gix085.

Korstian JM, Paulat NS, Platt RN, Stevens RD, Ray DA. SINE-based phylogenomics reveal extensive introgression and incomplete lineage sorting in myotis. *Genes*. 2022:13:399. https://doi.org/10.3390/genes13030399.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, *arXiv*, arXiv:1303.3997 [q-Bio]. 2013. https://doi.org/10.48550/arXiv.1303.3997.

Lilley TM, Sävilammi TM, Ossa G, Blomberg AS, Vasemägi A, Yung V, Vendrami D, Johnson JS. Population connectivity predicts vulnerability to white-nose syndrome in the Chilean Myotis (*Myotis chiloensis*)—a genomics approach. *G3 Genes Genom Genet*. 2020a:10:2117–2126. https://doi.org/10.25387/G3.12173385.

Lilley TM, Wilson IW, Field KA, Reeder DM, Vodzak ME, Turner GG, Kurta A, Blomberg AS, Hoff S, Herzog CJ, et al. Genome-wide changes in genetic diversity in a population of *Myotis lucifugus* affected by white-nose syndrome. *G3 Genes Genom Genet*. 2020b:10:2007–2020. https://doi.org/10.1534/g3.119.400966.

Lin M, Escalona M, Sahasrabudhe R, Nguyen O, Beraut E, Buchalski MR, Wayne RK. A reference genome assembly of the bobcat, *Lynx rufus*. *J Hered*. 2022:113:615–623. https://doi.org/10.1093/jhered/esac031.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. (2011). A high-resolution map of human evolutionary constraint using 29

mammals. *Nature*, 478:476–482. https://doi.org/10.1038/nature10530.

Mammal Diversity Database. (2022). *Mammal diversity database (Version 1.9)*. Zenodo. http://doi.org/10.5281/zenodo.4139818.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38:4647–4654. https://doi.org/10.1093/molbev/msab199.

Moreno Santillán DD, Lama TM, Gutierrez Guerrero YT, Brown AM, Donat P, Zhao H, Rossiter SJ, Yohe LR, Potter JH, Teeling EC, et al. Large-scale genome sampling reveals unique immunity and metabolic adaptations in bats. *Mol Ecol*. 2021:30:6449–6467. https://doi.org/10.1111/mec.16027.

Nishimura O, Hara Y, Kuraku S. GVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*. 2017:33:3635–3637. https://doi.org/10.1093/bioinformatics/btx445.

Nishimura O, Hara Y, Kuraku S. Evaluating genome assemblies and gene models using gvolante. In: Kollmar M, editor. *Gene prediction: methods and protocols*. Vol. 1962. New York (NY): Springer; 2019. p. 247–256. https://doi.org/10.1007/978-1-4939-9173-0

Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna A A, Goloborodko A, Imakaev M, Venev SV. Pairtools: From sequencing data to chromosome contacts. *bioRxiv*. 2023. https://doi.org/10.1101/2023.02.13.528389.

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). *G3 Genes Genom Genet*. 2020:10:3047–3060. https://doi.org/10.1534/g3.120.401028.

Platt RN, Faircloth BC, Sullivan KAM, Kieran TJ, Glenn TC, Vandewege MW, Lee TE, Baker RJ, Stevens RD, Ray DA. Conflicting evolutionary histories of the mitochondrial and nuclear genomes in new world Myotis bats. *Syst Biol*. 2018:67:236–249. https://doi.org/10.1093/sysbio/syx070.

R Core Team. (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018:9:189. https://doi.org/10.1038/s41467-017-02525-w.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020:11:1432. https://doi.org/10.1038/s41467-020-14998-3.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021:592:737–746. https://doi.org/10.1038/s41586-021-03451-0.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020:21:245. https://doi.org/10.1186/s13059-020-02134-9.

Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, et al. Landscape genomics to enable conservation actions: the California conservation genomics project. *J Hered*. 2022:113:577–588. https://doi.org/10.1093/jhered/esac020.

Shakoori AR. Fluorescence in situ hybridization (FISH) and its applications. In: Bhat T, Wani A, editors. *Chromosome structure and aberrations*. New Delhi: Springer, 2017. https://doi.org/10.1007/978-81-322-3673-3_16.

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022:23:157. https://doi.org/10.1186/s12864-022-08375-1.

Simmons NB, Cirranello AL. *Bat species of the world: a taxonomic and geographic database*; 2018 [accessed 2023 April 5]. https://batnames.org/

Solari, S. *Myotis yumanensis*. The IUCN red list of threatened species 2019: e.T14213A22068335; 2019. http://doi.org/10.2305/IUCN.UK.2019-1.RLTS.T14213A22068335.en

Sullivan IR, Adams DM, Greville LJS, Faure PA, Wilkinson GS. Big brown bats experience slower epigenetic aging during hibernation. *Proc R Soc B Biol Sci*. 2022:289:20220635. https://doi.org/10.1098/rspb.2022.0635.

Toffelmier E, Beninde J, Shaffer HB. The phylogeny of California, and how it informs setting multi-species conservation priorities. *J Hered*. 2022:113:597–603. https://doi.org/10.1093/jhered/esac045.

Udell BJ, Straw BR, ChengTL, Enns K, Gotthold B, Irvine KM, Lausen C, Loeb S, Reichard J, Rodhouse T, et al. *Summer occupancy analysis 2010-2019 (status and trends of North American bats)*. Fort Collins, CO: North American Bat Monitoring Program; 2022. p. 248.

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium, Blaxter M, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, et al. MitoHiFi: A python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*. 2023:24:288. https://doi.org/10.1186/s12859-023-05385-y .

Walker FM, Williamson CHD, Sanchez DE, Sobek CJ, Chambers CL. Species from feces: order-wide identification of chiroptera from guano and other non-invasive genetic samples. *PLoS One*. 2016:11:e0162342. https://doi.org/10.1371/journal.pone.0162342.

Wickham, H. *ggplot2: elegant graphics for data analysis*. New York (NY): Springer-Verlag; 2016. https://ggplot2.tidyverse.org

Yohe LR, Devanna P, Davies KTJ, Potter JHT, Rossiter SJ, Teeling EC, Vernes SC, Dávalos LM. Tissue collection of bats for -omics analyses and primary cell culture. *J Vis Exp*. 2019:152:59505. https://doi.org/10.3791/59505.

Zepeda Mendoza ML, Xiong Z, Escalera-Zamudio M, Runge AK, Thézé J, Streicker D, Frank HK, Loza-Rubio E, Liu S, Ryder OA, et al. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nature Ecol Evol*. 2018:2:659–668. https://doi.org/10.1038/s41559-018-0476-8.

CHAPTER 3: How Myotis move: range-wide genomics of Yuma bats (Myotis yumanensis)

ABSTRACT

As we enter into a period of unprecedented species decline, it becomes more important than ever to identify the best ways to manage species across their range. For many species, there have been no modern attempts to establish management units or reevaluate subspecies groupings, and this can hamper conservation efforts. Furthermore, where efforts to establish management units have occurred, they have primarily been applied to rare species, despite the fact that many common species have experienced similar rangewide declines and provide the bulk of ecosystem services globally. Within this current context, we used high throughput genome resequencing data to evaluate management units and genomic health in a common and widespread species of vespertilionid bat, Yuma myotis (*Myotis yumanenesis*). We find evidence of fine scale population structure, patterns of landscape-level genetic diversity, and historical demography that generally correspond with a post-glacial expansion from southern parts of the species' range. Further, we find that existing subspecies designations are discordant with results using large whole genome marker sets, indicating that these putative subspecies may not represent the best management units for the species. Finally, we find high levels of genome-wide heterozygosity and high contemporary effective population size across most MYYU populations assessed, which overall provides a positive outlook for the conservation status of the species.

INTRODUCTION

In a period of 50 years between 1970 and 2020, global vertebrate populations declined by 73% (WWF 2024), and for most taxa the current rates of extinction far outpace historic baselines (Ceballos et al. 2015). Given this backdrop of decline and extinction, conservation biologists and resource managers must use a variety of tools to halt current extinctions and address the causes of declines to prevent them in the future. One of the foundational steps in this process involves determining if there are meaningful ways to partition the management of species across the landscape – for example, through the establishment of

management units – and to determine if some or all of these population partitions are healthy. Further, given that conservation resources are often limited, management should focus on species that have key functional roles in ecosystems and where proactive management can have the greatest impact (Lindenmayer et al. 2011). For example, species managers may opt to direct resources towards common and abundant species, since they experience similar rates of decline as rare species globally (Daskalova et al. 2020), but represent the bulk of the total biomass (Greenspoon et al. 2023) and disproportionately contribute to ecosystem function compared to rare species (Gaston 2010; Gaston et al. 2018).

There are several means of establishing management units that depend on the species being studied, and these can include designating units based on ecological and behavioral differences between populations (Morin et al. 2024), culturally significant management units (Lamb et al. 2022), and unique evolutionary lineages based on the distribution of genetic variation across a species' range (i.e., Evolutionarily Significant Units, ESUs; Funk et al. 2012; Hohenlohe et al. 2021). Following Moritz (1994), ESUs are populations that have been historically isolated, have distinct adaptive potential, and are characterized by reciprocal monophyly and divergence in allele frequencies. ESUs are an especially important means of managing and protecting species in light of changing climate, as they not only preserve evolutionary heritage but also encapsulate adaptive variation and thus evolutionary potential (Moritz 1994; Barbosa et al. 2018). Further, studying how past climatic fluctuations and human-caused disturbances have shaped modern patterns of genetic diversity across the landscape may provide insights into the future under current human-caused climate change (Hofreiter and Stewart 2009).

Given rapid advancement in genomic technologies and as genomic datasets become more readily available for a variety of non-model organisms, species managers are uniquely positioned to utilize large quantities of genomic data to best determine ESUs. Whole genome data provide a complete picture for the different aspects of the evolutionary history of a species, including both neutral and selective processes shaping modern genomic variation, and a comprehensive analysis of genetic variation at the whole genome level establishes a baseline for analyses now and in the future (Supple and Shapiro 2018). Furthermore, by increasing the numbers of markers used, managers may be able to detect cryptic and

finescale population structure that was not previously described (Emerson et al. 2010; McCartney-Melstad et al. 2018; Sunde et al. 2020; Gallego-García et al. 2021; Dufresnes et al. 2023), including for highly mobile migratory species, where finescale patterns of genomic differentiation can be obscured by recurrent gene flow and lack of significant barriers to dispersal (Moussy et al. 2013; Turbek et al. 2023).

One particularly understudied group of animals in terms of available genomic resources is bats (Order: Chiroptera). For example, efforts to survey the available reference genomes for bats have found incomplete coverage across the bat tree of life, with strong patterns of overrepresentation in just 3 of the 19 recognized bat families (Pteropodidae, Phyllostomidae, and Vespertilionidae; Teeling et al. 2018; Curti et al. 2024). This lack of genomic resources is in line with the general lack of knowledge about the status of bat populations globally, with 18% of bats considered data deficient according to the International Union for the Conservation of Nature (IUCN) – a higher percentage compared to all other mammals (13%) or birds (1%; Frick et al 2020). This knowledge gap hampers the ability of researchers to advance bat conservation, which is critically needed considering the ecosystem services provided by bats through pollination, seed dispersal, or consumption of invertebrate pests (Gallai et al. 2009; Boyles et al. 2011; Kunz et al. 2011; Postel et al. 2012; Frick et al. 2020).

In North America, emerging pathogens, particularly the fungal disease known as white nose syndrome (WNS), also represent a major threat to extant bat species (Hammerson et al. 2017). WNS was first introduced to bat populations in New York in 2006. To date, WNS has been detected in 40 U.S. states and nine Canadian provinces (www.whitenosesyndrome.org/where-is-wns), including most recently in California. In California, WNS is suspected to be present in four counties following the detection of the causative fungal pathogen *Pseudogymnoascus destructans* (Pd) on the carcasses of two Yuma bats, *Myotis yumanensis*, (H. Allen, 1864) in the northwestern corner of the state, in Humboldt county in 2023 (Osborn et al. 2024). These recent confirmed cases in *M. yumanensis* (MYYU) represent a major threat to bat populations in the western U.S., given that WNS has decimated the hibernation colonies of other *Myotis* species in the Eastern U.S. by as much as 73% (Frick et al. 2010). In anticipation of potential declines of MYYU, species managers must begin to develop a framework for guiding

conservation efforts where they are most needed, and a large part of this should involve characterizing the spatial distribution of genetic diversity across the landscape, establishing baselines for current effective population size to track the magnitude of future declines, and establishing ESUs.

Here, we use whole genome re-sequencing data to study rangewide patterns of gene flow and genetic diversity in the Yuma myotis bat. MYYU is ideal for this type of study for several reasons. First, MYYU is still an abundant and widely distributed bat species, with a distributional range from southern Alaska to Morales, Mexico (Braun et al. 2015). Second, MYYU are more closely tied to the presence of fresh water sources than any other bat species of the United States (Barbour and Davis 1969; Brigham et al. 1992; Metcalfe et al. 2023), making them particularly vulnerable to the impacts of human-caused climate change. Finally, while there are currently six recognized subspecies of MYYU (Braun et al. 2015) based primarily on morphological data, including *M. y. lambi* (Benson 1947)*, M. y. lutosus* (G. S. Miller and G. M. Allen 1928)*, M. y. oxalis* (Dalquest 1947)*, M. y. Saturatus* (Say in James, 1823)*, M. y. sociabilis* (H.W Grinnell 1914)*, and M. y. Yumanensis* (H. Allen 1864), their reality as taxonomic or management units has never been adequately tested.

Therefore, this study set out to fill in critical gaps in our knowledge of MYYU populations using thousands of single nucleotide polymorphisms (SNPs). Specifically, our objective was to characterize the following: 1) genetic structure among MYYU populations rangewide and the validity of current subspecies using genomic data; 2) conservation-related genetic measures such as heterozygosity, inbreeding coefficient, and kinship; 3) changes in contemporary and historical effective population size; and 4) if ESUs can be established based on the genomic data. Ultimately, we hope these findings will help to better understand the movement of MYYU (and therefore their diseases) across the western United States and assist species managers through the establishment of ESUs.

METHODS

Sampling, DNA Extraction, and Sequencing

Tissues were derived primarily from four sources for this project. Non-lethal 3 mm biopsy punches were collected from 12 wild bats across California. Bats were captured using mist nets primarily set over fresh water sources. Captured bats were removed from nets and 3 mm biopsy punches were collected from both the plagiopatagium and uropatagium (wing and tail membranes) and stored on silica gel beads until DNA extraction was performed. We also obtained 53 tissues from bats subject to rabies screening by the California Department of Public Health. Yearly, bats that test negative for rabies are surrendered to the California Department of Fish and Wildlife and tissue punches were performed on bat carcasses that did not show visible signs of decay and that had high coordinate resolution for the place of origin. An additional 18 wing punches were harvested from tissues preserved in existing museum collections (Table S3-1). Finally, collaborators from across the western US provided 32 additional MYYU samples, for a total of 115 samples.

DNA was extracted following the standard tissue extraction protocol of Mag-Bind Blood and Tissue DNA HDQ 96 kit (Omega Bio-Tek, Norcross, GA) automated on Nimbus liquid handling robot (Hamilton, Reno, NV). For samples with muscle tissue available, 20 - 30mg of tissue was input into the extraction. For samples without 3mm wing biopsy punches, we performed extractions using both punches collected from the same individual to maximize DNA yield.

We performed quality control on extracted DNA using three methods. First, DNA concentration was determined by using the M3 SpectraMax spectrophotometer (Molecular Devices, San Jose, CA) with AccuBlue Broad Range dsDNA Quantitation chemistry (Biotium, Fremont, CA). A minimum of 100 ng of DNA was submitted for whole genome sequencing, with 500 ng being the target mass. Second, DNA size was determined using gel electrophoresis, with a goal of sequencing primarily high molecular weight DNA >20 kb in length. Finally, in cases where species identity could not be confirmed in the field with acoustics (Rodhouse et al. 2008), we confirmed species identity using Sanger Sequencing of a 202 bp segment of the mitochondrial COI gene using primers (SFF_145f and SFF_351r) and PCR conditions

from Walker et al. (2015) and a ABI 3500XL Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA). Generated COI sequences were manually trimmed and analyzed with Sequencher 5.4.6. (Gene Codes Corp., Ann Arbor, MI). Trimmed sequences were compared to known sequences in the NCBI BLAST database v. 2.15.0 (Camacho et al. 2009) and confirmed that the collected tissue was from MYYU.

Whole genome libraries were constructed from genomic DNA using both a custom version of the plexWell™ WGS24 Library Preparation Kit (seqWell, Beverly, MA, USA) and the KAPA EvoPlus Library Preparation Kit (Roche Sequencing, Indianapolis, Indiana, USA). The Roche KAPA EvoPlus Kit utilized the manufacturers instructions with the following modifications: an initial digest time of 10 minutes, SPRI ratio of 0.8x, and using quarter reaction volumes. Whole-genome sequencing using 150 bp, paired-end reads was performed on Illumina Novaseq 6000 S4 at the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley and Illumina NovaSeq X 25B flow cell at the Center for Applied Technologies at University of California, San Francisco.

All mist netting activities were covered under a permit from the UCLA Institutional Animal Care and Use Committee (IACUC Protocol #: ARC-2021-024) and under a work authorization letter from the California Department of Fish and Wildlife.

Sequence Data Processing and Alignment

Raw sequencing reads were processed using the Snakemake pipeline (https://github.com/ccgproject/ccgpWorkflow/tree/main) which is modified from snpArcher workflow (Mirchandani et al. 2024) and which follows the Genome Analysis Toolkit (GATK) Best Practices (DePristo et al. 2011; Van der Auwera & O'Connor 2020). Quality control and adapter trimming for paired-end reads was performed using fastp (Chen 2023) by enabling the "-detect_adapter_for_pe" flag. Cleaned reads were then aligned to the MYYU reference genome (GCA_028538775.1[MYYU_CA2020_CCGP]; Curti et al. 2024) using bwa-mem (Li 2013).

Sex Chromosome Identification

Sex chromosomes were identified and excluded from all downstream analyses. Sex chromosomes were identified by performing a sequence-to-sequence alignment in minimap2 v. 2.24 (Li 2018) allowing for up to 5% sequence divergence (-asm5) between the MYYU reference genome and the reference genome for the pale spear-nosed bat (GCA_004126475.3; *Phyllostomus discolor*; Jebb et al. 2020). We chose this reference genome because it was highly contiguous and had named sex chromosomes in NCBI. The output of minimap2 was converted to a .bed file using the *sam2bed* function in BEDOPS v. 2.4.41 (Neph et al. 2012), and scaffolds in the MYYU reference genome that mapped to the sex chromosomes in the *P. discolor* reference genome were removed.

Genotype Calling and Filtering

We performed joint genotyping across the MYYU genome using GATK v. 3.8. Prior to joint genotyping, as part of the Snakemake pipeline, PCR duplicates were removed using picard *MarkDuplicates* and genotypes were called for individuals using GATK *HaplotypeCaller*. We then ran joint genotyping across all autosomal scaffolds > 1 Mb in length (1,804,940,391 bp of 1,952,500,671 bp total length, or 92.44%) using GATK *GenotypeGVCF*. We performed hard filtering instead of Base Quality Score Recalibration given that no reference database of known variants is available for the species.

We used genotype-level annotations in the .vcf to determine hard filtering cutoffs. We queried the INFO field annotations in the unfiltered .vcf file using GATK *SelectVariants* to select SNPs and GATK *VariantsToTable* to output annotations. We imported the resulting annotation tables into R studio v. 4.2.2 (R Studio Team 2020) and visualized the distributions to determine appropriate cutoffs. We also determined depth cutoffs by querying the distributions of depth for each individual using a custom python script and determining the 99th percentile of the distribution (Kyriazis et al. 2023). We then filtered our .vcf file using GATK *VariantFiltration* and a custom python script (Robinson et al. 2021) to remove SNPs with low quality and with site-level depth lower than 6x and higher than the 99th percentile of

depth for that individual. Following GATK best practices, we applied the following hard filtering parameters: QUAL < 30, QD < 4.0, FS > 12.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, SOR > 3.0. Finally, we removed repetitive regions identified using RepeatModeler v. 2 (Flynn et al. 2020) and RepeatMasker v. 4.1.2 (Smit et al. 2015).

Relatedness and Population Structure

We removed closely related individuals prior to performing principal component analysis (PCA) and ADMIXTURE analysis to disentangle family and population structure. To do this, we removed individuals with kinship coefficients > 0.0884, which is equivalent to a second-degree level of kinship (i.e., equivalent to the relatedness between children and grandparents). We estimated pairwise kinship between all samples using SNPRelate v.1.32.0 (Zheng et al. 2012) by first converting our filtered .vcf file to a .gds file using *snpgdsVCF2GDS*. Next, we used the King Kinship Estimator by calling *snpgdsIBDKING*. Unless otherwise stated by the package manual, we also pruned our dataset for linkage disequilibrium (LD) using *snpgdsLDpruning* and a r^2 threshold of 0.2 and for minor alleles with frequencies < 0.01 (MAF).

For each sample, we inferred ancestry proportions using ADMIXTURE v. 1.3.0 (Alexander et al. 2009). ADMIXTURE analysis was performed on all unrelated individuals that had been pruned for LD and MAF. We ran ADMIXTURE analysis inferring model fit for K = 1 - 10 population clusters, each run for six iterations. We determined the best fitting model by selecting the K value with the lowest cross validation score averaged across all six iterations. We visualized ADMIXTURE outputs using the R package pophelper v. 2.3.1 (Francis 2017) and mapped the ADMIXTURE proportions using the mapmixture shiny app (Jenkins 2024; https://tomjenkins.shinyapps.io/mapmixture/) To provide additional information on the population clusters identified in the ADMIXTURE analysis, we also ran PCA in SNPRelate using *snpgdsPCA*.

To investigate support for models of K=2-5 clusters produced by ADMIXTURE, we ran Analysis of Molecular Variation (AMOVA). First, we converted our .vcf to a genind object in vcfR v. 1.15.0

(Knaus and Grünwald 2017), assigned population stratifications for putative subspecies and K=2-5 clusters based on majority ancestry proportion for a given cluster, and ran AMOVA in poppr v. 2.9.6 (Kamvar et al. 2014) using the *poppr.amova* function. We tested for significance for these population stratifications using the built in function *randtest* using 999 permutations.

Genetic Diversity and Runs of Homozygosity

To estimate how genetic diversity varies across the MYYU distributional range, we calculated per base pair heterozygosity for all samples. Heterozygosity was calculated in 10Mb sliding windows across the autosomal genome. To control for variation in genotyping rate per scaffold and the potential impact this might have on heterozygosity estimates, we filtered out sliding windows where the number of called sites was below half the window length (5 Mb). We then visualized these results in a manhattan plot using a custom R script.

To estimate inbreeding in MYYU, we calculated the total proportion of the autosomal genome in runs of homozygosity (F_{ROH}). To determine ROHs we used BCFtools RoH v. 1.9 (Narasimhan et al. 2016; Caballero et al. 2020). We then calculated F_{ROH} by summing all ROHs greater than 100 Kb in length by the total length of the scaffolds used in this analysis (1,802,989,729 bp). Following Kyriazis et al. (2023), we estimated the approximate distribution of age classes of autozygous segments by binning the ROHs into size categories (0.1 - 1Mb, 1 - 10Mb, and 10 - 100Mb), where the largest binned group represents ROHs that likely occurred due to recent inbreeding within the past ~10 generations (Mulim et al. 2022).

We also evaluated any significant differences in heterozygosity and F_{ROH} between clusters identified in the PCA and ADMIXTURE analyses. We first assessed the relationship between depth and missingness and levels of genetic diversity, and found a significant relationship between heterozygosity and both depth and missingness as well as F_{ROH} and depth (Figure S3-1). Given this relationship, we used a subset of our data only containing individuals with >15x coverage and <20% missingness (i.e., where the slopes of these relationships began to plateau) for all analyses that set out to determine significant differences in genetic diversity and inbreeding between spatial clusters. For K=2 clusters, we evaluated

differences in heterozygosity and F_{ROH} using a nonparametric Wilcoxon test in the R package stats. For K=3-5 clusters, we used a nonparametric Kruskal-Wallis test in the R package stats, and post-hoc Dunn's Tests using the R package FSA v. 0.9.5 (Ogle et al. 2023) with the built-in function *dunnTest*. Finally, to assess if differences in heterozygosity were consistent with expectations given biogeographic theory, we assessed the relationship of heterozygosity and latitude using a standard regression in R.

Historical Demography

We estimated fluctuations in historical demography in two ways. First we estimated recent historical demography between 50 - 200 years ago using GONE (Santiago et al. 2020). We first downloaded all GONE scripts from github (https://github.com/esrud/GONE) and converted our filtered but not pruned .vcf to .ped format using VCFtools v. 0.1.16 (Danecek et al. 2011) for input into the program. GONE assumes that samples are not admixed and do not contain overlapping generations. Some work has demonstrated that violating the model assumption of non-overlapping generations can lead to downward biased effective population size estimates (Kardos and Waples 2024). To avoid this bias, Kardos and Waples (2024) suggested subsetting samples to only contain individuals from a single cohort, where individuals share the same parents (Kardos and Waples 2024). Unfortunately, the landscape approach of sampling used here prevents us from applying this recommendation. Therefore, we minimized the extent of temporal lumping by selecting samples from as narrow a date range as possible while also balancing the need for a sufficient number of samples to effectively analyze patterns of LD among individuals. We ran GONE using default parameters settings and assuming the average rate of recombination in mammals of 1 cM/Mb on three primary groups of samples: five samples from ID, MT and BC that ranged in sample collection data from 2001 - 2003, 26 samples from CA from north of Point Conception that ranged in collection dates from 2017 - 2021, and five samples from NM, TX, and MX that ranged in collection dates from 2008 - 2016. For all groupings we ran the program 20 times to estimate confidence intervals. In both Northwest US and Canada as well as Southwest US and Mexico samples we observed large drops in N_e in the first 10 - 30 years which are not reflective of best available

data on census population size in these bats. Given the unpredictable nature of GONE estimates in these recent time intervals due to the presence of population structure (Novo et al. 2023) or age structure (Waples et al. 2014), we only present data for the period of 50-200 years before present, where we believe we have the greatest confidence in estimates. For all of these GONE runs we extracted the total difference in effective population size between 50 - 200 years (ΔN_e) by taking the absolute value of the difference of N_e estimates from these time periods.

We also estimated deep historical fluctuations in demography between 10 kya and 10 mya ago using PSMC v. 0.6.5 (Liu and Hansen 2017). Using our filtered but not pruned .vcf, we created a consensus file using BCFtools v.1.11, masking uncalled regions of the genome, and used this as input into PSMC. Following Chattopadhyay et al. (2019), we used default parameters including" 4+25*2+4+6" as the setup for the number of atomic time intervals and free interval parameters. Given that PSMC is sensitive to low coverage and high missingness within samples (Nadachowska-Brzyska et al. 2016) and to population structure (Mazet et al. 2015), we selected 12 samples with >18x coverage, <10% missingness, and from non-admixed locations across the MYYU range. To estimate model robustness, we ran PSMC for 25 iterations, ensuring that log-likelihood reached a plateau by the final iteration and that the model was not overfit by verifying that the expected number of segments in the interval $[t_k,t_{k+1})$ was >10. We also performed 100 rounds of bootstrapping for each sample using the built in *splitfa* function in PSMC. We then imported the PSMC outputs into R to rescale estimates and visualize them using $\mu = 2.366x10^{-9}$ (Ray et al. 2008) and the estimated generation time of 2 years (You et al. 2010).

Phylogenetic Relationships Between Populations

To investigate phylogenetic relationships between samples and to determine if putative subspecies designations would cluster together in a phylogenetic analysis, we created a phylogenetic network using SplitsTree v. 6.0.0 (Huson and Bryant 2024). First, we subsetted our .vcf to 43 individuals to provide coverage across five of the six MYYU subspecies, including two *M. y. lutosus*, nine *M. y.*

yumanensis, five *M. y. oxalis*, six *M. y. sociabilis*, and 20 *M. y. saturatus* samples. We did not have any *M. y. lambi* samples. We then converted our .vcf to a .nexus format using vcf2phylip v. 2.0 (Ortiz 2019) for input into splitstree. In splitstree, we created a neighbor net tree of a P-distance matrix (Hamming 1950) using default parameter settings (Bryant and Moulton 2004; Bryant and Huson 2023). Additionally, we used IQ-TREE v. 2.3.6 (Minh et al. 2020) to construct a maximum likelihood tree based on the GTR model with 1,000 bootstrap replicates. For this analysis, we used the concatenated autosomal SNPs, filtered to include loci present in 95% of sampled individuals (866,389 SNPS) to accommodate computational limitations.

RESULTS

Genomic Dataset

After genotyping and filtering the dataset, we were left with a total of 115 samples with a mean sequencing coverage of 15.7 (range = 6.21–37.4). Missingness was generally low, with an average per sample percent missingness of 8% (range = 1–41%). After filtering, our dataset contained a total of 1,804,939,434 sites and 43,240,875 SNPs.

Relatedness and Population Structure

Across the 115 rangewide MYYU samples, we found high levels of shared ancestry between two California samples collected in San Diego county, with a first-degree relationship (kinship coefficient > 0.35) equivalent to a monozygotic twin. We only retained the sample with the higher coverage in all subsequent analyses. Across the remaining samples we detected low level kinship that did not reach our threshold for removal (kinship coefficient < 0.0884; Figure S3-2).

Principal Component Analysis (PCA) showed evidence of fine-scale population structure across the range of MYYU (Figure 3-1). PC axes 1 explained 2.55% of the variation and axis 2 1.67% of the variation in our dataset, and they revealed two primary clusters. Cluster 1 contains samples from California, Idaho, Washington, British Columbia, and Alaska and approximately recapitulates the shape

of Western North America, consistent with a general pattern of isolation by distance. Cluster 2 contains all samples from Nevada, New Mexico, Utah, Colorado, Oklahoma, and Mexico, with the sample from Oklahoma separating the furthest from all samples in PC space. Further, given the overrepresentation of California samples in our dataset, we randomly down sampled California and reran the PCA, and the results are generally consistent with those derived using the entire dataset (Figure S3-3), with a couple of exceptions. First, in the reduced dataset, samples from Catalina Island in California and Oklahoma separated away from the rest of all other samples along both PC axis 1 and 2. Further, samples from Colorado and Utah do not appear as distinct from Cluster 1 in the reduced dataset as they do in the full dataset.

Our ADMIXTURE analysis was concordant with the findings of the PCA. Although cross validation analysis shows that the model with the highest support is K=1 cluster (Figure S3-4), outputs of K=2 align with the clusters present in the PCA (Figure 3-2A), grouping all samples from the Western US and Canada together and all samples from the Southwestern US and Mexico together. We also visualized K=3 through K=5, or the number of putative subspecies of MYYU within our sample. In K=3, samples from Los Angeles County group together with a zone of introgression between southern California and the rest of the state present just north of Point Conception extending approximately 161 km (~100 miles) in either direction (Figure 3-2B). Outputs of K=4 further split northern California from the rest of the western US and Canada samples, and K=5 splits Baja California, Nevada, Utah and Colorado from the rest of the Southwest US and Mexico samples (Figure S3-4). Finally, we assessed whether or not putative subspecies for MYYU were captured in the ADMIXTURE outputs of K=5 clusters and this analysis provided little support for the subspecies groupings (Figure 3-3).

We ran AMOVA to estimate support for putative subspecies groupings as well as K=2-5 models of population structure output by ADMIXTURE. Subspecies groupings explained 3.24% of the variation between samples and the remaining 96.8% of variation was explained by variation within samples ($\Phi ST = 0.032$, $df = 4$, $p = 0.001$). Among the different values of K, K = 4 clusters explained the highest amount of variation between groups at 4.45%, while the remaining 95.5% of variation was explained by variation

within samples (ΦST = 0.045, df = 3, p = 0.01). AMOVA results for all values of K are presented in Table 3-1.

Genetic Diversity and Runs of Homozygosity

Per base pair heterozygosity varied across the MYYU range (mean = $1.91x10^{-3}$, range = $1.05x10^{-3}$ - $4.57x10^{-3}$; Figure 3-4A). Generally, heterozygosity decreased with latitude ($R^2 = 0.471$, $F(1,51) = 47.21$, $p < 0.001$; Figure 3-5). Using K=2 clusters, samples from the western US and Canada had significantly lower mean heterozygosity (mean = $1.65x10^{-3}$, range = $1.05x10^{-3}$ - $2.35x10^{-3}$) compared to samples from the Southwest US and Mexico (mean = $2.81x10^{-3}$, range = $1.68x10^{-3}$ - $4.57x10^{-3}$; Wilcoxon Rank Sum, $W = 193$, $p < 0.001$; Figure 3-6A). There were also significant differences in mean heterozygosity under K=3 clusters (Kruskal-Wallis, $\chi^2 = 42.2$, df = 2, $p < 0.001$; Figure 3-6C). Specifically, post-hoc Dunn's tests with Bonferroni corrections revealed that samples from the western US and Canada excluding southern California had significantly lower mean heterozygosity (mean = $1.45x10^{-3}$, range = $1.05x10^{-3}$ to $1.84x10^{-3}$) compared to samples from the Southwest US and Mexico (mean = $3.59x10^{-3}$, range = $2.65x10^{-3}$ to $4.57x10^{-3}$; Dunn's Test, z-value = 5.73, adj. $p < 0.001$), and that samples from southern California had significantly lower mean heterozygosity (mean = $2.06x10^{-3}$, range = $1.60x10^{-3}$ to $2.35x10^{-3}$) compared to samples from the Southwest US and Mexico (Dunn's Test, z-value = 4.71, adj. $p < 0.001$). We did not find a significant difference in mean heterozygosity between the western US and Canada and southern California (Dunn's Test, z-value = -2.12, adj. p = 0.101). This pattern was generally consistent at higher values of K (Figure 3-6E & 3-6G).

On average, estimates of inbreeding measured as F_{ROH} were low across the MYYU range (mean = 4.00%, range = 0.17 - 24.4%; Figure 3-4B). A few samples at the edge of the estimated MYYU range had elevated levels of F_{ROH}, including the samples from Montana (F_{ROH} = 24.4%), Alaska (F_{ROH} = 24.3%), Catalina Island, CA (F_{ROH} = 22.3%), Idaho (F_{ROH} = 16.3%) and Baja California (F_{ROH} = 15.1%). We also evaluated the extent to which F_{ROH} varied across population structure with K=2-5 and these did not reach the level of statistical significance with the exception of K=5 (Figure 3-6B, 3-6D, 3-6F, 3-6H). ROHs in

all samples except 3 were < 10Mb. Specifically, the MYYU sample collected from Oroville, CA had a single large ROH of 10.25 Mb, the sample from Catalina Island, CA had a single large ROH of 11.04 Mb and the sample from Chester, CA had a large ROH of 12.84 Mb in size (Figure 3-7).

Historical Demography

Estimates of recent effective population size (N_e) using GONE were large and stable in each of the MYYU populations we investigated. Specifically, GONE estimates of N_e were greater than 3,000,000 for all populations assessed (Figure 3-8). For samples from California north of Point Conception, N_e estimates did not fluctuate greatly between 50 ya to 200 ya, with a total change in N_e of 3,016, and this was two orders of magnitude larger in samples from NM, TX, and MX (ΔN_e = 337,530) and in samples from BC, ID, and MT (ΔN_e = 427,889).

To further contextualize modern estimates of N_e, we ran PSMC to estimate fluctuations in deep historical timescales (between 10 kya - 10 mya). These analyses revealed two distinct demographic patterns for rangewide MYYU populations during this period. Specifically, MYYU samples from the Southwest US and Mexico had higher N_e at the start of the Last Interglacial Period (mean = 517,903, range = 370,081 - 668,772) relative to MYYU samples from the western US and Canada (mean = 141,248, range = 123,452 - 161,707; Figure 3-9). Focusing on the period between the Last Interglacial Period and the Last Glacial Maximum, 67% (4/6) of MYYU sampled from the western US and Canada experienced a population expansion during this time, compared to 40% (2/5) from Southwest US and Mexico. Finally, we also observed that while historical N_e estimates were somewhat variable between samples (Table 3-1; Figure S3-5), MYYU from the Southwest US and Mexico generally entered the mid-Holocene with higher estimates of N_e (median = 9,903,876 , range = 713,236 - 38,033,338) when compared to samples from Western US and Canada (median = 6,402,832, range = 50,580 - 9.94×10^{10}).

Phylogenetic Analysis

Using splitstree we created an unrooted phylogenetic network that fit our data well (Fit parameter = 99.9; Figure 3-10). In this phylogenetic network, none of the MYYU subspecies are grouped in a

monophyletic clade, and generally two clusters are apparent separating *M. y. yumanensis* and *M. y. lutosus* from all other MYYU subspecies. This generally corresponds to the geographic clustering observed in ADMIXTURE and PCA results. We also constructed a phylogenetic tree using a maximum likelihood approach with 1,000 bootstrap replicates, and branch placements had high support based on bootstrap support values >86%. We observed concordant results between the splitstree phylogenetic network and our maximum likelihood tree, where samples of Southwest US and Mexico are paraphyletic to a single fully supported clade containing all samples from the western US and Canada, with none of the MYYU subspecies grouped in a monophyletic clade.

DISCUSSION

Bats are critical components of functioning ecosystems worldwide and bat populations are threatened by a multitude of primarily human-caused stressors (Frick et al. 2020). As many bat populations decline in abundance, it becomes more important than ever to utilize new and innovative ways to direct conservation resources in the most efficacious manner possible. Here we use whole genome resequencing data to provide the first genomic assessment of MYYU to investigate rangewide population structure, genetic diversity, and historical demography. We found evidence of finescale population structure across the range of MYYU that roughly corresponds to geography. Further, we find high levels of genetic diversity across most populations of MYYU studied, with some interesting regions of low diversity and inbreeding at the leading edge of the species' range. We also found that methods of measuring fluctuations in historical population size were generally concordant between deep historical time and the recent past, and that overall these estimates demonstrate high effective population size for MYYU. Finally, we find that phylogenetic methods provide little support for current subspecies designations for MYYU. These data come in light of recent news that WNS was recently detected in MYYU populations in previously WNS-free regions of California, further emphasizing the need for baseline genomic data prior to anticipated declines due to this advancing pathogen.

Subtle population structure in a mobile bat species

Previous morphometric work has demonstrated substantial phenotypic variation in MYYU across its distributional range which has been used to establish six putative subspecies (Braun et al. 2015). While no landscape genomic methods have been applied to this species to date, Korstian et al. (2024) found evidence of genetic differentiation in a few samples of MYYU derived from ultraconserved elements across the genome, although they did not describe the spatial variation in this structure or its relationship with putative subspecies. It is well established that population structure can be challenging to detect in migratory species such as bats, especially given that bats typically exhibit uneven gene flow across the landscape (e.g., strong female philopatry and male-biased dispersal), which can obscure signals of population structure (Moussy et al. 2013; Collevati et al. 2020). While data on sex-biased dispersal is not known for MYYU and consensus on migratory behavior has not yet been reached (Braun et al. 2015), MYYU are capable of nightly foraging flights up to 4-km away from their roosting structures (Nagorsen and Brigham 1993; Evelyn et al. 2004). Despite their vagile nature, we found that MYYU had subtle, yet detectable, population structure across their distributional range based on PCA, ADMIXTURE, and phylogenetic approaches. Specifically, we found that across all clustering methods used, samples from the Southwest US and Mexico clustered together and separated out from the rest of our MYYU samples. Increasing values of K in ADMIXTURE provided finer scale resolution on genetic differentiation between groups of MYYU, including separating out southern California from the rest of the samples in the state at K=3, separating out samples from AK, BC, WA, ID, and MT from CA at K=4, and separating MYYU in Southwest US and Mexico into two clusters of bats from NV, UT, CO, and Baja California Sur, and bats from NM, TX, and mainland MX in K = 5 clusters. These results contrast with those from little brown bats (*Myotis lucifugus*), a congener with a similar widespread nearctic distribution and well studied migratory behavior, which show little evidence of genetic differentiation across much of their distributional range (Vonhof et al. 2015; Gignoux-Wolfsohn et al. 2021).

It is possible that the clustering observed in the genomic data can be explained by transitions in ecoregions throughout the MYYU range. Given that MYYU are so closely tied to the presence of open

water sources and that geographical barriers such as the Strait of Gibraltar have been shown to be important in genetically structuring other *Myotis* species (García-Mudarra et al. 2009; Ruedi et al. 2013), it is possible that the arid landscape at the confluence of the Central, Sonoran and Mojave Basins acts as a substantial barrier to MYYU populations, which may explain the apparent clustering of bats east and west of this divide in the PCA and the lack of admixture between these populations in population structure analysis at K=3-5. Efforts to model habitat suitability in North American bat species have demonstrated that drought and habitat represent major factors shaping past, present and future bat distributions (Chattopadhyay et al. 2019; Piccioli Cappelli 2021; Smith et al. 2021; Hamilton et al. 2022), and thus future studies can build upon these hypotheses by combining genomic sequencing data provided here with species distribution modeling, assessing whether regions separating these two primary clusters in our data represent areas of reduced habitat suitability. Further, as the distance that bats travel between roosting and foraging sites is highly dependent on the heterogeneity of vegetation close to their roosting sites (Laforge et al. 2021), it is possible that the xeric nature of these landscapes make it impossible for MYYU to perform longer distance movements to forage and reproduce, and studies utilizing radio telemetry methods in this region could help illuminate how landscape factors relate to apparent genetic connectivity of MYYU populations.

Modern genomic data as an echo from the past

The currently accepted biogeographic theory of the origins of *Myotis* in the nearctic is through colonization of North America in the mid-miocene across the Bering Strait followed by vicariant evolution between nearctic and neotropical myotis leading to many extant lineages including *M. yumanensis* (Ruedi et al. 2013; Korstian et al. 2024). Best available evidence indicates that MYYU lineages were also shaped by changing geographic and climatic landscapes during the Quaternary Period, with a presumed single climatic refugia for the species in the lower Colorado River basin that later expanded into eastern Mexico and into northwestern North America following the retreat of the Laurentide Ice Sheet (Harris 1974).

Under this biogeographic scenario, we might expect to see a directional gradient of genetic variation decreasing outward from the lower Colorado River basin, into Mexico and up the western coast of North America (Hewitt 1996; Eckert 2008). Our data provide partial support for this biogeographical explanation. First, genome-wide heterozygosity levels decrease as a function of increasing latitude. Counter to our expectation, this lack of a bidirectional signal might indicate that the broader region in the Southwestern US and Mexico represented one large glacial refugia during the last glacial maximum, or that a more complex refugia scenario is possible (e.g., Shafer et al. 2010). In fact, Chattopadhyay et al. (2019) modeled habitat suitability for the big brown bat (*Eptesicus fuscus*), another insectivorous vespertilionid bat in North America, and found that suitable habitat during the Last Glacial Maximum extended into Mexico and throughout much of California for this species. The extent to which this applies to MYYU remains to be determined and will require additional work to specifically tailor this modeling approach given the unique natural history of MYYU. Another possible explanation for these patterns of genetic diversity could be potential hybridization in Mexican MYYU, which would increase heterozygosity in hybrid individuals and obscure patterns of historical biogeography in the genomes of these animals. Some unpublished work from the region (Cortazar 2020) provides evidence that MYYU are capable of hybridization with several other species including California myotis (*M. californicus*) and peninsular myotis (*M. peninsularis*), providing some support to this hypothesis. While we do not explicitly investigate hybridization, the high coverage whole genome resequencing data provided here in combination with re-sequencing of other *Myotis* species will enable investigation into exactly how widespread hybridization is across the species' range.

In addition to patterns of heterozygosity, past climate can also leave a signature on historical demography of organisms (Mestre et al. 2022), and various methods to estimate Ne have been used to provide insights into vertebrate biogeography related to quaternary fluctuations (Dussex et al. 2020; Yang et al. 2022; McCarthy et al. 2024). Under this biogeographical explanation we might also expect to see differing historical demography in populations of MYYU within the region of the historic climatic refugia compared to the leading edge of expansion. Specifically, we might expect lower effective population size

at the range edge compared to the core (Eckert et al. 2008), perhaps as a result of increased mutational load at the range edge that can reduce fitness and decrease population size, thus enforcing the species range boundaries (Willi et al. 2018). Based on PSMC analysis, we found that populations of MYYU within the Southwest US and Mexico had consistently higher effective population sizes during the Last Interglacial Period compared to MYYU from the rest of the range, and that this pattern was observed across all samples assessed. We also found that MYYU within the Southwest US and Mexico also had consistently higher effective population size going into the mid-Holocene, although these estimates for the most recent time period in PSMC analysis bear the highest uncertainty (Li and Durbin 2011) reflected in increased size of confidence intervals in these results. Taken together, these results also suggest that MYYU in the Southwest US and Mexico represent a unique demographic population that corresponds to quaternary climatic fluctuations. Compared to other similar efforts, Chattopadhyay et al. (2019) found that insectivorous bats generally exhibited overall high effective population size and a decline in N_e between the LIG and LGM, which agrees with our findings for MYYU. However, our results disagree for values of N_e going into the Holocene, which represents the period of lowest overall N_e for other bat species in their study but represents the period of highest N_e for MYYU, with the exception of one bat from Catalina Island which we discuss further below. Our estimates of modern N_e based on GONE outputs provides further support for our PSMC results, as we generally observe high effective population sizes in the last 50 - 200 years, although we lack any estimates for the period between 200 ya - 10 kya.

Our findings here represent novel insights into the role of Quaternary fluctuations on the evolutionary history of MYYU in North America. These findings are important in the face of climate change that will likely lead to range contractions and possibly extirpation of many bat species worldwide (Festa et al. 2023). For example, information about the historic movement and recolonization of bats following periods of global warming can provide critical Information required to estimate the rates at which species are tracking the movement of habitats in response to climate change (Provan and Bennett 2008). Further, information on historical fluctuations in population size for species can have implications for conservation management in the present, as population size prior to bottlenecks can have a direct

impact on the amount of deleterious variation that a population harbors which can be expressed when populations go through sudden contractions (Kyriazis et al. 2021). Information about the mutational load of species is being readily used in conservation as genomic data become easier and more cost-effective to generate (Robinson et al. 2023). Of course, the correspondence of our estimates of N_e to climatic oscillations is based on rescaling using the inferred mutation rate for another *Myotis* spp., therefore future work could build upon these estimates by deriving a mutation rate for MYYU and assessing if these effective population size estimates change. Ultimately, our findings here provide a baseline for more in depth analysis aimed at further elucidating the role of historical climate on patterns on MYYU genetic diversity and effective population size.

Do subspecies designations represent the most meaningful management unit for MYYU?

In the absence of genomic data, subspecies designations based on morphometric data alone are often used by species managers to direct conservation actions across a species' range. We asked whether or not our rangewide genomic dataset for MYYU support existing morphometric-based subspecies designations. Based on ADMIXTURE and phylogenetic analyses, we found little support for existing subspecies designations. Specifically, in the ADMIXTURE analysis, when individuals are grouped by putative subspecies there is evidence of admixture between and within putative subspecies groupings. Further, these admixed individuals are not just found along boundaries between putative subspecies but in the core of their purported distributions. Similarly, based on the results of the phylogenetic network analysis and the maximum likelihood tree, subspecies did not group exclusively in any of the major clusters identifiable in the network or clades in the phylogenetic tree. Instead, samples are generally clustered in a manner consistent with the major groupings identifiable in the PCA and ADMIXTURE analysis. Finally, AMOVA results similarly demonstrate that subspecies designations are among the lowest determinants of genetic structure based on the percentage of variation explained between groups. Taken together, these results suggest that there are inconsistencies between putative subspecies for MYYU and molecular data that bring the validity of these subspecies into question. This overall lack of

concordance between molecular and morphometric data is not completely surprising, as several recent studies in bats have revised formerly accepted taxonomy in light of molecular data (Brown et al. 2011; Vonhof et al. 2015; Morales et al. 2016). These studies demonstrate that for a traditionally understudied group of organisms such as bats, we still have much to learn regarding the evolutionary histories of species. Although the results here converge on a general lack of support for putative subspecies designations in MYYU, we encourage future studies to take an integrative approach to defining conservation management units, incorporating both acoustic and morphometric data with the genomic data presented here (e.g., Solari et al. 2019).

In the absence of subspecies designations to group and manage MYYU, the logical next step is to ask 'what are the best means of grouping MYYU across their range?' Based on the genomic data provided here we have strong evidence for the presence of at least two major groups of MYYU, between Southwest US and Mexico samples and the remainder of samples across the range, and this split is generally supported by differences in genome-wide heterozygosity and demographic independence between the groups. However, low levels of differentiation seen in the AMOVA results between clusters as well as disagreement between the optimal K value in ADMIXTURE analyses and the separation of these clusters in PCA space might indicate that there is recurrent gene flow between these groups. Additional work to evaluate connectivity is needed in order to properly assess the degree of distinctiveness of these clusters. We also identify several samples across the range of MYYU that exhibit elevated levels of inbreeding, including samples on Catalina Island and at the range edge in Alaska, Montana, and Baja California, Mexico. When summed across the entire genome, average levels of F_{ROH} for MYYU (mean = 4.00%, range = 0.17 - 24.4) far exceeds documented average F_{ROH} for other bat species that have been assessed (mean = 0.096%, range = .00020 - 0.978; Brüniche-Olsen et al. 2018). To some extent, these results are not surprising, as inbreeding on the California Channel Islands is well established in other mammal species (Robinson et al. 2016) and the sample in Alaska represents a recent range expansion (Olson et al. 2014) where census populations sizes are expected to be exceedingly low, thus increasing the probability of inbreeding. However, the degree of inbreeding in these samples may be

of concern for conservation, as MYYU in these regions show increased levels of medium (> 1Mb) and long (>10 Mb) ROHs, which can be enriched in putatively deleterious variation (Szpiech et al. 2013) and have direct fitness and demographic consequences in wild mammal populations (Stoffel et al. 2021).

Overall, high levels of genome-wide heterozygosity and high effective population size estimates for the past 50 - 200 years paints a positive picture for MYYU conservation, and these conclusions would not be possible without the comprehensive genomic dataset we present here. We show that genomic datasets can reveal interesting patterns of rangewide population structure for a highly mobile species, contrasting with results for other bat species using reduced marker sets. As bat species continue to experience impacts due to habitat destruction and human-caused climate change, it is important to use all available resources to document common species prior to declines in order to establish appropriate baselines for the future.

FIGURES

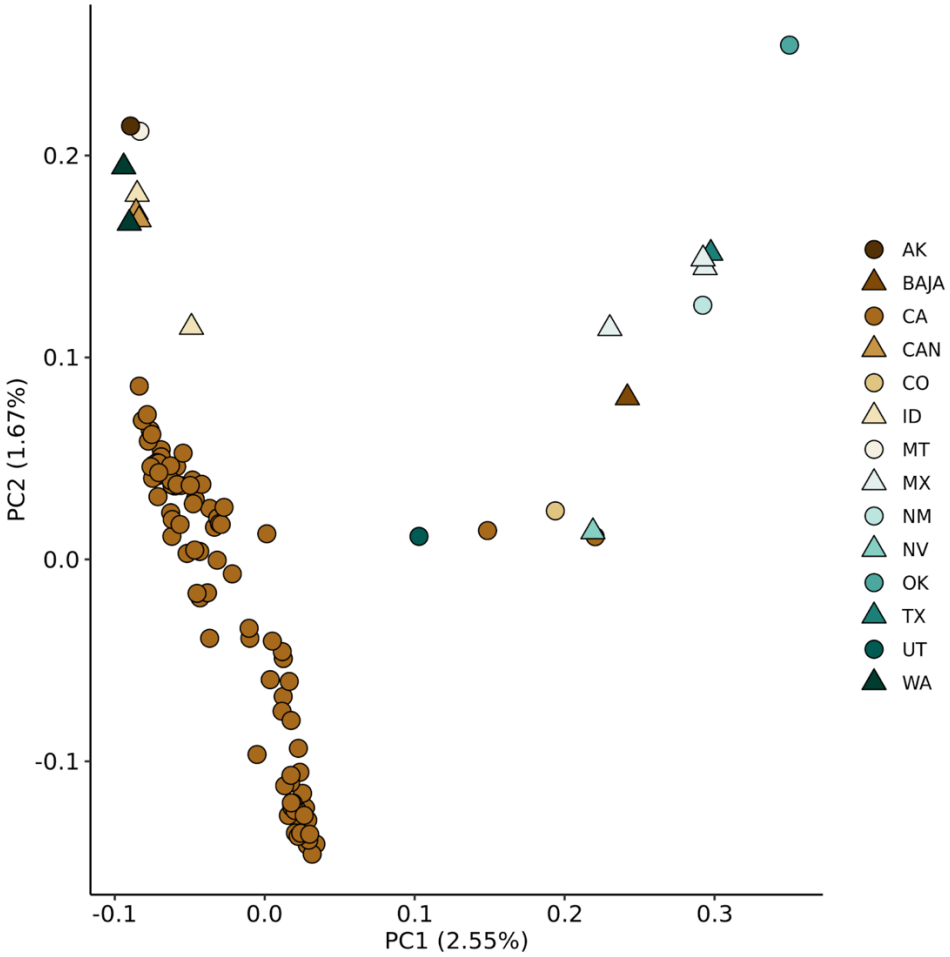

Figure 3-1. Principal Component Analysis of 115 MYYU samples using 143,496 unlinked SNPs. Shape and color indicate sampling location.

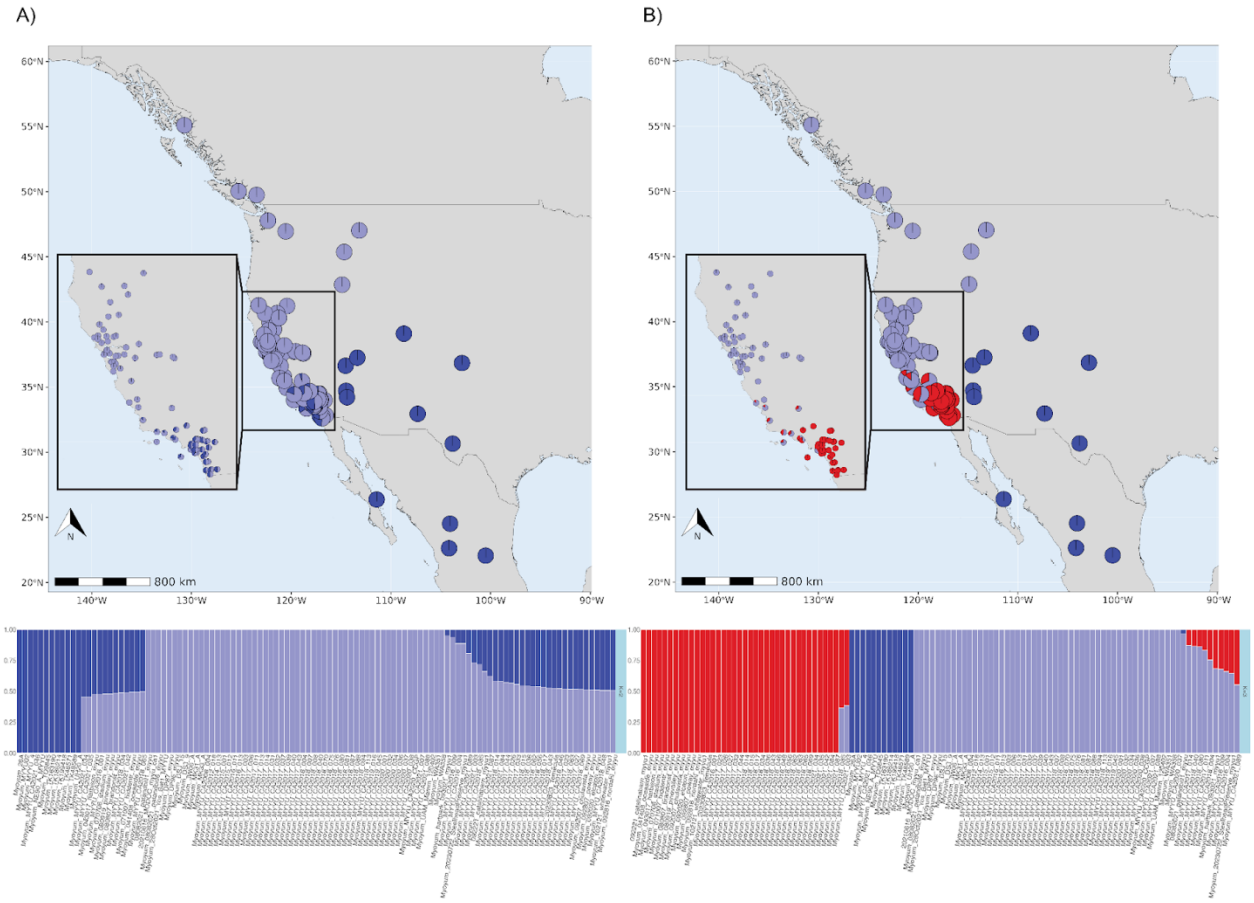

Figure 3-2. ADMIXTURE analysis outputs of sample-level ancestry proportions for A) K=2, and B) K=3 populations. Color of bar indicates proportion of ancestry belonging to each cluster for each individual.

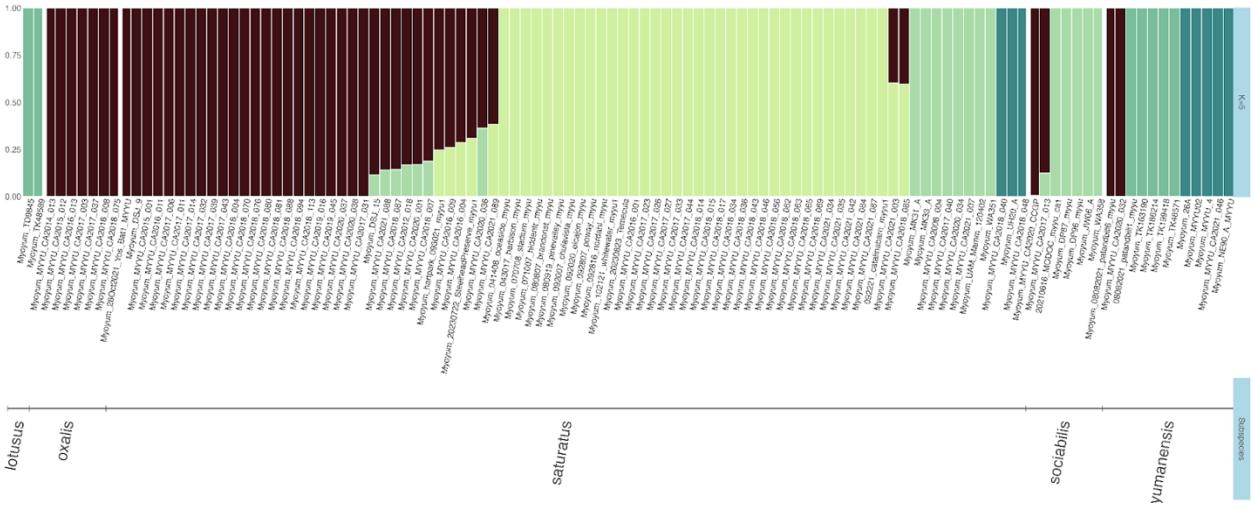

Figure 3-3. ADMIXTURE output for K=5 populations, grouped by putative subspecies. Color of bar indicates proportion of ancestry belonging to each cluster for each individual. Grouping labels indicate putative subspecies for samples.

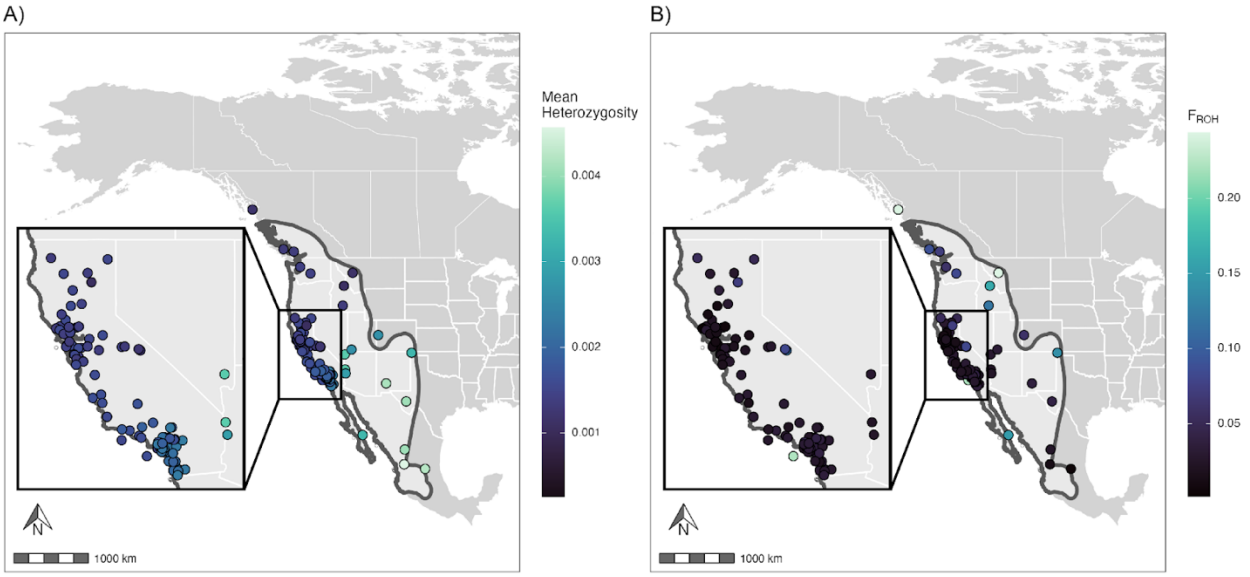

Figure 3-4. Map of study samples detailing rangewide A) levels of per-base pair heterozygosity, and B) proportion of the genome in runs of homozygous alleles greater than 100 Kb (F_{ROH}).

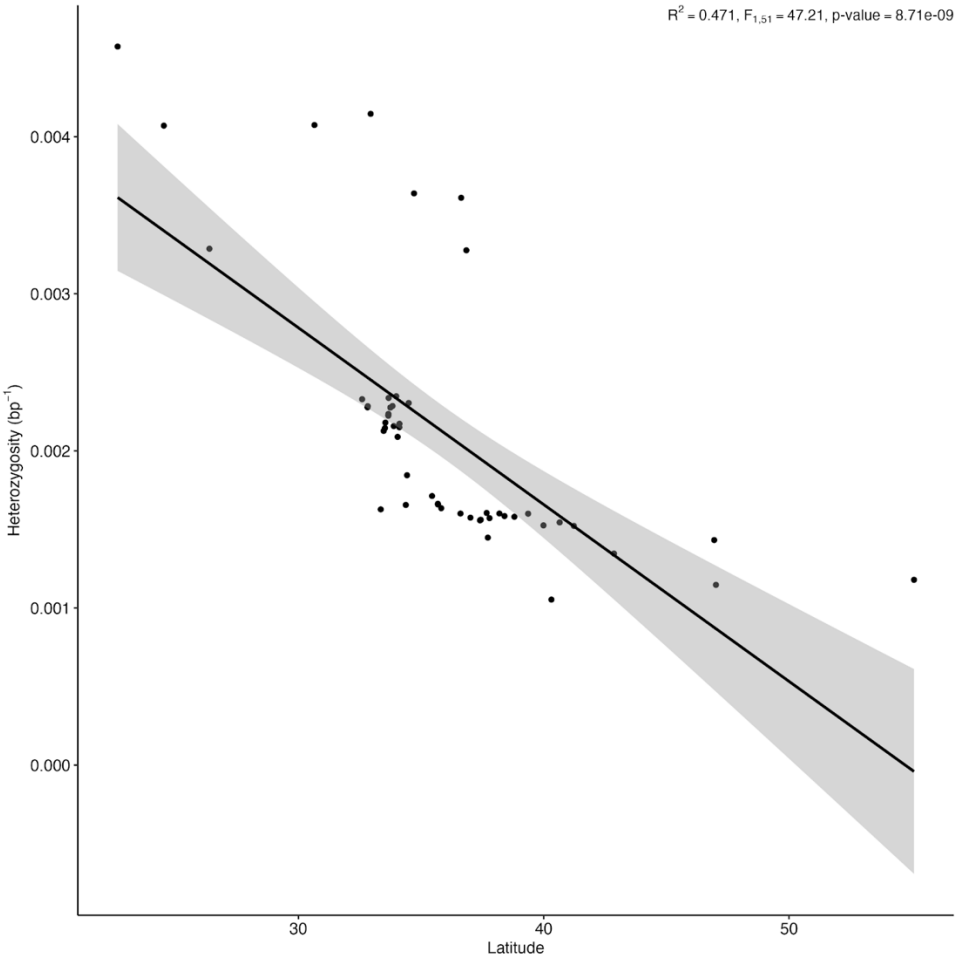

Figure 3-5. Scatterplot of per base pair heterozygosity 53 MYYU samples with >15x coverage and <20% missingness plotted against latitude.

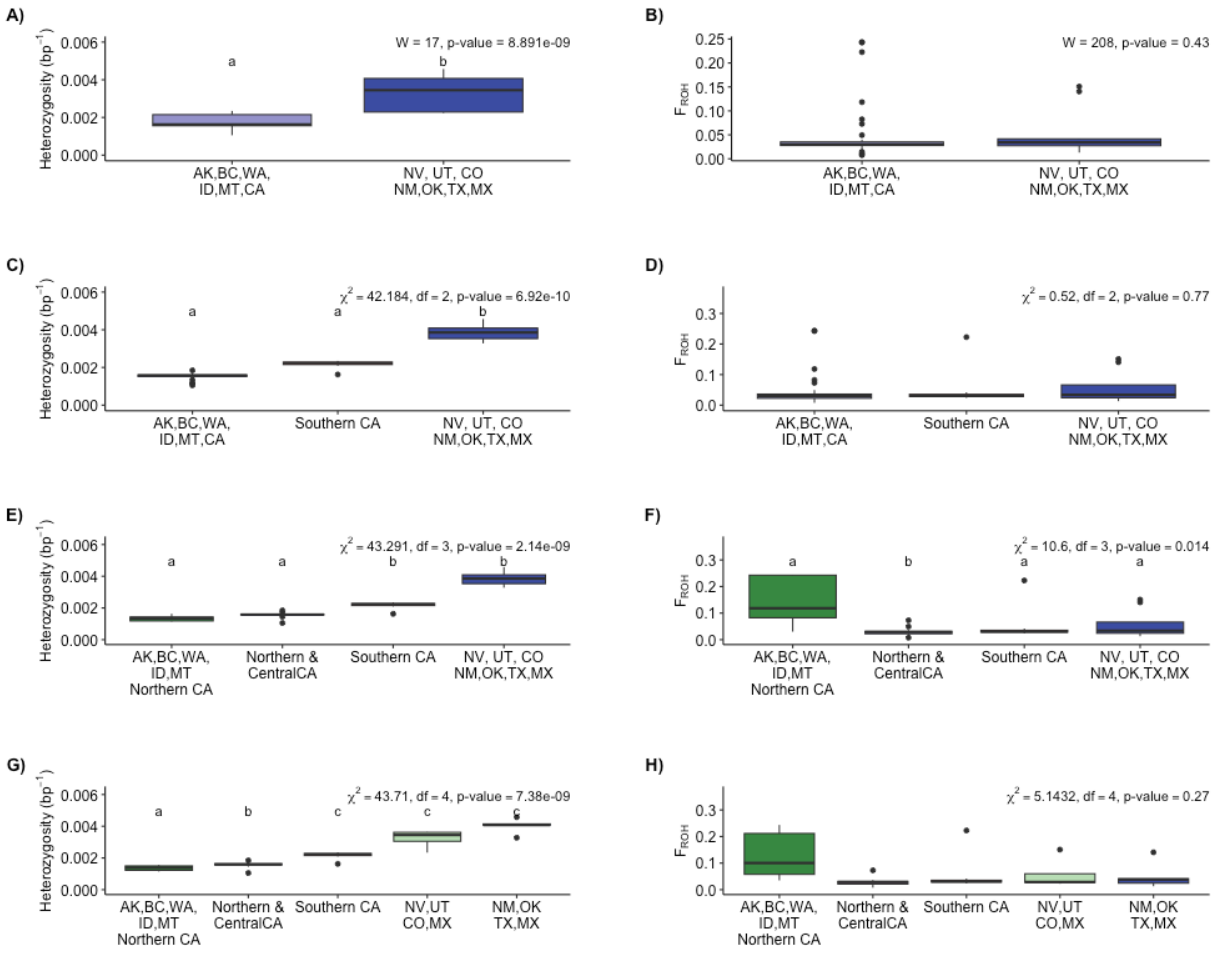

Figure 3-6. Barplots of per base pair heterozygosity and F_{ROH} between clusters assuming A-B) K=2 clusters, C-D) K=3 clusters, E-F) K=4 clusters, and G-H) K=5 clusters. Letters indicate significant differences between groupings.

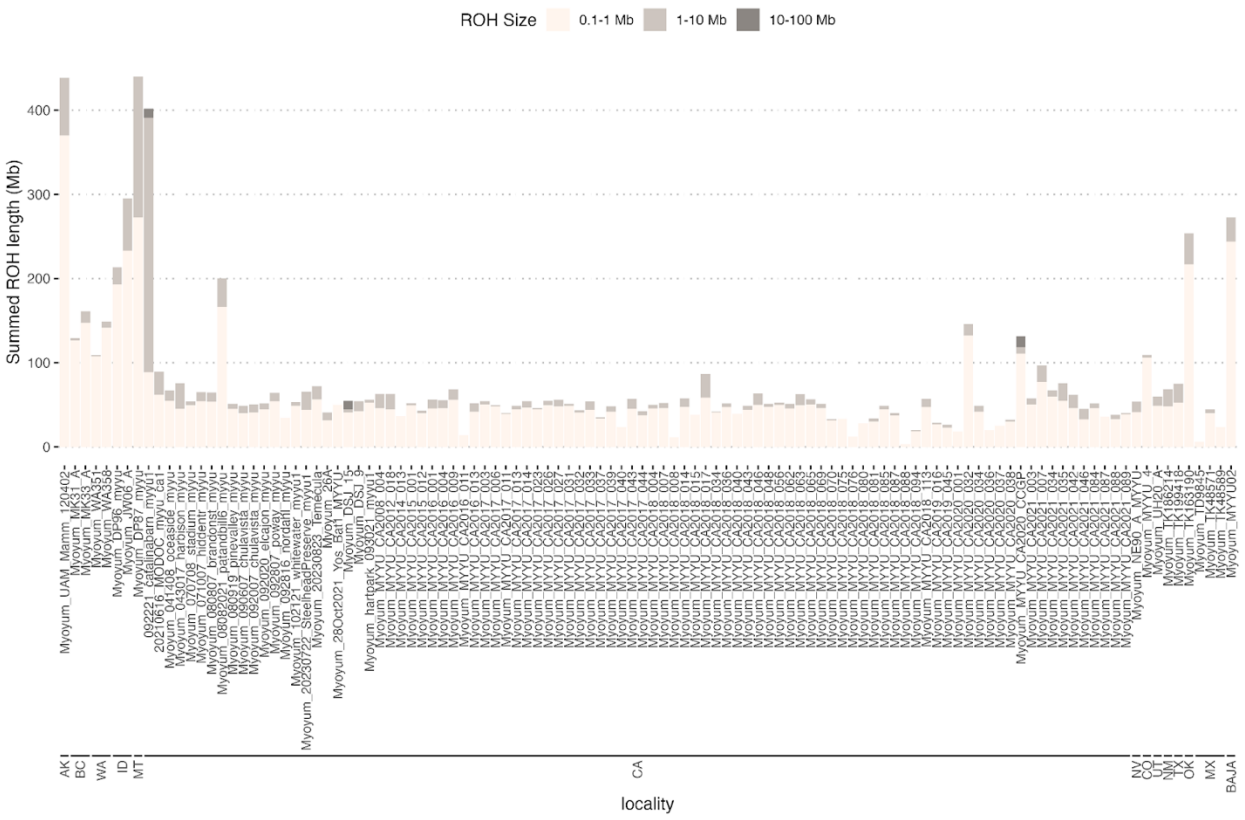

Figure 3-7. Sample-level summed lengths of runs of homozygosity, binned into three groups: [100 Kb - 1Mb), [1Mb - 10 Mb), ≥ 10Mb.

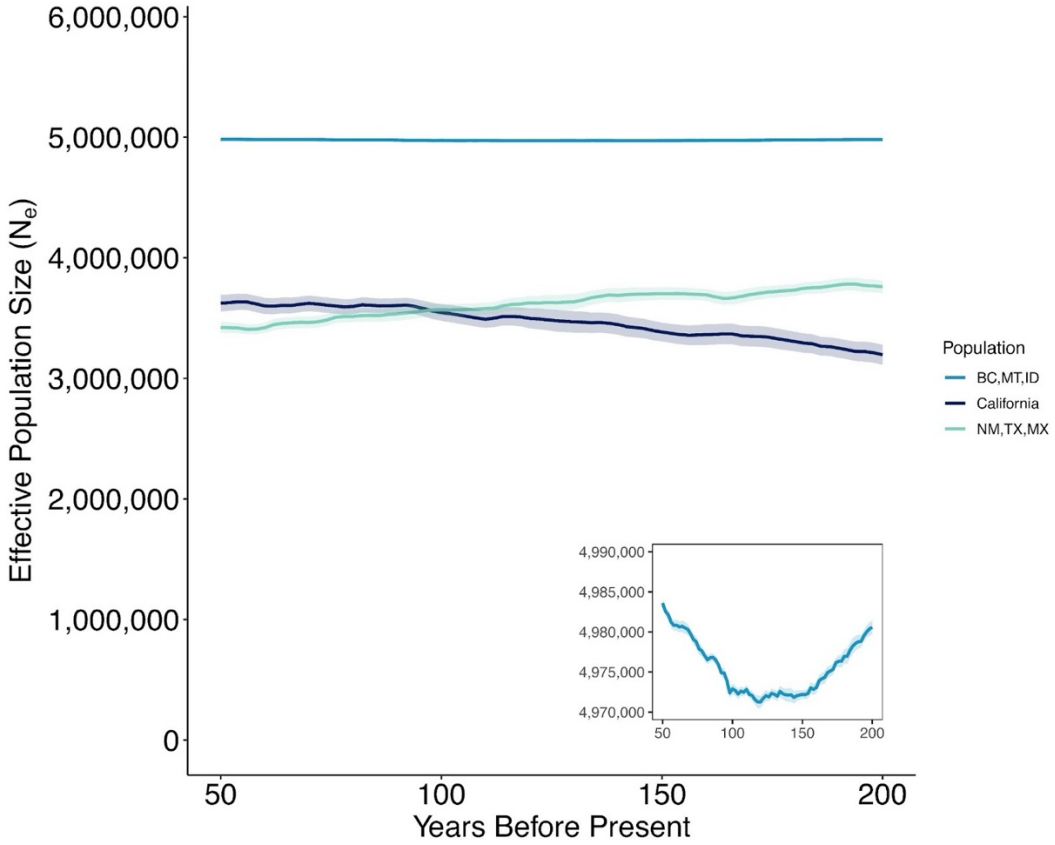

Figure 3-8. Estimated fluctuations in effective population size (N_e) using GONE between present day and 200 years ago for three populations of MYYU. Populations include British Columbia, Montana, and Idaho samples (medium blue), California samples (dark blue), and New Mexico, Texas, and Mexico samples (teal). Inset shows additional detail on California MYYU population N_e trajectories.

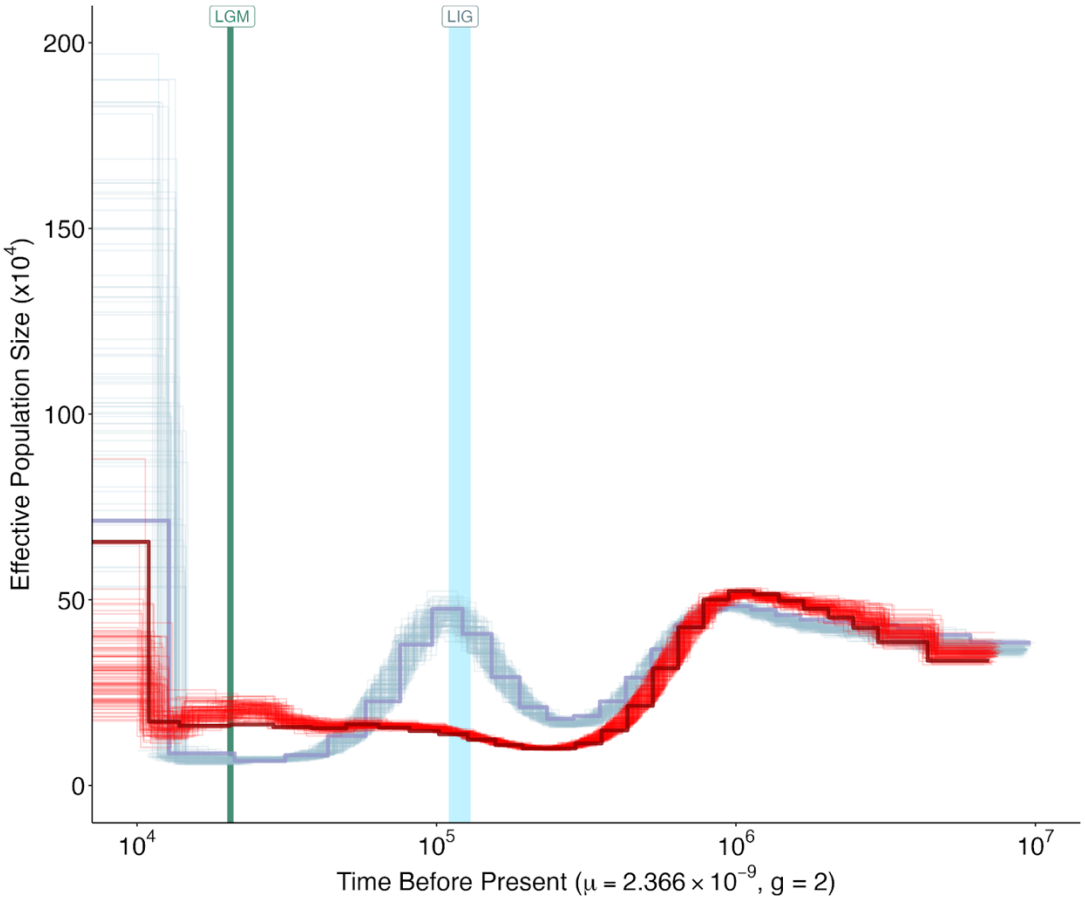

Figure 3-9. Estimated fluctuations in N_e using PSMC between 10 kya and 10 mya for two individuals. The red sample is representative of the general demographic trajectory of MYYU sampled from the western US and Canada, while the blue sample is representative of the trajectory of MYYU sampled from the Southwest US and Mexico. Thick line is the full run of PSMC while the thinner lines are 100 bootstrap replicates. LGM = Last Glacial Maximum, LIG = Last Interglacial Period.

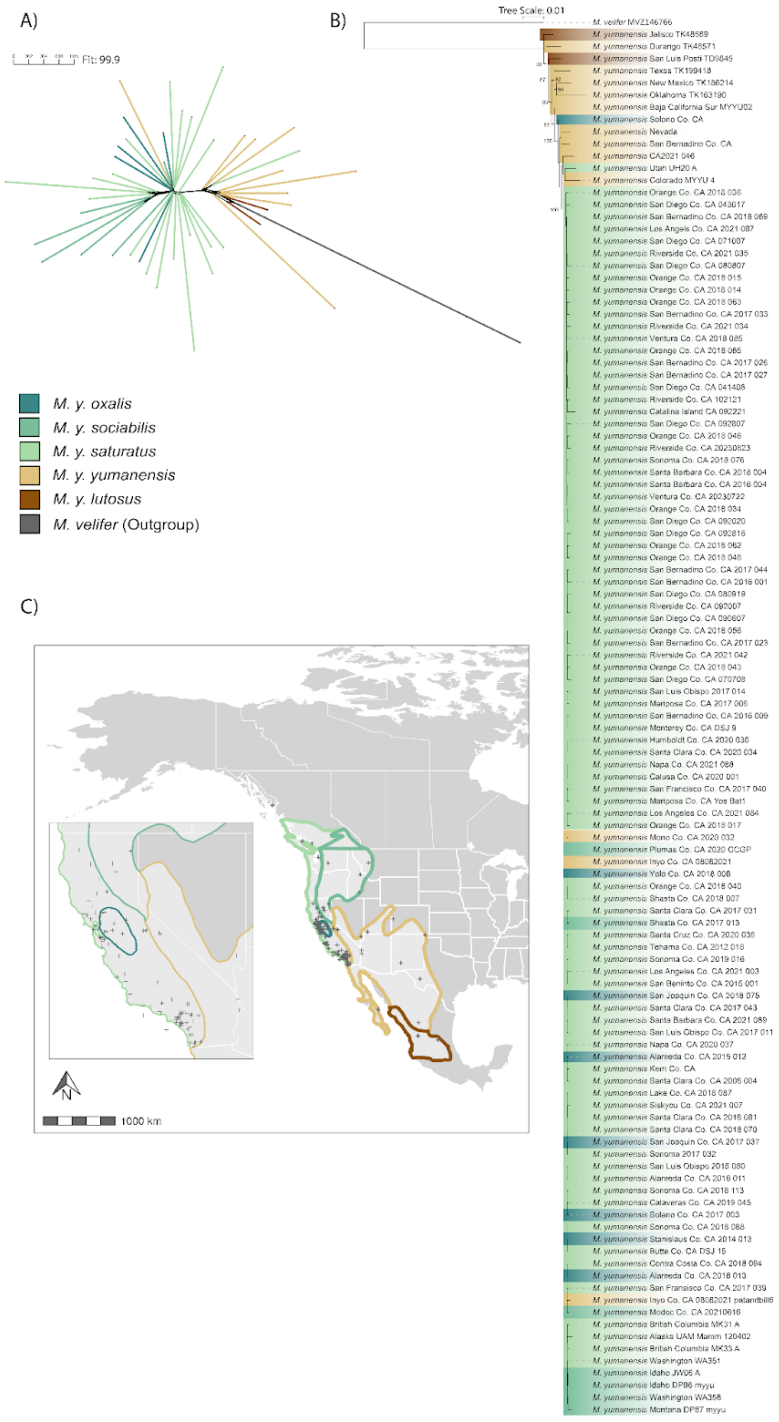

Figure 10. Phylogenetic analysis of MYYU, including A) neighbor net phylogenetic network of 43 MYYU samples, and B) maximum likelihood estimate tree of 110 MYYU samples. Branches colored by putative subspecies assignment and outgroup. Map of geographic distribution of individual subspecies in C), adapted from Braun et al. 2015.

TABLES

Table 3-1. Summary table of AMOVA outputs by cluster assignment for different numbers of clusters (K) between K = 2 and K = 5. Table details the percentage of variation explained by clusters and within samples, along with the amount of genetic differentiation between groups (ΦST), degrees of freedom (df), and p-values.

K	% Explained by Clusters	% Explained within Samples	ΦST	df	p-value
2	3.15	96.9	0.031	1	0.001
3	4.35	95.7	0.043	2	0.01
4	4.45	95.5	0.045	3	0.01
5	3.47	96.5	0.035	4	0.01

Table 3-2. Table of PSMC outputs for the mid-Holocene (ca. 6 kya), the Last Glacial Maximum (LGM, ca. 20 kya) and the Last Interglacial Period (LGI, 110 - 130 kya). Tk parameter is the scaled time parameter and Ne is the scaled λk from PSMC. Trend determined by taking the difference in effective population size between the LGI and the LGM.

Sampling Location	T_k mid-Holocene	N_e mid-Holocene	T_k LGM	N_e LGM	T_k LGI	N_e LGI	LGI - LGM Trend
Alaska, US	5190	8939020	21272	34141	124259	136569	Decreasing
Montana, US	5699	99391470000	19821	102771	127217	151723	Decreasing
N. California, US	5938	656191	20477	164150	127015	123452	Increasing
N. California, US	6535	8266275	18698	280111	125225	138256	Increasing
N. California, US	6445	4539390	18425	285093	122962	135778	Increasing
Catalina Is., California, US	6314	50580	21704	272756	133195	161707	Increasing
Nevada, US	7979	2417635	17496	411867	128175	370081	Increasing
New Mexico, US	4289	9903876	19666	400375	127765	668772	Decreasing
Oklahoma, US	6352	16946634	18325	174076	131487	488388	Decreasing
Baja California Sur, Mexico	5799	713236	21169	66022	136846	408578	Decreasing
Jalisco, Mexico	5053	38033338	23065	13312462	130313	653695	Increasing

APPENDIX 3-I: SUPPLEMENTAL FIGURES

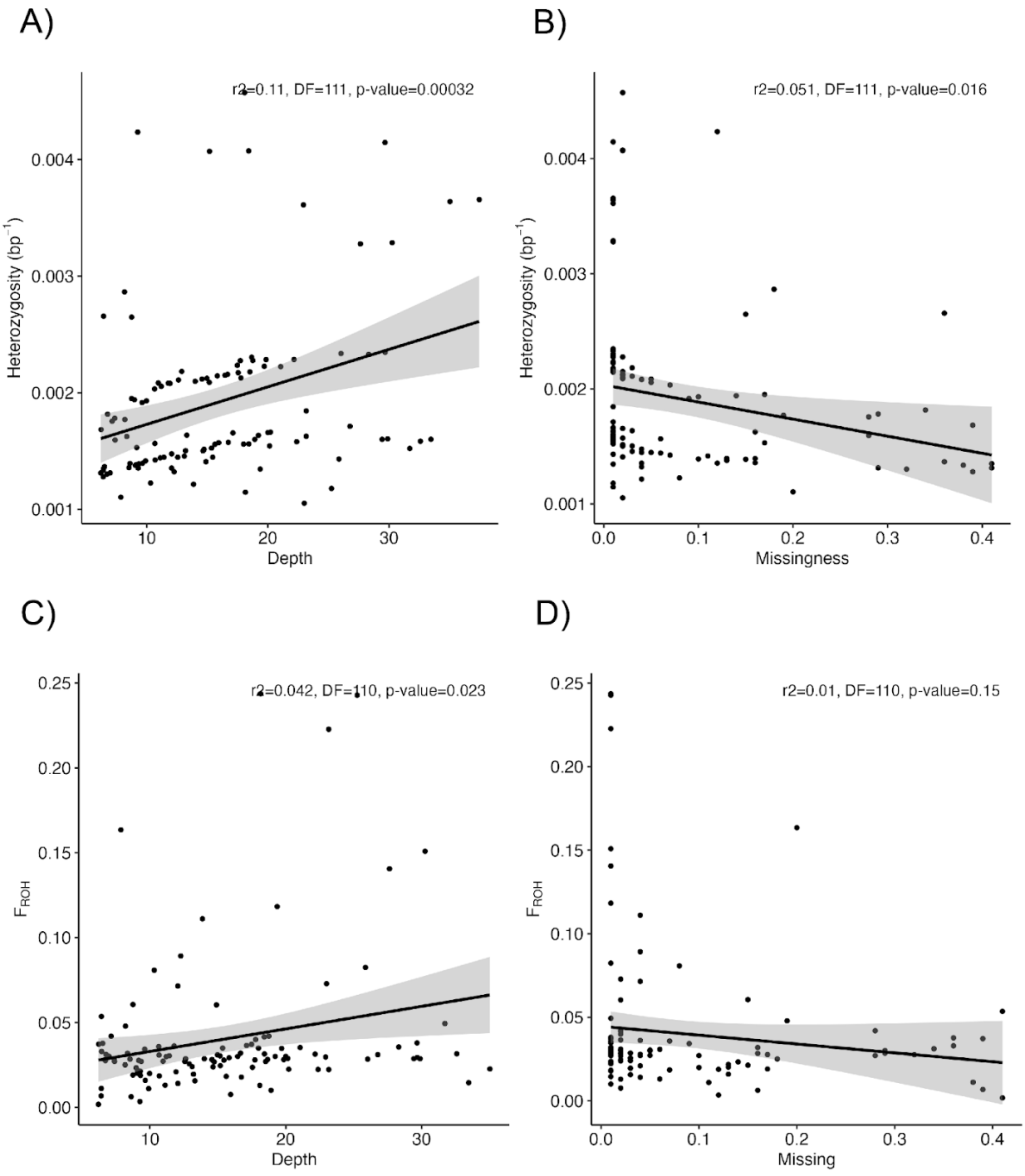

Figure S3-1. Regression between A) per base pair heterozygosity and depth, B) per base pair heterozygosity and missingness, C) FROH and depth, and D) FROH and missingness for 115 MYYU samples.

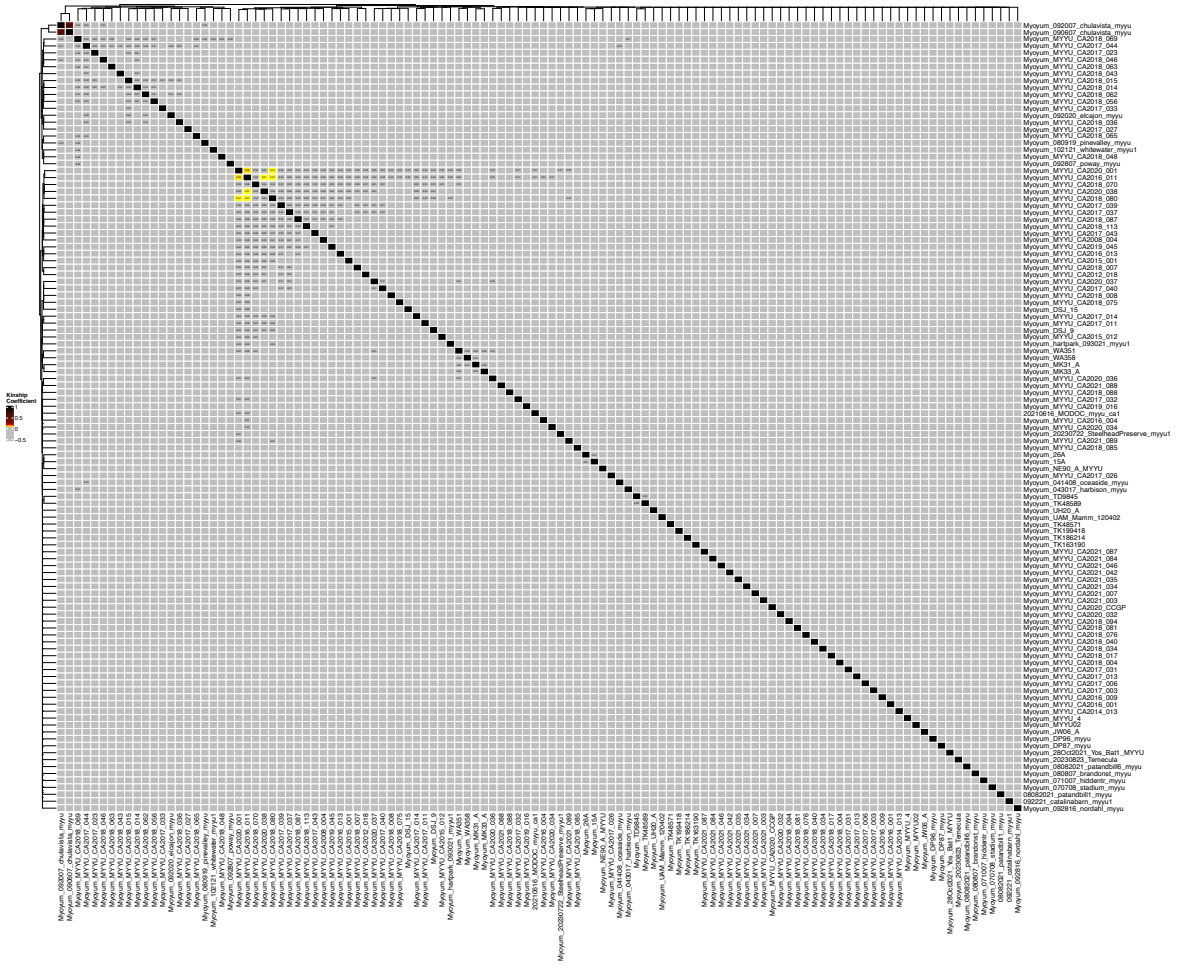

Figure S3-2. Pairwise identity by descent (IBD) matrix of MYYU samples. Samples in yellow indicate individuals approaching second-degree relationships and samples in red indicate samples with first-degree relationships. Self comparisons in dark gray.

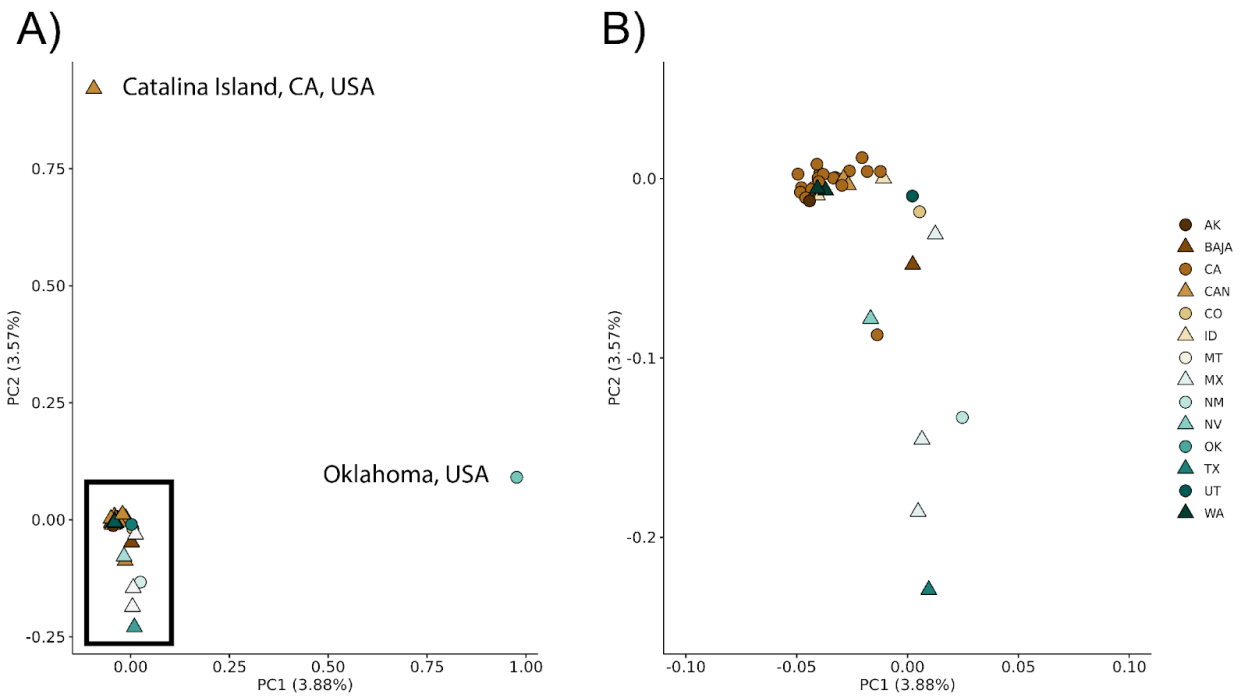

Figure S3-3. Principal Component Analysis of A) 38 downsampled MYYU samples, black box indicating samples from panel B, and B) PCA of 36 MYYU samples excluding two outlier samples. Shape and color indicate sampling location.

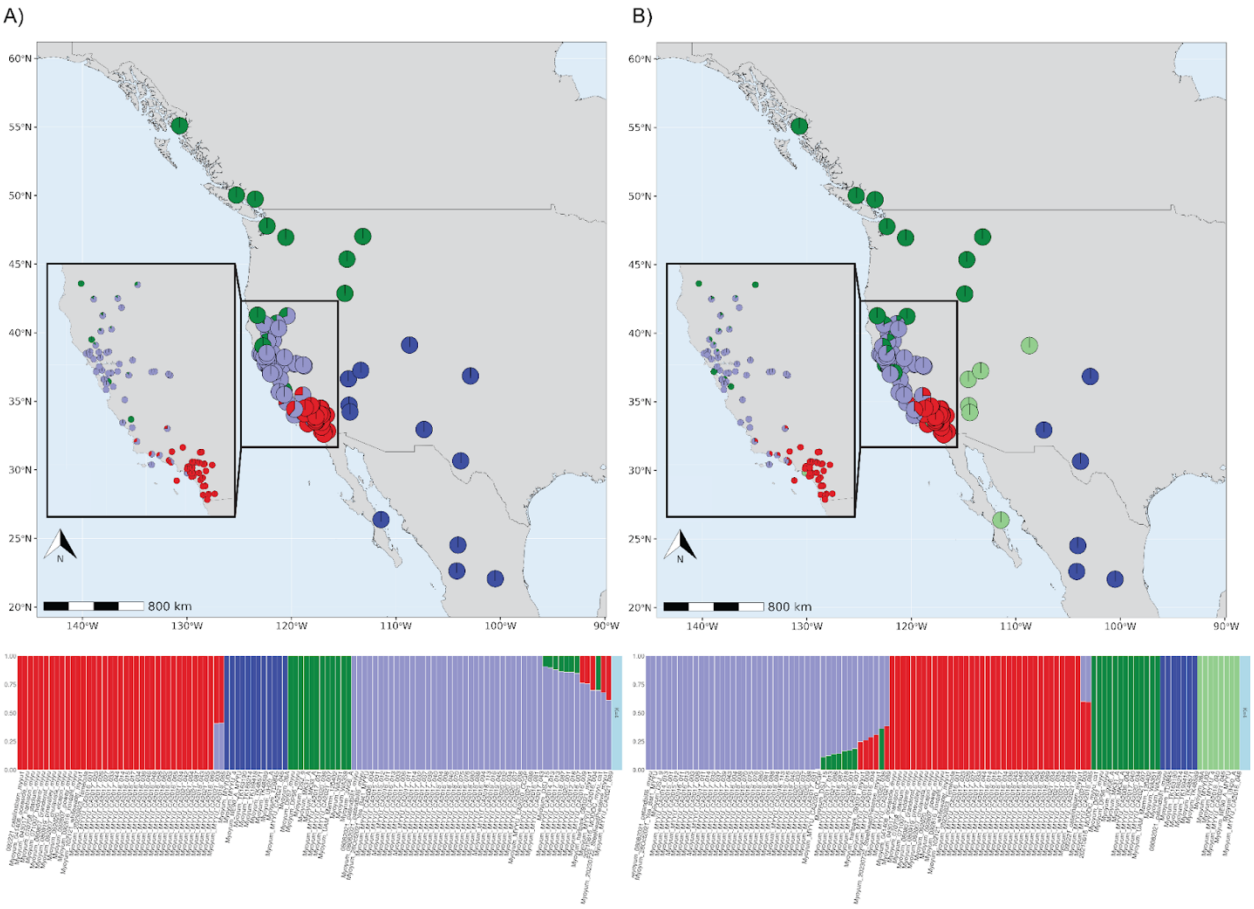

Figure S3-4. ADMIXTURE analysis outputs of sample-level ancestry proportions for A) K=4, and B) K=5 populations.

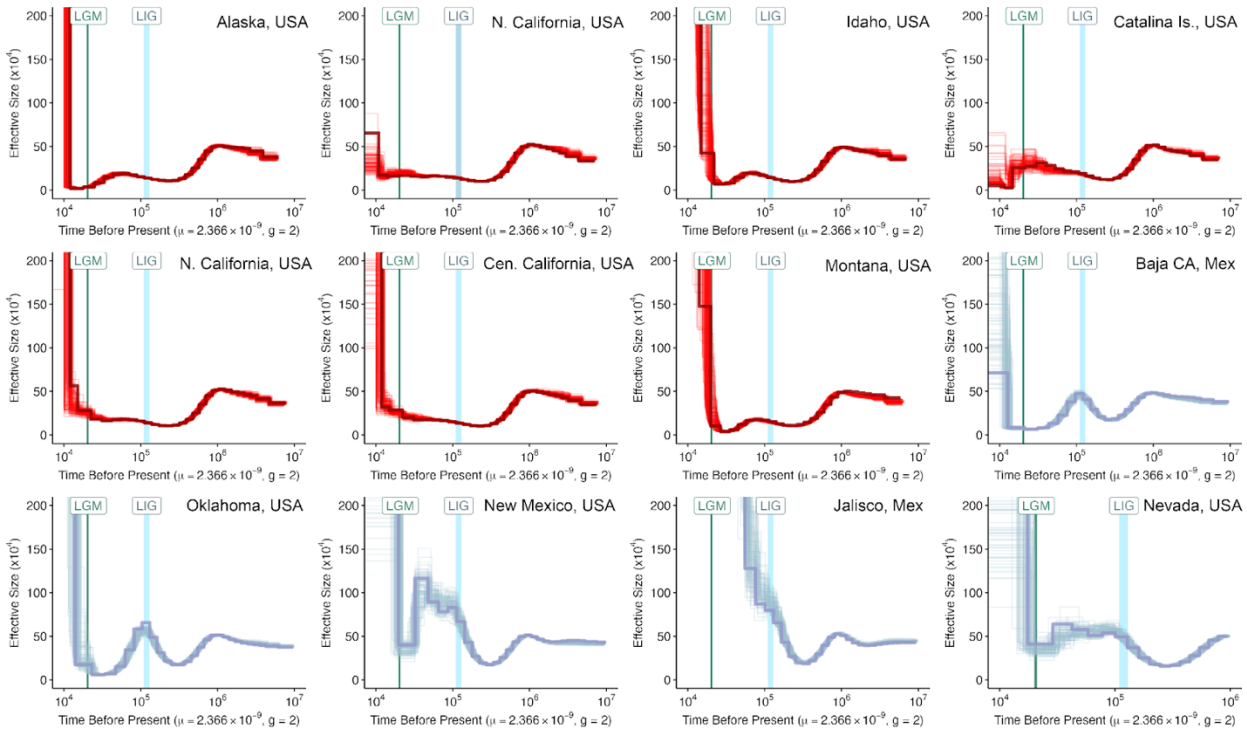

Figure S3-5. Estimated fluctuations in Ne using PSMC between 10 kya and 10 mya for twelve MYYU. Red samples indicate MYYU sampled from the western US and Canada, while the blue indicates MYYU sampled from the Southwest US and Mexico. Thick line is the full run of PSMC while the thinner lines are 100 bootstrap replicates. LGM = Last Glacial Maximum, LIG = Last Interglacial Period.

APPENDIX 3-II: SUPPLEMENTAL TABLES

Table S3-1. Table of MYYU samples detailing species, putative subspecies, forearm length (mm) sex, age, sampling locality and collector.

Sample Name	SPP	SubSPP	FA	Sex	Age	Year	Locality	Contributor
041408_oceaside_myyu	MYYU	saturatus	NA	NR	NR	2008	Escondido, San Diego County, CA	SDMNH
043017_harbison_myyu	MYYU	saturatus	NA	M	A	2017	El Cajon, San Diego County, CA	SDMNH
070708_stadium_myyu	MYYU	saturatus	NA	NR	NR	2008	Escondido, San Diego County, CA	SDMNH
071007_hiddentr_myyu	MYYU	saturatus	NA	M	J	2007	Escondido, San Diego County, CA	SDMNH
080807_brandonst_myyu	MYYU	saturatus	NA	NR	NR	2007	Spring Valley, San Diego County, CA	SDMNH
08082021_patandbill1_myyu	MYYU	yumanensis	NA	F	J(?)	2021	Lodge storage near Bishop, Inyo County, CA	P. Brown, W. Rainey
08082021_patandbill6_myyu	MYYU	yumanensis	NA	F	A	2021	Lodge storage near Bishop, Inyo County, CA	P. Brown, W. Rainey
080919_pinevalley_myyu	MYYU	saturatus	NA	NR	A	2019	Pine Valley, San Diego County, CA	SDMNH
090607_chulavista_myyu	MYYU	saturatus	NA	F	NR	2007	Carlsbad, San Diego County, CA	SDMNH
092007_chulavista_myyu	MYYU	saturatus	NA	M	NR	2007	Hemet, Riverside County, CA	SDMNH
092020_elcajon_myyu	MYYU	saturatus	NA	M	A	2020	El Cajon, San Diego County, CA	SDMNH
092221_catalinabarn_myyu1	MYYU	saturatus	33.30	M	A	2021	Quail Valley, Catalina Island, CA	J. Curti, E. Hamblin
092807_poway_myyu	MYYU	saturatus	NA	F	NR	2007	Poway, San Diego County, CA	SDMNH
092816_nordahl_myyu	MYYU	saturatus	NA	NR	NR	2016	San Marcos, San Diego County, CA	SDMNH
102121_whitewater_myyu1	MYYU	saturatus	35.00	M	A	2021	Whitewater Preserve, 9160 Whitewater Canyon Rd, Whitewater, CA 92282	J. Curti, M. Bourne, L. Basulto
15A	MYYU	oxalis	NA	M	A	2021	Suisun Creek, Solono County, CA	D. Johnston
20210616_MODOC_myyu_ca1	MYYU	sociabilis	NA	F	A	2021	Bridge by West Valley Reservoir, Modoc County, CA	J. Curti, N. Hamilton, C. Ramirez
20230722_SteelheadPreserve_myyu1	MYYU	saturatus	NA	M	A	2023	Steelhead Preserve, Ojai, CA, USA	J. Curti
20230823_Temecula	MYYU	saturatus	NA	F	A	2023	Temecula Creek under I15, Temecula, CA, USA	J. Curti, J. Carpenter, L. Harris
26A	MYYU	yumanensis	NA	F	A	2021	Bat Cave Wash, Topock, San Bernadino County, CA	D. Johnston
28Oct2021_Yos_Bat1_MYYU	MYYU	saturatus	NA	NR	NR	2021	Cascade Picnic Area, Yosemite National Park, Mariposa County, CA, USA	A. Waag
DP87_myyu	MYYU	sociabilis	NA	F	A	2002	Ovando, MT	Dixie Pierson, Maarten Vonhof
DP96_myyu	MYYU	sociabilis	NA	F	NR	2003	Malad Gorge, ID	Dixie Pierson, Maarten Vonhof
DSJ_15	MYYU	saturatus	NA	F	A	2001	Butte Creek, Butte County, CA	Von Hoff
DSJ_9	MYYU	saturatus	NA	F	J	2001	Bradley Bridge, Monterey, CA	Von Hoff
hartpark_093021_myyu1	MYYU	saturatus	34.90	M	A	2021	Hart Park, Bakersfield, CA	J. Curti, C. Ramirez, E. Noel
JW06_A	MYYU	sociabilis	NA	F	J	2001	Corn Creek, ID	Jo Wenger, Maarten Vonhof
MK31_A	MYYU	saturatus	NA	NR	J	2002	Clowhom Lake, British Columbia	Mandy Kellner, Maarten Vonhof
MK33_A	MYYU	saturatus	NA	F	A	2002	Campbell River, British Columbia	Mandy Kellner, Maarten Vonhof
MVZ146766a_MYVE	MYVE	NA	44.00	M	NR	1974	Minas Armalilla, 6 mi NW Alamos, State of Sonora, MX	Berkeley MVZ, Maarten Vonhof
MYYU_4	MYYU	yumanensis	35.60	NR	NR	2022	Monument Ponds, CO	Dan Neubaum, Maarten Vonhof
MYYU_CA2008_004	MYYU	saturatus	NA	F	A	2008	Alviso Cannery, Santa Clara County, CA	D. Johnson
MYYU_CA2012_018	MYYU	saturatus	NA	F	A	2012	Dye Creek Preserve, Tehama County, CA	D. Johnson
MYYU_CA2014_013	MYYU	oxalis	NA	NR	NR	2014	Hughson (95326), Stanislaus County, CA	CDPH
MYYU_CA2015_001	MYYU	saturatus	NA	A	F	2015	San Benito County, CA	CDPH
MYYU_CA2015_012	MYYU	oxalis	NA	NR	NR	2015	Livermore, Alameda County, CA	CDPH
MYYU_CA2016_001	MYYU	saturatus	NA	NR	NR	2016	Apple Valley, San Bernadino County, CA	CDPH
MYYU_CA2016_004	MYYU	saturatus	NA	NR	NR	2016	Santa Maria, Santa Barbara County, CA	CDPH
MYYU_CA2016_009	MYYU	saturatus	NA	NR	NR	2016	Rialto, San Bernadino County, CA	CDPH
MYYU_CA2016_011	MYYU	saturatus	NA	NR	NR	2016	Oakland, Alameda County, CA	CDPH
MYYU_CA2016_013	MYYU	oxalis	NA	NR	NR	2016	Pleasanton, Alameda County, CA	CDPH
MYYU_CA2017_003	MYYU	oxalis	NA	F	A	2017	Putah Creek Oxbow, Solano Count, CA	Harris
MYYU_CA2017_006	MYYU	saturatus	NA	NR	NR	2017	Mariposa County, CA	Harris
MYYU_CA2017_011	MYYU	saturatus	NA	NR	NR	2017	San Luis Obispo County, CA	Harris
MYYU_CA2017_013	MYYU	sociabilis	NA	NR	NR	2017	Big Spring, Lassen National Park, Shasta County, CA	Angerer
MYYU_CA2017_014	MYYU	saturatus	NA	NR	NR	2017	San Luis Obispo County, CA	Unknown
MYYU_CA2017_023	MYYU	saturatus	NA	NR	NR	2017	Chino Hills, San Bernadino County, CA	CDPH
MYYU_CA2017_026	MYYU	saturatus	NA	NR	NR	2017	Ontario, San Bernadino County, CA	CDPH
MYYU_CA2017_027	MYYU	saturatus	NA	NR	NR	2017	Yucaipa, San Bernadino County, CA	CDPH
MYYU_CA2017_031	MYYU	saturatus	NA	NR	NR	2017	San Jose, Santa Clara County, CA	CDPH
MYYU_CA2017_032	MYYU	saturatus	NA	NR	NR	2017	Windsor, Sonoma County, CA	CDPH
MYYU_CA2017_033	MYYU	saturatus	NA	NR	NR	2017	Rancho Cucamonga, San Bernadino County, CA	CDPH
MYYU_CA2017_037	MYYU	oxalis	NA	NR	NR	2017	Stockton (95209), San Joaquin County, CA	CDPH
MYYU_CA2017_039	MYYU	saturatus	NA	NR	NR	2017	San Francisco, San Francisco County, CA	CDPH
MYYU_CA2017_040	MYYU	saturatus	NA	NR	NR	2017	San Francisco, San Francisco County, CA	CDPH
MYYU_CA2017_043	MYYU	saturatus	NA	NR	NR	2017	Palo Alto, Santa Clara County, CA	CDPH
MYYU_CA2017_044	MYYU	saturatus	NA	NR	NR	2017	Upland, San Bernadino County, CA	CDPH
MYYU_CA2018_004	MYYU	saturatus	NA	F	A	2018	Cañada Del Puerto, Santa Cruz Island, Santa Barbara County, CA	P. Brown
MYYU_CA2018_007	MYYU	saturatus	NA	NR	NR	2018	Whiskeytown NRA, Shasta County, CA	Chung-McCoubry
MYYU_CA2018_008	MYYU	oxalis	NA	NR	NR	2018	Yolo County, CA	Unknown
MYYU_CA2018_014	MYYU	saturatus	NA	F	J	2018	Yorba Linda, Orange County, CA	CDPH
MYYU_CA2018_015	MYYU	saturatus	NA	M	J	2018	Rancho Santa Margarita, Orange County, CA	CDPH
MYYU_CA2018_017	MYYU	saturatus	NA	M	J	2018	Brea, Orange County, CA	CDPH
MYYU_CA2018_034	MYYU	saturatus	NA	F	A	2018	Costa Mesa, Orange County, CA	CDPH
MYYU_CA2018_036	MYYU	saturatus	NA	M	A	2018	Orange, Orange County, CA	CDPH
MYYU_CA2018_040	MYYU	saturatus	NA	F	A	2018	Costa Mesa, Orange County, CA	CDPH
MYYU_CA2018_043	MYYU	saturatus	NA	F	A	2018	Rancho Mission Viejo, Orange County, CA	CDPH
MYYU_CA2018_046	MYYU	saturatus	NA	F	A	2018	Trabuco Canyon, Orange County, CA	CDPH
MYYU_CA2018_048	MYYU	saturatus	NA	F	A	2018	Irvine, Orange County, CA	CDPH
MYYU_CA2018_056	MYYU	saturatus	NA	F	A	2018	Laguna Niguel, Orange County, CA	CDPH
MYYU_CA2018_062	MYYU	saturatus	NA	F	A	2018	Foothill Rance, Orange County, CA	CDPH
MYYU_CA2018_063	MYYU	saturatus	NA	F	A	2018	Lake Forest, Orange County, CA	CDPH
MYYU_CA2018_065	MYYU	saturatus	NA	F	A	2018	Anaheim, Orange County, CA	CDPH
MYYU_CA2018_069	MYYU	saturatus	NA	NR	NR	2018	Spring Valley Lake, San Bernadino County, CA	CDPH
MYYU_CA2018_070	MYYU	saturatus	NA	NR	NR	2018	Mountain View, Santa Clara County, CA	CDPH
MYYU_CA2018_075	MYYU	oxalis	NA	NR	NR	2018	Tracy (95304), San Joaquin County, CA	CDPH
MYYU_CA2018_076	MYYU	saturatus	NA	NR	NR	2018	Monte Rio, Sonoma County, CA	CDPH
MYYU_CA2018_080	MYYU	saturatus	NA	NR	NR	2018	Atascadero, San Luis Obispo County, CA	CDPH
MYYU_CA2018_081	MYYU	saturatus	NA	NR	NR	2018	Monte Sereno, Santa Clara County, CA	CDPH
MYYU_CA2018_085	MYYU	saturatus	NA	NR	NR	2018	Thousand Oaks, Ventura, CA	CDPH
MYYU_CA2018_087	MYYU	saturatus	NA	NR	NR	2018	Hidden Valley Lake, Lake County, CA	CDPH
MYYU_CA2018_088	MYYU	saturatus	NA	NR	NR	2018	Monte Rio, Sonoma County, CA	CDPH
MYYU_CA2018_094	MYYU	saturatus	NA	NR	NR	2018	Richmond, Contra Costa County, CA	CDPH
MYYU_CA2018_113	MYYU	saturatus	NA	NR	NR	2018	Sebastapol, Sonoma County, CA	CDPH
MYYU_CA2019_016	MYYU	saturatus	NA	NR	NR	2019	Petaluma, Sonoma County, CA	CDPH
MYYU_CA2019_045	MYYU	saturatus	NA	NR	NR	2019	San Andreas (95249), Calaveras County, CA	CDPH
MYYU_CA2020_001	MYYU	saturatus	NA	Unk	Unk	2020	Colusa County, CA	S. Osborn
MYYU_CA2020_032	MYYU	yumanensis	NA	Unk	Unk	2020	Hot Creek Hatchery, Mono County, CA	Heide
MYYU_CA2020_034	MYYU	saturatus	NA	NR	NR	2020	Morgan Hill, Santa Clara County, CA	CDPH
MYYU_CA2020_036	MYYU	saturatus	NA	NR	NR	2020	Arcada (95521), Humboldt County, CA	CDPH
MYYU_CA2020_037	MYYU	saturatus	NA	NR	NR	2020	Napa County, CA	CDPH
MYYU_CA2020_038	MYYU	saturatus	NA	NR	NR	2020	Santa Cruz, Santa Cruz County, CA	CDPH
MYYU_CA2020_CCGP	MYYU	sociabilis	NA	M	J	2020	Plumas County, CA	D. Fraser
MYYU_CA2021_003	MYYU	saturatus	NA	M	A	2021	Rock Pools near Century Dam, Malibu Creek State Park, Los Angeles, Los Angeles County, CA	J. Curti, N. Hamilton
MYYU_CA2021_007	MYYU	saturatus	NA	M	A	2021	Along Sawpit Creek near Somes Bar, Siskyou County, CA	J. Curti, N. Hamilton
MYYU_CA2021_034	MYYU	saturatus	NA	M	A	2021	Culvert, Redlands, Riverside County, CA	J. Curti, C. Ramirez
MYYU_CA2021_035	MYYU	saturatus	NA	F	A	2021	E Street Culvert, Riverside County, CA	J. Curti, C. Ramirez
MYYU_CA2021_042	MYYU	saturatus	NA	F	A	2021	Green River Golf Course, Riverside County, CA	J. Curti, C. Ramirez
MYYU_CA2021_046	MYYU	yumanensis	NA	F	A	2021	Riverview Mine, San Bernadino County, CA	J. Curti, C. Ramirez, R. Peat
MYYU_CA2021_084	MYYU	saturatus	NA	NR	NR	2021	Lancaster, Los Angeles County, CA	CDPH
MYYU_CA2021_087	MYYU	saturatus	NA	NR	NR	2021	Castaic (91384), Los Angeles County, CA	CDPH
MYYU_CA2021_088	MYYU	saturatus	NA	NR	NR	2021	Saint Helena, Napa County, CA	CDPH
MYYU_CA2021_089	MYYU	saturatus	NA	NR	NR	2021	Santa Barbara County, CA	CDPH
MYYU02	MYYU	yumanensis	32.02	NR	NR	2008	San Basilio, Baja California Sur, Mexico	Winnifred Frick
NE90_A_MYYU	MYYU	yumanensis	NR	M	A	2002	Gypsum mine near I-15 and Moapa Exit, NV	Mike O'Farrell, Jason Williams, Maarten Vonhof
TD9845	MYYU	lotusus	NR	F	NR	2016	Las Charcas, San Nicolás Tolentino Municipality, San Luis Potosí State, MX	Instituto Politécnico Nacional/Celia López-González
TK163190	MYYU	yumanensis	34.00	M	NR	2009	BLACK MESA STATE PARK, Cimarron County, OK	Texas Tech Museum
TK186214	MYYU	yumanensis	34.00	F	NR	2014	TRUJILLO CREEK RD AND NM HWY 187, Sierra County, NM	Texas Tech Museum
TK199418	MYYU	yumanensis	34.00	M	NR	2016	TX HWY 17, N FORT DAVIS, JEFF DAVIS CO	Texas Tech Museum
TK48571	MYYU	yumanensis	32.00	M	NR	1995	20 KM N LAS HERRERAS, State of DURANGO, MX	Texas Tech Museum
TK48589	MYYU	lotusus	35.00	F	NR	1995	30 KM W HUEJUQUILLA DEL ALTO, State of Jalisco, MX	Texas Tech Museum
UAM_Mamm_120402	MYYU	saturatus	36.00	F	J	2014	Hugh Smith Lake at Sockeye Creek, Ketchikan Quad, AK	University of Alaska Museum
UH20_A	MYYU	saturatus	NR	F	A	2002	Silver Reef WH48, Washington County, UT	Kate Grandison, Maarten Vonhof
WA351	MYYU	saturatus	NA	F	A	2021	Snohomish County, WA	Abigail Tobin
WA358	MYYU	sociabilis	NA	F	A	2021	Kittitas County, WA	Abigail Tobin

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Allen, Harrison. (1864). Monograph of the Bats of North America. *Smithsonian Miscellaneous Collections*, *7*, 1–85.

Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Molecular Ecology*, *27*(17), 3452–3465. https://doi.org/10.1111/mec.14806

Barbour, R. W., & Davis, W. H. (1969). *Bats of America*. University Press of Kentucky; WorldCat.

Benson, Seth B. (Seth Bertram). (1947). Description of a subspecies of *Myotis yumanensis* from Baja California, Mexico. *Proceedings of the Biological Society of Washington*, *60*, 45–46.

Boyles, J. G., Cryan, P. M., McCracken, G. F., & Kunz, T. H. (2011). Economic Importance of Bats in Agriculture. *Science*, *332*(6025), 41–42. https://doi.org/10.1126/science.1201366

Braun, J. K., Yang, B., Gonzalez-Perez, S. B., & Mares, M. A. (2015). *Myotis yumanensis* (Chiroptera: Vespertilionidae). *Mammalian Species*, *47*(918), 1–14. https://doi.org/10.1093/mspecies/sev001

Brigham, R. M., Aldridge, H. D. J. N., & Mackey, R. L. (1992). Variation in Habitat Use and Prey Selection by Yuma Bats, *Myotis yumanensis*. *Journal of Mammalogy*, *73*(3), 640–645. https://doi.org/10.2307/1382036

Brown, V. A., Brooke, A., Fordyce, J. A., & McCracken, G. F. (2011). Genetic analysis of populations of the threatened bat *Pteropus mariannus*. *Conservation Genetics*, *12*(4), 933–941. https://doi.org/10.1007/s10592-011-0196-y

Brüniche-Olsen, A., Kellner, K. F., Anderson, C. J., & DeWoody, J. A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics*, *19*(6), 1295–1307. https://doi.org/10.1007/s10592-018-1099-y

Bryant, D., & Huson, D. H. (2023). NeighborNet: Improved algorithms and implementation. *Frontiers in Bioinformatics*, *3*, 1178600. https://doi.org/10.3389/fbinf.2023.1178600

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, *21*(2), 255–265. https://doi.org/10.1093/molbev/msh018

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. https://doi.org/10.1186/1471-2105-10-421

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Chattopadhyay, B., Garg, K. M., Ray, R., & Rheindt, F. E. (2019). Fluctuating fortunes: Genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1911), 20190304. https://doi.org/10.1098/rspb.2019.0304

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, *2*(2), e107. https://doi.org/10.1002/imt2.107

Collevatti, R. G., Vitorino, L. C., Vieira, T. B., Oprea, M., & Telles, M. P. C. (2020). Landscape changes decrease genetic diversity in the Pallas' long-tongued bat. *Perspectives in Ecology and Conservation*, *18*(3), 169–177. https://doi.org/10.1016/j.pecon.2020.06.006

Cortazar, L. A. N. (2020). *Ecological genomics of Myotis bats and their ectoparasites in the Baja California peninsula*. University of Leeds. https://etheses.whiterose.ac.uk/28383/

Curti, J. N., Fraser, D., Escalona, M., Fairbairn, C. W., Sacco, S., Sahasrabudhe, R., Nguyen, O., Seligmann, W., Sudmant, P. H., Toffelmier, E., Vazquez, J. M., Wayne, R., Shaffer, H. B., & Buchalski, M. R. (2024). A genome assembly of the Yuma myotis bat, *Myotis yumanensis*. *Journal of Heredity*, *115*(1), 139–148. https://doi.org/10.1093/jhered/esad053

Dalquest, W. W. (1947). Notes on the Natural History of the Bat, *Myotis yumanensis*, in California, with a Description of a New Race. *American Midland Naturalist*, *38*(1), 224. https://doi.org/10.2307/2421638

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daskalova, G. N., Myers-Smith, I. H., & Godlee, J. L. (2020). Rare and common vertebrates span a wide spectrum of population trends. *Nature Communications*, *11*(1), 4394. https://doi.org/10.1038/s41467-020-17779-0

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dufresnes, C., Dutoit, L., Brelsford, A., Goldstein-Witsenburg, F., Clément, L., López-Baucells, A., Palmeirim, J., Pavlinić, I., Scaravelli, D., Ševčík, M., Christe, P., & Goudet, J. (2023). Inferring genetic structure when there is little: Population genetics versus genomics of the threatened bat Miniopterus schreibersii across Europe. *Scientific Reports*, *13*(1), 1523. https://doi.org/10.1038/s41598-023-27988-4

Dussex, N., Alberti, F., Heino, M. T., Olsen, R.-A., van der Valk, T., Ryman, N., Laikre, L., Ahlgren, H., Askeyev, I. V., Askeyev, O. V., Shaymuratova, D. N., Askeyev, A. O., Döppes, D., Friedrich, R., Lindauer, S., Rosendahl, W., Aspi, J., Hofreiter, M., Lidén, K., … Díez-del-Molino, D. (2020). Moose genomes reveal past glacial demography and the origin of modern lineages. *BMC Genomics*, *21*(1), 854. https://doi.org/10.1186/s12864-020-07208-3

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, *107*(37), 16196–16200. https://doi.org/10.1073/pnas.1006538107

Evelyn, M. J., Stiles, D. A., & Young, R. A. (2004). Conservation of bats in suburban landscapes: Roost selection by *Myotis yumanensis* in a residential area in California. *Biological Conservation*, *115*(3), 463–473. https://doi.org/10.1016/S0006-3207(03)00163-0

Festa, F., Ancillotto, L., Santini, L., Pacifici, M., Rocha, R., Toshkova, N., Amorim, F., Benítez-López, A., Domer, A., Hamidović, D., Kramer-Schadt, S., Mathews, F., Radchuk, V., Rebelo, H., Ruczynski, I., Solem, E., Tsoar, A., Russo, D., & Razgour, O. (2023). Bat responses to climate change: A systematic review. *Biological Reviews*, *98*(1), 19–33. https://doi.org/10.1111/brv.12893

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frick, W. F., Kingston, T., & Flanders, J. (2020). A review of the major threats and challenges to global bat conservation. *Annals of the New York Academy of Sciences*, *1469*(1), 5–25. https://doi.org/10.1111/nyas.14045

Frick, W. F., Pollock, J. F., Hicks, A. C., Langwig, K. E., Reynolds, D. S., Turner, G. G., Butchkoski, C. M., & Kunz, T. H. (2010). An Emerging Disease Causes Regional Population Collapse of a Common North American Bat Species. *Science*, *329*(5992), 679–682. https://doi.org/10.1126/science.1188594

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gallai, N., Salles, J.-M., Settele, J., & Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, *68*(3), 810–821. https://doi.org/10.1016/j.ecolecon.2008.06.014

Gallego-García, N., Caballero, S., & Shaffer, H. B. (2021). Are Genomic Updates of Well-Studied Species Worth the Investment for Conservation? A Case Study of the Critically Endangered Magdalena River Turtle. *Journal of Heredity*, *112*(7), 575–589. https://doi.org/10.1093/jhered/esab063

García-Mudarra, J. L., Ibáñez, C., & Juste, J. (2009). The Straits of Gibraltar: Barrier or bridge to Ibero-Moroccan bat diversity? *Biological Journal of the Linnean Society*, *96*(2), 434–450. https://doi.org/10.1111/j.1095-8312.2008.01128.x

Gaston, K. J. (2010). Valuing Common Species. *Science*, *327*(5962), 154–155. https://doi.org/10.1126/science.1182818

Gaston, K. J., Cox, D. T. C., Canavelli, S. B., García, D., Hughes, B., Maas, B., Martínez, D., Ogada, D., & Inger, R. (2018). Population Abundance and Ecosystem Service Provision: The Case of Birds. *BioScience*, *68*(4), 264–272. https://doi.org/10.1093/biosci/biy005

Gignoux-Wolfsohn, S. A., Pinsky, M. L., Kerwin, K., Herzog, C., Hall, M., Bennett, A. B., Fefferman, N. H., & Maslo, B. (2021). Genomic signatures of selection in bats surviving white-nose syndrome. *Molecular Ecology*, *30*(22), 5643–5657. https://doi.org/10.1111/mec.15813

Greenspoon, L., Krieger, E., Sender, R., Rosenberg, Y., Bar-On, Y. M., Moran, U., Antman, T., Meiri, S., Roll, U., Noor, E., & Milo, R. (2023). The global biomass of wild mammals. *Proceedings of the National Academy of Sciences*, *120*(10), e2204892120. https://doi.org/10.1073/pnas.2204892120

Grinnell, H. W. (1914). Three new races of vespertilionid bats from California. *University of California Publications in Zoology*, *12*, 317–320.

Hamilton, N. M., Morrison, M. L., Harris, L. S., Szewczak, J. M., & Osborn, S. D. (2022). Predicting habitat suitability for Townsend's big-eared bats across California in relation to climate change. *Ecology and Evolution*, *12*(12), e9641. https://doi.org/10.1002/ece3.9641

Hammerson, G. A., Kling, M., Harkness, M., Ormes, M., & Young, B. E. (2017). Strong geographic and temporal patterns in conservation status of North American bats. *Biological Conservation*, *212*, 144–152. https://doi.org/10.1016/j.biocon.2017.05.025

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, *29*(2), 147–160.

Harris, A. H. (1974). Myotis yumanensis in Interior Southwestern North America, with Comments on *Myotis lucifugus*. *Journal of Mammalogy*, *55*(3), 589–607. https://doi.org/10.2307/1379548

Hewitt, G. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–276. https://doi.org/10.1006/bijl.1996.0035

Hofreiter, M., & Stewart, J. (2009). Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. *Current Biology*, *19*(14), R584–R594. https://doi.org/10.1016/j.cub.2009.06.030

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Huson, D. H., & Bryant, D. (2024). The SplitsTree App: Interactive analysis and visualization using phylogenetic trees and networks. *Nature Methods*, *21*(10), 1773–1774. https://doi.org/10.1038/s41592-024-02406-3

James, E., Long, S. H., Say, T., United States War Department & Thwaites, R. G. (1905) James's Account of S. H. Long's expedition, -1820. Cleveland, Ohio: The Arthur H. Clark Company. https://www.loc.gov/item/05026896/

Jebb, D., Huang, Z., Pippel, M., Hughes, G. M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L. S., Skirmuntt, E. C., Katzourakis, A., Burkitt-Gray, L., Ray, D. A., Sullivan, K. A. M., Roscito, J. G., Kirilenko, B. M., Dávalos, L. M., Corthals, A. P., Power, M. L., Jones, G., … Teeling, E. C. (2020). Six reference-quality genomes reveal evolution of bat adaptations. *Nature*, *583*(7817), 578–584. https://doi.org/10.1038/s41586-020-2486-3

Jenkins, T. L. (2024). MAPMIXTURE: An R package and web app for spatial visualisation of admixture and population structure. *Molecular Ecology Resources*, *24*(4), e13943. https://doi.org/10.1111/1755-0998.13943

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). *Poppr*: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kardos, M., & Waples, R. S. (2024). Low-coverage sequencing and Wahlund effect severely bias estimates of inbreeding, heterozygosity and effective population size in North American wolves. *Molecular Ecology*, e17415. https://doi.org/10.1111/mec.17415

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Korstian, J. M., Stevens, R. D., Lee, T. E., Baker, R. J., & Ray, D. A. (2024). Unraveling the *Myotis* morass: Ultraconserved-element analysis reveals introgression, cryptic diversity, and taxonomic trouble. *Journal of Mammalogy*, *105*(2), 259–276. https://doi.org/10.1093/jmammal/gyad119

Kunz, T. H., Braun De Torrez, E., Bauer, D., Lobova, T., & Fleming, T. H. (2011). Ecosystem services provided by bats. *Annals of the New York Academy of Sciences*, *1223*(1), 1–38. https://doi.org/10.1111/j.1749-6632.2011.06004.x

Kyriazis, C. C., Beichman, A. C., Brzeski, K. E., Hoy, S. R., Peterson, R. O., Vucetich, J. A., Vucetich, L. M., Lohmueller, K. E., & Wayne, R. K. (2023). Genomic Underpinnings of Population Persistence in Isle Royale Moose. *Molecular Biology and Evolution*, *40*(2), msad021. https://doi.org/10.1093/molbev/msad021

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, *5*(1), 33–47. https://doi.org/10.1002/evl3.209

Laforge, A., Archaux, F., Coulon, A., Sirami, C., Froidevaux, J., Gouix, N., Ladet, S., Martin, H., Barré, K., Roemer, C., Claireau, F., Kerbiriou, C., & Barbaro, L. (2021). Landscape composition and life-history traits influence bat movement and space use: Analysis of 30 years of published telemetry data. *Global Ecology and Biogeography*, *30*(12), 2442–2454. https://doi.org/10.1111/geb.13397

Lamb, C. T., Willson, R., Richter, C., Owens-Beek, N., Napoleon, J., Muir, B., McNay, R. S., Lavis, E., Hebblewhite, M., Giguere, L., Dokkie, T., Boutin, S., & Ford, A. T. (2022). Indigenous-led conservation: Pathways to recovery for the nearly extirpated KLINSE-ZA mountain caribou. *Ecological Applications*, *32*(5), e2581. https://doi.org/10.1002/eap.2581

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Lindenmayer, D. B., Wood, J. T., McBurney, L., MacGregor, C., Youngentob, K., & Banks, S. C. (2011). How to make a common species rare: A case against conservation complacency. *Biological Conservation*, *144*(5), 1663–1672. https://doi.org/10.1016/j.biocon.2011.02.022

Liu, S., & Hansen, M. M. (2017). PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular Ecology Resources*, *17*(4), 631–641. https://doi.org/10.1111/1755-0998.12606

Mazet, O., Rodríguez, W., & Chikhi, L. (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theoretical Population Biology*, *104*, 46–58. https://doi.org/10.1016/j.tpb.2015.06.003

McCarthy, M., Martínez, A. R., Ferguson, S., Rosing-Asvid, A., Dietz, R., Cahsan, B. D., Schreiber, L., Lorenzen, E., Hansen, R., Stimmelmayr, R., Bryan, A., Quakenbush, L., Lydersen, C., Kovacs, K., & Olsen, M. T. (2024). *Circumpolar population structure, diversity and recent evolutionary history of the bearded seal in relation to past and present icescapes*. https://doi.org/10.22541/au.172114675.57316404/v1

McCartney-Melstad, E., Gidiş, M., & Shaffer, H. B. (2018). Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity*, *121*(2), 112–125. https://doi.org/10.1038/s41437-018-0097-7

Mestre, F., Barbosa, S., Garrido-García, J. A., Pita, R., Mira, A., Alves, P. C., Paupério, J., Searle, J. B., & Beja, P. (2022). Inferring past refugia and range dynamics through the integration of fossil, niche modelling and genomic data. *Journal of Biogeography*, *49*(11), 2064–2076. https://doi.org/10.1111/jbi.14492

Metcalfe, A. N., Fritzinger, C. A., Weller, T. J., Dodrill, M. J., Muehlbauer, J. D., Yackulic, C. B., Holton, P. B., Szydlo, C. M., Durning, L. E., Sankey, J. B., & Kennedy, T. A. (2023). Insectivorous bat foraging tracks the availability of aquatic flies (Diptera). *The Journal of Wildlife Management*, *87*(5), e22414. https://doi.org/10.1002/jwmg.22414

Miller, G. S., & Allen, G. M. (1928). The American bats of the genera Myotis and Pizonyx. *US Government Printing Office*, *144*.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., & Sackton, T. B. (2024). A Fast, Reproducible, High-throughput Variant Calling Workflow for Population Genomics. *Molecular Biology and Evolution*, *41*(1), msad270. https://doi.org/10.1093/molbev/msad270

Morales, A., Villalobos, F., Velazco, P. M., Simmons, N. B., & Piñero, D. (2016). Environmental niche drives genetic and morphometric structure in a widespread bat. *Journal of Biogeography*, *43*(5), 1057–1068. https://doi.org/10.1111/jbi.12666

Morin, P. A., McCarthy, M. L., Fung, C. W., Durban, J. W., Parsons, K. M., Perrin, W. F., Taylor, B. L., Jefferson, T. A., & Archer, F. I. (2024). Revised taxonomy of eastern North Pacific killer whales (*Orcinus orca*): Bigg's and resident ecotypes deserve species status. *Royal Society Open Science*, *11*(3), 231368. https://doi.org/10.1098/rsos.231368

Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, *9*(10), 373–375.

Mulim, H. A., Brito, L. F., Pinto, L. F. B., Ferraz, J. B. S., Grigoletto, L., Silva, M. R., & Pedrosa, V. B. (2022). Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics*, *23*(1), 209. https://doi.org/10.1186/s12864-022-08384-0

Moussy, C., Hosken, D. J., Mathews, F., Smith, G. C., Aegerter, J. N., & Bearhop, S. (2013). Migration and dispersal patterns of bats and their influence on genetic structure. *Mammal Review*, *43*(3), 183–195. https://doi.org/10.1111/j.1365-2907.2012.00218.x

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nagorsen, D. W., Brigham, R. M., & Museum, R. B. C. (1993). *Bats of British Columbia*. UBC Press. https://books.google.com/books?id=J9MpeByPvjwC

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: High-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Novo, I., Ordás, P., Moraga, N., Santiago, E., Quesada, H., & Caballero, A. (2023). Impact of population structure in the estimation of recent historical effective population size by the software GONE. *Genetics Selection Evolution*, *55*(1), 86. https://doi.org/10.1186/s12711-023-00859-2

Ogle, D. H., Doll, J. C., Wheeler, A. P., & Dinno, A. (2023). FSA: simple fisheries stock assessment methods. *R Package Version 0.9*, *4*.

Olson, L. E., Gunderson, A. M., MacDonald, S., & Blejwas, K. M. (2014). First Records of Yuma Myotis (*Myotis yumanensis*) in Alaska. *Northwestern Naturalist*, *95*(3), 228–235. https://doi.org/10.1898/13-29.1

Ortiz, E. M. (2019). *vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis* (Version v2.0). Zenodo. https://doi.org/10.5281/zenodo.2540861

Osborn, P. S., Tauber, A., Winkler, D., & Clifford, D. (2024). *2023 CDFW Spring White-nose Syndrome Surveillance Summary Report*. 1–9.

Piccioli Cappelli, M., Blakey, R. V., Taylor, D., Flanders, J., Badeen, T., Butts, S., Frick, W. F., & Rebelo, H. (2021). Limited refugia and high velocity range-shifts predicted for bat communities in drought-risk areas of the Northern Hemisphere. *Global Ecology and Conservation*, *28*, e01608. https://doi.org/10.1016/j.gecco.2021.e01608

Postel, S., Bawa, K., Kaufman, L., Peterson, C. H., Carpenter, S., Tillman, D., Dayton, P., Alexander, S., Lagerquist, K., Goulder, L., & others. (2012). *Nature's services: Societal dependence on natural ecosystems*. Island Press.

Provan, J., & Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, *23*(10), 564–571. https://doi.org/10.1016/j.tree.2008.06.010

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ray, D. A., Feschotte, C., Pagan, H. J. T., Smith, J. D., Pritham, E. J., Arensburger, P., Atkinson, P. W., & Craig, N. L. (2008). Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Research*, *18*(5), 717–728. https://doi.org/10.1101/gr.071886.107

Robinson, J. A., Bowie, R. C. K., Dudchenko, O., Aiden, E. L., Hendrickson, S. L., Steiner, C. C., Ryder, O. A., Mindell, D. P., & Wall, J. D. (2021). Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology*, *31*(13), 2939-2946.e5. https://doi.org/10.1016/j.cub.2021.04.035

Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., Lohmueller, K. E., & Wayne, R. K. (2016). Genomic Flatlining in the Endangered Island Fox. *Current Biology*, *26*(9), 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062

Robinson, J., Kyriazis, C. C., Yuan, S. C., & Lohmueller, K. E. (2023). Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annual Review of Animal Biosciences*, *11*(1), 93–114. https://doi.org/10.1146/annurev-animal-080522-093311

Rodhouse, T. J., Scott, S. A., Ormsbee, P. C., & Zinck, J. M. (2008). Field Identification of Myotis yumanensis and Myotis lucifugus: A Morphological Evaluation. *Western North American Naturalist*, *68*(4), 437–443. https://doi.org/10.3398/1527-0904-68.4.437

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Ruedi, M., Stadelmann, B., Gager, Y., Douzery, E. J. P., Francis, C. M., Lin, L.-K., Guillén-Servent, A., & Cibois, A. (2013). Molecular phylogenetic reconstructions identify East Asia as the cradle for the evolution of the cosmopolitan genus Myotis (Mammalia, Chiroptera). *Molecular Phylogenetics and Evolution*, *69*(3), 437–449. https://doi.org/10.1016/j.ympev.2013.08.011

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution*, *37*(12), 3642–3653. https://doi.org/10.1093/molbev/msaa169

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x

Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0. 2013–2015*. Seattle, USA.

Smith, T. N., Furnas, B. J., Nelson, M. D., Barton, D. C., & Clucas, B. (2021). Insectivorous bat occupancy is mediated by drought and agricultural land use in a highly modified ecoregion. *Diversity and Distributions*, *27*(7), 1152–1165. https://doi.org/10.1111/ddi.13264

Solari, S., Sotero-Caio, C. G., & Baker, R. J. (2019). Advances in systematics of bats: Towards a consensus on species delimitation and classifications through integrative taxonomy. *Journal of Mammalogy*, *100*(3), 838–851. https://doi.org/10.1093/jmammal/gyy168

Stoffel, M. A., Johnston, S. E., Pilkington, J. G., & Pemberton, J. M. (2021). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nature Communications*, *12*(1), 2972. https://doi.org/10.1038/s41467-021-23222-9

Sunde, J., Yıldırım, Y., Tibblin, P., & Forsman, A. (2020). Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (*Esox lucius*) and a Synthesis of Previous Studies. *Frontiers in Genetics*, *11*, 218. https://doi.org/10.3389/fgene.2020.00218

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131. https://doi.org/10.1186/s13059-018-1520-3

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long Runs of Homozygosity Are Enriched for Deleterious Variation. *The American Journal of Human Genetics*, *93*(1), 90–102. https://doi.org/10.1016/j.ajhg.2013.05.003

Teeling, E. C., Vernes, S. C., Dávalos, L. M., Ray, D. A., Gilbert, M. T. P., Myers, E., & Bat1K Consortium. (2018). Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annual Review of Animal Biosciences*, *6*(1), 23–46. https://doi.org/10.1146/annurev-animal-022516-022811

Turbek, S. P., Funk, W. C., & Ruegg, K. C. (2023). Where to draw the line? Expanding the delineation of conservation units to highly mobile taxa. *Journal of Heredity*, *114*(4), 300–311. https://doi.org/10.1093/jhered/esad011

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the Cloud*. O'Reilly Media, Inc.

Vonhof, M. J., Russell, A. L., & Miller-Butterworth, C. M. (2015). Range-Wide Genetic Analysis of Little Brown Bat (*Myotis lucifugus*) Populations: Estimating the Risk of Spread of White-Nose Syndrome. *PLOS ONE*, *10*(7), e0128713. https://doi.org/10.1371/journal.pone.0128713

Walker, F. M., Williamson, C. H. D., Sanchez, D. E., Sobek, C. J., & Chambers, C. L. (2016). Species From Feces: Order-Wide Identification of Chiroptera From Guano and Other Non-Invasive Genetic Samples. *PLOS ONE*, *11*(9), e0162342. https://doi.org/10.1371/journal.pone.0162342

Waples, R. S., Antao, T., & Luikart, G. (2014). Effects of Overlapping Generations on Linkage Disequilibrium Estimates of Effective Population Size. *Genetics*, *197*(2), 769–780. https://doi.org/10.1534/genetics.114.164822

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of Mutational Load at the Edges of a Species Range. *Molecular Biology and Evolution*, *35*(4), 781–791. https://doi.org/10.1093/molbev/msy003

WWF. (2024). *Living Planet Report 2024 – A System in Peril*. World Wildlife Fund. Gland, Switzerland.

Yang, W., Feiner, N., Salvi, D., Laakkonen, H., Jablonski, D., Pinho, C., Carretero, M. A., Sacchi, R., Zuffi, M. A. L., Scali, S., Plavos, K., Pafilis, P., Poulakakis, N., Lymberakis, P., Jandzik, D., Schulte, U., Aubret, F., Badiane, A., Perez i De Lanuza, G., … Uller, T. (2022). Population Genomics of Wall Lizards Reflects the Dynamic History of the Mediterranean Basin. *Molecular Biology and Evolution*, *39*(1), msab311. https://doi.org/10.1093/molbev/msab311

You, Y., Sun, K., Xu, L., Wang, L., Jiang, T., Liu, S., Lu, G., Berquist, S. W., & Feng, J. (2010). Pleistocene glacial cycle effects on the phylogeography of the Chinese endemic bat species, *Myotis davi*dii. *BMC Evolutionary Biology*, *10*(1), 208. https://doi.org/10.1186/1471-2148-10-208

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

CHAPTER 4: Using unstructured crowd-sourced data to evaluate urban tolerance of terrestrial native animal species with a California mega-city

PLOS ONE

RESEARCH ARTICLE

Using unstructured crowd-sourced data to evaluate urban tolerance of terrestrial native animal species within a California Mega-City

Joseph N. Curti[1]*, Michelle Barton[2], Rhay G. Flores[1], Maren Lechner[1], Alison Lipman[1], Graham A. Montgomery[1], Albert Y. Park[1], Kirstin Rochel[2], Morgan W. Tingley[1]*

1 Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, United States of America, **2** LA Sanitation and Environment, Los Angeles City, CA, United States of America

* jcurti3@g.ucla.edu (JNC); mtingley@g.ucla.edu (MWT)

OPEN ACCESS

Citation: Curti JN, Barton M, Flores RG, Lechner M, Lipman A, Montgomery GA, et al. (2024) Using unstructured crowd-sourced data to evaluate urban tolerance of terrestrial native animal species within a California Mega-City. PLoS ONE 19(5): e0295476. https://doi.org/10.1371/journal.pone.0295476

Editor: Dárius Pukenis Tubelis, UFERSA: Universidade Federal Rural do Semi-Arido, BRAZIL

Received: November 28, 2023

Accepted: March 18, 2024

Published: May 29, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0295476

Data Availability Statement: All relevant data and code are made available via a Zenodo repository (https://zenodo.org/records/10807319).

Abstract

In response to biodiversity loss and biotic community homogenization in urbanized landscapes, there are increasing efforts to conserve and increase biodiversity within urban areas. Accordingly, around the world, previously extirpated species are (re)colonizing and otherwise infiltrating urban landscapes, while other species are disappearing from these landscapes. Tracking the occurrence of traditionally urban intolerant species and loss of traditionally urban tolerant species should be a management goal of urban areas, but we generally lack tools to study this phenomenon. To address this gap, we first used species' occurrences from iNaturalist, a large collaborative dataset of species observations, to calculate an urban association index (UAI) for 967 native animal species that occur in the city of Los Angeles. On average, the occurrence of native species was negatively associated with our composite measure of urban intensity, with the exception of snails and slugs, which instead occur more frequently in areas of increased urban intensity. Next, we assessed 8,348 0.25 x 0.25 mile grids across the City of Los Angeles to determine the average grid-level UAI scores (i.e., a summary of the UAIs present in a grid cell, which we term Community Urban Tolerance Index or CUTI). We found that areas of higher urban intensity host more urban tolerant species, but also that taxonomic groups differ in their aggregate tolerance of urban areas, and that spatial patterns of tolerance vary between groups. The framework established here has been designed to be iteratively reevaluated by city managers of Los Angeles in order to track the progress of initiatives to preserve and encourage urban biodiversity, but can be rescaled to sample different regions within the city or different cities altogether to provide a valuable tool for city managers globally.

Introduction

The Earth is experiencing an extinction crisis, with modern species extinction rates, based on vertebrate taxa, estimated to exceed background rates of extinction by at least an order of magnitude [1,2]. In this contemporary era of species loss, there are a multitude of factors driving global declines including habitat loss, invasive species, disease, direct exploitation, pollution,

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

and human-caused climate change [3–7]. Many of the effects of these extinction drivers are increased due to synergistic interactions [8]; particularly, urbanization is well known to compound all of these drivers of extinction [8–10]. Globally, urban cover is predicted to increase by 2.5% between 2000 and 2030, such that urbanization will continue to increase as a driver of biodiversity loss [11]. Increases in urban cover are predicted to grow especially fast within global biodiversity hotspots, potentially by 200% between 2000 and 2030 [11], which could further exacerbate rates of species decline. With this predicted increase in urban areas, city managers and conservation biologists can work collaboratively to make cities more hospitable to native biodiversity in order to help avert increasing levels of extinctions.

The impacts of urbanization on biodiversity are well documented, especially in birds [12], arthropods [13,14], and plants [15]. Research on urban biodiversity has traditionally focused on quantifying changes in species richness along the urban-rural gradient (e.g., [16]). For example, studies demonstrate that native woodland bird species tend to be replaced by urban-adapted species in more urbanized habitats nearer to the city center [17,18], and that this pattern appears to be stronger for migratory species compared to residents [19]. Patterns of bird diversity are also tied to factors such as vegetation composition and structure. For example, canopy cover and native plant species diversity correlate with increases in native bird species richness in urban woodlands [20–22]; whereas, increases in lawn cover are related to increases in non-native and synanthropic species richness [23]. In other taxonomic groups, climatic variation has been shown to modulate the strength of the relationship between urbanization and species diversity. For example, in mammals, larger-bodied animals have more negative relationships with urbanization than smaller-bodied animals, but this relationship is intensified in cities that are characterized by warmer climates [24]. Furthermore, the level of taxonomic organization studied can impact the relationship between biodiversity and levels of urbanization. For example, a metanalysis of arthropod community responses to urbanization found an overall negative effect of increased urbanization on arthropod abundance and diversity, but the magnitude of this effect was much greater for specific orders of arthropods, with an increased negative effect in Coleoptera and Lepidoptera and a neutral response (i.e., mean effect size crossing zero) in Araneae [14]. These studies both demonstrate that the mechanisms underlying urban biodiversity are complex, and highlight the need for studies with wide geographic and taxonomic breadth to help us generalize patterns in ways that help cities improve their biodiversity [13,15,25,26].

Considering the projected increase in urban land cover within the next decade, the future of urban biodiversity will ultimately rely on the ability of global cities to attract and maintain populations of species that are largely considered urban intolerant. Well-planned cities can preserve and restore the habitat requirements of native species by facilitating heterogeneous landscapes, migratory stopover sites, and increased gene flow [27], among other initiatives. Likely as a result of urban planning efforts, some species' ranges have increased in urban areas over the past century [28–31]. Several studies have examined what factors increase and maintain urban biodiversity, for example, by evaluating the minimum number of native trees in urban residential yards needed to maintain diverse bird communities [32] or quantifying native species gained in planted rooftop gardens [33]. While these projects can help inform policy geared towards supporting and enhancing urban biodiversity, city managers still lack a comprehensive tool that can track spatio-temporal changes in urban biodiversity at the community level. As more native species are either threatened with extirpation or expand their ranges into urban environments, creating a tool that can track changes in urban diversity and community composition is more important than ever before.

In order to monitor diversity patterns and quantify the effects of varying levels of urbanization on different groups of native species, large amounts of data are needed across multiple

taxonomic groups and across broad areas of the urban environment. To correct for researcher biases that lead to datasets with limited geographic scope and taxonomic coverage, many studies have turned to large crowd-sourced datasets [15,34,35]. Such datasets are often referred to as 'unstructured' in that there is no required protocol for data collection, resulting in data that vary widely in their quality, organization, and information content [36]. One such platform, iNaturalist, has over 74 million observations for over 342,000 different species globally, 58% of which come from developed (i.e. urbanized) areas [37]. The size of the iNaturalist dataset gives it great potential for tracking and managing urban biodiversity. For example, Callaghan et al. [38] used community science data from metropolitan regions around Boston to quantify species- and community-level biodiversity responses to multiple urban gradients. These data were at a small-enough spatial unit to influence local policymaking. Large-scale public participatory datasets make urban biodiversity assessments at large spatial scales possible, even in cities, which tend to contain private lands that are largely excluded from structured biodiversity surveys [39].

Despite the abundance of data points from programs like iNaturalist, there are challenges associated with using these unstructured datasets to measure and manage urban biodiversity. For example, opportunistic sampling may lead to biases in data, as sampling effort is not equal across space, time, and taxonomic groups [40,41], potentially causing differences in user methodology to be misinterpreted as temporal or spatial changes in populations [42]. Analytical methods to best mitigate these inherent biases in unstructured data continue to be developed, including using both models and data processing to better account for unequal observations across space and time [43–45]. Particularly, these new methods use higher order taxa as indicators of survey effort which can inform negative occurrences and thus convert presence-only data to robust presence-absence format.

Here, we describe and implement an approach to spatially and temporally characterize urban tolerance of native species within the city of Los Angeles, California, USA using unstructured species occurrence data from iNaturalist. This approach was initially conceived to support the LA Biodiversity Index Baseline Report published by the Los Angeles Department of Sanitation and Environment [46] through the creation of an evaluative metric (Metric 1.2b; [47,48] that represents and monitors "Native Species Presence in Urban Areas." We refer to this index as a "Community Urban Tolerance Index" (CUTI), as it broadly aims to track how well native species that are often urban intolerant occur within Los Angeles by rating spatial units on the average urban association index or "UAI" (based on levels of urban tolerance) of their species assemblages. To assess this metric, we used iNaturalist data to estimate a species-level UAI for 967 species across six broad taxonomic focal groups that occur in Southern California. We then applied these indices to spatiotemporally thinned species occurrence data in order to calculate the CUTI for a spatial grid covering the city of Los Angeles. The CUTI represents the degree that the terrestrial animal community is composed of species that are either tolerant or intolerant to urbanization within a spatial unit. We then calculated a mean CUTI across all grid cells in Los Angeles, resulting in a single score for Metric 1.2b in the LA City Biodiversity Index. The methodology provided herein provides a framework for establishing repeated measures over time of urban tolerance within the city of Los Angeles and is applicable to other urban areas. Ultimately, these methods can help local managers and city officials across the region, state, or country understand and track the success (or failures) of local initiatives to support biodiversity and attract historically urban intolerant species to their cities. As urbanization poses a continued threat to biodiversity, particularly in biodiversity hotspots, the methods presented here will enable local governments to better manage and protect native biodiversity.

Methods

Study area

Our study was focused on Southern California, with an emphasis on the greater Los Angeles area, situated in the California Floristic Province, one of 36 biodiversity hotspots in the world [49,50]. The region is also one of five Mediterranean ecosystems in the world, which occur only on the western margins of landmasses between 30 and 40˚ latitude and which are typically characterized by cold and wet winters and warm and dry summers. Southern California has diverse topography, including the transverse mountain ranges to the north and east and the peninsular mountains ranges to the south, as well as diverse habitat types, including chaparral, coastal sage scrub, oak woodlands, coastal dunes and bluffs, riparian woodlands, and a variety of wetland habitats, which host over 2,200 species of vascular plants [51]. While the calculation of a formal CUTI was limited to areas within the city of Los Angeles, for the estimation of species-level UAI, our study area included all land within a 200 km buffer of the Los Angeles City boundary (approximate centroid: 34.031656 N, 118.241716 W), including the cities of Los Angeles, San Diego, Bakersfield, and Santa Barbara. We chose to focus on this broad geographic region because we were interested in creating a metric that could be measured repeatedly over time and would be robust to species that do not currently reside in our focal area of Los Angeles, but could colonize in the future. Furthermore, we treat urban tolerance here (as measured by the UAI) as a species-level trait, which is best estimated using occurrence data from a broader geographic region than just Los Angeles. As such, we aimed to include areas with a wide range of levels of urbanization, including multiple manifestations (i.e., cities) of urban land-use within the broader Southern California region.

Urban intensity

In order to estimate urban associations, we first had to define a continuous spatial layer of the urban intensity of our study region. To do this, we used Principal Component Analysis (PCA), which decomposes multivariate datasets into major axes of variation, to create a single composite index of urban intensity from multiple sources. Following [38], this index included the Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime lights data layer, but we added additional environmental variables related to urbanization as different taxa are likely to respond to different aspects of the urban environment. We initially tested a set of six data layers to depict urban intensity across our study area, but we removed several layers including pollution of fine particles smaller than 2.5 micrometers (PM2.5) [52], Average Traffic Volume [52], and Population Density [53] due to collinearity and coarser resolution. Thus, our PCA index represents a composite of three layers: (1) light pollution from VIIRS Version 4 DMSP-OLS Nighttime Lights Time Series (https://eogdata.mines.edu/products/dmsp/#v4_dmsp_download) [54], (2) "Percent Impervious Surfaces" from National Land Cover Database (NLCD) 2016 Percent Developed Imperviousness (CONUS) (https://www.mrlc.gov/data/nlcd-2016-percent-developed-imperviousness-conus) [55], and (3) "Noise Pollution" from National Park Service (NPS) Geospatial Sound Modeling 2013–2015 (https://irma.nps.gov/DataStore/Reference/Profile/2217356) [56]. All spatial layers were reprojected and resampled, as needed, to a 0.25 x 0.25 mile grid prior to combination. The first axis of the PCA explained 86.5% of the variance in all three layers (S1 Table), indicating as expected that the three layers are all indicative of the same general process (i.e., "urban intensity") yet individually add unique information. Because PCA axis 1 ("PC1") explained >70% of the variation, we retained it as our sole spatial index of urban intensity (Fig 1).

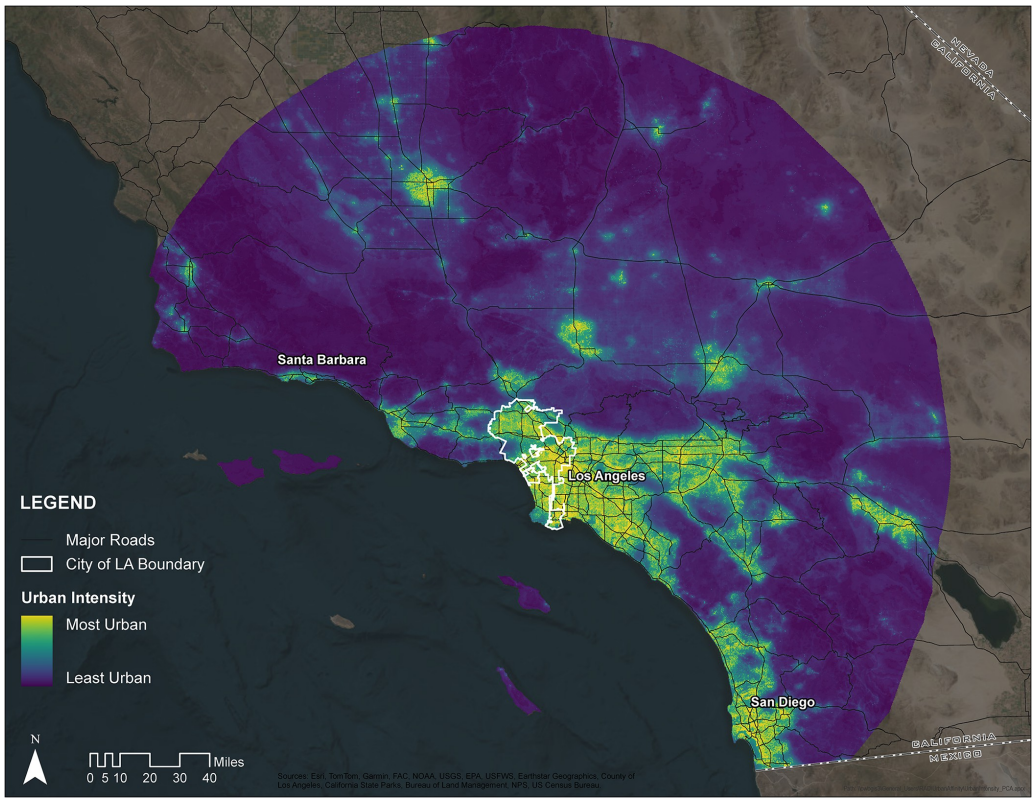

Fig 1. Map of urban intensity measured across a broad study region in Southern California. Our metric of urban intensity was determined as the first PCA axis of three different variables. Warmer colors indicate higher levels of urban intensity. Solid lines detail major roadways within California. Map created using data from [57–61].

https://doi.org/10.1371/journal.pone.0295476.g001

iNaturalist records and data quality filtering

We focused our analysis on selected target taxonomic groups, which were picked a priori with expert input to represent 12 taxonomic groups that are generally well-detected and well-surveyed by community scientists on the iNaturalist platform. The 12 groups include: snails and slugs (Order: Stylommatophora); spiders (Order: Araneae); dragonflies and damselflies (Order: Odonata); grasshoppers, locusts, and crickets(Order: Orthoptera); leafhoppers (Family: Cicadellidae); lady beetles (Family: Coccinellidae); hoverflies (Family: Syrphidae); bees and wasps (Family: Apidae and Vespidae); butterflies and moths (Families: Papilionidae, Pieridae, Lycaenidae, Nymphalidae, Sphingidae and Erebidae); herpetofauna (Classes: Amphibia and Reptilia); mammals (Class: Mammalia); and birds (Class: Aves). We queried the iNaturalist API for occurrence data on 13 January 2022 using the 'rinat' package version 0.1.8 [62]. We downloaded all iNaturalist records between 2011–2021 for the higher order taxa groups defined above, only limiting records to "research quality" georeferenced occurrences (i.e.,

those with a consensus taxonomic ID and location coordinates) bound within our broader study region (i.e., 200 km radius around Los Angeles). After downloading, we further filtered our data to remove species that had more than 60% of all observations records marked as "Geoprivacy = obscured" (a situation where iNaturalist provides spatial coordinates of sightings, but these coordinates are randomly offset by up to 22 km from the true location). Although we additionally filtered out all obscured spatial records for all species, we wholly excluded species meeting this arbitrary threshold as we believed that such a widespread degree of geoprivacy indicated species for which remaining iNaturalist data would not likely represent the species' true distribution within the study area. Finally, we reclassified all records identified to the subspecies level to the species level following other similar work using related datasets [63,64].

Expert review

After downloading iNaturalist records and applying our initial hard filters for data quality, we engaged the LASAN Biodiversity Expert Council, a regional group of scientists and taxonomic specialists who advise the annual biodiversity report, to further assist in data curation and QA/QC. Five experts in areas of specific taxonomic focus for the species included in this study (i.e., arthropods, mollusks, mammals, birds, and herpetofauna) were asked to evaluate occurrence data for the specific higher order taxonomic grouping they specialized in by using the following questions: 1) Is this species native to the study area?; 2) Is the species terrestrial?; 3) Is the natural history of this species so different from others in its taxonomic grouping that records for this particular species should not be used as indicators of search effort for other similar species?; and 4) Is there any other reason why we should exclude this species from this study? The third question refers to the issue that iNaturalist data are presence-only and do not, on their own, provide information on absence or non-detection. Increasingly, however, ecologists are using multi-taxa presence-only surveys to bin species into 'detection groups', whereby an observation of one species at a location provides inference on the non-detection of other species [44,65]. This assumption of substitutability is justified as natural history observers are often searching broadly within taxonomic groups; for example, a birdwatcher's positive record of one bird species says more about the non-detection of another bird species than it does about the non-detection of a butterfly. In the context of the present study, we did not require species' occurrences to be perfect indices of non-detection for other species, but simply sought taxonomic groups where the presence of one species in that group would serve as a broad index of survey effort for all species in that group. Thus, we sought via expert review to exclude taxa that differed so much from the rest of their grouping (e.g., diurnal versus nocturnal; or identifiable via photography versus identifiable only via microscope) that they should not be treated as survey effort proxies. For the fourth question, some common reasons for excluding species based on expert review included species with extremely limited distributions that would otherwise be uninformative to urban tolerance (e.g., a species of plethodontid salamander limited to a single remaining population on Mt. Baldy, Los Angeles, USA), or misidentifications based on recent taxonomic splits.

Following data review by taxonomic specialists, we curated their responses to make sure that experts interpreted these questions similarly. We filtered observations based on these responses to exclude non-native species, species unlikely to be detected by typical observers, non-terrestrial (i.e. marine or freshwater) species, and species according to additional criteria as determined by the taxonomic group specialists.

Controlling for differences in sampling effort

We took a number of steps to control for differences in sampling evenness and effort in iNaturalist data. First, to address the inherent biases associated with sampling evenness, we performed a broad spatiotemporal thin [66,67]. We thinned species-specific data to one observation per year within each 0.25 x 0.25 mile grid cell. This produced a database where every species is recorded as present or not detected in each grid cell and for each year between 2011–2021 (i.e. the number of yearly detections out of 11 years). Second, to address the additional bias associated with varied sampling effort, we defined site-specific sampling effort for each of the twelve focal taxonomic groups. We did this because observers may not equally record observations for all taxonomic groups (e.g., an observer may not observe spiders while recording birds). Taxonomic group-specific effort grids were calculated based on the observation effort per year for the corresponding focal taxonomic groups, summing within grid cells the number of years with at least one observation in a year of a species from within a taxonomic group. Thus, taxonomic group-specific effort layers indicate the number of years (0–11) in which observers obtained at least one record of a target group, which serves as the maximum potential number of thinned presences for any given species of that group. As such, the thinned species presence layer and the matching effort layer represent a spatially-varying binomial response, where the number of binomial trials is the effort in a grid cell and the number of binomial successes is the thinned number of species' presences.

Measuring species-level relationship to urban intensity layer in study area

Species-level indices of urban association were calculated based on thinned occurrence data and taxon-matched effort data across the entire Southern California study region. We calculated a UAI for each species that had at least 25 thinned annual occurrences across our study region. Specifically, using the 'stats' package [68], for each species we modeled the number of thinned occurrences, given effort per cell, as a binomial process that varied as a function of a single covariate: the urban intensity (PC1) of the grid cell. This model allowed us to estimate the number of thinned species occurrences as a binomial variable, where the number of successes (i.e., 'occurrences') was capped by the number of years with non-zero effort for the taxonomic group in each cell. In this way, our model accounted for taxon-specific sampling effort over time in each grid cell. The resulting logit-linear slope of the trend line, which indicates the relationship between species' occurrence and urban intensity, was stored as the UAI for a given species. Positive slopes indicate urban tolerance, negative slopes suggest urban intolerance, and slopes of zero indicate no relationship of occurrence to urban intensity.

Calculating a Community Urban Tolerance Index (CUTI)

After calculating a UAI for each species, we quantified a CUTI for each grid cell in Los Angeles by taxonomic group, as well as a composite score for the entire city (i.e., metric 1.2b for the city of Los Angeles). To calculate a taxonomic group-specific CUTI for each grid cell in Los Angeles, we matched species' UAIs to species occurrences in individual grid cells and calculated a raw CUTI score per cell by taking the mean UAI score of all species within any given cell, weighted by the thinned temporal occurrence of each species (i.e., a value of 1–11 for the number of years that the species occurred in that cell). This resulted in 12 grids (one for each taxonomic group) with a group-specific CUTI score for every cell in the city in which the group was detected. To interpret these results at a broader taxonomic scale, we also calculated the mean CUTI across all pixels for each of the 12 taxonomic groups. Finally, to calculate a composite CUTI across all taxonomic groups, we averaged the 12 taxonomic group grids and city-wide scores. In all cases, the raw CUTI scores were binned into a 5-point scale as follows:

-Infinity to -0.5 = 5; -0.5 to -0.25 = 4; -0.25 to 0 = 3; 0 to 0.25 = 2; 0.25 to 0.5 = 1; and 0.5 to Infinity = 0. On this scale, a cell with a CUTI index of 4 or 5 suggests that species in aggregate are more natural-area associated, while a cell with an index of 0 or 1 suggests that species are more urban tolerant. To test for an association between UIA and urban intensity (with the hypothesis that areas of higher urban intensity have lower CUTI scores), we used an ANOVA and Tukey HSD test in the 'stats' package [69] with urban intensity as the response variable and categorical binned CUTI scores for each grid cell as an independent variable.

Results

Our iNaturalist query yielded a total of 958,624 observations from 127,553 observers (Table 1; [64]). After filtering these observations, we retained 567,996 observations from 71,120 observers. Our filtered query included a total of 967 unique native species found within the study area, of which 563 occurred at least once within the city of Los Angeles. We were able to calculate UAI for 510 species in our dataset, of which 408 occurred at least once within the city of Los Angeles. The species assessed were on average negatively associated with our measure of urbanization, although there was variation across species (cross-species mean = -0.21, range = -2.93 to 0.62; Figs 1 and 2, & S2 Table). UAI varied between the 12 taxonomic groupings (Table 2), with snails and slugs having the highest (i.e. more urban tolerant) score (group mean = 0.24, range = -0.096 to 0.62), and butterflies and moths having the lowest (i.e. more urban intolerant) UAI (group mean = -0.40, range = -2.93 to 0.46). The most urban associated species in our study was the slipper snail (*Cochlicopa lubrica*) (UAI = 0.62), and the least urban associated species was the greenish blue butterfly (*Icaricia saepiolus*) (UAI = -2.93).

Table 1. Counts of observations and observers between unfiltered dataset downloaded from iNat API using specific identifiers for higher order groupings (i.e., iNat Taxa ID) and iNaturalist data subject to exclusion by expert review (see methods for criteria) and hard filters resulting in a filtered dataset of native species.

Taxa Information			Unfiltered		Filtered		
Higher Order Grouping	Taxon	iNat Taxa ID	# Observations	# Observers	# Observations	# Observers	# Species
Snails and Slugs	Stylommatophora	47485	22,539	5,507	919	213	5
Spiders	Araneae	47118	28,148	8,897	16,029	5,778	144
Dragonflies and Damselflies	Odonata	47792	16,391	3,431	16,366	3,426	73
Grasshoppers, Locusts, and Crickets	Orthoptera	47651	17,267	5,670	13,816	4,483	140
Leafhoppers	Cicadellidae	53237	2,585	849	796	236	41
Lady Beetles	Coccinellidae	48486	18,030	5,090	9,246	2,677	37
Hoverflies	Syrphidae	49995	10,654	2,200	8,212	1,832	40
Bees and Wasps	Apidae	47221	34,748	11,105	11,499	3,166	68
	Vespidae	52747	5,113	1,826			
Butterflies and Moths	Erebidae	121850	3,939	1,714	28,480	6,973	68
	Lycaenidae	47923	15,199	2,479			
	Nymphalidae	47922	38,786	10,023			
	Papilionidae	47223	8,514	3,963			
	Pieridae	48508	9,528	2,504			
	Sphingidae	47213	6,988	3,967			
Herpetofauna	Amphibia	20978	19,892	5,229	72,312	13,992	42
	Reptilia	26036	70,637	13,601			
Mammals	Mammalia	40151	56,914	11,763	34,107	7,514	49
Birds	Aves	3	572,752	27,735	356,214	20,830	260
		TOTAL	958,624	127,553	567,996	71,120	967

https://doi.org/10.1371/journal.pone.0295476.t001

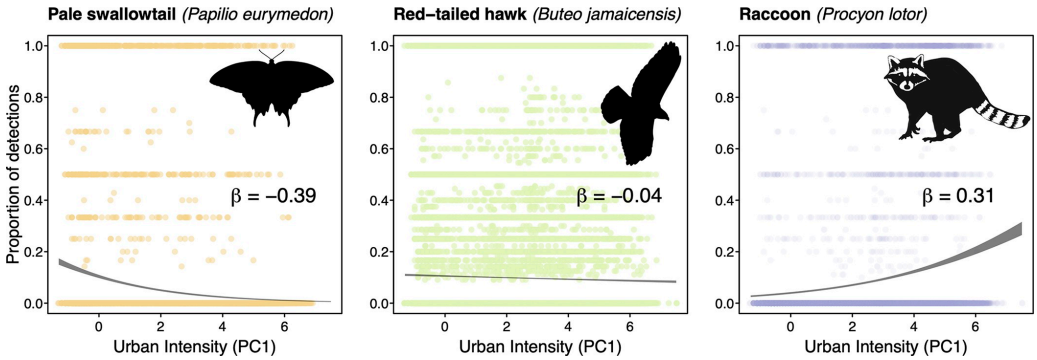

Fig 2. Species vary widely in their response of occurrence to urbanization. Scatterplots of three different taxa showing urban intolerance (left), urban ambivalence (center), and urban tolerance (right). Scatterplots show the proportion of detections (out of a maximum of 11 years) for each species across each grid cell in the broader Southern California study region (Fig 1). Trend lines show the 95% confidence interval surrounding a binomial regression of detection frequency as a function of urban intensity. Species' UAI scores (β) are the logit-linear slope of the trend line.

https://doi.org/10.1371/journal.pone.0295476.g002

We assessed 8,348 0.25 x 0.25 mile grid cells across the city of Los Angeles for their CUTI, weighted by the temporally thinned occurrence of each species. A total of 2,010 grid cells did not have records for any target taxa after filtering and were not included in our calculation of summary scores. Averaging across all higher order taxonomic groupings, the city had an average CUTI of 2.01 (raw, unbinned score = 0.129; Fig 3). Average CUTI varied between higher order taxonomic groupings (Table 3), with snails and orthoptera demonstrating the highest average CUTI (snails: 1.40 binned, 0.30 raw; orthoptera: 1.54 binned, 0.17 raw), and odonates and mammals showing the lowest average CUTI (odonates: 2.25 binned, 0.06 raw; mammals: 2.16 binned, 0.07 raw) (Fig 4). There was a significant relationship between urbanization values and the cross-taxa average CUTI of those cells, with areas of higher urban intensity holding taxa that, on the whole, were more urban tolerant (i.e, have higher UAI values) (ANOVA, $p < 0.001$; S1 Fig). This general relationship held true for every individual taxonomic group (ANOVA, $p < 0.001$), except for snails (ANOVA, $p = 0.99$).

Table 2. Average urban association index (UAI) scores for each of the 12 higher order taxonomic groupings. Species-level UAI scores ranged from -2.9 to 0.62, with more negative numbers indicating more urban intolerant species and more positive scores indicating more urban tolerant species. Range of values in parentheses.

Higher Order Grouping	Average UAI Score (range of all species)
Snails and Slugs	0.24 (-0.10 to 0.62)
Spiders	-0.14 (-0.90 to 0.62)
Dragonflies and Damselflies	-0.20 (-1.32 to 0.37)
Grasshoppers, Locusts, and Crickets	-0.37 (-2.20 to 0.36)
Leafhoppers	-0.08 (-0.51 to 0.34)
Lady Beetles	-0.08 (-0.63 to 0.37)
Hoverflies	-0.11 (-0.54 to 0.27)
Bees and Wasps	-0.16 (-0.83 to 0.43)
Butterflies and Moths	-0.40 (-2.93 to 0.46)
Herpetofauna	-0.34 (-1.32 to 0.35)
Mammals	-0.39 (-2.20 to 0.31)
Birds	-0.15 (-1.79 to 0.47)

https://doi.org/10.1371/journal.pone.0295476.t002

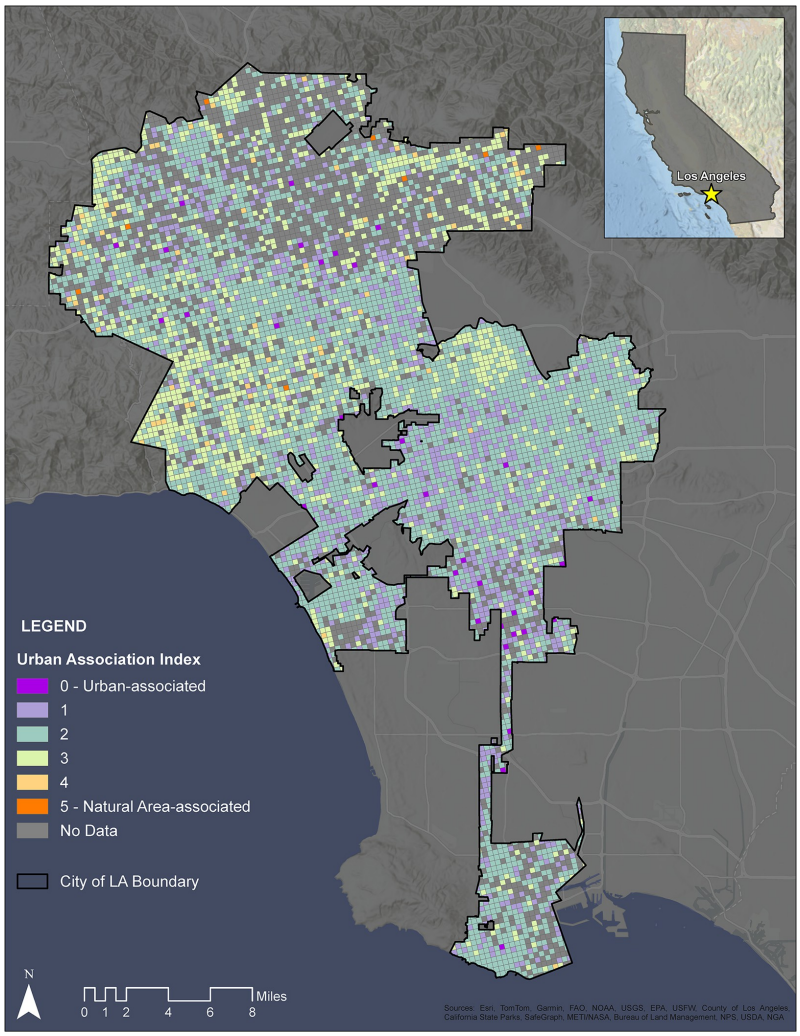

Fig 3. Map of the City of Los Angeles with overlaid mean composite urban association (UAI) scores. Warmer colors indicate more natural area tolerant species, whereas cooler colors indicate more urban tolerant species. Areas within Los Angeles city boundaries with insufficient data to calculate the score are colored gray. Map uses a 0.25 x 0.25 mile scale.

https://doi.org/10.1371/journal.pone.0295476.g003

Discussion

Using over 567,000 publicly available community science records from iNaturalist, we present the first comprehensive species-level evaluation of urban tolerance for Southern California taxa. For the City of Los Angeles, we found that on average, native species within the city were negatively associated with our measure of urban intensity (i.e. light pollution, impervious

Table 3. Community urban tolerance index (CUTI) scores for each of the 12 higher order taxonomic groupings. Raw CUTI values are the average of weighted average of species-level UAI scores, while binned CUTI values rescale to a 5-point index, where a CUTI index of 4–5 suggests that species in aggregate are more natural-area associated, while an index of 0–1 suggests that species are more urban tolerant.

Taxa	Raw CUTI	Binned CUTI
Snails and Slugs	0.30	1.40
Spiders	0.22	1.63
Dragonflies and Damselflies	0.06	2.25
Grasshoppers, Locusts, and Crickets	0.17	1.54
Leafhoppers	0.10	2.15
Lady Beetles	0.12	1.98
Hoverflies	0.10	1.95
Bees and Wasps	0.20	1.81
Butterflies and Moths	0.13	1.96
Herpetofauna	0.14	1.99
Mammals	0.07	2.16
Birds	0.10	2.15

https://doi.org/10.1371/journal.pone.0295476.t003

surfaces, and noise pollution), and that this pattern was conserved across the higher order taxonomic groupings with the exception of snails and slugs, which were positively associated with urban intensity. We then averaged this species-level response to our measure of urban intensity on a ¼ mile grid across the City of Los Angeles to understand the geographic distribution of urban intolerant species. Ultimately, this metric of native species urban tolerance (CUTI) can be reassessed regularly as a means of evaluating change in urban tolerance over time and following specific biodiversity improvement measures in the city.

Extrinsic and intrinsic factors leading to differences in UAI

The most urban tolerant species in our study was the slipper snail (*C. lubrica*). This species is known to be widespread and euryhygric (i.e. able to withstand a broad range of moisture conditions), and it is possible that urbanized areas such as Los Angeles provide year-round access to a variety of moisture regimes through the addition of ornamental landscapes and lawns [70–72]. Other studies have found high native snail abundance in areas of high urban intensity in Tennessee [73]. Given that there were only five species of snails and slugs that remained in our dataset post-filtering, this pattern could largely be due to a small sample size of species, and perhaps more purposeful sampling is required to truly ascertain the affinity of this entire taxonomic group for urban environments.

The most urban intolerant species in our study was the greenish blue butterfly (*Icaricia saepiolus*). Compared to other North American butterfly families, Lycaenidae is overrepresented in terms of number of species proposed for listing [74]. This is largely due to host plant specificity of Lycaenids, primarily for plants in the genera *Lupinus* and *Eriogonum*, and the fact that these plants are adapted to disturbance regimes that are infrequent in the urban context [74–76]. Conservation of many of the special status butterfly species including several Lycaenids therefore relies on maintaining and expanding critical segments of habitat that contain host plant species within urban settings and maintaining habitat fragments of varying sizes through deliberate disturbance as a management tool [76]. For example, the Palos Verdes blue butterfly (*Glaucopsyche lygdamus palosverdesensis*) is an endangered subspecies of Lycaenid butterfly in Los Angeles County, and findings from the US Fish and Wildlife Service demonstrate that the species appears to be establishing in reintroduction sites due primarily to efforts

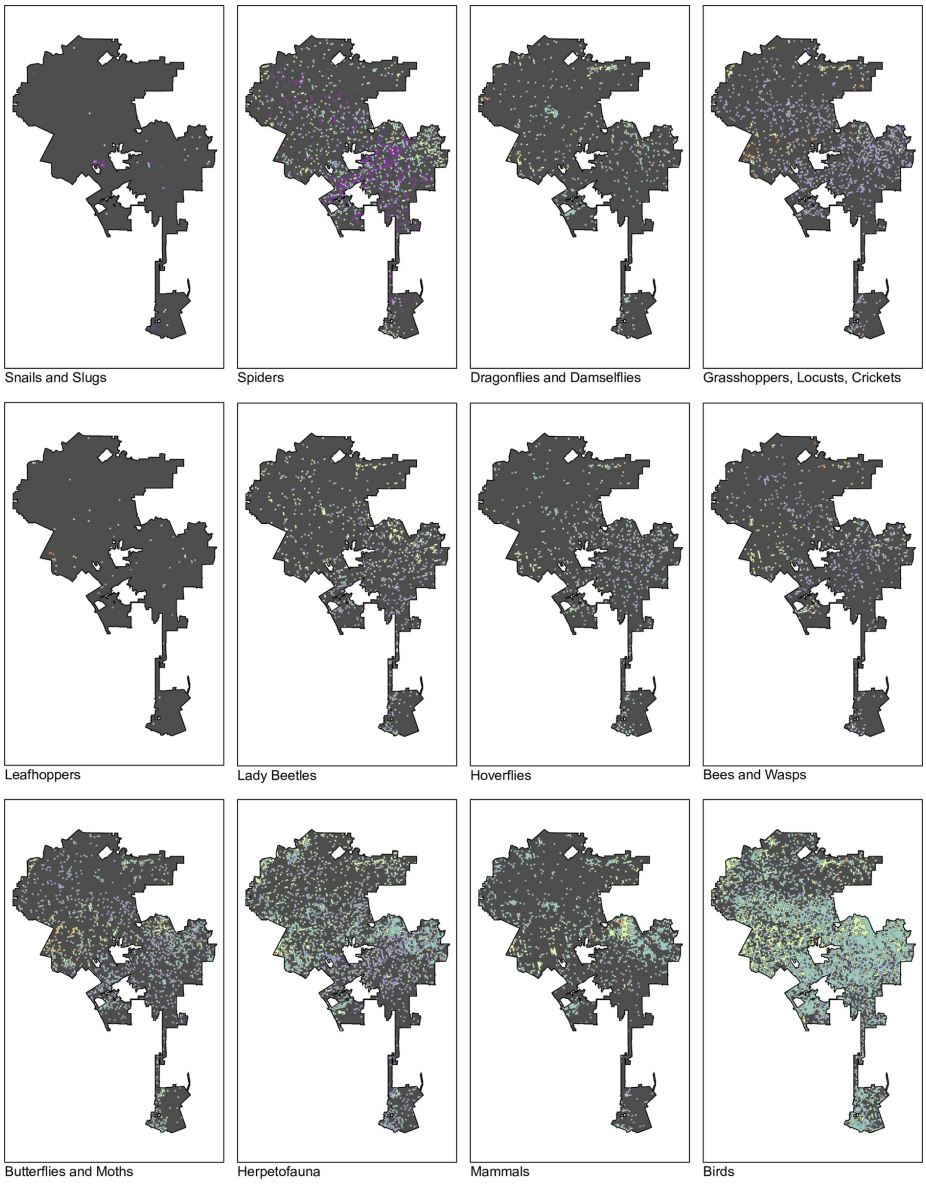

Fig 4. Map of the urban association scores separated by higher order taxonomic groupings. Levels correspond to Fig 3, where warmer colors indicate more urban intolerant species, cooler colors indicate more urban tolerant species, and gray indicates cells with insufficient data to calculate a score.

https://doi.org/10.1371/journal.pone.0295476.g004

to remediate historical habitat through mechanical disturbance, non-native plant removal, and planting of successional host plants [77]. Based on the findings from this study, the City of Los Angeles could use existing butterfly observations to identify target areas for conservation for the Palos Verdes blue butterfly and other imperiled butterfly species. In areas where observations overlap with property owned and managed by the City of Los Angeles, the city can focus restoration efforts on increasing host plant abundance through direct plantings of early successional plants and targeted mechanical disturbance to create conditions necessary for early successional habitat needed by the butterflies. Following restoration activities like these, city managers can evaluate success through a reevaluation of this metric on an annual basis.

All other higher order taxonomic groups had higher presence in less urbanized regions within the city. Previous research in Los Angeles, research has demonstrated that the presence of many taxonomic groups is negatively affected by increased levels of urbanization. For example, coyotes and bobcats in mixed urban/natural areas have home ranges that primarily utilize natural areas [78], and some regionally common amphibian species are markedly absent from streams within urbanized areas of the city [79], but within taxonomic groupings there was high levels of variation in individual species-level responses. Many factors contribute to these varied responses of species to different levels of urban intensity, including species-specific functional traits. Specifically, in mammals and birds, functional traits help explain urban tolerance of species, including body size, dietary breadth, clutch size, and nesting strategy, among other predictors in birds [64] and litter size in mammals [80]. While our study does not seek to evaluate all individual species-level traits and how they relate to values of the UAI, some previously noted relationships between species' functional traits and responses to urbanization are recapitulated in our findings. For example, Cooper et al. [81] found that urban raptors in Los Angeles including Cooper's hawk (*Accipiter cooperii*) and red-shouldered hawk (*Buteo lineatus*) responded to increases in urban cover within their home ranges by increasing nesting in urban sites, while other species of urban raptors such as American kestrels (*Falco sparverius*) were more likely to nest in less-urban areas within their home ranges. They concluded that nesting strategy may play a role in the response of urban raptors to levels of urbanization, as American kestrels are cavity nesters and dead trees are likely to be removed by homeowners [81]. Notably, these previously-reported responses to urbanization for Cooper's hawks, red-shouldered hawks, and American kestrel are confirmed in our estimated values of UAI (0.11, 0.19, and -0.03 respectively). Future studies could further explore the relationships between species-specific functional traits and UAI for the 510 native species studied here using existing databases of functional traits such as AVONET [82], AnimalTraits [83], or COMBINE [84].

In addition to intrinsic characteristics of species, many extrinsic factors may contribute to species-level responses to varying levels of urban intensity. For example, research on urban insect populations in Los Angeles found that diurnal temperature range had a consistent negative effect on occurrence of all arthropod species studied [85]. We found a negative association with urbanization across all native arthropods, although we also observed high levels of variation within higher order taxonomic groupings. It is possible that spatial variation in water resources across the city can partly explain this within-group variation, as has been demonstrated for desiccation-sensitive invasive insect species in Los Angeles [86]. The heterogeneous presence of water on the landscape also strongly influences the vegetation community present, which in turn strongly influences arthropod community composition. Additional research could build off our findings to investigate the relationship between our measures of UAI and both spatial variation in abiotic gradients and functional traits of individual species, such as desiccation tolerances of insects. The large amount of variation within groups can perhaps also be explained by different species-level responses to the anthropogenic stressors that come with urbanization, namely light pollution and noise pollution. Because spatial datasets of artificial

light at night (ALAN) and anthropogenic noise were used in the creation of our urban intensity layer for this study, we know that these measures of anthropogenic stressors explain much of the variation in our dataset when considered in aggregate (S1 Table). This is unsurprising given the broad impacts of ALAN [87] and anthropogenic noise [88] for many different groups of organisms, including many of the higher-order taxonomic groupings studied here, such as birds (light: [89]; noise:[90]), mammals (light: [87]; noise: [91]), and invertebrates (light: [92]; noise: [93]). These stressors can lead to a variety of responses that are specific to individual species or communities, including some positive responses (e.g., some bats species have increased foraging opportunities as a results of ALAN; [94]). Finally, large within-group variation in UAI values may also be related to the variable, and relatively artificial, taxonomic levels that we used to aggregate data; future studies with more fine resolution groupings (e.g., at the genus level) may reveal more taxonomically conserved and biologically relevant relationships.

Limitations of and future directions for using crowd-sourced data

While this study includes over 500 native species observed within the study area, approximately 59% of the total community science records were excluded from the analyses because they were considered undetectable to the general public (e.g. small insects, nocturnal mammals, etc.), the data were not at the "research grade" level, or the records for a given species were geoobscured due to species status or user preference. Other studies have noted similar data quality issues and biases in iNaturalist records [95–97]. In order to circumvent data quality issues, this study relied on expert review to identify higher order taxonomic groupings that could be reliably identified by the general public, but in doing so, may have increased ascertainment bias and decreased the overall scope of the data. In an effort to reduce the amount of data lost, future assessments of this metric may benefit from developing a relationship with iNaturalist in order to obtain user-obscured data en masse, which is not currently possible without requesting thousands of individual records from each iNaturalist user. This study may also be limited by bias within higher order taxonomic groups for species that are common or more easily observed, as has been reported previously [98,99]). While we were unable to confirm whether any given species is present within our dataset more or less often than "true" occurrence, due to a relationship with their abundance on the landscape or other factors, we believe that this bias should act randomly across the study area and therefore not impact the overall interpretability of our findings. Additional work could greatly benefit the field by investigating the potential over- or underrepresentation of common species within community science datasets.

While we present several limitations to the available crowd-sourced species presence dataset within our sampling area, these data limitations also provide targets for local environmental managers to improve these datasets and therefore biodiversity monitoring in their regions. Based solely on the number of assessed grid cells across the City of Los Angeles in this study, it is clear that there needs to be substantial effort placed on bolstering community science projects that focus on underrepresented taxonomic groups (e.g. snails and slugs and leafhoppers). Findings from other community science projects indicate that local city residents are underrepresented contributors to community science datasets [100], yet for community science-based biodiversity monitoring to be successful, it must be built from a bottom-up approach that includes both participatory and contributory opportunities for the communities where biodiversity monitoring is to take place [101]. While some efforts in Los Angeles to involve residents in taxonomically focused community science projects have led to increased knowledge of urban biodiversity for these taxa groups (e.g. the BioSCAN project, see https://nhm.org/community-science-nhm/bioscan; [102]), these projects are limited in geographic scope, are

often short-term assessments, and the primary role of residents is that of data collection, which may not lead to sustained participation by community members in the future [103]. Moreover, efforts to educate the public on specific indicator species is underway in Los Angeles [104], and these efforts could be directed at underrepresented taxonomic groups highlighted in this study. It has been demonstrated that community science datasets can match if not surpass traditional biodiversity assessment methods in data quantity, and do so in a fraction of the time [105,106]. Therefore, developing long-term and mutually beneficial partnerships with local communities to assess urban biodiversity should be a primary focus of city managers who plan to use large unstructured community science datasets to measure the efficacy of city-wide biodiversity measures.

Conclusions

Herein we present a broad taxonomic assessment of the urban tolerance of native animal species for the City of Los Angeles. We found that there are clear differences in species level responses to our measures of urban intensity and that native species within Southern California are largely urban intolerant. This is even more true within the City of Los Angeles. This study provides a baseline assessment of the degree of presence of urban intolerant species within the City of Los Angeles in a ten year study period. Repeated assessment of this metric will allow stakeholders such as the City of Los Angeles to monitor success of its stated goal of no-net biodiversity loss by 2035. An important metric within the City's Biodiversity Index is the ability of this urban system to attract and maintain healthy populations of urban intolerant species.

Supporting information

S1 Fig. Regression plot of urban intensity (first PCA axis) and CUTI scores for all 510 species.
(TIF)

S1 Table. Table detailing the loadings and variance explained for the composite urban intensity layer separated by contributing layers and PC axes.
(DOCX)

S2 Table. Table of all species considered in this paper and their associated urban affinity scores.
(CSV)

Acknowledgments

We would like to acknowledge Dan Cooper, Miguel Ordenana, and Jessica West for providing us access to iNaturalist project data. The study authors would like to specifically thank Gary Bucciarelli, Rachel Chock, and Jann Vendetti for their expert review of iNaturalist data of herpetofauna, mammals, and slugs and snails, respectively. The study authors would also like to acknowledge the LA City Biodiversity Expert Council, and particularly the members who attended the workshop for Metric 1.2b. Finally, we thank the iNaturalist observers and identifiers for their contributions that made this research possible.

Author Contributions

Conceptualization: Joseph N. Curti, Michelle Barton, Rhay G. Flores, Maren Lechner, Alison Lipman, Albert Y. Park, Morgan W. Tingley.

Data curation: Joseph N. Curti, Graham A. Montgomery, Morgan W. Tingley.

Formal analysis: Joseph N. Curti, Morgan W. Tingley.

Methodology: Michelle Barton, Morgan W. Tingley.

Project administration: Joseph N. Curti, Michelle Barton, Alison Lipman, Morgan W. Tingley.

Supervision: Joseph N. Curti, Michelle Barton, Alison Lipman, Morgan W. Tingley.

Visualization: Kirstin Rochel, Morgan W. Tingley.

Writing – original draft: Joseph N. Curti.

Writing – review & editing: Joseph N. Curti, Michelle Barton, Rhay G. Flores, Maren Lechner, Alison Lipman, Graham A. Montgomery, Albert Y. Park, Morgan W. Tingley.

References

1. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. Accelerated modern human–induced species losses: entering the sixth mass extinction. Sci Adv. 2015; 1: e1400253. https://doi.org/10.1126/sciadv.1400253 PMID: 26601195
2. McCallum ML. Vertebrate biodiversity losses point to a sixth mass extinction. Biodivers Conserv. 2015; 24: 2497–2519. https://doi.org/10.1007/s10531-015-0940-6
3. Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 2003; 421: 37–42. https://doi.org/10.1038/nature01286 PMID: 12511946
4. Sodhi NS, Bickford D, Diesmos AC, Lee TM, Koh LP, Brook BW, et al. Measuring the meltdown: drivers of global amphibian extinction and decline. Freckleton R, editor. PLoS One. 2008; 3: e1636. https://doi.org/10.1371/journal.pone.0001636 PMID: 18286193
5. Szabo JK, Khwaja N, Garnett ST, Butchart SHM. Global patterns and drivers of avian extinctions at the species and subspecies level. PLoS One. 2012; 7: e47080. https://doi.org/10.1371/journal.pone.0047080 PMID: 23056586
6. IPBES. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Zenodo; 2019. https://doi.org/10.5281/zenodo.6417333
7. Munstermann MJ, Heim NA, McCauley DJ, Payne JL, Upham NS, Wang SC, et al. A global ecological signal of extinction risk in terrestrial vertebrates. Conserv Biol. 2022; 36: e13852. https://doi.org/10.1111/cobi.13852 PMID: 34668599
8. Brook B, Sodhi N, Bradshaw C. Synergies among extinction drivers under global change. Trends Ecol Evol. 2008; 23: 453–460. https://doi.org/10.1016/j.tree.2008.03.011 PMID: 18582986
9. Fenoglio MS, Calviño A, González E, Salvo A, Videla M. Urbanisation drivers and underlying mechanisms of terrestrial insect diversity loss in cities. Ecol Entomol. 2021; 46: 757–771. https://doi.org/10.1111/een.13041
10. Ruas RDB, Costa LMS, Bered F. Urbanization driving changes in plant species and communities–a global view. Glob Ecol Conserv. 2022; 38: 1–17. e02243. https://doi.org/10.1016/j.gecco.2022.e02243
11. Seto KC, Güneralp B, Hutyra LR. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc Natl Acad Sci USA. 2012; 109: 16083–16088. https://doi.org/10.1073/pnas.1211658109 PMID: 22988086
12. Batáry P, Kurucz K, Suarez-Rubio M, Chamberlain DE. Non-linearities in bird responses across urbanization gradients: a meta-analysis. Glob Chang Biol. 2018; 24: 1046–1054. https://doi.org/10.1111/gcb.13964 PMID: 29080260
13. Faeth SH, Bang C, Saari S. Urban biodiversity: patterns and mechanisms. Ann NY Acad Sci. 2011; 1223: 69–81. https://doi.org/10.1111/j.1749-6632.2010.05925.x PMID: 21449966
14. Fenoglio MS, Rossetti MR, Videla M. Negative effects of urbanization on terrestrial arthropod communities: a meta-analysis. Global Ecol Biogeogr. 2020; 29: 1412–1429. https://doi.org/10.1111/geb.13107

15. Rega-Brodsky CC, Aronson MFJ, Piana MR, Carpenter E- S, Hahs AK, Herrera-Montes A, et al. Urban biodiversity: state of the science and future directions. Urban Ecosyst. 2022; 25: 1083–1096. https://doi.org/10.1007/s11252-022-01207-w

16. McDonnell MJ, Pickett STA. Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology. 1990; 71: 1232–1237. https://doi.org/10.2307/1938259

17. Blair R. The effects of urban sprawl on birds at multiple levels of biological organization. Ecol Soc. 2004; 9: art2. https://doi.org/10.5751/ES-00688-090502

18. Blair RB, Johnson EM. Suburban habitats and their role for birds in the urban–rural habitat network: points of local invasion and extinction? Landsc Ecol. 2008; 23: 1157–1169. https://doi.org/10.1007/s10980-008-9267-y

19. Husté A, Boulinier T. Determinants of local extinction and turnover rates in urban bird communities. Ecol Appl. 2007; 17: 168–180. https://doi.org/10.1890/1051-0761(2007)017[0168:doleat]2.0.co;2 PMID: 17479843

20. Fontana S, Sattler T, Bontadina F, Moretti M. How to manage the urban green to improve bird diversity and community structure. Landsc Urban Plan. 2011; 101: 278–285. https://doi.org/10.1016/j.landurbplan.2011.02.033

21. de Toledo MCB, Donatelli RJ, Batista GT. Relation between green spaces and bird community structure in an urban area in Southeast Brazil. Urban Ecosyst. 2012; 15: 111–131. https://doi.org/10.1007/s11252-011-0195-2

22. Wood EM, Esaian S. The importance of street trees to urban avifauna. Ecol Appl. 2020; 30: e02149. https://doi.org/10.1002/eap.2149 PMID: 32340072

23. Paker Y, Yom-Tov Y, Alon-Mozes T, Barnea A. The effect of plant richness and urban garden structure on bird species richness, diversity and community structure. Landsc Urban Plan. 2014; 122: 186–195. https://doi.org/10.1016/j.landurbplan.2013.10.005

24. Haight JD, Hall SJ, Fidino M, Adalsteinsson SA, Ahlers AA, Angstmann J, et al. Urbanization, climate and species traits shape mammal communities from local to continental scales. Nat Ecol Evol. 2023; 7: 1654–1666. https://doi.org/10.1038/s41559-023-02166-x PMID: 37667002

25. Magle SB, Hunt VM, Vernon M, Crooks KR. Urban wildlife research: past, present, and future. Biol Conserv. 2012; 155: 23–32. https://doi.org/10.1016/j.biocon.2012.06.018

26. Beninde J, Veith M, Hochkirch A. Biodiversity in cities needs space: a meta-analysis of factors determining intra-urban biodiversity variation. Ecol Lett. 2015; 18: 581–592. https://doi.org/10.1111/ele.12427 PMID: 25865805

27. Spotswood EN, Beller EE, Grossinger R, Grenier JL, Heller NE, Aronson MFJ. The biological deserts fallacy: cities in their landscapes contribute more than we think to regional biodiversity. BioScience. 2021; 71: 148–160. https://doi.org/10.1093/biosci/biaa155 PMID: 33613128

28. Wehtje W. The range expansion of the great-tailed grackle (*Quiscalus mexicanus* Gmelin) in North America since 1880. J Biogeogr. 2003; 30: 1593–1607. https://doi.org/10.1046/j.1365-2699.2003.00970.x

29. Veech JA, Small MF, Baccus JT. The effect of habitat on the range expansion of a native and an introduced bird species: habitat and range expansion. J Biogeogr. 2011; 38: 69–77. https://doi.org/10.1111/j.1365-2699.2010.02397.x

30. Ancillotto L, Santini L, Ranc N, Maiorano L, Russo D. Extraordinary range expansion in a common bat: the potential roles of climate change and urbanisation. Sci Nat. 2016; 103: 15. https://doi.org/10.1007/s00114-016-1334-7

31. Urošević A, Tomović L, Ajtić R, Simović A, Džukić G. Alterations in the reptilian fauna of Serbia: introduction of exotic and anthropogenic range expansion of native species. Herpetozoa. 2016; 28: 115–132.

32. Lerman SB, Warren PS. The conservation value of residential yards: linking birds and people. Ecol Appl. 2011; 21: 1327–1339. https://doi.org/10.1890/10-0423.1 PMID: 21774433

33. Wooster EIF, Fleck R, Torpy F, Ramp D, Irga PJ. Urban green roofs promote metropolitan biodiversity: a comparative case study. Build Environ. 2022; 207: 108458. https://doi.org/10.1016/j.buildenv.2021.108458

34. Yang J. Big data and the future of urban ecology: from the concept to results. Sci China Earth Sci. 2020; 63: 1443–1456. https://doi.org/10.1007/s11430-020-9666-3

35. Uchida K, Blakey RV, Burger JR, Cooper DS, Niesner CA, Blumstein DT. Urban biodiversity and the importance of scale. Trends Ecol Evol. 2021; 36: 123–131. https://doi.org/10.1016/j.tree.2020.10.011 PMID: 33168154

36. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage. 2015; 35: 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

37. Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. BioScience. 2021; 71: 1179–1188. https://doi.org/10.1093/biosci/biab093

38. Callaghan CT, Ozeroff I, Hitchcock C, Chandler M. Capitalizing on opportunistic citizen science data to monitor urban biodiversity: a multi-taxa framework. Biol Conserv. 2020; 251. https://doi.org/10.1016/j.biocon.2020.108753

39. Li E, Parker SS, Pauly GB, Randall JM, Brown BV, Cohen BS. An urban biodiversity assessment framework that combines an urban habitat classification scheme and citizen science data. Front Ecol Evol. 2019; 7. https://doi.org/10.3389/fevo.2019.00277

40. Kamp J, Oppel S, Heldbjerg H, Nyegaard T, Donald PF. Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. Divers Distrib. 2016; 22: 1024–1035. https://doi.org/10.1111/ddi.12463

41. Rapacciuolo G, Young A, Johnson R. Deriving indicators of biodiversity change from unstructured community-contributed data. Oikos. 2021; 130: 1225–1239. https://doi.org/10.1111/oik.08215

42. Bayraktarov E, Ehmke G, O'Connor J, Burns EL, Nguyen HA, McRae L, et al. Do big unstructured biodiversity data mean more knowledge? Front Ecol Evol. 2019; 6: 239. https://doi.org/10.3389/fevo.2018.00239

43. Van Strien AJ, Van Swaay CAM, Termaat T. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Devictor V, editor. J Appl Ecol. 2013; 50: 1450–1458. https://doi.org/10.1111/1365-2664.12158

44. Isaac NJB, Strien AJ, August TA, Zeeuw MP, Roy DB. Statistics for citizen science: extracting signals of change from noisy ecological data. Anderson B, editor. Methods Ecol Evol. 2014; 5: 1052–1060. https://doi.org/10.1111/2041-210X.12254

45. Steen VA, Elphick CS, Tingley MW. An evaluation of stringent filtering to improve species distribution models from citizen science data. Albright T (Oe), editor. Divers Distrib. 2019; 25: 1857–1869. https://doi.org/10.1111/ddi.12985

46. LA Biodiversity Index Baseline Report. Los Angeles Sanitation and Environment; 2022. Available: https://www.lacitysan.org/san/sandocview?docname=cnt076756

47. 2018 Biodiversity Report. Los Angeles Sanitation and Environment; 2018. Available: https://www.lacitysan.org/cs/groups/public/documents/document/y250/mdi0/~edisp/cnt024743.pdf

48. 2020 Biodiversity Report. Los Angeles Sanitation and Environment; 2020. Available: https://www.lacitysan.org/cs/groups/public/documents/document/y250/mduy/~edisp/cnt052553.pdf

49. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000; 403: 853–858. https://doi.org/10.1038/35002501 PMID: 10706275

50. Burge DO, Thorne JH, Harrison SP, O'Brien BC, Rebman JP, Shevock JR, et al. Plant diversity and endemism in the California Floristic Province. Madroño. 2016; 63: 3–206. https://doi.org/10.3120/madr-63-02-3-206.1

51. Rundel P, Gustafson RJ. Introduction to the plant life of southern California: coast to foothills. University of California Press; 2005.

52. CalEnviroScreen 4.0. California Office of Environmental Health Hazard Assesment; 2021. Available: https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-40

53. Updated Demographics Variables 2023. ESRI; 2023. Available: https://www.arcgis.com/home/item.html?id=ddd1510ed1964e2e8fb1ded93e9b03b1

54. Elvidge CD, Baugh KE, Zhizhin M, Hsu F-C. Why VIIRS data are superior to DMSP for mapping nighttime lights. APAN Proceedings. 2013; 35: 62. https://doi.org/10.7125/APAN.35.7

55. Dewitz J. National land cover database (NLCD) 2016 products: US Geological Survey data release. 2019.

56. Mennitt D, Sherrill K, Fristrup K. A geospatial model of ambient sound pressure levels in the contiguous United States. J Acoust Soc Am. 2014; 135: 2746–2764. https://doi.org/10.1121/1.4870481 PMID: 24815258

57. Esri. "World Imagery" [basemap]. Scale Not Given. "World Imagery". December 12, 2009. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9. (Feb 22, 2024).

58. Esri. "World Hillshade" [basemap]. Scale Not Given. "World Hillshade". July 9, 2015. https://www.arcgis.com/home/item.html?id=1b243539f4514b6ba35e7d995890db1d. (Feb 22, 2024).

59. Esri. "Dark Gray Base" [basemap]. Scale Not Given. "Dark Gray Canvas". November 13, 2015. https://www.arcgis.com/home/item.html?id=358ec1e175ea41c3bf5c68f0da11ae2b. (Feb 22, 2024).

60. U.S. Census Bureau. (n.d.). *cb_2018_us_state_500k.zip*. U.S. Department of Commerce. Retrieved April 15, 2019, from https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

61. U.S. Census Bureau. (n.d.). *tl_2020_06_prisecroads.zip*. U.S. Department of Commerce. Retrieved January 11, 2021, from https://www2.census.gov/geo/tiger/TIGER2020/PRISECROADS/

62. Barve V, Hart E (2024). Rinat: access 'iNaturalist' data through APIs. R package version 0.1.9. https://docs.ropensci.org/rinat/ (website) https://github.com/ropensci/rinat (devel).

63. Feng M-LE, Che-Castaldo J. Comparing the reliability of relative bird abundance indices from standardized surveys and community science data at finer resolutions. Silva D de P, editor. PLoS One. 2021; 16: e0257226. https://doi.org/10.1371/journal.pone.0257226 PMID: 34506572

64. Neate-Clegg MHC, Tonelli BA, Youngflesh C, Wu JX, Montgomery GA, Şekercioğlu ÇH, et al. Traits shaping urban tolerance in birds differ around the world. Curr Biol. 2023; 33: 1677–1688.e6. https://doi.org/10.1016/j.cub.2023.03.024 PMID: 37023752

65. Outhwaite CL, Gregory RD, Chandler RE, Collen B, Isaac NJB. Complex long-term biodiversity change among invertebrates, bryophytes and lichens. Nat Ecol Evol. 2020; 4: 384–392. https://doi.org/10.1038/s41559-020-1111-z PMID: 32066888

66. Boria RA, Olson LE, Goodman SM, Anderson RP. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol Modell. 2014; 275: 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

67. Steen VA, Tingley MW, Paton PWC, Elphick CS. Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. Methods Ecol Evol. 2021; 12: 216–226. https://doi.org/10.1111/2041-210X.13525

68. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: https://www.R-project.org/

69. Observations of multiple species from Southern California, U.S.A. observed on/between 2011–2021. Exported from https://www.inaturalist.org on Jan 2022.

70. Barker GM. Naturalised terrestrial Stylommatophora: Mollusca: Gastropoda. Lincoln: Manaaki Whenua Press; 1999.

71. Georgiev D. Habitat distribution of the land snails in one village area of the upper Thracian valley (Bulgaria). Anniversary Scientific Conference of Ecology, Proceedings. 2008; 1.

72. Čejka T, Hamerlík L. Land snails as indicators of soil humidity in Danubian woodland (SW Slovakia). Pol J Ecol. 2009; 57: 741–747.

73. Hodges MN, McKinney ML. Urbanization impacts on land snail community composition. Urban Ecosyst. 2018; 21: 721–735. https://doi.org/10.1007/s11252-018-0746-x

74. Cushman J, Murphy D. Conservation of North American lycaenids–an overview. Conservation Biology of Lycaenidae(Butterflies). Gland, Switzerland: IUCN; 1993. pp. 716–717. Available: https://www.nature.com/articles/188716b0

75. MacDonald B, Longcore T, Weiss S. Status and variability of Mission blue butterfly populations. Golden Gate National Parks Conservancy; 2012.

76. Longcore T, Osborne KH. Butterflies are not grizzly bears: Lepidoptera conservation in practice. In: Daniels JC, editor. Butterfly Conservation in North America. Dordrecht: Springer Netherlands; 2015. pp. 161–192. https://doi.org/10.1007/978-94-017-9852-5_9

77. Anderson A. Palos Verdes blue butterfly (Glaucopsyche lygdamus palosverdesensis) 5-Year Review. United States Fish and Wildlife Service; 2014. Available: https://ecosphere-documents-production-public.s3.amazonaws.com/sams/public_docs/species_nonpublish/2149.pdf

78. Riley SPD, Sauvajot RM, Fuller TK, York EC, Kamradt DA, Bromley C, et al. Effects of urbanization and habitat fragmentation on bobcats and coyotes in southern California. Conserv Biol. 2003; 17: 566–576. https://doi.org/10.1046/j.1523-1739.2003.01458.x

79. Riley SPD, Busteed GT, Kats LB, Vandergon TL, Lee LFS, Dagit RG, et al. Effects of urbanization on the distribution and abundance of amphibians and invasive species in southern California streams. Conserv Biol. 2005; 19: 1894–1907. https://doi.org/10.1111/j.1523-1739.2005.00295.x

80. Weiss KCB, Green AM, Herrera DJ, Hubbard TM, Rega-Brodsky CC, Allen ML. Effect of species-level trait variation on urban exploitation in mammals. Ecology. 2023; 104: e4055. https://doi.org/10.1002/ecy.4055 PMID: 37074821

81. Cooper DS, Yeh PJ, Blumstein DT. Tolerance and avoidance of urban cover in a southern California suburban raptor community over five decades. Urban Ecosyst. 2021; 24: 291–300. https://doi.org/10.1007/s11252-020-01035-w

82. Tobias JA, Sheard C, Pigot AL, Devenish AJM, Yang J, Sayol F, et al. AVONET: morphological, ecological and geographical data for all birds. Ecol Lett. 2022; 25: 581–597. https://doi.org/10.1111/ele.13898 PMID: 35199922

83. Herberstein ME, McLean DJ, Lowe E, Wolff JO, Khan MK, Smith K, et al. AnimalTraits—a curated animal trait database for body mass, metabolic rate and brain size. Sci Data. 2022; 9: 265. https://doi.org/10.1038/s41597-022-01364-9 PMID: 35654905

84. Soria CD, Pacifici M, Di Marco M, Stephen SM, Rondinini C. COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. Ecology. 2021; 102: e03344. https://doi.org/10.1002/ecy.3344 PMID: 33742448

85. Lewthwaite JMM, Baiotto TM, Brown BV, Cheung YY, Baker AJ, Lehnen C, et al. Drivers of arthropod biodiversity in an urban ecosystem. Sci Rep. 2024; 14: 390. https://doi.org/10.1038/s41598-023-50675-3 PMID: 38172148

86. Staubus WJ, Bird S, Meadors S, Meyer WM. Distributions of invasive arthropods across heterogeneous urban landscapes in southern California: aridity as a key component of ecological resistance. Insects. 2019; 10: 29. https://doi.org/10.3390/insects10010029 PMID: 30650585

87. Sanders D, Frago E, Kehoe R, Patterson C, Gaston KJ. A meta-analysis of biological impacts of artificial light at night. Nat Ecol Evol. 2021; 5: 74–81. https://doi.org/10.1038/s41559-020-01322-x PMID: 33139919

88. Kunc HP, Schmidt R. The effects of anthropogenic noise on animals: a meta-analysis. Biol Lett. 2019; 15: 20190649. https://doi.org/10.1098/rsbl.2019.0649 PMID: 31744413

89. Morelli F, Tryjanowski P, Ibáñez-Álamo JD, Díaz M, Suhonen J, Pape Møller A, et al. Effects of light and noise pollution on avian communities of European cities are correlated with the species' diet. Sci Rep. 2023; 13: 4361. https://doi.org/10.1038/s41598-023-31337-w PMID: 36928766

90. Arévalo C, Amaya-Espinel JD, Henríquez C, Ibarra JT, Bonacic C. Urban noise and surrounding city morphology influence green space occupancy by native birds in a Mediterranean-type South American metropolis. Sci Rep. 2022; 12: 4471. https://doi.org/10.1038/s41598-022-08654-7 PMID: 35296770

91. Slabbekoorn H, McGee J, Walsh EJ. Effects of man-made sound on terrestrial mammals. In: Slabbekoorn H, Dooling RJ, Popper AN, Fay RR, editors. Effects of Anthropogenic Noise on Animals. New York, NY: Springer; 2018. pp. 243–276. https://doi.org/10.1007/978-1-4939-8574-6_9

92. Desouhant E, Gomes E, Mondy N, Amat I. Mechanistic, ecological, and evolutionary consequences of artificial light at night for insects: review and prospective. Entomol Exp Appl. 2019; 167: 37–58. https://doi.org/10.1111/eea.12754

93. Morley EL, Jones G, Radford AN. The importance of invertebrates when considering the impacts of anthropogenic noise. Proc R Soc B Biol Sci. 2014; 281: 20132683. https://doi.org/10.1098/rspb.2013.2683 PMID: 24335986

94. Stone EL, Harris S, Jones G. Impacts of artificial lighting on bats: a review of challenges and solutions. Mamm Biol. 2015; 80: 213–219. https://doi.org/10.1016/j.mambio.2015.02.004

95. Hochmair HH, Scheffrahn RH, Basille M, Boone M. Evaluating the data quality of iNaturalist termite records. Barden P, editor. PLoS One. 2020; 15: e0226534. https://doi.org/10.1371/journal.pone.0226534 PMID: 32365126

96. Mesaglio T, Callaghan CT. An overview of the history, current contributions and future outlook of iNaturalist in Australia. Wildl Res. 2021; 48: 289–303. https://doi.org/10.1071/WR20154

97. Chesshire PR, Fischer EE, Dowdy NJ, Griswold TL, Hughes AC, Orr MC, et al. Completeness analysis for over 3000 United States bee species identifies persistent data gap. Ecography. 2023; 5. https://doi.org/10.1111/ecog.06584

98. Ward DF. Understanding sampling and taxonomic biases recorded by citizen scientists. J Insect Conserv. 2014; 18: 753–756. https://doi.org/10.1007/s10841-014-9676-y

99. Callaghan CT, Poore AGB, Hofmann M, Roberts CJ, Pereira HM. Large-bodied birds are over-represented in unstructured citizen science data. Sci Rep. 2021; 11: 19073. https://doi.org/10.1038/s41598-021-98584-7 PMID: 34561517

100. Dimson M, Gillespie TW. Who, where, when: observer behavior influences spatial and temporal patterns of iNaturalist participation. Appl Geogr. 2023; 153: 102916. https://doi.org/10.1016/j.apgeog.2023.102916

101. Pocock MJO, Chandler M, Bonney R, Thornhill I, Albin A, August T, et al. A vision for global biodiversity monitoring with citizen science. In: Bohan DA, Dumbrell AJ, Woodward G, Jackson M, editors. Next Generation Biomonitoring: Part 2. Academic Press; 2018. pp. 169–223. https://doi.org/10.1016/bs.aecr.2018.06.003

102. Brown BV, Hartop EA. Big data from tiny flies: patterns revealed from over 42,000 phorid flies (Insecta: Diptera: Phoridae) collected over one year in Los Angeles, California, USA. Urban Ecosyst. 2017; 20: 521–534. https://doi.org/10.1007/s11252-016-0612-7

103. West S, Pateman R. Recruiting and retaining participants in citizen science: what can be learned from the volunteering literature? Citiz Sci. 2016; 1: 15. https://doi.org/10.5334/cstp.8

104. Biodiversity Indicator Species: a guide to the City of Los Angeles' charismatic umbrella species. Los Angeles Sanitation and Environment; 2022. Available: https://www.lacitysan.org/cs/groups/public/documents/document/y250/mdc1/~edisp/cnt075161.pdf

105. Kittelberger KD, Hendrix SV, Şekercioğlu ÇH. The value of citizen science in increasing our knowledge of under-sampled biodiversity: an overview of public documentation of Auchenorrhyncha and the hoppers of North Carolina. Front Environ Sci. 2021; 9: 710396. https://doi.org/10.3389/fenvs.2021.710396

106. Shirey V, Belitz MW, Barve V, Guralnick R. A complete inventory of North American butterfly occurrence data: narrowing data gaps, but increasing bias. Ecography. 2021; 44: 537–547. https://doi.org/10.1111/ecog.05396