UNIVERSITY OF CALIFORNIA

Los Angeles

Expanding Genomics Toolkits to Facilitate the Conservation and Recovery of Two Protected Species, the Tricolored Blackbird and Burrowing Owl

A dissertation submitting in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Biology

by

Kelly Barr

2023

ABSTRACT OF THE DISSERTATION

Expanding Genomics Toolkits to Facilitate the Conservation and Recovery of Two Protected Species, the Tricolored Blackbird and Burrowing Owl

by

Kelly Barr

Doctor of Philosophy in Ecology and Evolutionary Biology

University of California, Los Angeles, 2021

Professor Thomas Bates Smith, Chair

Species across the globe are faced with unprecedented pressures due to human impacts. In what is now recognized as the Sixth Mass Extinction, habitat loss and climate change have driven many populations to and over the brink of extinction. Drawing from recent advances in genomic sequencing and analyses, I aimed to facilitate the conservation and recovery of two declining species, the Tricolored Blackbird (*Agelaius tricolor*) and the Western Burrowing Owl (*Athene cunicularius hypugaea*). I analyzed population structure and genetic diversity in both species, the former with both reduced representation and whole genome sequencing and the latter with low coverage, whole genome data. For the Tricolored Blackbird, I found no indications of

population structure in the species, an important finding for conservation as this indicates there is neither a need for individual management units for the species nor a concern about barriers to gene flow. Burrowing Owls are more complicated in having both resident and migratory breeding groups. I found genetic structure was linked to the migratory phenotype, with residents being highly structured and migrants having no structure. Furthermore, I found indications of high inbreeding in resident breeding populations. Using numerous metrics of environmental variation, I found significant genetic differences between residents and migrants. Among those regions of the genome most differentiated and correlating with environmental differences, genes associated with the processing of fats, or lipophagy, were found to be significantly more represented. Revisiting Tricolored Blackbirds with whole genome data, I found indications that perhaps the population is not in complete panmixia as previously indicated by reduced representation data analyses, but that any restrictions to gene flow are indeed low. Further, the higher resolution dataset detected that the recent known decline of the species is apparent in changes in effective population size over time. My work provided essential information for conservation efforts for both species and should be a foundation for continuing genomics research on these and other species of conservation concern as well.

The dissertation of Kelly Barr is approved by the following committee:

Thomas Bates Smith, Committee Chair

Kristen Ruegg

Kirk Edward Lohmuller

Michael Edward Alfaro

University of California, Los Angeles

2023

Table of Contents

Acknowledgements

I am grateful to my committee for their thoughts and help throughout my experience at UCLA. Drs. Tom Smith and Kristen Ruegg were patient and yet persistent mentors. I long admired Tom as a scientist from afar, and I know my career will arc higher because my path crossed his. Kristen, meanwhile, was someone I knew personally for years, and my admiration for her as a person and a scientist was the inciting incident I needed to make the commitment to pursue a PhD. I am lucky to have met her.

The science presented here took a village. Perhaps the most important members of that village were Winnie Le and Jasmine Rajbhandary for pushing through much of the considerable wetlab work needed for this work and assisting with fieldwork. I dragged Jasmine clear across the country to sample Burrowing Owls and Tricolored Blackbirds in Tom's 30-year-old Toyota 4-Runner. We even finished a day of fieldwork after finding ourselves stranded on the side of literally the loneliest road in California, hitchhiking to the nearest town 20 minutes away, getting a tow truck to give us a jump, and buying and getting a new alternator installed. After all of that, we still caught and sampled 20 Tricolored Blackbirds. What a day, Jasmine!

Two other key members of that village were Christen Bossu and Rachael Bay. I came to graduate school as an aged old student to learn bioinformatics skills, particularly in UNIX and R. While much of the learning came from toiling on my own, I would be toiling still if not for the kind and patient help of Christen and Rachael along the way. Those two were also always available as sounding boards for scientific ideas and really acted as unattributed committee members. I would happily work with them the rest of my scientific career if I could.

Two other people who were always available for scientific chats via email, phone, or in person were Robert Meese and David Johnson. When you start to work on a species, you should

speak to the experts on those species. For the Tricolored Blackbird and Burrowing Owl, respectively, no species could have better and more generous stewards than Bob and David. They are both working admirably and diligently towards the conservation and recovery of their respective species, and I appreciate the invaluable knowledge they provided to me over the years. I learned more about both species from these venerable gentlemen than any literature review could have ever revealed. And I appreciated the samples they provided too!

Most importantly, I thank my family. During my time at UCLA, I had an incredible wedding with my wife, Valine, witnessed the births of my son, Jasper, and daughter, Lucía, experienced the covid lockdown, and moved across the country twice for my wife's career. My beautiful and brilliant wife Valine has been such an amazing supporter in every way she could, from listening to my frustrations about school and science to taking the kids on daily adventures away from home during this last summer as I wrapped up. The pride I feel in my evolution as a scientist over the past seven years is exceeded by that in my transition into being a parent and husband. My kids rule my life and I would have it no other way. Completing my PhD was just as much for them as it was for me.

Chapter 1 was originally published in Evolutionary Applications.

Barr, Kelly, Annabel C. Beichman, Pooneh Kalhori, Jasmine Rajbhandary, Rachael A. Bay, Kristen Ruegg, and Thomas B. Smith. "Persistent panmixia despite extreme habitat loss and population decline in the threatened tricolored blackbird (*Agelaius tricolor*)." *Evolutionary applications* 14, no. 3 (2021): 674-684. This work is open access and may be reproduced with this citation.

KRB, ACB, PK, RAB, KR, and TBS helped design the research and write the paper. RAB designed the bioinformatic pipeline for processing raw data, and both assembled and assessed the quality of the *de novo* genome. KRB and JR performed fieldwork and labwork. ACB and PK ran SFS inferences. KRB executed all remaining analyses. KRB was supported by the California Conservation Genomics Consortium, the Audubon Society of Pasadena, and the La Kretz Center of UCLA for this project.

Chapter 2 is accepted for publication in Evolutionary Applications.

Barr, Kelly, Christen M. Bossu, Rachael A. Bay, Eric C. Anderson, Jim Belthoff, Lynne A. Trulio, Debra Chromzcak, Colleen Wisinski, Thomas B. Smith, and Kristen C. Ruegg. "Genetic and environmental drivers of migratory behavior in western burrowing owls and implications for conservation and management"

KRB, CMB, RAB, JB, LAT, TBS, and KCR helped design the research and write the paper. CMB and RAB designed the bioinformatic pipeline for processing raw data, and the latter both assembled and assessed the quality of the *de novo* genome. KRB, JB, DC, and CW performed fieldwork. KRB performed labwork. ECA created the 'SRS_stuff' r package. KRB executed all analyses. The project was supported by a California Energy Commission grant to

KCR and TBS, and National Geographic, NSF, and Santa Clara Valley Habitat Agency grants to KCR.

Chapter 3 is in preparation for submission.

Barr, Kelly, Thomas B. Smith. "Consistent Patterns of Panmixia in California's Near-Endemic Tricolored Blackbird in Multiple Genetic Datasets."

KRB and TBS designed the research and wrote the paper. KRB executed all analyses. KRB was supported by the California Conservation Genomics Program for this project.

Curriculum Vitae

Kelly Barr

Education:

2007 Master of Science in Biology. University of Louisiana. Lafayette, LA.

2001 Bachelor of Science in Public Health, Major in Environmental Sciences-Biology Concentration, Minor in Marine Sciences. University of North Carolina. Chapel Hill, NC.

Work Experience:

2022-now Lecturer: Applied Environmental Sciences and Biology & Chemistry Departments, California State University, Monterey Bay, Marina, CA.

2022-now Life Sciences Technician: Physiological Ecology Lab, Hopkins Marine Station, Stanford University, Pacific Grove, CA. PI: Jeremy Goldbogen.

2010-2014 Geneticist: USGS, San Diego, CA. PI: Amy Vandergast.

Publications:

2023 Ballare, Kimberly M., Merly Escalona, Kelly Barr, William Seligmann, Samuel Sacco, Ruta Madhusudan Sahasrabudhe, Oanh Nguyen, Christy Wyckoff, Thomas B. Smith, and Beth Shapiro. "A reference genome assembly of the declining tricolored blackbird, *Agelaius tricolor*." *Journal of Heredity*.

2021 Barr, Kelly, Annabel C. Beichman, Pooneh Kalhori, Jasmine Rajbhandary, Rachael A. Bay, Kristen Ruegg, Thomas B. Smith. "Persistent panmixia despite extreme habitat loss and population decline in the threatened tricolored blackbird (*Agelaius tricolor*)." *Evolutionary Applications*.

2019 Vandergast, Amy G., Barbara E. Kus, Kristine L. Preston, Kelly R. Barr. "Distinguishing recent dispersal from historical genetic connectivity in the coastal California gnatcatcher." *Scientific Reports*.

2018 Cade, David E., Kelly R. Barr, John Calambokidis, Ari S. Friedlaender, Jeremy A. Goldbogen. "Determining forward speed from accelerometer jiggle in aquatic environments." *Journal of Experimental Biology*.

2015 Barr, Kelly R., Barbara Kus, Kristine Preston, Scarlett Howell, Emily Perkins, Amy Vandergast. "Habitat fragmentation in coastal southern California disrupts genetic connectivity in the Cactus Wren (*Campylorhynchus brunneicapillus*)." *Molecular Ecology*.

2013 Richmond, Jonathan Q., Kelly R. Barr, Adam R. Backlin, Amy G. Vandergast, Robert N. Fisher. "Evolutionary dynamics of a rapidly receding southern range boundary in the threatened California Red-Legged Frog." *Evolutionary Applications*.

2012 Athrey, Giri, Kelly R. Barr, Richard F. Lance, Paul L. Leberg. "Birds in space and time: genetic changes accompanying anthropogenic habitat fragmentation in the endangered black-capped vireo (*Vireo atricapilla*)." *Evolutionary Applications*.

2011 Barr, Kelly R., Denise L. Lindsay, Giri Athrey, Scott A. Tweddale, Timothy J. Hayden, Richard F. Lance, Paul L. Leberg. "Missing the forest for the gene trees: conservation genetics is more than the identification of distinct population segments." *The Auk*.

2010 Leberg, Paul L., Giri Athrey, Kelly R. Barr, Denise L. Lindsay, and Richard F. Lance. "Implications of landscape alteration for the conservation of genetic diversity of endangered species." In *Molecular Insights Into Natural Resource Conservation and Management* (eds, J.A. DeWoody, J.W. Bickham, C. Michler, K. Nichols, O.E. Rhodes, and K. Woeste). Cambridge University Press.

2008 Barr, Kelly R., Denise L. Lindsay, Giri Athrey, Scott A. Tweddale, Timothy J. Hayden, Richard F. Lance, Paul L. Leberg. "Population structure in an endangered, migratory songbird: high differentiation despite significant population recovery and high vagility." *Molecular Ecology*.

2008 Lindsay, Denise L., Kelly R. Barr, Richard F. Lance, Scott A. Tweddale, Timothy J. Hayden, and Paul L. Leberg. "Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (*Dendroica chrysoparia*)." *Molecular Ecology*.

2007 Marko, Peter B. and Kelly R. Barr. "Basin-scale patterns of mtDNA differentiation and gene flow in the Bay Scallop, *Argopecten irradians concentricus* Say." *Marine Ecology Progress Series*.

2004 Collin, Rachel, Marjorie J. Wonham, and Kelly R. Barr. "*Crepidula convexa* Say 1822 (Caenogastropoda: Calyptraeidae) in Washington state, USA." *American Malacological Bulletin*.

Chapter I

Title: Persistent Panmixia Despite Extreme Habitat Loss and Population Decline in the Threatened Tricolored Blackbird (*Agelaius tricolor*)

Running Title: Tricolored Blackbird Genomics

Authors: Kelly Barr[1,2], Annabel C Beichman[2], Pooneh Kalhori[2], Jasmine Rajbhandary[2], Rachael A Bay[3], Kristen Ruegg[4], Thomas B Smith[1,2]

Contact Information:

[1]Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA

[2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA

[3]Department of Evolution and Ecology, University of California, Davis, Davis, CA

[4]Department of Biology, Colorado State University, Fort Collins, CO

Abstract: Habitat loss and alteration has driven many species into decline, often to the point of requiring protection and intervention to avert extinction. Genomic data provide the opportunity to inform conservation and recovery efforts with details about vital evolutionary processes with a resolution far beyond that of traditional genetic approaches. The tricolored blackbird (*Agelaius tricolor*) has suffered severe losses during the previous century largely due to anthropogenic impacts on their habitat. Using a dataset composed of a whole genome paired with reduced representation libraries (RAD-Seq) from samples collected across the species' range, we find evidence for panmixia using multiple methods, including PCA (no geographic clustering), admixture analyses (ADMIXTURE and TESS conclude K = 1), and comparisons of genetic differentiation (average FST = 0.029). Demographic modeling approaches recovered an ancient decline that had a strong impact on genetic diversity but did not detect any effect from the known recent decline. We also did not detect any evidence for selection, and hence adaptive variation, at any site, either geographic or genomic. These results indicate that species continues to have high vagility across its range despite population decline and habitat loss and should be managed as a single unit.

1 | INTRODUCTION

Rising anthropogenic pressures over the past century have created a global biodiversity crisis (Ceballos et al., 2015). Countless species have experienced significant population decline due to habitat reduction and alteration in the course of human activities, and many are now threatened with extinction (Zalasiewicz et al., 2011). Efforts to slow and reverse these trends are often limited by a deficiency of information regarding the evolutionary processes that dictate long-term species survival (Smith & Bernatchez, 2008). Historically, attempts to fill this information gap employed genetic markers with the capacity for evaluating only coarse genetic patterns (e.g.,

microsatellites or mitochondrial DNA sequences). A primary objective of these approaches, for instance, was the identification of evolutionarily significant units (ESUs) that may encompass unique and possibly adaptive variation and hence warrant targeted protection (Moritz, 1994; Ryder, 1986). Given the increasing accessibility of genome-wide data, we can now move towards more precise evaluations of evolutionary processes by directly assessing adaptive variation (Bay et al., 2018; Funk et al., 2012; K. Ruegg et al., 2018), analyzing fine-scale gene flow patterns and hierarchical genetic structure (Hendricks et al., 2017; K. C. Ruegg et al., 2014; Younger et al., 2017), and estimating recent and historical demographic trends (Beichman et al., 2018, 2019; Oh et al., 2019).

One such species experiencing severe impacts in the course of anthropogenic activities is the tricolored blackbird (tricoloreds; *Agelaius tricolor*), a colonial songbird that is near endemic to California (Beedy et al., 2018). Tricoloreds are now listed as threatened at the state level after declining by an estimated 63% from 1935 to 1975 (Graves et al., 2013) and another 34% from 2007 to 2016 (Robinson et al., 2018). These losses are primarily due to the destruction of the species' historically preferred habitats for nesting, wetlands, and foraging, grasslands, by extensive agricultural and urban development (Beedy et al., 2018). As a consequence, whereas 93% of surveyed colonies nested in wetlands in the 1930s (Neff, 1937), tricolored colonies today use a broad range of nesting substrates, often including croplands and invasive species (Meese, 2017). Led by a multiagency collaboration of public and private interests (The Tricolored Blackbird Working Group; Kester, 2007), substantial time and financial resources have been committed towards their conservation and recovery over the past two decades. The lone genetic study guiding these efforts, Berg et al. (2010), reported no differentiation and varying levels of genetic diversity among colonies using a small suite of microsatellites and mitochondrial

sequences. The limitations of these data leave many questions about range-wide genetic connectivity and the impacts of population decline on the overall genetic diversity in the species.

Here we offer a comprehensive examination of current levels of gene flow and genetic diversity in the tricolored blackbird using genome-wide data. We sample numerous colonies breeding at the range periphery that were not covered by Berg et al (2010) and where the earliest impacts of declining population sizes and restricted gene flow is expected. Using multiple demographic modeling approaches to distinguish between recent and historical events, we assess genetic diversity at multiple temporal scales. Finally, we investigate evidence for local adaptation using outlier and genotype-environment association (GEA) analyses. Our primary objectives are to 1) assess gene flow and genetic diversity, both neutral and adaptive (the latter being the product of local environmental selection), 2) estimate current and long-term effective population sizes (Ne), and 3) provide management recommendations based upon our results.

2 | MATERIALS AND METHODS

2.1 | Genetic Sampling

We obtained tissue samples from breeding tricolored colonies throughout their range (Fig. 1A). Detailed information about sample sites, numbers of individuals, tissue types, and sources are provided in Table S1. From tissue samples, we purified DNA using DNeasy Blood and Tissue Kits (Qiagen) and assessed extract quantity using a Qubit (Thermofisher) and quality with an agarose gel. We collected genetic data through two means: 1) whole genome sequencing with deep coverage (n=1) and, 2) restriction-site associated sequencing (RAD-Seq; n=329).

2.2 | Genome Sequencing

We prepared DNA for whole genome sequencing using the Illumina TruSeq DNA PCR-Free LT kit (Illumina). After fragmenting 1 μg of DNA to 400 bp using a Diagenode sonicator and cleaning with magnetic beads at a ratio of 105 μL of beads/79 μL of water to select for >400 bp fragments, bioanalyzer traces were collected by the University of California, Los Angeles GenoSeq Core to verify library quality. We sequenced a final library with fragments averaging ~500 bp using a 250 bp paired-end run on an Illumina HiSeq2500 at the University of California QB3 Vincent J. Coates Genomics Sequencing Laboratory. Scaffolds were assembled from resulting sequence data via the Discovar DeNovo assembler (Broad Institute), and those <5,000 bp were removed. We used BUSCO (Simão et al., 2015) to estimate genome completeness by searching for single copy orthologs common to all species in the class Aves.

2.3 | Variant Discovery

We developed genomic libraries using bestRAD (Ali et al., 2016). For these, DNA was digested using the SbfI restriction enzyme (New England Biolabs, NEB), cleaned using 1X Agencourt AMPure XP beads (Beckman Coulter), ligated with biotinylated adaptors, and sheared to 400bp fragments with a Bioruptor NGS sonicator (Diogenode). We filtered out non-ligated fragments using magnetic beads (Dynabeads M-280; Life Technologies). Blunt ends were repaired and ligated with adaptors via the Illumina NEBNext Ultra DNA Library Prep Kit (NEB), and 500bp fragments were selected with AMPure beads. PCR-enrichment was tested using 5 uL of library with a maximum of 15 cycles. Based upon product brightness on an agarose gel, 15 uL of library was then amplified for an appropriate number of cycles, cleaned with AMPure beads, and verified via bioanalyzer traces at the UCLA Technology Center for Genomics and Bioinformatics.

We sequenced RAD-Seq libraries over four lanes of 100bp paired-end reads on an Illumina HiSeq2500 at the UC-Davis DNA Technologies Core, and used the 'process_radtags' function in STACKS (Catchen et al., 2013) to demultiplex, filter, trim adapters, and remove low quality reads. PCR duplicates were removed using the 'clone_filter' function. We mapped reads to the genome assembly with bowtie2 (Langmead & Salzberg, 2012) and identified single nucleotide polymorphisms (SNPs) using the Haplotype Caller module in the Genome Analysis Toolkit (McKenna et al., 2010). We removed low quality variants (genotype quality<30, depth<8, minor allele frequency<0.01), indels, and non-biallelic SNPs with vcftools (Danecek et al., 2011). To determine filtering levels for missing data, we visualized and assessed missingness using the R package 'genoscapeRtools' (Anderson, 2019).

2.4 | Population Structure

Population structure is in part a consequence of recent gene flow, and hence is indicative both of a species' natural and recently developed changes in movement and dispersal patterns. Since closely related individuals can bias signatures of population structure and genetic diversity, we used KING (Manichaikul et al. 2010) to estimate kinship and removed individuals from pairs detected to have first order relationships (kinship >0.177). We conducted principal components analyses (PCA) with the R package SNPRelate (Zheng et al., 2012) and sequentially removed visual outliers . Heterozygosity, both observed (H_O) and expected (H_E), of detected variants and the inbreeding coefficient (F_{IS}) were estimated with the STACKS POPULATIONS module. We calculated global Tajima's D that is bias-corrected for missing data and tested for significance with 1000 simulations in the R package 'r2vcftools' (Pope, 2019). Using ADMIXTURE (Alexander et al., 2009) and the spatially-explicit Bayesian clustering algorithm TESS (Caye et

al., 2016), we estimated the number of genetic clusters in the dataset and assessed individual level admixture. We calculated pairwise F_{ST} among sample groups (with $N \geq 3$) using the POPULATIONS module and tested for isolation by distance (IBD) with a Mantel test.

2.5 | Historical Demography and Effective Population Size

While tricoloreds experienced a sharp decline through the 20th century, it is possible that older events also impacted genome-wide diversity patterns. We examined the species' demographic history using multiple approaches to understand the impacts of population declines on genetic diversity. These include inferences from the folded site frequency spectrum (SFS), via *∂a∂i* (Gutenkunst et al., 2010) and fastsimcoal2 (Excoffier et al., 2013), and scenario tests using approximate Bayesian computation (ABC) as implemented in DIY-ABC v2.0 (Cornuet et al., 2014). We also estimated Ne for the current generation based upon linkage disequilibrium (Waples & Do, 2010) using the program NEESTIMATOR v2 (Do et al., 2014).

For both *∂a∂i* and fastsimcoal2, we used an SFS generated via a modification of easySFS, (https://github.com/isaacovercast/easySFS) to smooth over missing data and maximize the total number of SNPs using a hypergeometric projection of a dataset filtered to remove loci with >75% heterozygosity. Focusing on historical population size changes for a single population (see results), we compared multiple demographic models: a nested "one epoch" model with no size changes, a "two epoch" model with a single size change, and a "three epoch" model with two size changes (Fig. 3). We assumed a mutation rate (μ) of 4.6 x 10-9 (Smeds et al., 2016), generation time (g) of 2 years for both *∂a∂i* and fastsimcoal2, and a sequence length (L) of 60,429,389 bps. This L is based on sites that had at least 190X coverage (10X * the number of individuals used in the SFS projection) across a merged bamfile composed of the individuals that

passed quality filters with and with no close relatives or PCA outliers (see results). Meanwhile g is based upon a robust estimation from another Passerine (Brommer et al., 2004), and the known ages of first breeding of one year for female and two years for male tricoloreds (Beedy et al., 2018).

In *∂a∂i*, demographic parameters for each model are determined by solving an approximation to the diffusion equation (Gutenkunst et al. 2009). We set extrapolation grid points for simulations as the haploid sample size of the projected SFS, and that plus 5, 15 and 25. Using 50 independent replicates, we carried out inferences with permuted starting parameter values and assessed the fit of the expected SFS under the inferred model parameters to the empirical SFS using a multinomial log-likelihood. For each model, the maximum likelihood estimate (MLE) for the set of 50 runs was selected. The best-fit θ (population scaled mutation rate) for each replicate's parameters was inferred using *∂a∂i*, and scaled by L and μ to calculate the ancestral size in diploids (Ne_{anc}):

$$\mathrm{Ne}_{anc} = \frac{\theta}{4\mu L}$$

Population sizes inferred in *∂a∂i* were then scaled by Ne_{anc}, and times were scaled by $2*Ne_{anc}*g$. After determining the best fit model, we used a grid-search approach to refine the bounds of our inferred parameters. We examined a 100x100 grid of values of *nu* (contraction size relative to Ne_{anc}) and *T* (contraction duration in terms of $2*Ne_{anc}*g$) spaced evenly along a $\log_{10}$ scale. We obtained the expected SFS for each of the 10,000 parameter pairs using *∂a∂i* and calculated the multinomial log-likelihood. We then plotted the delta log-likelihood between each parameter pair and the MLE as a heatmap.

Fastsimcoal2 offers an alternative analytical framework for using the SFS for demographic inferences using coalescent simulations. For each of the previously described

demographic models, we ran 100,000 coalescent simulations and used 50 Expected/Conditional Maximization (ECM) parameter-optimization cycles to estimate the expected SFS for each set of parameters.

We conducted ABC analyses in a sequential, hypothesis-testing manner with at least 100,000 simulations of each scenario (Fig. S6). First, we compared contraction and expansion scenarios, then single versus multiple contractions, and finally we compared scenarios with a single contraction at four time frames: recent (Ta = 0 – 99 g ago), older (Ta = 100 – 999 g ago), historical (Ta = 1000 – 9999 g ago), and deeply historical (Ta = 10000 – 99999 g ago). For computational efficiency, we used 1,000 randomly selected loci from the real dataset of 153 individuals and simulated the same number through the scenarios. Using proportion of monomorphic loci, Nei's (1987) mean gene diversity, variance of gene diversity across polymorphic loci, and mean gene diversity across all loci as summary statistics, we assessed scenario accuracy with a PCA and estimated posterior probabilities of scenarios using both direct and logistic regression approaches. Sampling priors used for simulations are provided in Fig. S6.

Estimations of contemporary, short-term Ne using the LD method are downwardly biased by the presence of overlapping generations (Waples et al., 2014). To limit this effect, we focused estimations on adults sampled in 2017 and 2018, removing nestlings and samples collected in 2002 and 2008. We considered the 132 remaining samples to be a single population based upon the lack of genetic structure (see results). For these samples, we used PLINK2.0 (Chang et al., 2015) to prune varying numbers of SNPs to assess the effect of using reduced representation of the genome on calculations of Ne and report results with a minimum allele frequency (MAF) of 0.01.

2.6 | Testing for Selection

The presence of adaptive variants would be critical information for conservation planning. We used both outlier analyses and genotype-environment association (GEA) methods to detect loci potentially under selection. For the former, we used PCAdapt (Luu et al., 2017) to jointly estimate genetic structure and assess significantly differentiated loci. We used the R package 'qvalue' (Storey et al., 2019) to adjust p-values using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). GEAs were assessed using both redundancy analyses (RDA; Forester et al. 2018) and a machine learning approach (gradient forests; Breiman, 2001). For RDA, we used the R package 'vegan' (Oksanen et al., 2019) to conduct a permutation test for constrained correspondence for significance testing. Our gradient forest approach detects areas of genotypic transitions associated with environmental conditions. For this, we used the R package 'gradientForest' (Ellis et al., 2012) using the following parameters: ntree=100, nbin=101, corr.threshold=0.5. We ran 10 additional gradient forests with randomized environmental variables for confidence testing. Both of these GEAs were based upon 19 climate variables downloaded from WorldClim (Hijmans et al., 2005), the vegetation indices NDVI and NDVIstd for May of 2018 (Carroll et al., 2004), tree cover (Sexton et al., 2013), elevations from the Global Land Cover Facility (www.landcover.org), and surface water measurements (QuickScat; from scp.byu.edu).

3 | RESULTS

3.1 | Data Quality

We obtained 389 samples from throughout the tricolored's breeding range, including many colonies sampled along range periphery (Fig. 1A; Table S1). The genome we assembled for the

species is 1.08 Gb in total length across 70,524 scaffolds with an N50 of 103,912 and >100X coverage. Of a total 4915 known single copy orthologs in Aves, the *de novo* genome assembly includes 87.2% of these represented completely, 8.3% are fragmented, and 4.5% missing. RAD-Seq libraries were created for 329 individuals that passed DNA quality standards. After filtering individuals and loci with missing data >10% (N = 219) to maximize quantity and quality (Fig. S1) and removing close relatives (N = 6) and PCA outliers (N = 11), a dataset with 153 tricoloreds genotyped at an average of 68,366 SNPs was used for analyses except where indicated. Additional data assessments are provided in Figures S14-S16.

3.2 | Gene Flow and Genetic Diversity

We found no evidence for population structure in either PCA (Fig. 1B) or clustering analyses—both ADMIXTURE and TESS indicated K = 1—suggesting high gene flow across the species' range. This pattern is supported by similar levels of heterozygosity (H_O: 0.23 – 0.24; H_E: 0.19 – 0.22) and inbreeding (F_{IS}: -0.012 - -0.002) across sample sites (Table S1), which also suggests no individual breeding colonies are in genetic isolation. An excess of rare alleles is suggested in a bias-corrected Tajima's D was significant ($p < 0.001$) and positive (1.95; CI: 1.929 – 1.972), which is a pattern usually attributed to a population expansion after a decline. Average pairwise F_{ST} between sampled colonies is quite low at 0.029 (0.02 – 0.049) and there was no correlation with geographic distance (Fig. 2, Mantel's r = -0.11, p = 0.84). This along with a lack of significant differentiation anywhere in the examined genome (Fig. S2) further indicates that gene flow is on-going with no restrictions by either habitat fragmentation or distance.

3.3 | Historical Demography and Effective Population Size

All three demographic modeling approaches indicated a strikingly similar pattern, with the strongest event shaping the species' genetic diversity being a single contraction that is much deeper in the tricolored's evolutionary history than the known 20th century decline. Using an SFS composed of 704,884 SNPs (Figs. S3-4) from a hypergeometric projection of 153 diploid individuals down to 19, both *∂a∂i,* and fastsimcoal2 rejected the single-epoch (no size change) model in favor of a two-epoch contraction (Fig. 3; *p*-value based on likelihood ratio test with two degrees of freedom < 0.00001). Additional size changes did not significantly improve the fit to the data beyond the two-epoch model (*p*-value > 0.05). ABC results were similar. The expansion and two-contraction scenarios were sequentially rejected with high confidence (Figs. S7-8). Meanwhile, the deeply historical contraction scenario was the strongest scenario, suggesting a population decline >10,000 generations ago (Fig. S9).

Each of the SFS-based approaches also arrived at similar parameter estimates of the time since the population contraction and Ne, both Ne_{anc} and long-term (Ne_{lt}). Specific parameter estimates for the two-epoch model from the grid search in *∂a∂i* suggest ~50% population size decline occurred 21,317 (19,541 – 23,018) g/ago from an Ne_{anc} of 174,455 (173,734 – 175,249) to a Ne_{lt} of 91,315 (89,385 – 92,500; parameter ranges are within 5 log-likelihood units of the MLE; Fig. 3 & S5). The parameters inferred using fastsimcoal2 were highly concordant, exhibiting a decline at 20,144 g/ago from an Ne_{anc} of 174,617 to a Ne_{lt} of 90,941.

While these methods infer a long-term Ne_{lt} of ~91,000, the LD method suggests current Ne is much lower at ~3,100 (Fig. 4). These methods are not directly comparable as SFS-based approaches are more influenced by ancient events and the LD estimation is the product of recent genetic drift. Further, all methods are influenced by model violations in different ways. SFS-

based inferences, for instance, may be impacted by incorrect mutation rates, and the LD method may be sensitive to cryptic linkage. For instance, if we examine the impacts on *∂a∂i* results of alternative μ used by previous authors for other birds (Nadachowska-Brzyska et al., 2015), it is apparent that a lower rate assumed for a domestic pigeon (*Columba livia*; mu = 4.598e-10) results in Ne_{lt} and g since the contraction are 10X as large; meanwhile, a higher rate such as that used for the rhinoceros hornbill (*Buceros rhinoceros*; mu = 6.999e-10) results in estimations that are 50% smaller (Table S3). The μ we choose to use here from Smeds et al. (2016) seems to be the most robustly estimated one available for a fellow songbird.

3.4 | Testing for Selection

There was no evidence for strong selection across the species' genome. The scree plot of eigenvalues calculated in PCAdapt exhibited no sharp changes in proportion of explained variation, which is the pattern expected when there is no population structure or significant outlier loci (Fig. S10). Arbitrarily selecting the first four eigenvalues, no loci were detected as significant outliers after accounting for multiple tests. The GEA correlation in the RDA also found to be insignificant ($p = 0.44$). Meanwhile, gradient forest analyses revealed both average r^2 and numbers of correlated SNPs from randomized environmental datasets were higher than those detected from the empirical data (Fig. S11), suggesting any correlations detected are spurious and likely variable between runs. Despite a sampling effort aimed to capture a broad range of habitat and environmental conditions experienced by tricoloreds, there were no adaptive variants apparent in our dataset.

4 | DISCUSSION

4.1 | Gene Flow and Genetic Diversity

Our data provide multiple lines of evidence indicating that tricoloreds persist as a single, panmictic population. The low genetic differentiation we observed both among breeding colonies and across the genome is the product of high gene flow that encompasses their complete range. Breeding colonies separated by the greatest distances, often with little suitable breeding or foraging habitat in between, are no more differentiated than the closest ones. This is particularly striking given the focus on peripheral, disparate colonies in the sample design. Most of the birds breed in the core of their contemporary range in the Central Valley of California, and colonies at the periphery tend to be comparatively small. If isolated, genetic drift would be expected to rapidly cause differentiation and loss of genetic diversity in these smaller colonies. Here, however, the high genetic connectivity we observed even at the extremes of the breeding range further supports the conclusion that the species is in panmixia.

Our genome-wide analyses also illustrate higher and more far-reaching vagility by tricoloreds than previously known, though multiple studies employing bands on thousands of individuals have revealed large-scale movements over major portions of the breeding range. These include, for instance, across the Central Valley (Beedy et al., 2018; DeHaven et al., 1975; Neff, 1942), along the coast (Wilson et al., 2016), and throughout southern California (Neff, 1942). However, because movements were not observed between the Central Valley and southern California, in particular by Neff (1942), the prevailing notion was that tricoloreds should be considered two separate demes split between these areas. This is illustrated in an equivocal presentation of genetic diversity calculations for these two areas by Berg et al. (2010) even though they observed no significant genetic differentiation. High vagility and panmixia are

important conclusions for conservation efforts, as the species as a whole can be considered a single deme.

It is notable that high genetic connectivity persists across the range despite the severe population decline and habitat loss experienced by the species over the past century. This may not be surprising for a vagile, volant species; however, genetic isolation associated with habitat fragmentation over similarly narrow extents was previously reported in other songbirds (Barr et al., 2008, 2015; Lindsay et al., 2008) and species with high dispersal distances would be expected to exhibit the earliest impacts of a barrier to gene flow (Landguth et al., 2010). It is possible that the species' historical preference for nesting in seasonal wetland habitat, which is inherently ephemeral in western North America, contributes to their capacity for maintaining genetic connectivity despite severe habitat loss. Since wetlands may not develop in the same geographic locations on an annual basis, the species is likely adapted for searching over greater areas for suitable habitat (Cerame et al., 2014). Indeed, the existence of the many colonies in extreme geographic isolation we sampled for this study further suggests a broadly ranging habitat searching behavior by tricoloreds.

4.2 | Historical Demography and Effective Population Size

Considering the documented extreme decline tricoloreds experienced during the past century, it is surprising that the multiple demographic modeling approaches we employed uniformly conclude an ancient contraction ~20,000 generations ago has more significant impact on shaping genetic diversity in the species. Similar patterns are observed in other species, such as yellow-bellied toads (*Bombina variegata*; Cornetti et al., 2016) and orcas (*Orcinus orca*; Moura et al. 2014), both of which experienced known recent declines but demographic modeling illustrates

events deeper in evolutionary time are more impactful for shaping long-term genetic diversity. Though ABC methods are frequently employed for examining recent bottlenecks (eg, Cammen et al., 2018; Richmond et al., 2013; Xenikoudakis et al., 2015), our dataset may be too course for detecting the impacts of the known 20th century decline in tricoloreds given its recency and the relatively high remaining population size (2017 Nc = 177,656; Meese, 2017). It is clear from our analyses, though, that an ancient contraction occurred was highly consequential in shaping overall species genetic diversity.

While there are no clear causes of the inferred ancient decline, plausible explanations include climate change or species divergence. During high glacial periods, precipitation was high in western North America (Allen & Anderson, 1993; Oster et al., 2015) resulting in more abundant wetland habitat and likely higher tricolored population sizes. Assuming a generation time of 2 years (Brommer et al., 2004), the decline was older than the Last Glacial Maximum (20,000 y/ago); however there were numerous climatic oscillations between 20,000 – 60,000 y/ago (Petit et al., 1999) that could have resulted in significant increases or decreases in breeding habitat. As for the contraction signal being a recent species divergence, it is notable that the node between tricolored and red-winged blackbirds (*Agelaius phoeniceus*) is less well resolved than most others in an Icteridae phylogeny (Powell et al., 2014). This apparent incomplete lineage sorting is suggestive of a recent divergence between the species and possibly post-divergence hybridization. We also collected RAD-Seq libraries for ten red-wingeds (sampled next to tricolored colonies) to test for the possibility of hybridization but found no evidence for admixture between these closely related species that often share breeding habitat (Figs. S12-13). TimeTree (Kumar et al., 2017) indicates these two blackbird species diverged long before (>3Mya; based upon (Barker et al., 2015; Powell et al., 2014)) the estimated time of population

contraction detected here (~40,000 years ago). This suggests the ancient cause of population decline is more likely associated with Pleistocene climate change rather than speciation.

Both the long-term Ne (~90,000) and recent Ne (~3100) estimated here are surprisingly low given the 400,000 birds reported in a 2008 census (Kelsey, 2011) and early 1900s estimates numbering in the millions. Ne is generally smaller than Nc, and the ratio between these varies between species based upon life history characteristics (Frankham, 1995). Long-term Ne, for instance, is influenced by population fluctuations over time, with small sizes having the strongest effect (Vucetich et al., 1997). This is relevant to tricoloreds as their populations likely fluctuated throughout its evolutionary history due to interannual variation in both in prey abundance (Meese, 2013) and habitat availability. Another factor that influences long-term Ne is variance in reproductive success, with high variance reducing the ratio of Ne to Nc (Sugg & Chesser, 1994). Thus, polygyny, which is thought to be relatively high in tricoloreds (Liu, 2014), may also impact Ne in an unpredictable direction that would require additional information about the mating system to ascertain (Liu, 2015). Finally, beyond the aforementioned biological causes, there is also an analytical component that should be considered. Our current N_e calculation may be reduced by the presence of overlapping generations (Waples et al., 2014), which is likely in our dataset because adults are long-lived (~12 years) and a lack of variation in molt and plumage beyond the second year limits age assessments (Beedy et al., 2018).

4.3 | Evidence for Selection

While full whole genome sequences are indisputably better for assessing subtle genetic variation patterns, our results illustrate no major selective sweeps affecting large regions of the tricolored genome. It is notable that such evidence has been reported with less sequencing effort in other

species with much larger genomes (e.g., Hohenlohe et al. 2010; White et al. 2013). This lack of evidence for selection may not be surprising given both the low standing genetic variation apparent in our Ne estimates and the recent, widescale shifts in habitat uses from primarily wetlands to highly variable alternative nesting substrates (Beedy et al., 2018; Meese, 2017). Another possible limitation to the development of adaptive diversity is gene swamping (Lenormand, 2002) across the range by dispersers from the Central Valley, where most of the species breeds. The high gene flow we detected would likely preclude the rise of large-effect alleles around the range periphery, where variance in environmental conditions is highest (Kirkpatrick & Barton, 1997). Additional sequencing effort focused on full genomes would be helpful for analyzing adaptive variation associated with alleles of weaker effect that may develop despite gene flow (Tigano & Friesen, 2016) and examining targeted regions directly relevant to a species' long-term viability, such as MHC loci (Agudo et al., 2012).

4.4 | Conclusions and Relevance to Management

Our results illustrate the analytical power and additional information gained from reexamining a system only informed by classical genetic markers with a modern genomic approach. Berg et al. (2010) were generally inconclusive about gene flow in the species, reporting at the same time a lack of genetic differentiation but also differences in genetic diversity among sample sites. Our data allow us to conclude that tricoloreds may be managed without concern for gene flow, directed preservation of unique genetic variation, or a focus on recovery of any particular local aggregation anywhere in their range. Genetic diversity, while seemingly low overall, is homogenous across breeding colonies. These results indicate the species as a whole may justifiably be considered a single management unit.

It seems that the tricolored's natural history modulates species-wide genetic diversity, and that the current level is quite a bit lower than one might predict given their recent census population sizes. While we detected no evidence for inbreeding, whether through population-level estimations of F_{IS} (Table S1) or analyses of runs of homozygosity within individuals (data not shown), the current Ne suggests that on-going genetic monitoring should occur to supplement censuses. Moreover, given the heterozygote excess we observed at all sample sites (Table S1), tricoloreds are likely in "drift debt" (Gilroy et al., 2017) and will experience further erosion of genetic diversity as they settle into mutation-drift equilibrium.

Our dataset and sample design should be quite powerful for the analyses we report here; however, we cannot entirely discount the possibility that our reduced representation dataset may miss weak or burgeoning genetic differentiation. Future additional sequencing effort aimed at whole genomes would significantly increase our power for detecting weak genetic differentiation or selection, and allow for a finer-scaled assessment of genetic diversity by estimating genome-wide heterozygosity. The relative impacts of the ancient and recent contraction events may be further examined through alternative analytical techniques that are less sensitive to departures from model assumptions, such as using identity by descent segments (Browning & Browning, 2015). Finally, museum samples may be used to better understand the impacts of recent population decline on the genetic diversity of the species.

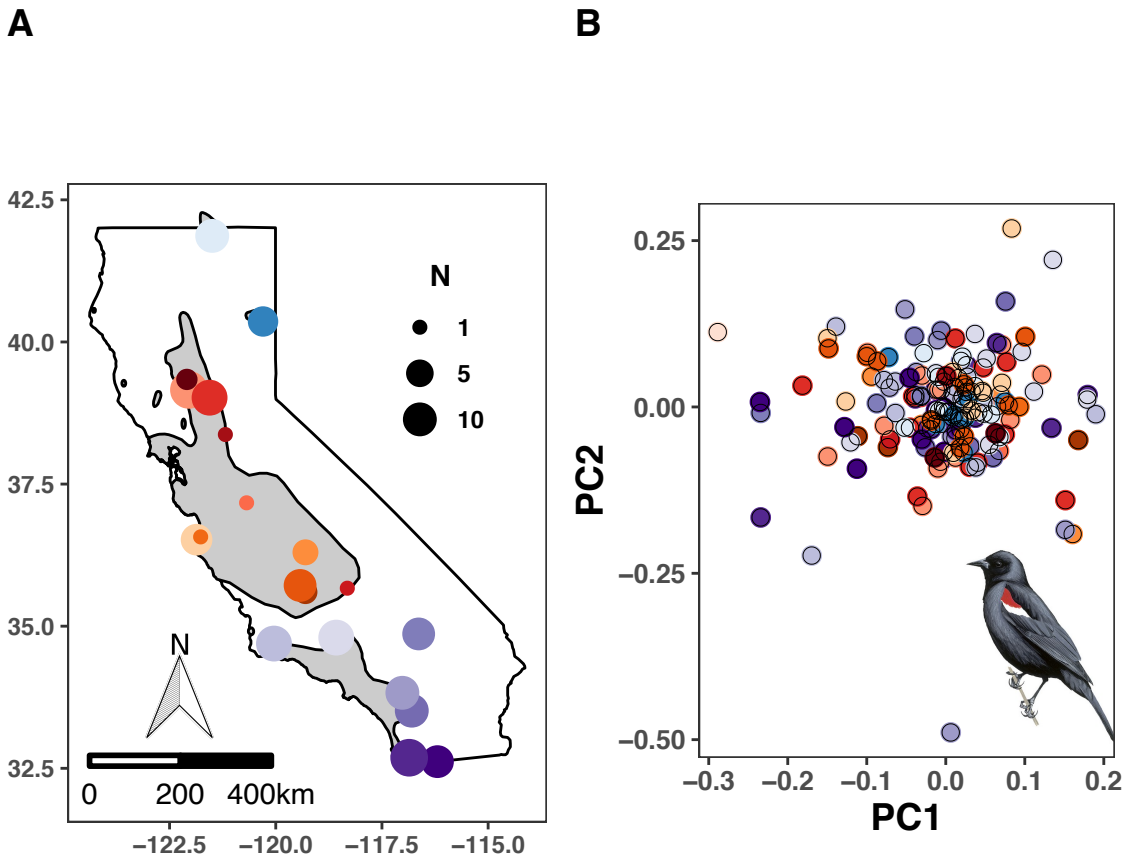


Figure 1. (A) Tricolored blackbird range (gray) and locations of breeding colonies where samples were collected (N = 153). Additional details about samples and collection are available in Table S1. (B) Plot of principal components analysis (PCA) of 153 tricoloreds genotyped at 70,933 single nucleotide polymorphisms (SNPs). Each point is an individual colored by general sample location indicated in A. Clear mixing of genotypes and very low loadings (both <1%) are consistent with panmixia across the breeding range. Warm colors represent colonies sampled in the range core and cool colors are peripheral or southern California.

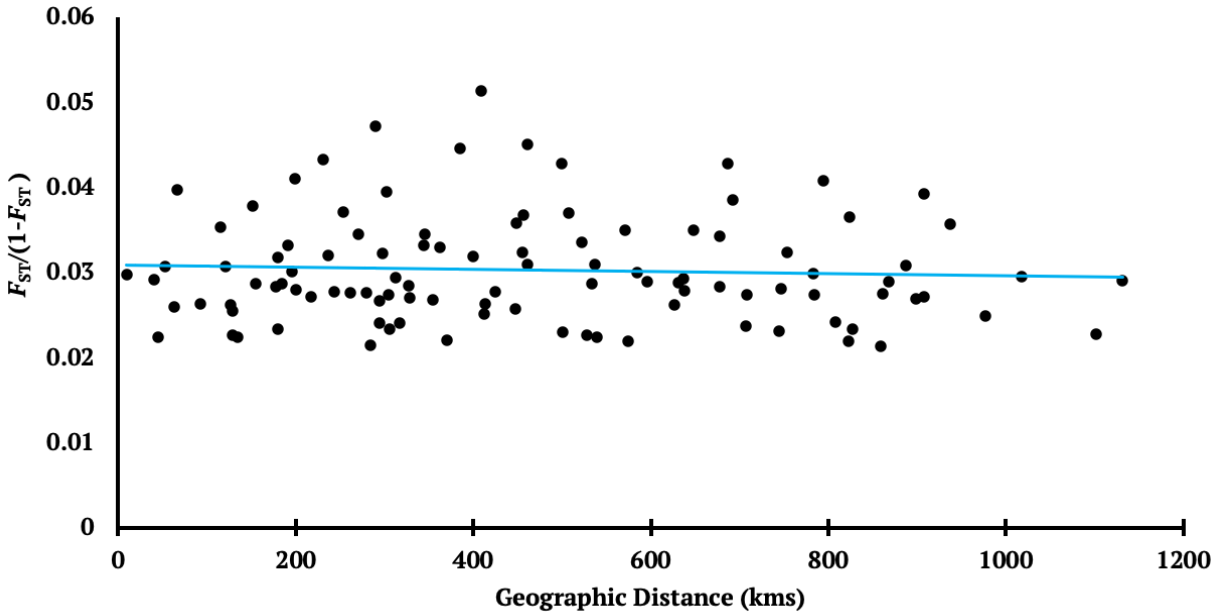


Figure 2. Plot of pairwise genetic distances (normalized F_{ST}) versus geographic distance (kms) between tricolored blackbird colonies with more than three samples. No significant relationship between differentiation and distances between colonies (Mantel's r = -0.11; p = 0.84) indicates high gene flow across the range.

MODEL		*∂a∂i*	fastsimcoal2
1 EPOCH	LL	-7,948	-2,506,605
Ne_a	Ne_a	151,228	152,490
2 EPOCHS	**LL**	**-121**	**-2,503,286**
Ne_a	**Ne_a**	**174,455**	**174,618**
Ne_{lt} T_{lt}	**T_{lt}**	**21,317**	**20,144**
	Ne_{lt}	**91,315**	**90,941**
3 EPOCHS	LL	-118	-2,503,287
Ne_{a2}	Ne_{a2}	174,932	180,143
Ne_{a1} T_{a1}	T_{a1}	24,101	428,032
Ne_{lt} T_{lt}	Ne_{a1}	101,739	172,688
	T_{lt}	2,501	21,093
	Ne_{lt}	78,023	92,648

Figure 3. Log-likelihoods (LL) and parameter estimations (including Time, T, in generations and effective population size, Ne) for the three demographic models examined using *∂a∂i* and fastsimcoal2. The "2 epoch" model was resolved to be best fitting as adding an additional change in Ne had a marginal impact on the LL. NB: The LLs are calculated differently between these two analytical frameworks and hence are not comparable.

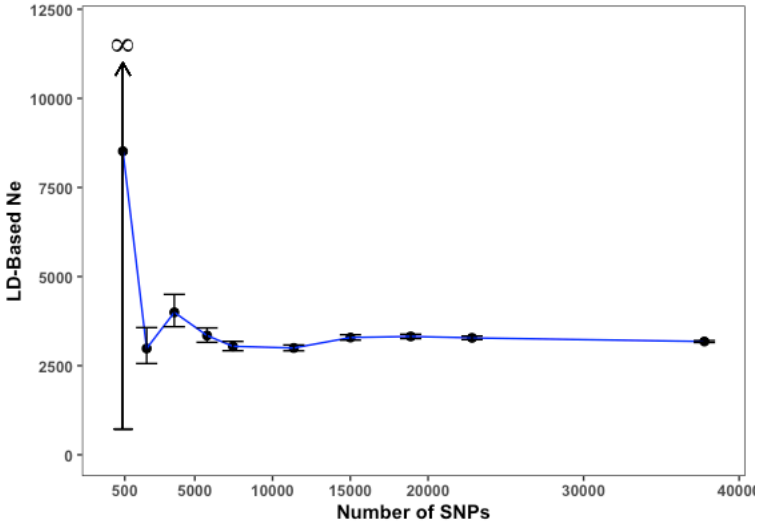


Figure 4. Plot of current effective population size (Ne) estimated from linkage disequilibrium among varying numbers of single nucleotide polymorphisms. The current Ne is ~3100.

SUPPLEMENTARY MATERIALS

Table S1. Sample site details, including locations* (map ID colors correlate with those in Figures 1A and 1B), samples included in most analyses (N), tissue types (if known), and diversity indices. These include observed and expected heterozygosities (H_O and H_E), nucleotide diversity (pi), and the inbreeding coefficient (F_{IS}).

Location	Map ID	N_{total}	N_{used}	Tissue Type	H_O	H_E	Pi	F_{IS}
Tulelake		17	11	Blood[1]	0.27	0.25	0.267	-0.010
Honey Lake		17	8	Blood[1]	0.27	0.25	0.268	-0.013
Capitol Outing Club		21	6	**	0.27	0.25	0.268	-0.010
Colusa		18	12	Blood[2]	0.27	0.26	0.267	-0.006
Plumas Lake		18	12	Blood[2]	0.27	0.26	0.268	-0.009
Carmel		4	9	Blood[3]	0.27	0.25	0.268	-0.008
Merced		6	2	**	--	--	--	--
Te Valde Ranch		6	5	**	0.27	0.24	0.269	-0.010
Pond Road		17	9	Muscle[4]	0.27	0.25	0.268	-0.008
Delevan NWR		16	2	Muscle[5]	--	--	--	--
Wilton		10	1	Muscle[5]	--	--	--	--
Costa Dairy		3	3	**	0.27	0.22	0.267	-0.006
Kern		11	1	**	--	--	--	--
Newberry Springs		25	7	Blood[1]	0.27	0.25	0.267	-0.006
Holiday Lake		23	12	Blood[1]	0.27	0.26	0.268	-0.007
Sedgewick Reserve		22	12	Blood[1]	0.28	0.26	0.270	-0.014
Ramona Farms		19	11	**	0.27	0.26	0.268	-0.012
Lake Riverside		23	9	Blood[1]	0.28	0.25	0.271	-0.019
Rancho Jamul		30	13	Blood[1]	0.28	0.26	0.269	-0.011
Jacumba		20	8	Blood[1]	0.27	0.25	0.267	-0.007

*Exact coordinates withheld at landowner request.
**Samples are from Berg et al. 2010.
[1]Samples collected using passive mistnets and via a puncture of the brachial vein (Stangel 1986; Sheldon et al. 2008) to obtain blood
[2]Samples were provided by Emily Graves, UC-Davis
[3]Samples were provided by Dr. Kristie Wychoff, Santa Lucia Conservancy
[4]Samples were provided by Jessie Bahm, USDA.
[5]Samples were provided by Dr. Robert Meese

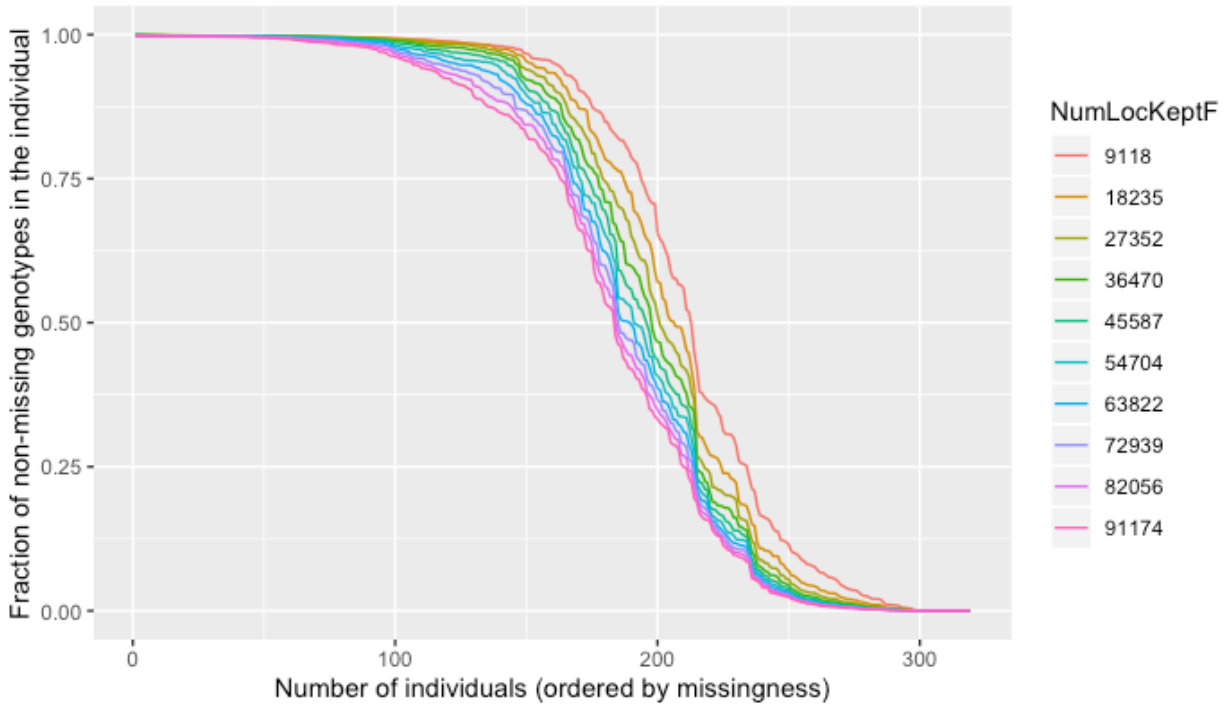


Figure S1. Levels of missingness across varying numbers of individuals. This is produced by the R package 'genoscapeRtools' (Anderson 2019).

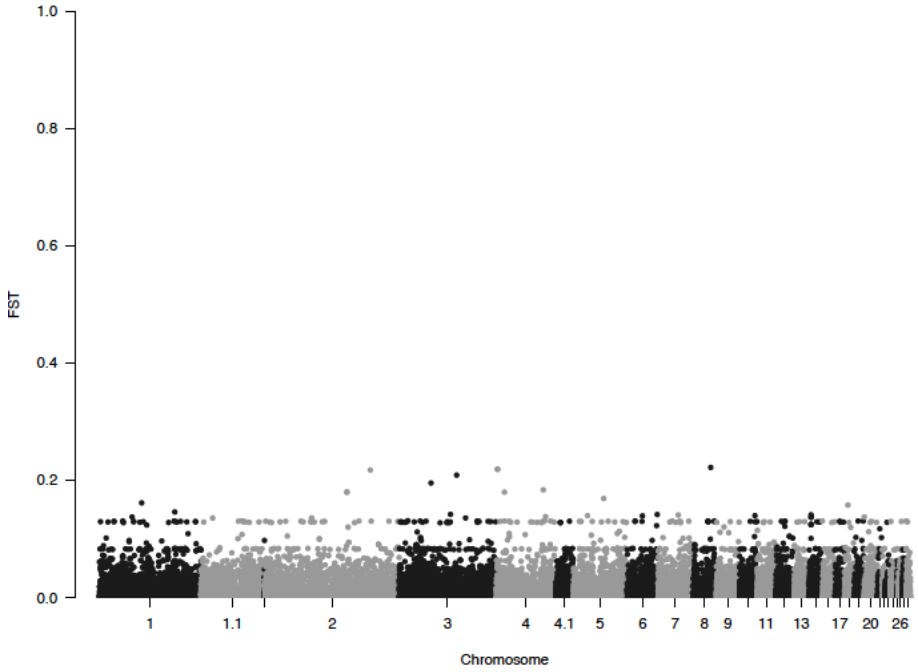


Figure S2. Differentiation (F_{ST}) across the Tricolored Blackbird genome based upon sites with three or more samples. Low genome-wide F_{ST} is indicative of panmixia in the species.

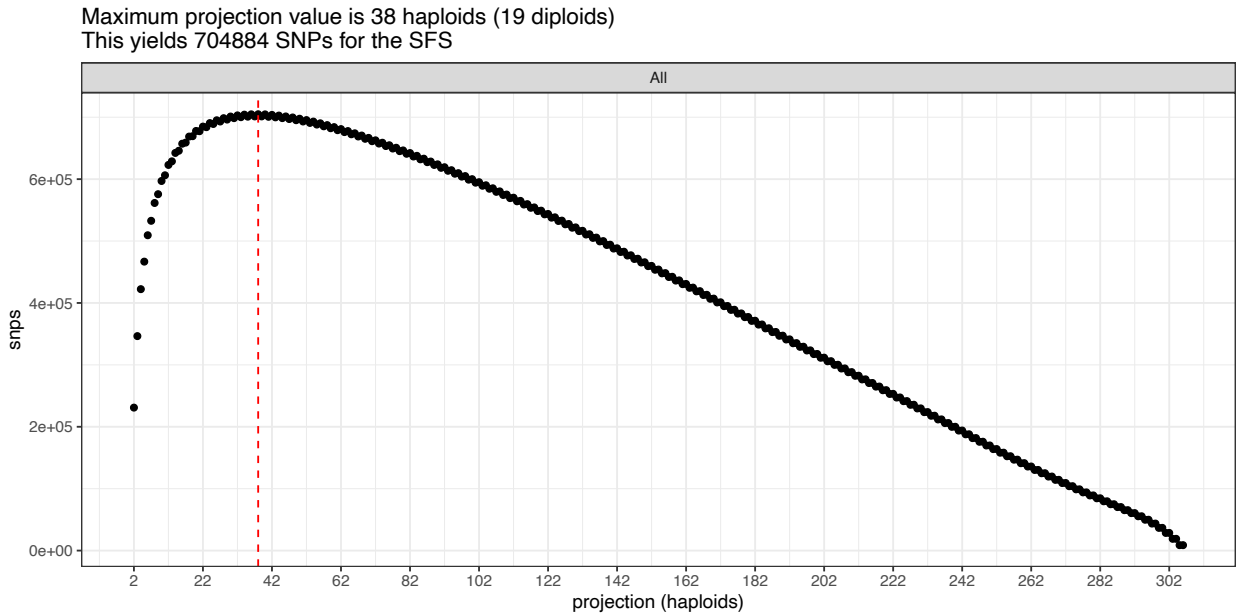


Figure S3. The projection preview from easySFS showing the number of SNPs for each possible projection value in terms of haploid individuals. The number of SNPs is maximized at 38 haploids (19 diploids), yielding 704,884 SNPs, so this value was chosen for the projection.

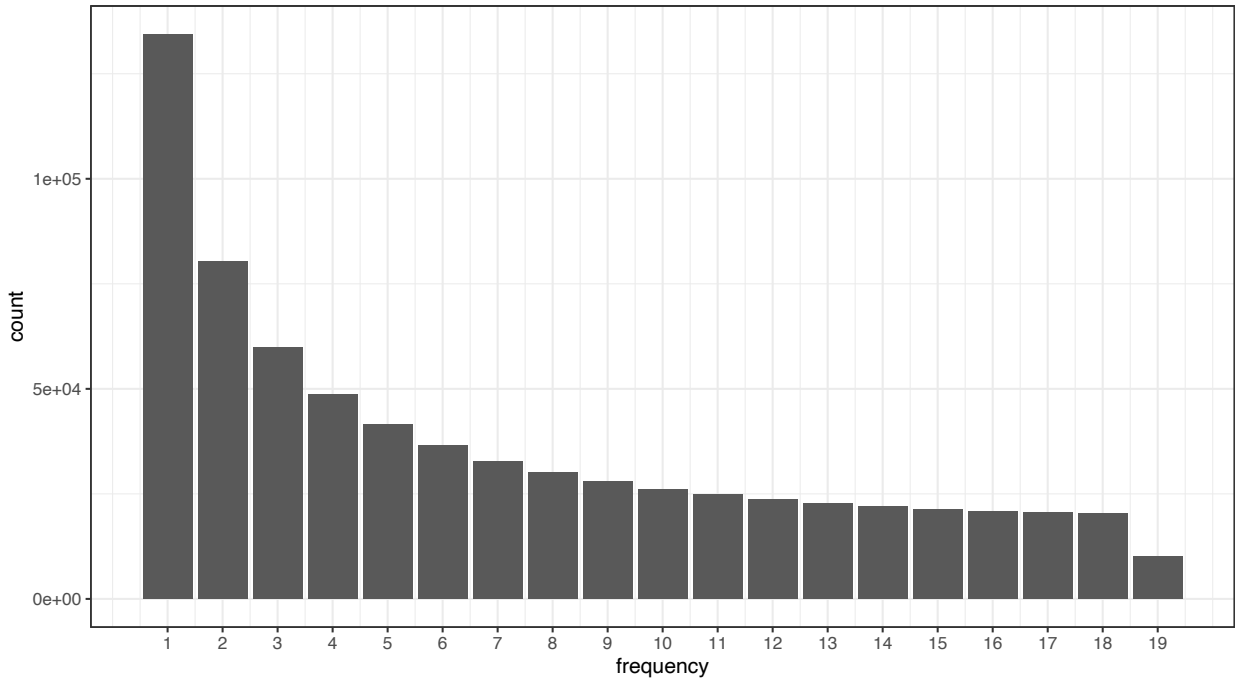


Figure S4. The projected folded site frequency spectrum, showing the number of SNPs along the y-axis that are present in the sample at each frequency (x-axis). The SFS was projected down from 153 individuals to 19 to maximize the number of SNPs and smooth over missing data.

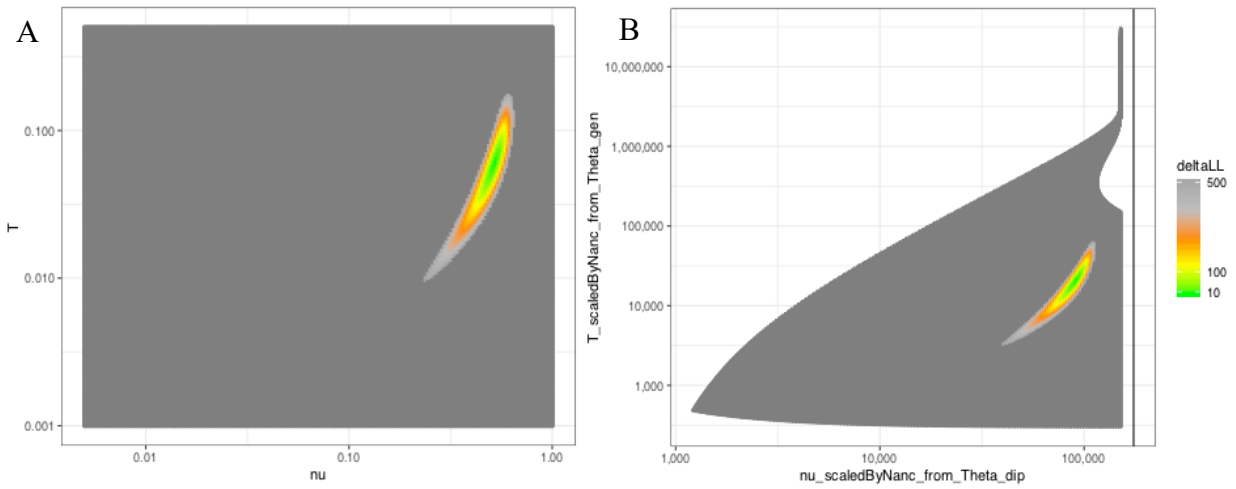


Figure S5. Heatmaps summarizing results of demographic inference in ∂a∂i. For the best fit "2 epoch" model, we examined a 100x100 range of parameters *nu* (contraction size scaled by N_{anc}) and T_{lt} (time since contraction scaled to 2*N_{anc} generations), spaced evenly along a log-scale, and colored by the delta log-likelihood (LL) between the expected site frequency spectrum (SFS) for each parameter pair and the maximum likelihood estimate (MLE). A ridge of high-likelihood estimates can be seen in green, with a correlation between contraction size and duration, corresponding to a ~50% contraction in population size ~20,000 generations ago. Long-term effective population is estimated to be ~91,000. (A) presents these results in terms of *nu* (x-axis; contraction size scaled by N_{anc}) and T_{lt} (y-axis; time since contraction scaled to 2*N_{anc} generations) and (B) shows the same results converted into units of diploid individuals (x-axis; Ne_{lt}) and generations (y-axis; g), based on the best-fit estimate of θ from ∂a∂i, and scaled using a mutation rate (μ) of 4.6 x 10-9 (Smeds et al., 2016) and a sequence length (L) of 60,429,389 bps.

P	Sampling Priors
N1	10 - 1000000
N2	10 - 1000000
t1	10 - 100000
N0	10 - 1000000
N2	1000 - 1000000
t2	2000 - 20000
N1	1000 - 1000000
t1	1 - 50
N3	1000 - 1000000
t3	2000 - 20000
N0	10 - 200000
N1	10000 - 10000000
t1	1 - 99
N2	10000 - 10000000
t2	100 - 999
N3	10000 - 10000000
t3	1000 - 9999
N4	10000 - 10000000
t4	10000 - 50000
N0	1000 - 300000

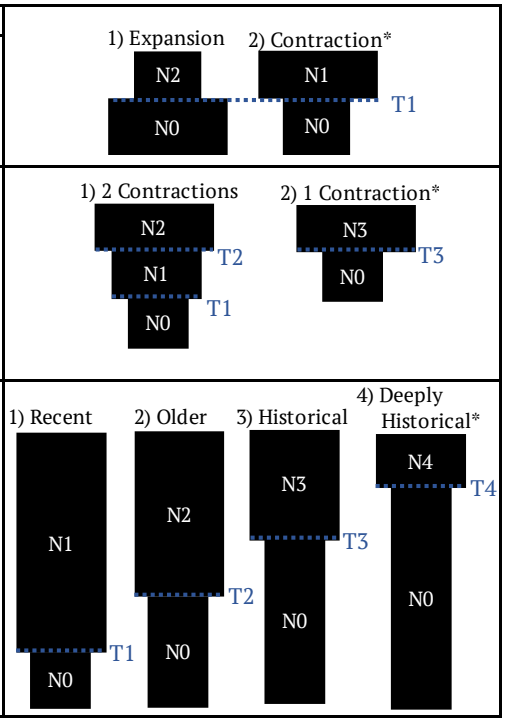
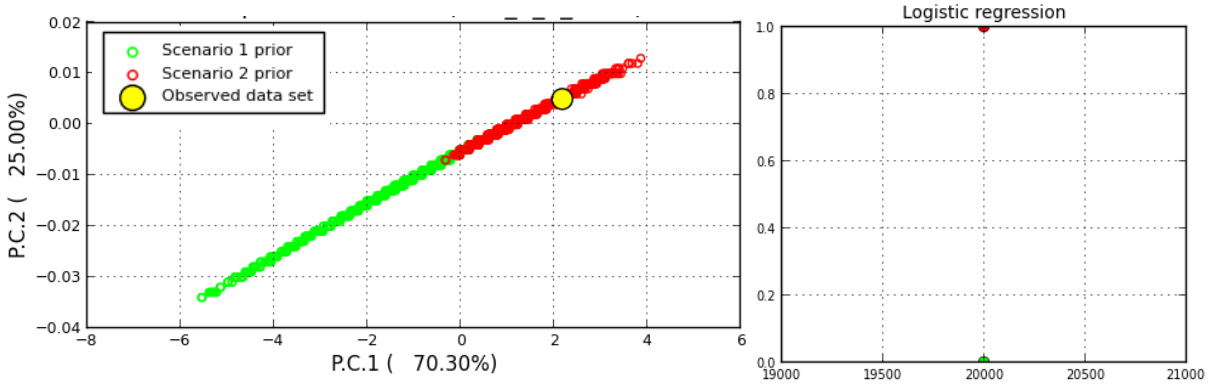
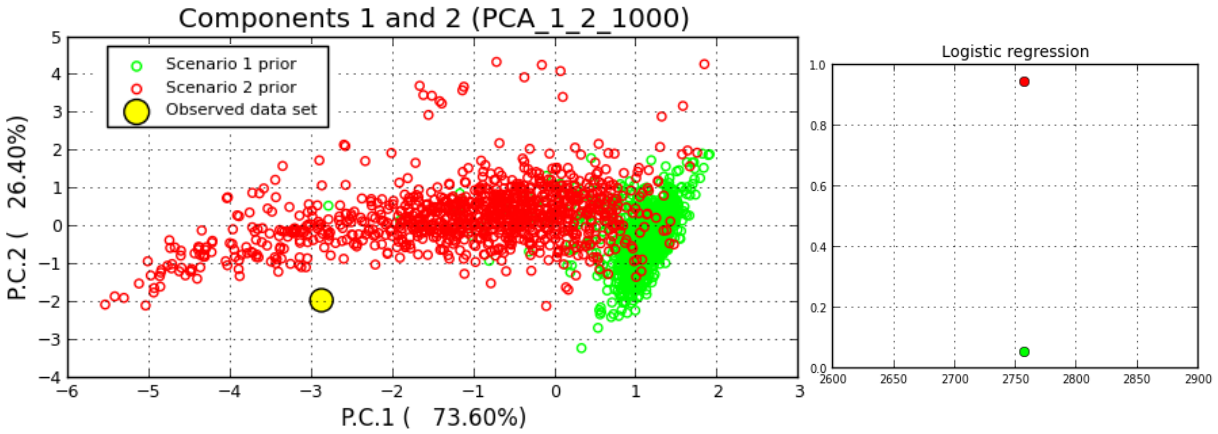


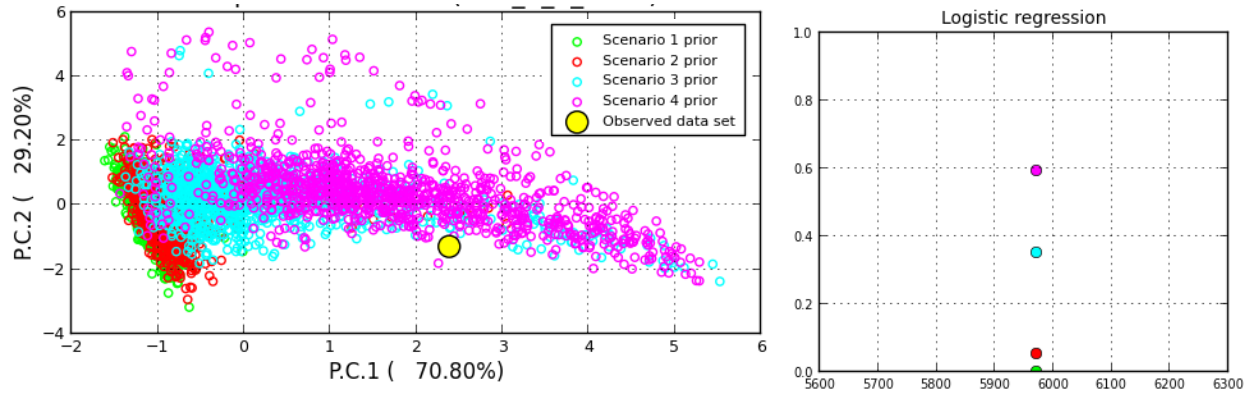
Figure S6. Approximate Bayesian Computation (ABC) parameter sampling priors on left and demographic scenarios simulated on right. Most likely scenarios as determined through direct and logistic regression are indicated with an asterisk (*).



FigureS7. Results of ABC comparison of (1) expansion and (2) contraction scenarios with PCA on left and logistic regression on right. The direct (not presented) and logistic results both suggest a contraction is the most likely scenario.



FigureS8. Results of ABC comparison of (1) two contraction and (2) one contraction scenarios with PCA on left and logistic regression on right. The direct (not presented) and logistic results both suggest a single contraction is the most likely scenario.



FigureS9. Results of ABC comparison of (1) recent (previous 1-99 generations/ago), (2) older (100 – 999 generations/ago), (3) historical (1000 – 9999 generations/ago), and (4) deeply historical (10000 – 99999 generations/ago) contraction scenarios with PCA on left and logistic regression on right. The direct (not presented) and logistic results both suggest a deeply historical contraction is the most likely scenario.

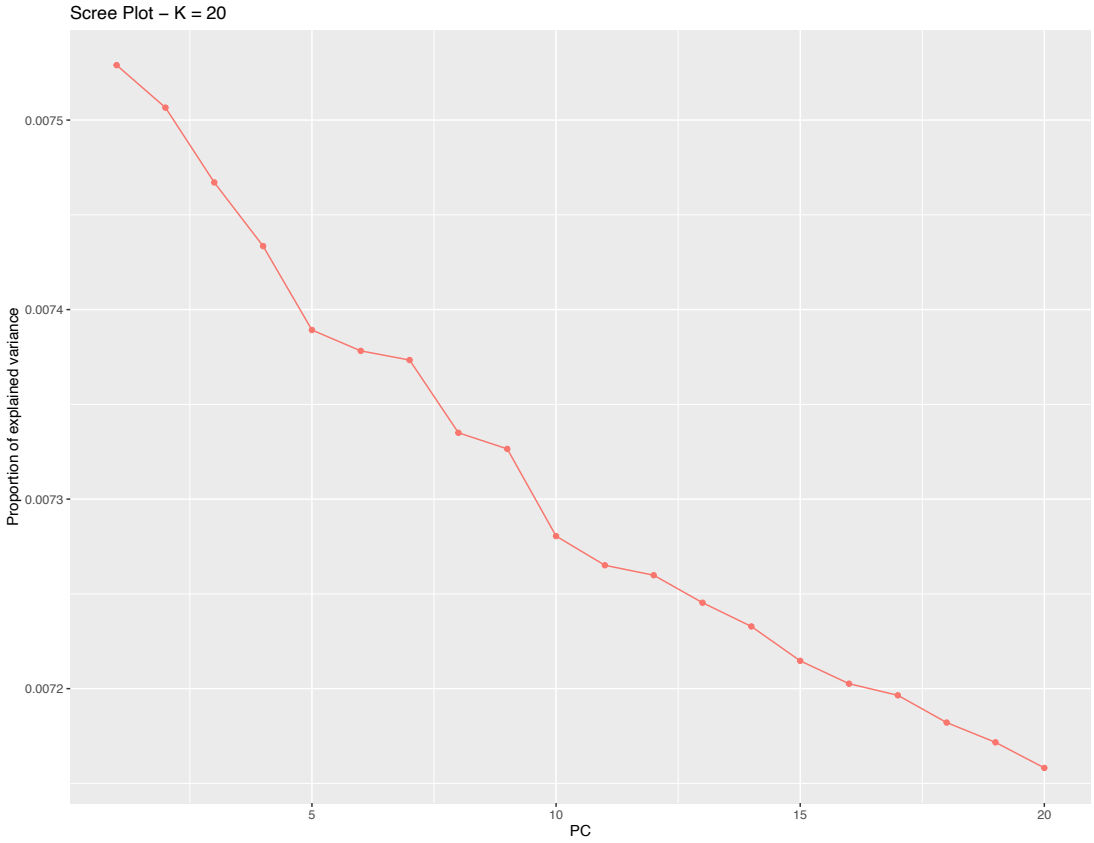


Figure S10. Scree plot of PCadapt analysis. Miniscule gains in proportion of explained variance as additional PCs are added is consistent with both a lack of genetic structure and no outlier loci.

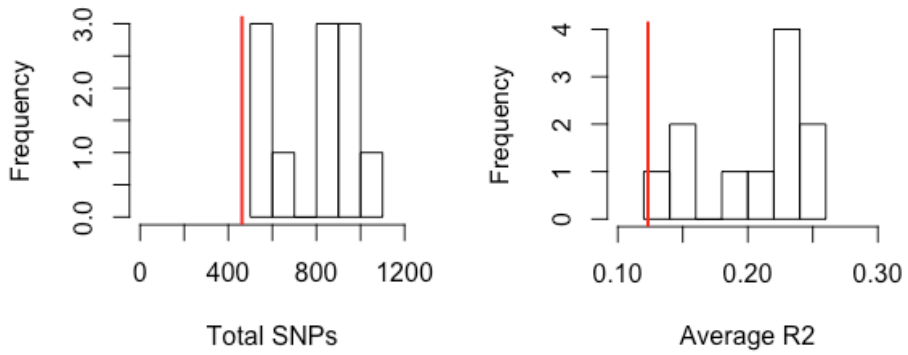


Figure S11. Results of ten randomized gradient forests compared to empirical results. Higher frequencies of total SNPs (left) and average r^2 (right) indicate the empirical results (red line) are spurious and the gradient forest model in insignificant.

Table S2. Diversity statistics after removing paralogous and high LD loci. All statistics are significantly correlated with their corollaries in table S1 (Ho: Pearson's $r = 0.99$, $p < 0.00001$; He: $r = 0.99$, $p < 0.00001$; pi: $r = 0.98$, $p < 0.00001$; Fis: $r = 0.95$, $p < 0.00001$).

Sample Site	Obs_Het	Exp_Het	Pi	Fis
Tulelake	0.261	0.249	0.262	0.002
Honey Lake	0.264	0.245	0.263	-0.002
Capitol Outing Club	0.262	0.240	0.262	0.001
Colusa	0.260	0.251	0.262	0.006
Plumas Lake	0.262	0.251	0.263	0.002
Carmel	0.261	0.246	0.262	0.003
Merced	--	--	--	--
Te Valde Ranch	0.263	0.235	0.263	0.000
Pond Road	0.261	0.247	0.262	0.003
Delevan NWR	--	--	--	--
Wilton	--	--	--	--
Costa Dairy	0.260	0.216	0.261	0.002
Kern	--	--	--	--
Newberry Springs	0.260	0.242	0.261	0.003
Holiday Lake	0.261	0.251	0.263	0.005
Sedgewick Reserve	0.266	0.252	0.264	-0.002
Ramona Farms	0.262	0.250	0.262	0.000
Lake Riverside	0.269	0.247	0.265	-0.008
Rancho Jamul	0.264	0.253	0.264	0.001
Jacumba	0.260	0.245	0.262	0.005

Table S3. Illustrating the effects of varying mutation rates on ∂a∂i inferences. The original rate is from Smeds et al. (2016) and the lower and higher rates are from Nadachowska-Brzyska et al. (2015). Parameters are as estimated in ∂a∂i for the two epoch model.

	μ	**nu**	**T**	**Nanc**	**Ncur**	**T (g)**	**theta**	**LL**	**LL_data**	**L**
Original	4.60E-09	0.5234	0.0611	174,455	91,315	21,317	193976	-120.83	-115.64	60429389
Low	4.598E-10	0.5234	0.0611	1,745,307	913,549	213,263	193976	-120.83	-115.64	60429389
High	6.999E-09	0.5234	0.0611	114,661	60,017	14,011	193976	-120.83	-115.64	60429389

Acknowledgements: We thank collaborators for providing samples, including Christy Wyckoff of the Santa Lucia Conservancy, Emilie Graves and Robert Meese of UC-Davis, and Jesse Bahm of the US Dept. of Agriculture. Our extensive sampling would have been impossible without property access approval by numerous owners and managers. We thank the Les Schwab Tire Center of Merrill, OR, for driving 30 minutes out onto the loneliest road in California to jump start our 30-year-old field vehicle. We thank Vinh Le for performing laboratory work and joining us for a day of sampling. We thank Christen Bossu for advice and bioinformatic assistance, and members of the Smith lab for comments on several manuscript draft versions. We thank Kirk E. Lohmuller for advice on demographic inferences and comments on the manuscript. Finally, we thank the California Conservation Genomics Consortium, the Audubon Society of Pasadena, and the La Kretz Center of UCLA for providing funds for this project.

DATA AVAILABILITY STATEMENT

of the genetic data collected for this study are available in public databases. These include the genome assembly in the NCBI Sequence Read Archive, accession SAMN16392922, and vcf file, both filtered and unfiltered, on Dryad, https://doi.org/10.5068/D1DM4H.

ANIMAL WELFARE AND PERMIT STATEMENTS

Samples were collected under Tom Smith's Federal Bird Banding Permit, #21901, and Kelly Barr's Scientific Collecting Permit, #SC-11568, and an MOU with California Fish and Wildlife. Animal handling and sampling protocols were conducted with the approval of UCLA's Animal Research Committee (ARC), agreement #2017-073-03.

REFERENCES

Agudo, R., Carrete, M., Alcaide, M., Rico, C., Hiraldo, F., & Donázar, J. A. (2012). Genetic diversity at neutral and adaptive loci determines individual fitness in a long-lived territorial bird. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3241–3249. https://doi.org/10.1098/rspb.2011.2606

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

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*, *202*(2), 389–400. https://doi.org/10.1534/genetics.115.183665

Allen, B. D., & Anderson, R. Y. (1993). Evidence from Western North America for Rapid Shifts in Climate During the Last Glacial Maximum. *Science*, *260*(5116), 1920–1923. https://doi.org/10.1126/science.260.5116.1920

Anderson, E. C. (2019). *GenoscapeRtools: Tools for building migratory bird genoscapes. R package version 0.1.0*.

Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and alliesNuevas Revelaciones sobre la Biogeografía del Nuevo Mundo: Una Visión Integrada desde la Filogenia de los Tordos, Cardenales, Gorriones, Tangaras, Reinitas y AliadosEmberizoid phylogeny and biogeography. *The Auk*, *132*(2), 333–348. https://doi.org/10.1642/AUK-14-110.1

Barr, K. R., Kus, B. E., Preston, K. L., Howell, S., Perkins, E., & Vandergast, A. G. (2015). Habitat fragmentation in coastal southern California disrupts genetic connectivity in the cactus wren (Campylorhynchus brunneicapillus). *Molecular Ecology*, *24*(10), 2349–2363. https://doi.org/10.1111/mec.13176

Barr, K. R., Lindsay, D. L., Athrey, G., Lance, R. F., Hayden, T. J., Tweddale, S. A., & Leberg, P. L. (2008). Population structure in an endangered songbird: Maintenance of genetic differentiation despite high vagility and significant population recovery. *Molecular Ecology*, *17*(16), 3628–3639. https://doi.org/10.1111/j.1365-294X.2008.03868.x

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86. https://doi.org/10.1126/science.aan4380

Beedy, E. C., Hamilton, III, W. J., Meese, R. J., Airola, D. A., & Pyle, P. (2018). Tricolored Blackbird (Agelaius tricolor). *The Birds of North America Online*. https://doi.org/10.2173/bna.tribla.03.1

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Beichman, A. C., Koepfli, K.-P., Li, G., Murphy, W., Dobrynin, P., Kliver, S., Tinker, M. T., Murray, M. J., Johnson, J., Lindblad-Toh, K., Karlsson, E. K., Lohmueller, K. E., & Wayne, R. K. (2019). Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter. *Molecular Biology and Evolution*, *36*(12), 2631–2655. https://doi.org/10.1093/molbev/msz101

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brommer, J. E., Gustafsson, L., Pietiäinen, H., & Merilä, J. (2004). Single-Generation Estimates of Individual Fitness as Proxies for Long-Term Genetic Contribution. *The American Naturalist*, *163*(4), 505–517. https://doi.org/10.1086/382547

Browning, S. R., & Browning, B. L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *The American Journal of Human Genetics*, *97*(3), 404–418. https://doi.org/10.1016/j.ajhg.2015.07.012

Cammen, K. M., Schultz, T. F., Bowen, W. D., Hammill, M. O., Puryear, W. B., Runstadler, J., Wenzel, F. W., Wood, S. A., & Kinnison, M. (2018). Genomic signatures of population bottleneck and recovery in Northwest Atlantic pinnipeds. *Ecology and Evolution*, *8*(13), 6599–6614. https://doi.org/10.1002/ece3.4143

Carroll, M. L., DiMiceli, C. M., Sohlberg, R. A., & Townshend, J. R. G. (2004). *250m MODIS Normalized Difference Vegetation Index, 250ndvi28920033435, Collection 4*.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. https://doi.org/10.1111/1755-0998.12471

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

Cerame, B., Cox, J. A., Brumfield, R. T., Tucker, J. W., & Taylor, S. S. (2014). Adaptation to Ephemeral Habitat May Overcome Natural Barriers and Severe Habitat Fragmentation in a Fire-Dependent Species, the Bachman's Sparrow (Peucaea aestivalis). *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0105782

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Cornetti, L., Benazzo, A., Hoban, S., Vernesi, C., & Bertorelle, G. (2016). Ancient, but not recent, population declines have had a genetic impact on alpine yellow-bellied toad populations, suggesting potential for complete recovery. *Conservation Genetics*, *17*(3), 727–743. https://doi.org/10.1007/s10592-016-0818-5

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., & Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187–1189. https://doi.org/10.1093/bioinformatics/btt763

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. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeHaven, R. W., Crase, F. T., & Woronecki, P. P. (1975). Movements of Tricolored Blackbirds Banded in the Central Valley of California, 1965-1972. *Bird-Banding*, *46*(3), 220–229. JSTOR. https://doi.org/10.2307/4512139

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLOS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

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

Gilroy, D. L., Phillips, K. P., Richardson, D. S., & Oosterhout, C. van. (2017). Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: Computer simulations see the 'ghost of selection past' and quantify the 'drift debt.' *Journal of Evolutionary Biology*, *30*(7), 1276–1287. https://doi.org/10.1111/jeb.13077

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2010). Diffusion Approximations for Demographic Inference: DaDi. *Nature Precedings*, *713*. https://doi.org/10.1038/npre.2010.4594.1

Hendricks, S., Epstein, B., Schönfeld, B., Wiench, C., Hamede, R., Jones, M., Storfer, A., & Hohenlohe, P. (2017). Conservation implications of limited genetic diversity and population structure in Tasmanian devils (Sarcophilus harrisii). *Conservation Genetics*, *18*(4), 977–982. https://doi.org/10.1007/s10592-017-0939-5

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLOS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Irene A. Liu. (2015). Conservation genetics and genetic mating system of the yellow-shouldered blackbird (Agelaius xanthomus), an endangered island endemic. *Conservation Genetics*, *16*(5), 1041–1053. https://doi.org/10.1007/s10592-015-0721-5

Kelsey, R. (2011). *Results of the Tricolored Blackbird 2008 Census* (p. 28).

Kester, S. (2007). *Tricolored Blackbird Working Group: Conservation Plan for the Tricolored Blackbird (Agelaius tricolor)*. Sustainable Conservation.

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a Species' Range. *The American Naturalist*, *150*(1), 1–23. https://doi.org/10.1086/286054

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKELVEY, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191. https://doi.org/10.1111/j.1365-294X.2010.04808.x

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Lindsay, D. L., Barr, K. R., Lance, R. F., Tweddale, S. A., Hayden, T. J., & Leberg, P. L. (2008). Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (Dendroica chrysoparia). *Molecular Ecology*, *17*(9), 2122–2133. https://doi.org/10.1111/j.1365-294X.2008.03673.x

Liu, I. A.-Y. (2014). *Molecular Causes and Consequences of Sperm Competition in Agelaius Blackbirds*. https://dukespace.lib.duke.edu/dspace/handle/10161/8795

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

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

Meese, R. J. (2017). *Results of the 2017 Tricolored Blackbird Statewide Survey.* [California Department of Fish and Wildlife, Nongame Wildlife Program Report 2017-04].

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, *9*(10), 373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Moura, A. E., Janse van Rensburg, C., Pilot, M., Tehrani, A., Best, P. B., Thornton, M., Plön, S., de Bruyn, P. J. N., Worley, K. C., Gibbs, R. A., Dahlheim, M. E., & Hoelzel, A. R. (2014). Killer Whale Nuclear Genome and mtDNA Reveal Widespread Population Bottleneck during the Last Glacial Maximum. *Molecular Biology and Evolution*, *31*(5), 1121–1131. https://doi.org/10.1093/molbev/msu058

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Neff, J. A. (1937). Nesting Distribution of the Tri-Colored Red-Wing. *The Condor*, *39*(2), 61–81. https://doi.org/10.2307/1363776

Neff, J. A. (1942). Migration of the Tricolored Red-Wing in Central California. *The Condor*, *44*(2), 45–53. https://doi.org/10.2307/1364223

Nei, M. (1987). *Molecular evolutionary genetics*. https://www.osti.gov/biblio/5943467

Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y., & Oyler-McCance, S. J. (2019). Conservation Genomics in the Sagebrush Sea: Population Divergence, Demographic History, and Local Adaptation in Sage-Grouse (Centrocercus spp.). *Genome Biology and Evolution*, *11*(7), 2023–2034. https://doi.org/10.1093/gbe/evz112

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. H. (2019). Vegan: Community Ecology Package. R package version 1.17-2. *R Package Version 2.5-4*, 2.

Oster, J. L., Ibarra, D. E., Winnick, M. J., & Maher, K. (2015). Steering of westerly storms over western North America at the Last Glacial Maximum. *Nature Geoscience*, *8*(3), 201–205. https://doi.org/10.1038/ngeo2365

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., PÉpin, L., Ritz, C., Saltzman, E., & Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, *399*(6735), 429–436. https://doi.org/10.1038/20859

Pope, N. (2019). *r2vcftools: An R interface for vcftools* (0.0.0.9000) [R].

Powell, A. F. L. A., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014). A comprehensive species-level molecular phylogeny of the New World blackbirds (Icteridae). *Molecular Phylogenetics and Evolution*, *71*, 94–112. https://doi.org/10.1016/j.ympev.2013.11.009

Richmond, J. Q., Barr, K. R., Backlin, A. R., Vandergast, A. G., & Fisher, R. N. (2013). Evolutionary dynamics of a rapidly receding southern range boundary in the threatened California Red-Legged Frog (Rana draytonii). *Evolutionary Applications*, *6*(5), 808–822. https://doi.org/10.1111/eva.12067

Robinson, O. J., Ruiz-Gutierrez, V., Fink, D., Meese, R. J., Holyoak, M., & Cooch, E. G. (2018). Using citizen science data in integrated population models to inform conservation. *Biological Conservation*, *227*, 361–368. https://doi.org/10.1016/j.biocon.2018.10.002

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., Paxton, E. H., & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096. https://doi.org/10.1111/ele.12977

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., Siegel, R. B., DeSante, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology*, *23*(23), 5726–5739. https://doi.org/10.1111/mec.12977

Ryder, OA. (1986). Species conservation and systematics: The dilemma of subspecies. *Trends Ecol. Evol.*, *1*, 9–10.

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., & Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, *6*(5), 427–448. https://doi.org/10.1080/17538947.2013.786146

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smeds, L., Qvarnström, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, *26*(9), 1211–1218. https://doi.org/10.1101/gr.204669.116

Smith, T. B., & Bernatchez, L. (2008). Evolutionary change in human-altered environments. *Molecular Ecology*, *17*(1), 1–8. https://doi.org/10.1111/j.1365-294X.2007.03607.x

Storey, J. D., Bass, A. J., Dabney, A., Robinson, D., & Warnes, G. (2019). *qvalue: Q-value estimation for false discovery rate control.* [R package version 2.16.0]. http://github.com/jdstorey/qvalue

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144–2164. https://doi.org/10.1111/mec.13606

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

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, *3*(3), 244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

White, T. A., Perkins, S. E., Heckel, G., & Searle, J. B. (2013). Adaptive evolution during an ongoing range expansion: The invasive bank vole (Myodes glareolus) in Ireland. *Molecular Ecology*, *22*(11), 2971–2985. https://doi.org/10.1111/mec.12343

Wilson, C. R., Meese, R. J., & Wyckoff, A. C. (2016). *Breeding chronology, movements, and life history observations of tricolored blackbirds in the California Central Coast*. 14.

Xenikoudakis, G., Ersmark, E., Tison, J.-L., Waits, L., Kindberg, J., Swenson, J. E., & Dalén, L. (2015). Consequences of a demographic bottleneck on genetic structure and variation in the Scandinavian brown bear. *Molecular Ecology*, *24*(13), 3441–3454. https://doi.org/10.1111/mec.13239

Younger, J. L., Clucas, G. V., Kao, D., Rogers, A. D., Gharbi, K., Hart, T., & Miller, K. J. (2017). The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Molecular Ecology*, *26*(15), 3883–3897. https://doi.org/10.1111/mec.14172

Zalasiewicz, J., Williams, M., Haywood, A., & Ellis, M. (2011). The Anthropocene: A new epoch of geological time? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1938), 835–841. https://doi.org/10.1098/rsta.2010.0339

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 II

Title: Genetic and environmental drivers of migratory behavior in western burrowing owls and implications for conservation and management

Running title: GEAs with migratory behavior

Authors: Kelly Barr[1,2]*, Christen M. Bossu[3], Rachael A. Bay[4], Eric C. Anderson[5,6], Jim Belthoff[7], Lynne A. Trulio[8], Debra Chromczak[9], Colleen L. Wisinski[10] , Thomas B. Smith[1,2], Kristen C. Ruegg[3]

Contact Information:

[1]Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA

[2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA

[3]Department of Biology, Colorado State University, Fort Collins, CO

[4]Department of Evolution and Ecology, University of California, Davis, Davis, CA

[5]Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA

[6]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO

[7]Raptor Research Center and Department of Biological Sciences, Boise State University, Boise, ID

[8]Department of Environmental Studies, San José State University, San Jose, CA

[9]Burrowing Owl Researcher & Consultant, Riegelsville, PA

[10]San Diego Zoo Wildlife Alliance, Escondido, CA

*Corresponding author: kellybarr@gmail.com

Abstract: Migration is driven by a combination of environmental and genetic factors, but many questions remain about those drivers. Potential interactions between genetic and environmental variants associated with different migratory phenotypes are rarely the focus of study. We pair low coverage whole genome resequencing with a *de novo* genome assembly to examine population structure, inbreeding, and the environmental factors associated with genetic differentiation between migratory and resident breeding phenotypes in a species of conservation concern, the western burrowing owl (*Athene cunicularia hypugaea*). Our analyses reveal a dichotomy in gene flow depending on whether the population is resident or migratory, with the former being genetically structured and the latter exhibiting no signs of structure. Among resident populations, we observed significantly higher genetic differentiation, significant isolation-by-distance, and significantly elevated inbreeding. Among migratory breeding groups, on the other hand, we observed lower genetic differentiation, no isolation-by-distance, and substantially lower inbreeding. Using genotype-environment association analysis, we find significant evidence for relationships between migratory phenotypes (i.e., migrant versus resident) and environmental variation associated cold temperatures during the winter and barren, open habitat. In the regions of the genome most differentiated between migrants and residents, we find significant enrichment for genes associated with the metabolism of fats. This may be linked to the increased pressure on migrants to process and store fats more efficiently in

preparation for and during migration. Our results provide a significant contribution toward understanding the evolution of migratory behavior and vital insight for ongoing conservation and management efforts for the western burrowing owl.

Keywords: genomics, migration, genotype-environment associations, inbreeding, genetic connectivity

INTRODUCTION

Migratory behavior has evolved repeatedly throughout the animal kingdom as species move to maximize their fitness in response to heterogeneous and changing environments (Dingle & Drake, 2007; Pulido, 2007; Shaw, 2016). While evolutionary theory commonly identifies seasonal fluctuations in climate and resources as the primary impetus (Alerstam et al., 2003; Cox, 1985; Winger et al., 2019), much remains to be learned about the relative contributions of environmental, genetic, and associations between genotypic and environmental variation in driving migratory phenotypes. Previous research into these drivers typically focused on the identification of individual environmental or genetic determinants. For example, significant environmental determinants include factors such as changing habitats (Gómez-Bahamón et al., 2020) and climates in birds (Winger et al., 2019), resource availability (Teitelbaum et al., 2015) and extreme weather events in mammals (Leclerc et al., 2021), and photoperiods and atmospheric pressure in insects (Chapman et al., 2015). Meanwhile, genetic determinants of migratory phenotypes are confirmed through both manipulation experiments, such as captive breeding and crossbreeding studies on both songbirds (Berthold & Pulido, 1994; Berthold & Querner, 1981; Pulido, 2007) and insects (Kent JR & Rankin, 2001), and, more recently, the

identification of numerous candidate genes associated with migratory behavior (Bossu et al., 2022; Jones et al., 2015; Mueller et al., 2011; Toews et al., 2019; Zhan et al., 2014).

Alleles of such candidate genes underlying differing migratory phenotypes may have important associations with environmental variation, but often these are not explicitly examined. In salmon, for instance, variations of the gene *GREB1* dictate the timing of migratory runs to upstream breeding grounds (e.g. spring/summer vs. fall; Narum et al., 2018; Thompson et al., 2020). Atlantic cod (*Gadus morhua*) are either migrants or resident facultatively based upon the orientation of a region of genes associated with increased movement performance, and post-glacial expansion of migratory populations are thought to be driven by the development of adaptive alleles in these regions that facilitated fitness in Arctic waters (Berg et al., 2017; Kirubakaran et al., 2016). In American Kestrels (*Falco sparverious*), migratory timing is significantly linked to variants in several known genes that regulate biological clocks (Bossu et al., 2022). Although the focus of these studies is strictly on identifying genomic regions associated with migratory phenotypes, the fact that migration occurs in each of these systems as a response to environmental stimuli suggests that genotype-environment interactions may be an important component.

Recent developments in genotype-environment association (GEA) analyses afford a promising opportunity to improve our identification of links between environmental and genotypic variation (Forester et al., 2018). For example, recent work employed GEA analyses to address adaptation-related questions, including identifying environmental and genetic drivers of adaptation (Capblancq et al., 2018; Dorant et al., 2020), predicting where rapid climate change may cause maladaptation in local populations (Bay et al., 2018; Fitzpatrick et al., 2021; K. Ruegg et al., 2018; Vanhove et al., 2021), and providing critical information to ensure the

success of increasingly necessary and intensive conservation actions such as assisted gene flow (Borrell et al., 2020). Another potential avenue for GEA analyses would be to address hypotheses about the environmental drivers of genetic variation linked to specific phenotypes. We can gain further insight into vital evolutionary phenomena such as variable migratory phenotypes within species by explicitly examining interactions among genotypes, phenotypes, and environmental variation using GEA analyses.

Here, we analyze links between genotypic and environmental variation underlying migratory phenotypes in the western burrowing owl (*Athene cunicularia hypugaea*), a species designated as being of conservation concern by the U.S. Fish and Wildlife Service (USFWS) and numerous states. Because interpreting GEAs and characterizing the genetic health of populations both require a detailed understanding of gene flow patterns, we also analyze genetic structure and inbreeding to inform our analyses and provide critical information for ongoing species conservation efforts. The western burrowing owl offers an ideal opportunity for this investigation because the subspecies is composed of both resident and migratory phenotypes across an extensive western North American breeding range where it is likely subject to a breadth of ecological variation (Fig. 1B).

As many migratory birds, *A. c. hypugaea* exhibits a cline of phenotypes along a latitudinal gradient with largely resident breeders in the southwestern U.S. and fully migratory populations farther north. Migrating burrowing owls are known to make relatively smaller movements versus that of other Neotropical migrants, with an average of 1800km (334km – 3541km; C. Conway *pers. comm.).* Resident breeding groups area also frequently exhibit partial non-breeding migration (Chapman, et al., 2011), meaning many individuals remain resident through the breeding cycle and others migrate to breed (Ogonowski & Conway, 2009).

Examining genetic relationships among resident and migratory breeding groups is an important goal for conservation and management of the species, particularly given the numerous on-going captive breeding and translocation projects (Doublet, 2020; Hennessy et al., 2022).

Using a high-resolution dataset composed of a *de novo* high coverage reference genome assembly and low coverage whole genome resequencing of samples from numerous migratory and resident breeding populations across the western burrowing owl's range, we address the following questions: 1) How does differential migratory behavior impact gene flow and inbreeding? 2) Are resident and migratory breeding groups genetically isolated from one another? 3) Are there correlations between genotypic and environmental variations that explain differences between migratory phenotypes? Our results not only reveal novel relationships between environmental variation, genotypes, and migratory phenotypes, but also provide critical information for ongoing species conservation efforts by reporting differences in two primary indices for understanding and predicting genetic health: population structure and inbreeding.

MATERIALS AND METHODS

Variant detection

Details regarding sample collection, genome sequencing, and sequence processing may be found in the Supplementary Methods; but notably we sequenced a reference genome to high coverage and 202 burrowing owl samples collected across their migratory and resident breeding range to low coverage. Because our resequencing dataset was low coverage, we used variant detection and analytical methods that largely did not require called genotypes. This included both genotype likelihoods as estimated in the program ANGSD (Korneliussen et al., 2014) and a single-read-sampling (SRS) method that randomly selects one read per variant to temper the bias

of high variation in locus-to-locus depths. Using these methods, files were prepared for analyses as described below using the following four filtering and genotyping frameworks and conditions: 1) Using ANGSD to produce genotype likelihood files for all individuals in the BEAGLE format (-doGlf 3) and a minor allele frequency file (-domaf 1) with restrictive filtering that uses a conservative minimum minor allele frequency (-minmaf 0.05), a low maximum likelihood of being polymorphic (-SNP_pval 1e-6), adjusting mapQ scores for excessive mismatches from the reference genome (-C 50), and confirming variants using a base alignment quality estimation (-baq 1). 2) For SRS analyses, we used the 'HaplotypeCaller' module in GATK (McKenna et al., 2010) to call genotypes for all individuals sequenced, filtered by removing insert/deletion variants, and kept only biallelic variants found in 50% of the individuals. 3) We used ANGSD to create population-specific site frequency spectra (SFSs) from site allele frequency (SAF) files using the reference genome to polarize allele calls (-anc), adjusting frequencies using individual F_{IS} (-indF), and with strict filtering conditions including discarding reads without unique mapping (-uniqueOnly 1), removing bad reads (-remove_bads 1), using only reads for which mates are mapped (-only_proper_pairs 1), discarding reads with low mapping quality (-minMapQ 1), keeping reads with high base quality (-minQ 20), dropping reads with low or high depth across samples (-setMinDepth 10 -setMaxDepth 500), keeping only biallelic sites (-skipTriallelic 1), and also previously described conditions (-minMaf 0.05 -C 50 -baq 1). 4) Minor allele frequency files (MAFs) were also generated for each sample site (-doMaf 1), sampling all the sites identified in the overall MAF file, and only generating minor allele frequencies for variants found in a minimum of four individuals in each population.

Population Structure and Inbreeding

We assessed population structure and gene flow patterns using multiple analytical frameworks, including principal components analyses (PCAs), Bayesian clustering analyses, estimation of genetic differentiation (F_{ST}), and calculating inbreeding (F_{IS}). Because the inclusion of related individuals can introduce bias in many of the analyses used here, we identified close relatives (*i.e.*, either parent-offspring or full sibships) using a combination of results from NgsRelate v2 (Hanghøj et al., 2019) and PCAs. A BEAGLE genotype likelihood file was prepared for NgsRelate v2 using the first set of conditions (1) described above for ANGSD. We removed the individual with the lowest coverage from each dyad of high relatedness as indicated by two of the following three analyses: 1) high relatedness across maximum likelihood estimates of the Jacquard's coefficients (k1 > 0.4; Jacquard, 2012), 2) a KING-robust estimator of kinship (r > 0.177; Waples et al., 2019), or 3) being paired outliers in PCAs. We calculated relatedness coefficients both across all samples and within sample sites.

For PCAs, a VCF file was prepared using GATK as previously described before implementing SRS. We obtained allele depths statistics (i.e. the "AD" field in the vcf file) for each subset of samples of interest using bcftools (Li et al., 2011), and filtered for sites that are bi-allelic (-m 2 -M 2), removed rare variants (--min-af 0.01) or fixed variants (--max-af 0.99), and eliminated sites with high levels of missing data (-i 'F_MISSING < 0.5'). With these allele depths, we used the R package 'SRS_Stuff' (https://github.com/eriqande/srsStuff) to identify population structure using a PCA with the SRS method.

We prepared genotype likelihood files both for estimating individual-level inbreeding coefficients (F_{IS}) and for Bayesian clustering analyses using the first (1) set of conditions described above for ANGSD. We estimated F_{IS} using ngsF (Vieira et al., 2013) and, based upon

our results, compared the means between residents and migrants using a Wilcoxon sign test in the R package 'ggpubr.' For clustering analyses, we ran 10 repetitions of NGSadmix (Skotte et al., 2013) each number of clusters (K) from 2-8 and compared these visually using the R package 'pophelper' (Francis, 2017) to assess consistency across multiple runs. Based upon our results (i.e., no structure among migrants; see Results), we reduced migratory breeding sites to three individuals each to make analyses more tractable. Note this reduction was only for the NGSadmix analysis and doing so would not be expected to impact results. Once we determined the most consistent K across runs, we mapped the results to create a GENOSCAPE (Ruegg et al., 2014, 2021). For this, we created a novel breeding range map for *hypugaea* using the R package 'ebirdst' (Fink et al., 2020) that utilizes citizen science observation data made available through the popular medium eBird (www.ebird.org), and smoothed this map by removing holes using the R package 'nngeo' (Dorman, 2018) and small polygons (<400km2) using the R package 'smoothr' (Strimas-Mackey, 2021). Then we used a modification of the R package 'tess3r' (Caye et al., 2016) as implemented in 'TESS3_encho_sen' (github.com/eriqande/TESS3_encho_sen) to map the cluster membership identified in NGSadmix using spatial kriging.

Finally, we calculated pairwise F_{ST} between all sample sites and tested for isolation by distance (IBD) among sites. For this, we estimated SFSs for each sample site as previously described (3) and then we used the 'realSFS' suite in ANGSD to create two dimensional SFSs and estimate F_{ST}s for each pair of sample sites. Using a Mantel test in the R package 'vegan', we assessed the significance of IBD across all sites and among either migratory or resident breeding sites only.

Identifying Candidate Loci

We used an available annotated genome for another burrowing owl subspecies, *A. c. cunicularia* (Mueller et al., 2020), to determine if loci that are highly differentiated between migrants and residents are in genic regions. We used this genome and the annotation because it is of higher quality than we sequenced for *A. c. hypugaea* (see results). For this analysis, we used the top 99.9% F_{ST} loci in a comparison of resident and migratory breeding groups. Using two pools of samples, one composed of the five resident populations that exhibit multiple lines of evidence for genetic structure and the other composed of all migratory sites, we created two-dimensional SFSs using ANGSD as previously described (genotyping conditions set 3), and calculated F_{ST} locus-by-locus using realSFS. Two sites were excluded from the "resident" pool because they were not genetically distinct from the migratory breeding sites in multiple analyses, and we were focused on detecting the genetic variants that differentiate migratory versus resident breeding groups. We then used BEDTools (Quinlan & Hall, 2010) to clip 200 bp segments around each locus and mapped them to the *A. c. cunicularia* genome, which was downloaded from genbank, using the 'aln' module in bwa because this outperforms 'mem' for such short segments. Using BEDTools, we then collected a candidate list of genes from the *A. c. cunicularia* annotation found within 25,000 bp segments around the mapped 200 bp segments. We compared this list of candidate genes to a growing list of genes known to be associated with migratory behavior (following Bossu et al., 2022). Finally, we used ShinyGO (Ge et al., 2020) to perform a gene ontology analysis using both the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttatus*) gene sets for comparisons.

Genotype-Environment Associations with Migratory Behavior

For GEAs of migratory phenotypes, we used gradient forest analyses to detect correlations between environmental variables and outlier loci in a comparison of migratory and resident genotypes. We defined outlier loci for this analysis as the top 1% most differentiated loci in the locus-by-locus F_{ST} calculation previously described between pools of migrants and residents (again, excluding unstructured resident sites based upon our results). Since low coverage whole genome data invariably include many missing sites per individual, we used population-specific allele frequencies for each sample site for these analyses. Outlier loci were subset from the sample-site specific MAF; however, since MAF files are calculated based upon the individual sample sites and filtering conditions vary by numbers of samples at each, not all outlier loci identified in the full migrant-resident comparison were present for individual sites. Those missing were dropped from further analysis. Numerous loci and one migratory site (NM) were removed from further analysis because of high levels of missing data (>50%). Remaining missing allele frequencies were imputed using the R package 'mice' (Buuren & Groothuis-Oudshoorn, 2011) using default settings.

We used the r packages 'gradientforest' and 'extendedforest' (Ellis et al., 2012) to test for correlations between these outlier loci and each of 19 bioclimatic variables (Hijmans et al., 2005), elevation, Normalized Difference Vegetation Index (NDVI), tree cover, migratory status (resident or migatory), and 20 landcover categories captured in the National Land Cover Database (NLCD; Dewitz, 2021). Details about these environmental variables can be found in Supplemental Material. Landcover was estimated for each category within 20km radii around the center points of sample sites to include most of the area being used by burrowing owls. We also included the first two PCs from PCAs using all sample sites to account for genetic structure, and,

to account for spatial biases, included the first two PCs from a principal coordinate of neighborhood matrix (PCNM) conducted using 'vegan.' For gradient forest analyses, we collected 2000 trees (nbin = 1, corr.threshold = 0.5) for each genetic variant and obtained a ranked list of environmental variables based upon their relative predictive power. This analysis was run 10 times to assess consistency in the top environmental variables identified. Using 10 replicates that each randomly permutes the observed environmental variation among sites, we assessed the significance of correlations by comparing the number of SNPs with a non-zero R^2 and the mean R^2 across these loci. For visualization of these results, the top four environmental variables were then collected from 10,000 random points from within the breeding range we predicted based upon observation data as previously described. This is bound to range of *A. c. hypugaea* within the United States due to the limitations of the NLCD, which only included information for the contiguous 48 states.

RESULTS

Data Quality

We produced an *A.c. hypugaea* reference genome assembly with an average depth of 49X. The total length of 1.25Gb is spread across 3,830 scaffolds at an N50 of 2.6Mb. BUSCO analyses revealed that 96.8% of the known genes from class Aves are captured by this reference genome. For the resequencing data, we removed low quality libraries (N=18) and one member of related pairs identified (N=23) and then conducted the following analyses with a dataset of 161 individuals sequenced to an average depth of 0.98X (range: 0.0104X – 2.132X).

Genetic Structure and Diversity

Population structure was strongly associated with migratory behavior. NGSadmix (Fig. 1A,B & S1; based upon 1,315,863 SNPs), the PCA (Fig. 1C; based upon 3,473,488 SNPs), and pairwise F_{ST}s (Fig. 2A,B & Table S2; based upon an average of 535,754 SNPs), all identified limited gene flow among resident breeding sites and no indications of limitations among migratory breeding sites. Other than two exceptions, the resident breeding sites are easily discriminated from one another and from the migratory sites when comparing PCs (Fig. 1C), exhibit higher relative levels of genetic differentiation (Fig. 2A), and form distinct clusters when considering 2 to 8 Ks (Fig. S1). Conversely, the migratory sites are not distinguishable from one another in a PCA (Fig. 1C), exhibit lower relative genetic differentiation (Fig. 2A), and exhibit high levels of admixture at all Ks considered (Fig. S1). We note that the NGS-Admix results are expectedly messy given the low coverage data being analyzed; but the overall trends in population structure are clear and supported by additional analyses.

These same patterns are further supported by analyses of isolation-by-distance (IBD) and inbreeding. While migratory sites do not follow a pattern of IBD, the correlation between genetic differentiation and geographic distance is significant and positive among the resident breeding sites (Fig. 2B; Mantel's $r = 0.675$, $p < 0.01$). This suggests high gene flow in the former and distance-restricted/stepping-stone gene flow in the latter. Meanwhile, restrictions on gene flow in residents are also evidenced by higher measures of individual inbreeding coefficients, F_{IS}, in resident versus the migrant populations (Fig. 2C; $W=1285$, $p < 0.001$).

Two sites, one composed of samples collected from burrowing owls around Phoenix, AZ (AZ-P) and another in the Imperial Valley, CA (CA-Imp), are exceptions among the resident breeding sites. These areas cannot be distinguished from the migrants using PCA (Fig. 1C) and do not form individual clusters in NGSadmix (Fig. 1A & S1); however, these sites exhibit higher

levels of inbreeding than observed in migratory sites, especially at AZ-P where inbreeding appears to be the highest of all study sites (Fig. S2). Genetic differentiation (F_{ST}) is also high between these two areas and the migratory sites (Table S2).

Phenotype-Genotype-Environment Analyses

After subsetting population-specific minor allele frequency files generated in ANGSD for the top 1% highest F_{ST} loci between residents and migrants and dropping loci related to missing data, we used 6,954 SNPs for gradient forest analyses. These were the highest of 815,438 total loci with positive F_{ST}s (mean F_{ST} = 0.024) considered for this analysis. Top loci were distributed across the genome and had an average F_{ST} of 0.205 (SD = 0.039). Of these, 3,458 were positively correlated with environmental variables. Repeated gradient forest runs (N=10) consistently identified the same top 10 environmental variables in terms of R^2 weighted importance in the same order between runs (Fig. S3). The top four uncorrelated variables were, in order of ranked importance, minimum temperature of the coldest month (Bioclim 6), pcnm1 (independent spatial variable pc1), barren/open land (landcover class 31), and mean temperature of the coldest quarter (Bioclim 11). These were the top four variables in order across all 10 empirical gradient forest runs, and comparisons of these variables between migratory and resident breeding sites illustrate apparent differences at each one of these (Figs S6A-S6D). Note that PC1, which accounts for population structure, is the 11th most important explanatory variable, and 9 of 10 variables that are higher are environmental (Fig. S3). Empirical observations of SNPs with positive R^2 and the average R^2 across loci were significantly higher than in randomizations (Fig. S4). It is not surprising that pcnm1, which accounts for spatial autocorrelation among the sample sites, would be a top variable as we examined breeding

aggregations with a clear spatial relationship (i.e., migratory populations in higher latitudes and resident populations in lower). Including pcnm1 in our analyses is a conservative approach, as genetic variants that might otherwise be associated with other environmental predictors are linked to it instead. Furthermore, the top four most important loci for each of the top four environmental variables identified by gradient forest analyses exhibit allele frequency differences between residents and migrants that trend with environmental variation (Fig. S5).

Candidate Loci

The *A. c. cunicularia* genome, with an N50 of 42M bp over 445 scaffolds, is significantly less fragmented than the one we produced for *A. c. hypugaea* (see Results). Hence, using the higher quality genome of this subspecies is well-justified. Of the 0.1% most differentiated loci between migrants and residents (N=1,009), 960 mapped successfully to the *A. c. cunicularia* genome. Within 25,000 bp regions around these successfully mapped sites, 457 unique genic regions were identified using the *cunicularia* annotation, and 116 of these had unique NCBI identifiers. We obtained a list of 24 recognized genes within these genic regions (Table S3). Gene ontology analyses using the chicken and zebra finch genomes both revealed enrichment in several pathways associated with processing fat (i.e., lipophagy; Table S4). This was the only pathway with significant enrichment. There were no genes in common with a list of previously identified genes relevant to migratory behavior in birds following Bossu et al. (2022).

DISCUSSION

While it is generally accepted that migratory behavior is an adaptation to life in seasonal environments, few studies have successfully identified genetic and environmental associations

underlying this key fitness-linked trait. Here we combined population and landscape genomic approaches to identify putative environmental associations with genetic differentiation between migrant and resident burrowing owls across North America. We found strong associations between the top differentiated loci between migrant and resident breeding burrowing owls and environmental variables related to cold, winter temperatures (i.e., the coldest month and the coldest quarter; Figs. S6A and S6D) and barren, open habitat (Fig. S6C). Further, gene flow, population structure, and inbreeding patterns largely could be explained by breeding strategy. Migratory breeders exhibited high gene flow and low inbreeding, and resident breeders exhibited limited gene flow and high inbreeding. Further investigation into the putative function of genes underlying migratory behavior provides further insight into differences in the forms, specifically indicating differences in genes linked to metabolic processes (i.e., liphophagy). Overall, our results have important implications for understanding links between genetic and environmental variation underlying migratory behavior across species and for the genetic health (i.e., inbreeding and gene flow) of western burrowing owl populations.

Population Structure, Gene Flow, and Inbreeding

While previous genetic studies failed to identify any significant limitations to gene flow among resident or migratory breeding groups in western burrowing owls (Desmond et al., 2001; Korfanta et al., 2005; Macías-Duarte et al., 2020), we detected clear population structure patterns associated with migratory phenotypes. Namely, we found distinct genetic clustering of residents by population and no limitations to gene flow among the migratory breeding groups. Many organisms exhibit similar gene flow regime differences based upon migratory behavior with genetic structure among resident breeders and high connectivity among migrants, including

brown trout (*Salmo trutto*; Lemopoulos et al., 2018), river lampreys (*Lampetra fluviatilis*; Bracken et al., 2015), European blackcaps (*Sylvia atricapilla*; Delmore et al., 2020), and numerous bat species (Moussy et al., 2013). This is also not unexpected for western burrowing owls given the propensity of individuals from migratory breeding groups to disperse to new areas, sometimes over great distances (Riding & Belthoff, 2018). On the other hand, significant IBD among residents suggests a stepping-stone pattern of gene flow that leaves distantly-spaced breeding areas more differentiated from one another. Due to limitations with our low coverage dataset, we did not assess whether resident populations are differentiated due to genetic drift and time or natural selection.

The differences in gene flow between residents and migrants have important implications for the relative genetic health of western burrowing owl populations. Notably, inbreeding is significantly higher in all resident populations than in what seems to be effectively one large migratory population. The level at which inbreeding might have fitness consequences for local populations is difficult to discern and likely varies species-to-species. Ralls et al. (2018), however, suggest that an inbreeding level of 0.1 is the point at which an isolated population should receive require an active management response, such as genetic rescue (Whiteley et al., 2015), to avoid decreased fitness. The highest F_{IS} we observed was a resident breeder near San Jose, CA at 0.11, and many resident breeding birds were just below this estimate. For species management purposes, it would be valuable to assess if elevated inbreeding has fitness consequences for *A. c. hypugaea,* such as lowered breeding success as observed in red deer (*Cervus elaphus*; Slate et al., 2000*)* or reduced survival as reported in song sparrows (*Melospiza melodia*; Keller, 1998). Further, captive breeding, genetic rescue, and translocation projects, the latter of which sometimes entails novel pairing of adults (Hennessy et al., 2022), would benefit

from using genomic data to guide pairing decisions and precisely reduce inbreeding in potential offspring (Bossu et al., n.d.).

Genotype-Environment Drivers of Migratory Behavior

Whereas previous studies focused individually on environmental or genetic variants underlying migratory behavior, we identified significant evidence for links between genotypic and environmental variation that differentiate migratory phenotypes within a species. Among the top four environmental predictors (Fig. S3) of genetic variation underlying migratory behavior are minimum temperature of the coldest month and mean temperature of the coldest quarter. These top climatic predictors may reflect extreme winter conditions associated with seasonality on migratory breeding grounds, which, in turn, results in the annual fluctuation of resources that is a primary driver of the evolution of migration in birds (Alerstam et al., 2003; Shaw, 2016; Winger et al., 2014, 2019). While the migratory birds are not directly experiencing selection from these environmental variables per se, examining correlations between this and genetic variation helps to further understand the differences between migratory phenotypes. The robustness of this result is illustrated by the randomizations we employed in the gradient forest analyses, as we observed significantly weaker associations and fewer variants with positive associations than those of the empirical dataset.

Another top environmental predictor of genotypic variation was barren land, which is defined as having less than 15% vegetation cover and may reflect low productivity of the desert or otherwise arid landscapes in which many of the resident populations are found. Migratory breeding groups are generally found in more productive grassland habitat--though it is notable that many areas in the migratory breeding range are subject to periods of drought that can also

leave the landscape visibly barren as well. It is possible that the connection to barren land is also linked to seasonality. After all, as do all Neotropical migrants during Spring, western burrowing owls migrate northward to take advantage of seasonal abundance. Burrowing owls that are migratory depart the open, barren habitat common to the American southwest to breed in more productive grasslands farther north.

The top outlier loci in our analysis were found to be associated with the regulation of fats, which suggests that migrants and residents may differ in metabolic processes linked to fat mobilization. Specifically, genic regions near the 960 outlier loci in our analysis were enriched for genes involved in the lipophagy pathway (Table S4), which regulates the presence of fat molecules in the body whether via accumulation or metabolism. This result aligns well with previous research into the physiological adaptations of migratory species. Not only are they uniquely able to cache fats for a ready energy source for migration, but migrants also more efficiently process them during extended movements (Guglielmo, 2018; Ramenofsky, 1990). A transcriptome study using livers from a passerine collected before, during, and after migration found that the lipophagy pathway specifically was active throughout (Frias-Soler et al., 2022). Given these observations in other species, one potential explanation for our result is that migratory and resident breeding western burrowing owls use their lipophagy pathways in different ways in relation to adaptation to their contrasting life cycles. Notably, most of the resident breeding sites represented in this study are occupied by migratory birds during the winter, lending further credence to the suggestion that the migratory breeders have experienced enrichment of the lipophagy pathway versus the residents.

Future genomic work on burrowing owls would benefit from a novel genome using long-read data that would better capture repeat or otherwise hard to sequence regions of the genome

that are likely not represented in the reference genome assembled here from short-read data. Pairing such a reference genome with higher depth data from individuals would be helpful for further understanding the potential fitness effects of elevated inbreeding at many of the resident sites and for further revealing associations between environmental and genetic variant underlying migratory phenotypes.

Lack of Structure in Two Resident Sites

Two sample sites, the Imperial Valley of CA (CA-Imp) and Phoenix, AZ (AZ-P), are exceptional in being resident breeding sites that cannot be distinguished from the migrants in either admixture analyses (Fig. 1A) or PCA (Fig. 1C). Arid regions subject to intense irrigation particularly for agriculture are known to support thriving populations of western burrowing owls (DeSante et al., 2004; Macias-Duarte, 2011). The Imperial Valley, for instance, experienced a 2.5X fold increase in burrowing owl population density from 1980 to 2000 as agricultural operations escalated in the area (Rosenberg & Haley, 2004), and it currently supports the majority of the total extant population in California (Poulin et al., 2020). Recent work suggests non-breeding partial migratory populations may be experiencing a switch to breeding partial migratory populations in desert areas heavily impacted by agriculture (Macías-Duarte et al., 2020). Increased gene flow resulting from this change may explain the lack of differentiation in the PCA between both CA-Imp and AZ-P and the migratory group. The recency of this phenomenon might be indicated by the fact that these two sites exhibit the same pattern of isolation by distance as other resident breeding sites as it is possible that the sites have yet to reach equilibrium, and there would be a longer lag effect in the F_{ST} calculation versus PCA. We cannot resolve the cause from our current dataset, however, and there are other complicating

factors. For example, AZ-P has long been subject to an on-going, intense translocation project without any guidance on population structure, genetic relatedness, or verification of migratory phenotypes (Doublet, 2020). Notably, AZ-P exhibits the highest levels of inbreeding for any of the sites which may be the product of inadvertent mixing of close relatives.

Conservation Implications

The results we report here have broad implications for our understanding of the evolution of migration and the management of western burrowing owls. Resident breeding populations show elevated inbreeding and may benefit from genetic rescue efforts. Based upon the low genetic differentiation among populations (Frankham et al., 2011), it is unlikely that translocations between structured populations would lead to outbreeding depression; however, our dataset is not sufficient for detecting signals of local adaptation that may exist within resident breeding groups. Future work on potential local adaptation in resident populations could be helpful for guiding source choices for genetic rescue. At many of the sites, genomic data might be used for distinguishing residents from overwintering migrants, which is a considerable difficulty for burrowing owl conservation programs. Given the significant genotype-environment associations underlying migratory behavior we detect, future work towards understanding the fitness consequences of retaining migrants to boost nonbreeding partial migratory populations would be a helpful next step for species conservation efforts as well. Further examination of these associations particularly in areas where burrowing owl migratory behavior may be shifting would be beneficial for understanding the links with changing climate and habitat, and also for predicting potential behavioral changes in the species.

Conclusions

Our study combines landscape and population genomic approaches to identify associations among genetic and environmental factors underlying migratory phenotypes. Additionally, our GO term analysis suggests enrichment of genes in the lipophagy pathway, lending further support to the idea that migrants and residents differ in their ability to process and store fats. Future work employing similar population and landscape genomic analyses across taxa will reveal the extent to which our findings are generalizable across species.

Acknowledgements

We thank the many collaborators who provided samples, site access, and often a hand in the extensive fieldwork undertaken for this effort. See Table S1 for a complete list of those collaborators and a description of their contributions. We thank Jasmine Rajbhandary for helping in the field, and both she and Vinh Le for help with labwork. We thank members of the Smith Lab for comments on several drafts of this manuscript. This work was made possible by a California Energy Commission grant to K. Ruegg and T. Smith (EPC-15-043), a National Geographic grant to K. Ruegg (WW-202R-17), a grant to K. Ruegg from the National Science Foundation (NSF-1942313), and a contract to K. Ruegg from the Santa Clara Valley Habitat Agency. We thank the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) for their assistance with the Next-Generation Sequencing. Computational allocations from the Extreme Science and Engineering Discovery Environment (Xsede), as well as UCLA's Shared Hoffman2 Cluster made this work possible.

Data Accessibility and Benefit-Sharing: Raw sequence reads for resequencing and the reference genome are deposited in the SRA. Genotype files are on DRYAD. No benefit-sharing statement is necessary.

Author Contributions: KRB, CMB, RAB, TBS, and KCR helped design the research and write the manuscript. KRB, CMB, and RAB contributed to the bioinformatic pipeline and analyses. JB, DC, and CLW provided a significant portion of the samples. LAT inspired the initiation of the research. KRB performed fieldwork and labwork.

Animal Welfare and Permit Statements: Samples were collected under Tom Smith's Federal Bird Banding Permit, #21901, and Kelly Barr's Scientific Collecting Permit, #SC-11568. Animal handling and sampling protocols were conducted with the approval of UCLA's Animal Research Committee (ARC), agreement #2017-073-03

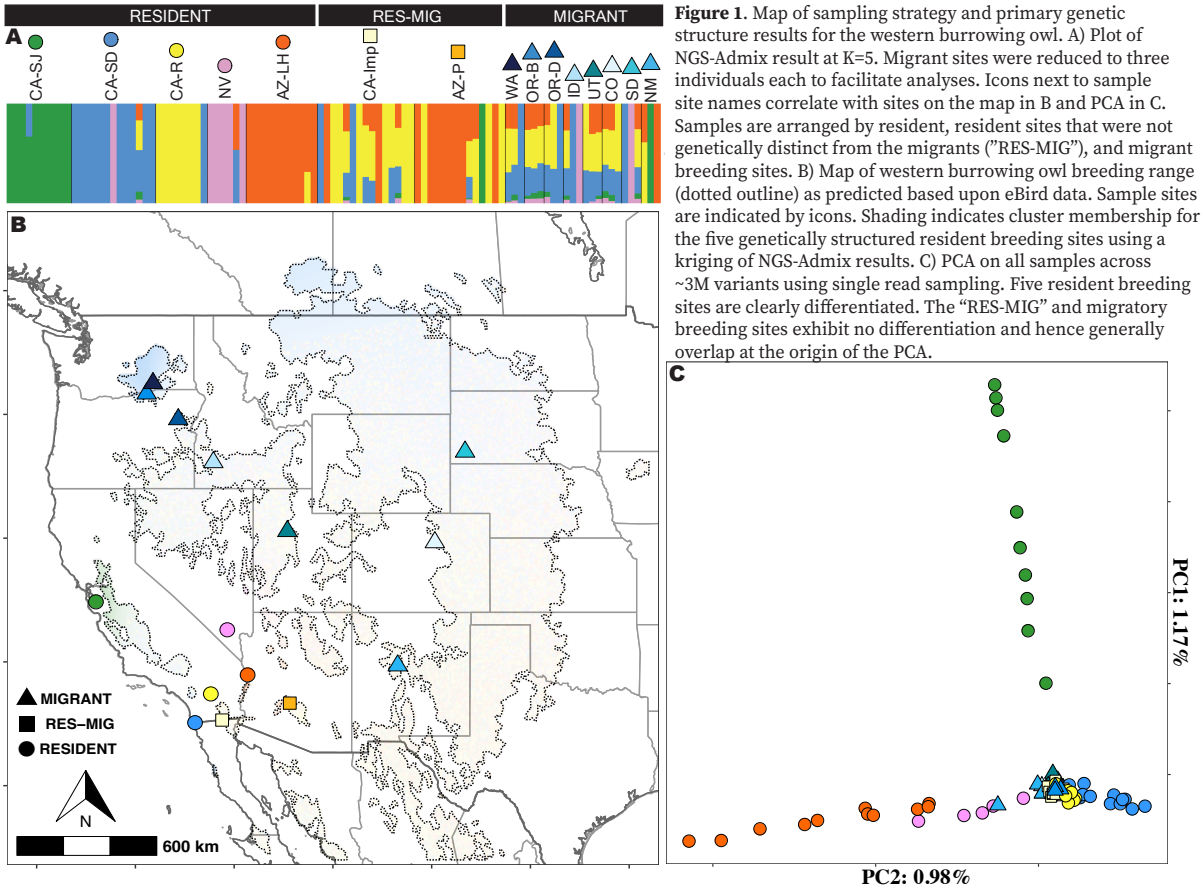


Figure 1. Map of sampling strategy and primary genetic structure results for the western burrowing owl. A) Plot of NGS-Admix result at K=5. Migrant sites were reduced to three individuals each to facilitate analyses. Icons next to sample site names correlate with sites on the map in B and PCA in C. Samples are arranged by resident, resident sites that were not genetically distinct from the migrants ("RES-MIG"), and migrant breeding sites. B) Map of western burrowing owl breeding range (dotted outline) as predicted based upon eBird data. Sample sites are indicated by icons. Shading indicates cluster membership for the five genetically structured resident breeding sites using a kriging of NGS-Admix results. C) PCA on all samples across ~3M variants using single read sampling. Five resident breeding sites are clearly differentiated. The "RES-MIG" and migratory breeding sites exhibit no differentiation and hence generally overlap at the origin of the PCA.

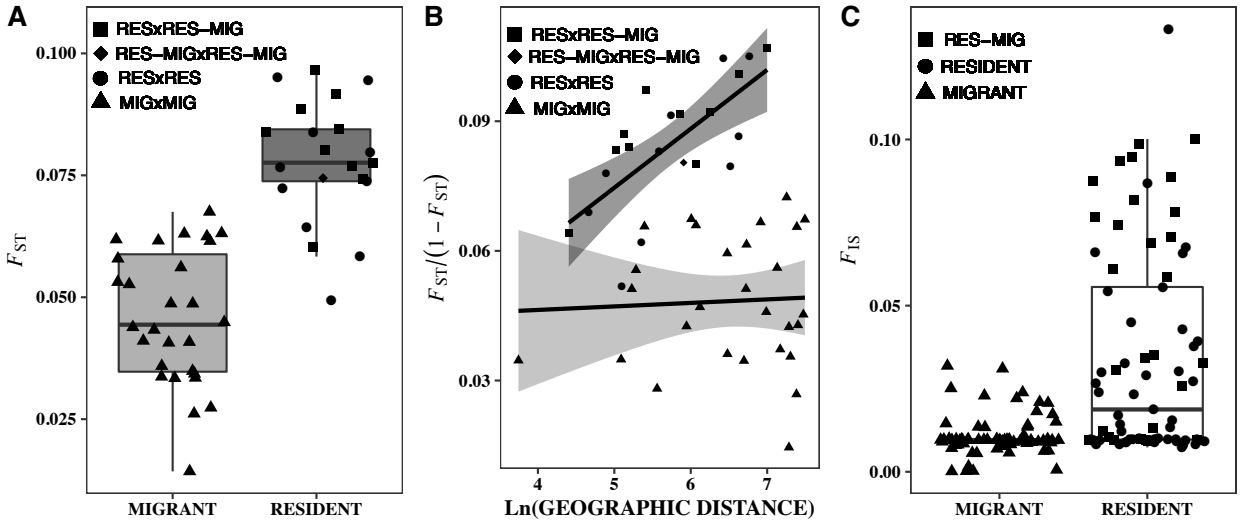


Figure 2. Comparisons of F statistics between BUOW migratory and resident breeding sites. Switcher sites are grouped with resident breeding sites. A) Residents are significantly more differentiated from one another than migrants (W=26, p<0.001). B) Residents exhibit significant isolation-by-distance (r=0.67, p=0.004) while migrants do not (r=-0.04, p=0.58). C) Inbreeding is significantly higher in residents than migrants (W=1285, p<0.001).

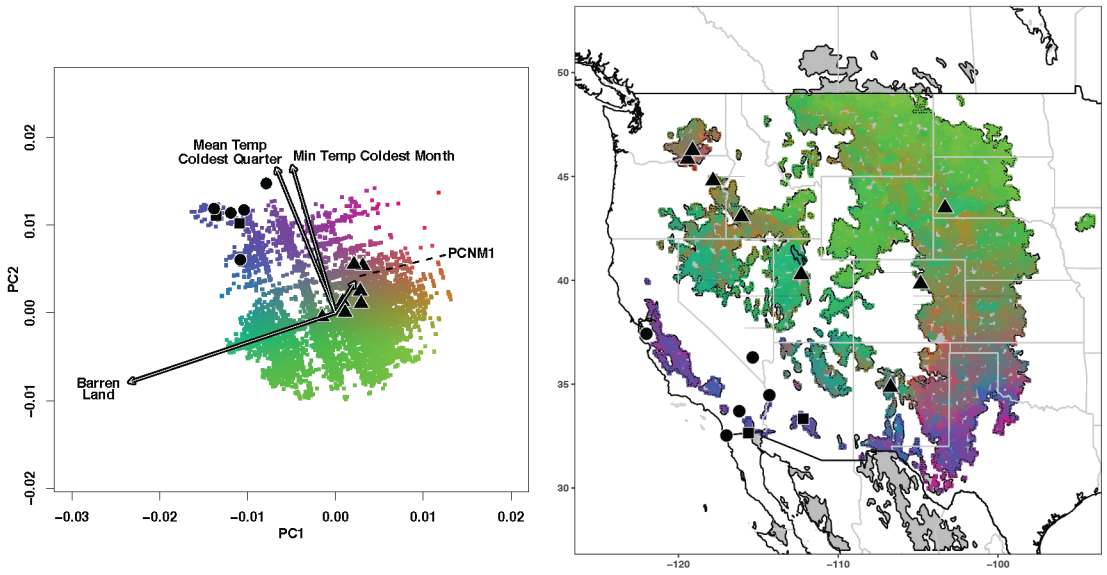


Figure 3. PCA (A) and map (B) portraying gene-environment correlations associated with migratory behavior across the BUOW range. Colors are based upon 10,000 random points across the breeding range, but is restricted to the U.S. due to the availability of the landcover data. A) PCA of climate variables with PC scores associated with sample sites indicated with symbols that match Fig. 1. Arrows indicate the loadings of top-ranked variables identified by gradient forest analysis. B) Map of projected GEA correlations across the BUOW range and sample sites indicated as in Fig. 1.

SUPPLEMENTAL MATERIALS

Table S1. Collaborators who either provided samples or provided access to sites for sampling ("collected with assistance"). Some collaborators provided samples that are not included in this study. *Specific location is the centerpoint GPS for the general sample sites, and these were used for genotype-environmental analyses.

Organization	Contact(s)	Location(s)	Specific Location*	Samples Provided	Samples Collected With Assistance
San Diego Zoo Wildlife Alliance	Colleen Wisinski	CA-SD	32.55, -116.98	70	0
		CA-Imp	32.65, -115.61		
		CA-R	33.71, -116.18		
BUOW Researcher & Consultant	Debra Chromczak	CA-SJ	37.429, -121.998	51	0
Boise State	Dr. Jim Beltoff	ID	43.065, -116.054	147	0
		ORB	44.8, -117.83		
		ORD	45.84, -119.43		
		WA	46.26, -119.11		
		SD	43.49, -103.31		
		CO	39.83, -104.84		
UCLA/Wild-at-Heart	Bob Fox, Beth Edwards	AZ-P	33.333, -112.183	0	21
UCLA/University of Idaho	Carl Lundblad	UT	40.281, -112.306	2	37
		LV	36.301, -115.346		
		NM	34.852, -106.719		
UCLA/Arizona State University at Lake Havasu	Kerrie Anne Loyd	AZ-LH	34.479, -114.317	2	12
Assiniboine Zoo	Stephen Petersen	Assiniboine Zoo	Winnipeg, MB, Canada	18	0
University of Florida	Elizabeth White	Florida	Florida, USA	6	0
Calgary Zoo	Graham Dixon-MacCallum	Calgary Zoo	Calgary, AB, Canada	18	0
The Burrowing Owl Conservation Society of BC	Lauren Meads	British Columbia/ Oregon	British Columbia, Canada Umatilla, OR, USA	11	0

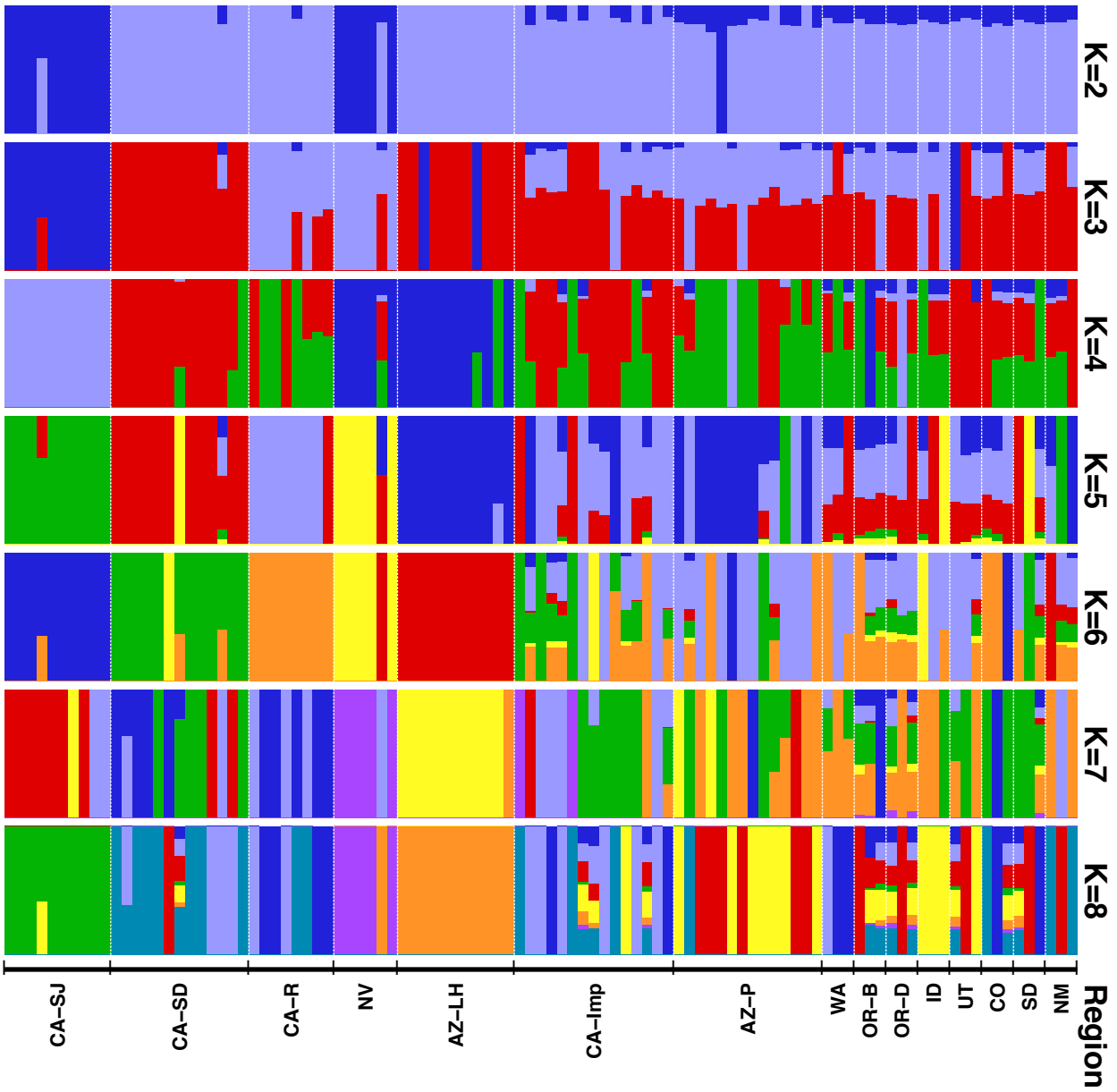


Figure S1. NGS-Admix results at Ks between 2 - 8, as indicated on the right. Migratory sites were reduced to 3 samples each to facilitate analyses.

Table S2. Pairwise F_{ST} between all sites sampled as calculated using the 'realSFS' module in ANGSD.

	CO	ID	CA-Imp	AZ-LH	NV	NM	CA-SJ	OR-B	OR-D	AZ-P	CA-R	SD	UT	WA
ID	0.0530													
CA-Imp	0.0551	0.0684												
AZ-LH	0.0587	0.0776	0.0801											
NV	0.0455	0.0730	0.0776	0.0732										
NM	0.0274	0.0439	0.0482	0.0441	0.0332									
CA-SJ	0.0661	0.0902	0.0923	0.0951	0.0797	0.0647								
OR-B	0.0407	0.0616	0.0651	0.0658	0.0515	0.0359	0.0756							
OR-D	0.0404	0.0642	0.0667	0.0681	0.0527	0.0343	0.0748	0.0488						
AZ-P	0.0590	0.0760	0.0745	0.0885	0.0839	0.0667	0.0961	0.0706	0.0735					
CA-R	0.0383	0.0597	0.0620	0.0583	0.0493	0.0254	0.0731	0.0446	0.0463	0.0739				
SD	0.0518	0.0675	0.0685	0.0766	0.0718	0.0449	0.0886	0.0615	0.0630	0.0756	0.0564			
UT	0.0487	0.0620	0.0641	0.0711	0.0650	0.0349	0.0848	0.0561	0.0589	0.0742	0.0503	0.0625		
WA	0.0261	0.0408	0.0458	0.0444	0.0379	0.0144	0.0619	0.0333	0.0335	0.0659	0.0238	0.0433	0.0334	
CA-SD	0.0596	0.0755	0.0770	0.0830	0.0766	0.0508	0.0941	0.0665	0.0680	0.0844	0.0624	0.0756	0.0707	0.0494

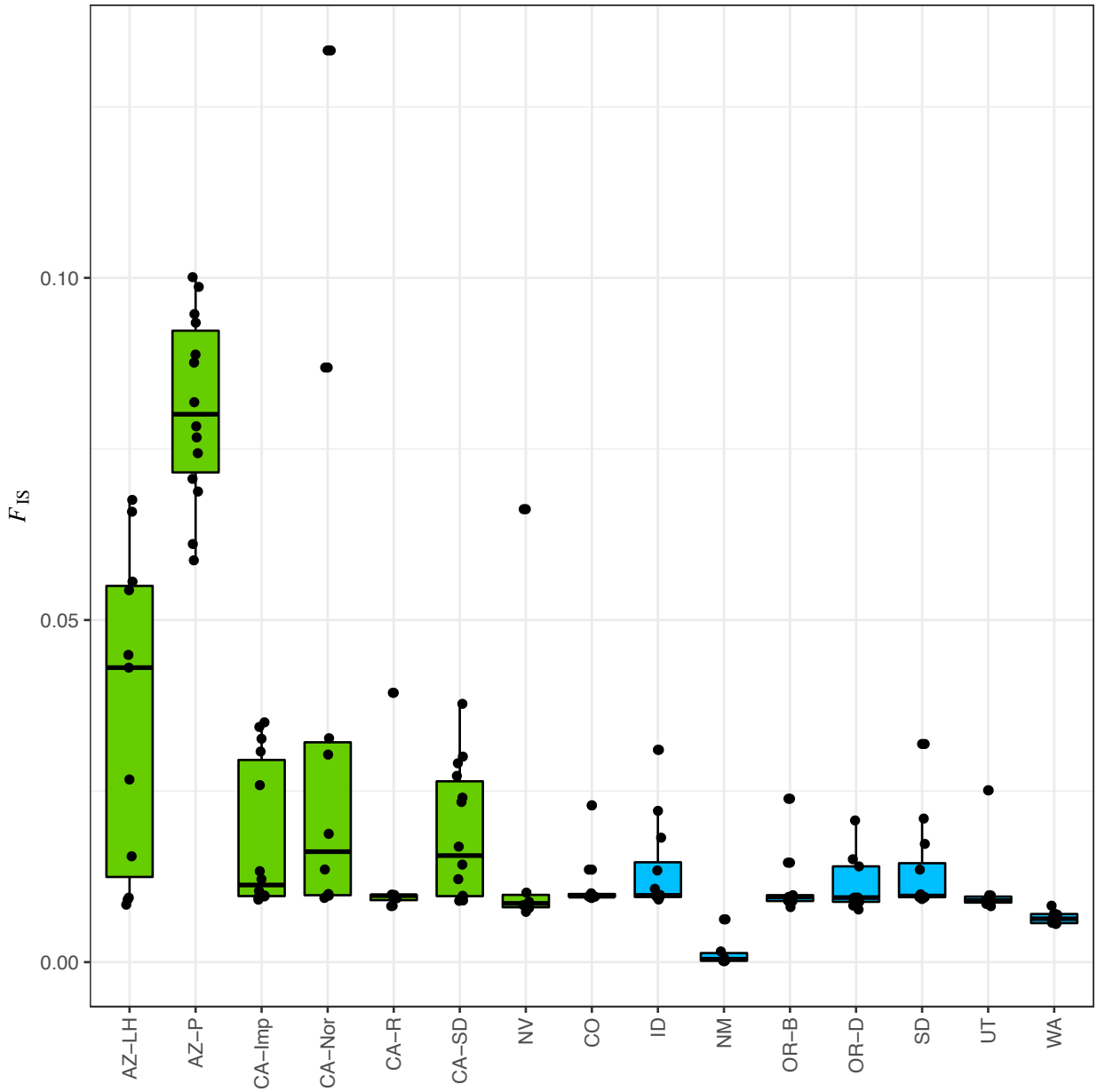


Figure S2. F_{IS} calculated by sample site. Residents are colored green and migrants blue. CA-Nor TO CA-SJ

Table S3. List of genes within 25kbps in *Athene cunicularia cunicularia* annotation using top 0.1% most differentiated loci between *A. c. hypugaea* migrants and residents.

Name	Ensembl Gene ID	Entrez	Gene Type	Species	Chr	Position (Mbp)	Description
CALCR	ENSTGUG00000001514	100223972	protein_coding	Zebra finch	2	24.236071	calcitonin receptor [Source:NCBI gene;Acc:100223972]
NOP14	ENSTGUG00000010455	100220875	protein_coding	Zebra finch	4	9.714675	NOP14 nucleolar protein [Source:NCBI gene;Acc:100220875]
HTT	ENSTGUG00000010399	NA	protein_coding	Zebra finch	4	9.78153	huntingtin [Source:HGNC Symbol;Acc:HGNC:4851]
SORCS2	ENSTGUG00000010129	NA	protein_coding	Zebra finch	4	11.48747	sortilin related VPS10 domain containing receptor 2 [Source:HGNC Symbol;Acc:HGNC:16698]
SPTLC2	ENSTGUG00000012332	NA	protein_coding	Zebra finch	5	40.177954	serine palmitoyltransferase long chain base subunit 2 [Source:HGNC Symbol;Acc:HGNC:11278]
GTF2A1	ENSTGUG00000012378	100219035	protein_coding	Zebra finch	5	41.766624	general transcription factor IIA subunit 1 [Source:NCBI gene;Acc:100219035]
PLCE1	ENSTGUG00000008633	100222852	protein_coding	Zebra finch	6	19.024883	phospholipase C epsilon 1 [Source:NCBI gene;Acc:100222852]
ARMH3	ENSTGUG00000009880	NA	protein_coding	Zebra finch	6	21.589012	armadillo like helical domain containing 3 [Source:HGNC Symbol;Acc:HGNC:25788]
STN1	ENSTGUG00000010354	NA	protein_coding	Zebra finch	6	23.35153	STN1 subunit of CST complex [Source:HGNC Symbol;Acc:HGNC:26200]
ZFYVE9	ENSTGUG00000008916	100231800	protein_coding	Zebra finch	8	7.365492	zinc finger FYVE-type containing 9 [Source:NCBI gene;Acc:100231800]
OSBPL9	ENSTGUG00000008771	100223171	protein_coding	Zebra finch	8	7.542853	oxysterol binding protein like 9 [Source:NCBI gene;Acc:100223171]
CFAP57	ENSTGUG00000007089	100225077	protein_coding	Zebra finch	8	12.666171	cilia and flagella associated protein 57 [Source:NCBI gene;Acc:100225077]
AK5	ENSTGUG00000007008	100229919	protein_coding	Zebra finch	8	13.150598	adenylate kinase 5 [Source:NCBI gene;Acc:100229919]
ZZZ3	ENSTGUG00000007000	100227013	protein_coding	Zebra finch	8	13.250444	zinc finger ZZ-type containing 3 [Source:NCBI gene;Acc:100227013]
TMEM266	ENSTGUG00000003531	100223195	protein_coding	Zebra finch	10	0.781096	transmembrane protein 266 [Source:NCBI gene;Acc:100223195]
ETFA	ENSTGUG00000003549	NA	protein_coding	Zebra finch	10	0.851982	electron transfer flavoprotein subunit alpha [Source:HGNC Symbol;Acc:HGNC:3481]
TLN2	ENSTGUG00000005039	100220667	protein_coding	Zebra finch	10	4.828676	talin 2 [Source:NCBI gene;Acc:100220667]
THSD4	ENSTGUG00000005809	100232194	protein_coding	Zebra finch	10	6.823963	thrombospondin type 1 domain containing 4 [Source:NCBI gene;Acc:100232194]
PEPD	ENSTGUG00000009480	100218638	protein_coding	Zebra finch	11	18.354231	peptidase D [Source:NCBI gene;Acc:100218638]
TRPC7	ENSTGUG00000001207	NA	protein_coding	Zebra finch	13	9.703912	transient receptor potential cation channel subfamily C member 7 [Source:HGNC Symbol;Acc:HGNC:20754]
CLINT1	ENSTGUG00000000611	100226122	protein_coding	Zebra finch	13	13.848802	clathrin interactor 1 [Source:NCBI gene;Acc:100226122]
EBF1	ENSTGUG00000000593	100228980	protein_coding	Zebra finch	13	14.182213	EBF transcription factor 1 [Source:NCBI gene;Acc:100228980]
DUSP1	ENSTGUG00000000299	100219518	protein_coding	Zebra finch	13	16.133668	dual specificity phosphatase 1 [Source:NCBI gene;Acc:100219518]
LOC113481779	Not mapped	NA	NA	NA	NA	NA	NA

Table S4. Gene ontology enrichment analysis using genes within 25kbps of top 0.1% most differentiated loci.

Pathway	Enrichment FDR	Pathway Genes	Fold Enrichment	Genes
Regulation of lipophagy	0.00792242	5	289.026087	HTT, SPTLC2
Positive regulation of lipophagy	0.00792242	5	289.026087	HTT, SPTLC2
Lipophagy	0.011072711	7	206.447205	HTT, SPTLC2

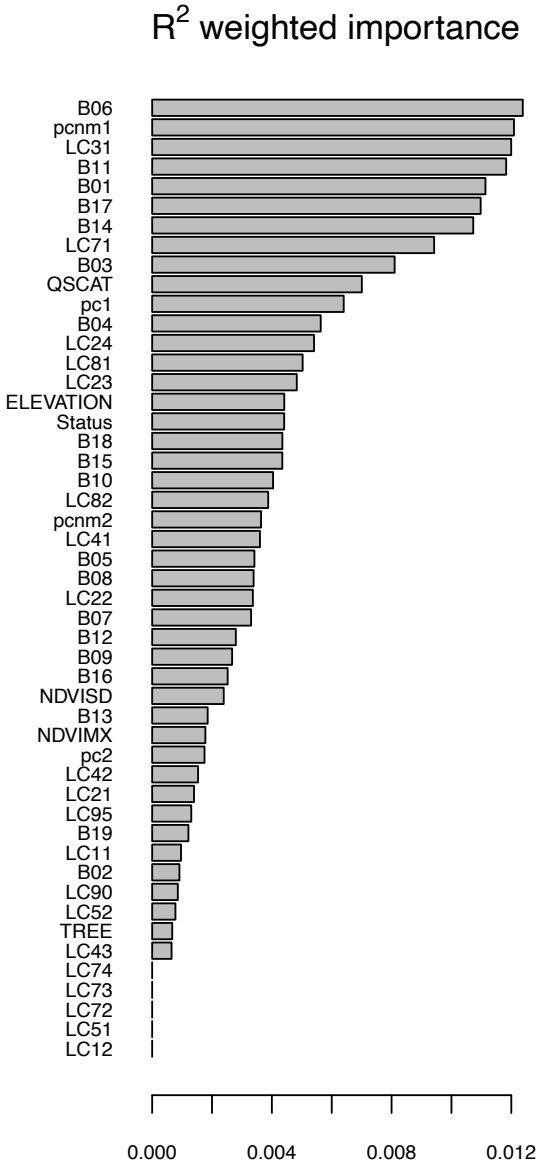


Figure S3. Results of gradient forest analyses using the top 1% of loci most differentiated loci between resident and migrant samples, excluding resident birds sampled at AZ-P and CA-Imp. PCNM1 and B01 are strongly correlated (r>0.75) with B06, hence B06, LC31, B11, and B17 are used for plotting.

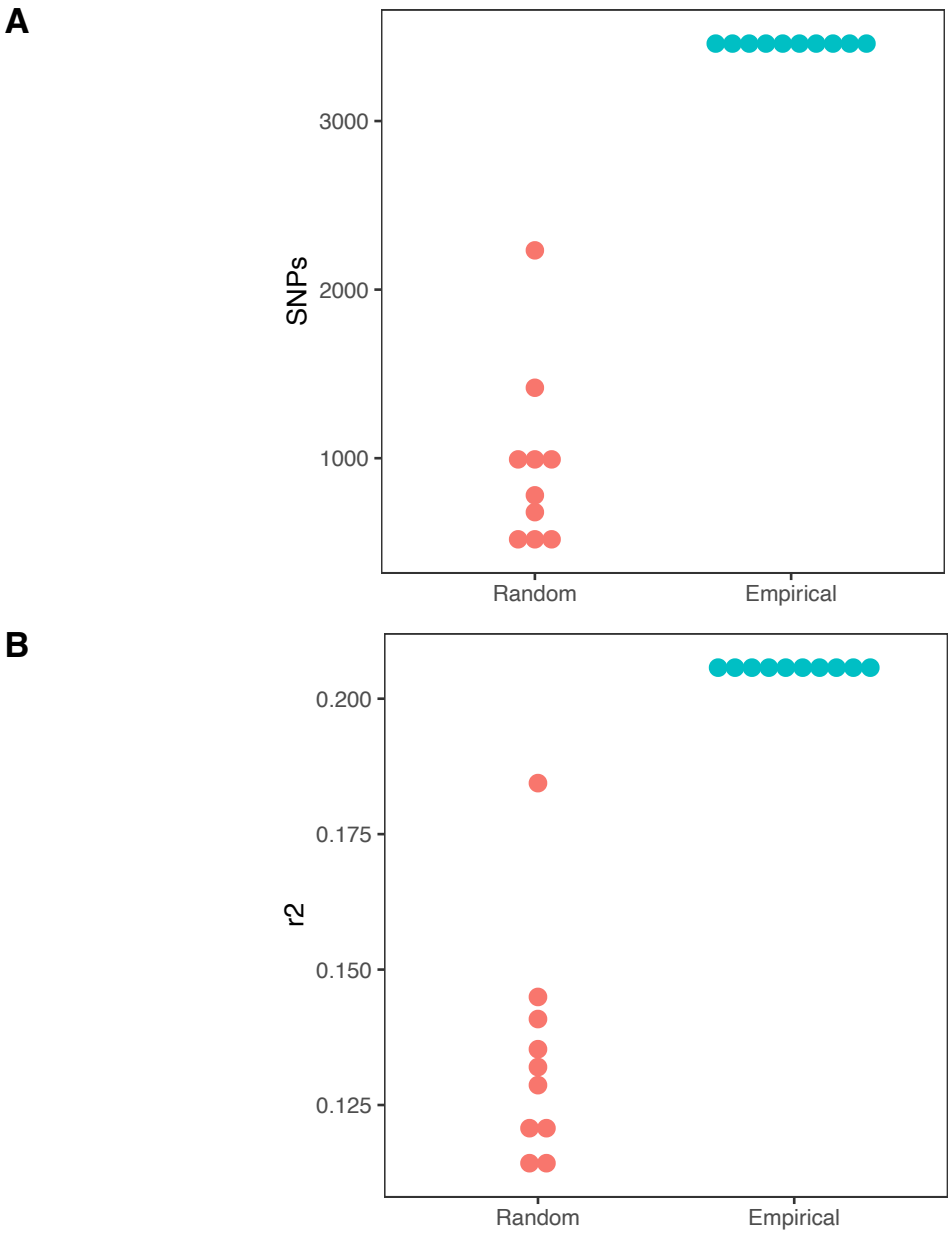


Figure S4. Results of comparisons between total SNPs with positive correlation coefficients (top) and average r2 of those SNPs (bottom) between 10 randomized and 10 empirical gradient forest analyses. Both plots illustrate consistency among empirical analyses, and clearly lower numbers of SNPs and r2s in randomized analyses.

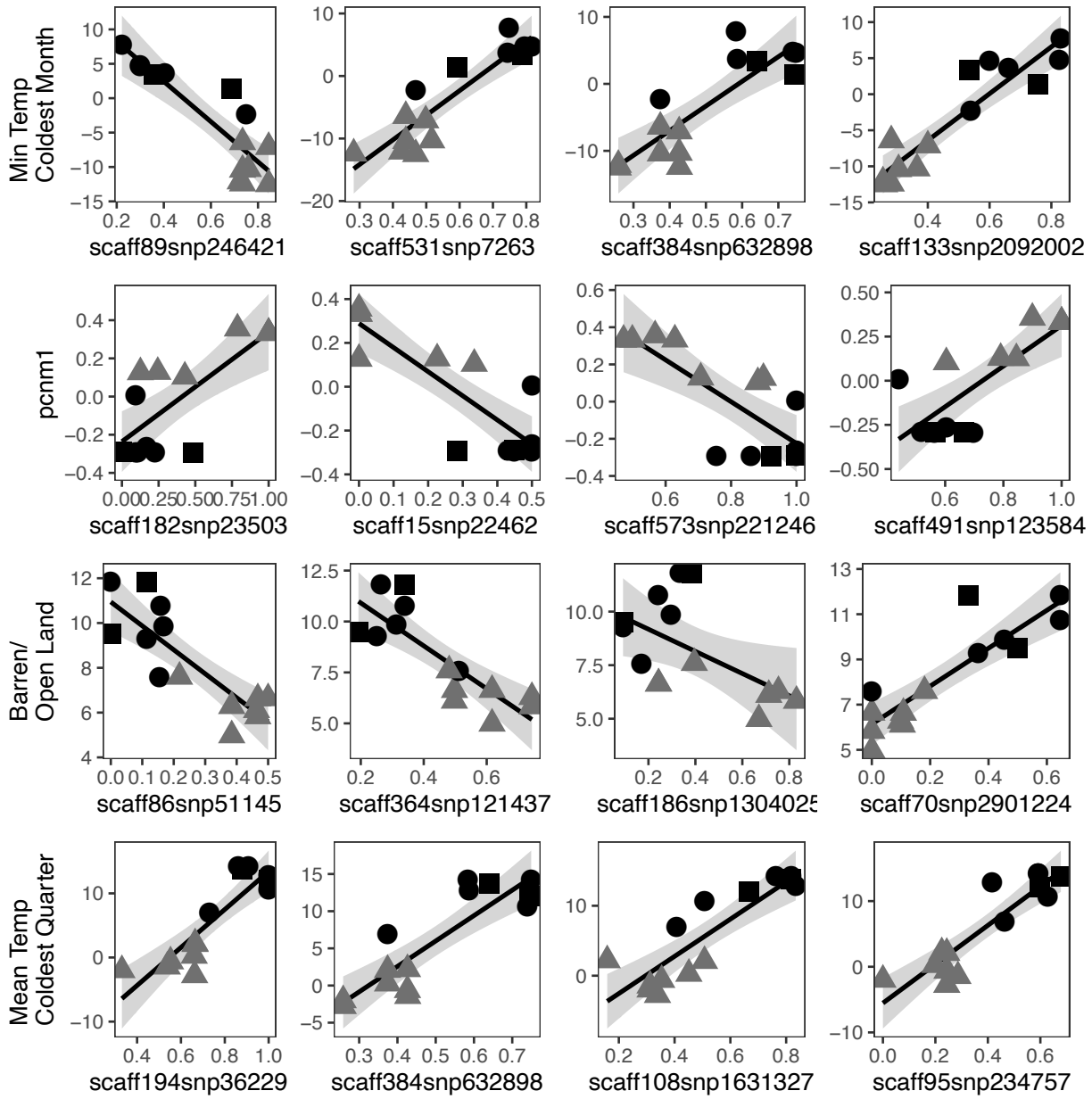


Figure S5. Minor allele frequency trends among resident (circles), migratory (triangles) and switcher (squares) breeding sites at the top four uncorrelated environmental variables identified in gradient forest analyses.

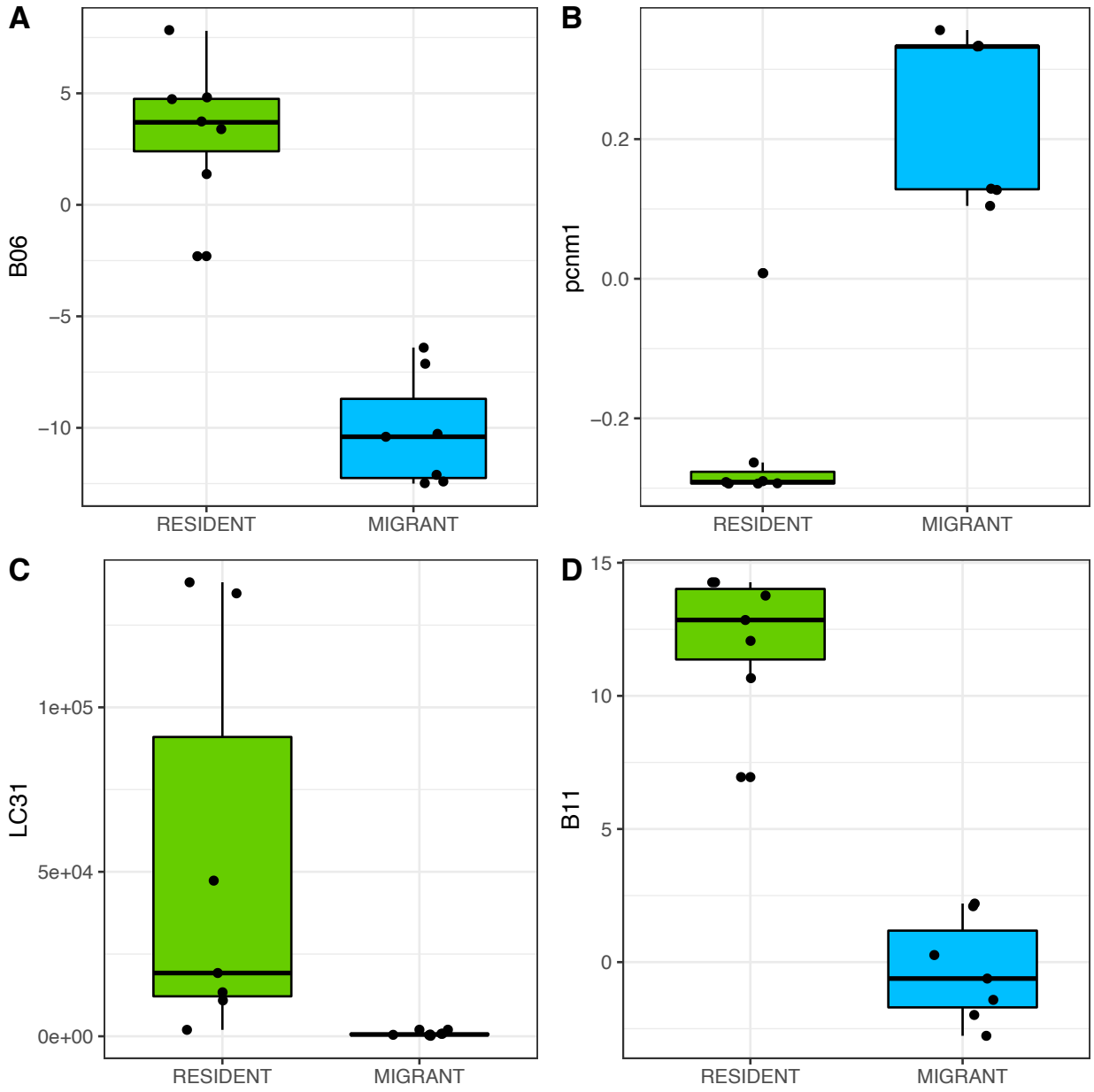


Figure S6. Summaries of top four environmental variables by site type. These do not include New Mexico as the site was excluded from gradient forest analyses. B06 and B11 are bioclimate variables representing, respectively, minimum temperature during the coldest month and mean temperature during the coldest quarter. LC31 is barren/open land as calculated within the National Landcover Dataset. PCNM1 is the first PC from a principal component analysis of neighborhood matrix.

SUPPLEMENTAL METHODS

Sample Collection

Tissue collection was made possible with the assistance of many collaborators (Table S1). We targeted seven resident sites and eight migratory breeding sites. Most sampling entailed capturing individuals in specialized traps (including bownet and box traps), and collecting 10 – 50 uL of blood with a non-heparinized capillary tube following a puncture of the brachial vein (Sheldon et al., 2008; Stangel, 1986) with a sterile 26G X 0.5 mm needle. Collected blood was stored either in Queen's lysis buffer, in EDTA, on filter paper treated with EDTA, or on Whatman blood cards. Two retrices (i.e., tailfeathers) or five breast feathers were also taken from most adults and these were added to the UCLA Center for Tropical Research's feather collection. Where possible (e.g., artificial burrows), we obtained growing feathers from nestlings and stored these in Queen's lysis buffer. We extracted DNA from tissues or snippets of Whatman blood cards using DNeasy extraction kits (Qiagen) following standard procedures, but with the addition of 20 uL of dithiothreitol (DTT) during tissue digestion for growing feathers. We quantified DNA concentrations using a Qubit (Thermofisher) and assessed quality on a 2% agarose gel.

Genome Assembly and Low Coverage Whole Genome Sequencing

Our genetic data includes a *de novo* reference genome for *A. c. hypugaea* and low coverage whole genome resequencing. For the former, we prepared DNA for whole genome sequencing using the Illumina TruSeq DNA PCR-Free LT kit (Illumina). After fragmenting 1 µg of DNA to 400 bp using a Diagenode sonicator and cleaning with magnetic beads at a ratio of 105 µL of beads/79 µL of water to select for >400 bp fragments, bioanalyzer traces were collected by the University of California, Los Angeles GenoSeq Core to verify library quality. We sequenced a final library with fragments averaging ~500 bp using a 250 bp paired-end run on

an Illumina HiSeq2500 at the University of California QB3 Vincent J. Coates Genomics Sequencing Laboratory. Additional sequence data was also collected using mate-pair libraries (Illumina). Two libraries with insert sizes 4kb and 8kb were prepared at the University of Utah Huntsman Cancer Center and 100 bp paired-end sequencing was run on one-third of an Illumina HiSeq 2500 lane. These data were trimmed and separated from reads sharing the same lane using NxTrim (O'Connell et al., 2015), and SSPACE (Boetzer et al., 2011) was then used to assemble the final scaffolds. Scaffolds were assembled from resulting sequence data via the Discovar DeNovo assembler (Broad Institute), and those <5,000 bp were removed because this improved N50 scores with little cost for assembly completeness. We used BUSCO (Simão et al., 2015) to estimate genome completeness by searching for single copy orthologs common to all species in the class Aves.

We also sequenced 202 samples to low coverage, targeting 1X. For these libraries, DNA concentrations were diluted to ≤ 2.5 ng/uL and prepared for sequencing using Nextera DNA Sample Preparation and Index Kits. DNA was fragmented and tagged with sequencing adaptors in a single step, and then a KAPA Library Amplification Kit (Roche) was used to attach indices. Libraries were cleaned using AMPure beads and 100 bp paired-end reads were sequenced on three lanes of on an Illumina HiSeq2500 at the University of California QB3 Vincent J. Coates Genomics Sequencing Laboratory. We removed duplicates using FastUniq (Xu et al., 2012) and used Trimmomatic (Bolger et al., 2014) to remove adaptors and low quality fragments (SLIDINGWINDOW: 4:15; MINLEN: 36). We then used FLASH (Magoč & Salzberg, 2011) to collapse overlapping reads into single reads and HISAT2 (Kim et al., 2019) to align these to the genome. The clipOverlap module in bamUtil (Jun et al., 2015) was used to clip overlapping

pairs, and BEDtools (Quinlan & Hall, 2010) was used for calculating coverage site-by-site for each individual.

List of environmental variables used in GEA analyses.

Environmental Data:

BIO1: Annual Mean Temperature
BIO2: Mean Diurnal Range
BIO3: Isothermality
BIO4: Temperature Seasonality
BIO5: Max Temperature of the Warmest Month
BIO6: Min Temperature of the Warmest Month
BIO7: Temperature Annual Range
BIO8: Mean Temperature of the Wettest Quarter
BIO9: Mean Temperature of the Driest Quarter
BIO10: Mean Temperature of the Warmest Quarter
BIO11: Mean Temperature of the Coldest Quarter
BIO12: Annual Precipitation
BIO13: Precipitation of the Wettest Month
BIO14: Precipitation of the Driest Month
BIO15: Precipitation Seasonality
BIO16: Precipitation of the Wettest Quarter
BIO17: Precipitation of the Driest Quarter
BIO18: Precipitation of the Warmest Quarter
BIO19: Precipitation of the Coldest Quarter
NDVI_Mean: Mean Normalized Difference Vegetation Index
NDVI StDev: Standard Deviation of Normalized Difference Vegetation Index
SRTM: Elevation
QuickSCAT: Surface Moisture
LC11: Open Water; areas of open water, generally with less than 25% cover of vegetation or soil.
LC12: Perennial Ice/Snow; areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
LC21: Developed, Open Space: areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
LC22: Developed, Low Intensity; areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units.
LC23: Developed, Medium Intensity: areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
LC31: Barren Land (Rock/Sand/Clay); areas of bedrock, desert pavement, scarps, talus,

slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.

LC41: Deciduous Forest; areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change.

LC42: Evergreen Forest; areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green foliage.

LC43: Mixed Forest; areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover.

LC51: Dwarf Scrub; Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.

LC52: Shrub/Scrub; areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.

LC71: Grassland/Herbaceous; areas dominated by gramanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.

LC72: Sedge/Herbaceous; Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.

LC73: Lichens; Alaska only areas dominated by fruticose or foliose lichens generally greater than 80% of total vegetation.

LC74: Moss; Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.

LC81: Pasture/Hay; areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.

LC82: Cultivated Crops; areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.

LC90: Woody Wetlands; areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

LC95: Emergent Herbaceous Wetlands; Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

REFERENCES

Alerstam, T., Hedenström, A., & Åkesson, S. (2003). Long-distance migration: Evolution and determinants. *Oikos*, *103*(2), 247–260. https://doi.org/10.1034/j.1600-0706.2003.12559.x

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86. https://doi.org/10.1126/science.aan4380

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., Jentoft, S., & Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, *119*(6), 418–428. https://doi.org/10.1038/hdy.2017.54

Berthold, P., & Pulido, F. (1994). Heritability of migratory activity in a natural bird population. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *257*(1350), 311–315. https://doi.org/10.1098/rspb.1994.0131

Berthold, P., & Querner, U. (1981). Genetic Basis of Migratory Behavior in European Warblers. *Science*, *212*(4490), 77–79. https://doi.org/10.1126/science.212.4490.77

Borrell, J. S., Zohren, J., Nichols, R. A., & Buggs, R. J. A. (2020). Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evolutionary Applications*, *13*(1), 161–175. https://doi.org/10.1111/eva.12883

Bossu, C. M., Heath, J. A., Kaltenecker, G. S., Helm, B., & Ruegg, K. C. (2022). Clock-linked genes underlie seasonal migratory timing in a diurnal raptor. *Proceedings of the Royal Society B*, *289*(1974), 20212507.

Bossu, C. M., Rodriguez, M., Rayne, C., Chromczak, D. A., Higgins, P. G., Trulio, L. A., & Ruegg, K. C. (n.d.). Genomic approaches to mitigating genetic diversity loss in declining populations. *Molecular Ecology*, *n/a*(n/a). https://doi.org/10.1111/mec.17109

Bracken, F. S. A., Hoelzel, A. R., Hume, J. B., & Lucas, M. C. (2015). Contrasting population genetic structure among freshwater-resident and anadromous lampreys: The role of demographic history, differential dispersal and anthropogenic barriers to movement. *Molecular Ecology*, *24*(6), 1188–1204. https://doi.org/10.1111/mec.13112

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*, 1–67. https://doi.org/10.18637/jss.v045.i03

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. https://doi.org/10.1111/1755-0998.12471

Chapman, J. W., Reynolds, D. R., & Wilson, K. (2015). Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters*, *18*(3), 287–302. https://doi.org/10.1111/ele.12407

Cox, G. W. (1985). The Evolution of Avian Migration Systems between Temperate and Tropical Regions of the New World. *The American Naturalist*, *126*(4), 451–474.

Delmore, K., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. *ELife*, *9*, e54462. https://doi.org/10.7554/eLife.54462

DeSante, D. F., Ruhlen, E. D., & Rosenberg, D. K. (2004). Density and abundance of burrowing owls in the agricultural matrix of the Imperial Valley, California. *Studies in Avian Biology*, *27*, 116–119.

Desmond, M. J., Parsons, T. J., Powers, T. O., & Savidge, J. A. (2001). An initial examination of mitochondrial dna structure in Burrowing Owl populations. *Journal of Raptor Research*, *35*(4), 274–281.

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

Dingle, H., & Drake, V. A. (2007). What Is Migration? *BioScience*, *57*(2), 113–121. https://doi.org/10.1641/B570206

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology*, *29*(24), 4765–4782. https://doi.org/10.1111/mec.15565

Dorman, M. (2018). *Nngeo: K-Nearest Neighbor Join for Spatial Data. R package version 0.2.4.* [Computer software].

Doublet, D. (2020). *Effects of Mitigation Translocation on Burrowing Owl Survival, Fidelity, and Reproduction in Arizona* [M.S., New Mexico State University]. https://www.proquest.com/docview/2412115874/abstract/8F2FD58F6F454F0EPQ/1

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168.

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M., & Kelling, S. (2020). Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecological Applications*, *30*(3), e02056. https://doi.org/10.1002/eap.2056

Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources*, *21*(8), 2749–2765. https://doi.org/10.1111/1755-0998.13374

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

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

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

Frias-Soler, R. C., Kelsey, N. A., Villarín Pildaín, L., Wink, M., & Bairlein, F. (2022). Transcriptome signature changes in the liver of a migratory passerine. *Genomics*, *114*(2), 110283. https://doi.org/10.1016/j.ygeno.2022.110283

Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Gómez-Bahamón, V., Márquez, R., Jahn, A. E., Miyaki, C. Y., Tuero, D. T., Laverde-R, O., Restrepo, S., & Cadena, C. D. (2020). Speciation Associated with Shifts in Migratory Behavior in an Avian Radiation. *Current Biology*, *30*(7), 1312-1321.e6. https://doi.org/10.1016/j.cub.2020.01.064

Guglielmo, C. G. (2018). Obese super athletes: Fat-fueled migration in birds and bats. *Journal of Experimental Biology*, *221*(Suppl_1), jeb165753. https://doi.org/10.1242/jeb.165753

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience*, *8*(5). https://doi.org/10.1093/gigascience/giz034

Hennessy, S. M., Wisinski, C. L., Ronan, N. A., Gregory, C. J., Swaisgood, R. R., & Nordstrom, L. A. (n.d.). Release strategies and ecological factors influence mitigation translocation outcomes for burrowing owls: A comparative evaluation. *Animal Conservation*, *n/a*(n/a). https://doi.org/10.1111/acv.12767

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Jacquard, A. (2012). *The Genetic Structure of Population* (Vol. 5). Springer Science & Business Media.

Jones, C. M., Papanicolaou, A., Mironidis, G. K., Vontas, J., Yang, Y., Lim, K. S., Oakeshott, J. G., Bass, C., & Chapman, J. W. (2015). Genomewide transcriptional signatures of

migratory flight activity in a globally invasive insect pest. *Molecular Ecology*, *24*(19), 4901–4911. https://doi.org/10.1111/mec.13362

Keller, L. F. (1998). Inbreeding and Its Fitness Effects in an Insular Population of Song Sparrows (melospiza Melodia). *Evolution*, *52*(1), 240–250. https://doi.org/10.1111/j.1558-5646.1998.tb05157.x

Kent JR, J. W., & Rankin, M. A. (2001). Heritability and physiological correlates of migratory tendency in the grasshopper Melanoplus sanguinipes. *Physiological Entomology*, *26*(4), 371–380. https://doi.org/10.1046/j.0307-6962.2001.00257.x

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., De Rosa, M. C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *25*(10), 2130–2143. https://doi.org/10.1111/mec.13592

Korfanta, N. M., McDonald, D. B., & Glenn, T. C. (2005). Burrowing Owl (Athene Cunicularia) Population Genetics: A Comparison of North American Forms and Migratory Habits. *The Auk*, *122*(2), 464–478. https://doi.org/10.1093/auk/122.2.464

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Leclerc, M., Leblond, M., Le Corre, M., Dussault, C., & Côté, S. D. (2021). Determinants of migration trajectory and movement rate in a long-distance terrestrial mammal. *Journal of Mammalogy*, *102*(5), 1342–1352. https://doi.org/10.1093/jmammal/gyab081

Lemopoulos, A., Uusi-Heikkilä, S., Huusko, A., Vasemägi, A., & Vainikka, A. (2018). Comparison of Migratory and Resident Populations of Brown Trout Reveals Candidate Genes for Migration Tendency. *Genome Biology and Evolution*, *10*(6), 1493–1503. https://doi.org/10.1093/gbe/evy102

Li, H., Handsaker, B., Danecek, P., McCarthy, S., & Marshall, J. (n.d.). *Bcftools: Utilities for variant calling and manipulating VCFs and BCFs*.

Macias-Duarte, A. (2011). *Change in Migratory Behavior as a Possible Explanation for Burrowing Owl Population Declines in Northern Latitudes*. https://repository.arizona.edu/handle/10150/145395

Macías-Duarte, A., Conway, C. J., & Culver, M. (2020). Agriculture creates subtle genetic structure among migratory and nonmigratory populations of burrowing owls throughout North America. *Ecology and Evolution*, *10*(19), 10697–10708. https://doi.org/10.1002/ece3.6725

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

Mueller, J. C., Carrete, M., Boerno, S., Kuhl, H., Tella, J. L., & Kempenaers, B. (2020). Genes acting in synapses and neuron projections are early targets of selection during urban colonization. *Molecular Ecology*, *29*(18), 3403–3412. https://doi.org/10.1111/mec.15451

Mueller, J. C., Pulido, F., & Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1719), 2848–2856. https://doi.org/10.1098/rspb.2010.2567

Narum, S. R., Di Genova, A., Micheletti, S. J., & Maass, A. (2018). Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883), 20180935. https://doi.org/10.1098/rspb.2018.0935

Poulin, R. G., Todd, L. D., Haug, E. A., Millsap, B. A., & Martell, M. S. (2020). Burrowing Owl (Athene cunicularia). *Birds of the World*. https://birdsoftheworld.org/bow/species/burowl/cur/introduction

Pulido, F. (2007). The Genetics and Evolution of Avian Migration. *BioScience*, *57*(2), 165–174. https://doi.org/10.1641/B570211

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ralls, K., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., Sunnucks, P., & Frankham, R. (2018). Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. *Conservation Letters*, *11*(2), e12412. https://doi.org/10.1111/conl.12412

Ramenofsky, M. (1990). Fat Storage and Fat Metabolism in Relation to Migration. In E. Gwinner (Ed.), *Bird Migration* (pp. 214–231). Springer. https://doi.org/10.1007/978-3-642-74542-3_15

Riding, C. S., & Belthoff, J. R. (2018). Breeding Dispersal By Burrowing Owls (Athene cunicularia) In Idaho. *Journal of Raptor Research*, *52*(2), 143–157. https://doi.org/10.3356/JRR-17-17.1

Rosenberg, D. K., & Haley, K. L. (2004). The ecology of burrowing owls in the agroecosystem of the Imperial Valley, California. *Studies in Avian Biology*, *27*, 120–135.

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., Paxton, E. H., & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096. https://doi.org/10.1111/ele.12977

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., Siegel, R. B., DeSante, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology*, *23*(23), 5726–5739. https://doi.org/10.1111/mec.12977

Ruegg, K. C., Brinkmeyer, M., Bossu, C. M., Bay, R. A., Anderson, E. C., Boal, C. W., Dawson, R. D., Eschenbauch, A., McClure, C. J. W., Miller, K. E., Morrow, L., Morrow, J., Oleyar, M. D., Ralph, B., Schulwitz, S., Swem, T., Therrien, J.-F., Van Buskirk, R., Smith, T. B., & Heath, J. A. (2021). The American Kestrel (Falco sparverius) genoscape: Implications for monitoring, management, and subspecies boundaries. *Ornithology*, *138*(2), ukaa051. https://doi.org/10.1093/auk/ukaa051

Shaw, A. K. (2016). Drivers of animal migration and implications in changing environments. *Evolutionary Ecology*, *30*(6), 991–1007. https://doi.org/10.1007/s10682-016-9860-5

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Slate, J., Kruuk, L. E. B., Marshall, T. C., Pemberton, J. M., & Clutton-Brock, T. H. (2000). Inbreeding depression influences lifetime breeding success in a wild population of red

deer (Cervus elaphus). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1453), 1657–1662. https://doi.org/10.1098/rspb.2000.1192

Strimas-Mackey, M. (2021). *Https://CRAN.R-project.org/package=smoothr* [Computer software].

Teitelbaum, C. S., Fagan, W. F., Fleming, C. H., Dressler, G., Calabrese, J. M., Leimgruber, P., & Mueller, T. (2015). How far to go? Determinants of migration distance in land mammals. *Ecology Letters*, *18*(6), 545–552. https://doi.org/10.1111/ele.12435

Thompson, N. F., Anderson, E. C., Clemento, A. J., Campbell, M. A., Pearse, D. E., Hearsey, J. W., Kinziger, A. P., & Garza, J. C. (2020). A complex phenotype in salmon controlled by a simple change in migratory timing. *Science*, *370*(6516), 609–613. https://doi.org/10.1126/science.aba9059

Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R., & Lovette, I. J. (2019). Selection on VPS13A linked to migration in a songbird. *Proceedings of the National Academy of Sciences*, *116*(37), 18272–18274. https://doi.org/10.1073/pnas.1909186116

Vanhove, M., Pina-Martins, F., Coelho, A. C., Branquinho, C., Costa, A., Batista, D., Príncipe, A., Sousa, P., Henriques, A., Marques, I., Belkadi, B., Knowles, L. L., & Paulo, O. S. (2021). Using gradient Forest to predict climate response and adaptation in Cork oak. *Journal of Evolutionary Biology*, *34*(6), 910–923. https://doi.org/10.1111/jeb.13765

Vieira, F. G., Fumagalli, M., Albrechtsen, A., & Nielsen, R. (2013). Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Research*, *23*(11), 1852–1861. https://doi.org/10.1101/gr.157388.113

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*(1), 35–48. https://doi.org/10.1111/mec.14954

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, *30*(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Winger, B. M., Auteri, G. G., Pegan, T. M., & Weeks, B. C. (2019). A long winter for the Red Queen: Rethinking the evolution of seasonal migration. *Biological Reviews*, *94*(3), 737–752. https://doi.org/10.1111/brv.12476

Winger, B. M., Barker, F. K., & Ree, R. H. (2014). Temperate origins of long-distance seasonal migration in New World songbirds. *Proceedings of the National Academy of Sciences*, *111*(33), 12115–12120. https://doi.org/10.1073/pnas.1405000111

Zhan, S., Zhang, W., Niitepõld, K., Hsu, J., Haeger, J. F., Zalucki, M. P., Altizer, S., de Roode, J. C., Reppert, S. M., & Kronforst, M. R. (2014). The genetics of monarch butterfly migration and warning colouration. *Nature*, *514*(7522), 317–321. https://doi.org/10.1038/nature13812

Chapter III

Title: Consistent Patterns of Panmixia in California's Near-Endemic Tricolored Blackbird in Multiple Genetic Datasets

Authors: Kelly R. Barr[1,2], Thomas B. Smith[1,2]

Contact Information:

[1]Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA

[2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA

ABSTRACT

The rise of genomics presents new and important avenues of research that can vastly improve both our understanding of the natural world and the conservation of it. Here, I draw on previous genetic research on the Tricolored Blackbird (*Agelaius tricolor*) using both microsatellites and RAD-Seq data to illustrate some of the improvements that may be attained using whole genome sequences (WGSs). I found that across datasets, including microsatellites, RAD-Seq, and now WGSs, the species has high enough gene flow range wide to prevent the development any significant genetic structure. Low Ne (~3,100) reported from the RAD-Seq data is confirmed here with WGSs (Ne=~2,700), and nucleotide diversity both is low across the genome and overall ($\prod$ = 0.0022) relative to other species. Whereas demographic analyses with the RAD-Seq data detected a single, very old population contraction, analyses using WGS data

found evidence for both a recent contraction on the time frame of the known 20th Century decline of the species and an ancient contraction. Overall, the results reported here provide further confidence in previous genetic analyses, and also improve our understanding of ongoing evolutionary processes impacting the Tricolored Blackbird. It is clear that the species can be managed overall as a single conservation unit, and there is no evidence that inbreeding is impacting the species.

INTRODUCTION

DNA sequencing and analyses have recently experienced major technological leaps. Whereas population genetic studies were previously designated as "genomic" when more than 50 genetic markers were used (Luikart et al. 2003), genomic research now frequently involves analyses of thousands or even millions of markers. Whole genome sequencing of numerous individuals allows for more advanced analyses than previously available for understanding the natural world, including detailed characterizations of genetic variation across the genome (Thompson et al. 2020), development of coalescent models for identifying historical demographic events and accounting for their impacts on current genetic diversity (Beichman et al. 2018), and inference of fine-scale, detailed gene flow patterns (Oyler-McCance et al. 2022). These advances have pushed conservation research and management into a higher echelon by providing far more granular information about inter- and intrapopulation evolutionary processes (Supple and Shapiro 2018; Zamudio 2023).

For decades, the field of conservation genetics could be distilled to essentially describing broadscale genetic patterns that have developed over hundreds, or often thousands of generations (King and Burke 1999; Moritz 1994). Much of this was focused on identifying "distinct

population segments" (DPUs) or "evolutionarily significant units" (ESUs) that were intended to better focus management and protection under the Endangered Species Act (Ryder 1986; Waples 1991). DPUs or ESUs are individual populations that are assumed to possess adaptive variation distinguishing them as unique products of evolutionary processes (Fallon 2007). As such, these populations are deemed acutely worthy of increased protection. One primary means for detecting ESUs was sequencing mitochondrial DNA and building genetic trees (Crandall et al. 2000). At the time, this was boon for conservation genetics and helped to protect many species populations; however, the ESUs detected through mitochondrial sequencing were limited to those formed after many generations of reproductive isolation that allowed reciprocal monophyly to form. This could take quite a while and is highly variable among species, as time to reciprocal monophyly is 2N for mitochondrial DNA (Moritz 1994). Also, processes contributing to the limitation of gene flow that leads to reciprocal monophyly among populations may have no bearing on underlying evolutionary dynamics and adaptation (Crandall et al. 2000; Funk et al. 2012).

Another advancement in sequencing technology allowed for the genotyping of highly polymorphic microsatellite markers, which increased statistical power and provided the opportunity to analyze more recently formed or finer-scaled patterns than previously possible in studies that only used mitochondrial sequences (Barr et al. 2011; Turlure et al. 2014; Flanders et al. 2009). Conservation genetics moved from the phylogeography era to broader applied population genetics. The field of landscape genetics developed as researchers could characterize how gene flow is impacted by recent processes such as human-caused habitat fragmentation (Lindsay et al. 2008; Manel et al. 2003) rather than just major vicariance events such as from orogeny or island formation. Microsatellites, however, also have many limitations. Elevated

mutation rates that contribute to high polymorphism among loci and enhanced statistical power for capturing recent or weak signals also leads to higher rates of homoplasy and bias in the calculation of primary statistical indices used for population genetics studies, such as F_{ST} and F_{IS} (Balloux et al. 2000; Putman and Carbone 2014). Importantly, inferences about adaptive variation in populations are also often inherently presumptive when relying on microsatellite data to detect genetic structure.

Characterizing genome-wide variation began in earnest with the rise of restriction enzyme-associated sequencing (RAD-Seq). RAD-Seq opened up the opportunity to examine variation patterns across the genome and proved particularly powerful when paired with a whole genome reference sequence that allows for the mapping and verification of variants (Shafer et al. 2017). As datasets turned from genetic to truly genomic in scope, processes may be tracked to specific genomic coordinates in non-model organisms (Bay et al. 2018; Thompson et al. 2020). Conservation efforts benefitted from this spectacular advancement as studies could move beyond the statistical limitations of microsatellite data to high-resolution genomic datasets (Seeb et al. 2011). Rather than suggesting adaptation may be occurring, genomic signatures of selection could be detected, mapped, and analyzed with much greater confidence (McKinney et al. 2017; Catchen et al. 2017). Now, researchers are no longer limited to detecting populations experiencing limited gene flow sufficient in time to allow for the development of reciprocal monophyly, and instead can directly identify populations with unique adaptive variation to be considered for management units (Funk et al. 2019) or other directed conservation effects such as assisted gene flow (Aitken and Bemmels 2016). RAD-Seq is limited, however, as a reduced representation data type (Lowry et al. 2017). Depending on the density of markers and the sizes of linkage blocks within species, the latter of which may be particularly small when gene flow is

high, important signatures of selection may be missed. This can be particularly problematic for populations experiencing soft sweeps (Ferrer-Admetlla et al. 2014).

Current population genomics moves beyond these limitations with whole genome sequences either at low depth, which are effective for analyses in genotypic likelihood frameworks (Lou et al. 2021), or high depth, which is necessary for calling individual genotypes with confidence (Jiang et al. 2019). Larger datasets may be filtered to the highest confidence variants for assessing genetic variation, differentiation, and demographic processes. Rather than relying on statistics like F_{IS} that are inherently dependent on data quality and are inferential about inbreeding, runs-of-homozygosity (RoHs) may be assessed to characterize with greater confidence levels of inbreeding individual-by-individual (Kardos et al. 2016). Signatures of selection can be assessed with far higher resolution, allowing for detecting more subtle processes such as weak selection or otherwise recently developing adaptation (Hohenlohe et al. 2021). Genetic diversity may be assessed across the genome to determine if acute effects are occurring that further threaten species or if local populations are genetically depauperate due to reduced population sizes and barriers to gene flow.

Tricolored Blackbirds (*Agelaius tricolor*) are one of the few non-model organisms that have been examined using both microsatellite (Berg et al. 2010) and RAD-Seq (Barr et al. 2021) approaches (but see Hauser et al. 2021). Plummeting populations through the 20th Century elicited the attention of conservation efforts and research. The species decline by an estimated 95% matches the loss of their primary nesting habitat, wetlands, in California (Beedy et al. 2018). Microsatellite and RAD-Seq datasets were developed for the species to assess current genetic diversity and population structure. Historical, large scale banding studies suggested the species has a barrier gene flow at the Transverse Ranges, with a northern grouping largely

centered in the Central Valley and a southern one extending to the international border with Mexico (DeHaven et al. 1975). As a consequence, these northern and southern groups are treated explicitly as separate entities for the triannual census of the species and for species management decisions. Using microsatellite data, Berg et al. (2010) found suggestive, if inconclusive, evidence for genetic structure in the species matching that of the prior banding studies based mainly upon significant differences in genetic diversity between the northern and southern portions of the species' range. More recently, Barr et al. (2021) reported the species to be in panmixia, with no evidence for genetic structure based on analyses of thousands of SNPs identified using RAD-Seq. The RAD-Seq study also found no evidence for selection, surprisingly low effective population size (Ne=~3,100) given recent census sizes (Nc=218,000; Meese 2022), and evidence for an old (T=~21,000 generations) population bottleneck that seemed to be far more important for explaining modern genetic diversity in the species than the known population decline of the 20th Century.

Here, I expand upon the previous genetic work on Tricolored Blackbirds with a novel, high-resolution reference genome (Ballare et al. 2023) and many (N=65) medium coverage (goal=10X) whole genome sequences. My goals are to: 1) illustrate the advances in analytical power and frameworks provided by the increased resolution of WGS beyond that of microsatellites and RAD-Seq data, 2) reassess previously reported patterns of genetic diversity and population structure in the species, and 3) improve upon the existing genomic resources to inform conservation efforts for the species. Results show that no matter the size of the dataset, the main findings remain the same. I find further Tricolored Blackbirds are in panmixia throughout their range and have a low current effective population size. Unlike the RAD-Seq dataset, demographic modeling using the whole genome sequencing dataset did detect the impact

of very recent population declines in the species. Further, differences in identity-by-descent segments (IBDsegs) and RoHs between northern and southern populations are suggestive of either weak barriers to gene flow or lingering historical genetic dynamics.

METHODS

Genome Sequencing and Data Filtering

Samples were collected as previously outlined in Barr et al. (2021) and Ballare et al. (2023), and the high quality reference genome was sequenced and assembled as detailed in the latter study. Seven samples per site for 10 sites spread throughout the Tricolored Blackbird's range (Fig. 1A) were chosen for whole genome resequencing based upon extraction quality, with a goal of 10X in coverage per sample for the resulting 70 samples. Libraries were prepared by the California Conservation Genomics Program (CCGP) Mini-Core at the University of California, Los Angeles (UCLA), and sequencing was conducted on an Illumina NovaSeq at the Vincent J. Coates Genomics Sequencing Lab at the University of California, Berkeley. Sequencing occurred across three lanes, as they were combined with samples from other projects for efficiency.

All initial processing and most of the following analyses were conducted on the Hoffman2 shared cluster at UCLA. I processed raw reads using standard procedures. Namely, raw reads were initially trimmed of sequencing adaptors using Trim Galore (Krueger 2015) and mapped to the reference genome using the 'mem' module in BWA (Li and Durbin 2010). I used samtools (Li et al. 2009) to sort the sequences by position and mark duplicated sequences.

Bird sex is determined as ZZ for males and ZW for females, and the Z chromosome is one of only a few macrochromosomes in the avian genome. Hence, biases may be introduced

into analyses as many variants may be homozygous or heterozygous based upon sex and coalescence times would differ from those of variants found on autosomes. For this reason, I used comparisons of coverage between males and females to determine the identities of sex chromosomes and excluded sex-linked loci from all analyses (See Supplemental Figs. 1 – 3).

Based upon wide variation in sequencing depths (see results), the following analyses were based upon either: 1) data that were downsampled to 2X using the 'DownsampleSam' module in Picard (https://broadinstitute.github.io/picard/) and mostly in genotype likelihoods using ANGSD (Korneliussen, Albrechtsen, and Nielsen 2014) or 2) a set limited to samples >10X in depth and with called genotypes using the 'HaplotypeCaller' module in GATK3.8 (DePristo et al. 2011). These are hereafter referred to as: 1) low coverage and 2) high coverage, respectively. Detected variants were filtered for most analyses as follows; but with exceptions as noted later in the methods.

Low Coverage

After downsampling, I used ANGSD to produce genotype likelihood files for all individuals in the BEAGLE format (-doGlf 3) and a minor allele frequency file (-domaf 1) with restrictive filtering that uses a conservative minimum minor allele frequency (-minmaf 0.05), discarding reads without unique mapping (-uniqueOnly 1), removing bad reads (-remove_bads 1), using only reads for which mates are mapped (-only_proper_pairs 1), discarding reads with low mapping quality (-minMapQ 1), keeping reads with high base quality (-minQ 20), keeping only biallelic sites (-skipTriallelic 1), using a high maximum likelihood of being polymorphic (-SNP_pval 1e-6), adjusting mapQ scores for excessive mismatches from the reference genome (-C 50), confirming variants using a base alignment quality estimation (-baq), and limiting variants to those present in half of the dataset (-minInd 33) and within a reasonable minimum and

maximum depth (-setMinDepth 25 -setMaxDepth 800). For F_{ST} analyses, I used ANGSD to create population-specific site frequency spectra (SFSs) from site allele frequency (SAF) files using the reference genome to polarize allele calls (-anc) and the same filtering conditions as previously outlined. NB, the -minInd filter was adjusted according to the number of individuals present in the dataset for each collection site.

High Coverage

For most of the high coverage analyses, I used vcftools (Danecek et al. 2011) to remove insert-deletion variants (--no-indels) and limit variants to a maximum missingness among samples of 20% (--max-missing 0.8), biallelic sites (--min-alleles 2 --max-alleles 2), minimum genotype quality scores of 30 (--minGQ 30), and minimum allele frequencies of 0.01 (--maf 0.01). Using VCF-kit (Cook and Andersen 2017), I limited variants to a maximum heterozygosity of 0.75 to avoid paralogs, and I used PLINK1.9 to both filter sites that were significantly out of Hardy Weinberg Equilibrium (--hwe midp 0.01) and prune sites for linkage disequilibrium (-indep 50 5 2).

Population Structure

Using multiple frameworks, I tested for indications of population structure. I used the R package 'srsStuff' (https://github.com/eriqande/srsStuff) for a principal components analysis (PCA). The package samples individual reads from each genomic site for pairs of samples in order to account for differences in sequencing depths. Data were prepared for this analysis following the same filtering conditions as outlined above for the high coverage samples, but with the following exceptions: 1) the analysis used the 2X downsampled dataset and hence included all samples, and 2) the filtering conditions used in vcftools were different (--maf 0.1 and –max-

missing 0.5). I used the Bayesian clustering algorithm ADMIXTURE (Alexander et al. 2009) to analyze the both the low and high coverage datasets for the presence of genetic clusters. For the former, I used the same filtering conditions as outline for the PCA. Using both vcftools and the realSFS module in ANGSD, I calculated pairwise F_{ST} between sample sites using the high coverage and low coverage datasets, respectively. Finally, I used IBDseq (Browning and Browning 2011) to detect long runs of identity-by-descent segments (IBDsegs) that would indicate recent gene flow among sites in the high coverage dataset. Using identified IBDsegs, I plotted a network using the R package 'iGraph' (Csardi 2006) to illustrate differences in connectivity among sites.

Genetic Diversity

I calculated numerous indices of genetic diversity using the high coverage data, including nucelotide diversity (∏), Tajima's D, contemporary effective population size (Ne), and runs-of-homozygosity (RoHs). For the first two indices, I used vcftools and the latter two Ne-Estimator 2.0 (Do et al. 2014) and the R package 'detectRuns' (CRAN), respectively. I plotted genomewide ∏ and Tajima's D calculated over 10,000 bp windows using the R package 'qqman' (CRAN). For Ne-Estimator, the high coverage data were filtered as outlined except no missing data were allowed (--max-missing 1); furthermore, given the total number of SNPs used in the analysis (see Results), I report the Ne at P_{crit}=0.01. I subsampled 20,000 SNPs five times to obtain a confidence interval for contemporary Ne. RoHs were calculated in sliding windows of 15 SNP, with minimum lengths of 250,000 bps, and otherwise default settings. I reviewed plots of locations in the genomes chromosome-by-chromosome, calculated F_{RoH} (ie, the sum of total

homozygous segments/total length of the genome), and used a Mann-Whitney U test to determine significance of comparisons.

Historical Demography

I examined effective population size and changes through multiple time frames using the high coverage data. In addition to contemporary Ne as described in the previous section obtained based on LD, I used the program GONE (Santiago et al. 2020) to estimate "recent" changes in Ne and fastsimcoal2 (Excoffier et al. 2021) to estimate significant changes in Ne over deep evolutionary time. For GONE, I ran analyses using default settings on the full high coverage dataset as well as just subsets including either north sites or southern sites. Data were prepared for faststimcoal2 as previously described except all non-sex-linked variants were retained for estimation of the site frequency spectrum. This method allows for the optimization of total SNPs to be retained using a projection as implemented in easySFS (https://github.com/isaacovercast/easySFS). Fastsimcoal2 was then run to compare three scenarios following Barr et al. (2021): 1) a no epoch, no population size change scenario, 2) a one epoch, single size change scenario, and 3) a two epoch, two size change scenario. Parameters were the same as Barr et al. (2020), and each scenario was simulated 1M times with 100 replicates of each and 50 ECM optimization cycles were used to determine best likelihood outcomes.

RESULTS

Sequencing Quality

After initial quality filters and removal of one mistakenly sampled Red-winged Blackbird (*Agelaius phoeniceus*), 65 Tricolored Blackbirds remained that were sequenced to a wide range of depths (Fig. 1; range: 0.05-27.92X). The lowest depths collected were from samples taken onto Whatman blood collection cards. These include sites represented as circles in Figure 2A. Triangles represent sites represented in analyses of data with >10X depth (ie, the "high coverage" dataset). Both triangles and circles were included in downsampled datasets (ie, the "low coverage" dataset). Except for the fastsimcoal2 analysis, the high coverage dataset included 3,433,002 SNPs. The BEAGLE produced for ANGSD analyses included 9,717,517 SNPs. The range of total SNPs retained in SAF files for populations was 1,140,909 – 1,908,867.

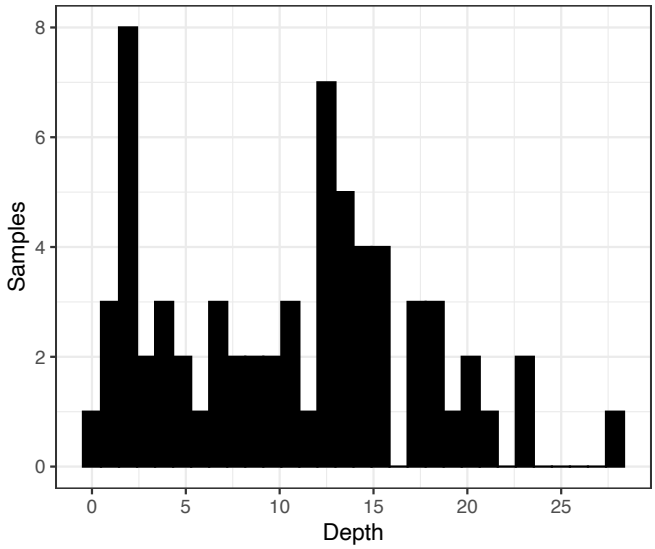


Figure 1. Depths of 65 Tricolored Blackbird samples sequenced for this study.

Genetic Structure

Multiple analyses for genetic structure indicate that Tricolored Blackbirds are in panmixia across their breeding range. These include low coverage analyses using a PCA (Fig 2B), which indicates little to no geographic clustering of individuals based upon shared allele frequencies, and ADMIXTURE (Fig. 3), which showed steadily increasing cross-validation scores from K=1. F_{ST}s based upon pairwise comparisons of sites in the high coverage dataset

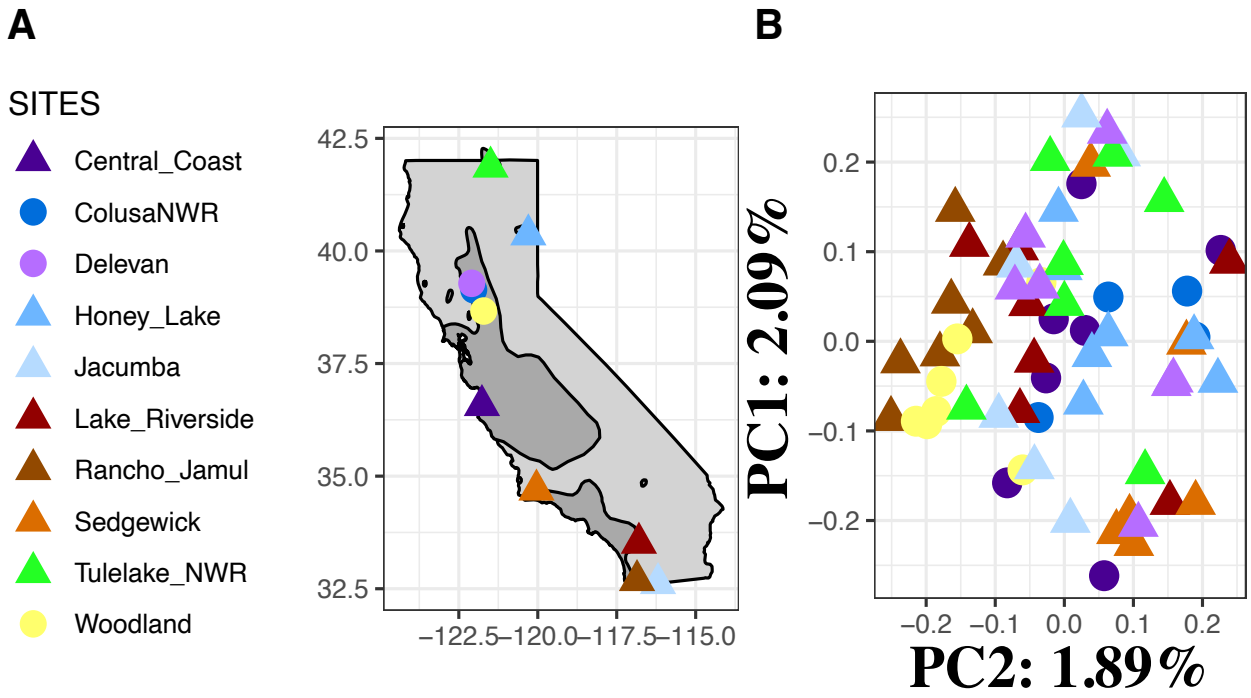


Figure 2. Map of sample sites (A) and PCA based upon single read sampling (B). Sites are demarked with circles and triangles. Triangles were included in >10X ("high coverage") analyses and circles exclusively in "low coverage" analyses. The PCA is based upon 34,000 SNPs that were strictly filtered for minor allele frequencies (--maf 0.1).

were low, with an average of 0.0014 (range: 0 – 0.0286; Supplemental Table 1). No pairwise populations in the low coverage dataset were differentiated by an F_{ST}>0 when analyzed with realSFS in ANGSD. Based upon shared IBDsegs among sites in the high coverage analysis (Fig. 4), however, genetic connectivity among the southern sample sites (ie, Sedgewick, Rancho Jamul, Jacumba, Lake Riverside) appears higher than among the northern sample sites (ie, Central Coast, Tulelake, Honey Lake). The 3,499 IBD segments detected among all samples were a median length of 241,743 bps, mean 378,853 bps, and ranged from 33,628 bps – 8,380,491 bps.

Genetic Diversity

Multiple indices indicate Tricolored Blackbirds as a species are relatively low in genetic diversity. Contemporary Ne based upon a strictly filtered dataset stands at a mean of 2709.66 (range: 2531.3 – 3305.3). ∏ and Tajima's D are relatively even across the genome, with no indications of outlying features that would be indicative of strong selection (Fig. 5A&B). Over 10,000 bp windows with more than 50 SNPs, mean ∏ was 0.0022 (range: 0.00027 – 0.011) and Tajima's D 0.506 (range: -2.595 – 4.35).

Inbreeding, on the other hand, does not seem to be developing in the species, though there were significant differences between northern and southern sampling locations. Average F_{RoH} across samples was 0.006 (range: 0.0006 – 0.0184). That for northern

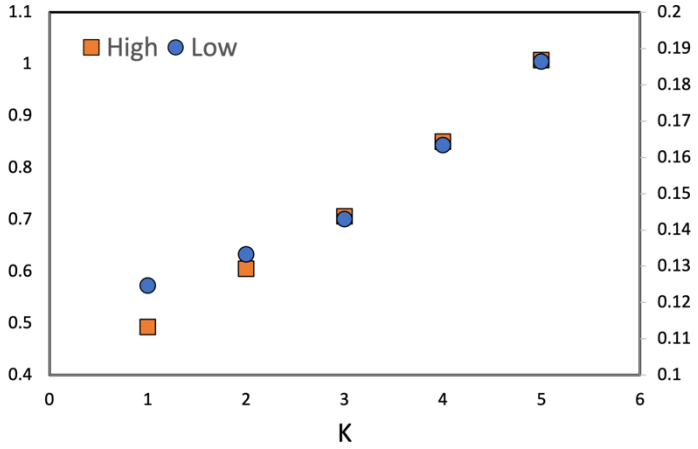


Figure 3. Results of admixture analyses for both low (blue circles) and high (orange squares) coverage datasets. Note increasing cross-validation scores indicate the best fitting model is K=1 for both datasets.

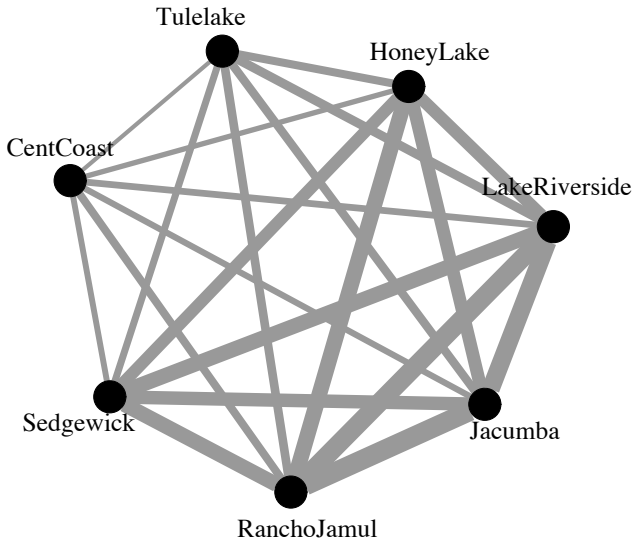


Figure 4. Network based upon shared isolation by descent segments (IBDsegs) among sites with samples at >10X. The thickness of the edges connecting nodes is scaled by the number of shared IBDsegs divided by the total number of samples between the two sites. Thinner lines mean for fewer shared IBDsegs and thicker more. Nodes are generally arranged by geographic relationships among sites.

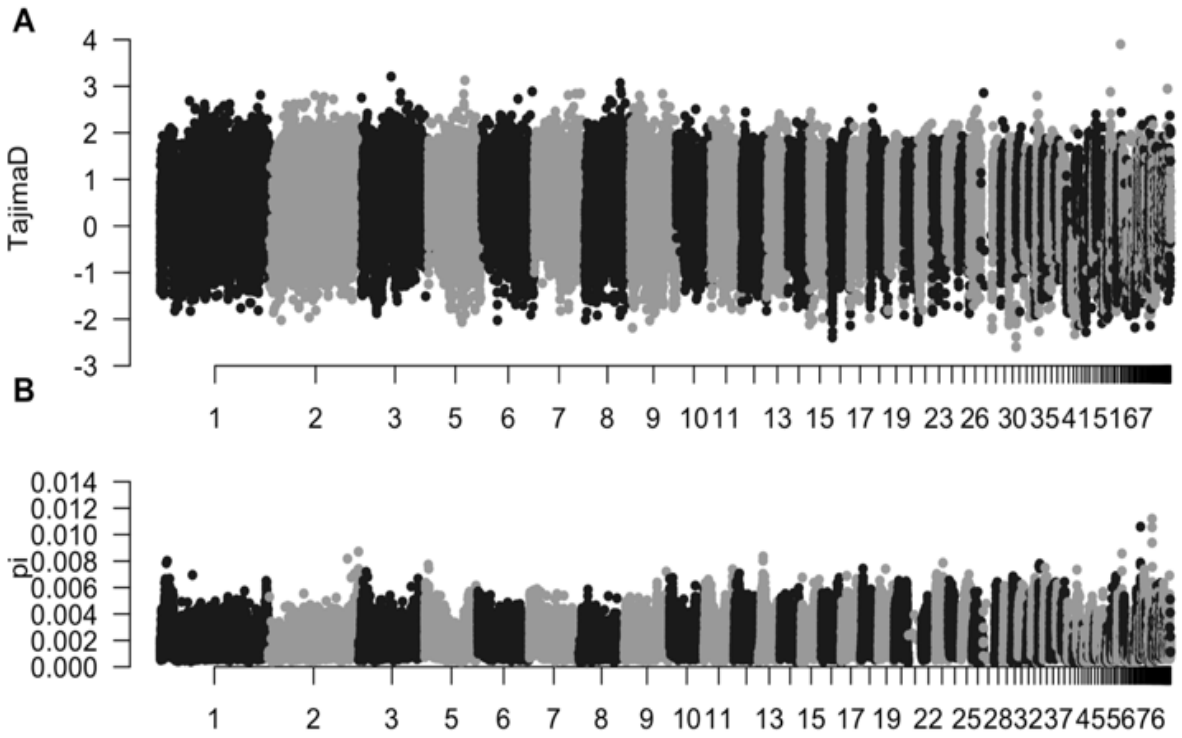


Figure 5. Genome-wide genetic diversity indices for sites with samples sequenced to >10X depth. A is Tajima's D and B is pi, both measured in 10,000 bp windows. Windows with >50 SNPs are plotted.

sample sites was 0.0044 (range: 0.0006 – 0.0144) and southern 0.0082 (range: 0.0016 – 0.0184).

Inbreeding was significantly higher in southern sites than northern sites (W=102, p=0.036; Fig. 6).

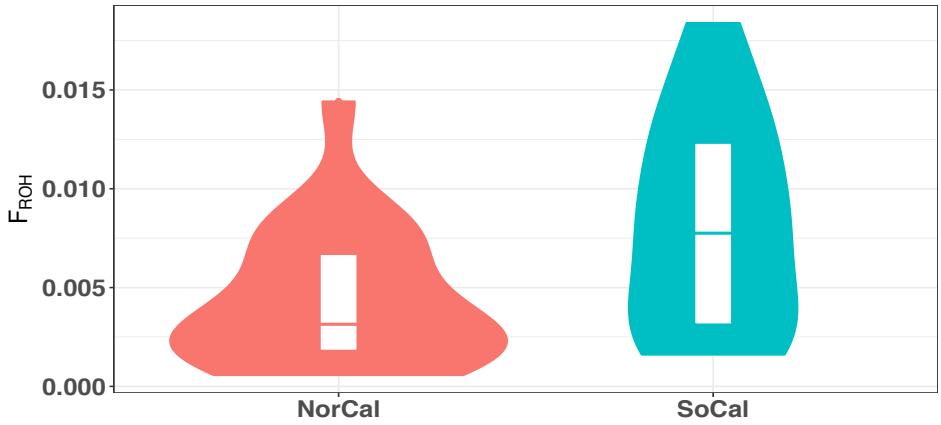


Figure 6. Violin plots summarizing F_{RoH} in northern and southern Tricolored Blackbird sample sites with high coverage data. The difference is significant (W=102,p=0.036).

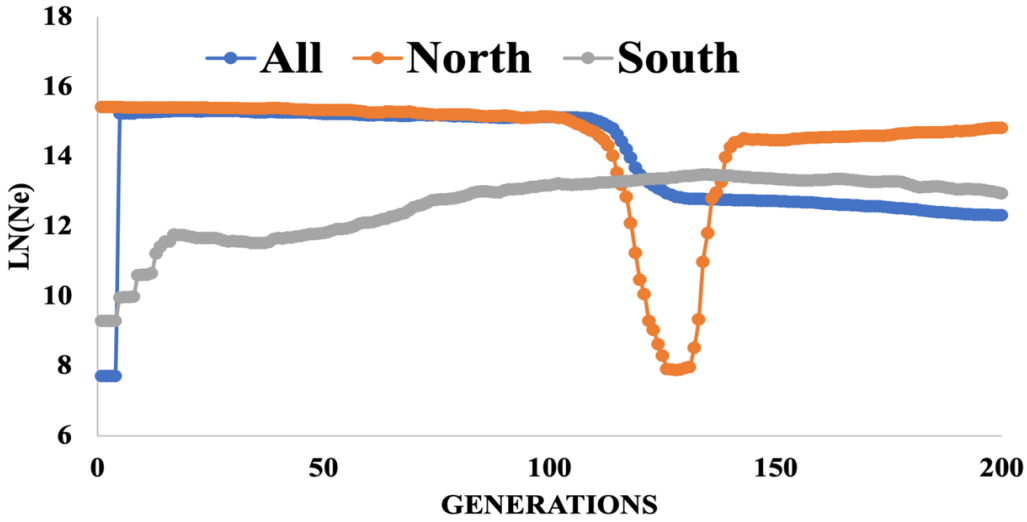


Figure 7. Recent Ne changes estimated using GONE. Analyses were conducted using the total high coverage dataset (all) and subsets of the northern sites (north) and southern sites (south).

Historical Demography

Multiple analytical frameworks suggest the species' recent decline has strongly impacted genetic diversity in Tricolored Blackbirds. When all high coverage samples are considered GONE detects a very recent decline in from a very large Ne (Fig. 7); however, when samples are split between north and south, there is a bottleneck signature detected ~120 generations ago in the former and a steady, low Ne with a recent decline in the latter. Meanwhile, analyses from fastsimcoal2 based on a projected SFS with 4,164,115 SNPs indicate the species has experienced population contractions that were both recent (NB: times are in generations (g); T1: mean = 50.6g, range = 8g – 168g) and deep (T2: mean = 3.45Mg, range = 3.28Mg – 3.59Mg) in the species' evolutionary history. The scenario of two contractions had a much higher likelihood

than those considering no decline or a single decline, and these two latter scenarios were similar in likelihood to one another (Fig. 8).

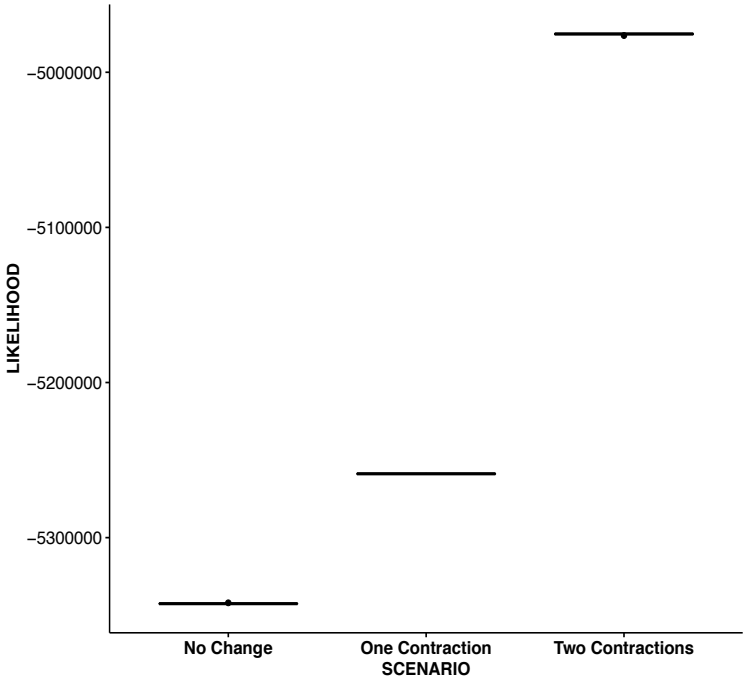


Figure 8. Comparisons of three historical demographic scenarios for Tricolored Blackbirds considered in fastsimcoal 2.0 simulations. 'No change' means the population size remained the same over time. 'One contraction' means the species experienced a single contraction. ‘Two contractions’ means the species experienced both recent and old contractions.

DISCUSSION

Many of the results I report here based upon whole genome sequences (WGSs) mirror those from previous microsatellite (Berg et al. 2010) and RAD-Seq (Barr et al. 2021) analyses. Namely, gene flow among sites is high with no major indications of limitations or the development of significant genetic structure. As observed with the RAD-Seq dataset, current effective population size (Ne) is low. Notably, however, genome-wide patterns of genetic diversity were accessible with the WGS dataset, and these provide further confidence in the results.

The WGS data, however, did reveal several novel results. First, whereas the RAD-Seq dataset suggested that a single, deeply historical population contraction was most impactful on genetic diversity in the species, with WGS data, I found that both recent and deep contractions were important. This matches what is known about the species, which has experienced an

estimated decline of 95% over the 20th Century. Notably, the WGS dataset had far more variants with >4M versus the 704,884 SNPs in the RAD-Seq fastsimcoal analysis (Barr et al. 2021) and hence there were many additional coalescence genealogies that provide more power for resolving demographic events.

Recent (ie, <200 generations) changes in Ne based on LD in GONE were less clear, but overall do indicate the species has experienced recent impacts on genetic diversity. It should be noted that without a linkage map, the timing of events detected in GONE are not accurate (Santiago et al. 2020). Overall, it seems that the species as a whole has experienced a very recent decline from a historically very large size to the quite small current size reported by both Ne-Estimator and GONE. This matches expectations given the species' population at the beginning of the 20th Century was an estimated 4M total birds; however, as noted in Barr et al. (2021), the current Ne is surprising given the recent census size of 218,000 birds (CI: 183,000 – 261,000; Meese 2022). The existence of a strong bottleneck in the northern portion of the range needs further examination to verify and identify a potential cause. One possible explanation is intense early century hunting, during which hundreds of birds at a time were killed at breeding colonies and sold in California bush meat markets. Neff (1942) reported that hunters took 400,000 Tricolored Blackbirds over five years from a small region of the Sacramento Valley, for instance.

Analyses of isolation by descent segments (IBDsegs) and runs-of-homozygosity (RoHs) also represent additions to our understanding of the species. Sharing of IBDsegs was greatest among sites sampled around southern California and lowest among sites in the far north, and F_{RoH} was significantly higher in the southern sites. Gene flow among southern sites appears to be higher than among northern sites and lower between north and south—which notably is consistent with Berg et al. (2010) and banding studies. The lack of gene flow with the northern

sites might explain the higher levels of inbreeding observed in the southern sites. Tricolored Blackbird populations in southern California are quite small with a total of an estimated 6,000 birds (Meese 2022).

Relevancy of Classical Genetic Markers

The availability of relatively cheap whole genome sequencing and a rapidly growing library of well-annotated reference genomes might suggest that the use of microsatellites and perhaps even RAD-Seq datasets is no longer relevant for conservation genetics. Indices such as F_{ST} and F_{RoH} are better quantified with a large number of biallelic loci spread across the genome, such as those available from RAD-Seq or WGS, versus microsatellites that are less precise for the former and cannot be used for the latter. While neither inbreeding nor selective sweeps were evident in Tricolored Blackbirds, WGS undoubtedly improves upon the accuracy and understanding of the impacts of these processes on the genome—indeed, mapping the genomic locations of RoHs, for instance, can be exceptionally helpful for improving the genetic health of populations experiencing intense conservation attention, such as through captive breeding (Bossu et al. 2023). In a comparison of microsatellites versus SNPs, however, Hauser et al. (2021) noted that in a genetically depleted species, the Black-capped Vireo (*Vireo atricapilla*), the former marker type provided the variation needed for parentage analyses that was not available in the latter. The authors attributed this to the lack of variation in the SNP dataset due to the lack of heterozygous sites. This issue may be averted using WGS that would allow for rigorous filtering, such as I used here for PCA and estimation of Ne in Ne-Estimator, and to retain enough markers with the information content needed for a parentage analysis.

Conservation Considerations

There are several important conservation implications from my analyses for Tricolored Blackbirds. As results show, the species should be managed as a single unit as now three different datasets had no evidence for notable genetic structure. Though the species has experienced a decline in genetic diversity, inbreeding is quite low especially compared to laboratory experiments that show that severe inbreeding depression effects occur at a threshold of F=0.2 (Hemmings et al. 2012). While surprisingly low given recent census sizes, Ne is well above the (albeit somewhat specious) thresholds of 50 to avoid inbreeding and 500 to avoid loss of genetic diversity through genetic drift (Frankham et al. 2017). π is also surprisingly low especially when compared to other species that are much smaller in population size and have experienced significant bottleneck events, such as the island fox (*Urocyon littoralis*; Robinson et al. 2016). These results indicate the primary focus for the Tricolored Blackbird should be to increase the overall population size with no regard for genetic connectivity or any particular focus on local breeding groups. Future genomic research capitalizing on the considerable population of preserved Tricolored Blackbird specimens in natural history collections would be helpful for further characterizing the effects of population decline and changes in drivers of selection on genomewide diversity in the species.

REFERENCES

Aitken, Sally N., and Jordan B. Bemmels. 2016. "Time to Get Moving: Assisted Gene Flow of Forest Trees." *Evolutionary Applications* 9 (1): 271–90. https://doi.org/10.1111/eva.12293.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Ballare, Kimberly M, Merly Escalona, Kelly Barr, William Seligmann, Samuel Sacco, Ruta Madhusudan Sahasrabudhe, Oanh Nguyen, Christy Wyckoff, Thomas B Smith, and Beth Shapiro. 2023. "A Reference Genome Assembly of the Declining Tricolored Blackbird, Agelaius Tricolor." *Journal of Heredity* 114 (1): 44–51. https://doi.org/10.1093/jhered/esac053.

Balloux, François, Harald Brunner, Nicolas Lugon-Moulin, Jacques Hausser, and Jéarôme Goudet. 2000. "MICROSATELLITES CAN BE MISLEADING: AN EMPIRICAL AND SIMULATION STUDY." *Evolution* 54 (4): 1414–22. https://doi.org/10.1111/j.0014-3820.2000.tb00573.x.

Barr, Kelly, Annabel C. Beichman, Pooneh Kalhori, Jasmine Rajbhandary, Rachael A. Bay, Kristen Ruegg, and Thomas B. Smith. 2021. "Persistent Panmixia despite Extreme Habitat Loss and Population Decline in the Threatened Tricolored Blackbird (Agelaius Tricolor)." *Evolutionary Applications* 14 (3): 674–84. https://doi.org/10.1111/eva.13147.

Barr, Kelly R., Giri Athrey, Denise L. Lindsay, Richard F. Lance, Timothy J. Hayden, Scott A. Tweddale, and Paul L. Leberg. 2011. "Missing the Forest for the Gene Trees:

Conservation Genetics Is More than the Identification of Distinct Population Segments." *The Auk* 128 (4): 792–94. https://doi.org/10.1525/auk.2011.128.4.792.

Bay, Rachael A., Ryan J. Harrigan, Vinh Le Underwood, H. Lisle Gibbs, Thomas B. Smith, and Kristen Ruegg. 2018. "Genomic Signals of Selection Predict Climate-Driven Population Declines in a Migratory Bird." *Science* 359 (6371): 83–86. https://doi.org/10.1126/science.aan4380.

Beedy, Edward C., William J. Hamilton, III, Robert J. Meese, Daniel A. Airola, and Peter Pyle. 2018. "Tricolored Blackbird (Agelaius Tricolor)." Edited by Paul G. Rodewald. *The Birds of North America Online*, July. https://doi.org/10.2173/bna.tribla.03.1.

Beichman, Annabel C., Emilia Huerta-Sanchez, and Kirk E. Lohmueller. 2018. "Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms." *Annual Review of Ecology, Evolution, and Systematics* 49 (1): 433–56. https://doi.org/10.1146/annurev-ecolsys-110617-062431.

Berg, Elena C., John P. Pollinger, and Thomas Benjamin Smith. 2010. "Population Structure of the Tricolored Blackbird (Agelaius Tricolor) in California : Are Northern and Southern Populations Genetically Distinct ?" In .

Browning, Brian L., and Sharon R. Browning. 2011. "A Fast, Powerful Method for Detecting Identity by Descent." *American Journal of Human Genetics* 88 (2): 173–82. https://doi.org/10.1016/j.ajhg.2011.01.010.

Catchen, Julian M., Paul A. Hohenlohe, Louis Bernatchez, W. Chris Funk, Kimberly R. Andrews, and Fred W. Allendorf. 2017. "Unbroken: RADseq Remains a Powerful Tool for Understanding the Genetics of Adaptation in Natural Populations." *Molecular Ecology Resources* 17 (3): 362–65. https://doi.org/10.1111/1755-0998.12669.

Cook, Daniel E, and Erik C Andersen. 2017. “VCF-Kit: Assorted Utilities for the Variant Call Format.” *Bioinformatics* 33 (10): 1581–82. https://doi.org/10.1093/bioinformatics/btx011.

Crandall, Keith A., Olaf R. P. Bininda-Emonds, Georgina M. Mace, Robert K. Wayne, Keith A. Crandall, Olaf R. P. Bininda-Emonds, Georgina M. Mace, et al. 2000. “Considering Evolutionary Processes in Conservation Biology.” *Trends in Ecology & Evolution* 15 (7): 290–95. https://doi.org/10.1016/S0169-5347(00)01876-0.

Csardi, G. 2006. “The Igraph Software Package for Complex Network Research.” *InterJournal Complex Syst*, 1695.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

DeHaven, Richard W., Frederick T. Crase, and Paul P. Woronecki. 1975. “Movements of Tricolored Blackbirds Banded in the Central Valley of California, 1965-1972.” *Bird-Banding* 46 (3): 220–29. https://doi.org/10.2307/4512139.

DePristo, Mark A., Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data.” *Nature Genetics* 43 (5): 491–98. https://doi.org/10.1038/ng.806.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. “NeEstimator v2: Re-Implementation of Software for the Estimation of Contemporary Effective Population Size (Ne) from Genetic Data.” *Molecular Ecology Resources* 14 (1): 209–14. https://doi.org/10.1111/1755-0998.12157.

Excoffier, Laurent, Nina Marchi, David Alexander Marques, Remi Matthey-Doret, Alexandre Gouy, and Vitor C Sousa. 2021. "Fastsimcoal2: Demographic Inference under Complex Evolutionary Scenarios." *Bioinformatics* 37 (24): 4882–85. https://doi.org/10.1093/bioinformatics/btab468.

Fallon, Sylvia M. 2007. "Genetic Data and the Listing of Species under the U.S. Endangered Species Act." *Conservation Biology: The Journal of the Society for Conservation Biology* 21 (5): 1186–95. https://doi.org/10.1111/j.1523-1739.2007.00775.x.

Ferrer-Admetlla, Anna, Mason Liang, Thorfinn Korneliussen, and Rasmus Nielsen. 2014. "On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure." *Molecular Biology and Evolution* 31 (5): 1275–91. https://doi.org/10.1093/molbev/msu077.

Flanders, Jon, Gareth Jones, Petr Benda, Christian Dietz, Shuyi Zhang, Gang Li, Mozafar Sharifi, and Stephen J. Rossiter. 2009. "Phylogeography of the Greater Horseshoe Bat, Rhinolophus Ferrumequinum: Contrasting Results from Mitochondrial and Microsatellite Data." *Molecular Ecology* 18 (2): 306–18. https://doi.org/10.1111/j.1365-294X.2008.04021.x.

Frankham, Richard, Jonathan D. Ballou, Katherine Ralls, Mark Derek Bruce Eldridge, Michele R. Dudash, Charles B. Fenster, Robert C. Lacy, and Paul Sunnucks. 2017. *Genetic Management of Fragmented Animal and Plant Populations*. Oxford University Press.

Funk, W. C., Brenna R. Forester, Sarah J. Converse, Catherine Darst, and Steve Morey. 2019. "Improving Conservation Policy with Genomics: A Guide to Integrating Adaptive Potential into U.S. Endangered Species Act Decisions for Conservation Practitioners and

Geneticists." *Conservation Genetics* 20 (1): 115–34. https://doi.org/10.1007/s10592-018-1096-1.

Funk, W. Chris, John K. McKay, Paul A. Hohenlohe, and Fred W. Allendorf. 2012. "Harnessing Genomics for Delineating Conservation Units." *Trends in Ecology & Evolution* 27 (9): 489–96. https://doi.org/10.1016/j.tree.2012.05.012.

Hauser, Samantha S., Giridhar Athrey, and Paul L. Leberg. 2021. "Waste Not, Want Not: Microsatellites Remain an Economical and Informative Technology for Conservation Genetics." *Ecology and Evolution* 11 (22): 15800–814. https://doi.org/10.1002/ece3.8250.

Hemmings, N.L., J. Slate, and T.R. Birkhead. 2012. "Inbreeding Causes Early Death in a Passerine Bird." *Nature Communications* 3 (1): 863. https://doi.org/10.1038/ncomms1870.

Hohenlohe, Paul A., Susan Bassham, Paul D. Etter, Nicholas Stiffler, Eric A. Johnson, and William A. Cresko. 2010. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLOS Genetics* 6 (2): e1000862. https://doi.org/10.1371/journal.pgen.1000862.

Hohenlohe, Paul A., W. Chris Funk, and Om P. Rajora. 2021. "Population Genomics for Wildlife Conservation and Management." *Molecular Ecology* 30 (1): 62–82. https://doi.org/10.1111/mec.15720.

Jiang, Yifan, Yao Jiang, Sheng Wang, Qin Zhang, and Xiangdong Ding. 2019. "Optimal Sequencing Depth Design for Whole Genome Re-Sequencing in Pigs." *BMC Bioinformatics* 20 (1): 556. https://doi.org/10.1186/s12859-019-3164-z.

Kardos, Marty, Helen R. Taylor, Hans Ellegren, Gordon Luikart, and Fred W. Allendorf. 2016. "Genomics Advances the Study of Inbreeding Depression in the Wild." *Evolutionary Applications* 9 (10): 1205–18. https://doi.org/10.1111/eva.12414.

King, Tim L., and Terry Burke. 1999. "Special Issue on Gene Conservation: Identification and Management of Genetic Diversity." *Molecular Ecology* 8 (s1): S1–3. https://doi.org/10.1046/j.1365-294X.1999.00893.x.

Korneliussen, Thorfinn Sand, Anders Albrechtsen, and Rasmus Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15 (1): 356. https://doi.org/10.1186/s12859-014-0356-4.

Krueger, Felix. 2015. "Trim Galore!: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data." *Babraham Institute*. https://cir.nii.ac.jp/crid/1370294643762929691.

Li, Heng, and Richard Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform." *Bioinformatics* 26 (5): 589–95. https://doi.org/10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078. https://doi.org/10.1093/bioinformatics/btp352.

Lindsay, Denise L., Kelly R. Barr, Richard F. Lance, Scott A. Tweddale, Timothy J. Hayden, and Paul L. Leberg. 2008. "Habitat Fragmentation and Genetic Diversity of an Endangered, Migratory Songbird, the Golden-Cheeked Warbler (Dendroica

Chrysoparia).” *Molecular Ecology* 17 (9): 2122–33. https://doi.org/10.1111/j.1365-294X.2008.03673.x.

Lou, Runyang Nicolas, Arne Jacobs, Aryn P. Wilder, and Nina Overgaard Therkildsen. 2021. “A Beginner’s Guide to Low-Coverage Whole Genome Sequencing for Population Genomics.” *Molecular Ecology* 30 (23): 5966–93. https://doi.org/10.1111/mec.16077.

Lowry, David B., Sean Hoban, Joanna L. Kelley, Katie E. Lotterhos, Laura K. Reed, Michael F. Antolin, and Andrew Storfer. 2017. “Breaking RAD: An Evaluation of the Utility of Restriction Site-Associated DNA Sequencing for Genome Scans of Adaptation.” *Molecular Ecology Resources* 17 (2): 142–52. https://doi.org/10.1111/1755-0998.12635.

Luikart, Gordon, Phillip R. England, David Tallmon, Steve Jordan, and Pierre Taberlet. 2003. “The Power and Promise of Population Genomics: From Genotyping to Genome Typing.” *Nature Reviews Genetics* 4 (12): 981–94. https://doi.org/10.1038/nrg1226.

Manel, Stéphanie, Michael K. Schwartz, Gordon Luikart, and Pierre Taberlet. 2003. “Landscape Genetics: Combining Landscape Ecology and Population Genetics.” *Trends in Ecology & Evolution* 18 (4): 189–97. https://doi.org/10.1016/S0169-5347(03)00008-9.

McKinney, Garrett J., Wesley A. Larson, Lisa W. Seeb, and James E. Seeb. 2017. “RADseq Provides Unprecedented Insights into Molecular Ecology and Evolutionary Genetics: Comment on Breaking RAD by Lowry et al. (2016).” *Molecular Ecology Resources* 17 (3): 356–61. https://doi.org/10.1111/1755-0998.12649.

Moritz, Craig. 1994. “Defining ‘Evolutionarily Significant Units’ for Conservation.” *Trends in Ecology & Evolution* 9 (10): 373–75. https://doi.org/10.1016/0169-5347(94)90057-4.

Neff, Johnson A. 1942. “Migration of the Tricolored Red-Wing in Central California.” *The Condor* 44 (2): 45–53. https://doi.org/10.2307/1364223.

Oyler-McCance, Sara J., Todd B. Cross, Jeffery R. Row, Michael K. Schwartz, Dave E. Naugle, Jennifer A. Fike, Kristopher Winiarski, and Brad C. Fedy. 2022. "New Strategies for Characterizing Genetic Structure in Wide-Ranging, Continuously Distributed Species: A Greater Sage-Grouse Case Study." *PLOS ONE* 17 (9): e0274189. https://doi.org/10.1371/journal.pone.0274189.

Putman, Alexander I., and Ignazio Carbone. 2014. "Challenges in Analysis and Interpretation of Microsatellite Data for Population Genetic Studies." *Ecology and Evolution* 4 (22): 4399–4428. https://doi.org/10.1002/ece3.1305.

Robinson, Jacqueline A., Diego Ortega-Del Vecchyo, Zhenxin Fan, Bernard Y. Kim, Bridgett M. vonHoldt, Clare D. Marsden, Kirk E. Lohmueller, and Robert K. Wayne. 2016. "Genomic Flatlining in the Endangered Island Fox." *Current Biology* 26 (9): 1183–89. https://doi.org/10.1016/j.cub.2016.02.062.

Ryder, OA. 1986. "Species Conservation and Systematics : The Dilemma of Subspecies." *Trends Ecol. Evol.* 1: 9–10.

Santiago, Enrique, Irene Novo, Antonio F Pardiñas, María Saura, Jinliang Wang, and Armando Caballero. 2020. "Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium." *Molecular Biology and Evolution* 37 (12): 3642–53. https://doi.org/10.1093/molbev/msaa169.

Seeb, J. E., G. Carvalho, L. Hauser, K. Naish, S. Roberts, and L. W. Seeb. 2011. "Single-Nucleotide Polymorphism (SNP) Discovery and Applications of SNP Genotyping in Nonmodel Organisms." *Molecular Ecology Resources* 11 (s1): 1–8. https://doi.org/10.1111/j.1755-0998.2010.02979.x.

Shafer, Aaron B. A., Claire R. Peart, Sergio Tusso, Inbar Maayan, Alan Brelsford, Christopher W. Wheat, and Jochen B. W. Wolf. 2017. "Bioinformatic Processing of RAD-Seq Data Dramatically Impacts Downstream Population Genetic Inference." *Methods in Ecology and Evolution* 8 (8): 907–17. https://doi.org/10.1111/2041-210X.12700.

Supple, Megan A., and Beth Shapiro. 2018. "Conservation of Biodiversity in the Genomics Era." *Genome Biology* 19 (1): 131. https://doi.org/10.1186/s13059-018-1520-3.

Thompson, Neil F., Eric C. Anderson, Anthony J. Clemento, Matthew A. Campbell, Devon E. Pearse, James W. Hearsey, Andrew P. Kinziger, and John Carlos Garza. 2020. "A Complex Phenotype in Salmon Controlled by a Simple Change in Migratory Timing." *Science* 370 (6516): 609–13. https://doi.org/10.1126/science.aba9059.

Turlure, Camille, Sofie Vandewoestijne, and Michel Baguette. 2014. "Conservation Genetics of a Threatened Butterfly: Comparison of Allozymes, RAPDs and Microsatellites." *BMC Genetics* 15 (1): 114. https://doi.org/10.1186/s12863-014-0114-7.

Waples, Robin S. 1991. "Definition of'species' under the Endangered Species Act: Application to Pacific Salmon."

Zamudio, Kelly R. 2023. "Conservation Genomics: Current Applications and Future Directions." *Journal of Heredity* 114 (4): 297–99. https://doi.org/10.1093/jhered/esad019.