AN EVOLUTIONARY GENOMICS STUDY FOR CONSERVATION OF THE MONTEZUMA QUAIL

by

Samarth Mathur

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biological Sciences

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. James Andrew DeWoody, Chair
Department of Forestry and Natural Resources
Department of Biological Sciences

Dr. Ximena E. Bernal
Department of Biological Sciences

Dr. Mark R. Christie
Department of Biological Sciences
Department of Forestry and Natural Resources

Dr. John M. Tomeček
Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station

Dr. Jeffrey R. Lucas
Department of Biological Sciences

Approved by:
Dr. Janice P. Evans

To my grandmas, Pushpa Mathur and Sudha Mathur

"Deyr fé,
deyja frændur,
deyr sjálfur ið sama;
ek veit einn at aldri deyr,
dómr um dauðan hvern."

"Animals die,
friends die,
and thyself, too, shall die;
but one thing I know that never dies
the tales of the one who died."
- Hávamál (Gestaþáttur, 77)

ACKNOWLEDGMENTS

I want to begin by sincerely thanking my research advisor Dr. J. Andrew DeWoody for his constant support, guidance, motivation, and most importantly, patience. Your belief in my scientific curiosities and in my capabilities to pursue them has undeniably been steadfast throughout the years, even if my performance hasn't. Your mentoring skills are exceptional and your encouragement invigorating. Thank you for adopting me into your lab and for constantly reminding me why it's called a doctorate of "philosophy".

I am also deeply indebted to Dr. Terry Blankenship, Dr. Selma Glasscock, and Welder Wildlife Foundation for awarding me their prestigious Graduate Research Fellowship and supporting me throughout my PhD. I would also like to thank my committee members Dr(s) Ximena Bernal, John M. Tomeček, and Mark Christie for their feedback and advice in defining my research goals. I would also like to thank our collaborators: Dr. Ryan Luna, Dr. Luis A. Tarango-Arámbula, Robert Perez, Kristyn Stewart, and Dr. Fidel Hernandez for their contribution. I would also like to thank Dr. Michael R. Gribskov and Phillip SanMiguel for providing me additional computational resources that were invaluable to my research. I am also indebted to entire team at Purdue Genomics Core Facility, ITaP Research Computing team, and folks at the AgIT department for their services and constant help with troubleshooting throughout my research.

People say education begins at home, and I have been so fortunate to have been raised by my supportive and loving parents, Soma and Sanjiva Mathur. You are my bedrock, my guardian angels, and my strength. Your adoration and care inspire me every day to be a better person. I want to thank my entire family for keeping me in their thoughts and prayers and sending me wishes and blessings throughout the years. I want to especially thank my brother, Sankalp Dayal, for being the textbook definition of ideal and making me realize how much discipline I still need in my life. Also, thank you Sugandha for making sure he doesn't ignore my texts. I am privileged to have found two brothers at Purdue: Dr. Gurneet Sangha and Dr. Richard Lie, who raised my spirits and lit my life. The adventures we had, and will continue to have, are sacred to me.

A person is defined by the company they keep, and my company makes me feel richer than Zuckerberg's. Anna Brüniche Olsen, Jennifer Antonides, Janna Willoughby, Nadia Fernandez, Rian Bylsma, Ashlyn Heniff, Morgan Elizabeth Kelly, Andrew Black, Andrew Mularo, and Erangi Heenkenda are some of the best people I know, and I am lucky to have shared my work and life

with you all. Brandon Quinby and Wes Flynn, you both have been there when I needed you the most, and for that, I am forever grateful. A special shout out to my academic rival and friend, Avril Harder; let us always keep pushing each other to greatness.

Everyone needs a friend who's there even when you don't need anyone, who's never out of reach, and in whose company, time loses its meaning. I am so fortunate to have friends like Eash Verma, Aashish Jain, Ankita Thawani, Sonali Srijan, Bethany and Val Schull, Mohammad Tayyab Adil, Surbhi Jain, Paul Lengemann, and Apeksha Yajnik who color my otherwise gray world. I am forever obliged to Mrigaunk Pillai for keeping me in his company. A huge holler to my bois: Abhishek Rao, Abin Ghosh, Amber Chugh, Chirag Jain, Mit Kotecha, Meet Mukesh Paswan, Rahul Gupta, Randhir Raj Singh, Siddhant Gupta, and Vineet Dalal. Cheers to our lifelong comradery.

I want to thank my friends and colleagues in Ecology and Evolutionary Biology (EEB) program and the administrative staff at the department of Biology for always having my back. Thank you for all the free coffee, lunches, and banter. A huge thanks to Georgina Rupp and Jennifer Spitzsnagle for helping me with the paperwork and untangling me from the webs of bureaucracy these past five years.

Lastly, I want to thank my past mentors and friends who gave me the right opportunities and encouragements that are still propelling me upwards even today. Dr. Naveen K. Navani, Dr. Ranjana Pathania, Dr. N.K. Goel thank you for your guidance and support; and Dr. Manasi Gupta, Dr. Tamoghna Ghosh, and Dr. Atin Sharma for all the good and educational times at the CBDD lab.

PREFACE

The overarching goal of this dissertation is to highlight how genetics and genomics are important tools with immediate applicability in wildlife conservation and a source to understand the evolutionary history of a species and its current risks for better management planning. This thesis is divided into three chapters. Each chapter is an independent publication unit. Chapter one of this dissertation has been published in peer-reviewed journal and chapter two has gone through peer review and is currently under revisions. The last chapter is currently under preparation for submission. Chapter 1 focuses on creating genetic assessment tools (genotyping arrays) using whole genome sequence data and their application to estimate population genetic metrics for different populations. Chapter 2 focuses on using low coverage whole genome data from multiple individuals to further validate the findings from Chapter 1 and compare genetic diversity between genic and inter-genic regions. From these two chapters, I establish that isolated populations suffer from loss of genetic diversity at a genome-wide scale, are significantly more inbred and are possibly locally adapted to its microhabitat. Moreover, initial whole genome analysis from multiple populations suggest that small populations survive at low genetic diversity and avoid inbreeding depression by purging the load of highly deleterious mutations (Chapter 2). These findings are further examined using deeper sequences from even more samples and populations in Chapter 3. Chapter 3 focuses on creating demographic models that shape the contemporary genomic diversity patterns first observed in Chapter 1 and Chapter 2. Additionally, I compare the distributions of deleterious mutations in small versus large populations to show that small populations are less efficient in purging newly rising deleterious mutations and thus, more likely to suffer from inbreeding depression.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Humans have altered natural landscape since the agricultural revolution, but it has been most destructive since human globalization and rampant industrialization in the last two centuries. These activities deteriorate and fragments natural habitat of many wild species that creates small isolated populations that lose genetic diversity over time. Loss of genetic diversity reduces the adaptive capacity of a population to respond to future environmental change and increases their extinction risks. Implementing strategies for wildlife conservation is a challenge primarily because of our lack of understanding of the biology of many wild species, the risks they are currently facing, and their evolutionary histories. With the advent of genomic and computational techniques, it is now possible to address these concerns. In my research, I used genomics to study the evolutionary history of the Montezuma Quail (*Cyrtonyx montezumae*) and created monitoring tools that can be readily applied by wildlife managers for its conservation. Montezuma Quail is a small gamebird found mostly in Mexico with peripheral populations existing in Arizona, New Mexico, and Texas. Montezuma Quail are going through species wide decline in the United States and are listed as vulnerable in the state of Texas due to their small population sizes and geographic isolation from rest of the range. My results show that Texas quail are genetically distinct and significantly less diverse than Arizona quail. Analysis of whole genome sequences from multiple individuals show that due to small population sizes and isolation, Texas quail are significantly more inbred and genetic drift is the major contributor for loss of genetic diversity we see today. Inbreeding is negatively impacting Texas quail as they carry more deleterious alleles within their genome that reduce fitness of the individuals. Demographic models predict that both Arizona and Texas populations were formed via founding bottlenecks around 20,000 years ago. Texas populations have maintained small population sizes since its split from the ancestral populations and are less efficient in purging new deleterious mutations that arise post-bottleneck. The inferences from my research not only carries direct implications for Montezuma Quail conservationists, but also illustrate the power of evolutionary genomics in implementing targeted management strategies for any species that face existential threats in today's waning world.

CHAPTER 1. THE GENOME AND POPULATION STRUCTURE OF MONTEZUMA QUAIL

The first chapter of this dissertation had three main goals: (i) to characterizing the Montezuma Quail genome which can be used as reference for future Montezuma Quail studies, (ii) to create a genotyping panel that can be used for individual identification of field collections and monitor genetic health of populations, and (iii) use the panel to study the population variation and structure of the all Montezuma Quail populations in the U.S: Arizona, New Mexico, and Texas. The results from our genotypic array showed that Texas quail are genetically distinct and less diverse at key fitness genes. This chapter contains published content from the journal *Conservation Genetics* entitled, "Evidence of genetic erosion in a peripheral population of a North American game bird: the Montezuma quail (*Cyrtonyx montezumae*)".

The published version of this chapter available at https://doi.org/10.1007/s10592-019-01218-9

1.1 Abstract

Population extirpations are often precursors to species extinctions. Anthropogenic activities often lead to smaller populations that are more prone to extirpations and advocates for active conservation management have recently called for the preservation and monitoring of genetic diversity, particularly with regard to the adaptive potential of vulnerable populations. We used genomics and curated arrays of molecular markers, including those expected to impact key fitness traits, to quantify evidence of genomic erosion in core and peripheral populations of a gallinaceous bird. The Montezuma quail (*Cyrtonyx montezumae*) is a game species considered vulnerable to extirpation in Texas, but core populations in Arizona and New Mexico are robust and have the potential to serve as genetic reservoirs. We sequenced the Montezuma quail genome then developed a single nucleotide polymorphism (SNP) assay to quantify genetic variation, effective population sizes, signatures of natural selection, and population structure. We genotyped SNPs from gene deserts and from genes associated with fitness traits and found the isolated Texas population exhibits an extremely small effective population size, is genetically distinct from our Arizona and New Mexico samples, and has reduced heterozygosity at the fitness-related markers. Thus, our samples from Texas exhibit symptoms of genetic erosion that could exacerbate future

risk of local extirpation. Management agencies must decide if active conservation efforts such as assisted gene flow or genetic rescue are now warranted. This decision may not be straightforward because the current conservation status of the Texas population reflects its isolated geographic locale on the periphery of the species' range.

1.2 Introduction

Peripheral populations often arise due to natural processes such as irregular dispersal and landscape heterogeneity, but they can also be due to anthropogenic activities (e.g. van der Valk et al. (2018)). Peripheral populations are often prone to extinction due to natural metapopulation dynamics (Hampe and Petit 2005), but they are biologically important because geographic range margins often harbor populations that will presumably be the source for future range expansion under climate change. Phylogeographic studies have shown that such peripheral populations are mostly constrained to specific habitats within unsuitable and heterogeneous landscapes (Hampe and Petit 2005). Populations on the edge of a species' range can suffer from long standing isolation from the core populations and thus, can exhibit high levels of genetic differentiation (Martin and McKay 2004). These populations may contain important genetic variants that are not present in the core of the range(Garner, Pearman, & Angelone, 2004; Hunter & Hutchinson, 1994; Lesica & Allendorf, 1995). If these polymorphisms include adaptive variation, then they contribute to the evolutionary potential of the species' gene pool(Barbosa et al., 2018; Lesica & Allendorf, 1995).

Population declines are often associated with the loss of genetic diversity that can lead to a feedback loop or vortex (Li et al., 2016; Spielman, Brook, & Frankham, 2004). Some authors have argued that the detection, monitoring, and preservation of genetic diversity in small isolated populations should be a primary goal of conservation (Hampe & Jump, 2011) because they are at extreme risk of extirpation and a lack of genetic information thwarts effective conservation efforts (Ralls et al., 2018). Conservation efforts usually target non-game species, but legally hunted species also warrant monitoring, especially peripheral or isolated populations (e.g., Epps, Crowhurst, and Nickerson (2018)).

The Montezuma quail (*Cyrtonyx montezumae*), sometimes called Mearns' quail, is a small, social, gallinaceous bird that mostly inhabits the oak-grassland landscapes in mountain ranges and sky islands of Mexico and the southwestern United States (Leopold and McCabe (1957); Figure 1.1) from 1500 to 1800 meters above sea level (Albers & Gehlbach, 1990). Like many

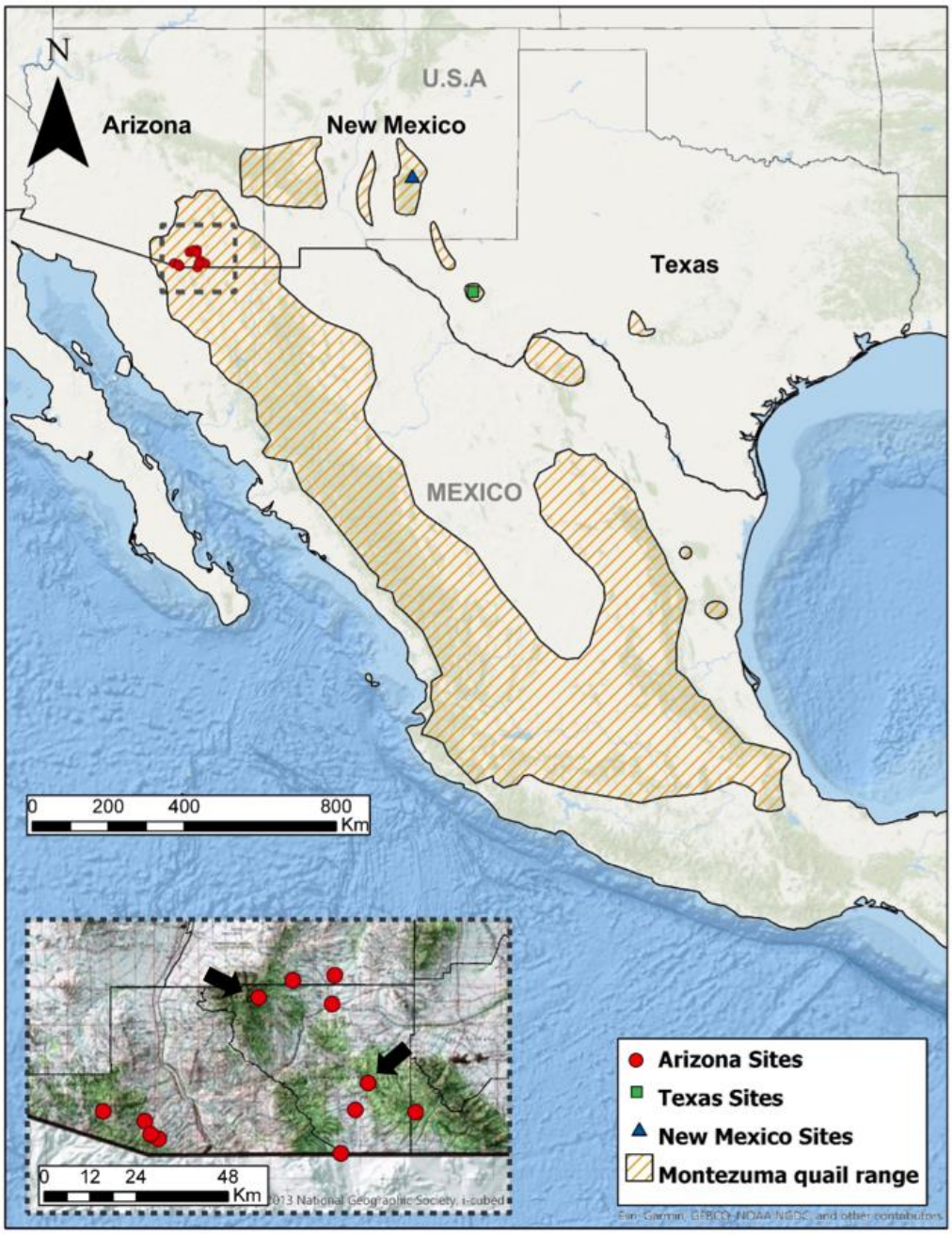


Figure 1.1: Map showing the contemporary geographic range of Montezuma Quail (shaded) and all collection sites, with an inset of Arizona (the source of most samples). Montezuma Quail samples were collected from hunter harvested wings in Arizona (red dots), voucher specimens from New Mexico (blue triangle), and road-kills in Texas (green square). The locations of samples used for genome sequencing are indicated with black arrows. The range data were collected from IUCN Red List of Threatened Species Data (IUCN 2018) and the map was generated using ArcGIS® Pro.

galliformes, their foraging habits and population dynamics are greatly influenced by stochastic environmental factors such as precipitation (Heffelfinger & Olding, 2000). Montezuma quail in the U.S. are considered a game bird in some states, subject to legal sport hunting (e.g., in Arizona and New Mexico), but there has been no open hunting season in Texas for more than a decade due to conservation concerns. The Montezuma quail is categorized as Vulnerable by the Texas Parks and Wildlife Department due to its restricted geographic range and recent widespread population declines that have been attributed to habitat degradation, habitat fragmentation, and overgrazing (Albers & Gehlbach, 1990; Brown, 1982; Rollins, 2002). Although relatively common in large portions of its range, this bird is extremely difficult to trap and monitor due to its tuberiferous diet (i.e., trapping with grain as bait is ineffective) and thus its biology is relatively unknown. A 28 August 2018 search of ISI's Web of Science using "Montezuma quail" or "*Cyrtonyx montezumae*" as the search terms yielded a total of only 9 published scientific papers on the species.

In this study, we aimed to create genetic resources that a) are robust enough for monitoring using opportunistically collected samples (e.g., roadkill, remains collected from hunters, etc.); and b) would enable us to identify populations that exhibit signatures of genomic erosion (Leroy et al., 2018). Herein, we characterized the genome of the Montezuma quail and then estimated its genetic diversity and population structure by genotyping individuals from different geographic regions within the United States. We queried a SNP panel that includes genetic markers from both annotated genes and gene deserts (i.e., intergenic regions) and subsequently genotyped Montezuma quail samples from Arizona, New Mexico, and Texas to characterize their genetic diversity and genetic structure. We frame our study in light of the degree of fragmentation within each state. The Arizona habitat is generally contiguous with the Mexican range of the species and supports a large, robust population of Montezuma quail. The Texas habitat is highly fragmented and peripheral populations are small, whereas in New Mexico the habitat is more disjunct and populations are likely intermediate in size (Figure 1.1). Finally, we interpret these data in light of the geographical distribution of functional genetic diversity (i.e., in genic SNPs) and the conservation status of the Texas birds.

1.3 Materials and methods

1.3.1 Sampling, DNA extraction, and genome sequencing

Montezuma quail tissue samples were collected from sites in Arizona, New Mexico, and Texas (Figure 1.1). This species is notoriously difficult to live trap (Hernandez, Garza, Harveson, & Brewer, 2009), so our sampling was largely opportunistic. Arizona samples were acquired from hunter harvested wings collected between November 2007 to February 2008. In contrast, New Mexico samples were collected from voucher specimens and Texas samples were collected as road-killed carcasses. All acquired samples had associated geographic co-ordinates of capture locations. Muscle tissues were isolated from each sample and were subsequently frozen at -80˚C until DNA was extracted.

Genomic DNA was extracted using a potassium acetate protocol (Nicholls, Double, Rowell, & Magrath, 2000), then purified using the Zymo® Genomic DNA Clean and Concentrator Kit. For genome sequencing, two Arizona samples (1 female E8452 and 1 male E8454) were individually barcoded, then both paired-end (2 x 100bp read lengths) and mate-pair libraries were constructed from each using Illumina® PCR free prep kits. Mate-pair libraries were constructed from 2-4kb fragments excised from an agarose gel. Samples were sequenced using 2 lanes of Illumina® HiSeq2500 platform, with one lane each for paired-end and mate pair libraries. Illumina® adapter sequences and low-quality bases (Phred < 20) were trimmed using Trimmomatic v.036 (Bolger, Lohse, & Usadel, 2014) prior to genome assembly and subsequent bioinformatics analyses. The quality of the reads was inferred by generating summary statistics using FastQC v0.11.7 (Andrews & others, 2010).

1.3.2 Genome assemblies, filtering, and annotation

The mitochondrial genome was assembled using MITObim 1.6 (Hahn, Bachmann, & Chevreux, 2013), using a baiting and iterative mapping approach. A draft mitochondrial assembly was created using the *Cyrtonyx montezumae* cytochrome b (cytb) gene (AF068192.1) as bait and was annotated using MITOS (Bernt et al., 2013).

The genome size was estimated with the Jellyfish (Marçais & Kingsford, 2011) k-mer counting method using the optimal k-mer of 60. Nuclear genome assemblies were constructed for both male and female individuals using multiple *de novo* assemblers (Supplementary Methods).

The most comprehensive assembly was chosen for downstream analyses by comparing assembly completeness using BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) and assembly statistics such as N50 using Quast (Gurevich, Saveliev, Vyahhi, & Tesler, 2013).

Only the male nuclear assembly was used for gene annotation. Prior to annotation, the assembly was filtered for xenobiotic sequences (i.e., exogenous sequences not of avian origin from pathogens, parasites, commensals, etc.) according to Antonides, Ricklefs, and DeWoody (2017) (Appendix A) and repetitive elements were masked. The assembly was annotated using the iterative MAKER 2.31 (Holt & Yandell, 2011) pipeline (Appendix A).

1.3.3 SNP identification and assay design

SNPs were identified using the assembly with contigs greater than 500bp using Genome Analysis Toolkit (GATK) pipeline following the "best practices protocol" (McKenna et al. (2010); Appendix A). Genomic nucleotide diversity was estimated from the SNPs identified using VCFtools v0.1.14 (Danecek et al., 2011). Suites of markers from protein coding genes (i.e. putatively "non-neutral" SNPs) and from non-coding gene deserts (putatively "neutral" SNPs) were chosen to help partition genetic differentiation into variance due to selection and to drift, respectively. Leveraging variation at potentially neutral and non-neutral markers, we can better assess the evolutionary histories of different populations (Barbosa et al. 2018; Funk et al. 2012). By analyzing variation at only neutral loci, we can determine the populations that are demographically independent as variations at those loci would be due to demographic stochasticity (Funk et al. 2012). On the other hand, variation at putative non-neutral markers, we can elucidate the adaptive distinctiveness and potential of the different populations (Barbosa et al. 2018). We had no *a priori* evidence that the putatively non-neutral markers are under strong selection, but as a group they are much more likely to be targets of selection than our putatively neutral markers. Hereafter, we implicitly assume "putatively" when referring to the two panels of markers as future studies will be required to validate this initial characterization. Ultimately, a SNP panel was designed with 192 SNPs partitioned into 96 SNPs from the non-neutral sites, 94 SNPs from the neutral sites, and 2 SNPs from the mitochondrial genome (see "mtDNA markers"; Appendix A).

For the non-neutral SNPs, a manually curated list of candidate genes with potential roles in fitness related traits was generated from published Galliform literature (Table A4). These include genes associated with reproduction, growth and development, immune and stress response,

behavior, and other aspects of avian life that might provide insights into adaptive variation across the geographic range we sampled (Table A4). Predicted amino acid changes and their effects from exonic SNPs were determined using SnpEff 4.2 (Cingolani et al., 2012).

The presumptively neutral SNPs were identified by generating a distribution of all nearest-neighbor intergenic distances using BEDOPS 2.0 (Neph et al., 2012) and choosing SNPs from the largest intergenic regions to maximize the distance between any marker and its nearest gene (Doyle et al., 2016). For neutral SNPs, the average distance between a SNP and nearest annotated gene was 1.4 ? 0.88Mb. IGV v2.3 (Thorvaldsdottir, Robinson, & Mesirov, 2013) was used to identify targeted SNPs with at least 60 nucleotides of flanking sequence upstream and downstream, GC content less than 65%, and no other variable sites within 20 nucleotides. Linkage disequilibrium was minimized by choosing a single SNP per annotated scaffold.

1.3.4 SNP genotyping and genetic variation

We genotyped 188 Montezuma quail samples (Arizona, N=165; Texas, N=18; New Mexico, N=5; Table 2) using the panel of 192 SNP markers. Our sampling was heavily skewed towards Arizona where samples were hunter harvested, whereas Texas samples were opportunisitically collected as roadkill. Samples were genotyped using a Fluidigm® BioMark™ platform (Fluidigm, South San Francisco, CA). This platform works well with low quantity and/or low quality of DNA (e.g. DeWoody et al. (2017)), such as our road-killed samples or hunter harvested wings. To assess the repeatability and genotyping error of the assay, two individuals were included in replicates, resulting in total of 190 DNA samples (Appendix A). We edited individual SNP calls using the Fluidigm® Genotyping Analysis Software using a confidence threshold of 0.65 for assigning genotypes for a particular SNP assay. Markers were removed from downstream analyses if they failed to cluster into distinct homozygous and heterozygous states or if a given marker had minor allele frequency less than 0.025. Samples that failed to amplify at fewer than 164 loci (i.e., 97% of our useful markers; see Results) were removed from further analysis. Allele frequencies were calculated using GENEPOP v4.2 (Rousset, 2008). GenAlEx 6.501 (Peakall & Smouse, 2012) was used to estimate the probability of identity (P_I), which quantifies the robustness of the SNP assay to correctly assign genotypes to randomly collected samples by quantifying the probability of the likelihood of two individuals from a population having the same genotype at all loci (Waits, Luikart, & Taberlet, 2001).

Observed and expected heterozygosity was measured for all the sampling regions using GENEPOP. Deviations from Hardy-Weinberg equilibrium (HWE) and LD between loci pair were evaluated using exact tests with Markov Chain approximation. P-values for each of these tests were corrected by applying a sequential Bonferroni correction to adjust for multiple tests (Holm, 1979).

1.3.5 Population structure, effective population sizes, and selection tests

We first performed a principle component analysis (PCA) using *dudi.pca* function in ADE4 package in R (Dray and Dufour 2007), a model free approach to assess the distribution of genetic variation in the different populations. Next, a Bayesian clustering analysis was performed using two complementary approaches: sparse non-negative matrix factorization algorithms (sNMF, Frichot, Mathieu, Trouillon, Bouchard, and François (2014)) implemented in the LEA package (Frichot & François, 2015) using program R, and the program STRUCTURE (Pritchard, Stephens, and Donnelly (2000); Appendix A). Unlike STRUCTURE, the sNMF algorithms in LEA are flexible approaches that do not rely on oversimplified population genetic hypotheses such as no genetic drift, no deviations from Hardy–Weinberg equilibrium, and no linkage disequilibrium in ancestral populations (Frichot et al., 2014). The algorithm requires a parameter α (regularization parameter) that controls the regularity of ancestry estimates over geographic space. Large values of α indicate the geographically closer samples having similar ancestry coefficients, whereas small values ignore spatial autocorrelation in observed allele frequencies (Caye, Jay, Michel, & François, 2018). The number of ancestral gene pools (K) and α are chosen to minimize the cross-entropy criterion as it indicates better algorithm outputs and estimates (Frichot et al. (2014); Appendix A).

NeEstimator v2.1 (Do et al., 2014) was used to provide comparative estimates of contemporary effective population size (N_e) of the Montezuma quail populations among the different geographic regions. The linkage disequilibrium method with random mating and molecular co-ancestry method were used. P_{crit} values were explicitly provided (0.02 for Arizona, 0.05 for Texas, 0.15 for New Mexico) due to significant difference in sample sizes for each region, as recommended by Waples and Do (2010).

Pairwise population differentiation was analyzed by calculating locus-specific and global pairwise F_{ST} values for individuals sampled in each geographic region using Genodive v2.0b27 (Meirmans & Van Tienderen, 2004). A 95% confidence interval was generated by bootstrapping

over loci of pairwise F_{ST} values 1000 times using hierfstat (Goudet, 2005). Even though we specifically targeted loci associated with fitness and more likely to be under selection, F_{ST} outlier tests were performed to identify explicit signatures of selection using an FDIST approach (Beaumont & Nichols, 1996) implemented in both LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) and Arlequin v3.5 (Excoffier & Lischer, 2010), and a Bayesian approach in BAYESCAN 2.1 (Foll and Gaggiotti (2008); Supplementary Methods).

1.4 Results

1.4.1 Genome sequencing, assembly, and annotation

The whole genome sequencing of the female individual (E8452) resulted in a total of 515,110,758 raw reads from the combined PE and MP libraries and spanned 40.65 Gb. Adapter trimming and quality control resulted in 0.6% and 34% loss in reads from PE and MP libraries, respectively (Table 1.1). For the male individual (E8454), a total of 492,547,194 raw reads (49.75 Gb) were generated. Trimming and quality filtering removed 1.0% PE and 33% MP reads (Table 1.1).

The K-mer counting method from both individuals estimated the mean genome size of Montezuma quail to be 1.23 Gbp with an overall mean coverage depth of 32.1X. The most comprehensive assemblies were generated from AbySS at a k-mer size of 60 for both individuals. The female assembly was 1.03 Gb in size with 2544 scaffolds greater than 5000 bp and N50 of 1,289,626. The male draft assembly was 1.01 Gb in size with 2479 scaffolds greater than 5000 bp and N50 of 1,179,040 (Table 1.1). BUSCO estimated both assemblies to be 94% complete relative to other avian genomes (aves_odb9; number of species: 40, number of BUSCOs: 4915) (Figure A1). The Montezuma quail mitochondrial genome assembly was 16,977 bp long and contained 13 protein coding genes, 2 ribosomal subunit genes, and 33 tRNA genes (Figure 1.2).

Only the scaffolds greater than 5000 bp (n = 2479) from the male genome assembly were used for the annotation pipeline. The final annotation set consisted 17,573 genes with a gene density of 1.86 genes per 100 kb. The mean gene length was 18,046 bp with 95% of genes containing multiple exons. Out of all annotated genes, 14,680 genes (84%) had a true Pfam domain hit (Table A1).

Table 1.1: Summary statistics for the whole genome sequencing and assembly of two Montezuma quail individuals, E8452 (female) and E8454 (male)

Genome sequencing				
	Female (E8452)		**Male (E8454)**	
	Filtered reads	**Read length (total bp)**	**Filtered reads**	**Read length (total bp)**
Paired-end library	259,148,992	26,019,844,111	234,806,556	23,570,217,596
Mate-paired library	168,399,062	14,621,365,037	169,984,720	14,820,360,988
Total	426,719,784	40,572,256,797	404,791,276	38,390,578,584
Mean coverage	32.9X		31.2X	
Genome assembly (scaffolds > 5000 bp)				
No. of scaffolds	2,544		2,479	
Total Scaffold length (bp)	940,660,442		940,567,648	
Max. scaffold length (bp)	6,720,811		6,032,256	
N50	1,289,626		1,179,040	
Assembly Completeness (%)	93.9		93.6	

The mean genome coverage depth was calculated from the estimated Montezuma quail genome size of 1.23 Gb. Sequences were assembled de novo using ABySS with a k-mer of 60 with paired-end reads used to assemble the contigs whereas mate pair reads were used for scaffolding. N50 is the median value where more than half of the assembly is contained in larger contiguous regions. Assembly completeness was assessed using BUSCO with the avian database (aves_odb9) which contained 4915 BUSC

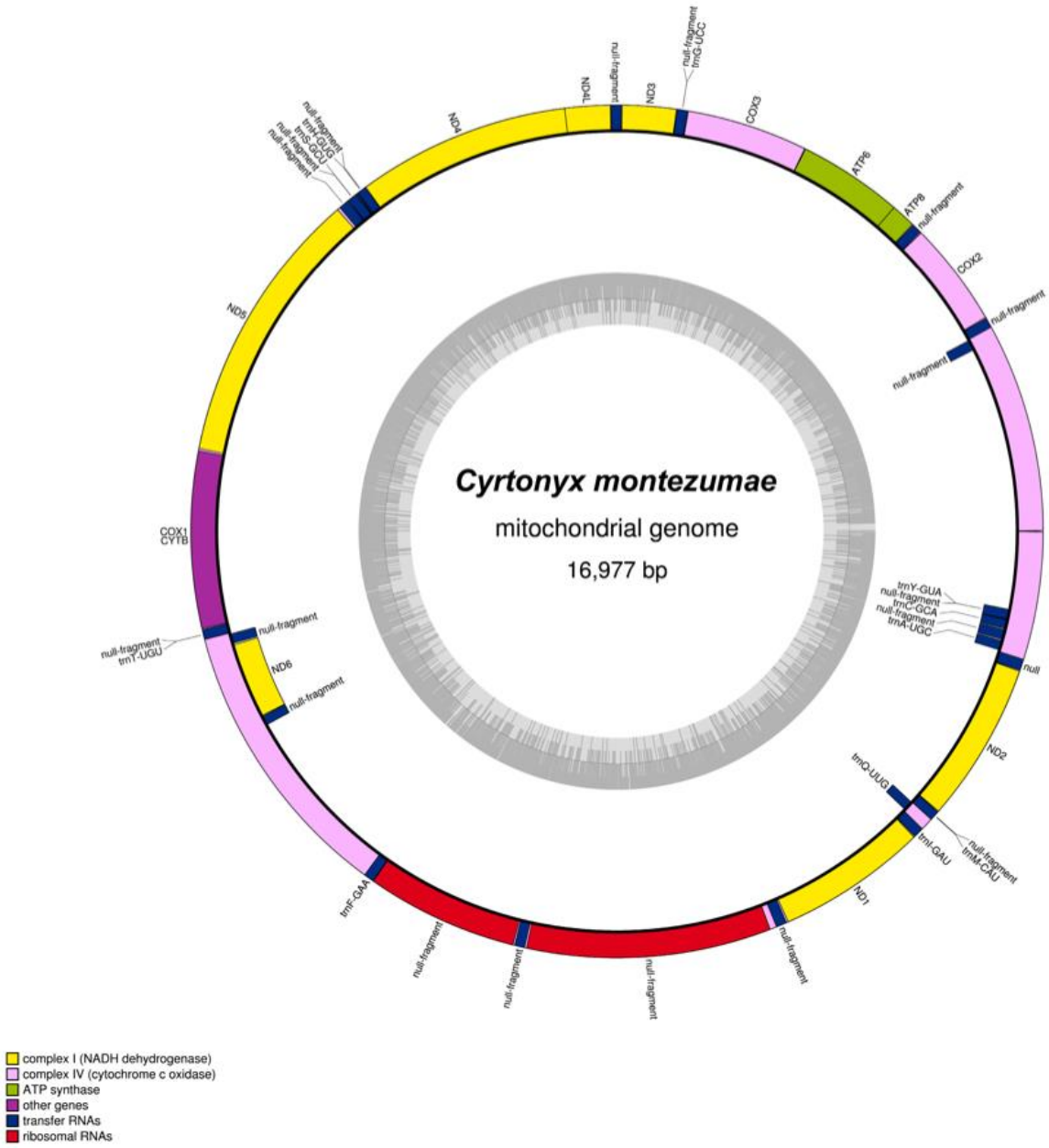


Figure 1.2: Montezuma quail (Cyrtonyx montezumae) mitochondrial genome map. The mitochondrial genome was assembled from the female individual's filtered paired-end reads using MITObim and annotated using MITOS. The Montezuma quail mitochondrial genome is highly similar to other avian genomes in size (16,977 bp) and content (13 protein coding genes, 2 ribosomal subunit genes, and 33 tRNA genes)

1.4.2 SNP identification and assay development

The variant calling pipeline resulted in over 3.2 million genome wide SNPs that passed quality-control filtering within the two sequenced individuals with an average rate of 1 variant per 289 bases (~ 3.5 per kb). Of the total SNPs discovered, 43% were found in annotated genes and of these, 41,990 SNPs were in exonic regions (i.e., the remainder were in introns or untranslated regions). Of those SNPs in coding genes, 33% were predicted to result in missense substitutions and 66% in silent substitutions (Missense / Silent ratio: 0.51). The transition to transversion (Ts/Tv) atio based on observed SNPs in coding genes was 2.37 and the nucleotide diversity across the genome was 0.55 ± 0.08 (mean ± SD).

Post-genotyping, we removed 23 SNPs from downstream analysis as they either did not consistently yield robust genotypes or had an overall minor allele frequency (MAF) < 0.025 (i.e., the marker conversion rate was 88%; only 12% of the designed markers failed in empirical tests). Two DNA samples from Texas were removed due to poor amplification and missing data. Ultimately, genotypes from 169 markers (87 non-neutral, 80 neutral, and 2 mtDNA) were successfully assigned to 186 individuals (Arizona, N=165; Texas, N=16; New Mexico, N=5). An overall genotyping error rate of 0.45% was calculated based on 4 DNA samples from 2 replicate individuals. The overall mean probability of identity (P_I) for the SNP array was estimated at 7.3E-47.

1.4.3 SNP genotyping and genetic variation

When all samples were considered together, the mean expected (H_E) and observed (H_O) heterozygosity at all SNP markers was 0.35 ± 0.14 and 0.32 ± 0.17, respectively. Between the two marker sets, observed heterozygosity at non-neutral markers (H_O = 0.38 ± 0.16) was significantly higher than observed heterozygosity at neutral markers (H_O = 0.28 ± 0.15; Mann–Whitney Rank Sum Test: W = 4326.5, p-value < 0.01; Figure A2). At non-neutral markers that presumably assess adaptive variation, mean H_O for both New Mexico and Arizona samples was significantly higher than for Texas samples ($p < 0.05$; Figure 1.3A; Table 1.2). At presumably neutral markers, there was no significant difference between mean H_O for Arizona and Texas samples ($p = 0.7$), whereas New Mexico samples had significantly higher mean H_O than either Arizona ($p = 0.02$) or Texas

Table 1.2: Number of samples, heterozygosity (HE = expected; HO = observed), and effective population size (Ne) estimates based on the Fluidigm SNP Array of 169 loci (87 non-neutral, 80 neutral, 2 mtDNA markers). Mean HO for Texas was significantly lower than Arizona ($p < 0.05$) at non-neutral loci but not at neutral loci ($p = 0.7$). Mean Ne for the Texas samples was 20 and 6.7 times lower than in Arizona and New Mexico samples, respectively.

	# Samples	Males	Females	Non-neutral markers		Neutral markers		N_E (95% CI)
				H_e (mean ± SD)	H_O (mean ± SD)	H_e (mean ± SD)	H_O (mean ± SD)	
Arizona	165	85	80	0.36 ± 0.12	0.36 ± 0.16	0.33 ± 0.15	0.28 ± 0.16	260.6 (188.4 – 399.0)
New Mexico	5	1	4	0.44 ± 0.12	0.48 ± 0.26	0.42 ± 0.12	0.37 ± 0.23	87.7 (23.7 - ∞)
Texas	16	7	9	0.31 ± 0.18	0.31 ± 0.25	0.30 ± 0.17	0.29 ± 0.21	12.6 (5.9-35.7)

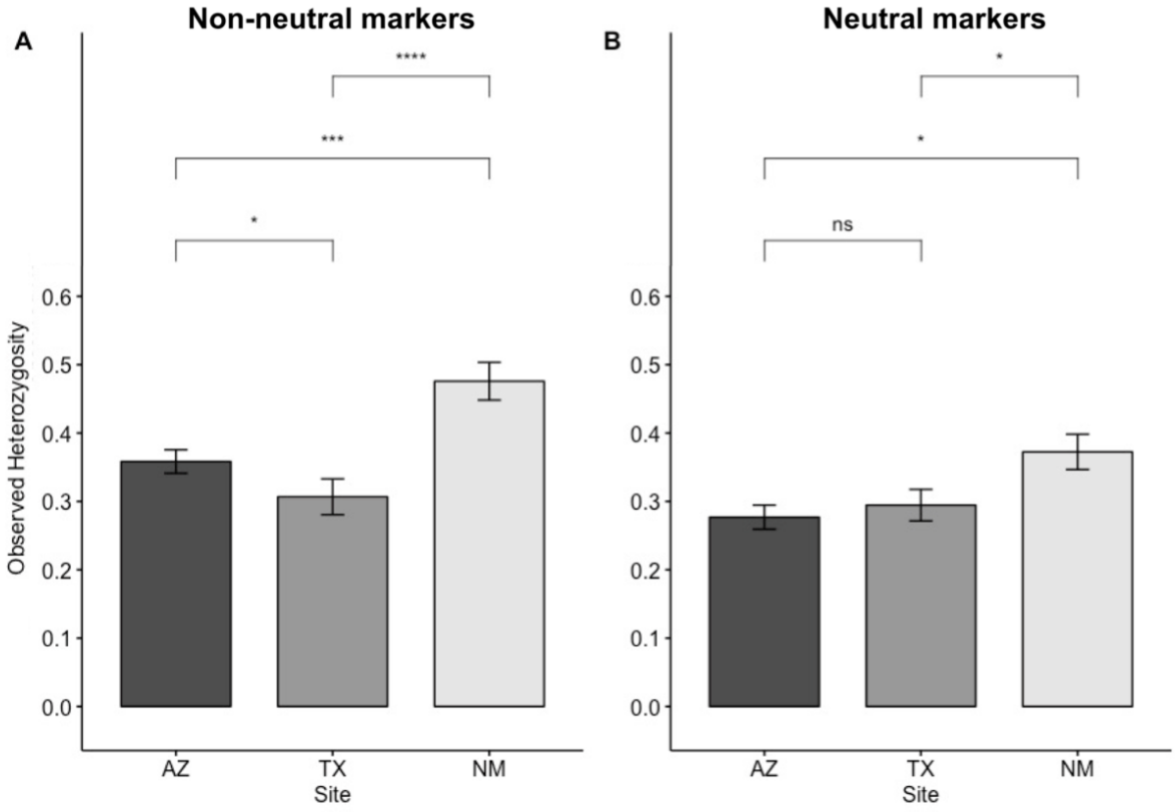


Figure 1.3: Mean observed heterozygosity for Montezuma quail samples among different sites at a non-neutral and b neutral markers. (A) At non-neutral markers, the mean observed heterozygosity in TX samples was significantly less than in AZ samples ($p < 0.05$) or in NM samples ($p < 0.001$). (B) At presumptively neutral markers, the mean observed heterozygosity in TX samples was not significantly different than from AZ samples ($p = 0.7$) but was significantly ($p < 0.05$) lower than in NM samples. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

samples ($p = 0.03$; Fig.3B; Table 1.2). After sequential Bonferroni correction, 28 out of 169 SNPs in the array were found to significantly deviate from HWE (13 from non-neutral, and 15 from neutral). No markers in the SNP array were found to be in significant pairwise LD after the p-value correction. A total of four mtDNA haplotypes were detected with our SNP assay, but there were no fixed differences among geographic sampling sites and very limited signal in the data so they are not discussed further.

1.4.4 Population structure, effective population sizes, and selection tests

Global pairwise F_{ST} values ranged from 0.08 to 0.14 (Table A2). Both sNMF and STRUCTURE analyses produced similar results indicating that when all loci were analysed together, all individuals clustered into two genetically distinct populations: Arizona-New Mexico and Texas (Figure 1.4; Figure A3). However, when only using the non-neutral loci (N= 87), all the three populations were found to be genetically distinct (Figure A4). These results were similar to the results obtained from the PCA analyses (Figure A6). We found no significant difference in genetic differentiation (F_{ST}) between non-neutral and neutral marker sets (p = 0.8; Figure A7).

We generated estimates of N_e to compare among sampling sites, not as absolutes. Estimates based on our SNP array show that N_e was highest in Arizona (261) and lowest in Texas (13), a 20-fold difference. We only had 5 samples from New Mexico but estimated an intermediate N_e there of 88 individuals (Table 1.2).

BAYESCAN failed to detect any locus under selection. In contrast, LOSITAN identified 5 non-neutral markers under significant positive selection and Arlequin found 1 non-neutral locus under positive selection (Table A3). We found no locus that was common to both analyses, but all identified markers exhibit high levels of differentiation between Arizona and Texas populations (Table A3; Figure A5).

1.5 Discussion

Adaptive genetic variation is critical to long-term population persistence and is often maintained in peripheral populations outside the core of a species' range (Lesica & Allendorf, 1995). Herein, we assessed putatively adaptive genetic variation in a peripheral population of Montezuma quail that is of conservation concern according to the Texas Parks and Wildlife Department. We sequenced the Montezuma quail genome and then genotyped a curated panel of SNPs associated with fitness related genes (Table A4) as well as a SNP panel not associated with any known gene. We were interested in the number and distinctiveness of extant populations represented in our samples, and whether signatures of genetic erosion correspond to the available habitat. Our results indicate the Arizona and New Mexico populations cluster together and contain similar levels of genetic diversity. In contrast, the small Texas population we sampled appears to be genetically

isolated, exhibits signs of genetic erosion (e.g., reduced heterozygosity at key genes), and has an alarmingly small effective population size relative to the samples from the other states.

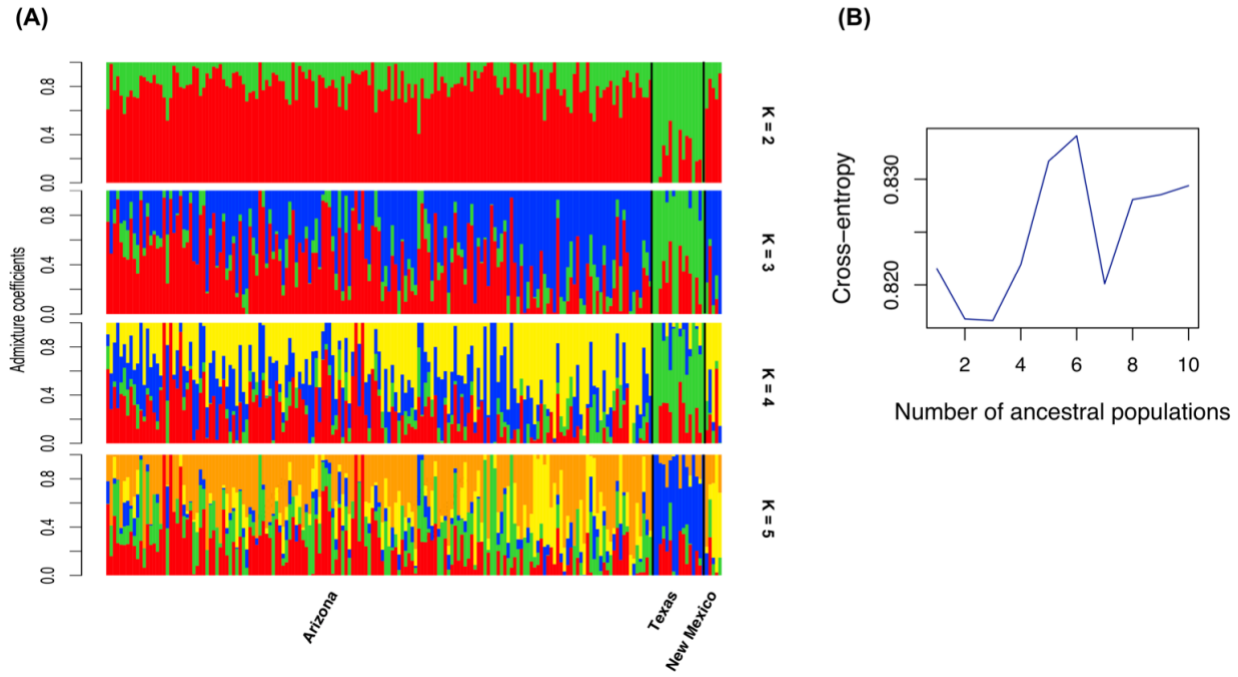


Figure 1.4: Results from Bayesian clustering analysis (K = 2–5) to determine population structure for all 186 Montezuma quail samples (AZ = 165, TX = 16, NM = 5) that were genotyped at 169 loci containing 87 non-neutral, 80 neutral, and 2 mitochondrial markers. A Samples from the Texas collection site form an independent genetic cluster as compared to Arizona and New Mexico. (B) The most parsimonious number of gene pools is estimated to be K = 3, but K = 2 is nearly as likely (See Methods for details)

1.5.1 Genome sequencing, assembly, and annotation

We sequenced the Montezuma quail genome to provide the infrastructure necessary to survey genetic variation at (in theory) both neutral and adaptive loci. Overall, the genome is unremarkably relative to other sequenced galliform genomes in terms of size and gene content. The genome assembly was comparable to scaled quail (*Callipepla squamata*, 1.08 Gb; Oldeschulte et al. (2017)), Japanese quail (1.04Gb; Wu et al. (2018)) and chicken (*Gallus gallus*, 1.05 Gb; Hillier et al. (2004)). Our assemblies were likely > 93% complete (Table 1.1, Figure A1) and we annotated 17,573 genes, similar to other birds (Zhang et al., 2014) including scaled quail (17,131 genes;

Oldeschulte et al. (2017)). The Montezuma quail mitochondrial genome assembly was also comparable to other birds (Figure 1.2; Doyle et al. (2018); Halley et al. (2014)).

1.5.2 SNP identification and assay development

We used single nucleotide variants (SNVs) within the genome assembly to a) compare diversity within the Montezuma quail genome to related species; and b) to serve as targets for a SNP genotyping array. The SNV density in the Montezuma quail genome (~ 3.5 SNVs per kb) is similar to other quail genomes (2.5 SNVs per kb in scaled quail and 4.2 SNVs per kb in Northern Bobwhite; Oldeschulte et al. (2017)). Furthermore, the ratio of non-synonymous to synonymous substitutions (D_N/D_S; 0.51 in Montezuma quail) was similar to Japanese quail (0.45; Wu et al. (2018)). Thus, at least on the surface, the Montezuma quail genome seems representative of other galliforms in terms of nucleotide diversity.

Our curated SNP assay was designed to capture variation due to neutral (e.g., drift) and to non-neutral processes (e.g., selection). The collective evidence (Figure 1.3 and A2; Table 1.2) suggests these markers generally behave as expected (e.g., the least variation at the neutral markers in the smallest population; Table 1.2). Furthermore, our array appears robust to DNA quantity and quality limitations, and has a low genotyping error rate and a low probability of identity.

1.5.3 Genetic variation revealed by the SNP assay

Our data reveal that a) overall, Montezuma quail have levels of H_O similar to other birds that have been surveyed with comparable SNP panels (Doyle et al., 2018; Doyle et al., 2016); b) in our samples, H_O was consistently higher in our array of presumptively non-neutral SNP markers than in neutral markers; and c) H_O at non-neutral markers was significantly lower in Texas individuals as compared to Arizona or New Mexico individuals.

We and others (e.g., Barbosa et al. (2018)) argue that the putatively non-neutral SNPs better reflect adaptive potential than do the neutral SNPs whereas neutral loci are better indicators of standing genetic variation due to historic demographic events. In our study, the mean overall H_O at the non-neutral SNPs was significantly higher than the neutral SNPs (Mann–Whitney Rank Sum Test: $U = 2633.5$, p-value < 0.05; Figure A2). We found no significant difference between H_O in Arizona and Texas at neutral markers ($p=0.7$) but a significant difference at non-neutral markers

($p < 0.05$; Fig. 3). This pattern could be due to balancing selection maintaining multiple alleles in the gene pool(s) or an indirect signal of over-dominance (Mitton (2000). Balancing selection should be more effective in Arizona populations due to their presumed larger size, but the small Texas population sampled still maintains some genetic diversity that may be maintained by associative overdominance (Frydenberg 1963; Schou et al. 2017). If true, these findings also indicate that the geographically isolated Texas population has reduced adaptive potential relative to the populations sampled from New Mexico and Arizona. This is less than ideal given a) the demographic vulnerability of Texas populations and b) the genetic isolation exhibited by the Texas population.

1.5.4 Population structure and effective population sizes

Limited vagility and small home ranges are one of the primary reasons for population isolation in Montezuma quail. Radiotelemetry studies show that Montezuma quail are consistently found within $50m^2$ area of their original capture (Hernandez et al., 2009; Stromberg, 1990) and thus gene flow among populations is expected to be limited. This is exactly the trend we found with our Bayesian population clustering analyses. When all sampled individuals were analyzed together, Texas individuals show evidence of a separate genetic population with no recent admixture (Figure 1.4; Figure A4). In contrast, the New Mexico samples clustered together with the Arizona samples. A more detailed analysis of spatial genetic and landscape structure in the Arizona-New Mexico population will be interesting, but our primary emphasis here is the need for local *in situ* conservation planning in Texas.

Long term conservation strategies require the mitigation of both extrinsic (e.g., habitat fragmentation) and intrinsic (e.g., genomic erosion) threats (Mussmann et al., 2017). Productive grasslands, savannas, and woodlands that are native habitat for Montezuma quail in Texas have been negatively affected by European settlements and practices (e.g., overgrazing) over the last 150 years. Consequently, the environment has been subject to erosion and species composition shifts such that their habitat is now relatively depauperate of a) tall native grasses required for escape cover and b) preferred open spaces due to brush encroachment on mid-elevation slopes (Albers & Gehlbach, 1990; Harveson et al., 2007b). The lack of contiguous contemporary habitat is likely a major barrier to dispersal and gene flow in Montezuma quail in West Texas. The Texas population that we sampled is on the periphery of the species' current range and is one of only a

few remaining small, isolated populations in the state (Figure 1.1). It seems as though substantive natural gene flow among Texas populations is unlikely in the foreseeable future.

An emerging paradigm is that conservation efforts require proactive, rather than reactive, genetic management strategies (Ralls et al., 2018). We generally agree, but think it is premature to recommend active population management (e.g., translocations) of Montezuma quail by state agencies until more strategic genetic sampling is conducted in light of the contemporary landscape (see e.g. Hoffman and Blouin 2004). Unfortunately, if conservation management is ultimately required, it may prove difficult for both intrinsic (e.g., low survival rates of translocated galliforms) and extrinsic (habitat degradation and fragmentation) reasons (Mussmann et al., 2017).

Our estimate of contemporary N_E using the SNP data was small in Texas (mean N_E = 13 individuals). In theory, an N_E = 50 is required to minimize the impacts of inbreeding and at least N_E = 500 to maintain evolutionary potential (Franklin & Frankham, 1998; Palstra & Ruzzante, 2008). These numbers suggest the Texas population is small enough that inbreeding is a real concern and that adaptive variation may be at risk due to drift. Beyond the absolute value of N_E *per se,* we note that despite a 3-fold greater sampling effort in Texas relative to New Mexico, our N_E estimate in Texas was 6.7X larger in New Mexico (mean N_E = 88) and 20-fold higher in Arizona (mean N_E = 261). Note there is only a 3-fold difference between estimates in Arizona and New Mexico despite large differences in sample sizes. Thus, Texas is the clear outlier with respect to relative N_E. Our data provide the first snapshot of population metrics in Montezuma quail from Texas and they emphasize the need for future monitoring and conservation given the multiple signs of genetic erosion there, including small N_E relative to the other states, reduced H_O relative to other states, reduced nonneutral Ho relative to other states, and the genetic isolation revealed in our population structure analyses.

We designed our SNP array with markers associated with putatively non-neutral genes (Table A3) mostly because such loci more quickly register signals of genetic differentiation relative to neutral loci (Freamo, O'Reilly, Berg, Lien, & Boulding, 2011; Helyar et al., 2011), but also because we wanted to leverage the genome sequence and previously published studies which have identified fitness-related genes in galliformes. Our global estimates of genetic differentiation indicate that there is a moderate level of differentiation between the Arizona and Texas populations (Table A3) and some markers which exhibit substantially higher F_{ST} values. (Figure A7). These could be indicative of local adaptation, but targeted research is required to address this question.

1.5.5 Caveats

Our characterization of the genome is probably representative of the species as a whole, but our population genetic analyses may not be, because all the samples analyzed in this study were opportunistically collected. For example, wings from Arizona hunters are not random samples from the landscape (e.g., quail coveys are often composed of relatives and hunters often take >1 bird per covey). This limitation also means that we did not attempt to critically assess inbreeding because our opportunistic sampling could produce substantial biases in within- and between-population estimates of identity-by-descent (e.g., F_{IS} or F_{IT}). Similarly, the road-killed samples from Texas are not randomly distributed across the available habitat and thus we did not account for within-population environmental variation. It is possible there is an ascertainment bias in our SNP data (due to marker selection from Arizona genome sequences), but we find this unlikely given that diversity at neutral loci did not differ between Arizona and Texas whereas functional diversity did. If present, such a bias would affect the allele frequency spectrum (Nielsen et al. 2004) and thus skew demographic inferences like estimates of divergence times and migration rates (Wakeley et al. 2001) so we have not considered such analyses herein.

1.6 Conclusions

We sequenced, assembled, and annotated the Montezuma quail genome to develop a SNP panel that could be used to demarcate genetic structure, estimate N_E, identify candidate genes, and detect genetic erosion. We found the genome size and content to be similar to related species, and we genotyped birds from across their U.S. range at unlinked adaptive and neutral SNP loci. Our results indicate that the Arizona-New Mexico populations are relatively homogenous (at least at the spatial scale considered herein) whereas the Texas population is differentiated and shows evidence of genetic erosion, including a very small N_E, genetic isolation, and reduced heterozygosity at fitness-related markers. Extirpations can lead to extinctions, and the signs of genetic erosion in Texas populations of Montezuma quail should be worrisome given their official (TPWD) status as demographically Vulnerable and the known importance of peripheral populations as evolutionary reservoirs for adaptive alleles. This study illustrates how modern genomic techniques can be used to proactively inform conservation efforts, especially for understudied and at-risk species.

1.7 Acknowledgements

We thank Dr. Louis Harveson at the Borderlands Research Institute, Sul Russ State University, for having the foresight to collect the Texas samples from the highway. We further thank the Arizona Department of Game and Fish (J. Heffelfinger) for the hunter harvested wings, Oldeschulte et al. and Drs. Nova J. Silvy and Roel Lopez, and C. Zachary Johnson of the Department of Wildlife and Fisheries Sciences at Texas A&M University, and Dr. Pedro Chavarria of Northern New Mexico College for stewarding them. This research was funded in part by the Texas A&M AgriLife Extension Service and the Reversing the Decline of Quail in Texas Initiative. This work was funded in part by the U.S. National Institute of Food and Agriculture. SM was supported by a Graduate Research Fellowship from the Welder Wildlife Foundation. This article represents publication #720 of the Rob and Bessie Welder Foundation.

1.8 Data accessibility

The datasets generated and/or analyzed during the current study are available in NCBI's Short Read Archive (BioProject accession # PRJNA525712; BioSample accession # SAMN11060579-80; SRA accession #SRX5471116-9) and Dryad digital repository: https ://doi.org/10.5061/dryad .4g6q6 3g.

CHAPTER 2. THE GENOMIC IMPACT OF ISOLATION AND SMALL SIZE IN TEXAS MONTEZUMA QUAIL

The results from the genotyping panel showed that Texas quail carry the signatures of genetic erosion i.e. small effective sizes, less diversity, and lack of admixture with rest of the quail. These inferences were drawn from a handful of *a priori* chosen markers and needed a genome-wide assessment of diversity and differentiation. In this chapter, we sequenced whole genomes of multiple individuals at a low depth of coverage and compared the patterns of heterozygosity, relatedness, inbreeding, and genetic load. Our results indicate that isolated Texas population is significantly more inbred and have differentiated gene pool due to higher impact of drift. Our results suggest that small populations have lower number of deleterious mutations, but drift and inbreeding is reducing the fitness by homogenizing those mutations within individuals. This chapter has gone through peer review and is currently under revision. The revised manuscript will be published in the journal *Evolutionary Applications* upon acceptance. An earlier draft of this chapter is available as a pre-print and can be accessed at: https://www.authorea.com/doi/full/10.22541/au.158941448.85174067

2.1 Abstract

Populations with higher genetic diversity and larger effective sizes have greater evolutionary capacity (i.e., higher adaptive potential) to respond to ecological stressors. We are interested in how the variation captured in protein-coding genes fluctuates relative to overall genomic diversity and whether smaller populations suffer greater genetic loads of deleterious mutations compared to larger populations. To this end, we analyzed individual whole genome sequences from different populations of Montezuma Quail (*Cyrtonyx montezumae*), a small ground-dwelling bird that is sustainably harvested in some portions of its range but is of conservation concern elsewhere. Our historical demographic results indicate that overall, Montezuma Quail populations in the U.S. exhibit low levels of genomic diversity due in large part to long-term declines in effective population sizes over nearly a million years. The smaller and more isolated Texas population is significantly more inbred and homozygous than the large Arizona and the intermediate-sized New Mexico populations. The Texas gene pool has a significantly lower proportion of deleterious

alleles than the Arizona gene pool, but Texas birds carry those deleterious alleles at higher frequencies. These results demonstrate that even in small populations, purifying selection effectively purges highly deleterious mutations that have a large effect on fitness. However, we also find that many slightly deleterious mutations rise in frequency due to drift and then become homozygous due to inbreeding in small populations, thereby elevating the overall genetic load realized in populations of conservation concern. Overall, our study illustrates how population genomics can be used to proactively assess both neutral and adaptive aspects of contemporary genetic diversity in a conservation framework while simultaneously considering deeper demographic histories.

2.2 Introduction

Many species and populations world-wide are declining at an alarming rate (Barnosky et al., 2011; Ceballos et al., 2015; Dirzo et al., 2014), mainly driven by human-mediated habitat loss and climate change (Loarie et al., 2009). Active prevention of population declines and extirpations is a priority for conservation (Cardinale et al., 2012; Thompson, Koshkina, Burgman, Butchart, & Stone, 2017) because reduction in population size is often followed by reduction in genetic diversity (Allendorf, Luikart, & Aitken, 2013; Soulé, 1985). The loss of genetic diversity has negative consequences on the future persistence of a species as it impedes its ability to adapt to environmental change (Bijlsma & Loeschcke, 2005; Bürger & Lynch, 1995; Reed & Frankham, 2003). Smaller and/or isolated populations exhibit a more rapid loss of within-population genetic variation as compared to their larger counterparts (Willi, Van Buskirk, & Hoffmann, 2006). The combined effects of drift, inbreeding, weak selection, and lack of gene flow in small, isolated populations may lead to "genetic erosion" (Bijlsma & Loeschcke, 2012). Genetic erosion is expected to reduce the mean fitness of a population and thus increase extinction risks (Bijlsma & Loeschcke, 2012; Leroy et al., 2018).

In theory, mean fitness is expected to progressively decrease in small isolated populations because of the accumulation of deleterious mutations that are ineffectively purged by selection. In large populations and/or when selection intensity is very strong (i.e., when $N_e(s)>1$ where N_e is effective population size and s is the selection coefficient), natural selection is an effective determinant of allelic fate (Kimura & Ohta, 1969). However, in small populations and/or when selection is weak (e.g., on small effect mutant alleles), genetic drift is more pronounced and allelic

fate is more stochastic (Lynch, Conery, & Burger, 1995). Thus, highly deleterious mutations are more likely to be purged by selection than to drift to high frequencies, whereas slightly deleterious mutations can actually increase in frequency in small populations (Hedrick & Garcia-Dorado, 2016). Because most of the genes underlying adaptation represent complex polygenic traits and most genetic load is probably due to small effect (i.e., only slightly deleterious) alleles (Charlesworth & Charlesworth, 1987), collectively this means that genetic erosion can impede future adaptive potential in small inbred populations (Keller, 2002) if small effect recessive deleterious alleles rise in frequency due to drift (Charlesworth, Morgan, & Charlesworth, 1993; Lynch, 2007).

In practice, empirical evidence for the purging of deleterious mutations is mostly experimental and there is far less evidence from natural populations, especially with respect to genomic sequence data (Bersabé & García-Dorado, 2013; Bijlsma, Bundgaard, & Putten, 1999; Crnokrak & Barrett, 2002; Grossen, Guillaume, Keller, & Croll, 2020; Rettelbach, Nater, & Ellegren, 2019). Economic and technical breakthroughs in whole genome resequencing now make such assessments in wild populations far more tractable. Beyond the basic evolutionary interest in allelic fates, the genetic erosion of adaptive potential is increasingly recognized as a major threat to modern conservation efforts (Holderegger et al., 2019; Ralls et al., 2018)

Much of the vertebrate genome is thought to evolve in a neutral or nearly-neutral fashion (Ohta, 1992) and is shaped by genome-wide processes such as inbreeding, migration, and demographic stochasticity (Pool & Nielsen, 2007). For example, contemporary genomic patterns of neutral diversity may be affected by the recent lack of gene flow due to anthropogenic habitat fragmentation (Lino, Fonseca, Rojas, Fischer, & Ramos Pereira, 2019) and historic demographic responses to glaciations (Nadachowska-Brzyska, Li, Smeds, Zhang, & Ellegren, 2015). Beyond neutrality, variants in genic regions often underlie evolutionary adaptations subject to natural selection, and the mode and strength of selection largely determines the phenotypic response (Ellegren & Sheldon, 2008). Hence, explicitly comparing whole genomes with defined genic regions should help with identifying the major contributors to overall genomic architecture and also gauge the adaptive potential of populations. In this study, we use whole genome sequences to quantify genic and whole genome variation from different sized populations of Montezuma Quail (*Cyrtonyx montezumae*), then estimate the degree of genetic erosion and its impact on adaptive

potential by investigating the genic load via biochemical predictions as inferred from coding regions throughout the genome.

The Montezuma Quail is a small gamebird that is hunted in portions of Mexico, New Mexico, and Arizona but of conservation concern in Texas (Figure 2.1). It is one of the least-studied avian species in North America (Gonzalez, Harveson, & Luna, 2015) due to its cryptic nature as well as difficulties associated with live trapping and monitoring (Hernandez, Harveson, & Brewer, 2006). Montezuma Quail are currently experiencing species-wide declines within the U.S. (Harveson et al., 2007b), and Texas populations are listed as Vulnerable by Texas Parks and Wildlife Department (TPWD) with no open hunting season due to growing concerns about extirpations (Harveson, 2009). Unlike other North American quails, Montezuma Quail are diet (Albers & Gehlbach, 1990) and habitat specialists (Brown, 1979) that heavily rely on grass cover for predator evasion (Bristow & Ockenfels, 2004). Their demography is strongly impacted by seasonal rainfall (Chavarria, Montoya, Silvy, & Lopez, 2012) and adequate grass cover (Brown, 1979) making habitat degradation and fragmentation major threats to Montezuma quail survival (Luna, Oaster, Cork, & O'Shaughnessy, 2017). Populations in Arizona are more genetically diverse than those from Texas or New Mexico (Mathur, Tomeček, Heniff, Luna, & DeWoody, 2019) and are expected to be the least impacted by genetic erosion due to larger sizes and more contiguous habitat (Figure 2.1). In contrast, the Texas population is expected to have the highest signature of genetic erosion due to a restricted geographic range and associated demographic isolation.

Herein, we report the data from whole genome sequencing (WGS) of 90 Montezuma Quail from Arizona, New Mexico, and Texas. We used these WGS data to quantify the levels of overall genomic diversity, genic variation, differentiation, individual inbreeding, and the inferred genetic load in each population. We do so in a conservation context by comparing populations of different sizes. Our results indicate that Montezuma Quail effective population sizes have decreased over much of the last million years, and their similar trajectories over time indicate that now-disjunct populations in the U.S. were long connected demographically. Furthermore, we find that the small Texas population is isolated, genetically depauperate, and that its genetic load is mostly due to small impact deleterious mutations that have drifted to higher frequencies. Because inbreeding is also more pronounced in the small Texas population compared to the larger populations, these

deleterious mutations are more likely to occur in homozygotes and thus contribute to declines in overall population fitness.

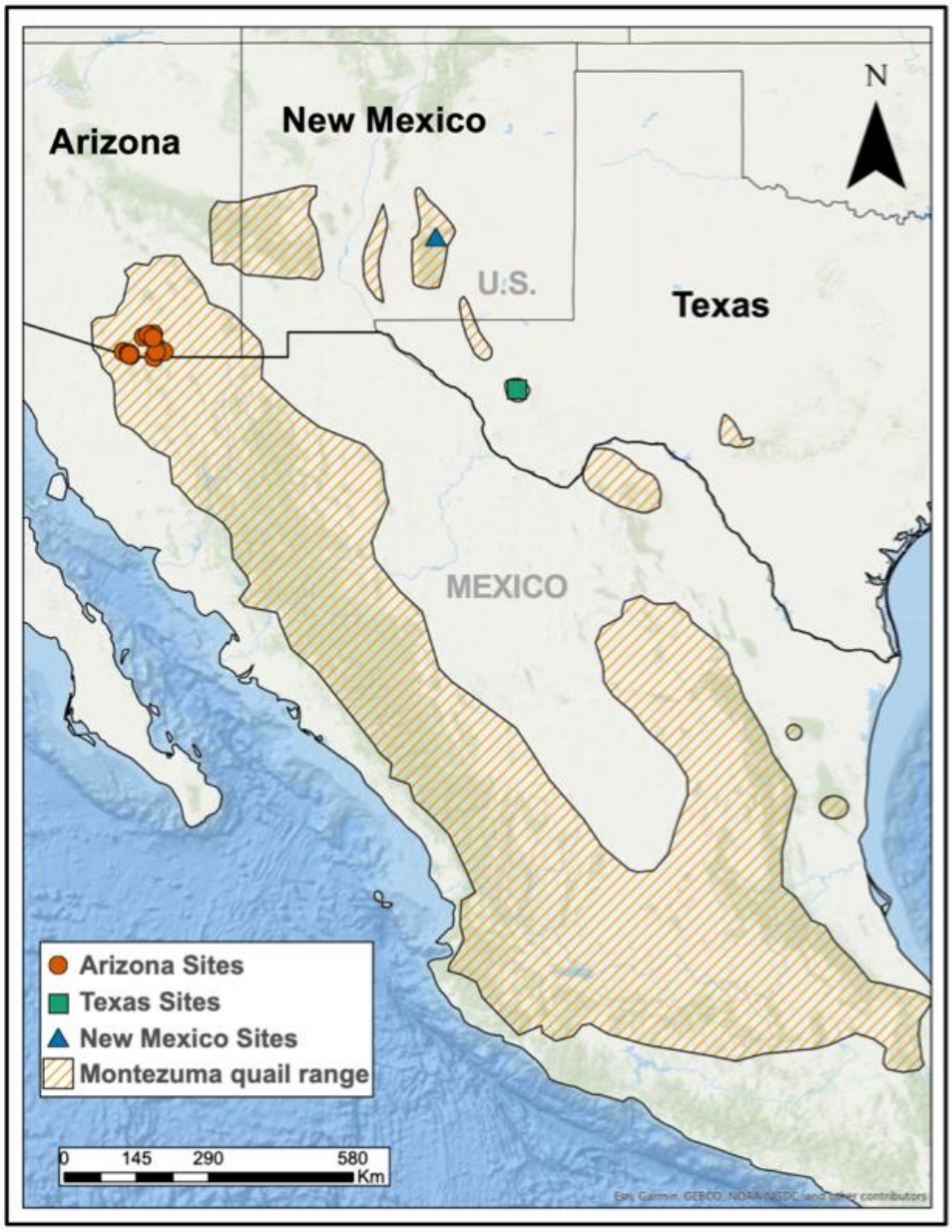


Figure 2.1: Montezuma Quail species range and sampling sites (from Mathur et al. 2019). Samples were collected from the larger and most contiguous Arizona sites (N=60), from an intermediate-sized population in New Mexico (N=13), and from a relatively isolated and small population in Texas (N=15).

2.3 Materials and methods

2.3.1 Samples, DNA extraction, and sequencing

Montezuma Quail samples were opportunistically collected from three representative geographic populations in the United States: Arizona (AZ), New Mexico (NM), and Texas (TX) as described earlier (Mathur et al. 2019; Figure 2.1). Based on the size of their geographic range in each state, on assessments by each state game agency, on eBird sightings, and on previous genetic analyses, we explicitly assume that AZ samples come from a large population, NM from a medium-sized population, and TX from a small population relative to each other (Mathur et al. 2019). Arizona samples were acquired from hunter harvested wings initially collected by Randel et al. (2019). New Mexico samples were acquired as voucher specimens by R. Luna, whereas Texas samples were collected as road-kill carcasses by L. Harveson. Sample handling and DNA extraction protocols are described in Mathur et al. (2019).

We sequenced whole genomes of 90 Montezuma quail samples (AZ=60, TX=17, NM=13) by creating individually barcoded dual-index libraries using Illumina® Nextera™ reagents following the manufacturer's protocol. The libraries were sequenced in 8 lanes of paired-end 150bp reads (2x150bp) on one S4 flow cell using Illumina® NovaSeq™ 6000 sequencing system in Purdue University's Genomics Core Facility. We removed any sample if they failed to generate more than 8 million reads (i.e. less than 1x mean read depth).

2.3.2 Sequencing filtering, alignment, and read preprocessing

We used FastQC v0.11.7 (Andrews & others, 2010) to quality check our raw reads and removed adapter sequences from trailing and leading edges of each read using Trimmomatic v.036 (Bolger et al., 2014). We also used Trimmomatic to remove low quality sequences (Phred < 20) and any read smaller than 30bp after clipping and quality filtering, prior to any further downstream analysis.

The filtered reads were mapped to a Montezuma quail draft genome assembly (Mathur et al. 2019) with BWA v.0.7.17 (Li & Durbin, 2009) using the *mem* algorithm. Samples with less than 50% mapped reads were removed from further analysis. Our final dataset contained 74 individuals (AZ=52, TX=15, NM=7). We used the Genome Analysis ToolKit (GATK) "Best Practice Workflow" (Auwera et al., 2018) to pre-process our mapped reads. We first sorted the reads by their co-ordinates and marked duplicates using PicardTools

(http://picard.sourceforge.net). We then used GATK v3.6.0 (McKenna et al., 2010) to realign our reads around indels to minimize misaligning with mismatches. We identified the regions to be realigned using RealignerTargetCreator and aligned bam files using IndelRealigner. The base quality score was recalibrated for all the reads using known variant sites discovered from high coverage genome reads (Mathur et al., 2019) using BaseRecalibrator. We finally used these filtered-realigned-recalibrated reads to get coverage statistics using samtools depth (Li et al., 2009), and for further downstream analyses.

In cases where we needed to polarize genomic variants as ancestral or derived (see below), we used the high-quality and contiguous chicken genome (*Gallus gallus* GRCg6a) as reference. Both Galliformes, Montezuma Quail belong to the New World quail Family Odontophoridae that diverged from junglefowl (*Gallus* spp.; Family: Phasianidae) approximately 30-40 million years ago (Cox, Kimball, & Braun, 2007; Hosner, Braun, & Kimball, 2015). Read mapping and preprocessing steps were same as above.

2.3.3 Mitogenome assembly and diversity

We mapped genomic reads to the previously published Montezuma Quail mitogenome (Mathur et al., 2019) and extracted the uniquely mapped reads (mito-reads) using BBMap v37.93 (Bushnell, 2014). Since nuclear copies of mitochondrial DNA (NUMTs) exist in nearly all eukaryotic genomes (Bensasson, 2001; Lopez, Yuhki, Masuda, Modi, & O'Brien, 1994), we tried to first identify the NUMTs in the nuclear genome assembly of the Montezuma Quail. We used a BLAST-based approach to query the Montezuma Quail reference mitogenome against a custom blast database of Montezuma Quail nuclear genome scaffolds. We extracted the NUMT sequences from genome assembly as fasta files using faSomeRecords (Kent et al., 2002). Any mito-read that also uniquely matched to the NUMT fasta sequences were removed using BBMap. This helped ensure that final mito-specific reads we retained belonged to the mitogenome and not NUMTs. We used samtools mpileup to align mito-specific reads to the reference mitogenome and used bcftools (Li et al., 2009) to call variants. We filtered the variants with a minimum base depth of 10 using vcflib (Garrison, 2012) and used bcftools consensus to create consensus mitogenomes for every individual. To avoid mismapping and errors introduced at the artificial ends created in the linearized mitogenome, we trimmed 40bp from either end of the mitochondrial sequence prior to analysis.

All mitogenomes were aligned as multiple sequence alignment using Clustalw v.2.1 (Thompson, Higgins, & Gibson, 1994) using default parameters. We calculated mitochondrial nucleotide diversity indices and haplotype statistics using Arlequin v3.5 (Excoffier & Lischer, 2010). We accounted for unequal sample sizes for each population by randomly subsampling mitochondrial genomes from each population (N=7) and recalculated nucleotide diversity indices using 100 independent permutations.

2.3.4 Genotype likelihood estimation, subsampling, and genotype calling

For the nuclear reads, we used the samtools model in ANGSD v0.929 (Korneliussen, Albrechtsen, & Nielsen, 2014) to estimate genotype likelihoods (GL) and call single nucleotide polymorphisms (SNPs). We filtered bam files to only include unique reads with a minimum mapping quality of 20. We excluded bases with a base quality score < 20 and only retained only proper pairs. Major and minor allele was inferred from the GL and triallelic sites were removed. Per-site allele frequencies (AF) were estimated using a combination of estimators i.e. first estimating allele frequency from GL assuming both major and minor alleles are known and then re-estimating AF by summing over the three possible minor alleles weighted by their probabilities. We used a p-value cut off of 10^{-6} to call a site polymorphic and a minimum minor allele frequency (MAF) of 0.05. We also used a maximum depth threshold of 500 to avoid calling SNPs from repetitive regions (Clucas, Lou, Therkildsen, & Kovach, 2019). Deviations from Hardy-Weinberg equilibrium were tested and sites with p-value < 0.01 were filtered out to remove potential paralogous sequences with an excess of heterozygotes due to erroneous mapping (Meisner & Albrechtsen, 2019).

When estimating GL across all samples (N=74), we used a threshold of minimum 60 individuals to ensure including segregating sites from more than one population, in other words, to prevent retaining sites from only the Arizona population (N=52) ("population dataset"). To avoid biases introduced due to uneven sample sizes, we re-estimated GL and discovered SNPs from an equal subset of Arizona and Texas samples (N=21; AZ=7, TX=7, NM=7). For our subsamples, we chose samples with the highest depth and breadth of coverage to maximize the genomic spread of our variants ("genomic dataset"). For the subset, we used a minimum individual threshold of 15 and maximum depth threshold of 100.

In the end, we analyzed our genotype likelihood data two ways: (a) retaining maximum individual information at the cost of markers per individual ("population dataset") and (b) retaining maximum genomic information on each population at the cost of individuals analyzed per population ("genomic dataset"). The population dataset was used for the estimation of inbreeding and genetic structure, both of which can be inferred from a smaller set of widespread markers from more individuals (McLennan, Wright, Belov, Hogg, & Grueber, 2019), whereas the genomic dataset with higher SNP density was used to estimate genome-wide diversity and for detecting signatures of selection (Benjelloun et al., 2019).

2.3.5 Relatedness, inbreeding coefficient, and population structure estimation

Assumptions of many population genetic estimators are violated if family members and closely related individuals are analyzed simultaneously. Related individuals among a sample set should thus be identified and removed prior to population structure analysis (Meisner & Albrechtsen, 2018, 2019). We estimated relatedness among our samples using IBSrelate (Waples, Albrechtsen, & Moltke, 2019). IBSrelate uses GL estimates to categorize a pair of individuals as either parent-offspring, full-siblings, half-siblings, first-cousins, or unrelated based on whether the pair share the same genotype or exhibit dissimilar genotypes at a particular site (Manichaikul et al., 2010). We compared all individual pairs (total of 2701 comparisons) and removed any pairwise comparison from relatedness estimates if the number of sites compared were less than 100,000.

We estimated individual inbreeding coefficients (F) using PCAngsd v.0.982 (Meisner & Albrechtsen, 2018) from inferred GL. This allows F-values at a site to vary between -1 and 1 where a negative value indicates an excess of heterozygotes and a positive value indicates an excess of homozygotes at a site. Since inbred individuals would have an excess of homozygous sites, they should have an overall $F > 0$. We used extremely low tolerance values (1×10^{-9}) and 5000 maximum iterations for estimation to assure a stricter stopping criterion and avoid convergence at a local minimum (Figure B11).

To identify genetic structure in our Montezuma Quail samples, we used two approaches. First, we used PCAngsd to calculate a covariance matrix and performed individual level PCA using princomp function in R (Team & others, 2013). Second, we used NGSAdmix (Skotte, Korneliussen, & Albrechtsen, 2013) to estimate individual admixture proportions. For PCAngsd, we used a minimum tolerance value for population AF estimation of 1e-9, a tolerance threshold

for updating individual AF of 1e-9 for 1000 iterations. For NGSAdmix, we ran 10 independent runs for each K from 1-10 with minimum MAF 0.05, 1e-9 tolerance for convergence, 1e-9 tolerance for log likelihood difference in 50 iterations, and maximum 50,000 iterations. The most likely number of subpopulations were determined based on first and second order rate of change of the likelihood distribution from the 10 runs (Evanno, Regnaut, & Goudet, 2005).

2.3.6 Nucleotide diversity, heterozygosity, and contemporary effective population size estimation

For nucleotide diversity estimates, we only used the genomic dataset to avoid biases in estimating site frequency spectrum (SFS) due to uneven sample sizes and heavy data pruning, which was the case for our population dataset. We used ANGSD to generate a folded SFS by using the Montezuma Quail reference genome and a minimum base quality of 20 and minimum mapping quality of 20 (Figure B12). Next, we obtained a maximum likelihood estimate of the SFS using realSFS by bootstrapping it 100 times and using the mean SFS for each population to estimate per-site Watterson's theta (θ_W). We estimated heterozygosity for each individual as the total proportion of heterozygous sites from its SFS.

To obtain an estimate of contemporary effective population sizes (N_e) from mean genomic θ_W, we first estimated the whole-genomic mutation rate (μ) for Montezuma Quail ($\theta_W = 4N_e\mu$). Since no linkage map exists for Montezuma Quail, we estimated μ following Zhan et al. (2013). The Montezuma Quail reference assembly was mapped to the Chicken genome (*Gallus gallus* GRCg6a) using LASTZ (Harris, 2007). The mean divergence time (t) between chicken and Montezuma Quail was derived from www.timetree.org and polymorphic loci were identified only if neither target nor query nucleotide was N/n and the locus was not in an alignment gap. The final μ per nt per year was calculated with the following formula: μ = (counts of mutated loci / sequence length) / 2t (Zhan et al., 2013).

2.3.7 Genetic differentiation and selection scans

Small populations in isolation can become genetically differentiated due to drift at neutral loci as well as positive selection at non-neutral loci (e.g., in response to local adaptation). Both processes lead to nucleotide divergence (D_{XY}) and divergence in allele frequencies (F_{ST}) (Matthey-Doret & Whitlock, 2019; Puzey, Willis, & Kelly, 2017; Rousset, 1997). We investigated genomic patterns

of genetic differentiation by estimating pairwise F_{ST} using a sliding window approach (window size=100kb, step=50kb) for each population pair (AZ-TX, TX-NM, AZ-NM). We used ANGSD to calculate the 2D SFS for each population pair using the chicken genome (GRCg6a) as reference to polarize alleles as derived or ancestral. We quantified the levels of nucleotide divergence (D_{XY}) using the calcDxy.R (https://github.com/mfumagalli/ngsPopGen/blob/master/scripts/calcDxy.R). In this case, we estimated GL for each population individually, but only retained sites that were shown to be segregating in all populations. This ensured that sites with a fixed allele in one population is still included in our per population D_{XY} calculations.

To identify candidate regions under putative selection due to local adaptation, we Z-transformed F_{ST} around the mean for each sliding window and examined the outliers that had $Z(F_{ST})$ values outside 5 standard deviations from the mean (Willoughby, Harder, Tennessen, Scribner, & Christie, 2018). After removing false positives that showed higher deviations due to lack of data, the remaining outlier windows were inspected for nearby genes. We blasted the 100-kb outlier window to the chicken genome using default parameters and only retained windows that contained annotated genes with known function.

2.3.8 Population trends and historic demographic sizes

Neutral alleles with rare initial frequencies are more likely to be lost during bottlenecks whereas more common alleles tend to increase in frequencies more than expected under an equilibrium demographic model. This shift from rarity in the allele frequency spectrum results in an overall positive value of Fu's F statistic (Fu, 1997). On the other hand, new mutants tend to increase in frequency in expanding populations and thus produce an excess of rare variants and a negative mean value of Fu's F. Fu's F is more sensitive to demographic changes than Tajima's D (Ramos-Onsins & Rozas, 2002) but requires ancestral sequences for unbiased estimations. Thus, we estimated mean Fu's F statistic for every population over a sliding window in ANGSD using the chicken genome as an ancestral reference with 100kb window size and 50kb step.

We reconstructed ancestral demographic histories using SMC++ v.1.15.2 (Terhorst, Kamm, & Song, 2017) which uses unphased whole genome data to infer population size histories using sequential Markov coalescent (SMCs) simulations. The reads that mapped to the first 10 chicken chromosomes (NC_006088.5- NC_006097.5) comprising ~750 Mbp were used to create composite likelihoods for each population individually by varying the identity of the distinguished

individual while keep other individuals within the population as undistinguished. We used cross-validation to estimate population size changes using the Powell algorithm with a tolerance of $1x\ 10^{-5}$ and a mutation rate of $3.14x10^{-09}$ (estimated as above). We ran our model using 5000 iterations and used different parameter values for thinning and regularization penalty to avoid degeneracy in the likelihood and overfitting (Terhorst et al., 2017) with final model generated using thinning parameter of 1300 and regularization penalty of 6. A generation time of 1.5 was used to convert generations into years.

2.3.9 Genic diversity and estimation of genetic load

The Montezuma Quail genome consists of ~17,500 genes (Mathur et al., 2019), and here we compared levels of nucleotide variation across the entire genome to levels of variation in just the genic regions in order to help partition the effects of drift and selection. We used BEDOPS (Neph et al., 2012) to convert the gene annotation file (.gff) to a BED file and filtered BAM files to only include reads that overlapped with the genic co-ordinates using samtools *view*. The genotype likelihoods and diversity indices were estimated for the genic regions following the same methods and parameters as above.

We quantified the potential genetic load of a population as the proportion of deleterious variants of different impact classes across all annotated protein-coding genes. We did so by predicting the effect of each nucleotide variant on the resulting amino acid sequence, then quantifying its putative deleterious impact using SnpEff 4.2 (Cingolani et al., 2012) where we analyzed only those variants the algorithm considered high-quality. A variant was classified as either high, moderate, low, or modifier based on its inferred effect on protein translation. High impact variants should have the most disruptive (i.e., deleterious) effect on protein structure such as premature termination or other loss of function mutations, whereas low impact mutations were mostly synonymous substitutions with little to no impact on protein sequences. Individuals and populations that bear the highest ratio of highly deleterious mutations to total genic variants have the highest potential genetic load. We compared the among-population differences in impact proportions using chi-squared tests with Yates' continuity correction (Newcombe, 1998; Yates, 1934). We note, however, that the proportion of potential load that is actually realized in individuals also depends on the mode of dominance and on zygosity. To test whether individuals in different populations have inferred differences in genetic load, we compared the average number

of deleterious alleles per individual relative to the reference genome sequence. We computed the per-individual proportion of deleterious variants of each impact class as the total number of deleterious alleles present within an individual divided by twice the number of segregating sites (Simons, Turchin, Pritchard, & Sella, 2014) and for convenience generally refer to this quantity as the realized load of an individual, recognizing that dominance, zygosity, and other factors also impact genetic load. To assess the impact of a mutation in genic regions and its zygosity, we called genotypes at SNPs within the genes from GL estimates. Genotypes were only called at sites with minimum individual depth of 5X to minimize technical biases (Benjelloun et al., 2019). Allele frequencies from the genotype calls at each of the genic variants were calculated using vcftools v.0.1.16 (Danecek et al., 2011).

2.4 Results

In this study, we collected WGS data from 90 Montezuma Quail (AZ=60, TX=17, NM=13; Figure 2.1). We generated more than 1.65 billion reads (mean = 18.5 million reads per individual) corresponding to approximately 250 billion bases (mean = 2.8 billion bases per individual; >2x individual coverage). Since these samples were opportunistically collected (i.e. either hunter-harvested wing tissues or roadkill carcasses), we found significant variability in the quality and quantity of DNA sequenced. This stochasticity was evident from sequences generated per individual (Table B1) and their depth and breadth of coverage (Table B1, Figure B1). We removed samples that failed to generate the threshold of 8 million bases (N=10) or where less than 50% of the total reads mapped to the Montezuma Quail assembly (N=6). However, we achieved a high level of read mapping for the remainder of the samples (84.4% ± 18.1%; Table S1). Ultimately, we analyzed genomic information from 74 individuals (AZ=52, TX=15, NM=7) that covered 65.1 ± 22.1% (mean ± SD) of the Montezuma Quail genome at 2.1 ± 1.3X depth (Table 2.1).

Our complete mitogenome analysis detected 39 unique haplotypes in the Arizona population with 239 parsimony-informative sites shared among them. There were 11 unique Texas haplotypes sharing 171 parsimony-informative sites, and we found only 3 unique haplotypes for the New Mexico population with 167 such sites. We found per-site nucleotide diversity (Π) and Kimura 2-P pairwise distances to be smaller in the Texas and New Mexico mitogenomes ($p=0.03$ and $p=0.04$ respectively) as compared to Arizona. Haplotype diversity (H_d) did not significantly

differ between Texas and Arizona mitogenomes ($p = 0.70$) but was significantly smaller in New Mexico as compared to Arizona ($p=0.02$; Figure B2).

For the nuclear genome analysis, we partitioned our data into two datasets: population and genomic. The population dataset consisted of genotype likelihoods from 456,373 SNPs retained from all individuals (N=74). The genomic dataset contained genotype likelihood information from 6,696,145 SNPs sampled across an equal subset of each representative population (N=21). Using the population dataset, we first estimated the relatedness among our samples to determine if we had close relatives in the study. Pairwise relatedness was measured for 2,341 individual pairs. Almost all the pairs analyzed were either unrelated (99.5%) or 3rd-degree relatives (0.21%). We found no full-sibling or parent-offspring relationships (1st-degree) in our samples; however, 5 pairs from Arizona, 1 pair from Texas, and 1 pair from New Mexico had 2nd degree or half-sibling relationship (Figure 2.2A). Overall, our kinship analysis indicates that, consistent with our opportunistic field sampling and broad survey range, close relatives were only rarely sampled and thus, should not impact our population structure results. Inbreeding coefficient estimates (Table 2.1) showed significantly higher levels of mean inbreeding in Texas birds as compared to Arizona birds (Figure 2.2B; Table B2) whereas inbreeding in Texas was only slightly elevated relative to New Mexico birds. Both PCA and admixture analyses produced similar results indicating that the Arizona, Texas, New Mexico populations are genetically distinct (Figure 2.2C, D). However, based on the ΔK method (Evanno et al., 2005), the most likely number of ancestral populations is K=4 (Figure B3), splitting Arizona populations into two subpopulations (Figure 2.2C). The population-level trends for relatedness, inbreeding and genetic differentiation were concordant between the two datasets (Fig. B4) and thus it seems clear that sampling issues have not biased our interpretations.

We used genomic dataset to quantify the levels of genome-wide nucleotide diversity as estimated by per-site Watterson's theta (θ_W). Mean genome-wide θ_W was significantly lower for the Texas population ($\theta_W = 4.05 \times 10^{-4}$; SE = 1.67×10^{-7}) as compared to both Arizona ($\theta_W = 5.37 \times 10^{-4}$; SE = 1.93×10^{-7}) and New Mexico ($\theta_W = 4.57 \times 10^{-4}$; SE = 1.80×10^{-7}) (Table 2.1; Table B3). The genome-wide distribution of per scaffold diversity had higher a mean in the Arizona population than in Texas or New Mexico (Fig. B5).

Table 2.1: Summary statistics for sequence coverage, inbreeding coefficients (F), per-site Watterson's theta (θw), heterozygosity (H) and effective population sizes (Ne for Montezuma quail populations analyzed in this study. The diversity indices were calculated for either the whole genome or just the genic regions. Long-term (evolutionary) Ne was calculated using an estimated mutation rate of 3.14x10-9 with 95% CI calculated using standard error in θw estimates. Sequence depth is measured in fold-coverage and breadth is measured as percentage of Montezuma quail assembly mapped by the reads.

	N	Sequence depth (X) (mean ± SD)	Sequence breadth (%) (mean ± SD)	F (mean ± SD)	Whole genome		Genic regions		N_e (95% CI)
					θ_W	H	θ_W	H	
Arizona	52	2.14 ± 0.78	69.45 ± 14.51	0.05 ± 0.08	$5.37x10^{-4}$	0.0014	5.23×10^{-4}	0.0012	42,795 (42,764 - 42,825)
Texas	15	1.45 ± 1.82	42.69 ± 30.17	0.33 ± 0.28	$4.05x10^{-4}$	0.0009	3.94×10^{-4}	0.0007	32,208 (32,182 - 32,234)
New Mexico	7	3.48 ± 1.78	84.16 ± 11.51	0.07 ± 0.08	$5.57x10^{-4}$	0.0013	4.47×10^{-4}	0.0011	36,417 (36,390 - 36,446)

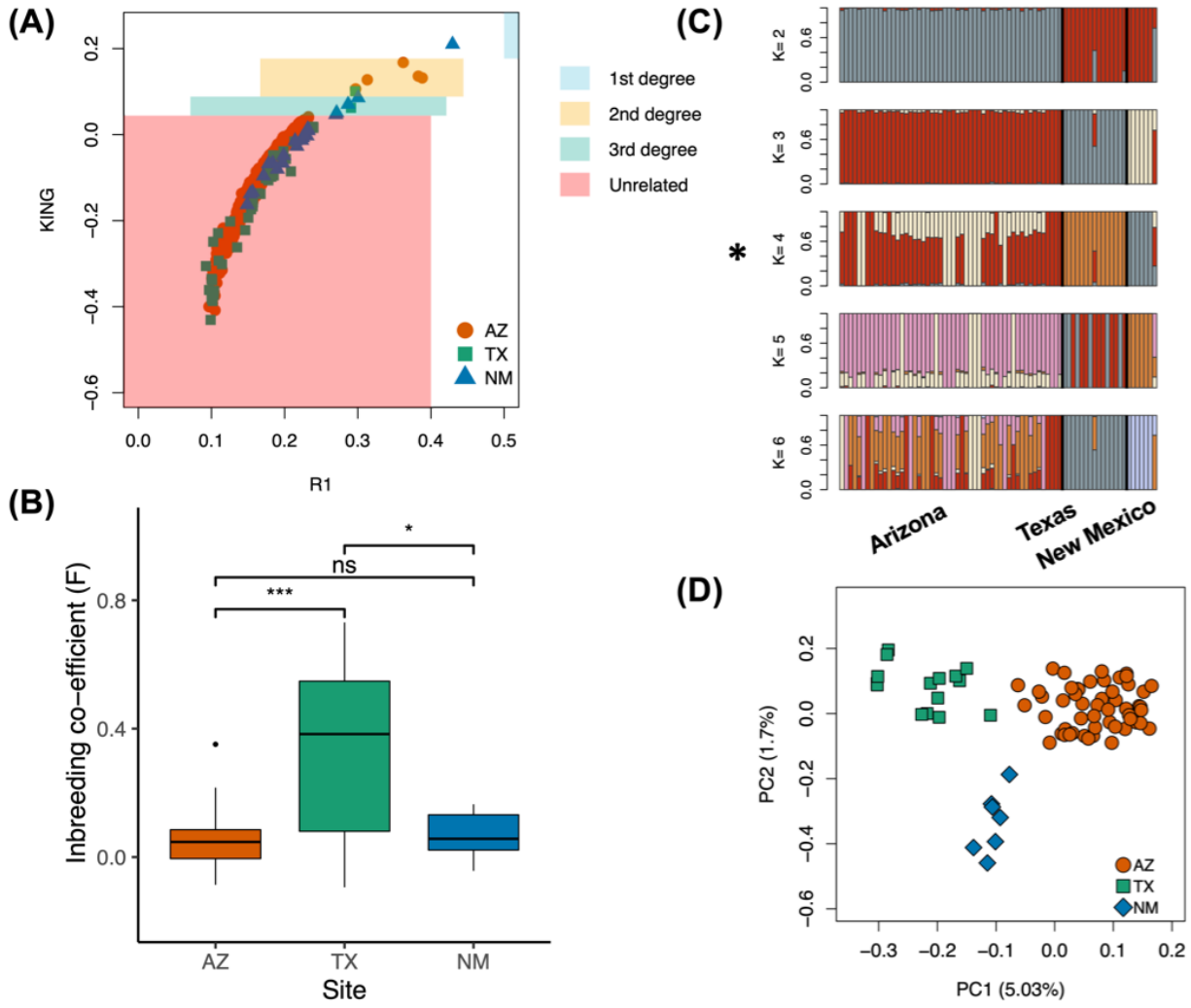


Figure 2.2: Inbreeding and population structure in Montezuma Quail. Samples analyzed in this study were mostly unrelated based on (A) kinship analysis. (B) Mean individual inbreeding coefficients (F) were significantly higher in the Texas population with no significant difference between Arizona and New Mexico populations. Results from both (C) admixture and (D) PCA analysis clearly demarcate samples from the three collecting sites into independent genetic clusters. However, likelihood estimates indicate the most likely number of ancestral populations in our data is K=4 (indicated with asterisk), where Arizona is sundered into two subpopulations.

Contemporary estimates of N_e were quantified using whole-genomic μ of 3.14×10^{-9} $bp^{-1}year^{-1}$ (CI: 2.59×10^{-9} - 3.34×10^{-9}) (Table 2.1). Thus, Texas quail show a ~30% reduction in their overall genomic diversity with a mean, long-term evolutionary N_e reduction of ~25% relative to Arizona. The genomic heterozygosity was also significantly reduced for Texas birds (Table 2.1) as compared to either Arizona or New Mexico birds (Figure 2.3A; Table B4). This indicates that smaller Montezuma Quail population in Texas

is more severely impacted by genetic erosion with contemporary diversity equivalent to those reported in endangered and vulnerable avian species, whereas the larger Arizona population has heterozygosity estimates similar to other more common avian species (Figure 2.3B).

Global estimates of F_{ST} between each population pair showed low to moderate levels of genetic differentiation at the whole genome level (Table 2.2). However, we found significant variation in F_{ST} values across the genome for each population pair (Figure 2.4; Figure B6). One interesting observation was large $Z(F_{ST})$ scores for loci on chromosome 16 (NC_006103.5) for all population comparisons (Figure 2.4; Figure B6). This is probably due to low synteny between quail and chicken at chromosome 16 (Morris et al., 2020), perhaps due to an inversion (Clucas et al., 2019) but this needs further validation using longer sequence scaffolds (Lamichhaney & Andersson, 2019). There is a similar discontinuity at one end of chicken chromosome 26 (Figure B6). We examined the windows that were highly differentiated in both AZ-TX and TX-NM comparisons to look for genes and assess their functionality. Genes or a gene clusters associated with the outlier peaks are shown in Figure 2.4 and their known functions are listed in Table B5. Per-site F_{ST} and D_{XY} values for SNPs located in those genes are in shown in Figure B7. In total, we found 12 genes that exhibited very high levels of differentiation (> 5 SD) with known function in immunity and/or development related traits (Table B5). These genes are candidates for those under strong selection and could underlie local adaptations in Texas quail.

Demographic analysis indicated that the Arizona population have been expanding with Fu's F = -0.23 ± 0.01 (mean ± SE) whereas both the Texas and New Mexico populations have been declining with Fu's F = 0.11 ± 0.02 and 0.22 ± 0.02 respectively (Fig. 2.5A). We tracked N_e estimates over the last ~1 million years using the pairwise sequentially Markov coalescent method (Fig. 2.5B). The three populations display concordant trajectories for most of their evolutionary history over that timeframe. We observed a decline in N_e from in the period of 10^6 - 10^5 years before present (YBP) followed by a more stable period. A subsequent re-expansion

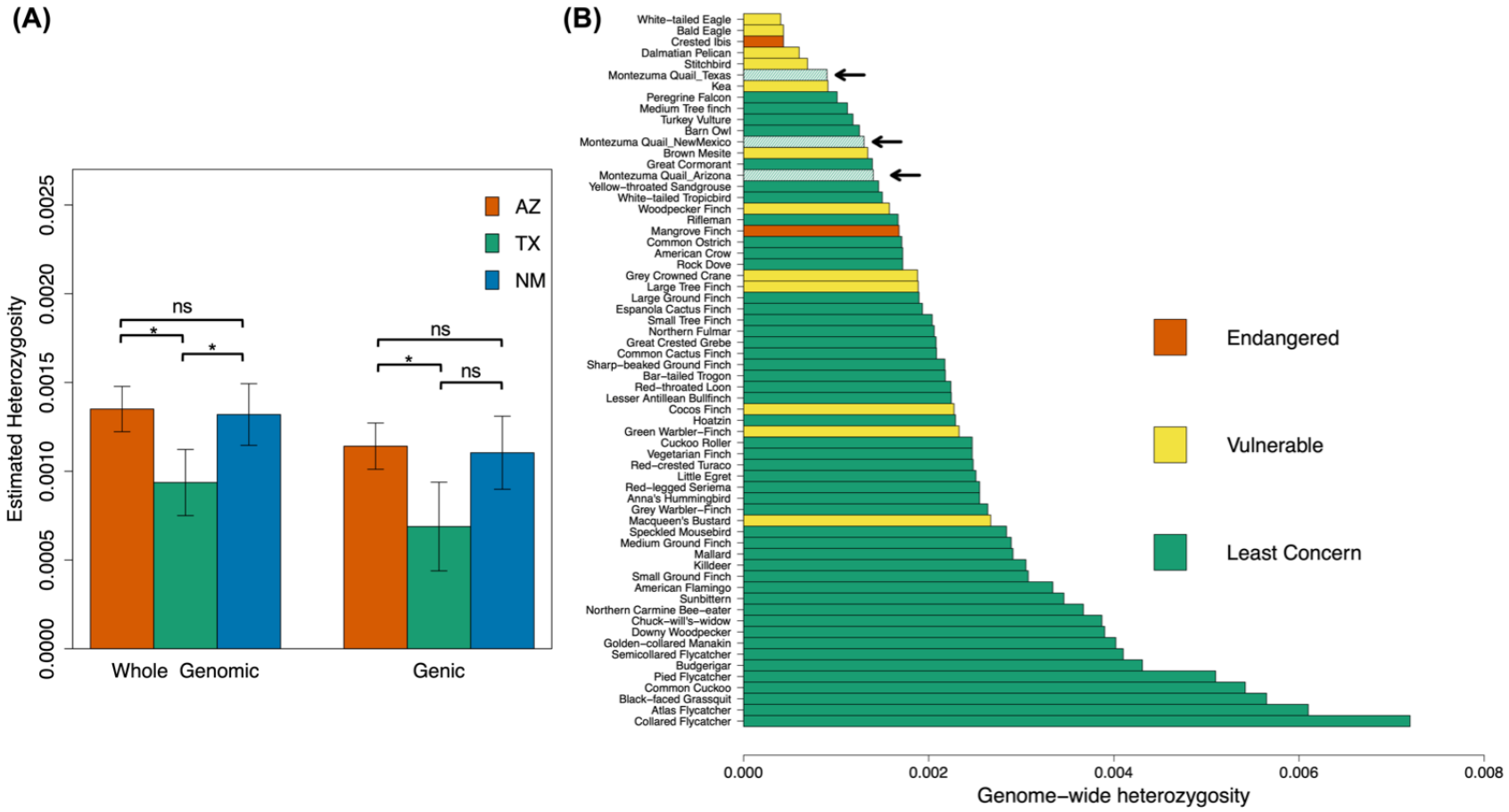


Figure 2.3: Estimated levels of heterozygosity in Montezuma quail. (A) Genic heterozygosity is comparable to genome-wide heterozygosity in individuals from all three populations but Texas quail exhibit significantly lower levels of both as compared to Arizona quail (B) Comparison of genome-wide heterozygosity with other birds indicates that smaller Montezuma Quail populations in Texas and New Mexico have genomic diversity comparable to vulnerable species (Brüniche-Olsen, Kellner, & DeWoody, 2019; de Villemereuil et al., 2019; Li et al., 2014). Heterozygosity was measured as the mean proportion of heterozygous sites per individual genome.

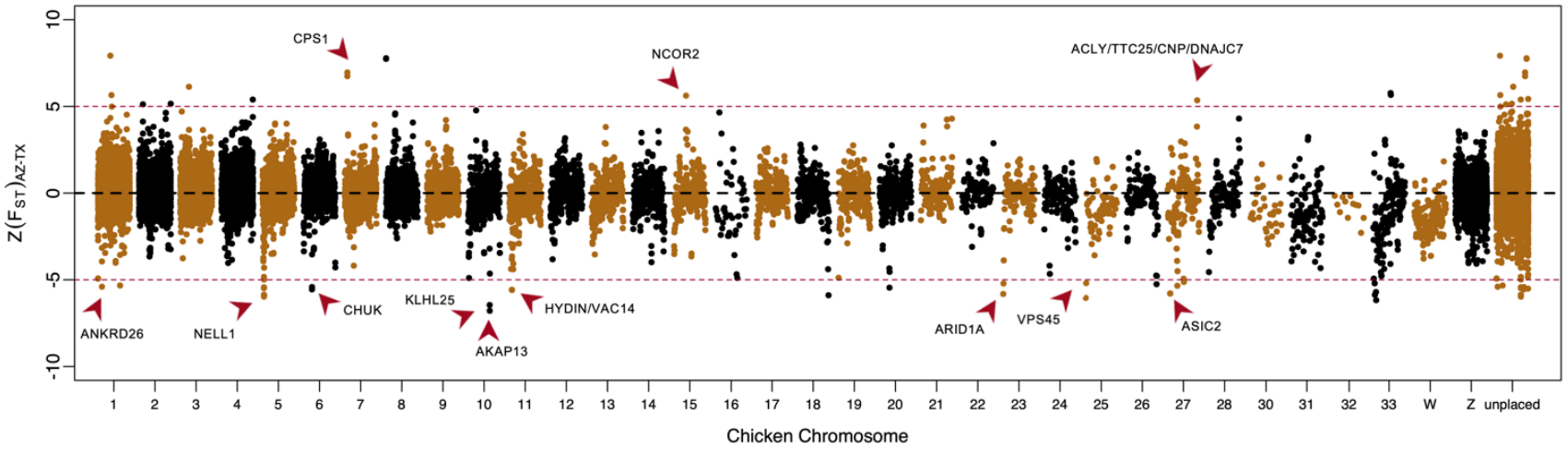


Figure 2.4: Z-transformed FST estimates for comparisons between Arizona and Texas samples. The reads were mapped to the chicken genome and the windows (100 kb width with 50 kb steps) were arranged according to chicken autosomal (1-33) or sex (Z, W) chromosomes. Scaffolds that were not part of the major chicken chromosomes were binned together as unplaced. We found windows within each chromosome that had high (>5 SD) levels of differentiation and many of those windows contained genes with known function (red arrows). These data illustrate the heterogeneous landscape of genomic differentiation in Montezuma quail.

Table 2.2: Estimates of global FST between the different population pairs measured for either the whole genome or just the genic regions. 95% CI was calculated using standard error in FST estimates by 100 bootstraps of the 2D-SFS for each population pair.

Population Pair	Mean Global F_{ST} (95% CI)	
	Whole genome	Genic regions
Arizona - Texas	0.1287 (0.1286 - 0.12878)	0.2042 (0.2018 – 0.2065)
Texas - New Mexico	0.0962 (0.0961 - 0.0962)	0.1762 (0.1744 – 0.1781)
Arizona - New Mexico	0.0972 (0.0972 - 0.0973)	0.1217 (0.1213 – 0.1220)

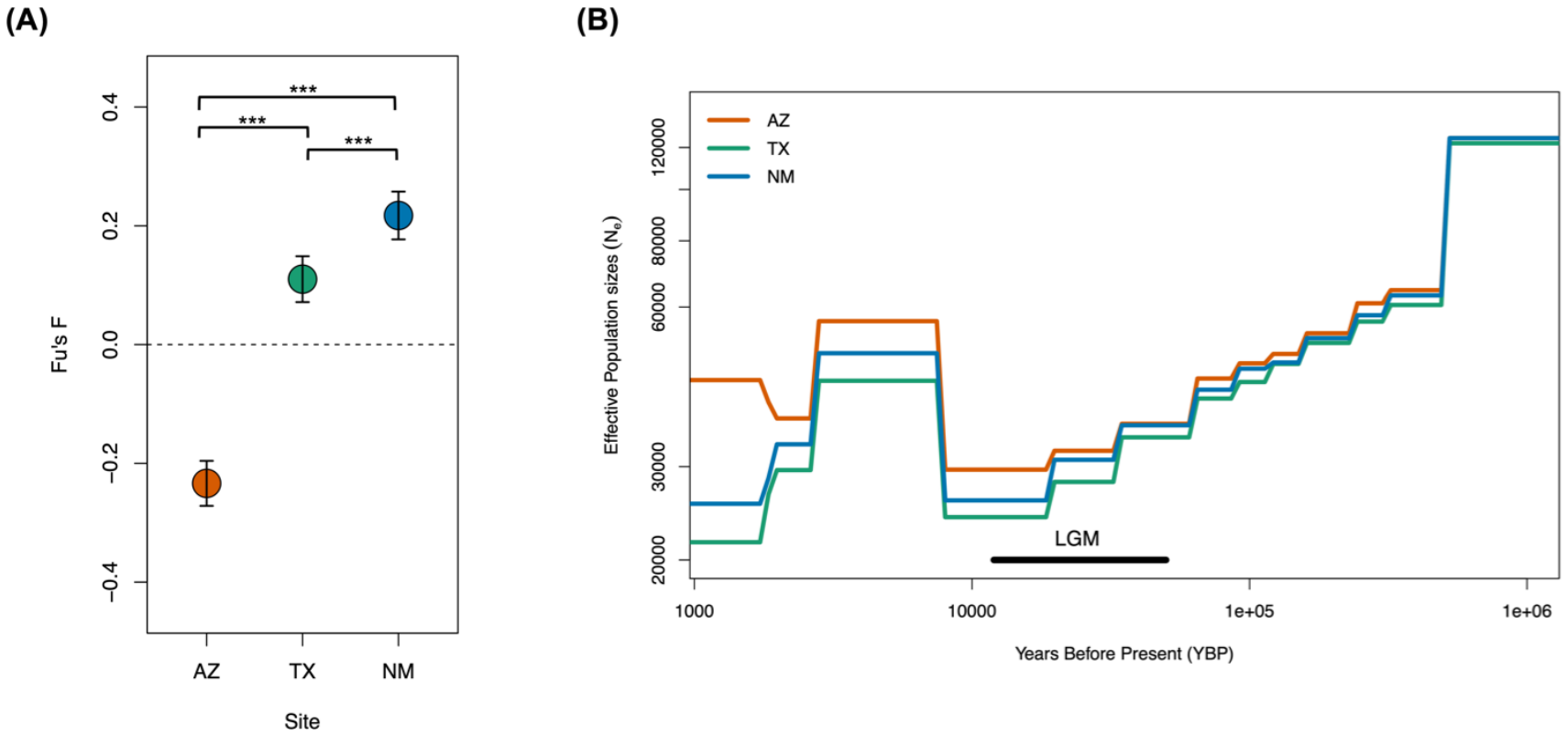


Figure 2.5: (A) Population trends and (B) demographic histories of Montezuma quail. Population trends indicate that only the Arizona population(s) has been expanding (Fu's F < 0) whereas both Texas and New Mexico populations are declining (Fu's F > 0). Error bars indicate 95% CI around the estimate. The data indicate that Montezuma quail experienced a strong historic bottleneck during the last glacial maxima (LGM) followed by re-expansion, and the similar demographic trajectories of each population prior to the LGM suggests that genomic differentiation (Figure 2.4) is relatively recent in origin.

occurred around 10,000 years ago, then populations began to rebound until growth rates became negative around 3000-5000 YBP (Figure 2.5B).

One of the major emphases of our study was to assess the adaptive potential of Montezuma Quail, particularly in the small, isolated Texas population. Variation in protein-coding genes has the capacity to accurately gauge adaptive potential (Barbosa et al., 2018). The trend we observed for the subset of genic diversity was similar to the whole genome data; in both cases there was a ~25% reduction of nucleotide diversity in Texas quail (Table 2.1). In particular, the Texas population had significantly lower (θ_W = 3.94 x 10^{-4}; SE = 2.87 x 10^{-7}) genic nucleotide diversity as compared to both Arizona (θ_W = 5.23 x 10^{-4}; SE = 3.33 x 10^{-7}) and New Mexico (θ_W = 4.47 x 10^{-4}; SE = 3.09 x 10^{-7}; Table S6). Mean heterozygosity (i.e. proportion of heterozygous sites per individual) in the genic regions of Texas quail was significantly reduced relative to Arizona quail whereas Texas and New Mexico samples showed similar levels of genic heterozygosity (Figure 2.3A; Table B7). Our F_{ST} estimates from the genic regions show significantly higher levels of differentiation among the three populations as compared to the whole genomic background (Table 2.2) which indicates that selection as well as drift is contributing to population structure.

To quantify selection and the potential genetic load associated with the genic variants, we compared the deleterious mutations within protein-coding genes (Fig. B8) and their predicted change on translation (Figure 2.6A). Most (82.1%) of the genic variation was due to non-coding intronic sites upstream and downstream of the transcription unit; both of these sources of variation can impact gene expression levels and thus serve as sources of regulatory variation. Exonic sites harbored about 4.5% of the genic variation. Within the exonic SNPs, the Arizona population had a significantly higher proportions of high, moderate, and low impact deleterious mutations, and lower proportions of non-coding variants, when compared to either the Texas or New Mexico populations (Figure 2.6A; Table B8). Our estimates of realized genetic load showed that Texas quail had no significant difference in the mean observed heterozygosity (Fig. B9) or frequencies of highly deleterious mutations per individual ($p>0.05$) as compared to Arizona quail. Most exonic variants were classified as moderate or low impact deleterious mutations, and we found them more homozygous and at higher frequencies in Texas quail as compared to Arizona quail ($p=0.04$ and $p=0.03$; Figure 2.6B; Figure B9; Table B9,10). We realize that frequency estimates based on called genotypes may be biased due to low coverage and sample size (Benjelloun et al., 2019), but we

note that the trends we observe here among different impact classes have also been observed in other natural populations (Grossen et al., 2020).

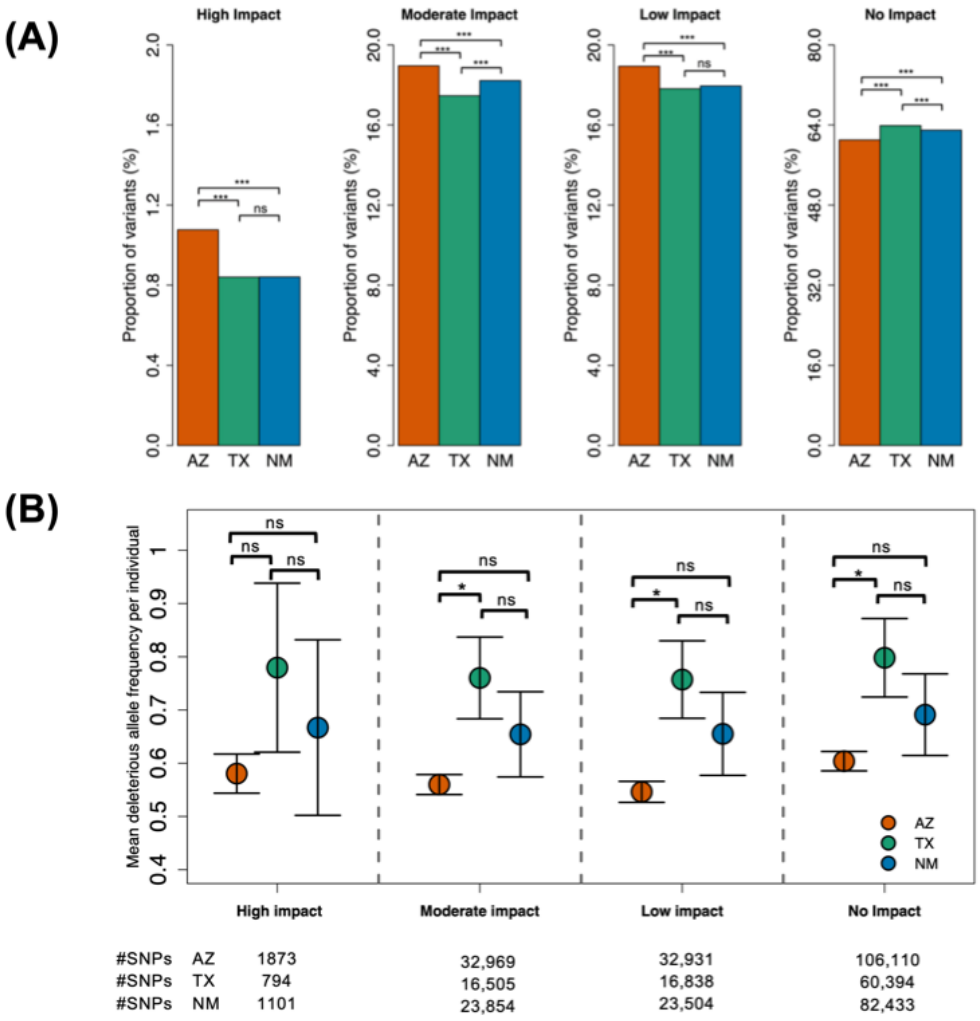


Figure 2.6: Larger populations have higher potential genetic load, but load is more realized in smaller, inbred populations. (A) Potential genetic load was estimated for each population as the proportion of deleterious mutations within annotated protein-coding genes. The Arizona samples had the highest proportions of high impact, moderate impact, and low impact variants. The number of variants in each impact class for each population are noted at the bottom. Note the difference in scales on y-axis. (B) Realized load was measured as the mean frequency of deleterious alleles found within individual genomes for each impact class. No significant difference was found in the frequency of highly deleterious mutations between Texas and Arizona quail, but the small Texas population has a higher frequency of weakly deleterious and non-coding variants coupled with more inbreeding and more homozygosity (Figs. 2, 3, and S9) than the larger outbred Arizona populations. Error bars indicate 95% CI around the estimates.

2.5 Discussion

In this study, we analyzed whole genome sequences from three natural populations of Montezuma Quail that vary in size and habitat continuity (Figure 2.1) to understand how drivers of genetic erosion (e.g., small sizes and isolation) can affect genomic diversity and reservoirs of future adaptive potential. Small populations are predicted to have lower levels of diversity (Soulé, 1985) and recessive deleterious alleles should have a more pronounced impact on fitness than in large populations due to inbreeding (Charlesworth & Charlesworth, 1999). Populations that have experienced declines and are restricted to smaller habitats tend to have lower levels of overall genomic heterozygosity (Brüniche-Olsen et al., 2019; Palkopoulou et al., 2015; Rogers & Slatkin, 2017), but how these factors affect the adaptive potential is far less explored. By comparing levels of genome-wide diversity, genic (i.e., potentially adaptive) diversity, and quantifying genetic load in different populations, our aim was to gain a better understanding of how the adaptive potential of a species is affected by genetic erosion.

2.5.1 Genetic erosion reduces genomic diversity

Our genomic diversity estimates are consistent with predictions for small declining populations that are expected to be most impacted by genetic erosion (Bijlsma & Loeschcke, 2012; Leroy et al., 2018). Species with small populations sizes have lower diversity (Frankham, 1996) and less adaptive potential (Hedrick, Robinson, Peterson, & Vucetich, 2019) than larger populations, and our population genomic data are consistent with these expectations. Montezuma Quail exhibit lower levels of whole genomic heterozygosity than many other avian species (Figure 2.3). The reduction of genomic diversity in Montezuma Quail is reflective of long-term declines in N_e over the last million years (Figure 2.5B). More specifically, Montezuma Quail from Texas are the most genetically depauperate of the populations we surveyed, with genomic diversity similar to vulnerable and endangered birds (Figure 2.3B). Our Texas samples had genome-wide heterozygosity similar to raptors and other large birds (Table 2.1, Figure 2.3B) even though small birds typically have more genetic diversity (Eo, Doyle, & DeWoody, 2011). Overall, we think the data reveal that genomic erosion has likely reduced the evolutionary potential of Montezuma Quail in Texas and that this reduction is unlikely to abate in the absence of gene flow through assisted translocation or other means.

2.5.2 Isolation leads to more inbreeding

A lack of migration among populations limits gene flow and accelerates inbreeding (Frankham, 1996; Gong, Gu, & Zhang, 2010; Hedrick & Garcia-Dorado, 2016; Keller, 2002; Madsen, Stille, & Shine, 1996; Pulanić et al., 2008). Our samples from Montezuma quail populations in the U.S. form independent genetic clusters (Figure 2.2C,D), which is unsurprising given the geographic distances among sampling sites and the limited dispersal capacity of this ground-dwelling bird (Stromberg, 1990). These results are in general accordance with our previous findings based on a small SNP panel (Mathur et al. 2019), but the divide in Arizona (Figure 2.2C; Figure B3) was undetected with that same SNP panel. Our kinship analysis suggests that very few of our samples were derived from related individuals (Figure 2.2A), and our inbreeding estimates show that the Texas population is highly inbred as compared to Arizona and New Mexico (Figure 2.2B). Our samples were acquired opportunistically and that likely reduced the probability of collecting related individuals. However, inbreeding itself can reduce estimates of kinship as inbred individuals may have elevated number of alternate homozygous genotypes and a reduced number of shared heterozygous genotypes (Waples et al., 2019). We observed an elevated incidence of alternative homozygotes for within-Texas comparisons (Figure B10) and we think the collective evidence shows that the small, isolated population of Montezuma quail in West Texas is relatively inbred. This is key, as elevated inbreeding means more of the potential genetic load will be realized (see below).

2.5.3 Impact of genetic drift on population divergence

One of the major drivers of genetic erosion in small populations is genetic drift. In the absence of migration, genetic drift can fix common alleles or lose rare alleles from the gene pool. Isolated populations with historically low sizes can become phenotypically distinct over time (Holycross & Douglas, 2007; Van Belleghem et al., 2018) due to differences in nucleotide composition (D_{XY}) (Wakeley, 1996) or allele frequencies (F_{ST}) (Beaumont, 2005). The intensity of genetic differentiation due to drift is generally expected to be the same for all neutral loci in the nuclear genome due to lack of selection pressures, but it is complicated by linked selection (Cai, Macpherson, Sella, & Petrov, 2009; Rettelbach et al., 2019). Recent population genomic studies have shown that different populations exhibit a heterogeneous differentiation landscape

("differentiation islands") across the genome (Burri et al., 2015; Ellegren et al., 2012). We observe similar results in Montezuma Quail populations (Figure 2.4; Figure B6) where many regions show highly significant values of F_{ST} even though global estimates seem biologically insignificant (Table 2.2). Some of these high- F_{ST} windows no doubt represent statistical artefacts, but many of these highly differentiated regions contain functional genes (Figure 2.4) that could impact various fitness traits (Table B5) and could be signatures of local adaptation (Willoughby et al., 2018). This idea is bolstered by our global estimates of genic differentiation which suggest that coding genes are more diverged than the genome overall (Table 2.2). These analyses suggest that local adaptation could constrain genetic rescue due to the possible reduction in fitness of interpopulation hybrids (Bell et al., 2019; Whiteley, Fitzpatrick, Funk, & Tallmon, 2015). On the other hand, such analyses have the potential to identify source populations that have adaptive genetic signatures most similar to the recipient population and thus the greatest likelihood of success from a long-term, evolutionary perspective.

2.5.4 The adaptive potential of small populations

Understanding the adaptive response of a species to future environmental changes is a high priority for conservation (Holderegger et al., 2019) as this response impacts the long-term probability of persistence (Hedrick et al., 2019), but such an assessment is not straightforward. Genetic erosion is expected to affect adaptive potential by either reducing the overall standing variation in genic regions or by the accumulation of deleterious mutations (Lynch, Conery, & Burger, 1995; Ohta, 1992). We evaluated these two detractors of adaptive capacity by considering variation contained exclusively in genic regions and assessing their possible phenotypic impact. Montezuma Quail have over 17,000 genes and our results show that both nucleotide diversity and heterozygosity in genic regions is lower relative to the whole genomic background (Table 2.1; Figure 2.3A). This is not entirely unexpected as many genes might be evolving neutrally or nearly so, but some are highly conserved and mutations arising at these genes will be deleterious and subject to purifying selection (Rettelbach et al., 2019). Our study thus documents a reduction in both the "nearly neutral" (all) and "adaptive" (genic) fractions of genomic diversity in progressively smaller wild quail populations. These reductions in genomic diversity, including both nucleotide diversity and heterozygosity, are likely to diminish the evolutionary potential of the small, isolated Texas population.

The proportion of deleterious mutations present in the genic regions should reflect the potential genetic load (Charlesworth et al., 1993; Ellegren & Sheldon, 2008; Hedrick & Garcia-Dorado, 2016). Our results show Arizona quail carry significantly more high impact deleterious variants as compared to Texas quail and this difference tends to diminish with variant impact (Figure 2.6A). Most of the genic variants are non-coding (Figure B8) and thus do not impact amino acid sequences but we expect that many serve as regulatory variants that impact expression levels (Harder, Willoughby, Ardren, & Christie, 2020). Recent population genomics studies have shown via simulations (Balick, Do, Cassa, Reich, & Sunyaev, 2015) and empirical data (Ávila, Amador, & García-Dorado, 2010; Do et al., 2015; Rettelbach et al., 2019) that most deleterious genic variants are eventually culled by strong purifying selection, but small effect recessive mutations can persist as seen in our Texas quail. Overall, population genomics data are revealing that most populations can efficiently purge highly deleterious mutations, but small effect deleterious mutants are difficult to purge in small populations where drift predominates (i.e., when $N_e(s) < 1$). In addition to drift, individuals from smaller inbred populations tend to carry these small effect deleterious mutants as homozygotes whereas they tend to be heterozygous in larger outbred populations (Figure B9). It seems clear that large effect deleterious mutations (e.g., *FOXQ1*; (Rogers & Slatkin, 2017)) can have a major impact on fitness. However, most adaptive traits are polygenic and based on many small effect mutations, so small effect deleterious alleles in homozygotes may disproportionately contribute to the overall genetic load of small and declining populations like in Montezuma quail from West Texas.

2.5.5 Conservation considerations

Our results indicate that Montezuma quail populations in the U.S. exhibit low genomic diversity comparable to a number of threatened and endangered species ((Brüniche-Olsen et al., 2019; de Villemereuil et al., 2019; Zhan et al., 2013); Figure 2.3B). Our genomic diversity estimates are consistent with predictions for small declining populations, and we argue that our estimates of genic diversity serve as a reasonable proxy for the evolutionary potential of the species. This study adds to the growing body of literature urging conservation organizations like IUCN to add genetic diversity estimates as a consideration in the listing process (Allendorf, Hohenlohe, & Luikart, 2010; Brüniche-Olsen, Kellner, Anderson, & DeWoody, 2018; Ralls et al., 2018; Willoughby et al., 2015)). Theory suggests that deleterious mutations should be more abundant in small populations

and empirical data support this prediction for species like wooly mammoths (Rogers & Slatkin, 2017) and Iberian lynx (Abascal et al., 2016), with critically low population sizes and ineffective purifying selection. However, most of the species that are declining due to recent anthropogenic activities (like Montezuma quail; Fig. 5B) have maintained relatively large N_e with previous cycles of bottlenecks and re-expansions (Nadachowska-Brzyska et al., 2015). This study and a recent overview of mammals (van der Valk, de Manuel, Marques-Bonet, & Guschanski, 2019) suggest that smaller populations have significantly lower proportions of deleterious mutations as compared to larger, more genetically diverse populations. These deleterious variants are maintained at lower frequencies and presumably represent a major fraction of the potential genetic load. This pattern exists in part because purifying selection against partially recessive deleterious recessive alleles is relaxed in large populations where higher heterozygosity effectively hides these alleles from selection. In contrast, small populations are only likely to purge strongly deleterious mutations, but the collective genetic load of mildly deleterious mutations still impacts individual fitness when these variants are homogenized due to inbreeding and/or drift. Thus, our genomic data illustrate and quantify the incidence of potential genetic load in large populations (Arizona) relative to the realized genetic load in small, inbred populations like Texas.

2.6 Conclusions

We analyzed whole genome sequences from different populations of Montezuma Quail in the U.S and compared the relative impact of genetic erosion between populations of various sizes. Our results indicate that Montezuma Quail populations in the U.S. have mean genome-wide heterozygosity comparable to other avian taxa of conservation concern. We found that inbreeding and random drift due to isolation are the major driving force behind these observed patterns of reduced genomic diversity, but we also identified highly differentiated (candidate) genes that may underlie local adaptations. More interestingly, we find that larger populations carry a larger proportion of deleterious mutations (potential genetic load) than small populations. However, small populations are most susceptible to reduced adaptive potential because small effect deleterious alleles are homogenized due to drift and inbreeding (realized genetic load). Overall, we think these data will be useful to those interested in the conservation of Montezuma Quail, and that they illustrate the power of population genomics in evaluating adaptive potential in light of fragmented landscapes and rapid environmental change.

2.7 Acknowledgements

We thank Dr. Louis Harveson for collecting the Texas samples and Dr. Ryan Luna for the New Mexico samples. We thank Dr. John M. Tomeček and Arizona Department of Game and Fish (J. Heffelfinger) for the hunter-harvested wings. This research was funded in part by the Texas A&M AgriLife Extension Service, the Reversing the Decline of Quail in Texas Initiative, and the National Institute for Food and Agriculture. SM was supported by a Graduate Research Fellowship from the Welder Wildlife Foundation. This article represents publication #### of the Rob and Bessie Welder Foundation. We thank Drs. John W. Bickham, H. Lisle Gibbs, Mark Christie, Ximena Bernal, and members of the DeWoody laboratory group for constructive criticism on an earlier draft of the manuscript.

2.8 Data accessibility

The sequence datasets generated during the current study are available in NCBI's Short Read Archive BioProject accession # PRJNA623948, BioSample accession # SAMN14562436-509 and SRA accession # SRR11514056-129. The scripts developed for analysis can be publicly accessed at https://github.com/samarth8392/MQU_PopGenomics

CHAPTER 3. EVOLUTIONARY HISTORY AND PRESENT LOAD IN THE MONTEZUMA QUAIL

The final chapter of this dissertation will highlight the application of genomics in reconstructing evolutionary history and describing how demography shapes the contemporary genomic diversity and the load of deleterious mutations. In this chapter, we analyzed genome-wide data with higher coverage depth from previous chapter and new samples including individuals from Mexico and Central Texas. The aims of this study are to: i) evaluate how different demographic histories shape the levels of overall GD and shared ancestry among different populations; ii) empirically characterize the distribution of deleterious mutations in small and large populations; iii) determine how these mutations arise and segregate over evolutionary time. The contents of this chapter are currently under preparation for publication. Feedback from the committee members will be used to modify the manuscript prior to publication submission.

3.1 Abstract

The implementation of effective conservation strategies is a challenge due to our limited understanding of how natural populations evolve. Contemporary gene pools and distribution of adaptive mutations with individual genomes depend on past demographic events, effective population sizes, and the history of gene flow. In this study, we analyzed whole genome resequencing data from 98 Montezuma Quail (*Cyrtonyx montezumae*) representing three major populations across the species' range, including locally threatened and isolated populations in Texas. We evaluated the genomic imprint of evolutionary history in different populations and the current distribution of deleterious mutations in small and large populations from assessment of genome-wide variants. We estimated the age of deleterious mutations that are present in populations of different sizes to understand when such mutations arise in evolutionary time and how demographic history shapes their segregation in small versus large populations. We show that Texas quail are significantly more inbred and have maintained low genomic diversity and population sizes after they went through a strong bottleneck ~20,000 years ago. Using empirical data, we demonstrate that when populations undergo a bottleneck, they lose genetic load from the ancestral population and the mutations that survive tend to segregate at higher frequencies in

smaller populations due to the combined effects of drift and ineffective purifying selection. We demonstrate that due to these factors, mutations that arise post-bottleneck (including deleterious) are likely to become fixed at a faster rate in smaller populations. We also highlight that even though smaller populations have fewer deleterious mutations ("potential load") in the gene pool, those mutations still accumulate within individual genomes as homozygotes ("realized load") and can contribute to overall loss of individual fitness. This study illustrates how genetic load in a population is a dynamic characteristic varying over time and thus needs an evolutionary context, and why conservation efforts like assisted gene flow could be beneficial to small populations in alleviating the risks of inbreeding depression.

3.2 Introduction

Humans have altered natural landscapes since the agricultural revolution, but it has been most rapid and at a global scale since industrialization and urbanization (Li et al., 2016). Anthropogenic activities reduce and subdivide suitable habitat for wild species (Fahrig, 2003), creating isolated populations that become smaller with time due in large part to lower mean fitness as genetic/genomic diversity (GD) is lost due to drift and inbreeding (Frankham, 2005). Many modern conservation efforts are targeted towards increasing GD of small populations by introducing individuals from larger and genetically diverse populations to provide additional variation necessary for future adaptation (Whiteley, Fitzpatrick, Funk, & Tallmon, 2015). Successful examples of genetic rescue exist (Frankham, 2015; Ralls et al., 2018; Ralls, Sunnucks, Lacy, & Frankham, 2020), but conservation efforts are complicated by the choice of appropriate source population that maximizes the benefits of genetic rescue but also minimizes the risks associated with migration (Bell et al., 2019). Some recent studies claim that isolated populations could potentially be robust to the loss of GD or that choosing small populations as source population would more likely be more beneficial as they have purged themselves of deleterious alleles (Robinson, Brown, Kim, Lohmueller, & Wayne, 2018; Robinson et al., 2019). Such claims and uncertainties with genetic rescue call for a more critical understanding of the evolutionary histories of natural populations in diverse systems as each species has its own unique evolutionary history shaping contemporary load and vulnerabilities to future extinctions.

The distribution of deleterious mutations within different populations is key to conservation genetics as it contributes to the mean fitness of a population (Barrett & Charlesworth, 1991;

Kimura, Maruyama, & Crow, 1963; Muller, 1950). Genetic load can be defined as the loss of mean fitness due to the accumulation of deleterious mutations (Henn, Botigué, Bustamante, Clark, & Gravel, 2015). By comparing the genetic load, we can categorize population fitness and prioritize efforts towards populations that are most vulnerable to extinction risks. Genetic load depends on the number of deleterious alleles, the magnitude of their effect, their frequency in the population, and zygosity (a function of the breeding system) (Lohmueller, 2014). The genetic load can be estimated at a population level or individual level (i.e. the number of deleterious alleles present in a population vs in an individual genome). Population genetic load could be similar between expanding populations and a population that went through a bottleneck (Do et al., 2015; Simons, Turchin, Pritchard, & Sella, 2014); however, many highly deleterious mutations are recessive (Agrawal & Whitlock, 2011; Mukai, Chigusa, Mettler, & Crow, 1972), so the diploid genotype composition will determine how much of the overall genetic load is realized within individuals (Fu, Gittelman, Bamshad, & Akey, 2014; Lohmueller, 2014). Individuals in a population with significant load may still be relatively more fit if recessive deleterious mutations exist as heterozygotes as compared to a population where recessive alleles are homozygous in individuals due to drift and inbreeding. Thus, the difference between potential genetic load harbored by populations and the reduction in individual fitness due to realized load is an important distinction as it dictates the impact of inbreeding depression.

Herein, our major goals were to: i) evaluate how different demographic histories shape the levels of overall GD and shared ancestry among different populations; ii) empirically characterize the distribution of deleterious mutations in small and large populations; iii) determine how these mutations arise and segregate over evolutionary time. Thus, we analyzed whole genome sequence data for 98 Montezuma Quail (*Cyrtonyx montezumae*) from four wild populations that exist in varying degrees of habitat contiguity across the entire species range (Figure 3.1A). Montezuma Quail are small Galliform birds in the New World Family Odontophoridae; they are mostly found in Mexico along the Sierra Madre Occidental Mountain Range and their range extends into the United States (Leopold & McCabe, 1957). Unlike other North American quail, Montezuma Quail are habitat and diet specialists that rely on underground bulbs and tubers in the climatic belt generally associated with pine-oak forest (Albers & Gehlbach, 1990; Stromberg, 1990). The Montezuma Quail is one of the least studied North American birds and little is known about their biology relative to many other avian species (Hernandez, Garza, Harveson, & Brewer, 2009). In

the U.S., they are found in Arizona, New Mexico, and Texas. Two isolated populations exist within Texas, one Western population in the Trans-Pecos ecoregion and the other population in Central Texas on the Edwards Plateau (Figure 3.1A). Like many montane species, Montezuma Quail populations are declining (Harveson et al., 2007), and local extirpation are major concerns in Texas as their habitat has dramatically reduced in the last century due in large part to land use practices (e.g., domestic livestock grazing)(Brown, 1979; Harveson, 2009; Harveson et al., 2007). Conservation efforts like translocations (e.g., from Arizona or Mexico to Texas) are possible in theory, but there is little information about interpopulation dynamics, and an assessment of genetic load (as a measure of loss of fitness) is desirable. Our aim here is to elucidate how genomics can recover recent evolutionary histories of different populations and to determine how past demographic changes help shape the distribution of contemporary genetic load. We frame our hypotheses in light of recent evidence suggesting that small populations purge the load of high impact mutations (Grossen, Guillaume, Keller, & Croll, 2020; Robinson et al., 2018). Using empirical data, we critically evaluate these competing ideas in a wild non-migratory avian species and discuss the implications of translocations for wildlife conservation. We specifically test whether: (a) larger populations have more deleterious mutations and if they are more abundant as compared to overall functional mutations (i.e. potential load); (b) small populations effectively purge their load of deleterious mutations or if small populations are more vulnerable to loss of fitness due to inbreeding depression. The results from our study not only carry direct implications for the conservation of Montezuma Quail, but also demonstrate how evolutionary principles can be used to winnow competing conservation strategies for at-risk species.

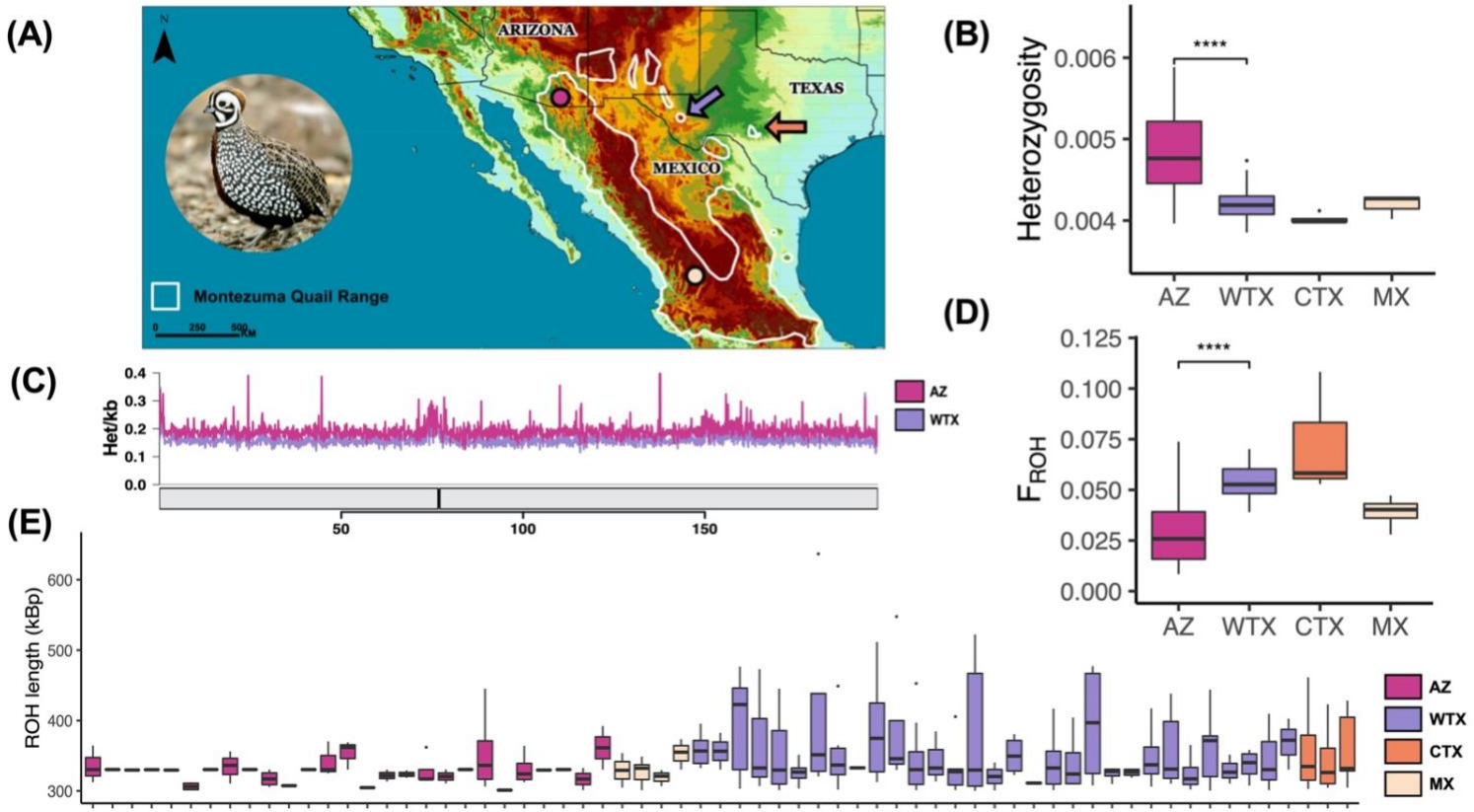


Figure 3.1: Genomic diversity and inbreeding in Montezuma Quail. (A) For this study, we analyzed 66 whole genomes from Arizona (AZ=28, pink circle), West Texas (WTX=31, purple arrow), Central Texas (CTX=3, orange arrow), and Mexico (MX=4, light pink circle). (B) Mean genomic heterozygosity estimates indicate isolated population in WTX are significantly less heterozygous than AZ population and CTX has the lowest mean. (C) Sliding window analysis of heterozygosity across non-overlapping 1kb windows along chicken chromosome 1 show overall reduction in genomic heterozygosity in WTX population (break represents centromere position; see Fig. S2 for all chromosomes). (D) Genome wide estimates of runs of homozygosity (ROH) show that WTX individuals have significantly greater proportion of their genome in ROHs with CTX individuals having highest inbreeding coefficient (FROH). (E) The distribution of ROHs across the genome shows individuals (each column along the x-axis) in WTX and CTX populations have longer ROHs as compared to AZ or MX individuals. Montezuma Quail image courtesy: Bob Gress, BirdsInFocus

3.3 Materials and Methods

3.3.1 Samples & sequencing

DNA was extracted from 98 Montezuma Quail samples from Arizona (AZ; N=60), West Texas (WTX; N=31), Central Texas (CTX; N=3), and Mexico (MX; N=4). Paired end libraries (2x150 bp) were created using a hyper Library construction kit from Kapa biosystems (Roche) and sequenced using an Illumina NovaSeq 6000. The genomes were sequenced at a mean depth of 11.7x +/- 3.3x (mean +/- SD) coverage. See Supplementary materials and methods for sampling and sequencing details (Table C1).

3.3.2 Read alignment, preprocessing, and variant discovery

Raw sequences were trimmed to remove adapter sequences & low-quality bases (Phred < 20) using Trimmomatic v.0.36 (Bolger et al., 2014). Filtered reads were mapped to the chicken genome (*Gallus gallus* assembly GRCg6a) with BWA v.0.7.17 using the *mem* algorithm (Li & Durbin, 2009). We used the Genome Analysis ToolKit (GATK) "Best Practice Workflow" (Auwera et al., 2013) to pre-process our mapped reads (Appendix C). We used both GATK and ANGSD Samtools model (Korneliussen et al., 2014) to discover variants and only retained sites that were concordant between the two methods. For GATK, we used HaplotypeCaller and GenotypeGVCF and only retained Single Nucleotide Polymorphisms (SNPs) using hard filters (QD < 2.0 || FS > 60.0 || MQ < 20.0 || MQRankSum <-3.0 || MQRankSum > 3 || ReadPosRankSum < -4.0 || ReadPosRankSum > 4.0 || SOR > 3.0). For ANGSD, we identified SNPs with minQ = 20 and pval = 1e-6. Our initial combined SNP calling and filtering pipeline identified 32,745,572 SNPs in 98 individuals but there was a high variance in the fraction of genotypes missing for AZ individuals (Table C1), possibly due to degraded DNA and over-representation of GC-rich sequences (Appendix C). To minimize biases in downstream analyses, we removed any individual with >= 20% missing genotypes. We also only retained variants identified in major chicken autosomes and removed any variants from the sex chromosomes, mitogenome, and unplaced scaffolds. We also removed singletons and private doubletons to avoid biases due to genotyping errors. Ultimately, we analyzed genotypes at 12,943,838 SNPs in 66 individuals (AZ=28, WTX=31, ETX=3, MX=4) placed in chicken auto-chromosomes 1-33. Variant statistics, diversity indices, and heterozygosity measured as the

number of heterozygous genotypes per individual were calculated using vcftools (Danecek et al., 2011). We imputed missing genotypes and phased genotypes into haplotypes using Beagle v.5.1 (Browning, Zhou, & Browning, 2018). We used the chicken recombination map (Groenen et al., 2008) to estimate the extent of linkage disequilibrium (LD) between markers and assumed a large diverse ancestral population (N_e =100,000).

3.3.3 Identification of runs of homozygosity (ROHs)

We used PLINK v.1.9 (Chang et al., 2015) to first identify ROHs of size >100 kb using multiple sets of values for parameters associated with scanning window size, SNP density, minimum number of homozygous SNPs within a ROH, minimum number of heterozygotesus allowed, and maximum internal gap between ROHs. We computed the total proportion of genome within ROH (F_{ROH}) (total chicken autosomal genome length = 1,042,168,264 bp) for each set of parameters. To understand the effect of each parameter on ROH identification, we performed a Standardized Regression Coefficients (SRC) analysis to compute the sensitivity indices based on the linear model using sensitivity package in R (Supplementary Methods). For final identification, we used the parameters --homozyg-window-het 2, --homozyg-snp 50, --homozyg-kb 100, --homozyg-window-snp 20. We used the detectRUNS package (Biscarini, Cozzi, Gaspa, & Marras, 2018) in R to visualize PLINK output files.

3.3.4 Relatedness, population structure, and shared ancestry

We examined the degree of relatedness among individual pairs by quantifying R0, R1 and KING-robust kinship statistics using IBSRelate (Waples et al., 2019). For estimating population structure using a model-based approach, we used the genotype likelihood information from all 98 individuals for NGSadmix (Skotte et al., 2013) and genotype calls from best 66 individuals in ADMIXTUREv1.3.0 (Alexander & Lange, 2011). The best value for K was estimated from the rate of change of the likelihood distribution from the 10 runs in NGSAdmix and 10 cross validation steps in ADMIXTURE. Both analyses estimated that all Montezuma Quail individuals derive from k=2 ancestral populations (but higher K also explains the structure between populations; Figure C9). We next used fineSTRUCTURE (Copenhaver, Lawson, Hellenthal, Myers, & Falush, 2012) to estimate the co-ancestry among individuals using phased haplotypes from 66 individuals

(N=132 haplotypes). fineSTRUCTURE implements the chromosome painting algorithm and reconstructs each chromosome of a "recipient" individual as a patchwork of genetic material inherited from potential donors. Individuals from the same or related populations are expected to share more recent common ancestors than are pairs of individuals from historically separated groups, so co-ancestry would be smaller in the latter case. We considered all individuals as donor-recipient pairs and ran the analysis for each chromosome independently using chicken recombination maps for the first 28 chromosomes except chromosome 16 because of lack of loci (Groenen et al., 2008). For each chromosome, fineSTRUCTURE pipeline was executed in 4 steps: the effective population size and mutation rate was first estimated using expectation maximization (EM) algorithm with 10 iterations and then optimized values were used to calculate expected number and the length of inherited segments. Next, an MCMC scheme was used to assign individuals to populations with 10^5 iterations and 50,000 iterations for burn-in, and finally, population trees were built from the converged MCMC runs.

3.3.5 Functional variant annotation and fitness effect classification

We first used both Ensembl Variant Effect Predictor (VEP) (v101.0) (McLaren et al., 2016) and SNPEff v4.3 (Cingolani et al., 2012) to annotate variant sites. Both software provide a subjective impact classification of a SNP ("Modifier", "Low", "Moderate", and "High") based on predicted consequences to protein structure and function. We only retained annotations for protein-coding variants where the impact classes were concordant between the two methods and were without warnings (either due to lack of data or low prediction score). To further classify non-synonymous mutations, we used SIFT (Sorting Intolerant from Tolerant) scores (Ng, 2003) as determined by VEP for all possible amino acid changes in the chicken reference genome (assembly GRCg6a). SIFT scores are generated from sequence homology and the physical properties of amino acids and rank functional i.e. non-synonymous variants based on their deleteriousness. We classified the variants into two categories: "Deleterious", "Weakly deleterious". Deleterious mutations were missense mutations with SIFT score < 0.05, weakly deleterious mutations were missense mutations with SIFT score $\in [0.05, 0.1]$. Variants were categorized as synonymous if they if they had no predicted effect on protein translation or function.

3.3.6 Reconstructing recent demographic history

To infer a demographic model that best fits the joint distributions of the observed site frequency spectra (SFS), we simulated SFS under multiple possible demographic models that could best explain the observed SFS. We only considered synonymous mutations for calculating SFS as they are expected to segregate purely due to demographic changes and not due to selective changes. We used Genetic Algorithm for Demographic Model Analysis (GADMA) (Dobrynin, O'Brien, Koepfli, Ulyantsev, & Noskova, 2020) to automatically find the best demographic model for 2 populations: Arizona and Texas. Here, we combined CTX and WTX samples into a single population (TX) due to high co-ancestry between the two populations. We used AZ (N=28) and TX (N=34) samples and projected the observed SFS in gridof 25 samples from each population [25x25]. To calculate the composite likelihood Akaike information criterion (CLAIC) for goodness-of-fit estimations and confidence intervals (CI) for estimated parameters, we bootstrapped our SFS 100 times. For each model, we started with a series of simple initial structure that includes the number of time intervals that occur before and after each single splitting event (Supplementary Methods) and ran forward time simulations using moments within GADMA. The demographic parameters were locally optimized using default algorithms and the global optimization of the best demographic model was done using 20 iterations. We used a generation time of 1 year as Montezuma Quail are best considered as an annual species. We first chose the final model based on log likelihood estimates and then performed multiple local optimizations to estimate CLAIC scores and CI.

3.3.7 Estimating genetic load and age of deleterious alleles

Genetic load can be viewed from a gene pool level or at individual level. To distinguish the two perspectives, we introduce the terms potential load and realized load. Potential genetic load is defined as the proportion of variants present in the coding sequence that are classified as deleterious or weakly deleterious.

$$Load_P = \frac{\text{Total number of mutations of impact class } i \text{ in individual } j_k}{\text{Total number of non-synonymous mutations in population } k}$$

where $i \in$ (deleterious, weakly deleterious) and k = (AZ, WTX). $Load_P$ is similar to segregating load as described by van Oosterhout (2020) , but instead of comparing the absolute number of deleterious SNPs, we defined $Load_P$ as a proportion conditioned on all non-synonymous SNPs present in a population to avoid bias due to differences in overall GD and to account for all mutations that could possibly be involved in adaptation, directly or indirectly. Thus, population 1 has a higher potential load than population 2 if a larger proportion of the non-synonymous variants in population 1 are deleterious. More diverse populations are expected to have more deleterious mutations, but they may or may not have more $Load_P$ depending on whether those mutations significantly outnumber other non-neutral mutations that indirectly affect fitness.

To estimate how $Load_P$ is being expressed and decreasing the absolute fitness of individuals, we calculated realized load as the proportion of impactful variants that exist as homozygotes within individual genomes. So,

$$Load_R = \frac{\text{Total number of alternate homozygous mutations of class } i \text{ in individual } j_k}{2 \text{ x Total number of sites of impact class } i \text{ in individual } j_k}$$

where $i \in$ (deleterious, weakly deleterious, synonymous) and k = (AZ, WTX). Realized load of synonymous mutations can be viewed as the impact of drift on neutral variation. An individual would have higher $Load_R$ if it carries higher proportion of deleterious alleles in a homozygous state (e.g., due to inbreeding) as compared to another individual where most of the deleterious alleles are heterozygous. We demarcate load as "potential" and "realized" to incorporate the effect dominance plays in the expression of mutational load within individuals and to show that the genetic load of deleterious recessive alleles that segregates at the population level may or may not be widely realized, depending on zygosity (Fu et al., 2014). The inverse relationship between dominance and selection coefficient means that highly deleterious mutations that arise in a population are mostly recessive (Agrawal & Whitlock, 2011) and would rarely homogenize in large outbred populations whereas inbred individuals in smaller populations would have more of the load from those deleterious mutations realized.

How old are deleterious alleles and how long have they segregated in wild populations? We leveraged the phased data from 132 haplotypes to estimate the age of deleterious alleles by inferring the time to the most recent common ancestor (TMRCA) of deleterious mutation between

a pair of haplotypes using Genealogical Estimation of Variant Age (GEVA; Barton, Albers, and McVean (2020)) . GEVA reconstructs genealogical trees of variable sites along each chromosome and estimates the age of any mutation by identifying genomic regions that are identical by descent (IBD) that are shared among haplotypes and broken via recombination. The allele is assumed to be derived from a mutation event in the genome of the common ancestor and shared by all descendent haplotypes. A probabilistic estimate of the time of origin of a mutation is obtained by combining the cumulative distributions for pairs of haplotypes that share the mutation ("concordant pairs") and the pairs that do not ("discordant pairs"). A mutation is expected to be older than concordant and younger than discordant pairs. We used recombination rate estimates from the chicken and an estimated point mutation rate of $3.14x10^{-9}$ as estimated in Mathur and DeWoody (2020). We used the effective population sizes inferred from EM optimization by chromopainter (see above) as a scaling parameter and joint molecular clock to estimate the age of deleterious mutations in different Montezuma Quail populations.

3.4 Results

3.4.1 Isolated Texas populations exhibit less genomic diversity and lack shared co-ancestry

The very few studies that describe the ecology, habitat use, and population dynamics of Montezuma Quail highlight the importance of precipitation and grass cover for their survivability and their vulnerability to landscape alterations (Albers & Gehlbach, 1990; Bristow & Ockenfels, 2011; Brown, 1979; Leopold & McCabe, 1957; Luna et al., 2017; Randel et al., 2019; Stromberg, 1990). Within the US, Montezuma Quail exist in Arizona, New Mexico and Texas (Figure 3.1A) but in this study, we focus on contrasting the Arizona and Texas populations. In Arizona, Montezuma Quail are thought to be abundant on many federal and state-managed public lands to the point where open season for recreational hunting occurs in a sustainable fashion (Heffelfinger & Olding, 2000). In contrast, Texas populations are more isolated, exist largely on private lands, and though listed as a game animal the state has allowed no legal hunting of Montezuma Quail for decades (Harveson, 2009). Since Montezuma Quail are considered habitat specialists with annual survivorship strongly correlated with seasonal rainfall and adequate grass cover, isolated populations in Texas are also suspected to be locally adapted to their micro-climate (Harveson et al., 2007a; Mathur & DeWoody, 2020) Our previous efforts to characterize the population genetics

of Montezuma Quail indicated that Texas quail are genetically distinct from Arizona quail and have significantly lower GD and effective population sizes (N_e; Mathur et al. (2019)). Here, we report whole genome sequences from 98 Montezuma Quail individuals from four distinct geographic regions: Arizona (AZ; N=60), Mexico (MX; N=4), West Texas (WTX; N=31), and Central Texas (CTX; N=3; Figure 3.1A). AZ and WTX habitat are the most similar where AZ marks the western periphery of the Sierra Madre in the Maderean Sky Island landscape in the Sonoran Desert whereas the WTX range is on the eastern periphery in the Trans-Pecos sky island landscape in the Chihuahuan Desert (Leopold & McCabe, 1957). The CTX population is the most unique of all Montezuma Quail habitats and is characterized by the relatively low elevation and less arid Edwards Plateau region (Albers & Gehlbach, 1990). We removed 32 samples from AZ due to a high fraction of missing genotypes (see methods and supplementary info for details associated with these hunter-harvested samples) and ultimately analyzed 12,943,838 variants in the remaining 66 samples. We found a similar pattern of significant reduction in nucleotide diversity (Figure C1 and Table C1) and reduced observed heterozygosity across the whole genome (Fig. 1B) in the WTX samples ($4.2x10^{-3} \pm 2.0x10^{-4}$; mean ± SD) as compared to AZ samples ($4.8x10^{-3} \pm 4.9x10^{-4}$). Our MX samples had comparable levels of mean genomic heterozygosity ($4.2x10^{-3} \pm 1.5x10^{-4}$) to WTX whereas CTX samples had the lowest mean heterozygosity ($4.0x10^{-3} \pm 6.7x10^{-5}$). Our results show that CTX quail from the Edwards Plateau, the smallest population of most concern in Texas, is the least genetically diverse. The lack of statistical power for CTX or MX samples is due to low sample sizes, but Montezuma Quail in Central Texas are the most isolated with only a few hundred individuals thought to persist. Genomic scans in 100kb non-overlapping windows show a remarkably similar diversity pattern between AZ and WTX samples and consistent reduction of heterozygosity throughout the genome (Figure 3.1C, Figure C2). On average, AZ samples have 1 heterozygous site per 5.43 kb of their genome, whereas WTX only had 1 heterozygous site per 6.45 kb. In other words, WTX samples had only 85% of the heterozygosity found in the AZ samples.

Our ROH analysis also captured the loss of GD in Texas quail with WTX samples having significantly higher proportion of their genome within ROHs ($F_{ROH} = 0.054 \pm 0.008$; mean ± SD) than Arizona (0.029 ± 0.016) and CTX samples having the highest mean F_{ROH} (0.073 ± 0.031; Figure 3.1D). Thus, the mean ROH burden is 85% higher in WTX samples than in AZ and 250% higher for CTX birds compared to AZ birds. Longer ROHs are characteristic of recent inbreeding

among close relatives whereas shorter ROHs describe more ancestral inbreeding that has been decaying over generations of recombination events and recent *de novo* mutations (Ceballos, Joshi, Clark, Ramsay, & Wilson, 2018). Across all individuals, we found no ROH > 1Mb (Figure 3.1E) but the average length of ROH (L_{ROH}) was significantly higher for WTX (L_{ROH} = 138 ± 3 kb) as compared to AZ (132 ± 4 kb; Figure C3) and they carry higher number of homozygous SNPs that fall within ROHs (Figure C4). Most ROHs in all populations have relatively short sizes (100-200 kb; Figure C5) and we found only 6 individuals that carry ROHs longer than 500kb (Figure C6) and they all belonged to Texas populations (WTX=5, CTX=1). These ROH distributions indicate that Montezuma Quail populations carry signature of ancient inbreeding, but recombination over time has shortened ROHs and their presumed detrimental effects. The patterns of genome-wide reduction in heterozygosity and higher F_{ROH} in Texas quail reflects their smaller effective population sizes and the impact of drift and inbreeding in reducing diversity over generations.

Consanguineous mating increases the risks of inbreeding depression and thus, it is important to estimate relatedness and shared ancestry among populations as conservation efforts are considered. We estimated relatedness by pairwise comparisons of shared regions that are identical by descent (IBD). IBD fragments tend to coalesce in a more recent common ancestor and are shared between relatives (Copenhaver et al., 2012). The population level genealogy tree for Montezuma Quail shows that Texas quail are more closely related to the ancestors of AZ and MX than to the contemporary populations (Figure 3.2A). The CTX samples are more closely related to one of the WTX lineages. Population average co-ancestry clearly highlights these results demarcating AZ, MX, and TX populations (Figure C7) and are concordant with model-free PCA (Figure C8) and Bayesian clustering (Figure C9) methods. We found of few pair of WTX samples (N=8 pairs) that share highest co-ancestry than any other pairs (Figure C7) and indicate potentially familial relationships. Based on R0-R1-KING kinship analysis (Figure C10), we assigned 1 pair as parent-offspring and 7 pairs as full-siblings (Figure C11). Our kinship analysis indicates that AZ and TX share the least amount of co-ancestry and the MX population is more closely related to AZ population. Additionally, within-population mean co-ancestry is higher and more homogeneous for AZ samples than TX population. We think these patterns likely reflect the habitat disparity between Arizona and Texas with more intra-population gene flow within AZ, whereas extant birds in Texas are more locally isolated and are remnants of more previously diverse

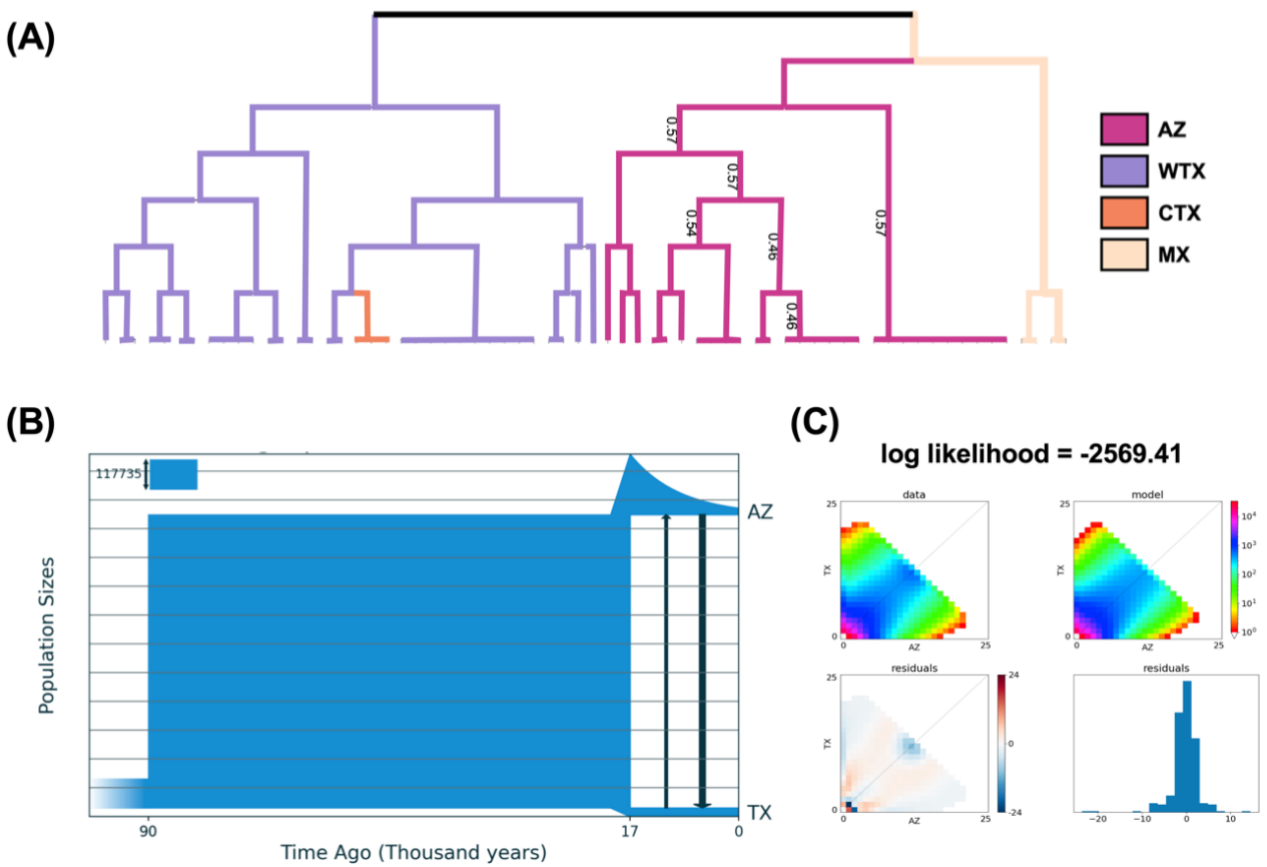


Figure 3.2: Genealogy tree and demographic models of Montezuma Quail populations. (A) Genealogical relationships based on pairwise analysis of identical-by-descent (IBD) segments show that Texas population are genetically distinct with WTX (purple) and CTX (orange) more closely related than AZ (pink) or MX (light pink) populations. (B) The most likely demographic model of Montezuma Quail populations show a demographic expansion 90 kya followed by split 17kya. TX populations have maintained small sizes since their split from the ancestral population whereas contemporary AZ populations are derived from a larger ancestral population formed after a series of bottlenecks. The model also predicts an unequal migration rate between the two populations as indicated by different sized arrows. Gridlines indicate ancestral population size (N_{A0} = 117, 735). (C) The model accuracy was based on highest likelihood values and better simulation of the observed SFS. See Table S2 for model parameters. The SFS was simulated using only synonymous mutations to avoid confounding effects of selection.

populations that were extirpated as habitat degraded over time. Genetic analyses of museum collections from historic TX quail and landscape genetic analyses (Manel, Schwartz, Luikart, & Taberlet, 2003) could shed more light into these observations.

3.4.2 Demographic models predict historic expansion followed by bottlenecks due to founder effects

Previous reconstruction of demographic history using coalescent simulations indicated that Montezuma Quail populations went through a long history of bottleneck in the past 1 million years and Arizona and Texas populations had similar demographic histories until the Last Glacial Maxima (Mathur & DeWoody, 2020). To build the demographic history of the recent past (<100 kya), we performed forward simulations of the Site Frequency Spectrum (SFS) for the synonymous mutations. The most probable demographic model (Figure 3.2B) that most accurately simulated the observed SFS (Figure 3.2C) predicts that ancestral Montezuma Quail populations (N_{A0} = 117,735) expanded 90 kya and maintained large population sizes (N_{A1} = 1,199,304) until the AZ and TX populations split about 17 kya (Figure 3.2B; Table C2). Our models estimate that AZ and TX split in an ~ 80:20 ratio from the ancestral population (i.e. the TX population underwent a more severe bottleneck). Since splitting, the TX populations have maintained smaller population sizes (N_{TX1} = 27,325) until present whereas, larger sized AZ population has been exponentially declining to the current population size (N_{AZ1} = 35,523). The model also predicted an unequal rate of gene flow between the two populations with AZ->TX migration rate 10x higher than TX->AZ migration rate (Table C2). Our estimates of the contemporary population sizes from the demographic model are smaller than previous estimates in Montezuma Quail populations (Mathur & DeWoody, 2020) but still indicates a ~30% demographic reduction in TX as compared to AZ. Collectively, based on the genomic patterns of either inferred or simulated demography, our results imply that after the core populations in southern Mexico expanded demographically, it split into two populations, one ancestral to TX quail and other to AZ-MX quail. The contemporary AZ population was formed as Montezuma Quail ranges expanded in Western Mexico. We think that these results points towards AZ and TX populations representing the two ends of the species "ring" (Mayr, 1999) with the Sierra Madre Occidental Mountain ranges acting as the major landscape barrier between the two lineages (Figure 3.1A) resulting in less shared ancestry (Figure 3.2A, C7) and limited geneflow (Figure 3.2B; Table C2) between the contemporary populations. Future

studies comparing the habitat use between the two populations and a more continuous geographic sampling from the core ranges in Mexico could shed further light into the patterns of spatial expansions.

3.4.3 Despite higher potential load in larger populations, individuals in smaller populations have higher realized load

The genetic contributions to overall fitness are a function of beneficial mutations that increase and deleterious mutations that decrease fitness. Genetic load is typically evaluated based on three metrics: the number of deleterious mutations, their fitness impact, and their frequency within the population (Crow, 1958). To evaluate the loss of fitness in Montezuma Quail populations, we compared the proportion and frequencies of different impact class variants that are present in AZ and WTX populations. One of the arguments against genetic rescue from large populations is the higher number of deleterious mutations carried by larger populations could "infect" small populations via translocations (Robinson et al., 2019). Our results indicate that this could potentially be the case as larger sized AZ population have significantly more non-synonymous SNPs than WTX (~1.12x higher; Table C3) and higher number of deleterious variants (Figure C12). However, we argue that deleterious mutations are not being over-represented in the genome as $Load_P$ (see methods for definitions) is not significantly different between AZ and WTX populations ($p = 0.06$). However, the $Load_P$ of weakly deleterious mutations is significantly higher in AZ ($p=0.002$; Figure 3.3A) possibly because they are less susceptible to purging via purifying selection and can segregate more freely in larger populations. The mean minor allele frequency (MAF) was significantly higher for all types of mutations in TX population than AZ but the difference gets more pronounced for weakly deleterious or synonymous mutations where MAF rise even higher as drift gets more influential (Figure 3.3B; Figure C12). This means overall, larger populations have higher $Load_P$ as they carry more sites that could potentially be deleterious or evade selection, but higher frequencies of deleterious mutations in small populations indicate their prevalence in more in smaller populations.

Population-level productivity, or fitness, is one aspect of genetic health but the other aspect is individual-level fitness in terms of individual reproductive success. By comparing the proportion of the segregating deleterious mutations that exist in a population accumulate within an individual

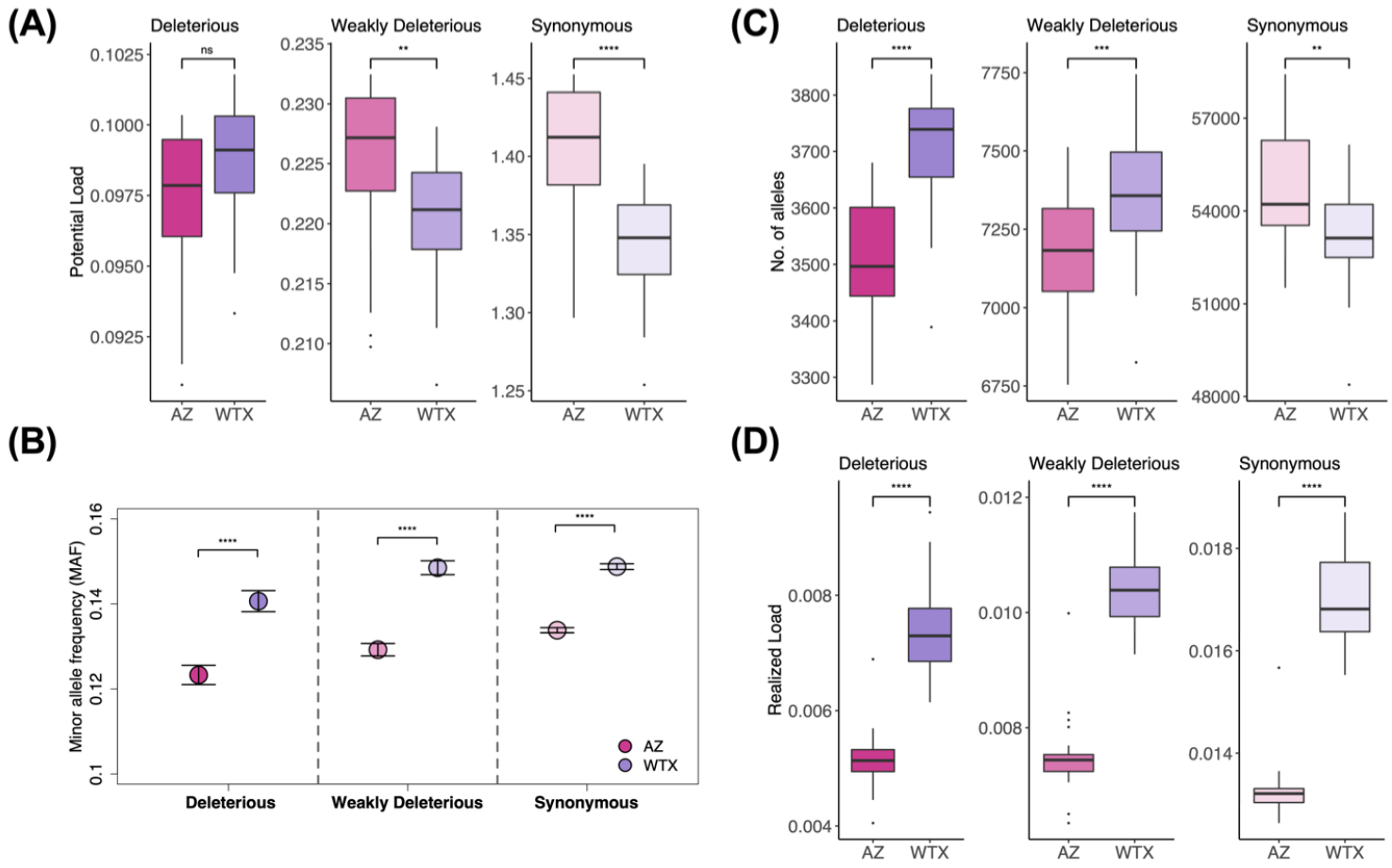


Figure 3.3: Estimates of potential and realized load in Montezuma Quail populations. (A) Potential load is estimated as the proportion of non-synonymous mutations in a gene pool that are categorized as deleterious. We see the potential load of deleterious mutations is similar between the larger AZ and smaller WTX population, but AZ has a higher potential load of weakly deleterious mutations. (B) As expected, the smaller population harbors all impact class mutations at higher mean frequencies. Error bars indicate standard error in the mean estimate. (C) Despite having a lower number of deleterious mutations (Figure C12), individuals in the smaller WTX population have more deleterious and weakly deleterious alleles within their genome whereas WTX birds have fewer synonymous alternate alleles. (D) Realized load is measured as the proportion of deleterious mutations that exist as homozygotes in individual genomes. Smaller WTX individuals have higher Realized load that contributes to inbreeding depression.

genome, we can get a better picture of genomic vulnerabilities and possible inbreeding depression in individuals of small populations. We observed that WTX quail not only have a significantly higher number of deleterious and weakly deleterious alleles within their genome (Figure 3.3C), but also have more detrimental alleles that exist in homozygous states and significantly increases their $Load_R$ (Figure 3.3D).

3.4.4 Populations purge ancestral load during bottlenecks, but smaller populations accumulate contemporary load

Recent studies suggest that strongly deleterious mutations should be purged when wild populations undergo strong human-induced bottlenecks (e.g., due to overharvesting or domestication) (Bortoluzzi et al., 2019; Grossen et al., 2020; Robinson et al., 2018). Our aim here was to elucidate the role of purging (or lack thereof) in shaping the contemporary distribution of deleterious mutations segregating in wild populations that underwent historic and natural bottlenecks. We estimated the age (i.e. the point in time when a mutation first appeared) of deleterious and weakly deleterious mutations carried by both populations and whether they are shared between the two populations or are segregating privately. We then created linear regression models to understand the relationship between the age of a mutation and its frequency in a population. Our results show that almost all the deleterious mutations in Montezuma Quail populations coalesce within the past 150 kya (Figure C13). Both populations have similar age distributions but with deleterious mutation in WTX relatively older (median age = 30,457 years; p-value = 0.012) than the deleterious mutations in AZ (median age = 25,987 years; Figure C13). Most of the deleterious mutations are shared between the two populations (Figure C14) and are significantly older (median age = 50,086; p-value < 2.2e-16) than private mutations (Figure 3.4A). Shared deleterious mutations are segregating at similar frequencies in the two populations (Figure C15) and are likely to rise at the same rate in the two populations (Figure 3.4B). The higher value observed in the smaller WTX population is due to higher starting frequencies (i.e., initial allele frequencies are 1/2N in diploid populations of size N, as indicated by intercepts in Figure 3.4B).

When we look at the SNPs that arose in the two populations in the last 100kya (Figure 3.4C), AZ had more deleterious mutations (N= 11,102) as compared to WTX (N= 10,308). In both populations, most of the deleterious mutations either belong to age class 25kya or older (~50%),

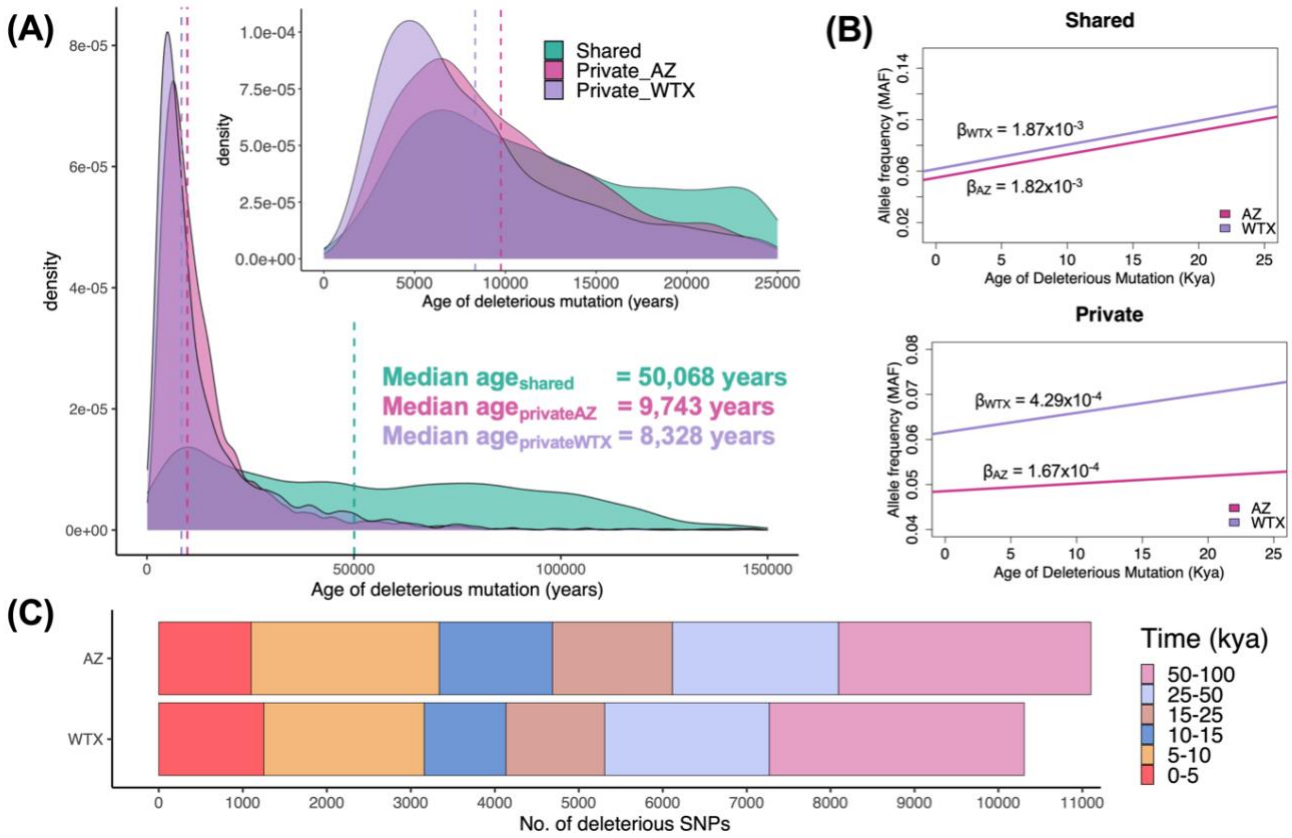


Figure 3.4: Small populations have ineffective purging against *de novo* rising mutations. (A) The age distribution of deleterious mutations that are shared between AZ and WTX population or privately segregating in the past 150kya and 25kya (Inset). Deleterious mutations that arose during the bottleneck were more efficiently removed in small WTX populations but purging has been inefficient against the most recent mutations. (B) Linear model of mutation age and frequency predict similar trajectories for shared deleterious mutations between the two populations but deleterious mutations privately segregating in WTX have higher rate of increase with time. Higher intercept for WTX population indicated higher starting frequency in smaller population. (C) The age distribution of contemporary deleterious mutations in the two populations in the last 100kya. Much of the contemporary load originated in the large pre-bottleneck ancestral population (>50kya) with a smaller proportion of deleterious mutations in the age range corresponding to bottlenecks in Montezuma Quail populations (10-25 kya). The smaller WTX population has higher number of deleterious mutations that arose in the last 5000 years.

or 10 kya or younger (~30%; Fig. S13). Only 20% of the total mutations arose around the bottleneck event 10-25 kya (Figure C13). Most of the younger mutations (age <10 kya) are privately segregating in both populations (Fig. 4A) and we see a rise in deleterious mutations that arose in WTX population in the last 5,000 years (N= 1250) as compared to AZ (N= 1099; Figure 3.4C). Private deleterious mutations in WTX are significantly younger (median age = 8,328; p = 8.872e-12) as compared to private deleterious mutations in AZ (median age = 9,743 years) and are accumulating at the individual level as part of $Load_R$ (Fig. S15). Private mutations in the smaller WTX population also start at higher frequencies (1/2N) but unlike shared mutations, have a ~2.5x faster rate of allele frequency increase as compared to larger AZ population (Figure 3.4B). This pattern is probably because more recent deleterious mutations (i.e., those that arose since the Pleistocene WTX population bottleneck) persist and accumulate because purging is inefficient in the small contemporary population of Texas quail.

3.5 Discussion

Genetic rescue is a promising conservation strategy as the introduction of new alleles to small populations via assisted gene flow can alleviate inbreeding depression in small populations and potentially add to their adaptive capacity to future environmental change (Ralls et al., 2018; Whiteley et al., 2015). Arguments against genetic rescue are based on the idea that a) small populations have likely purged their load of deleterious mutations (Grossen et al., 2020; Robinson et al., 2018) and that b) larger populations carry more deleterious mutations (Lohmueller et al., 2008; Tishkoff & Williams, 2002) so translocations could potentially "infect" recipient populations and inadvertently decrease the already diminished fitness of small populations (Robinson et al., 2019). We tested these ideas in an avian species with large historic sizes but now facing elevated extinction risks due to anthropogenic activities (e.g., human-induced climate change)(Ceballos et al., 2015). We first discuss how the demographic histories of different populations determine the contemporary levels of overall GD to provide an evolutionary background for small populations in need of conservation. We next discuss why, despite carrying more deleterious mutations, deleterious alleles are not overrepresented in larger populations. We also discuss the evidence for purging of deleterious mutations during the bottleneck, but inefficient purging in smaller populations against deleterious mutations that rise post-bottleneck. We

conclude that as deleterious mutations tend to accumulate in genomes of smaller populations, they are more at the risk of inbreeding depression that could be alleviated via genetic rescue.

3.5.1 Evolutionary history shapes contemporary diversity and ancestry among different populations

Peripheral populations are more prone to extinction due to stochastic natural population dynamics (Hampe & Petit, 2005), but they are biologically relevant because marginal populations along geographic clines are presumably more adapted to future range expansion under climate change (Lesica & Allendorf, 1995). Range expansions go hand in hand with demographic bottlenecks as new habit is colonized by a small number of founders, creating a genetic gradient with the least diversity on the leading edge (Excoffier, Foll, & Petit, 2009). Genome-wide analysis of diversity (Figure 3.1), shared ancestry (Figure 3.2A, C7) and the demographic model (Figure 3.2B) suggest that contemporary Montezuma Quail populations in AZ and TX are genetically smaller populations compared to their ancestral populations and were formed recently (< 20,000 generations ago) at the end of the Pleistocene. These signatures of founder effects are buttressed by the ROH distributions which indicate most of the ROHs are short (Figure C5). The TX quail were formed from fewer founders (Table C2) as they went through a stronger bottleneck, whereas the AZ is a more genetically diverse population formed by more founders. We think the Montezuma Quail evolutionary history mimics the patterns of a "ring species" (Mayr, 1999). Our data suggest a larger, relatively diverse ancestral population expanded into the west side of the Sierra Madre Occidental and now forms the contemporary AZ population, and a smaller population on the east of the mountain range formed the TX populations (Figure 3.1A). The relative lack of shared ancestry between AZ and TX populations (Figure 3.2A, C7) and the higher divergence in genic regions between the two populations (Mathur & DeWoody, 2020) all suggest that Montezuma Quail range expansion and colonization that ultimately formed a ring. Ring species are thought to form when populations expand around a geographical barrier and leading edges diverge over time due to lack of gene flow and subsequently become reproductively isolated (Irwin, 2005). However, if there is limited gene flow between the two terminal populations, as was in the case of Montezuma Quail populations (Figure 3.2B; Table C2), such populations are also expected to be unstable and more susceptible to extinction due to environmental change (de Brito Martins & de Aguiar, 2017). Texas populations are currently more at-risk of extirpations partly

due to habitat degradation (Harveson, 2009) and also possibly due to significantly less GD than Arizona quail (Figure 3.1B). The overall levels of individual F_{ROH} ranged from 0.008 – 0.108 per individual, which seems to be comparable to many wild mammalian (Brüniche-Olsen, Kellner, Anderson, & DeWoody, 2018) and avian genomes (Kardos, Qvarnström, & Ellegren, 2017). However, Texas birds are significantly more inbred than AZ (Figure 3.1D) and have longer stretches of autozygous regions (Figure 3.1E, C6). The mean ROH burden is 85% higher in WTX samples than in AZ and 250% higher for CTX birds compared to AZ birds. As longer ROHs are enriched for strongly deleterious variants that disproportionally reduce the overall individual fitness (Szpiech et al., 2013), we argue that Texas quail genomes are more susceptible to future inbreeding depression as they accumulate genetic load (see below).

3.5.2 Distribution of genetic load in small vs. large populations

Genetic load is defined as the loss of fitness due to the multiplicative effects of all deleterious mutations. Since more deleterious mutations segregate in larger populations, recent studies caution against using large, stable populations for genetic rescue. We argue that these studies ignore the genomic background in which these mutations arise (i.e. what proportion of the total variants are predicted to be deleterious vs. benign). To account for all the coding variants, we introduce the term "potential load" ($Load_P$). Theoretically, a population that has a higher proportion of deleterious mutations is more likely to have lower fitness as compared to a population with similar overall GD but a smaller proportion of deleterious mutants. We show that the Arizona population has a higher $Load_P$ of weakly deleterious mutations, but the $Load_P$ of deleterious mutations is comparable between the two populations (Figure 3.3A). This indicates that if Arizona quail were outcrossed with Texas quail, deleterious mutations would not overrepresent the total influx of alleles but would also bring in other non-synonymous mutations that could directly or indirectly increase fitness. Extra functional variants entering the population could decrease the load of deleterious mutations that are currently segregating at higher frequencies in Texas quail (Figure 3.3B).

Purifying selection plays an important role in maintaining GD and fitness by purging deleterious mutations from the populations. We show that when Arizona and Texas populations were formed, both populations lost ancestral load (Figure 3.4C) as the fewest currently segregating mutations belong to the age range that corresponds to the late Pleistocene bottleneck event (15-25

kya; Figure C13). Immediately after the Pleistocene bottleneck (15 kya), the smaller WTX populations presumably accrued fewer deleterious mutations than AZ (10-15 kya) but deleterious mutations that arose in the last 5kya are now overrepresented in WTX populations (Figure 3.4C). These privately segregating mutations are not only significantly younger in the smaller WTX population (Figure 3.4A) but are likely to rise in frequency at a faster rate due in part to starting at higher frequencies (Figure 3.4B). These results highlight the inefficiency of purifying selection in smaller populations to cull "contemporary mutations even though selection might have previously culled ancestral load. The evidence of weakened purifying selection in small populations also comes from higher $Load_R$ of deleterious mutation (Figure 3.3D) where both standing deleterious variants and *de novo* deleterious variants are homozygous in individual genomes (Figure C15). Since deleterious mutations are mostly recessive and detrimental to fitness, their presence in Texas genomes is indicative of the potential for inbreeding depression in small populations.

In conclusion, we argue that genetic rescue remains a promising conservation strategy, but it needs an evolutionary context before implementation. Our empirical genomic data show that populations tend to purge ancestral deleterious mutations during bottlenecks that occur over evolutionary timespans, but recent deleterious mutations persist and add to the genetic load of small contemporary populations where purifying selection is ineffective. In principle, translocating individuals from a population with less $Load_P$ is less likely to increase overall load and more likely to reduce $Load_R$ of vulnerable populations. Besides translocations, habitat restoration and proactive genetic monitoring (Leroy et al., 2018) of source and recipient populations should also be implemented as a part of overall management plan (Holderegger et al., 2019). Advancements in genomic and computational resources provide an excellent opportunity to understand the evolution of wild species, assess loss of fitness due to load of deleterious mutations, and the risk of future inbreeding. These assessments are necessary to devise a more robust strategy to help prevent extinctions in today's world.

3.6 Acknowledgements

We thank J. Heffelfinger at the Arizona Department of Game and Fish for the hunter-harvested wings. We also thank Genaro Olmos, Erasmo Olmos, Juan Felipe Montoya for their Mexican field work support; Alberto Macías Duarte, Angel B. Montoya, James Weaver, Ryan Schmidt, Joyce Moore and Mike Miller for Texas field work support; and Kristyn Stewart, Dr. Ryan Luna and Dr.

Fidel Hernandez from Texas A & M University-Kingsville. This research was funded in part by the Texas A&M AgriLife Extension Service, the Reversing the Decline of Quail in Texas Initiative, the National Institute for Food and Agriculture, Colegio de Postgraduados Campus San Luis Potosí, Caesar Kleberg Wildlife Research Institute and Texas Parks and Wildlife Department. SM was supported by a Graduate Research Fellowship from the Welder Wildlife Foundation. We thank Drs. John W. Bickham, Mark Christie, Ximena Bernal, and members of the DeWoody laboratory group for constructive criticism on an earlier draft of the manuscript.

3.7 Data accessibility

The sequence datasets generated during the current study will be available in NCBI's Short Read Archive upon publication. The scripts developed for analysis will also be publicly available on GitHub after publication.

APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 1

A.1 Genome assemblies and genome size estimation

After culling low quality and adapter sequences, both paired-end and mate pair reads were used for contig building, whereas mate pair sequences were used for scaffolding and gapfilling. AbySS v2.0.2 (Simpson et al., 2009) was used to assemble the reads with the most optimum k-mer of 60 estimated using KmerGenie (Chikhi & Medvedev, 2013). SOAPdenovo2 (Luo et al., 2012) was also used to assemble reads using a multi k-mer approach from 35 to 65. GapFiller (Boetzer & Pirovano, 2012) was used for filling the gaps in the SOAPdenovo assembly. A third assembly was created by merging the AbySS assembly and the SOAPdenovo assembly into a single assembly using GAM-NGS (Vicedomini et al. 2013) with a block size of 50. All the assemblies were then assessed for completeness using BUSCO (Simão et al., 2015) and Quast (Gurevich et al. 2013). The nuclear genome size was estimated using the K-mer approach by counting all the k-mers of length 60 and calculating the mean coverage from the frequency histogram.

A.2 Xenobiotic removal and repeat masking

Once the most complete assembly (i.e., with the highest N50) was chosen, the assembly was filtered prior annotation and variant calling. Xenobiotics were removed according to Antonides et al. (2017) in a two-step process. Firstly, the Montezuma quail assembly was compared to all publicly available avian genomes (top-level sequences unmasked) downloaded from Ensemble using NCBI BLASTN. All the sequences that did not significantly match any of the reference avian genomes ($e < 10\text{-}6$) were culled. Next, these filtered sequences were further compared to all the vertebrate sequences from GeneBank database. Ultimately, any sequence in the draft assembly that did not match with any other avian or vertebrate species was removed as putative xenobiotic in origin. Repetitive elements were identified and classified using both homology-based and de novo approaches. For homology-based approach, the repeats were masked by RepeatMasker (Smit et al., 2015) using the chicken genome. RepeatModeler (Smit & Hubley, 2015) was used as a de novo approach for interspersed repeat discovery, and Tandem Repeats Finder (Benson 1999) was used to detect tandem repeats. The results from the two approaches were combined and the overlaps were removed using the perl script used by Doyle et al. (2014).

A.3 Genome annotation

The filtered assembly with masked repeats was used for genome annotation using the iterative MAKER 2.31 (Holt & Yandell, 2011) pipeline, which includes BLAST for alignment of supplied protein and EST sequences to the assembly, and the gene predictors SNAP (Cantarel et al., 2007) and Augustus (Stanke & Waack, 2003). First, all the manually annotated and reviewed proteins from all avian species found in the Swiss-Prot database of UniProtKB (http://www.uniprot.org) were provided for the initial MAKER run to create gene models, along with assembled sequences greater than 2Mb. The *ab initio* models created from the initial run was used to train the two gene predictors: SNAP and Augustus. Augustus was run with the "chicken" model. Subsequently, the Montezuma quail assembly consisting of scaffolds greater than 5000 bp was supplied to MAKER along with ESTs from the chicken (*Gallus gallus*) and Japanese quail (*Coturnix japonica*) for an additional source of gene modeling. The quality metrics Annotation Edit Distance (AED) and Quality Index summary (QI) were used to assess the strength of each final gene model (Cantarel et al., 2007). The next step was to functionally annotate the genome with putative protein domain and gene functions using Interproscan. InterPro, Pfam, and Gene Ontology (GO) terms were assigned to the gene models. Lastly, the final annotation set consisted of MAKER gene models with an AED <1 and/or at least one identified protein domain.

A.4 SNP identification

Nuclear SNPs were identified using the male Montezuma quail draft assembly with contigs greater than 500bp using Genome Analysis Toolkit (GATK) pipeline (McKenna et al., 2010). The filtered paired-end reads from both male and female Montezuma quail were aligned to the male assembly with the BWA (Li & Durbin, 2009). Mapped reads were sorted and marked for duplicates using PicardTools (http://picard.sourceforge.net). GATK 3.6 was used to identify and realign reads around insertions/deletions (indels). To find sites polymorphic in both male and female, a joint genotyping analysis of the genomic VCFs produced for both individuals was performed. Variants were called in each sample individually using HaplotypeCaller in genomic VCF (GVCF) mode to produce a comprehensive record of genotype likelihoods and annotations for each site in the genome. Target SNPs were filtered for quality (Phred >30) and read depth (>20).

To identify mitochondrial markers, we initially identified mitochondrial sequences by mapping paired end reads to the mitochondrial assembly using BBSplit tool in BBMap (https://github.com/BioInfoTools/BBMap). We then identified the variants in those sequences using the same protocol as described above and retained SNPs with coverage depth of > 200X.

A.5 SNP panel design and error rate

Most of our road-killed samples from Texas were degraded and the DNA samples were of low quality. Thus, a specific target amplification (STA) step was incorporated to increase template material for downstream amplification by using a low molar concentration of each primer and limited thermal cycles (e.g., DeWoody et al., 2017). Genotyping error rate was calculated using replicate DNA samples (n = 4 in total) from 2 individuals following the equation $e = \frac{m}{(d)(s)}$, where *m* represents the total number of mismatches between each replicate sample, *d* represents the total number of loci per replicate sample, and *s* represents the total number of replicate samples (Doyle et al., 2016; DeWoody et al., 2017).

A.6 Population structure analysis

To estimate population structure using sNMF algorithms in the LEA package, we need to optimize the regularization parameter (α) that controls the regularity of ancestry estimates over geographic space. To choose the appropriate value of α, the mean values of the cross-entropy criterion for each cluster (K = 1 to 10) were plotted as a function of α and the value of α that had the minimum mean cross-entropy was chosen. Once the regularization parameter was set, the algorithm was run again with number of clusters set to K = 1-10 with 10 iterations each.

For the STRUCTURE analysis of all samples, loci that significantly deviated from Hardy-Weinberg equilibrium in our (relatively large) Arizona samples were excluded. Then, the analysis was run for K = 1–10, running each value 10 times with an initial burn-in of 100,000 MCMC (Markov chain Monte Carlo iterations) and 1,000,000 subsequent iterations. Due to unbalanced sampling for each geographic region, the relative admixture levels between populations parameter (α) was *a priori* set at 0.33 and an individual α for each population with uniform prior and maximum value of 10 (Wang 2017). For the detailed analysis of Arizona samples, all the SNPs were analyzed the STRUCTURE parameters were set at default using the admixed ancestry model

and correlated allele frequencies. The results of analysis were interpreted using mean likelihood values of K and ΔK (Evanno et al., 2005).

A.7 Outlier tests for selection

We removed samples from New Mexico (N=5) from analysis for outlier tests due to small samples size and only considered Arizona and Texas samples. To identify signatures of natural selection, we used multiple approaches: an FDIST approach (Beaumont & Nichols, 1996) implemented in both LOSITAN (Antao et al., 2008) and Arlequin v3.5 (Excoffier & Lischer, 2010), and a Bayesian approach in BAYESCAN 2.1 (Foll & Gaggiotti, 2008).

FDIST methods estimate the expected distribution of F_{ST} vs. expected heterozygosity under an island model of migration, and outliers are identified based on significant deviations from neutral expectations (Beaumont & Nichols 1996).

In BAYESCAN, a logistic regression model is used that implements a locus effect and a population effect. A locus is assumed to be under selection when the locus-specific component is necessary to explain the observed pattern of diversity (Foll & Gaggiotti, 2008).

LOSITAN was run with 1,000,000 simulations assuming an infinite allele mutation model. The outliers were identified assuming a confidence interval of 0.995 and a false discovery rate (FDR) rate of 0.05. For Arlequin, we ran 20,000 simulations with 100 demes with all samples as a single group and minimum and maximum expected heterozygosities of 0 and 0.8, respectively. Only the loci beyond the 1% quantile were assumed to be significant.

BAYESCAN was initialized with 10 pilot runs of 5000 iterations and an additional burn-in of 50,000 iterations with a thinning factor of 20 (total 150,000 iterations) to identify outlier loci by F_{ST} amongst the geographically distinct regions.

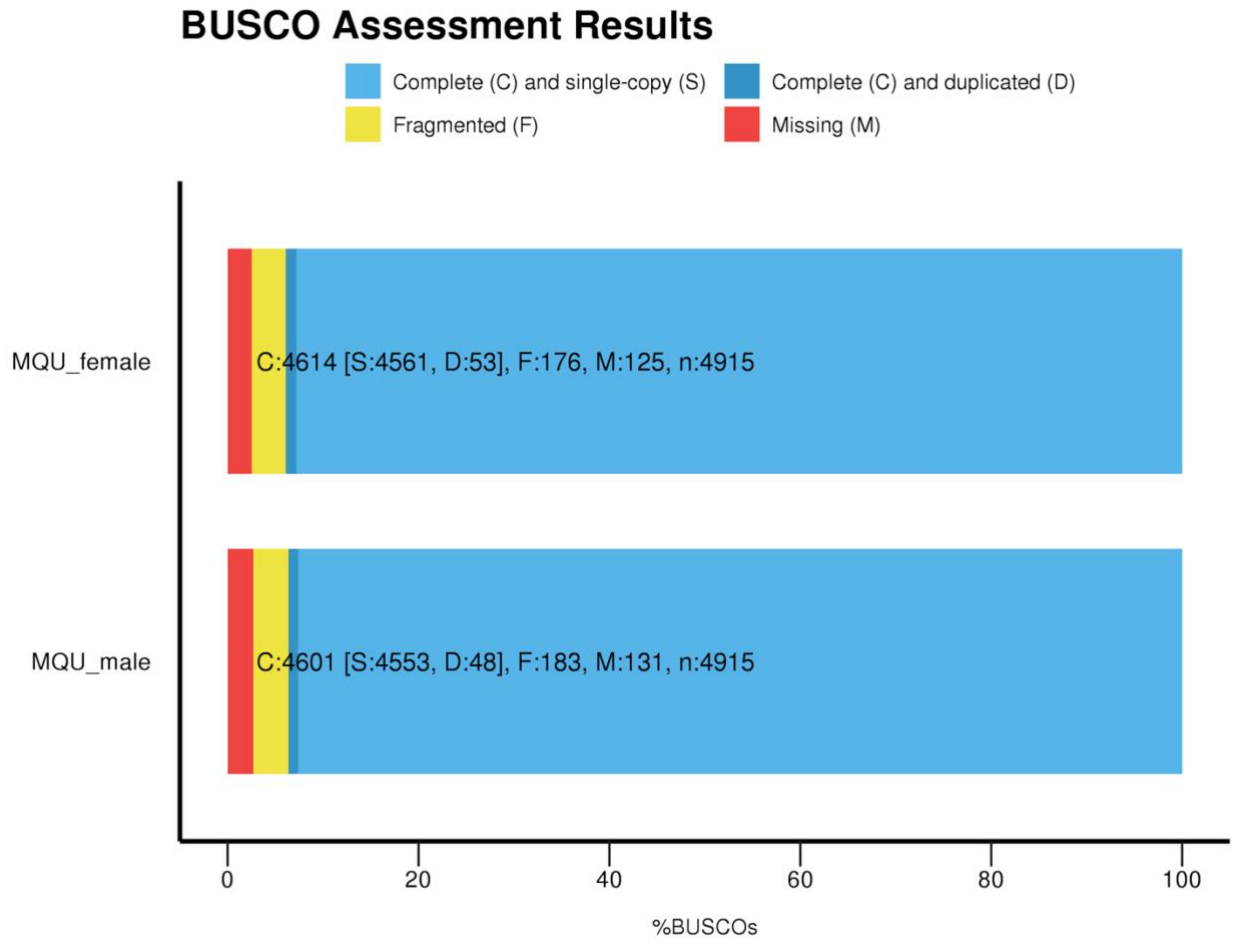


Figure A1 BUSCO single-copy vertebrate ortholog detection summary for the female (MQU_female) and male (MQU_male) Montezuma quail genome assemblies. The x-axis represents the percentage of BUSCOs relative to the avian dataset (aves_odb9). Both assemblies contain the vast majority of genes known in other avian species. The numbers in bars indicate the absolute numbers of Complete (C), Fragmented (F), Missing (M) BUSCOs identified in the assembly.

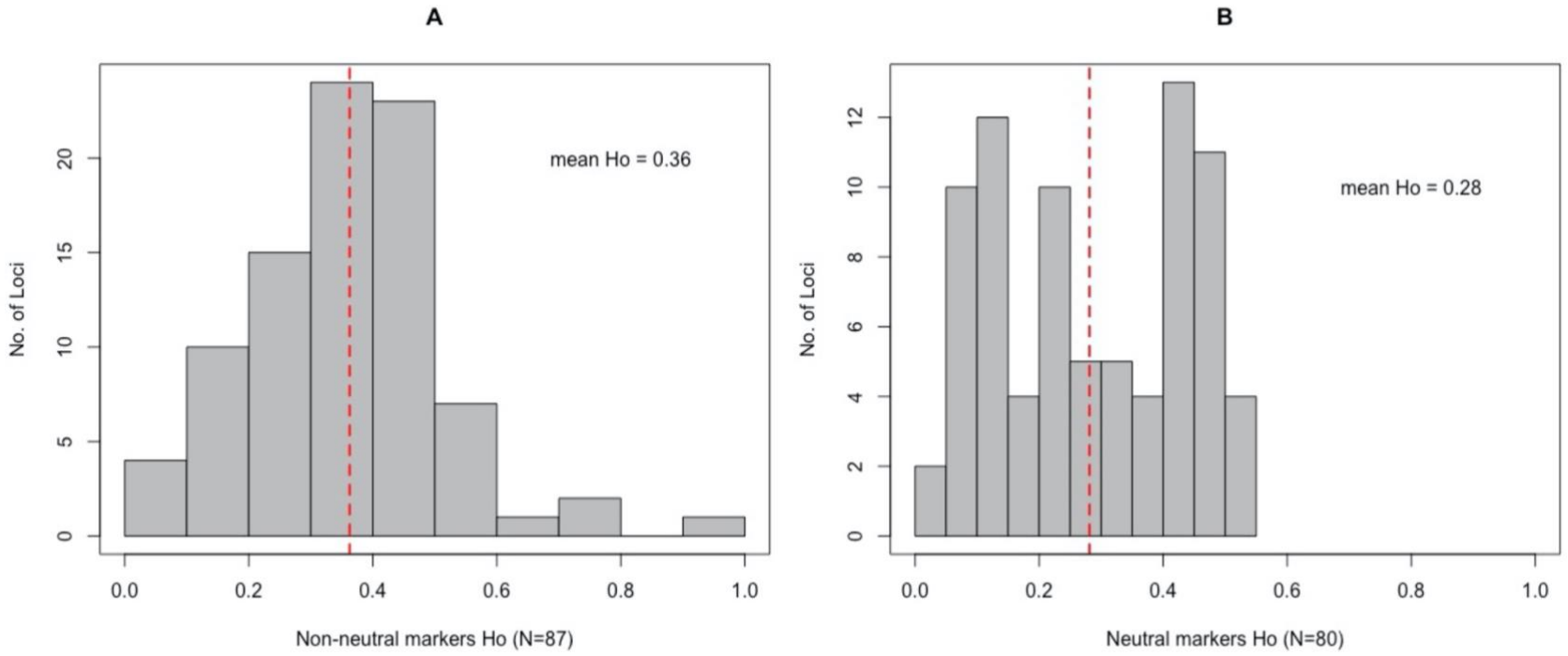


Figure A2 Histograms of observed (Ho) heterozygosity for (A) non-neutral markers and (B) neutral markers. Heterozygosity observed at non-neutral loci (mean Ho = 0.36 ± 0.16) was significantly higher than the heterozygosity observed at presumptively neutral markers (mean Ho = 0.28 ± 0.15; Mann–Whitney Rank Sum Test: U = 2633.5, p-value < 0.05) (Red dashed lines). This observation of higher heterozygosity at non-neutral markers is consistent with the idea of overdominance in Montezuma quail populations.

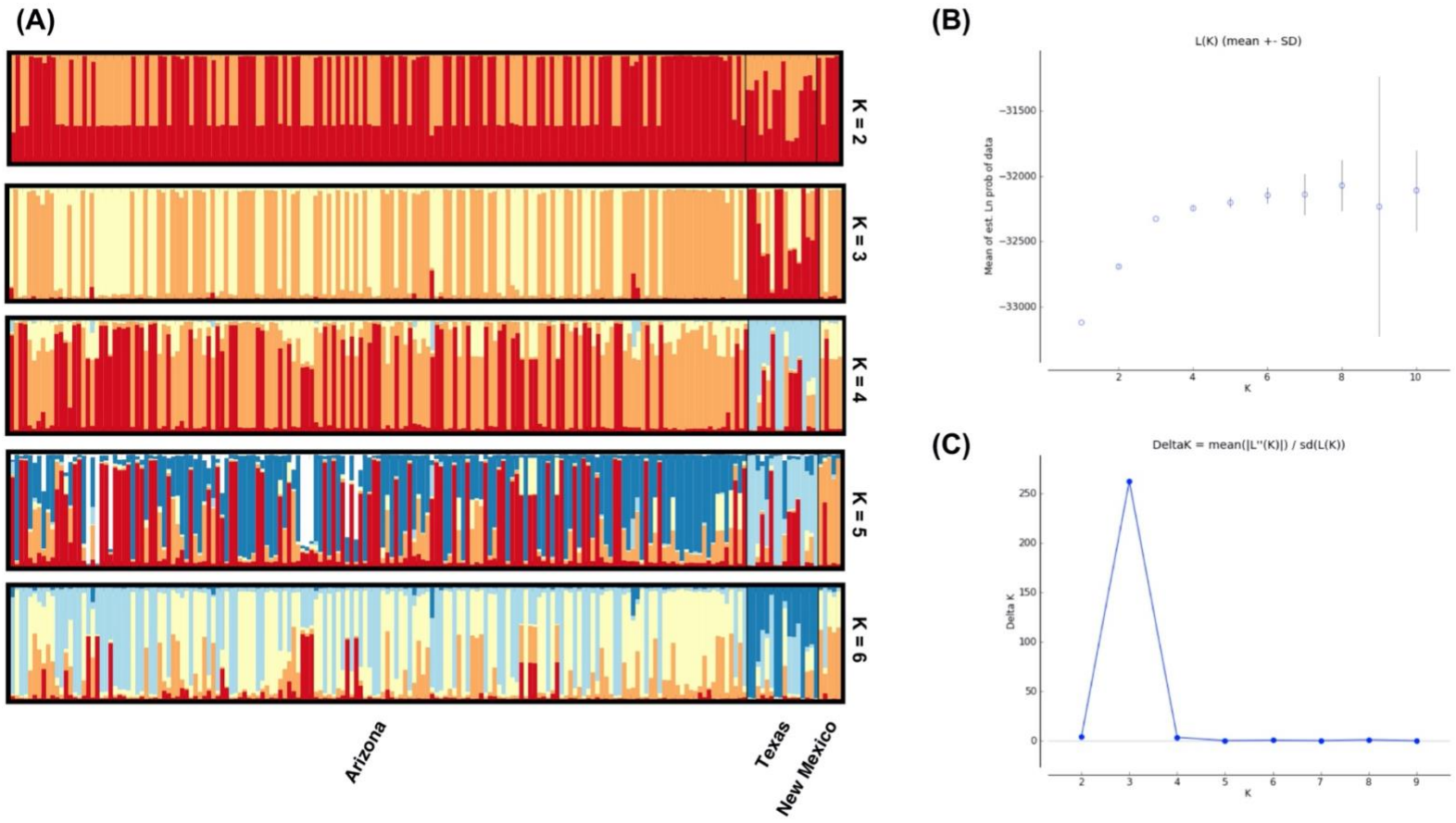


Figure A3 Results from STRUCTURE analyses for all Montezuma quail samples (N=186) genotyped at 141 loci (i.e., using those loci consistent with HWE). (A) At K=3 and K=4, the Texas population is genetically divergent from samples collected in Arizona and New Mexico. (B) Mean estimated Ln probability of data ± SD. (C) Delta K values for each K = 1-10 indicates the best estimated number of ancestral population is K = 3 based on Evanno et al. (2005). Based on these data, the Arizona and New Mexico sampling sites are genetically more similar to each other than either is to the Texas site.

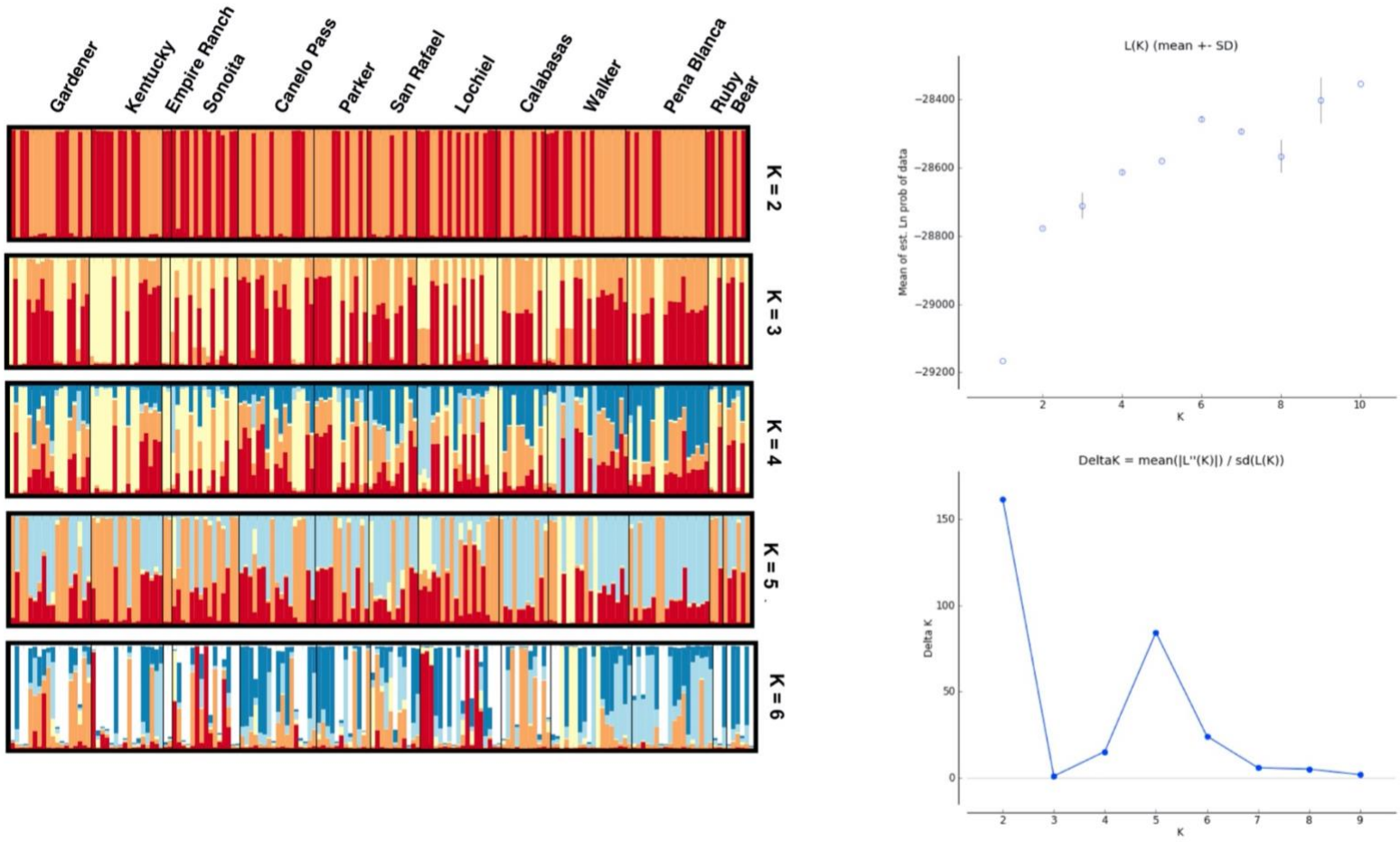


Figure A4 Results from STRUCTURE analyses for Montezuma quail samples from Arizona (N=165) genotyped at 169 loci. Overall, the samples from Arizona are relatively homogenous. These data suggest that the genetic distinctiveness of the Texas samples (Fig. S3, K=3 or K=4) is not an artifact of structure within Arizona

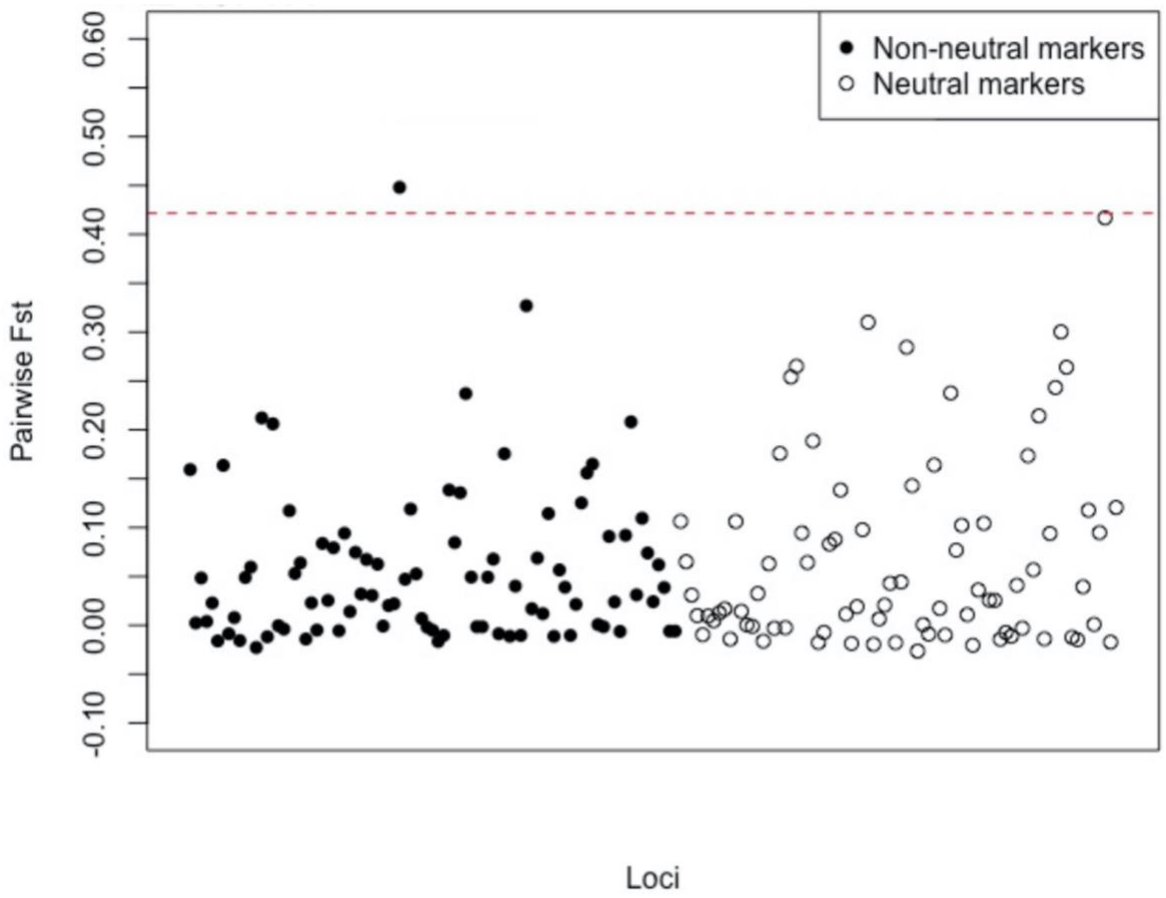


Figure A5 Pairwise FST values for Arizona versus Texas samples. SNP loci (N=169) are ordered on the x-axis with 87 non-neutral markers (black dots) and 80 neutral markers (white dots). Red dotted line indicates top 0.5 % (99.5th quantile) of FST values. We found high levels of locus specific differentiation between Arizona and Texas samples with no significant difference between non-neutral and neutral FST values (p = 0.8)

Table A1 Top 50 Pfam domain hits and their counts in the Montezuma quail genome

Pfam Domain	Counts
WD domain, G-beta repeat	502
Cadherin domain	415
Fibronectin type III domain	403
Immunoglobulin I-set domain	363
Protein kinase domain	329
Leucine rich repeat	277
7 transmembrane receptor (rhodopsin family)	265
Ankyrin repeats (3 copies)	259
Collagen triple helix repeat (20 copies)	249
RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	229
Kelch motif	196
EGF-like domain	180
C2 domain	175
Homeobox domain	172
PDZ domain (Also known as DHR or GLGF)	172
Immunoglobulin domain	161
Calcium-binding EGF domain	159
BTB/POZ domain	157
Tetratricopeptide repeat	145
Ion transport protein	143
Sushi repeat (SCR repeat)	135
Spectrin repeat	133
Ras family	130
PH domain	128
Thrombospondin type 1 domain	124
LIM domain	121
Low-density lipoprotein receptor domain class A	120
C2H2-type zinc finger	113
CUB domain	110
Protein tyrosine kinase	107
Immunoglobulin V-set domain	105
Variant SH3 domain	101
Leucine Rich repeat	100
Scavenger receptor cysteine-rich domain	99
Laminin EGF domain	98
Mitochondrial carrier protein	92
Zinc finger, C2H2 type	88
EF-hand domain pair	87
Armadillo/beta-catenin-like repeat	87
Nebulin repeat	81
SH2 domain	79
Helicase conserved C-terminal domain	79
Helix-loop-helix DNA-binding domain	79
SH3 domain	77
von Willebrand factor type A domain	77
SAM domain (Sterile alpha motif)	77

Table A2 Candidate loci that appear to be targets of positive selection as inferred by outlier analyses in both LOSITAN and Arlequin

LOSITAN			
Assay ID	**Gene**	**Gene Name**	**Pairwise Fst Arizona – Texas**
GTA0201149	ACAN	Aggrecan core protein	0.03
GTA0200995	AQP9	Aquaporin-9	0.05
GTA0200999	MSTN	Myostatin	0.11
GTA0201113	CLOCK	Circadian locomoter output cycles protein kaput	0.13
GTA0201114	FTO	Alpha-ketoglutarate-dependent dioxygenase	0.11
Arlequin			
GTA0201098	MECOM	MDS1 and EVI1 complex locus protein EVI1	0.45

Table A.3 Description of 96 Montezuma quail SNPs associated with genes under selection in other avian species.

Fluidigm® Assay ID	Scaffold Position	Gene	Gene Name
GTA0200985	scaffold.2419218_392333	TGFA	Protransforming growth factor alpha
GTA0200986	scaffold.2420599_1685901	MBL	Mannose-binding protein
GTA0200987	scaffold.2419560_292758	TH	Tyrosine 3-monooxygenase
GTA0200991	scaffold.2419672_841757	IGF1R	Insulin-like growth factor 1 receptor
GTA0200992	scaffold.2420575_3714577	SPP1	Osteopontin
GTA0200995	scaffold.2420503_1541415	AQP9	Aquaporin-9
GTA0200999	scaffold.2420000_352123	MSTN	Myostatin
GTA0201001	scaffold.2420758_896004	Snap25	Synaptosomal-associated protein 25
GTA0201003	scaffold.2419915_166790	TRH	Pro-thyrotropin-releasing hormone
GTA0201004	scaffold.2420075_160697	FECH	Ferrochelatase, mitochondrial
GTA0201005	scaffold.2420129_675643	GABRG2	Gamma-aminobutyric acid receptor subunit gamma-2
GTA0201006	scaffold.2419827_1273517	LEPR	Leptin receptor
GTA0201007	scaffold.2419481_281608	BLNK	B-cell linker protein
GTA0201011	scaffold.2419704_24613	PMEL	Melanocyte protein PMEL
GTA0201018	scaffold.2420934_1575034	CRY2	Cryptochrome-2
GTA0201019	scaffold.2419691_559449	GHR	Growth hormone receptor
GTA0201022	scaffold.2420813_600376	NOS2	Nitric oxide synthase, inducible
GTA0201023	scaffold.2419135_70105	KRT6A	Keratin type II cytoskeletal 6A
GTA0201026	scaffold.2420058_94346	HBB	Hemoglobin subunit beta/beta'
GTA0201027	scaffold.2419201_81592	NPY	Pro-neuropeptide Y
GTA0201031	scaffold.2419496_865363	LYZ	Lysozyme C
GTA0201033	scaffold.2421001_1115644	GHRHR	Growth hormone-releasing hormone receptor
GTA0201036	scaffold.2419192_19188	CCK	Cholecystokinin
GTA0201037	scaffold.2420541_3716953	PRKAA2	5'-AMP-activated protein kinase catalytic subunit alpha-2
GTA0201040	scaffold.2419861_205633	PRL	Prolactin
GTA0201041	scaffold.2419774_409716	ADAM17	Disintegrin and metalloproteinase domain-containing protein 17
GTA0201042	scaffold.2419870_19178	ARNT	Aryl hydrocarbon receptor nuclear translocator
GTA0201045	scaffold.2419193_19305	Feather	Feather keratin 4
GTA0201046	scaffold.2420736_1062	MC1R	Melanocyte-stimulating hormone receptor
GTA0201049	scaffold.2419236_1490598	PPARG	Peroxisome proliferator-activated receptor gamma
GTA0201051	scaffold.2419889_766209	SLCO1C1	Solute carrier organic anion transporter family member 1C1
GTA0201052	scaffold.2419821_342438	CTGF	Connective tissue growth factor
GTA0201053	scaffold.2419996_643476	FABP1	Fatty acid-binding protein, liver
GTA0201054	scaffold.2419904_206589	Muc2	Mucin-2
GTA0201056	scaffold.2419464_308854	AGRP	Agouti-related protein

Table A3 continued

GTA0201057	scaffold.2419794_788999	MMP2	Matrix metallopeptidase 2
GTA0201059	scaffold.2420270_504235	LHCGR	Lutropin-choriogonadotropic hormone receptor
GTA0201060	scaffold.2419234_102446	PDIA3	Protein disulfide-isomerase A3
GTA0201064	scaffold.2420711_480278	COCH	Cochlin
GTA0201065	scaffold.2419273_218258	TNC	Tenascin
GTA0201066	scaffold.2419824_65515	STAT3	Signal transducer and activator of transcription 3
GTA0201068	scaffold.2419563_279544	CALB1	Calbindin
GTA0201069	scaffold.2419287_935738	VIM	Vimentin
GTA0201076	scaffold.2420484_534110	HSPD1	60 kDa heat shock protein, mitochondrial
GTA0201078	scaffold.2420267_724729	Cldn1	Claudin-1
GTA0201079	scaffold.2420920_870248	VEGFA	Vascular endothelial growth factor A
GTA0201082	scaffold.2419542_5611	ABCG2	ATP-binding cassette sub-family G member 2
GTA0201088	scaffold.2419025_1502396	calm2b	Calmodulin-2 B
GTA0201090	scaffold.2419529_29868	Cd74	H-2 class II histocompatibility antigen gamma chain
GTA0201091	scaffold.2420222_2234036	PCK1	Phosphoenolpyruvate carboxykinase, cytosolic [GTP]
GTA0201096	scaffold.2419810_91555	EXFABP	Extracellular fatty acid-binding protein
GTA0201098	scaffold.2419041_41475	MECOM	MDS1 and EVI1 complex locus protein EVI1
GTA0201099	scaffold.2420501_307779	OPN5	Opsin-5
GTA0201100	scaffold.2372192_495	RPE65	Retinoid isomerohydrolase
GTA0201102	scaffold.2420615_2539713	ESR1	Estrogen receptor
GTA0201104	scaffold.2419418_267683	PARP1	Poly [ADP-ribose] polymerase 1
GTA0201105	scaffold.2420379_209806	SHH	Sonic hedgehog protein
GTA0201107	scaffold.2420084_1087802	GDF9	Growth/differentiation factor 9
GTA0201108	scaffold.2420963_235259	ACACB	Acetyl-CoA carboxylase 2
GTA0201109	scaffold.2420062_1441996	ASTL	Astacin-like metalloendopeptidase
GTA0201110	scaffold.2419231_1025841	BRCA2	Breast cancer type 2 susceptibility protein
GTA0201112	scaffold.2419902_1843835	ROBO1	Roundabout homolog 1
GTA0201113	scaffold.2420361_317748	CLOCK	Circadian locomoter output cycles protein kaput
GTA0201114	scaffold.2419407_242816	FTO	Alpha-ketoglutarate-dependent dioxygenase
GTA0201115	scaffold.2418980_311832	NTS	Neurotensin/neuromedin N
GTA0201118	scaffold.2420740_525483	PGR	Progesterone receptor
GTA0201120	scaffold.2420754_475821	KCNH6	Potassium voltage-gated channel subfamily H member 6
GTA0201122	scaffold.2419198_948773	NAMPT	Nicotinamide phosphoribosyltransferase
GTA0201124	scaffold.2419962_205525	PRKAA1	5'-AMP-activated protein kinase catalytic subunit alpha-1
GTA0201125	scaffold.2420081_518670	HMOX1	Heme oxygenase 1
GTA0201127	scaffold.2420999_36136	CASP1	Caspase-1
GTA0201129	scaffold.2420326_1103027	BCL2	Apoptosis regulator Bcl-2

Table A3 continued

GTA0201130	scaffold.2420958_614552	FGF2	Fibroblast growth factor 2
GTA0201135	scaffold.2419292_300168	Edn1	Endothelin-1
GTA0201137	scaffold.2420264_2300657	NGF	Beta-nerve growth factor
GTA0201138	scaffold.2420928_384544	PPARA	Peroxisome proliferator-activated receptor alpha
GTA0201139	scaffold.2419560_280705	INS	Insulin
GTA0201140	scaffold.2419197_25948	IFNL3	Interferon lambda-3
GTA0201142	scaffold.2419746_1961703	TYR	Tyrosinase
GTA0201144	scaffold.2419188_2133072	COLEC11	Collectin-11
GTA0201147	scaffold.2420462_314560	CDH13	Cadherin-13
GTA0201148	scaffold.2419129_85609	FLRT3	Leucine-rich repeat transmembrane protein
GTA0201149	scaffold.2419461_158392	ACAN	Aggrecan core protein
GTA0201152	scaffold.2420832_658918	LGALS2	Galectin-2
GTA0201154	scaffold.2419526_26842	HIF1A	Hypoxia-inducible factor 1-alpha
GTA0201156	scaffold.2419692_601099	AR	Androgen receptor
GTA0201158	scaffold.2419354_183098	Trpm8	Transient receptor potential cation channel subfamily M member 8
GTA0201159	scaffold.2419429_656029	TNN	Tenascin-N
GTA0201160	scaffold.2420770_1209742	SEMA3A	Semaphorin-3A
GTA0201161	scaffold.2418924_70694	BKJ	Beta-keratin-related protein
GTA0201163	scaffold.2419825_546536	CDH1	Cadherin-1
GTA0201165	scaffold.2420818_2912542	ABCB1	ATP binding cassette subfamily B member 1
GTA0201166	scaffold.2420283_7714	CPOX	Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial
GTA0201197	scaffold.2419443_219448	PTH	Parathyroid hormone
GTA0201201	scaffold.2419408_480576	PAX7	Paired box protein Pax-7
GTA0201202	scaffold.2419295_783492	GNRH1	Progonadoliberin-1

APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table B1 Whole genome sequence read statistics for each individual (N=74) trimming adapters and low quality sequences Arizona samples are highlighted in light blue, Texas samples are highlighted in pink and New Mexico samples are highlighted in green.

Sample ID	Total Reads	Total Bases	Min Len	Mean Len	Max Len	% seq lost	% bases lost	Mapping %	Breadth of coverage (%)	Depth of coverage (X)
E6537	33,600,122	4,274,242,469	30	127	151	0	16.43590294	96.429	91.8536	3.99176
E6538	16,482,094	2,057,428,202	30	124	151	0	18.07030504	95.806	72.4103	1.91611
E6539	14,528,498	1,786,460,364	30	122	251	1	19.48874053	89.643	47.0595	1.43745
E6541	20,129,040	2,679,752,197	30	133	251	1	12.82387168	46.580	37.6452	0.81683
E6543	25,414,610	3,214,919,818	30	126	251	1	17.08177431	81.209	71.3493	2.20447
E6545	23,589,836	2,717,050,993	30	114	251	1	24.77170793	89.768	63.3578	1.99691
E6548	18,907,440	2,337,892,781	30	123	151	1	19.07479279	50.670	27.0938	0.682041
E6566	19,174,870	2,196,258,563	30	114	251	1	25.21643812	95.967	63.7276	1.9232
E6597	27,394,358	3,403,617,829	30	124	151	0	18.48732046	96.313	84.721	3.12776
E6598	26,252,348	3,275,918,821	30	124	151	0	18.07011672	96.577	85.7205	3.0762
E6599	28,926,254	3,868,730,275	30	133	151	0	12.08514672	95.148	91.7807	3.57133
E6600	23,699,300	2,841,877,036	30	119	151	1	21.44906733	95.889	80.7189	2.63539
E6609	24,873,758	3,273,051,783	30	131	151	0	13.47750122	94.182	85.6522	2.91976

Table B1 continued

E6705	21,014,564	2,567,496,803	30	122	251	1	20.06150528	89.300	65.6898	1.83042
E6715	19,342,052	2,329,540,171	30	120	251	1	21.20189571	95.523	64.5881	1.86892
E6716	15,439,024	1,881,125,334	30	121	151	1	20.14143974	94.915	66.2256	1.69422
E6718	18,686,522	2,228,796,760	30	119	251	1	22.07742769	86.688	61.0875	1.63069
E6758	17,096,568	2,224,836,949	30	130	151	0	14.48065886	97.251	76.6621	2.13067
E6759	28,932,710	3,653,246,481	30	126	251	0	17.00499497	97.061	87.3612	3.37574
E6762	18,225,844	2,010,931,735	30	110	251	1	27.84035543	97.147	61.7902	1.84956
E6799	22,244,404	2,823,617,543	30	126	151	0	16.67511558	93.894	80.2249	2.5287
E6800	18,236,814	2,259,191,138	30	123	251	1	18.85298064	69.118	52.3194	1.26126
E6801	25,572,556	3,331,496,821	30	130	251	0	14.40190715	87.986	83.0296	2.69863
E6803	33,241,370	4,407,669,511	30	132	251	0	12.87317077	90.539	89.4336	3.56039
E6841	18,517,834	2,246,285,149	30	121	251	1	20.55069929	88.430	64.5698	1.77449
E6843	19,370,072	2,198,511,271	30	113	251	1	25.868553	95.307	56.4973	1.73955
E6847	18,482,182	2,202,056,693	30	118	151	1	21.92476168	95.860	69.8905	1.99843
E6963	33,110,208	4,483,974,539	30	135	251	0	10.92455982	94.837	92.3482	3.91862
E6964	23,792,780	3,065,819,963	30	128	151	0	15.39894896	96.457	84.1642	2.81308
E6966	17,474,332	2,206,802,478	30	126	151	0	17.11722216	93.732	73.9285	1.98149
E7010	15,537,242	1,956,164,954	30	125	251	1	17.48555854	81.536	52.2074	1.34422

Table B1 continued

E7011	24,013,338	3,194,700,379	30	132	251	0	12.66704089	75.421	72.0282	1.95008
E7012	15,047,108	1,916,571,683	30	127	251	0	16.45366863	82.791	56.2622	1.38327
E7035	21,848,376	2,801,671,533	30	128	151	0	15.77403418	94.806	78.2445	2.46568
E7048	17,471,158	2,102,506,216	30	120	251	1	21.27985227	87.481	51.9475	1.43534
E7049	17,993,798	2,657,534,292	30	147	251	0	2.465671319	86.799	80.4312	2.25097
E7159	15,107,394	1,884,886,783	30	124	251	0	18.13328492	96.354	67.5206	1.70692
E7161	17,010,834	2,057,255,210	30	120	251	1	20.75178688	96.584	67.2025	1.85173
E7175	18,742,570	2,197,303,953	30	117	251	1	23.49579741	92.125	49.7387	1.47077
E7177	30,251,566	3,944,254,228	30	130	251	0	14.31669848	96.274	88.0516	3.53435
E7219	17,194,322	2,264,874,267	30	131	151	0	13.36150975	67.415	54.9096	1.27549
E7755	16,995,676	2,147,627,109	30	126	151	0	17.12450629	95.202	73.6911	1.97215
E7757	25,794,068	3,308,977,700	30	128	151	0	15.68578099	96.425	87.5735	3.13268
E7879	21,256,178	2,636,978,688	30	123	251	1	18.70614593	82.189	57.7679	1.70622
E7910	17,971,200	2,171,787,423	30	120	251	1	20.93717563	95.763	62.1332	1.7536
E7915	17,578,272	2,240,711,607	30	127	251	1	16.58785958	77.516	57.8872	1.37476
E8002	20,709,886	2,711,406,782	30	130	151	0	13.98346026	94.272	82.1617	2.48988
E8036	17,775,628	2,289,208,833	30	128	151	0	15.35558051	96.213	76.4024	2.1513
E8063	23,354,646	2,885,458,056	30	123	251	1	19.03492102	79.793	69.0365	2.01307

Table B1 continued

E8064	18,753,610	2,380,572,322	30	126	251	0	16.74568544	84.285	66.2246	1.78377
E8065	22,153,744	2,828,033,476	30	127	251	0	16.21585754	83.147	69.8572	2.06117
E8066	15,503,686	1,875,468,084	30	120	251	1	20.94536433	94.478	57.35	1.52586
E8429	32,323,714	4,342,013,603	30	134	151	0	11.46613643	44.596	21.3903	0.617508
E8430	19,245,050	2,505,088,373	30	130	251	0	14.51041762	76.161	62.0729	1.59527
E8431	15,180,192	2,039,074,782	30	134	251	0	11.65678354	67.267	50.8396	1.13806
E8434	18,990,972	2,583,733,637	30	135	151	0	10.43850378	39.695	10.6877	0.29867
E8436	14,149,156	2,114,118,056	30	149	251	0	1.121292552	18.401	6.20641	0.110517
E8438	21,916,948	2,842,528,190	30	129	151	0	14.80478683	92.805	82.7523	2.5479
E8440	17,872,270	2,282,899,619	30	127	251	1	16.58721341	48.525	21.1236	0.519821
E8441	14,925,446	2,040,757,229	30	136	251	0	9.918530559	39.504	5.12778	0.170257
E8442	22,680,876	2,945,703,537	30	129	251	0	14.76058099	52.957	31.5934	0.803266
E8444	17,829,658	2,366,276,178	30	132	251	0	12.75230469	85.652	69.0892	1.78475
E8445	15,107,500	1,853,800,182	30	122	251	1	19.66198845	85.174	51.9213	1.31947
E8447	15,574,392	2,007,149,772	30	128	251	0	15.31504714	89.384	62.7057	1.57462
E8448	19,254,870	2,529,861,729	30	131	251	0	13.67618165	38.060	8.75886	0.263198
E8450	14,767,776	1,898,582,498	30	128	251	0	15.64508845	63.352	36.7484	0.81188
E8740	59,409,300	7,588,382,444	30	127	151	0	15.99661505	97.302	82.1234	7.2751

Table B1 continued

E8451	29,702,054	4,115,007,059	30	138	251	0	8.657388114	97.563	98.0159	3.91939
E8452	16,023,926	2,077,335,982	30	129	151	0	14.6884511	97.381	93.023	1.98967
E8453	49,140,128	6,601,298,792	30	134	251	0	11.53922044	97.796	74.7906	6.20812
E8454	17,572,070	2,311,277,780	30	131	151	0	13.38441737	97.365	97.3018	2.2415
E8455	19,539,982	2,562,345,697	30	131	151	0	13.65431739	96.875	79.94	2.44945
E8741	46,867,732	5,925,996,446	30	126	151	0	16.94725225	97.583	95.3667	5.55919
E8748	18,800,044	2,227,006,429	30	118	151	1	22.55251279	96.226	66.5808	2.00603

Table B2 Test for significant differences in mean individual inbreeding coefficients (F) using Wilcoxon rank sum test with continuity correction

Population Comparison	Test Statistic (W)	p-value
AZ- TX	149	<0.0003
TX-NM	82	0.039
AZ-NM	149	0.446

Table B3 Test for significant differences in mean whole-genomic nucleotide diversity estimates (θ_w) using Wilcoxon rank sum test with continuity correction

Population Comparison	Test Statistic (W)	p-value
AZ- TX	185570578	< 2.2e-16
TX-NM	99585686	< 2.2e-16
AZ-NM	174571874	< 2.2e-16

Table B4 Test for significant differences in mean individual genome-wide heterozygosity using Wilcoxon rank sum test with continuity correction

Population Comparison	Test Statistic (W)	p-value
AZ- TX	42	0.02622
TX-NM	8	0.03788
AZ-NM	33	0.3176

Table B5 Genes annotated within highly differentiated outlier regions with $Z(F_{ST}) > 5SD$ and their chromosomal location in the chicken genome

Gene title	**Gene name**	**Chr**	**Position (bp)**	**Functional description**
ANKRD26	Ankyrin repeat domain 26	1	24435899 - 24484038	Interacts with proteins or protein complexes
NELL1	Neural EGFL like 1	5	2461252 - 2747574	Involved in cell growth regulation and differentiation
CHUK	Component of inhibitor of nuclear factor kappa B kinase complex	6	9980703 -10003570	Has a role in NF-kappa-B signaling pathway which is activated in response to cellular stresses
CPS1	Carbamoyl-phosphate synthase 1	7	2842238 - 2944989	Involved in amino acid and nitrogen metabolism
KLHL25	Kelch like family member 25	10	14396164 - 14410877	Related role in innate immune system for antigen processing and presentation
AKAP13	A-Kinase Anchoring Protein 13	10	14421126 -14487012	Functions as scaffolding proteins to coordinate a Rho signaling pathway
HYDIN	Axonemal central pair apparatus protein	11	1545049 - 1649225	Required for ciliary motility and neural cell development
VAC14	Component of PIKFYVE Complex	11	1649800 - 1700757	Encodes a scaffold protein for components of cellular membranes
NCOR2	Nuclear receptor corepressor 2	15	4857712 - 5022472	Mediates transcriptional silencing of certain target genes

Table B5 continued

ARID1A	AT-rich interaction domain 1A	23	142792 - 202595	Involved in transcriptional activation and repression of select genes
VPS45	Vacuolar protein sorting 45 homolog	25	83517 - 105457	Involved in vesicle transport to vacuoles suggesting a role in protein trafficking
ASIC2	Acid sensing ion channel subunit 2	27	4186435 - 4599845	Encodes sodium channels that play a role in neurotransmission
ACLY	ATP citrate lyase	27	7520780 - 7549243	Primary enzyme responsible for fatty acid synthesis and carbohydrate metabolism
TTC25	Tetratricopeptide repeat domain 25	27	7545735 - 7553230	Localizes to ciliary axonmenes and required for cell signaling and cell motility
CNP	2',3'-cyclic nucleotide 3' phosphodiesterase	27	7553422 - 7558601	Expressed exclusively by oligodendrocytes in the Central Nervous System
DNAJC7	DnaJ heat shock protein family (Hsp40) member C7	27	7558532 - 7577845	Involved in cellular response to heat stress

Table B6 Test for significant differences in mean genic nucleotide diversity estimates (θ_w) using Welch Two Sample t-test

Population Comparison	Test Statistic (t)	Degree of Freedom (df)	p-value
AZ- TX	293.8	605670000	< 2.2e-16
TX-NM	-124.58	614770000	< 2.2e-16
AZ-NM	168.5	615040000	< 2.2e-16

Table B7 Test for significant differences in mean individual genic heterozygosity using Wilcoxon rank sum test with continuity correction

Population Comparison	Test Statistic (W)	p-value
AZ- TX	42	0.02622
TX-NM	9	0.05303
AZ-NM	32	0.3829

Table B8 Test for significant differences in mean proportion of different impact deleterious mutations (potential load) and non-coding variants using chi-square test with Yates' continuity correction

Variant Impact	Population Comparison	Test Statistic (χ^2)	Degree of Freedom (df)	p-value
High Impact (Most deleterious)	AZ- TX	34.616	1	4.016e-09
	TX-NM	9.6379e-28	1	1
	AZ-NM	42.753	1	6.209e-11
Moderate Impact (Mildly deleterious)	AZ- TX	95.165	1	< 2.2e-16
	TX-NM	21.274	1	3.981e-06
	AZ-NM	29.46	1	5.709e-08
Low Impact (Least deleterious)	AZ- TX	49.539	1	1.944e-12
	TX-NM	0.66378	1	0.4152
	AZ-NM	46.714	1	8.215e-12
Non-coding (Non-deleterious)	AZ- TX	214.56	1	2.2e-16
	TX-NM	18.76	1	1.483e-05
	AZ-NM	124.38	1	2.2e-16

Table B9 Test for significant differences in mean deleterious allele frequency per individual (realized load) of different impact deleterious mutations and non-coding variants using Welch Two Sample t-test

Variant Impact	Population Comparison	Test Statistic (t)	Degree of Freedom (df)	p-value
High Impact (Most deleterious)	AZ- TX	-2.3947	6.6428	0.06971
	TX-NM	0.96428	11.982	0.354
	AZ-NM	-1.0027	6.5956	0.3513
Moderate Impact (Mildly deleterious)	AZ- TX	-2.5349	6.7177	0.04031
	TX-NM	0.9555	11.979	0.3582
	AZ-NM	-1.1473	6.6602	0.2908
Low Impact (Least deleterious)	AZ- TX	-2.7997	6.8769	0.02702
	TX-NM	0.95632	11.943	0.3579
	AZ-NM	-1.3562	6.7647	0.2186
Non-coding (Non-deleterious)	AZ- TX	-2.5581	6.7285	0.03894
	TX-NM	0.95632	11.943	0.3579
	AZ-NM	-1.3562	6.7647	0.2186

Table B10 Test for significant differences in mean observed heterozygosity per individual of different impact deleterious mutations and non-coding variants using Welch Two Sample t-test

Variant Impact	Population Comparison	Test Statistic (t)	Degree of Freedom (df)	p-value
High Impact (Most deleterious)	AZ- TX	1.3477	10.321	0.2066
	TX-NM	-1.0673	8.7382	0.3145
	AZ-NM	0.01867	11.153	0.9854
Moderate Impact (Mildly deleterious)	AZ- TX	3.2757	11.728	0.006823
	TX-NM	-1.3719	7.5206	0.2096
	AZ-NM	0.087255	7.1271	0.9329
Low Impact (Least deleterious)	AZ- TX	5.0392	11.134	0.0003646
	TX-NM	-1.5721	6.9705	0.1601
	AZ-NM	0.15961	6.5498	0.878
Non-coding (Non-deleterious)	AZ- TX	3.6726	9.3357	0.004825
	TX-NM	-1.518	8.2024	0.1666
	AZ-NM	0.16772	6.6905	0.8718

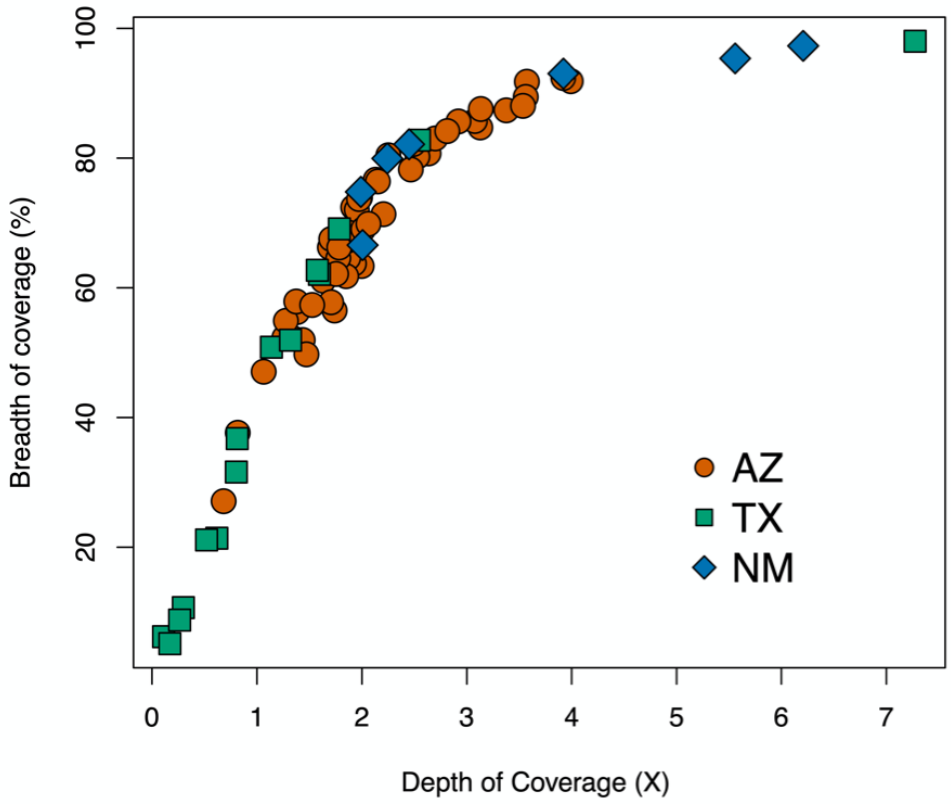


Figure B1 Individual depth of coverage (X) and breadth of coverage (%). Depth of coverage is calculated as the mean base coverage covered by the reads and breadth is calculated as the percentage of draft Montezuma quail covered by the reads.

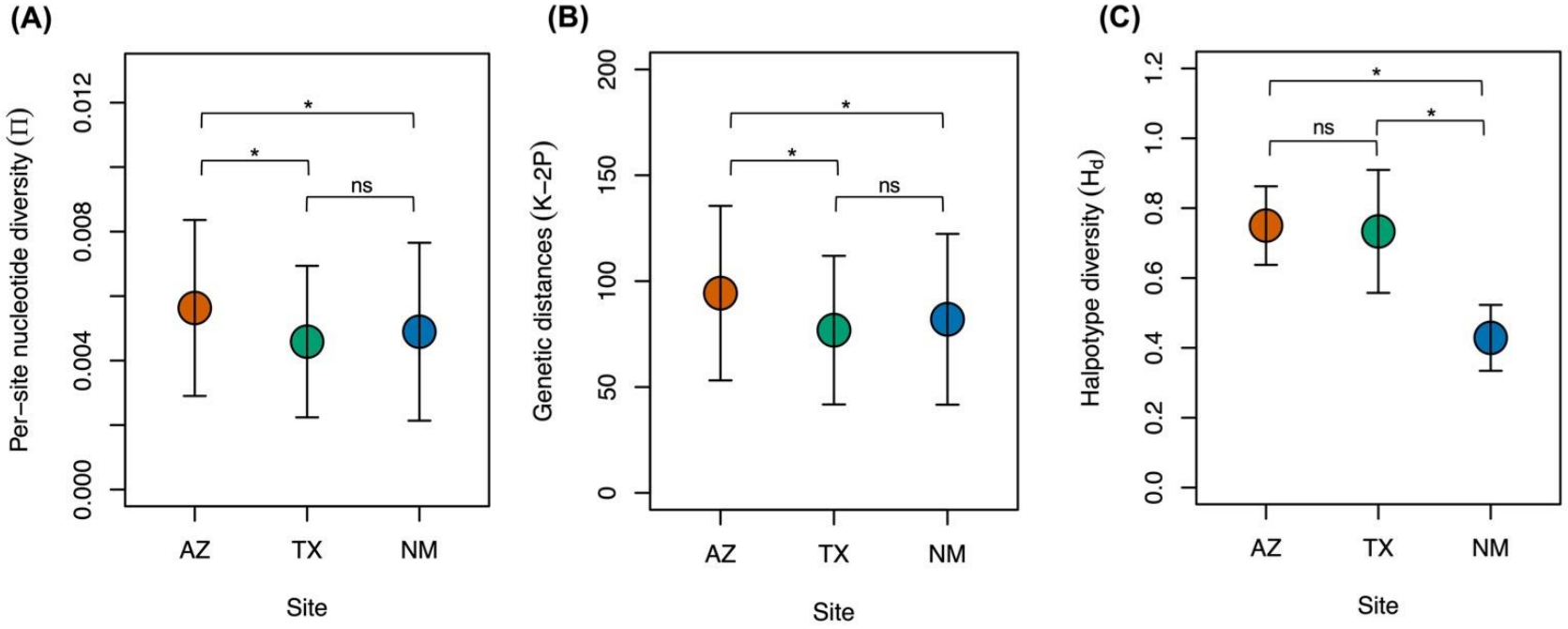


Figure B2 Mitogenome diversity statistics for each population. (A) Per-site nucleotide (Π), (B) Within population genetic distances calculated using Kimura-2P parameter (K-2P), and (C) Haplotype diversity in Montezuma quail populations in Arizona, Texas, New Mexico. Π and Kimura 2-P pairwise distances were smaller in the Texas and New Mexico mitogenomes (p=0.034 and p=0.041 respectively) as compared to Arizona but no significant difference in haplotype diversity between Texas and Arizona mitogenomes (H_d ; p = 0.7) but significantly smaller in New Mexico (p=0.02).

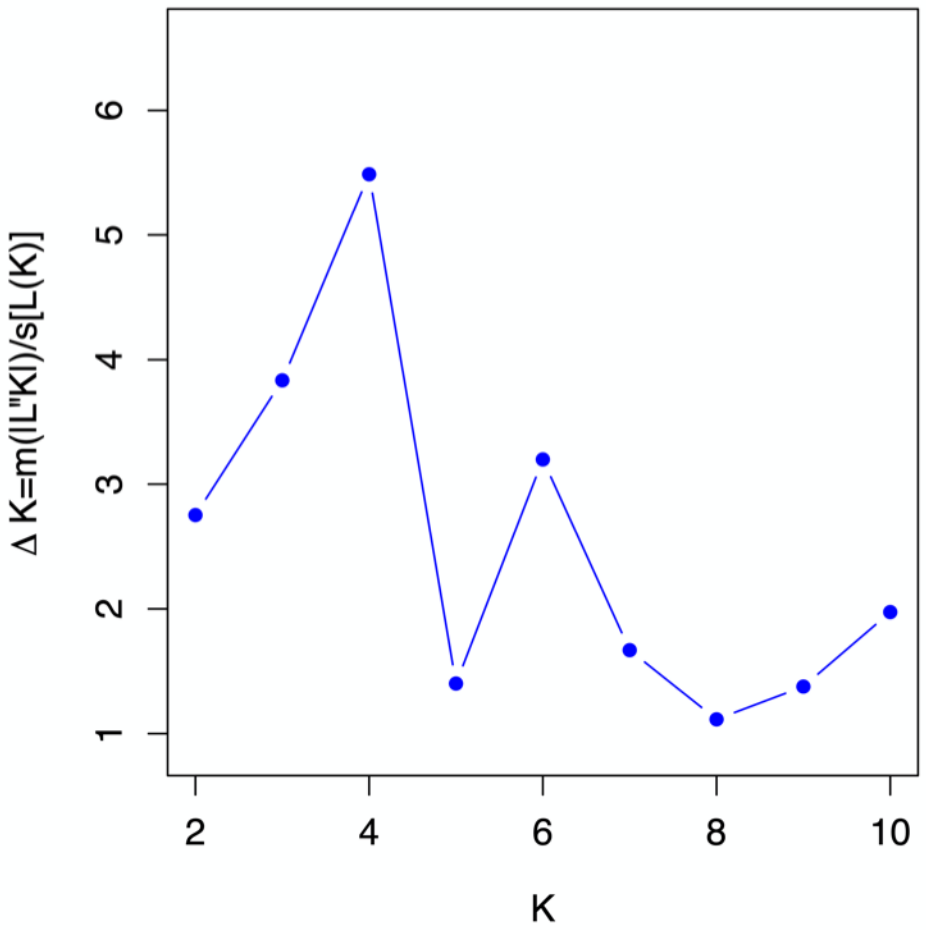


Figure B3 Delta K values (for each K = 1-10) indicate the best estimated number of ancestral populations is K = 4 based on Evanno et al. (2005). At K=4, the Arizona, Texas, and New Mexico populations form independent clusters with Arizona populations being split into two subpopulations.

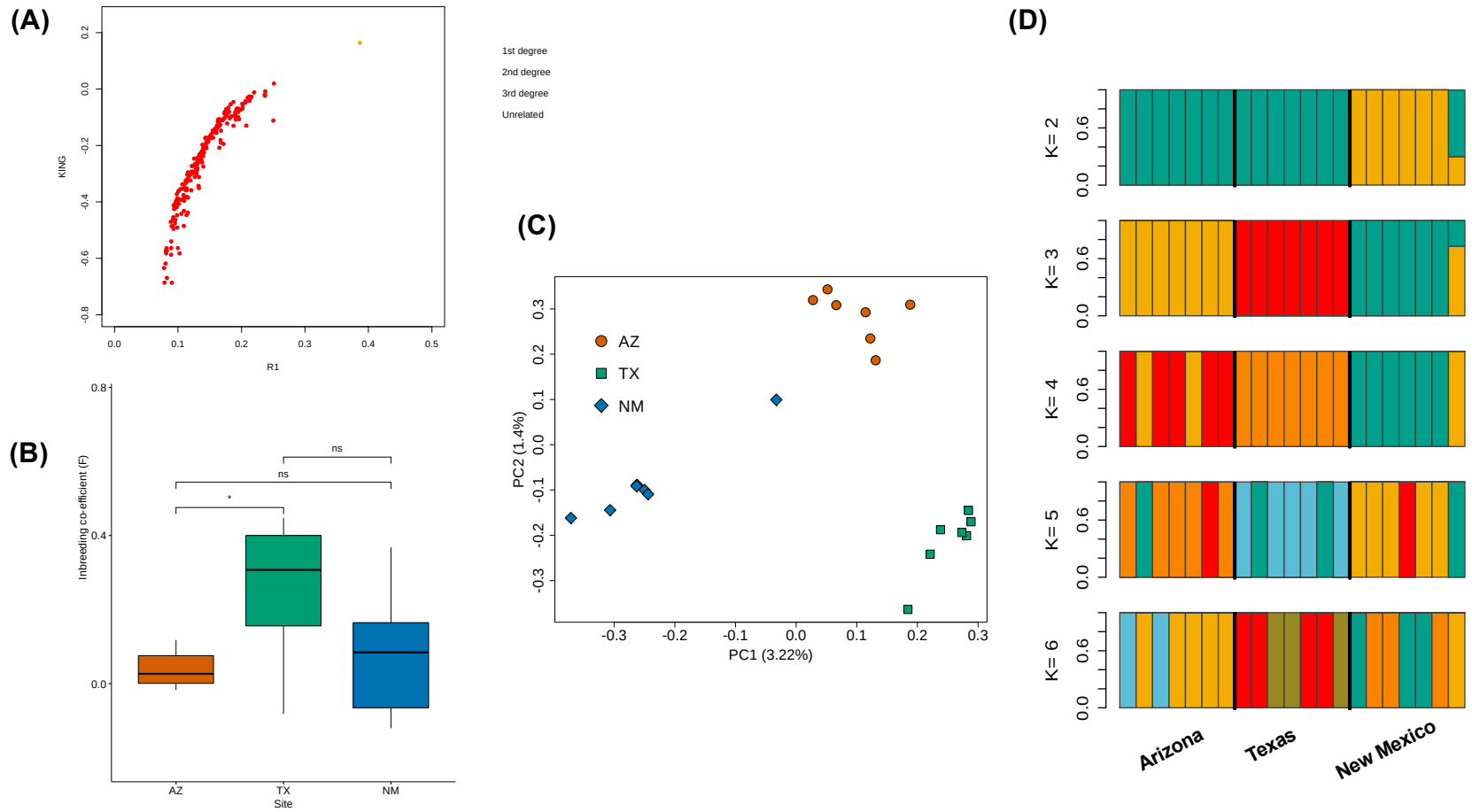


Figure B4 (A) Pairwise Relatedness comparisons, (B) Inbreeding coefficient estimation, (C) PCA, and (D) Admixture, estimates *using equal subsamples* (N=21; AZ=7, TX=7, NM=7). These results are similar to what was estimated using population dataset (N=74; AZ=52, TX=15, NM=7) indicating that the results shown in the main text are biologically relevant and not due to our sampling scheme.

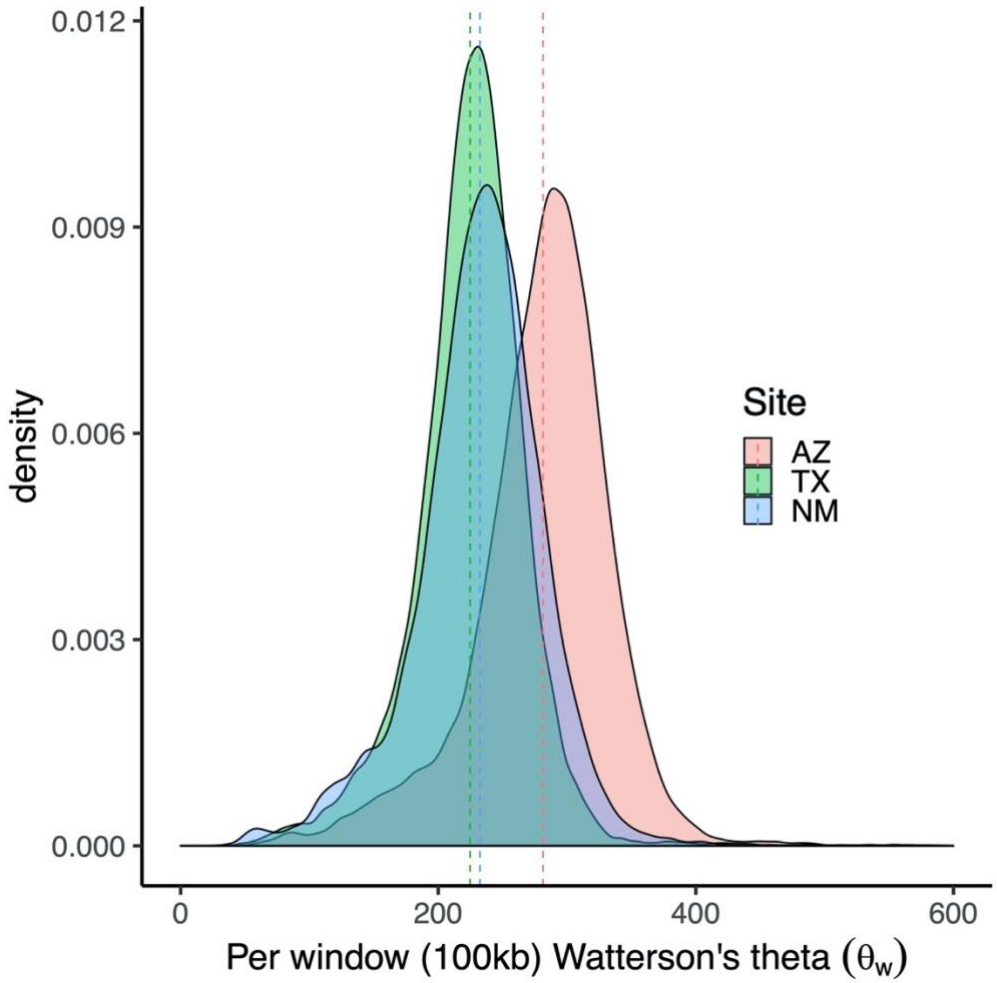


Figure B5 Distribution of Watterson's theta (θ_W) across the whole genome. θ_W was calculated for every window (100kb size, 50kb step

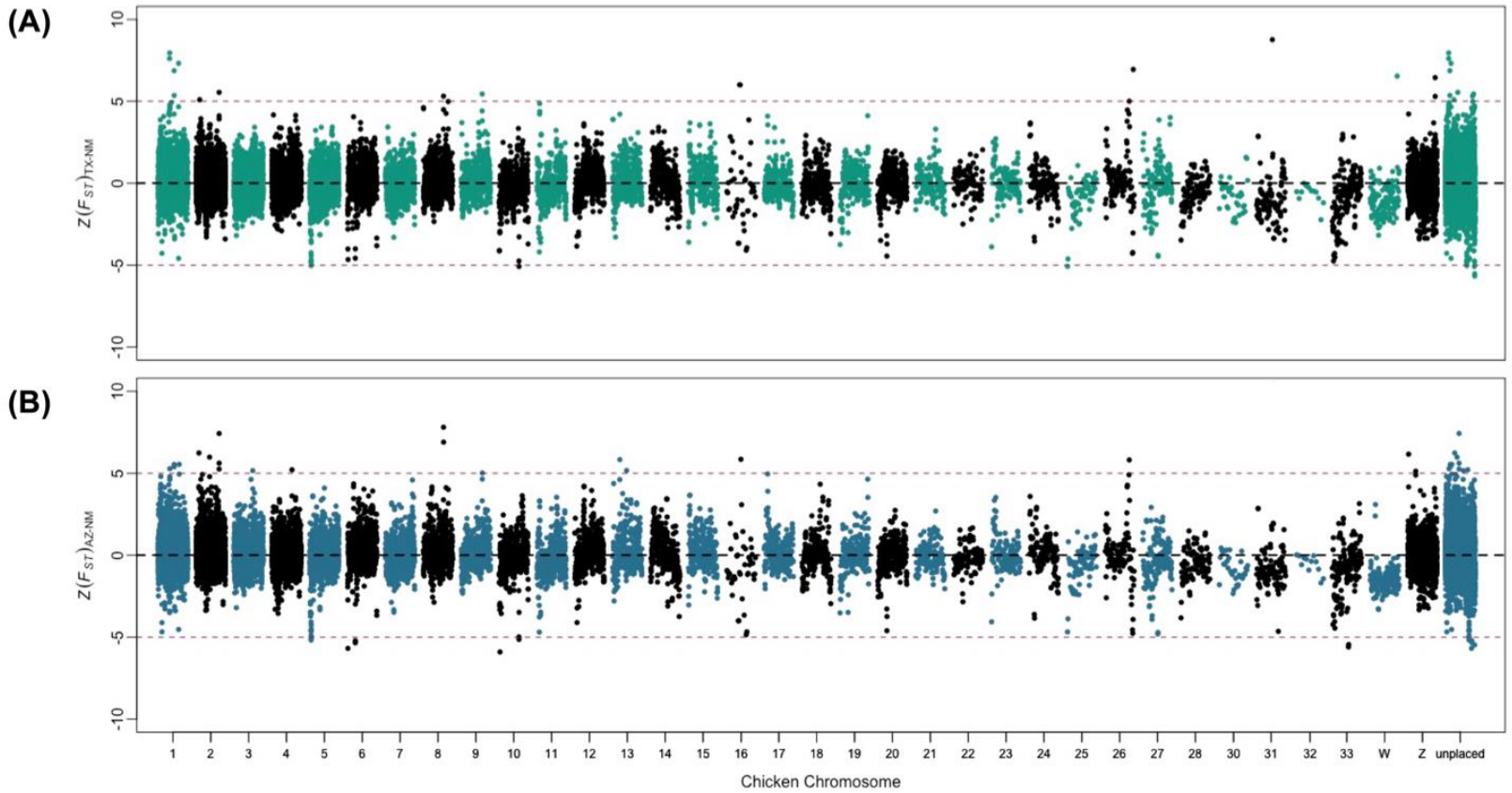


Figure B6 Z-transformed FST estimates for comparisons made between (A) Texas and New Mexico Montezuma quail, and (B) Arizona and New Mexico Montezuma quail populations for every 100 kb window (50 kb steps). The reads were mapped to the chicken genome and the windows were arranged according to chicken autosomal (1-33) or sex (Z, W) chromosomes. Scaffolds that were not part of the major chicken chromosomes were binned together as unplaced. This figure shows a heterogeneous landscape of genetic differentiation and drift is the primary evolutionary driver behind the observed patterns.

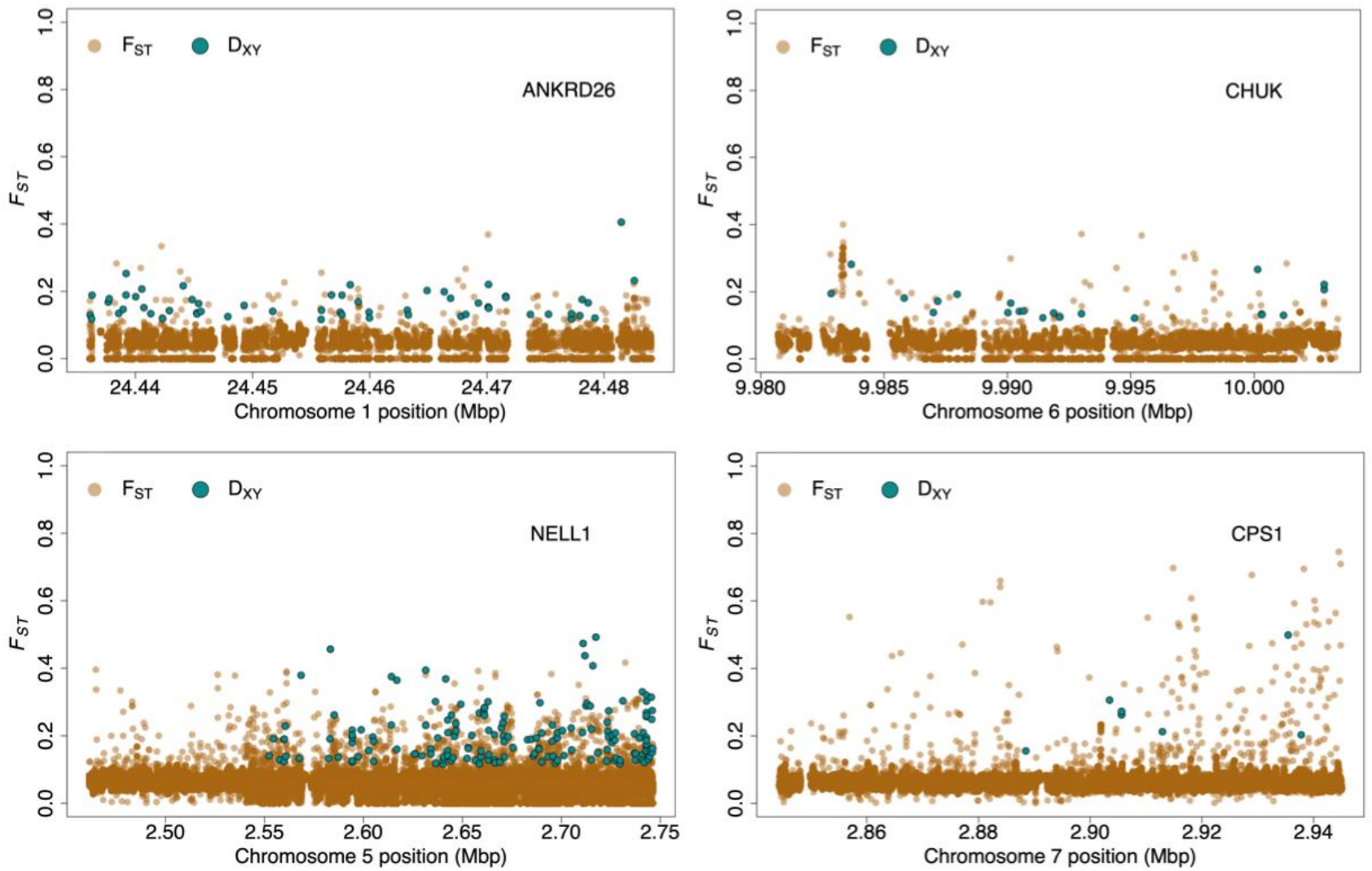


Figure B7 Pairwise FST for each polymorphic site in the genes associated with outlier windows in Arizona and Texas populations and DXY for the SNPs that were segregating in all three populations. The functional description of each gene is listed in Table S5

Figure B7 continued

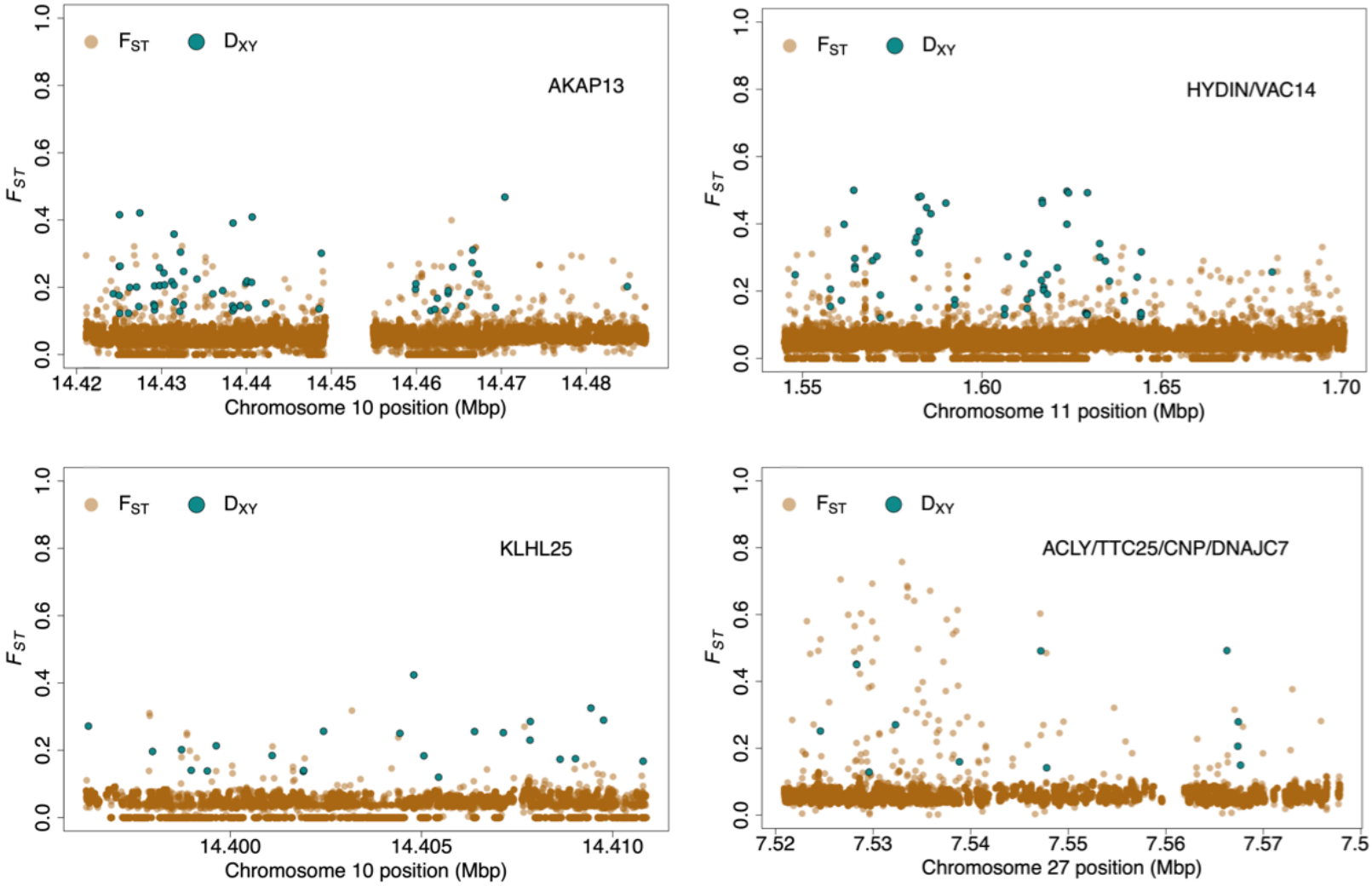
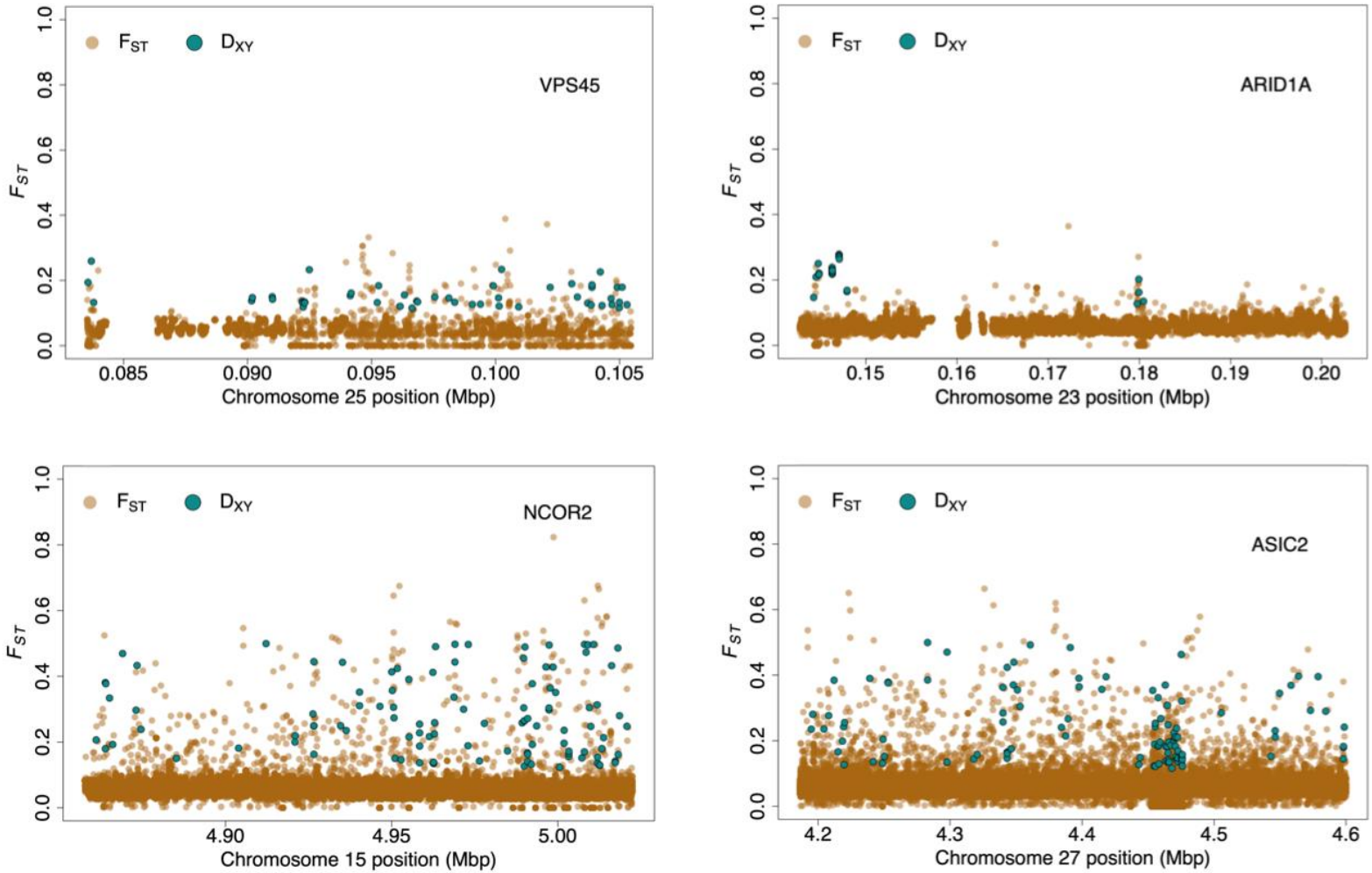


Figure B7 continued



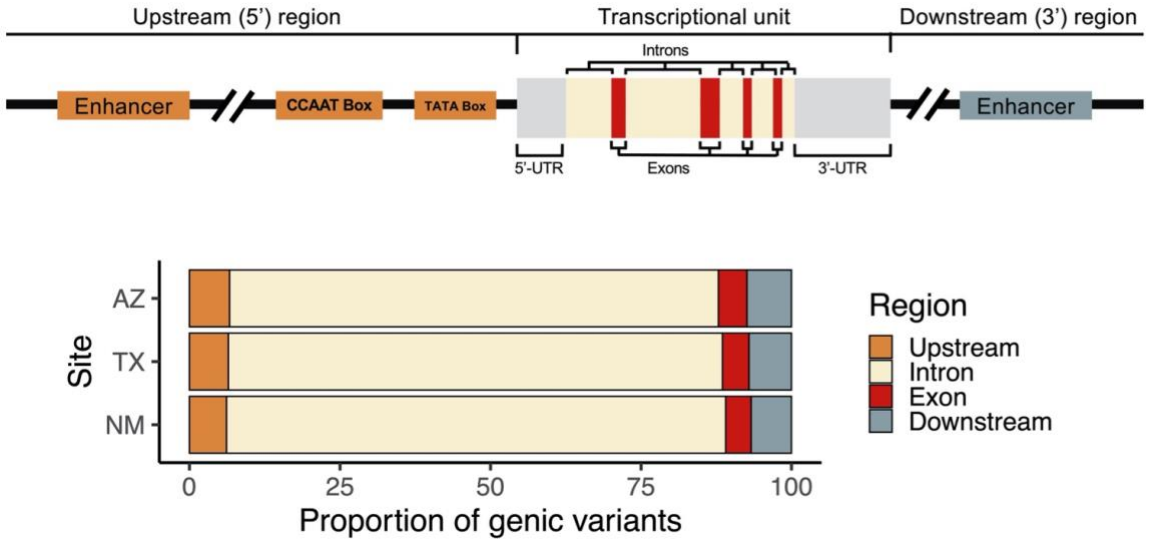


Figure B8 Schematic of eukaryotic gene structure and proportion of variants in different genic regions where the colors in each panel correspond to one another. Much of the genic variation exists outside the transcription unit

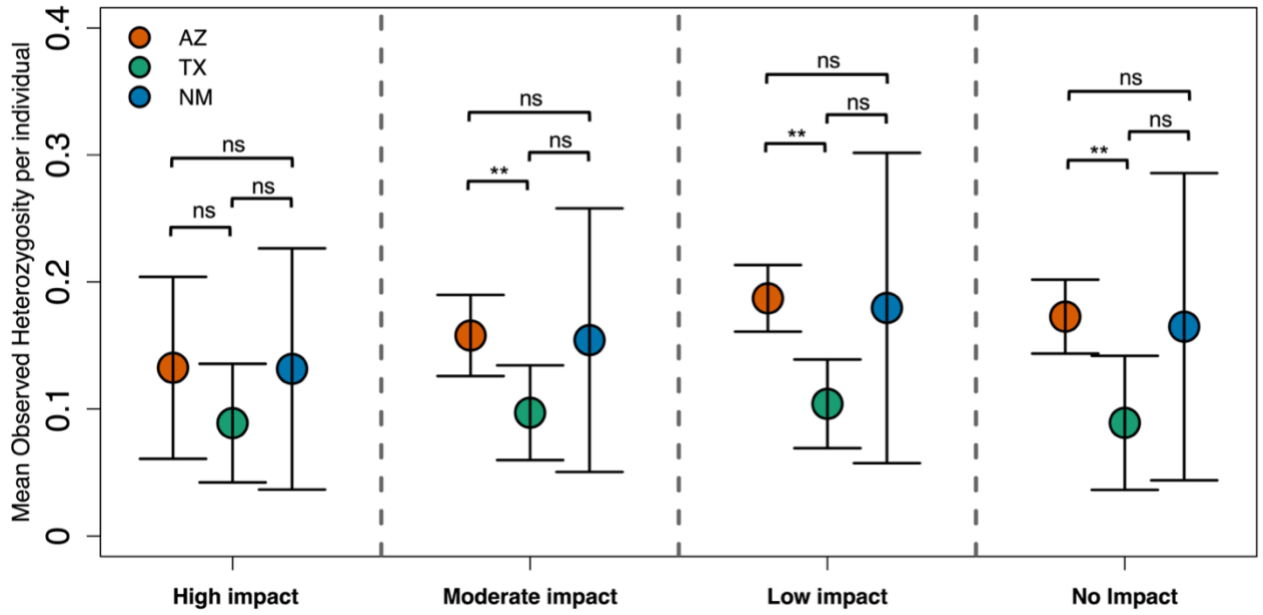


Figure B9 Mean observed heterozygosity per individual for different impact class variants

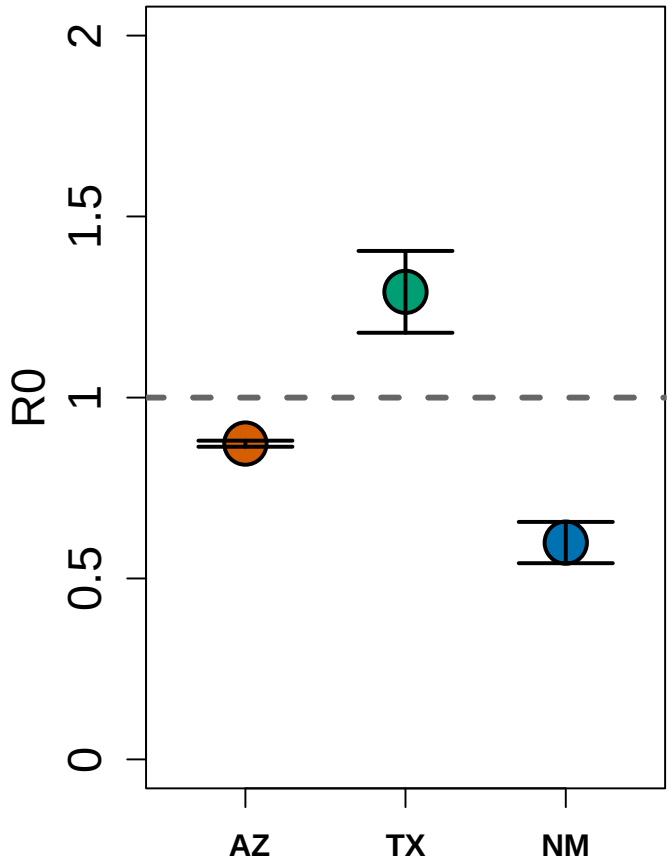


Figure B10 Within population comparisons for R0 statistics from kinship analysis. This statistic is estimated from genome-wide pattern of two individuals sharing identity by state. R0 is the ratio of alternate homozygous genotype likelihoods to shared heterozygotes observed at each polymorphic site between a comparison of two individuals. A mean higher than 1 indicates that the two individuals have more alternate homozygote alleles observed across the genome than shared heterozygote genotypes. Inbreeding between individuals elevates the number of alternate homozygous genotypes and reduces number of shared heterozygous genotypes and was found to be in case of Texas comparisons. Error bars indicate 95% CI around mean.

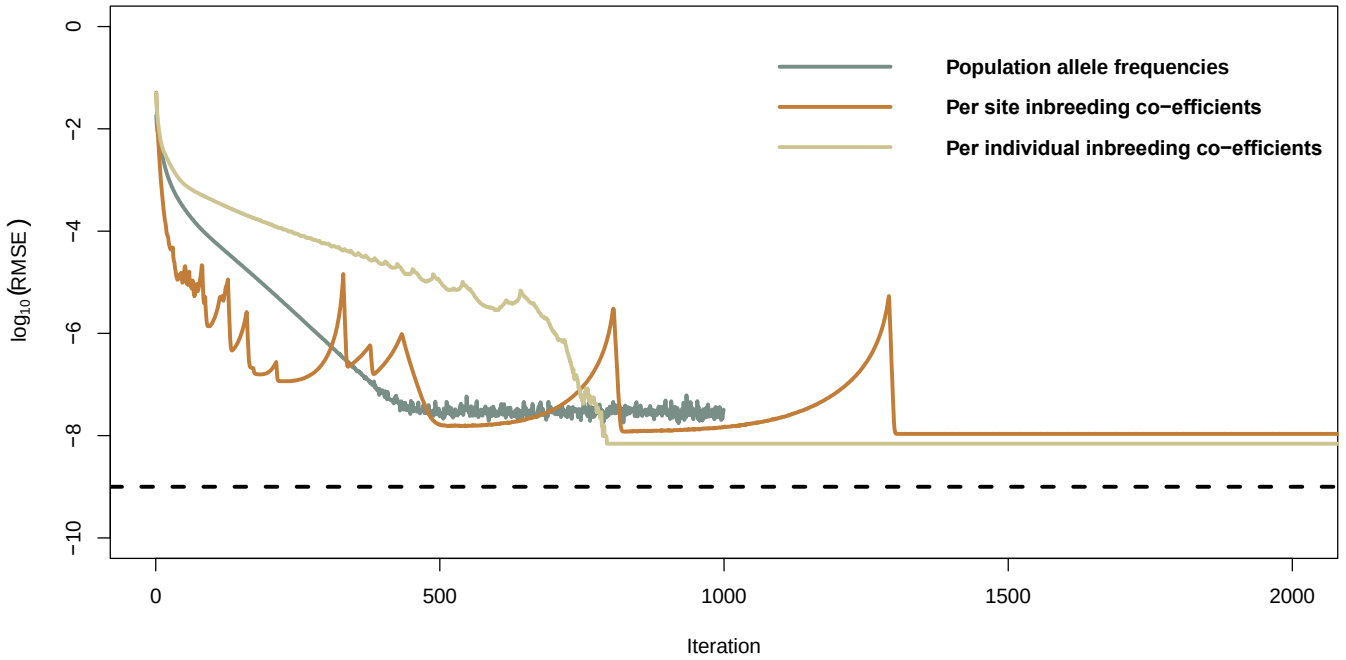


Figure B11 Root mean square error (RMSE) minimization for inbreeding estimation. One issue with iterative algorithms is the stopping criteria. Since we had low coverage WGS data, we used low thresholds (1e-9) and high number of iterations (5000) as stopping criteria of maximization (EM) algorithm for estimating population allele frequencies as well as per-site and per-individual estimates of inbreeding co-efficient. This ensured that the estimation did not stop at the local minima and the estimates were quantified only after the RMSE values were at their global minimum

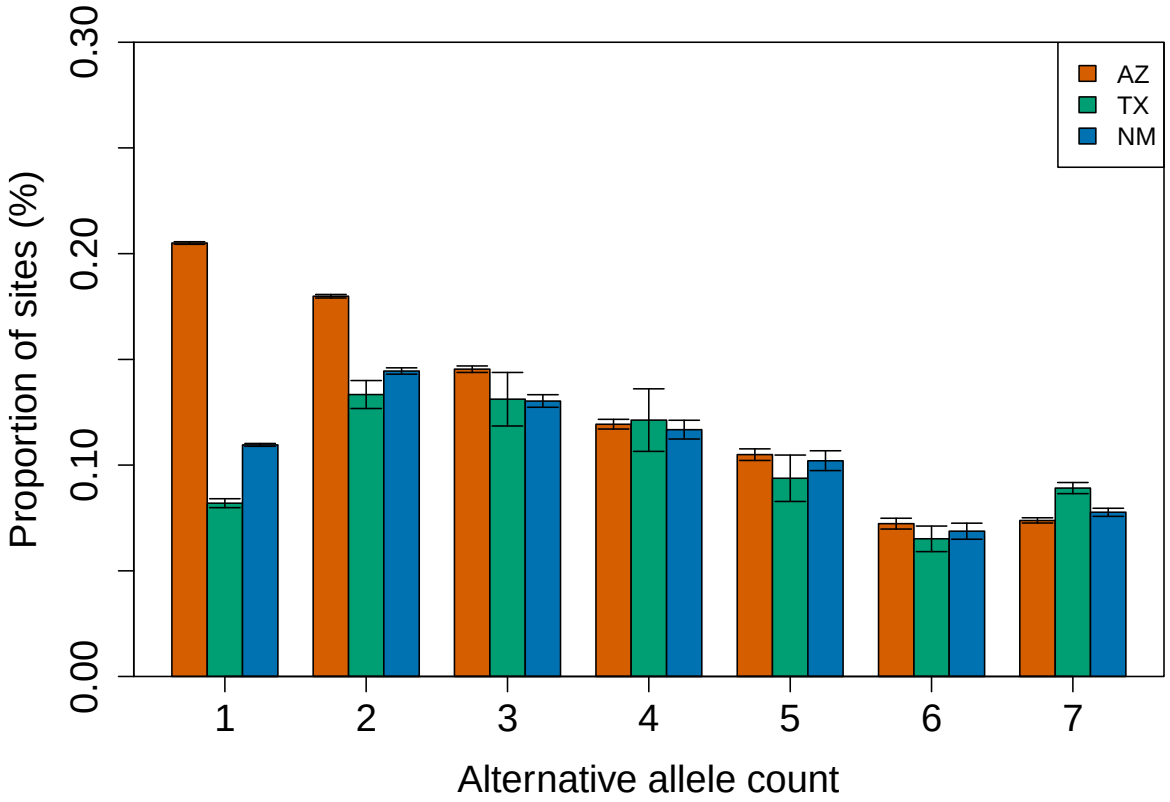


Figure B12 Folded site-frequency spectrum (SFS) of different Montezuma quail populations. We only used the genomic dataset (N=7 for each population) to avoid biases SFS due to uneven sample sizes and heavy data pruning. We used ANGSD to generate a folded SFS by using the Montezuma Quail reference and bootstrapped it 100 times. Barplot shows the mean and error bars represent 1 standard deviation (SD) from the mean. The figure represents the proportion of polymorphic sites within the population based on sampling k-derived alleles. Lower allele count represent rarer mutations vs higher alleles counts represent most common mutations. Arizona (AZ) population has higher proportions of rarer mutations as compared to Texas (TX) or New Mexico (NM) populations, which is expected in larger increasing populations. Whereas, TX population has been declining which removed rarer mutations while maintaining intermediate mutations within the genome.

APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table C1 Whole genome sequence reads, coverage and variant statistics for the best 66 Montezuma Quail samples analyzed

Sample ID	Population	Total Reads	Total bases (Bp)	Depth of Coverage	Total SNPs	SNPs identified	SNP Depth	Missing SNPs	% SNPs missing
E6536	Arizona	185178890	27776833500	22.58279146	12943838	12845978	16.6863	97860	0.75603542
E6609	Arizona	137110856	20566628400	16.7208361	12943838	12628497	12.7162	315341	2.43622487
E6628	Arizona	121802426	18270363900	14.85395439	12943838	12323312	11.0769	620526	4.79398769
E6846	Arizona	98301236	14745185400	11.98795561	12943838	10482670	7.57255	2461168	19.0142058
E6877	Arizona	114686320	17202948000	13.98613659	12943838	12309848	10.7935	633990	4.8980063
E7031	Arizona	111478372	16721755800	13.59492341	12943838	12116579	10.1661	827259	6.39114148
E7125	Arizona	146619834	21992975100	17.88046756	12943838	12697514	13.3485	246324	1.90302135
E7146	Arizona	161807926	24271188900	19.7326739	12943838	12813019	15.224	130819	1.01066623
E7208	Arizona	94077838	14111675700	11.47290707	12943838	10354240	7.7236	2589598	20.0064154
E7220	Arizona	102158574	15323786100	12.45836268	12943838	11679634	9.21478	1264204	9.76684041
E7563	Arizona	111106904	16666035600	13.54962244	12943838	11867314	9.66863	1076524	8.31688407
E7746	Arizona	141991628	21298744200	17.3160522	12943838	12054369	10.3234	889469	6.87175628
E7747	Arizona	137730300	20659545000	16.79637805	12943838	12359471	11.773	584367	4.51463469
E7751	Arizona	92480368	13872055200	11.27809366	12943838	11020768	8.09017	1923070	14.8570308
E7752	Arizona	117732082	17659812300	14.35757098	12943838	12421911	11.2764	521927	4.03224299
E7927	Arizona	98151402	14722710300	11.96968317	12943838	11211484	8.37311	1732354	13.3836193
E7932	Arizona	130151520	19522728000	15.87213659	12943838	12268329	10.7613	675509	5.21876896
E7934	Arizona	128517028	19277554200	15.67280829	12943838	12540699	12.131	403139	3.11452446
E7946	Arizona	155434834	23315225100	18.95546756	12943838	12748027	14.6051	195811	1.51277388
E7969	Arizona	92190924	13828638600	11.24279561	12943838	-6389739	8.0305	19333577	149.365103
E8013	Arizona	131934000	19790100000	16.0895122	12943838	12561977	12.1835	381861	2.95013736
E8017	Arizona	130276824	19541523600	15.88741756	12943838	12535277	11.8789	408561	3.15641311

Table C1 continued

E8024	Arizona	134157622	20123643300	16.36068561	12943838	12660971	12.7602	282867	2.18534101
E8025	Arizona	132408576	19861286400	16.14738732	12943838	12657681	12.9121	286157	2.21075851
E8030	Arizona	97433292	14614993800	11.88210878	12943838	11439808	8.92644	1504030	11.6196603
E8031	Arizona	100730906	15109635900	12.28425683	12943838	11635546	9.16484	1308292	10.1074504
E8032	Arizona	117371682	17605752300	14.31361976	12943838	12352714	10.8434	591124	4.56683713
E8142	Arizona	109406486	16410972900	13.34225439	12943838	11470804	9.14819	1473034	11.380195
E8946	Mexico	122213486	18332022900	14.90408366	12943838	12577002	12.0374	366836	2.83405896
E8947	Mexico	98253366	14738004900	11.9821178	12943838	12062335	9.75707	881503	6.81021348
E8948	Mexico	89454674	13418201100	10.90910659	12943838	11829242	9.16661	1114596	8.6110163
E8949	Mexico	102656188	15398428200	12.51904732	12943838	12164445	10.0568	779393	6.0213439
E8954	Texas (West)	114815894	17222384100	14.00193829	12943838	12444986	11.3858	498852	3.85397283
E9030	Texas (West)	79914678	11987201700	9.745692439	12943838	11122809	7.90628	1821029	14.0686943
E9031	Texas (West)	102525124	15378768600	12.5030639	12943838	12196577	10.1749	747261	5.77310223
E9032	Texas (West)	101415580	15212337000	12.36775366	12943838	12145689	9.98901	798149	6.16624683
E9033	Texas (West)	85527246	12829086900	10.43015195	12943838	11581056	8.89278	1362782	10.5284229
E9034	Texas (West)	95445534	14316830100	11.63969927	12943838	12020281	9.79772	923557	7.13510939
E9035	Texas (West)	83440134	12516020100	10.1756261	12943838	11422301	8.54267	1521537	11.7549138
E9036	Texas (West)	92236562	13835484300	11.24836122	12943838	11942350	9.59604	1001488	7.73717965
E9037	Texas (West)	102644342	15396651300	12.51760268	12943838	12163444	10.1572	780394	6.02907731
E9038	Texas (West)	87940724	13191108600	10.72447854	12943838	11723980	9.0032	1219858	9.42423723
E9039	Texas (West)	97522896	14628434400	11.8930361	12943838	12064246	9.79062	879592	6.7954497
E9040	Texas (West)	94882882	14232432300	11.57108317	12943838	11957626	9.5253	986212	7.61916211
E9041	Texas (West)	96182136	14427320400	11.72952878	12943838	12059319	9.87183	884519	6.83351414
E9042	Texas (West)	92566188	13884928200	11.28855951	12943838	11851473	9.2479	1092365	8.43926662
E9043	Texas (West)	107428744	16114311600	13.10106634	12943838	12359806	10.9296	584032	4.51204658
E9044	Texas (West)	89193526	13379028900	10.87725927	12943838	11771509	9.24577	1172329	9.05704321
E9045	Texas (West)	114293712	17144056800	13.93825756	12943838	12379545	10.9222	564293	4.35954931

Table C1 continued

E9046	Texas (West)	84839202	12725880300	10.34624415	12943838	11431827	8.46697	1512011	11.6813189
E9047	Texas (West)	87207620	13081143000	10.63507561	12943838	11438643	8.3426	1505195	11.6286607
E9048	Texas (West)	91966144	13794921600	11.21538341	12943838	11940178	9.45948	1003660	7.75395984
E9049	Texas (West)	110136570	16520485500	13.43128902	12943838	12367914	10.8713	575924	4.44940674
E9050	Texas (West)	92697098	13904564700	11.30452415	12943838	11875774	9.26797	1068064	8.25152478
E9051	Texas (West)	114499720	17174958000	13.96338049	12943838	12516441	11.7593	427397	3.30193409
E9052	Texas (West)	105227944	15784191600	12.8326761	12943838	12342975	10.8717	600863	4.64207757
E9053	Texas (West)	94642240	14196336000	11.54173659	12943838	11902521	9.39218	1041317	8.04488591
E9054	Texas (West)	84248368	12637255200	10.27419122	12943838	11370250	8.25921	1573588	12.1570434
E9055	Texas (West)	77469266	11620389900	9.447471463	12943838	10783375	7.3467	2160463	16.6910541
E9056	Texas (West)	84581900	12687285000	10.31486585	12943838	11379794	8.25995	1564044	12.0833095
E9057	Texas (West)	96119376	14417906400	11.72187512	12943838	11944440	9.46599	999398	7.72103297
E9058	Texas (West)	89158566	13373784900	10.87299585	12943838	11754101	9.04613	1189737	9.19153191
E9059	Texas (West)	106671248	16000687200	13.00868878	12943838	12221768	10.6294	722070	5.57848453
E9067	Texas (East)	135368044	20305206600	16.50829805	12943838	12699622	13.2156	244216	1.8867356
E9569	Texas (East)	114965880	17244882000	14.02022927	12943838	12579503	12.3918	364335	2.81473702
E9570	Texas (East)	84839998	12725999700	10.34634122	12943838	11747324	9.03245	1196514	9.24388887

Table C2 Maximum likelihood parameters inferred from the demographic model (Fig. 2B) for the two populations: Arizona (AZ) and Texas (TX). The composite likelihood Akaike information criterion (CLAIC) for goodness-of-fit estimations and confidence intervals (CI) for estimated parameters were inferred from 100 bootstrapped SFS.

No. of parameters	8
Log Likelihood	-2569.41
CLAIC ($\in = 10^{-4}$)	454475
Population size (95% CI)	
NA_0	117,735 (100,691 – 136,727)
NA_1	1,199,304 (989,482 – 1,341,419)
AZ_0	951,272 (741,844 – 987,410)
TX_0	248,032 (247,638 – 354,009)
AZ_1	35,523 (34,983 – 42,123)
TX_1	27,325 (26,145 – 32,832)
Migration rate per generation (95% CI)	
m_AZ-TX	3.99×10^{-4} (3.43×10^{-4} – 4.53×10^{-4})
m_TX-AZ	3.84×10^{-5} (2.67×10^{-5} – 4.93×10^{-4})
T_NA (years)	87,301 (84,435 – 96,134)
T_AZ-TX (years)	16,539 (15,145 – 26,187)

NA_0: size of ancestral population pre-expansion; NA_1: size of ancestral population after growth; AZ_0: size of Arizona population after divergence from ancestral population; TX_0: size of Texas population after divergence from ancestral population; AZ_1: current size of Arizona population; TX_1: current size of Texas population; m_AZ-TX: migration rate from Arizona to Texas population; m_TX-AZ: migration rate from Texas to Arizona population; T_NA: time of ancestral population size growth; T_AZ−TX: time of divergence.

Table C3 Results from annotations of variants

Consequence type	AZ	WTX
splice_donor_variant	1,876	1,768
splice_acceptor_variant	2,125	2,011
stop_gained	6,161	5,541
stop_lost	3,514	3,229
start_lost	2,968	2,660
missense_variant	219,144	199,020
splice_region_variant	40,144	36,476
stop_retained_variant	571	526
synonymous_variant	200,014	177,245
coding_sequence_variant	12	15
mature_miRNA_variant	548	522
5_prime_UTR_variant	439,142	403,741
3_prime_UTR_variant	424,037	390,316
non_coding_transcript_exon_variant	212,927	194,770
intron_variant	1,556,751	1,438,833
upstream_gene_variant	658,555	606,129
downstream_gene_variant	318,601	294,344
intergenic_variant	7,168,265	6,602,812
TOTAL	11,255,355	10,359,958

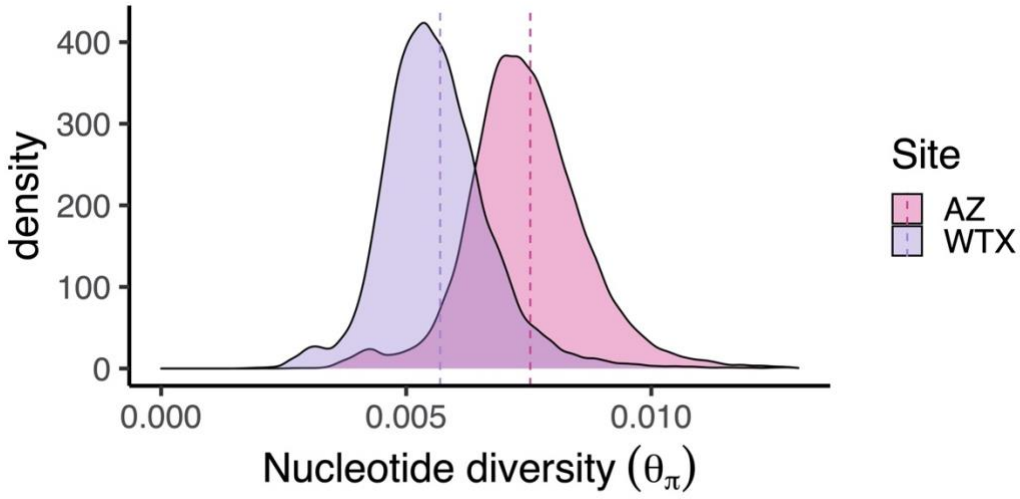


Figure S1 Distribution of nucleotide diversity (θπ) across the whole genome. θπ was calculated for every non-overlapping 1kb windows

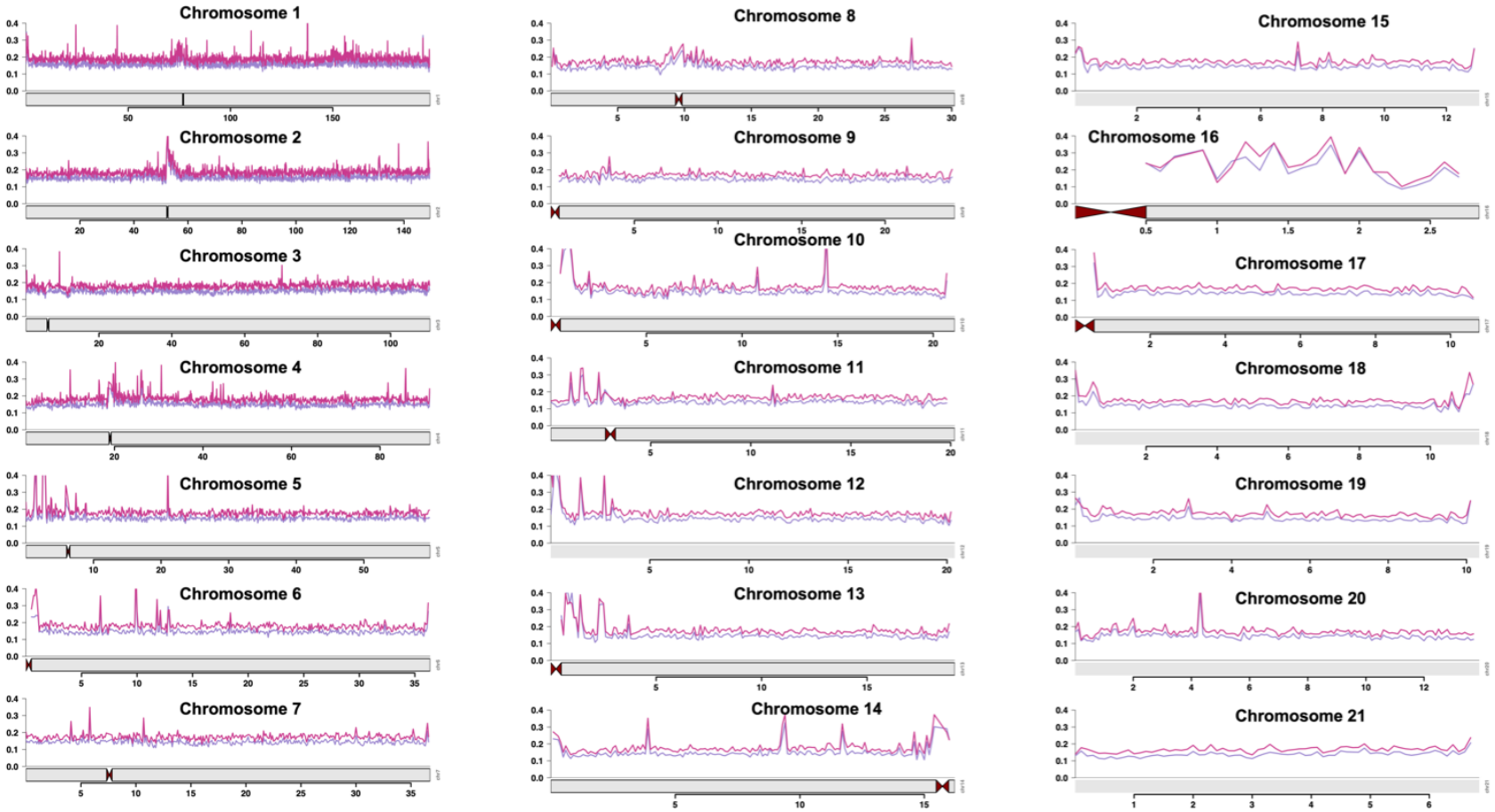
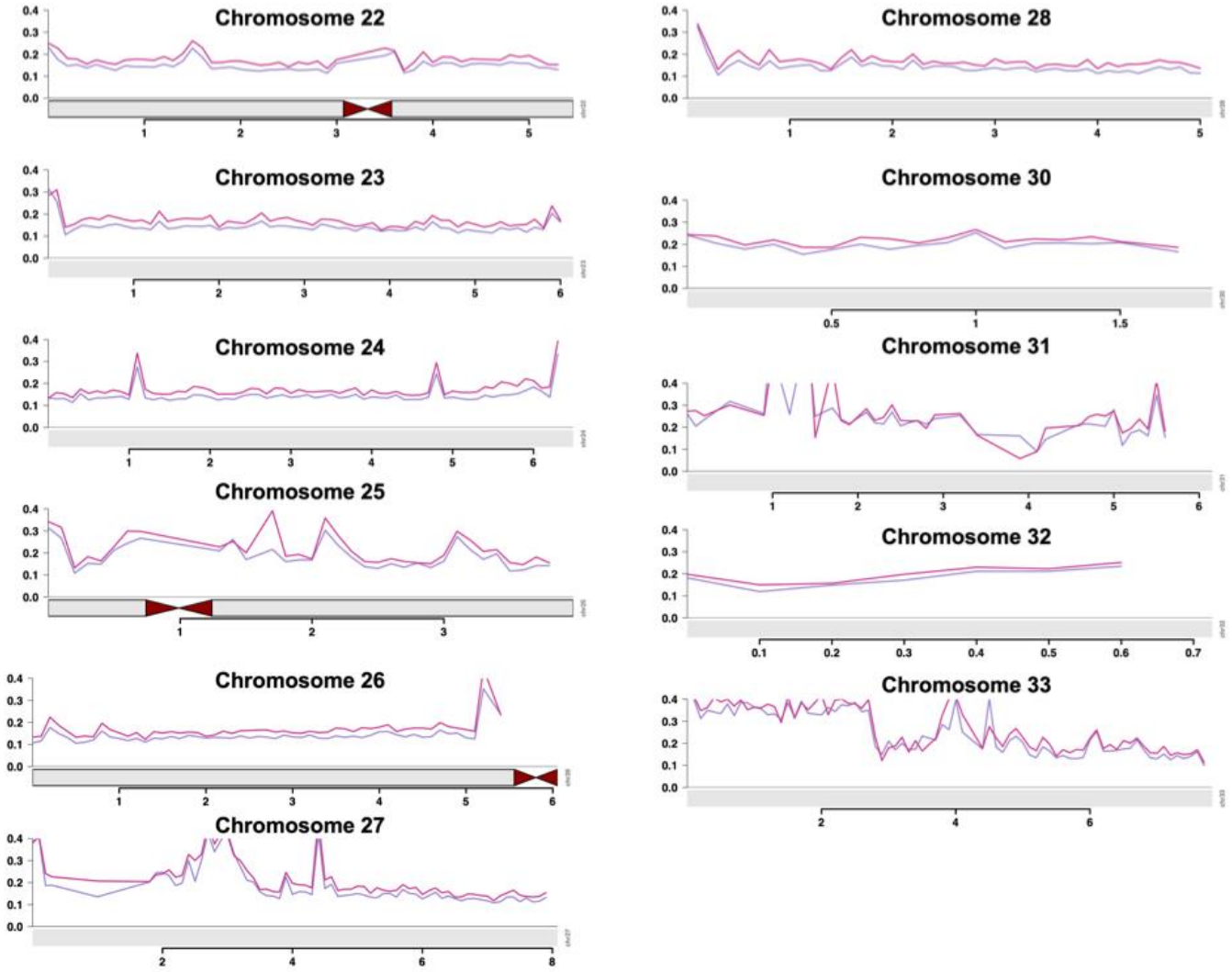


Figure C2 Sliding window analysis of heterozygosity across non-overlapping 1kb windows along each chicken chromosome for Arizona (Pink) and West Texas (purple) populations. In each figure, X-axis represents the genomic coordinates along the chromosome and Y-axis represent heterozygosity/kb. Heterozygosity was estimated as proportion of heterozygotes per 1kb of genome.

Figure C2 continued



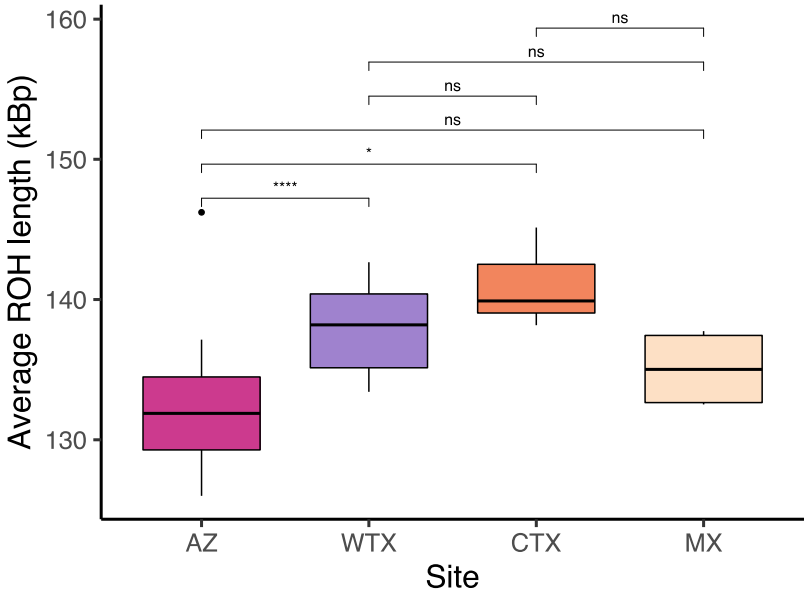


Figure C3 Average length of runs of homozygosity (ROHs) regions identified in each population. AZ=Arizona; WTX=West Texas; CTX= Central Texas; MX= Mexico. The lack of statistical significance is due to small sample sizes for CTX (N=3) and MX (N=4) populations.

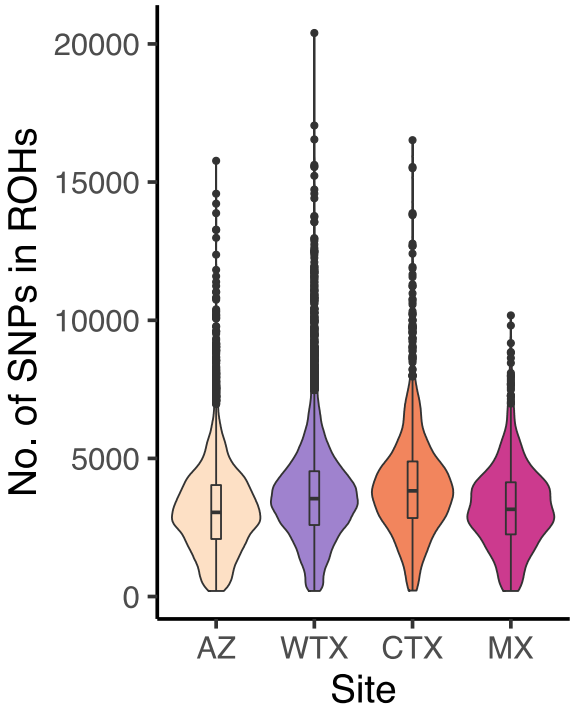


Figure C4 Violin plots showing the distribution of homozygous SNPs identified in ROHs within each population. Average length of runs of homozygosity (ROHs) regions identified in each population. AZ=Arizona; WTX=West Texas; CTX= Central Texas

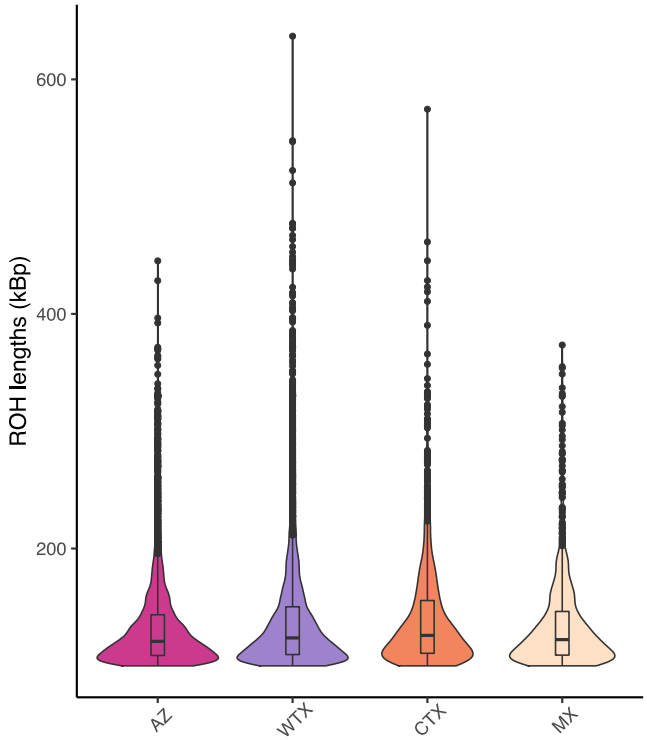


Figure C5 Violin plots showing the genome wide distribution of all ROHs (> 100kb) in each population. For all populations, most of the ROHs were between 100-200kb in size.

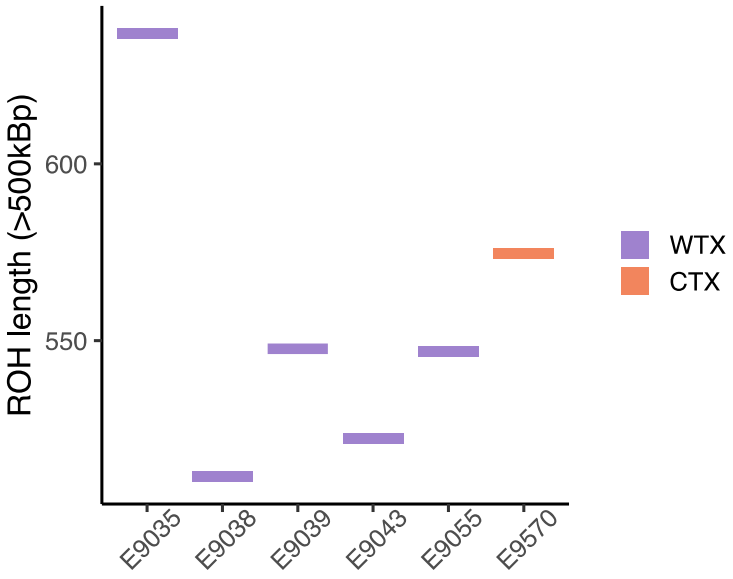


Figure C6: 6 individuals that carry ROHs longer than 500kb all belonged to Texas populations (WTX=5, CTX=1).

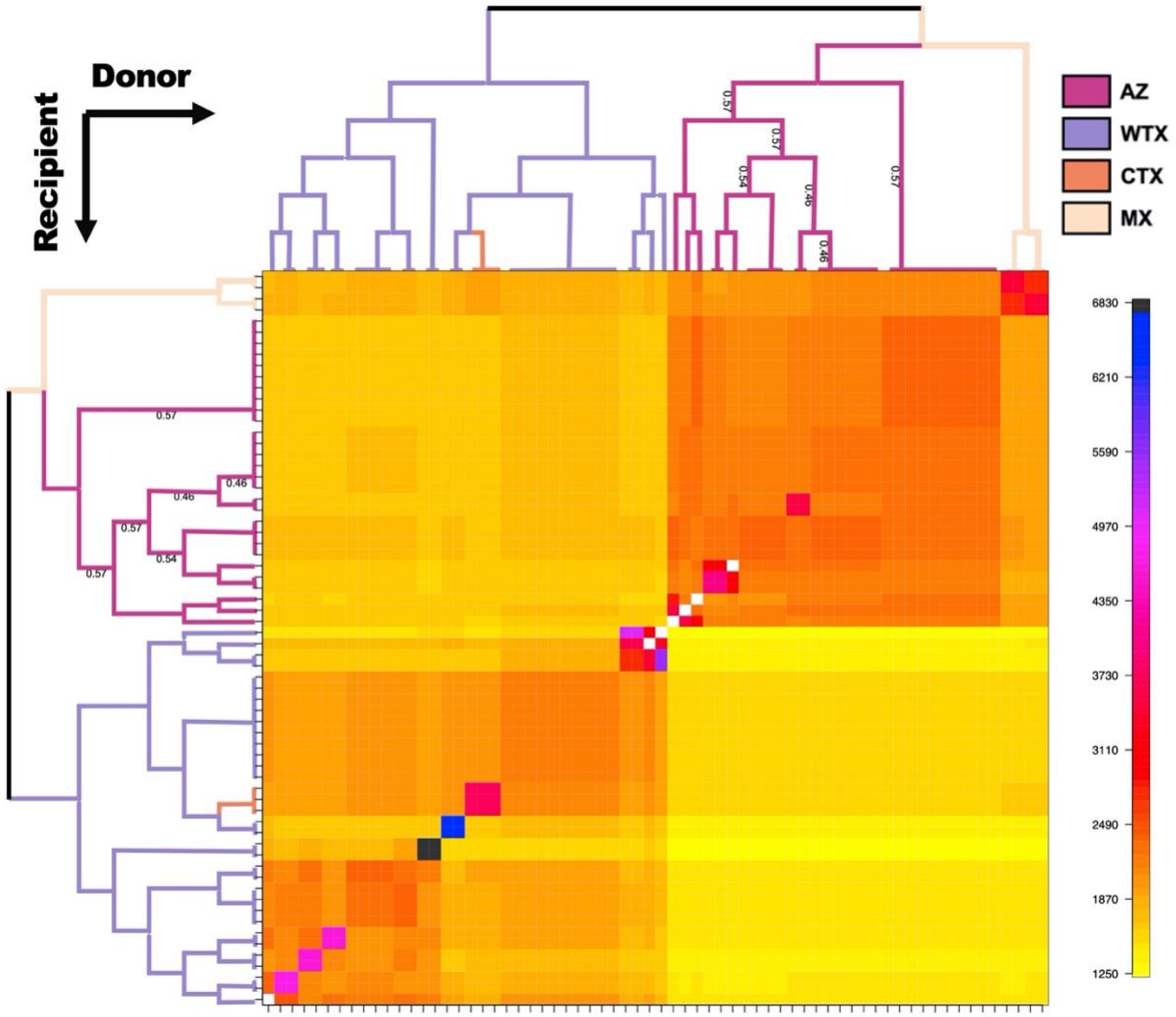


Figure C7 Co-ancestry heat map for the Montezuma Quail populations. Genealogical relationship based on pairwise analysis of identical by descent (IBD) segments in chicken chromosome 1 show that Texas population are genetically distinct with WTX (purple) and CTX (orange) more closely related than AZ (pink) or MX (light pink) populations. Co-ancestry matrix can be visualized as all individuals as recipients (rows) whose chromosome made up of genomic segments contributed by all donor individuals (columns). The heatmap shows the number of IBD chunks shared between a pair of individuals. We see that both AZ and MX samples share higher within population co-ancestry whereas CTX samples are more related to each other than rest of WTX samples. Highest co-ancestry was shared by a few pair of WTX samples with a pair identified as parent offspring (black square, Fig. S37). See methods for additional details. We saw similar co-ancestry patterns for all chromosomes (Fig. S7-33) and show that AZ samples have relatively higher co-ancestry with TX population (upper left diagonal) where TX samples share minimum co-ancestry with AZ samples (bottom right diagonal).

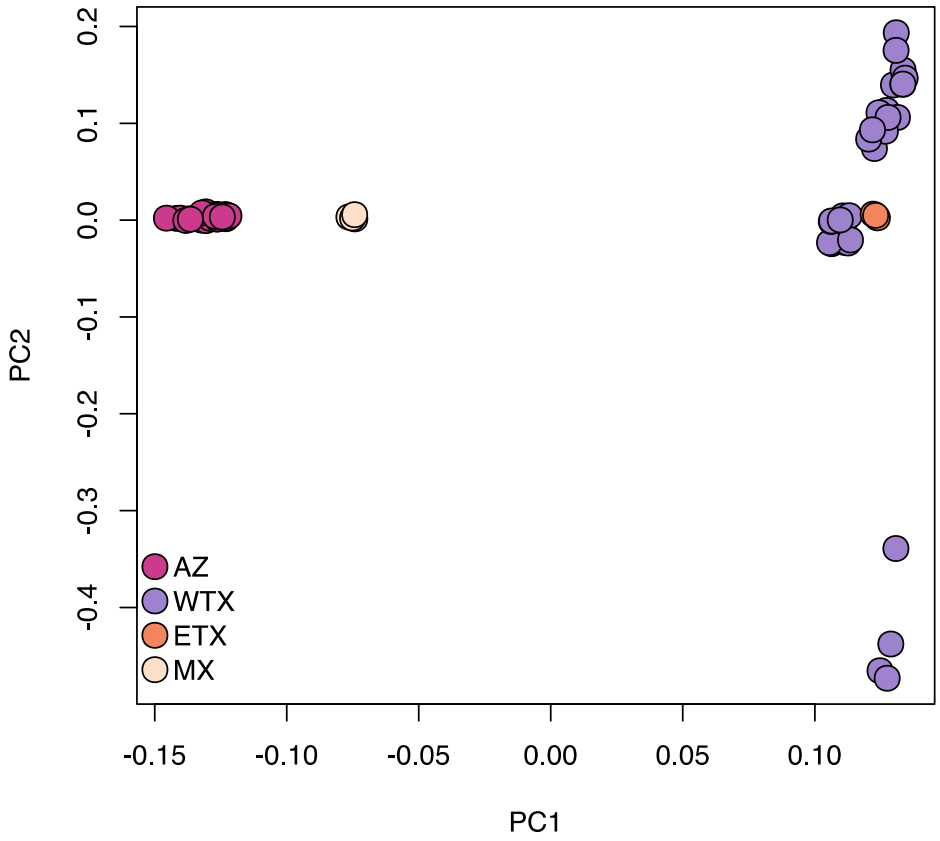


Figure C8 Principal Component Analysis (PCA) of Montezuma Quail individuals clearly demarcates the Arizona (dark pink), Texas (purple + orange), and Mexico populations (light pink). Central Texas individuals (orange) cluster with Western Texas populations (purple). The four WTX individuals in the bottom right corner belong to the same familial group (Fig. C11)

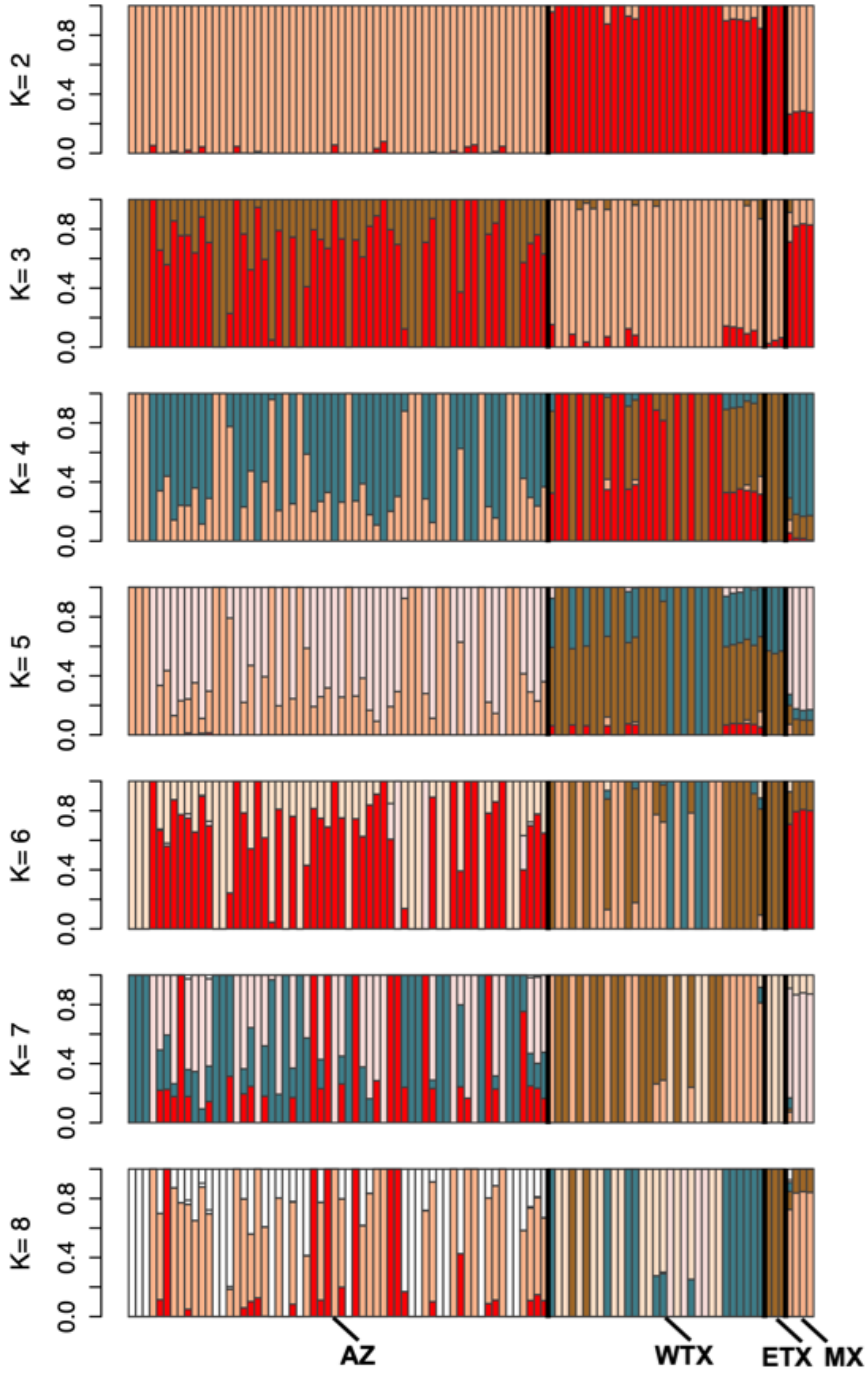


Figure C9 Admixture analysis using genotype likelihoods of all 98 Montezuma Quail individuals. Each column represents a single individual and Y-axis represent the admixture proportions for each number of ancestral populations (K). The most likely value for the number of ancestral populations based on K=2.

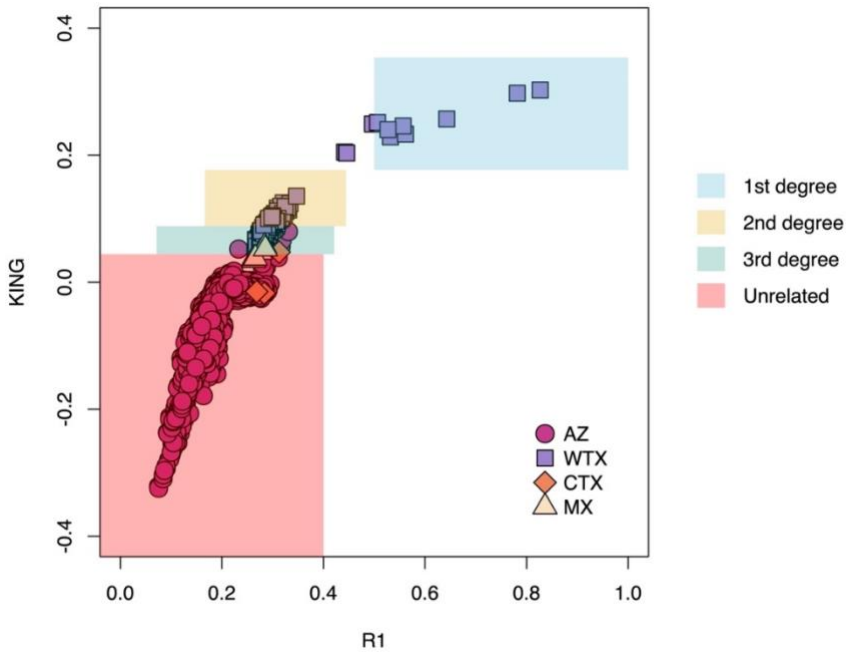


Figure C10 Relatedness analysis for each pair of Montezuma Quail individuals (N=98; number of pairwise comparisons = 4,753) using genotype likelihoods. All CTX and MX and most AZ individuals were unrelated whereas most of the WTX samples were either 2nd degree relatives or 1st degree relatives.

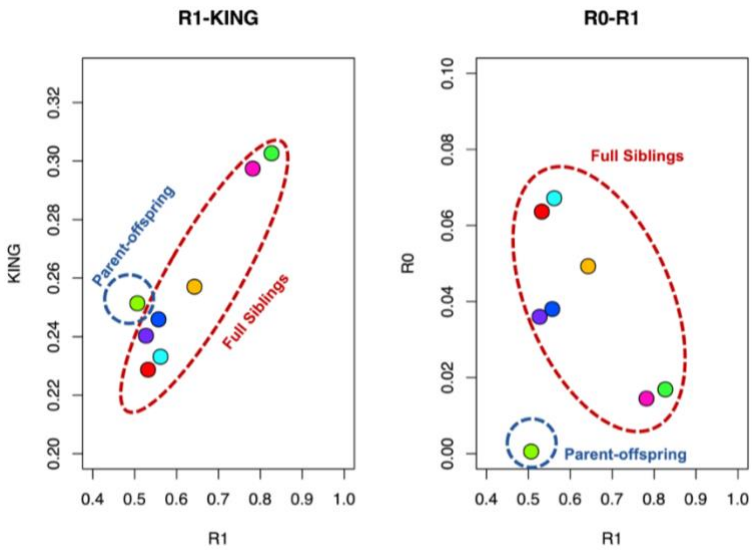


Figure C11 Classifying the 8 pair of WTX individuals that were identified as 1st degree as either parent-offspring or full sibling pairs based on R1-KING and R0-R1 relationships.

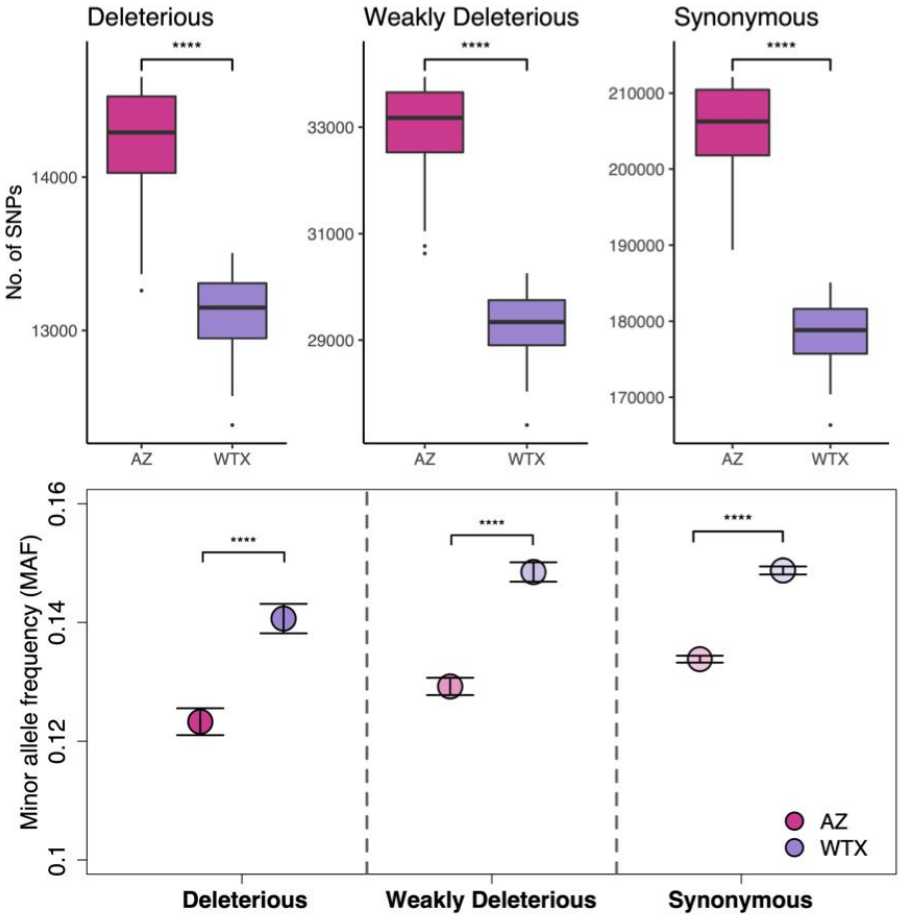


Figure C12 Number of SNPs identified as deleterious, weakly deleterious, and synonymous (upper panel) and their frequency (lower panel) within Arizona (AZ) or West Texas (WTX) populations. Since AZ is genetically more diverse, they carry more mutations of each impact type but since WTX populations are small in size, they carry those mutations at higher frequencies. Highest mean frequency for synonymous mutations and lowest mean frequency for deleterious mutations indicate the stronger impact of drift acting on neutral and nearly-neutral mutations.

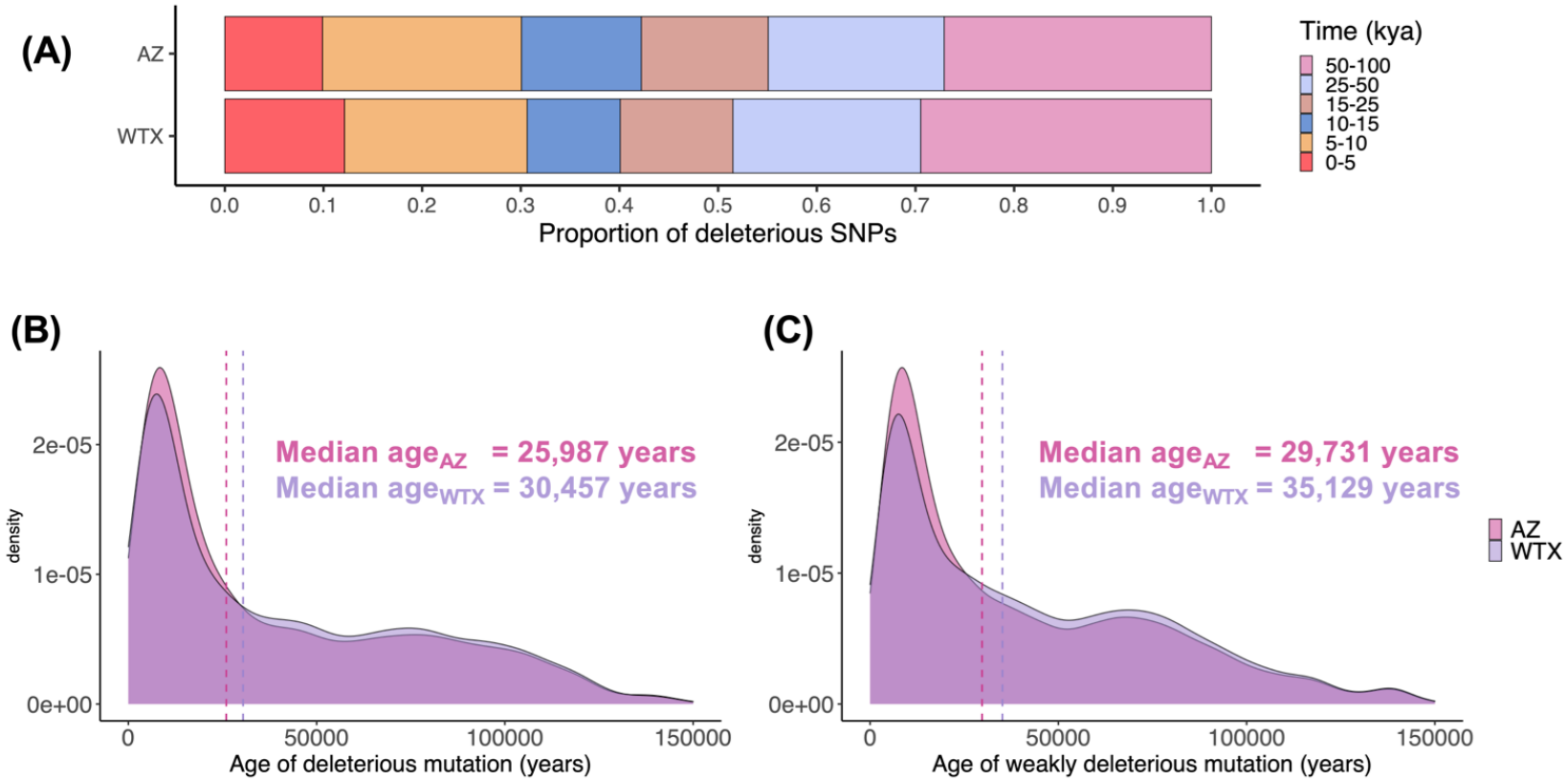


Figure C13 (A) The proportion of deleterious mutations that rose in the last 100kya that currently segregating in the two populations. Much of the contemporary load originated in the large pre-bottleneck ancestral population (>50kya) with a smaller proportion of deleterious mutations in the age range corresponding to bottlenecks in Montezuma Quail populations (10-25 kya). The smaller WTX population has higher proportion of deleterious mutations that arose in the last 5000 years. The age distribution of (B) deleterious mutations and (C) weakly deleterious mutations that are currently segregating in AZ and WTX populations

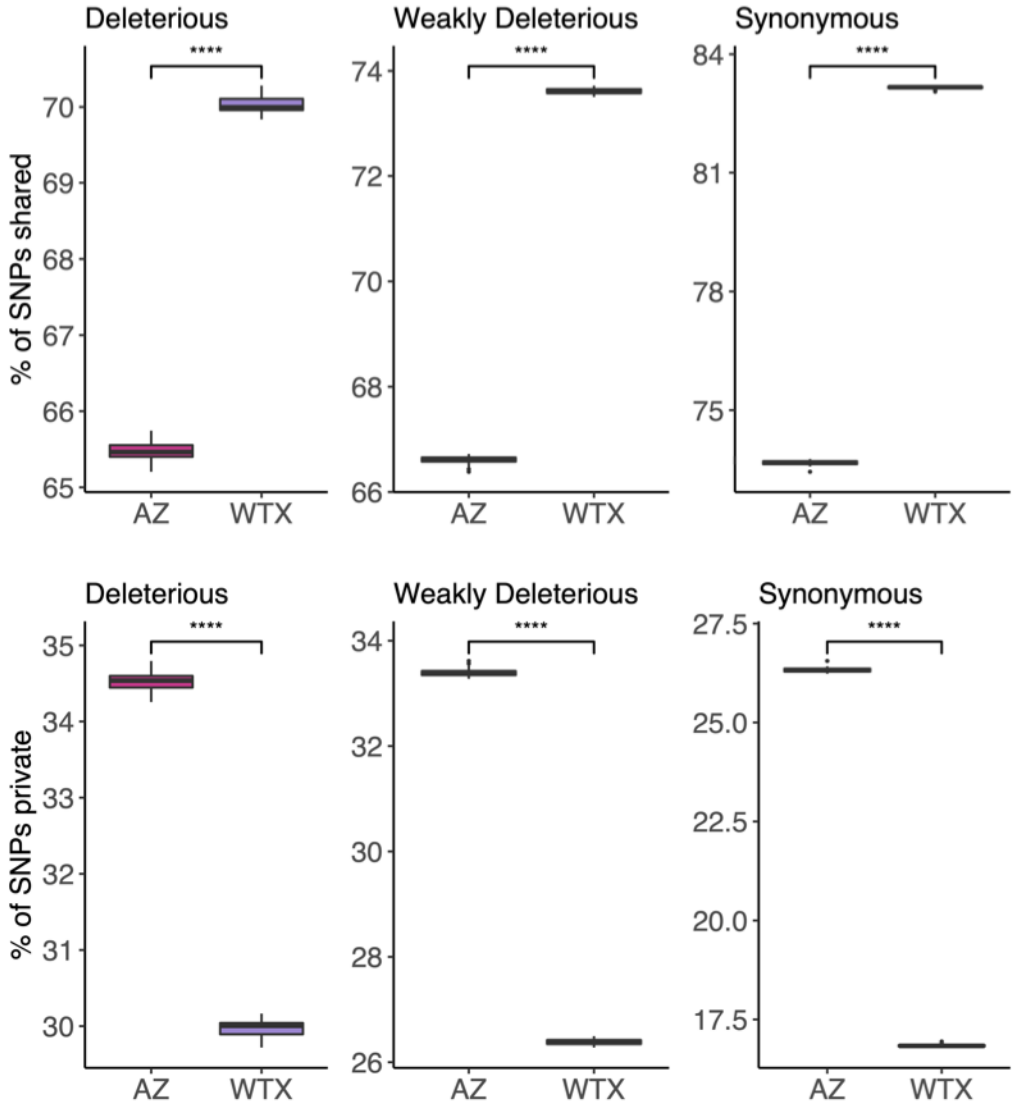


Figure C14 Proportion of contemporary mutations of each impact class that are either shared (upper panel) or privately segregating (lower panel) in Arizona (AZ) and West Texas (WTX) populations.

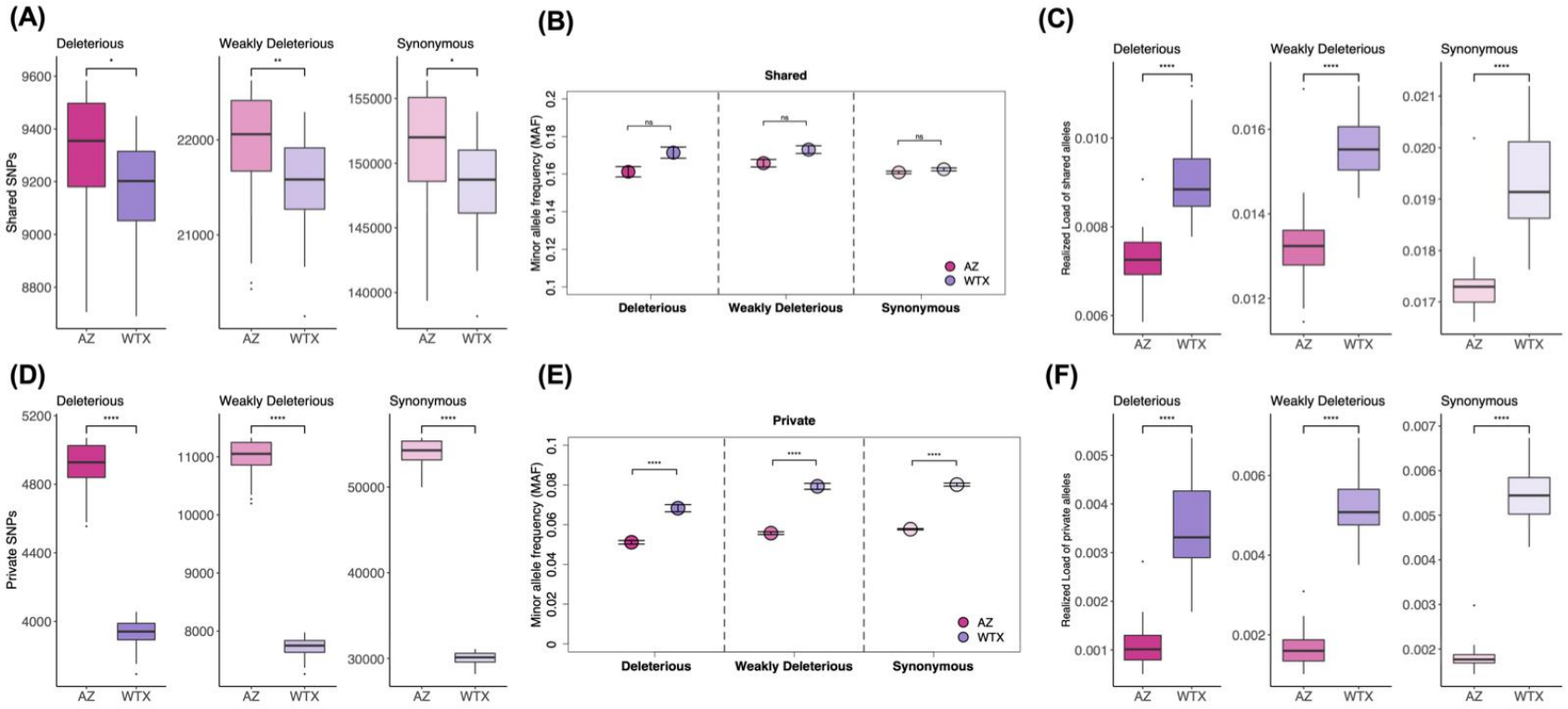


Figure C15 (A) Small WTX population carry significantly lower number of deleterious, weakly deleterious, and synonymous mutations that are shared with larger AZ population. (B) Mutations that are shared between the two populations segregate at comparable frequencies. (C) Even though both populations have similar number and frequency of shared mutations, they are significantly more homozygous in the small populations and increase the realized load of ancestral mutations. (D) Larger AZ population carry significantly higher number of mutations that are privately segregating than WTX populations. (E) The mean allele frequencies of private mutations are significantly higher in smaller populations and (F) increase the realized load of private mutations. These results show that small populations purge ancestral load but accumulate load of deleterious mutations that arise post-bottleneck.

REFERENCES

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., . . . Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology, 17*(1), 251. doi:10.1186/s13059-016-1090-1

Agrawal, A. F., & Whitlock, M. C. (2011). Inferences About the Distribution of Dominance Drawn From Yeast Gene Knockout Data. *Genetics, 187*(2), 553-566. doi:10.1534/genetics.110.124560

Albers, R. P., & Gehlbach, F. R. (1990). Choices of feeding habitat by relict Montezuma quail in central Texas. *The Wilson Bulletin*, 300–308.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics, 12*(1). doi:10.1186/1471-2105-12-246

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics, 11*(10), 697-709. doi:10.1038/nrg2844

Allendorf, F. W., Luikart, G., & Aitken, S. N. (2013). *Conservation and the genetics of populations* (2nd ed ed.). Hoboken: John Wiley & Sons.

Andrews, S., & others. (2010). *FastQC: a quality control tool for high throughput sequence data*: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F st -outlier method. *BMC Bioinformatics, 9*(1), 323. doi:10.1186/1471-2105-9-323

Antonides, J., Ricklefs, R., & DeWoody, J. A. (2017). The genome sequence and insights into the immunogenetics of the bananaquit (Passeriformes: Coereba flaveola). *Immunogenetics, 69*(3), 175-186. doi:10.1007/s00251-016-0960-8

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics, 43*(1). doi:10.1002/0471250953.bi1110s43

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2018). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics, 43*(1). doi:10.1002/0471250953.bi1110s43

Ávila, V., Amador, C., & García-Dorado, A. (2010). The purge of genetic load through restricted panmixia in a Drosophila experiment: The purging efficiency of restricted panmixia. *Journal of Evolutionary Biology, 23*(9), 1937-1946. doi:10.1111/j.1420-9101.2010.02058.x

Balick, D. J., Do, R., Cassa, C. A., Reich, D., & Sunyaev, S. R. (2015). Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. *PLOS Genetics, 11*(8), e1005436. doi:10.1371/journal.pgen.1005436

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

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., . . . Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature, 471*(7336), 51-57. doi:10.1038/nature09678

Barrett, S. C. H., & Charlesworth, D. (1991). Effects of a change in the level of inbreeding on the genetic load. *Nature, 352*(6335), 522-524. doi:10.1038/352522a0

Barton, N. H., Albers, P. K., & McVean, G. (2020). Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biology, 18*(1). doi:10.1371/journal.pbio.3000586

Beaumont, M. A. (2005). Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution, 20*(8), 435-440. doi:10.1016/j.tree.2005.05.017

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 263*(1377), 1619-1626. doi:10.1098/rspb.1996.0237

Bell, D. A., Robinson, Z. L., Funk, W. C., Fitzpatrick, S. W., Allendorf, F. W., Tallmon, D. A., & Whiteley, A. R. (2019). The Exciting Potential and Remaining Uncertainties of Genetic Rescue. *Trends in Ecology & Evolution, 34*(12), 1070-1079. doi:10.1016/j.tree.2019.06.006

Benjelloun, B., Boyer, F., Streeter, I., Zamani, W., Engelen, S., Alberti, A., . . . Pompanon, F. (2019). An evaluation of sequencing coverage and genotyping strategies to assess neutral and adaptive diversity. *Molecular Ecology Resources, 19*(6), 1497-1515. doi:10.1111/1755-0998.13070

Bensasson, D. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution, 16*(6), 314-321. doi:10.1016/S0169-5347(01)02151-6

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., . . . Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution, 69*(2), 313-319. doi:10.1016/j.ympev.2012.08.023

Bersabé, D., & García-Dorado, A. (2013). On the genetic parameter determining the efficiency of purging: an estimate for *Drosophila* egg-to-pupae viability. *Journal of Evolutionary Biology, 26*(2), 375-385. doi:10.1111/jeb.12054

Bijlsma, Bundgaard, & Putten, V. (1999). Environmental dependence of inbreeding depression and purging in Drosophila melanogaster. *Journal of Evolutionary Biology, 12*(6), 1125-1137. doi:10.1046/j.1420-9101.1999.00113.x

Bijlsma, R., & Loeschcke, V. (2005). Environmental stress, adaptation and evolution: an overview. *Journal of Evolutionary Biology, 18*(4), 744-749. doi:10.1111/j.1420-9101.2005.00962.x

Bijlsma, R., & Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments: Genetic erosion and adaptive responses. *Evolutionary Applications, 5*(2), 117-129. doi:10.1111/j.1752-4571.2011.00214.x

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes.

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

Bortoluzzi, C., Bosse, M., Derks, M. F. L., Crooijmans, R. P. M. A., Groenen, M. A. M., & Megens, H. J. (2019). The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. *Evolutionary Applications, 13*(2), 330-341. doi:10.1111/eva.12872

Bristow, K., & Ockenfels, R. (2011). Pairing season habitat selection by Montezuma quail in southeastern Arizona. *Journal of Range Management, 57*(5). doi:10.2458/azu_jrm_v57i5_bristow

Bristow, K. D., & Ockenfels, R. A. (2004). Pairing season habitat selection by Montezuma quail in southeastern Arizona. *Rangeland Ecology & Management, 57*(5), 532-538. doi:10.2111/1551-5028(2004)057[0532:Pshsbm]2.0.Co;2

Brown, D. E. (1979). Factors Influencing Reproductive Success and Population Densities in Montezuma Quail. *The Journal of Wildlife Management, 43*(2), 522. doi:10.2307/3800365

Brown, R. L. (1982). Effects of Livestock Grazing on Mearns Quail in Southeastern Arizona. *Journal of Range Management, 35*(6), 727. doi:10.2307/3898250

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics, 103*(3), 338-348. doi:10.1016/j.ajhg.2018.07.015

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

Brüniche-Olsen, A., Kellner, K. F., & DeWoody, J. A. (2019). Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Molecular Ecology, 28*(22), 4914-4925. doi:10.1111/mec.15266

Bürger, R., & Lynch, M. (1995). Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution, 49*(1), 151-163. doi:10.1111/j.1558-5646.1995.tb05967.x

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., . . . Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum ofFicedulaflycatchers. *Genome Research, 25*(11), 1656-1665. doi:10.1101/gr.196485.115

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner.

Cai, J. J., Macpherson, J. M., Sella, G., & Petrov, D. A. (2009). Pervasive Hitchhiking at Coding and Regulatory Sites in Humans. *PLOS Genetics, 5*(1), e1000336. doi:10.1371/journal.pgen.1000336

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., . . . Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature, 486*(7401), 59-67. doi:10.1038/nature11148

Caye, K., Jay, F., Michel, O., & François, O. (2018). Fast inference of individual admixture coefficients using geographic data. *The Annals of Applied Statistics, 12*(1), 586-608. doi:10.1214/17-AOAS1106

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics, 19*(4), 220-234. doi:10.1038/nrg.2017.109

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. doi:10.1126/sciadv.1400253

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

Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetical Research, 74*(3), 329-340. doi:10.1017/S0016672399004152

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics, 134*(4), 1289-1303.

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics, 18*(1), 237-268. doi:10.1146/annurev.es.18.110187.001321

Chavarria, P. M., Montoya, A., Silvy, N. J., & Lopez, R. R. (2012, 2012). *Impact of inclement weather on overwinter mortality of Montezuma quail in southeast Arizona*.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ; iso-2; iso-3. *Fly, 6*(2), 80-92. doi:10.4161/fly.19695

Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications, 12*(10), 1971-1987. doi:10.1111/eva.12861

Copenhaver, G. P., Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLOS Genetics, 8*(1). doi:10.1371/journal.pgen.1002453

Cox, W. A., Kimball, R. T., & Braun, E. L. (2007). Phylogenetic Position of the New World Quail (Odontophoridae): Eight Nuclear Loci and Three Mitochondrial Regions Contradict Morphology and the Sibley-Ahlquist Tapestry. *The Auk, 124*(1), 71-84. doi:10.1093/auk/124.1.71

Crnokrak, P., & Barrett, S. C. H. (2002). Perspective: purging the genetic load: a review of the experimental evidence. *Evolution, 56*(12), 2347-2358. doi:10.1111/j.0014-3820.2002.tb00160.x

Crow, J. F. (1958). Some possibilities for measuring selection intensities in man. *Hum Biol, 30*(1), 1-13. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13513111

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

de Brito Martins, A., & de Aguiar, M. A. M. (2017). Barriers to gene flow and ring species formation. *Evolution, 71*(2), 442-448. doi:10.1111/evo.13121

de Villemereuil, P., Rutschmann, A., Lee, K. D., Ewen, J. G., Brekke, P., & Santure, A. W. (2019). Little Adaptive Potential in a Threatened Passerine Bird. *Current Biology, 29*(5), 889-894.e883. doi:10.1016/j.cub.2019.01.072

Department, T. P. a. W. (2018). Species of Greatest Conservation Need.

DeWoody, J. A., Fernandez, N. B., Brüniche-Olsen, A., Antonides, J. D., Doyle, J. M., San Miguel, P., . . . Bickham, J. W. (2017). Characterization of the Gray Whale *Eschrichtius robustus* Genome and a Genotyping Array Based on Single-Nucleotide Polymorphisms in Candidate Genes. *The Biological Bulletin, 232*(3), 186-197. doi:10.1086/693483

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science, 345*(6195), 401-406. doi:10.1126/science.1251817

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 (N_e) from genetic data. *Molecular Ecology Resources, 14*(1), 209-214. doi:10.1111/1755-0998.12157

Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S., & Reich, D. (2015). No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature Genetics, 47*(2), 126-131. doi:10.1038/ng.3186

Dobrynin, P., O'Brien, S. J., Koepfli, K.-P., Ulyantsev, V., & Noskova, E. (2020). GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience, 9*(3). doi:10.1093/gigascience/giaa005

Doyle, J. M., Bell, D. A., Bloom, P. H., Emmons, G., Fesnock, A., Katzner, T. E., . . . Andrew DeWoody, J. (2018). New insights into the phylogenetics and population structure of the prairie falcon (Falco mexicanus). *BMC Genomics, 19*(1), 233. doi:10.1186/s12864-018-4615-z

Doyle, J. M., Katzner, T. E., Roemer, G. W., Cain, J. W., Millsap, B. A., McIntyre, C. L., . . . Andrew DeWoody, J. (2016). Genetic structure and viability selection in the golden eagle (Aquila chrysaetos), a vagile raptor with a Holarctic distribution. *Conservation Genetics, 17*(6), 1307-1322. doi:10.1007/s10592-016-0863-0

Ellegren, H., & Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. *Nature, 452*(7184), 169-175. doi:10.1038/nature06737

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., . . . Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature, 491*(7426), 756-760. doi:10.1038/nature11584

Eo, S. H., Doyle, J. M., & DeWoody, J. A. (2011). Genetic diversity in birds is associated with body mass and habitat type: Microsatellite diversity in birds. *Journal of Zoology, 283*(3), 220-226. doi:10.1111/j.1469-7998.2010.00773.x

Epps, C. W., Crowhurst, R. S., & Nickerson, B. S. (2018). Assessing changes in functional connectivity in a desert bighorn sheep metapopulation after two generations. *Molecular Ecology, 27*(10), 2334-2346. doi:10.1111/mec.14586

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology, 14*(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics, 40*(1), 481-501. doi:10.1146/annurev.ecolsys.39.110707.173414

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

Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics, 34*(1), 487-515. doi:10.1146/annurev.ecolsys.34.011802.132419

Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics, 180*(2), 977-993. doi:10.1534/genetics.108.092221

Frankham, R. (1996). Relationship of Genetic Variation to Population Size in Wildlife. *Conservation Biology, 10*(6), 1500-1508. doi:10.1046/j.1523-1739.1996.10061500.x

Frankham, R. (2005). Genetics and extinction. *Biological Conservation, 126*(2), 131-140. doi:10.1016/j.biocon.2005.05.002

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

Franklin, I. R., & Frankham, R. (1998). How large must populations be to retain evolutionary potential? *Animal Conservation, 1*(1), 69-70. doi:10.1111/j.1469-1795.1998.tb00228.x

Freamo, H., O'Reilly, P., Berg, P. R., Lien, S., & Boulding, E. G. (2011). Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs: SNP GENOTYPING AND APPLICATIONS. *Molecular Ecology Resources, 11*, 254-267. doi:10.1111/j.1755-0998.2010.02952.x

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution, 6*(8), 925-929. doi:10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics, 196*(4), 973-983. doi:10.1534/genetics.113.160572

Fu, W., Gittelman, Rachel M., Bamshad, Michael J., & Akey, Joshua M. (2014). Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations. *The American Journal of Human Genetics, 95*(4), 421-436. doi:10.1016/j.ajhg.2014.09.006

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics, 147*(2), 915-925.

Garner, T. W. J., Pearman, P. B., & Angelone, S. (2004). Genetic diversity across a vertebrate species' range: a test of the central-peripheral hypothesis: GENETIC DIVERSITY OF RANA LATASTEI POPULATIONS. *Molecular Ecology, 13*(5), 1047-1053. doi:10.1111/j.1365-294X.2004.02119.x

Garrison, E. (2012). Vcflib: A C++ library for parsing and manipulating VCF files. In.

Gong, W., Gu, L., & Zhang, D. (2010). Low genetic diversity and high genetic divergence caused by inbreeding and geographical isolation in the populations of endangered species Loropetalum subcordatum (Hamamelidaceae) endemic to China. *Conservation Genetics, 11*(6), 2281-2288. doi:10.1007/s10592-010-0113-9

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes, 5*(1), 184-186. doi:10.1111/j.1471-8286.2004.00828.x

Groenen, M. A. M., Wahlberg, P., Foglio, M., Cheng, H. H., Megens, H. J., Crooijmans, R. P. M. A., . . . Andersson, L. (2008). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research, 19*(3), 510-519. doi:10.1101/gr.086538.108

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications, 11*(1), 1001. doi:10.1038/s41467-020-14803-1

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research, 41*(13), e129-e129. doi:10.1093/nar/gkt371

Haller, B. C., Messer, P. W., & Hernandez, R. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution, 36*(3), 632-637. doi:10.1093/molbev/msy228

Halley, Y. A., Dowd, S. E., Decker, J. E., Seabury, P. M., Bhattarai, E., Johnson, C. D., . . . Seabury, C. M. (2014). A Draft De Novo Genome Assembly for the Northern Bobwhite (Colinus virginianus) Reveals Evidence for a Rapid Decline in Effective Population Size Beginning in the Late Pleistocene. *PLoS ONE, 9*(3), e90240. doi:10.1371/journal.pone.0090240

Hampe, A., & Jump, A. S. (2011). Climate Relicts: Past, Present, Future. *Annual Review of Ecology, Evolution, and Systematics, 42*(1), 313-333. doi:10.1146/annurev-ecolsys-102710-145015

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters: Rear edges and climate change. *Ecology Letters, 8*(5), 461-467. doi:10.1111/j.1461-0248.2005.00739.x

Harder, A. M., Willoughby, J. R., Ardren, W. R., & Christie, M. R. (2020). Among-family variation in survival and gene expression uncovers adaptive genetic variation in a threatened fish. *Molecular Ecology, 29*(6), 1035-1049. doi:10.1111/mec.15334

Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA.*

Harveson, L. A. (2009). *Management of Montezuma quail in Texas: barriers to establishing a hunting season.* Paper presented at the National Quail Symposium Proceedings.

Harveson, L. A., Allen, T. H., Hernández, F., Holdermann, D. A., Mueller, J. M., & Whitley, M. S. (2007a). Montezuma quail ecology and life history. *Texas quails. Texas A&M University, College Station, USA*, 23-29.

Harveson, L. A., Allen, T. H., Hernández, F., Holdermann, D. A., Mueller, J. M., & Whitley, M. S. (2007b). Montezuma quail ecology and life history. *Texas quails. Texas A&M University, College Station, USA*, 23–29.

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding Inbreeding Depression, Purging, and Genetic Rescue. *Trends in Ecology & Evolution, 31*(12), 940-952. doi:10.1016/j.tree.2016.09.005

Hedrick, P. W., Robinson, J. A., Peterson, R. O., & Vucetich, J. A. (2019). Genetics and extinction and the example of Isle Royale wolves. *Animal Conservation, 22*(3), 302-309. doi:10.1111/acv.12479

Heffelfinger, J. R., & Olding, R. J. (2000). *Montezuma quail management in Arizona.* Paper presented at the National Quail Symposium Proceedings.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., . . . Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: Analytical Approaches. *Molecular Ecology Resources, 11*, 123-136. doi:10.1111/j.1755-0998.2010.02943.x

Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G., & Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Reviews Genetics, 16*(6), 333-343. doi:10.1038/nrg3931

Hernandez, F., Garza, E., Harveson, L. A., & Brewer, C. E. (2009, 2009). *Fate and survival of radio-marked Montezuma quail*.

Hernandez, F., Harveson, L. A., & Brewer, C. E. (2006). A comparison of trapping techniques for Montezuma Quail. *Wildlife Society Bulletin, 34*(4), 1212–1215.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., . . . Wilson, R. K. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature, 432*(7018), 695-716. doi:10.1038/nature03154

Holderegger, R., Balkenhol, N., Bolliger, J., Engler, J. O., Gugerli, F., Hochkirch, A., . . . Zachos, F. E. (2019). Conservation genetics: Linking science with practice. *Molecular Ecology, 28*(17), 3848-3856. doi:10.1111/mec.15202

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics, 12*(1), 491. doi:10.1186/1471-2105-12-491

Holycross, A. T., & Douglas, M. E. (2007). Geographic isolation, genetic divergence, and ecological non-exchangeability define ESUs in a threatened sky-island rattlesnake. *Biological Conservation, 134*(1), 142-154. doi:10.1016/j.biocon.2006.07.020

Hosner, P. A., Braun, E. L., & Kimball, R. T. (2015). Land connectivity changes and global cooling shaped the colonization history and diversification of New World quail (Aves: Galliformes: Odontophoridae). *Journal of Biogeography, 42*(10), 1883-1895. doi:10.1111/jbi.12555

Hunter, M. L., & Hutchinson, A. (1994). The Virtues and Shortcomings of Parochialism: Conserving Species That Are Locally Rare, but Globally Common. *Conservation Biology, 8*(4), 1163-1165. doi:10.1046/j.1523-1739.1994.08041163.x

Irwin, D. E. (2005). Speciation by Distance in a Ring Species. *Science, 307*(5708), 414-416. doi:10.1126/science.1105201

Kardos, M., Qvarnström, A., & Ellegren, H. (2017). Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences. *Genetics, 205*(3), 1319-1334. doi:10.1534/genetics.116.198861

Keller, L. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution, 17*(5), 230-241. doi:10.1016/S0169-5347(02)02489-8

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, a. D. (2002). The Human Genome Browser at UCSC. *Genome Research, 12*(6), 996-1006. doi:10.1101/gr.229102

Kim, B. Y., Huber, C. D., & Lohmueller, K. E. (2017). Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. *Genetics, 206*(1), 345-361. doi:10.1534/genetics.116.197145

Kimura, M., Maruyama, T., & Crow, J. F. (1963). The Mutation Load in Small Populations. *Genetics, 48*, 1303-1312. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14071753

Kimura, M., & Ohta, T. (1969). The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics, 61*(3), 763-771.

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

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2020). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *bioRxiv*.

Lamichhaney, S., & Andersson, L. (2019). A comparison of the association between large haplotype blocks under selection and the presence/absence of inversions. *Ecology and Evolution, 9*(8), 4888-4896. doi:10.1002/ece3.5094

Leopold, A. S., & McCabe, R. A. (1957). Natural History of the Montezuma Quail in Mexico. *The Condor, 59*(1), 3-26. doi:10.2307/1364613

Leroy, G., Carroll, E. L., Bruford, M. W., DeWoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications, 11*(7), 1066-1083. doi:10.1111/eva.12564

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Subgroup, G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, H., Xiang-Yu, J., Dai, G., Gu, Z., Ming, C., Yang, Z., . . . Zhang, Y.-P. (2016). Large numbers of vertebrates began rapid population decline in the late 19th century. *Proceedings of the National Academy of Sciences, 113*(49), 14079-14084. doi:10.1073/pnas.1616804113

Li, S., Li, B., Cheng, C., Xiong, Z., Liu, Q., Lai, J., . . . Yan, J. (2014). Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology, 15*(12), 557. doi:10.1186/s13059-014-0557-1

Lino, A., Fonseca, C., Rojas, D., Fischer, E., & Ramos Pereira, M. J. (2019). A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. *Mammalian Biology, 94*, 69-76. doi:10.1016/j.mambio.2018.09.006

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. *Nature, 462*(7276), 1052-1055. doi:10.1038/nature08649

Lohmueller, K. E. (2014). The distribution of deleterious genetic variation in human populations. *Current Opinion in Genetics & Development, 29*, 139-146. doi:10.1016/j.gde.2014.09.005

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., . . . Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature, 451*(7181), 994-997. doi:10.1038/nature06611

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W., & O'Brien, S. J. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution, 39*(2), 174-190. doi:10.1007/BF00163806

Luna, R. S., Oaster, E. A., Cork, K. D., & O'Shaughnessy, R. (2017, 2017). *Changes in Habitat Use of Montezuma Quail in Response to Tree Canopy Reduction in the Capitan Mountains of New Mexico*.

Lynch, M. (2007). *The origins of genome architecture*. Sunderland, Mass: Sinauer Associates.

Lynch, M., Conery, J., & Burger, R. (1995). Mutation Accumulation and the Extinction of Small Populations. *The American Naturalist, 146*(4), 489-518. doi:10.1086/285812

Madsen, T., Stille, B., & Shine, R. (1996). Inbreeding depression in an isolated population of adders Vipera berus. *Biological Conservation, 75*(2), 113-118. doi:10.1016/0006-3207(95)00067-4

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution, 18*(4), 189-197. doi:10.1016/s0169-5347(03)00008-9

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics, 26*(22), 2867-2873. doi:10.1093/bioinformatics/btq559

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics, 27*(6), 764-770. doi:10.1093/bioinformatics/btr011

Mathur, S., & DeWoody, J. A. (2020). Genetic load has potential in large populations but is realized in small inbred populations. *Authorea*. doi:10.22541/au.158941448.85174067

Mathur, S., Tomeček, J. M., Heniff, A., Luna, R., & DeWoody, J. A. (2019). Evidence of genetic erosion in a peripheral population of a North American game bird: the Montezuma quail (Cyrtonyx montezumae). *Conservation Genetics, 20*(6), 1369-1381. doi:10.1007/s10592-019-01218-9

Matthey-Doret, R., & Whitlock, M. C. (2019). Background selection and F_{ST} : Consequences for detecting local adaptation. *Molecular Ecology, 28*(17), 3902-3914. doi:10.1111/mec.15197

Mayr, E. (1999). *Systematics and the origin of species, from the viewpoint of a zoologist*: Harvard University Press.

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

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., . . . Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology, 17*(1). doi:10.1186/s13059-016-0974-4

McLennan, E. A., Wright, B. R., Belov, K., Hogg, C. J., & Grueber, C. E. (2019). Too much of a good thing? Finding the most informative genetic data set to answer conservation questions. *Molecular Ecology Resources, 19*(3), 659-671. doi:10.1111/1755-0998.12997

Meirmans, P. G., & Van Tienderen, P. H. (2004). genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes, 4*(4), 792-794. doi:10.1111/j.1471-8286.2004.00770.x

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics, 210*(2), 719-731. doi:10.1534/genetics.118.301336

Meisner, J., & Albrechtsen, A. (2019). Testing for Hardy–Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Molecular Ecology Resources, 19*(5), 1144-1152. doi:10.1111/1755-0998.13019

Mitton, J. B. (2000). *Selection in natural populations* (1. paperback ed ed.). New York, NY: Oxford Univ. Press.

Morris, K. M., Hindle, M. M., Boitard, S., Burt, D. W., Danner, A. F., Eory, L., . . . Smith, J. (2020). The quail genome: insights into social behaviour, seasonal biology and infectious disease response. *BMC Biology, 18*(1), 14. doi:10.1186/s12915-020-0743-4

Mukai, T., Chigusa, S. I., Mettler, L. E., & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. *Genetics, 72*(2), 335-355. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4630587

Muller, H. J. (1950). Our load of mutations. *Am J Hum Genet, 2*(2), 111-176. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14771033

Mussmann, S. M., Douglas, M. R., Anthonysamy, W. J. B., Davis, M. A., Simpson, S. A., Louis, W., & Douglas, M. E. (2017). Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. *Royal Society Open Science, 4*(2), 160736. doi:10.1098/rsos.160736

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. doi:10.1016/j.cub.2015.03.047

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., . . . Stamatoyannopoulos, J. A. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics, 28*(14), 1919-1920. doi:10.1093/bioinformatics/bts277

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine, 17*(8), 873-890. doi:10.1002/(sici)1097-0258(19980430)17:8<873::aid-sim779>3.0.co;2-i

Ng, P. C. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research, 31*(13), 3812-3814. doi:10.1093/nar/gkg509

Nicholls, J. A., Double, M. C., Rowell, D. M., & Magrath, R. D. (2000). The evolution of cooperative and pair breeding in thornbills *Acanthiza* (Pardalotidae). *Journal of Avian Biology, 31*(2), 165-176. doi:10.1034/j.1600-048X.2000.310208.x

Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics, 23*(1), 263-286. doi:10.1146/annurev.es.23.110192.001403

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

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., . . . Dalén, L. (2015). Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. *Current Biology, 25*(10), 1395-1400. doi:10.1016/j.cub.2015.04.007

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology, 17*(15), 3428-3447. doi:10.1111/j.1365-294X.2008.03842.x

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

Pool, J. E., & Nielsen, R. (2007). POPULATION SIZE CHANGES RESHAPE GENOMIC PATTERNS OF DIVERSITY. *Evolution, 61*(12), 3001-3006. doi:10.1111/j.1558-5646.2007.00238.x

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

Pulanić, D., Polašek, O., Petrovečki, M., Vorko-Jović, A., Peričić, M., Lauc, L. B., . . . Rudan, I. (2008). Effects of Isolation and Inbreeding on Human Quantitative Traits: An Example of Biochemical Markers of Hemostasis and Inflammation. *Human Biology, 80*(5), 513-533. doi:10.3378/1534-6617-80.5.513

Puzey, J. R., Willis, J. H., & Kelly, J. K. (2017). Population structure and local selection yield high genomic variation in *Mimulus guttatus*. *Molecular Ecology, 26*(2), 519-535. doi:10.1111/mec.13922

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

Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation, 251*. doi:10.1016/j.biocon.2020.108784

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical Properties of New Neutrality Tests Against Population Growth. *Molecular Biology and Evolution, 19*(12), 2092-2100. doi:10.1093/oxfordjournals.molbev.a004034

Randel, C. J., Johnson, C. Z., Chavarria, P. M., Lopez, R. R., Silvy, N. J., & Tomeček, J. M. (2019). Estimating Montezuma quail hatch date using primary molt at harvest. *Wildlife Society Bulletin, 43*(4), 766-768. doi:10.1002/wsb.1017

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology, 17*(1), 230-237. doi:10.1046/j.1523-1739.2003.01236.x

Rettelbach, A., Nater, A., & Ellegren, H. (2019). How Linked Selection Shapes the Diversity Landscape in *Ficedula* Flycatchers. *Genetics, 212*(1), 277-285. doi:10.1534/genetics.119.301991

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology, 28*(21), 3487-3494.e3484. doi:10.1016/j.cub.2018.08.066

Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances, 5*(5). doi:10.1126/sciadv.aau0757

Rogers, R. L., & Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genetics, 13*(3), e1006601. doi:10.1371/journal.pgen.1006601

Rollins, D. (2002, 2002). *Sustaining the'quail wave'in the southern Great Plains*.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics, 145*(4), 1219-1228.

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

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. doi:10.1093/bioinformatics/btv351

Simons, Y. B., Turchin, M. C., Pritchard, J. K., & Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. *Nature Genetics, 46*(3), 220-224. doi:10.1038/ng.2896

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

Soulé, M. E. (1985). What Is Conservation Biology? *BioScience, 35*(11), 727-734. doi:10.2307/1310054

Spielman, D., Brook, B. W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences, 101*(42), 15261-15264. doi:10.1073/pnas.0403809101

Stromberg, M. R. (1990). Habitat, Movements and Roost Characteristics of Montezuma Quail in Southeastern Arizona. *The Condor, 92*(1), 229-236. doi:10.2307/1368404

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

Team, R. C., & others. (2013). *R: A language and environment for statistical computing*: Vienna, Austria.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics, 49*(2), 303-309. doi:10.1038/ng.3748

Thompson, C. J., Koshkina, V., Burgman, M. A., Butchart, S. H. M., & Stone, L. (2017). Inferring extinctions II: A practical, iterative model based on records and surveys. *Biological Conservation, 214*, 328-335. doi:10.1016/j.biocon.2017.07.029

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research, 22*(22), 4673-4680. doi:10.1093/nar/22.22.4673

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics, 14*(2), 178-192. doi:10.1093/bib/bbs017

Tishkoff, S. A., & Williams, S. M. (2002). Genetic analysis of African populations: human evolution and complex disease. *Nature Reviews Genetics, 3*(8), 611-621. doi:10.1038/nrg865

Uricchio, L. H., & Hernandez, R. D. (2014). Robust Forward Simulations of Recurrent Hitchhiking. *Genetics, 197*(1), 221-236. doi:10.1534/genetics.113.156935

Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., . . . Hendrickx, F. (2018). Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLOS Genetics, 14*(11), e1007796. doi:10.1371/journal.pgen.1007796

van der Valk, T., de Manuel, M., Marques-Bonet, T., & Guschanski, K. (2019). *Estimates of genetic load in small populations suggest extensive purging of deleterious alleles*. Retrieved from

van der Valk, T., Sandoval-Castellanos, E., Caillaud, D., Ngobobo, U., Binyinyi, E., Nishuli, R., . . . Guschanski, K. (2018). Significant loss of mitochondrial diversity within the last century due to extinction of peripheral populations in eastern gorillas. *Scientific Reports, 8*(1), 6551. doi:10.1038/s41598-018-24497-7

van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology & Evolution, 4*(8), 1004-1006. doi:10.1038/s41559-020-1204-8

Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology, 10*(1), 249-256. doi:10.1046/j.1365-294X.2001.01185.x

Wakeley, J. (1996). Distinguishing Migration from Isolation Using the Variance of Pairwise Differences. *Theoretical Population Biology, 49*(3), 369-386. doi:10.1006/tpbi.1996.0018

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. doi:10.1111/mec.14954

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

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. doi:10.1016/j.tree.2014.10.009

Willi, Y., Van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics, 37*(1), 433-458. doi:10.1146/annurev.ecolsys.37.091305.110145

Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T., & Christie, M. R. (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology, 27*(20), 4041-4051. doi:10.1111/mec.14726

Willoughby, J. R., Sundaram, M., Wijayawardena, B. K., Kimble, S. J. A., Ji, Y., Fernandez, N. B., . . . DeWoody, J. A. (2015). The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation, 191*, 495-503. doi:10.1016/j.biocon.2015.07.025

Wu, Y., Zhang, Y., Hou, Z., Fan, G., Pi, J., Sun, S., . . . Du, J. (2018). Population genomic data reveal genes related to important traits of quail. *GigaScience, 7*(5). doi:10.1093/gigascience/giy049

Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ 2 Test. *Supplement to the Journal of the Royal Statistical Society, 1*(2), 217. doi:10.2307/2983604

Zhan, X., Pan, S., Wang, J., Dixon, A., He, J., Muller, M. G., . . . Bruford, M. W. (2013). Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics, 45*(5), 563-566. doi:10.1038/ng.2588

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., . . . Froman, D. P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science, 346*(6215), 1311-1320. doi:10.1126/science.1251385