PATTERNS AND PROCESSES OF GENETIC DIVERSITY IN THE ENDANGERED SAN JOAQUIN KIT FOX

by

Tammy R Wilbert

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Environmental Science and Public Policy

Committee:

Dr. Patrick M. Gillevet, Dissertation Director

Dr. Jesus E. Maldonado, Research Advisor

Dr. Larry Rockwood, Committee Member

Dr. Suzanne Robbins, Committee Member

Dr. Albert Torzilli, Graduate Program Director

Dr. Robert Jonas, Department Chairperson

Dr. Richard Diecchio, Interim Assistant Dean, Student and Academic Affairs, College of Science

Dr. Peggy Agouris, Interim Dean, College of Science

Date: December 5, 2013

Fall 2013
George Mason University
Fairfax, VA

Patterns and Processes of Genetic Diversity in the Endangered San Joaquin Kit Fox

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Tammy R Wilbert
Bachelor of Science
University of Maryland Baltimore County, 2002

Director: Patrick M. Gillevet, Professor
Department of Environmental Science and Public Policy

Fall Semester 2013
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to all of the wild things that nurture my soul and make me happy and to the not-so-wild but amazing animals I have cared for.

ACKNOWLEDGEMENTS

I would like to thank the many people who have made this possible and even enjoyable! First and foremost, I really want to thank my mentors over the last 7 years. I can't even imagine going through this PhD without the guidance and friendship of Jesus E. Maldonado. He has been a wonderful mentor, leading me to be a better scientist by pushing me to take on challenging research and believing in my abilities every step of the way. He gave excellent advice, whether it was about the details of lab work or the implications of our research. Jesus has truly been an inspiration and unforgettable part of my life.

Patrick M. Gillevet was also crucial to my success. He regularly brainstormed with me on the bioinformatics of the MHC pyrosequencing, thinking about new strategies for the analysis, and leading me towards better research. With every difficult situation over the years, he always gave great perspective with a funny story about some scientist, ultimately giving me the advice I needed and telling me that I could do it.

Katherine Ralls has incredibly deep knowledge of the San Joaquin kit fox and has been very important in helping me understand the results and implications of this research. She has been a great mentor about the career of a research scientist and the various challenges along the way. It has been wonderful to work with Kathy, knowing she always tries to improve our research and advocates for me whenever necessary.

Elizabeth Archie was the first person to mentor me in a genetics lab, and she has forever changed the way I teach people. She is brilliant and thorough with any information she shares, but also patient and willing to help at any time. She showed me how important it is to be systematic and organized with lab work and data analysis. Ever since I worked with Beth, I have been motivated to ask dynamic research questions about wildlife disease, and I hope to collaborate with them again.

I would like to thank my committee members, Larry Rockwood and Suzanne Robbins, for helpful feedback throughout the research process. The PhD process can be quite grueling and it has been wonderful to have supportive committee members. I could not have completed any of this research without the many San Joaquin kit fox collaborators. Deborah (Smith) Woollett from Working Dogs for Conservation (WDC) provided over a thousand scat samples for me to process (wow!) as well as incredible guidance throughout our projects. Besides the amazing work by Debbie and her conservation detection dog-handler teams, I am truly grateful for our conversations about SJKF and life in science. I would also like to thank all of the people from WDC, including Alice Whitelaw, Aimee Hurt and Ngaio Richards, and a special thanks to Camas, Tsavo, Tia, Orbee and Seamus, the conservation dogs whose hard work made this research possible. I would like to thank Topaz Solar Farms, LLC, for logistical support

throughout the collection of scat samples by Working Dogs for Conservation. In addition, Daniel Meade and Jason Dart have been extremely helpful in understanding TSF and the kit foxes on site. Thank you as well for the beautiful mapping provided by Jacqueline Tilligkeit from Althouse and Meade, Inc. In addition, Michael Westphal from the Bureau of Land Management has been key to SJKF research in the Ciervo-Panoche area, and I feel lucky to be a part of such important research. His insights into the movements and population dynamics of kit foxes have been very helpful. Last but definitely not least, I am very thankful for my collaborators from the Endangered Species Recovery Program (ESRP) at California State University - Stanislaus – Brian Cypher, Christine Van Horn Job, and Samantha Bremner-Harrison. ESRP has been doing SJKF research for over 30 years, and the incredible sample collection in this dissertation is largely due to their hard work. In addition, they have helped me to understand the SJKF in a very fundamental way, taking me out to various sites, allowing me to help with trapping, and sharing the information they've gathered over studying kit foxes for many, many years. All of these kit fox researchers are unbelievable and I am honored to say that I've worked with them.

I am also very grateful to my collaborators in the lab, both at the Center for Conservation and Evolutionary Genetics (CCEG) and the Microbiome Analysis Center (MBAC). Christine Bozarth has been a great mentor in canid conservation and genetics and gave me a solid foundation for my research on kit foxes. Joseph Hoyt and Spencer Galen are two of the hardest working technicians, and both 'bird people' that I'd like to think I converted to loving kit foxes if even for a little while. I am really grateful for all of their work with the scat samples, especially when only Ace of Base could get us through the long stretches of Genemapper! Masoumeh Sikaroudi was incredibly helpful in helping me find my footing when I first started in genetics and has been a great collaborator ever since. She has extremely disciplined lab techniques that I admire, and I have always enjoyed learning from her. I am very appreciative of the time she has spent with me and the time she has contributed to this research.

CCEG has been an amazing influence in my life, with many incredibly smart and diverse people that have provided inspiration, guidance, and support over the years. In short, I would like to thank Robert Fleischer, Michael Schwartz, Lori Eggert, Nancy Rotzel, Andreanna Welch, Frank Hailer, Heather Lerner, Emily Latch, and Marissa Ahlering. These conservation geneticists have been friends, mentors, and role models, and I can't envision my degree without them. In addition, my fellow graduate students have played an important role throughout this process, and I have learned a great deal with and from them. I am particularly grateful to Jolanda Luksenburg, Trsihna Dutta, Sandeep Sharma, Katherine Bryant, Ryan Peters, Herlitz Davis, Melissa (Roberts) Hawkins, Courtney Hoffman, Lillian Parker, and Mirian Tsuchiya. I hope to have the opportunity to work some of the amazing people that I have connected with at CCEG in the future.

I want to acknowledge a few additional people who have provided me with opportunities and support over the last few years. Joseph Kolowski has given me great teaching opportunities at the Smithsonian Conservation Biology Institute, and I have deeply enjoyed teaching the course on Non-Invasive Genetic Techniques in Conservation

with Christine Bozarth and Jesus Maldonado. I appreciate the confidence and encouragement that all three of them have given me throughout the courses. Rajesh Ganesan gave me the wonderful opportunity to be a NSF GK-12 SUNRISE fellow and learn about teaching practices and STEM education in the US and abroad. I have a broader perspective of science and teaching and will use the information from my experiences with GK-12 throughout my career. In addition, Rajesh has been a great friend and given great advice about scientific careers, and I am very thankful for these experiences, personally and professionally. Next I would like to thank Stacy Lance for the experience at the 2011 NSF-AGA workshop on Evolutionary Genomics of Non-Model Species: Next-Gen Sequencing, Data Management, & Hypothesis Testing. This workshop greatly helped my understanding of next-generation sequencing and analysis. Additionally I would like to thank Stacy, Kenneth Jones, Maria Ruiz-Lopez and Jennifer Bollmer for great discussion and suggestions on MHC data analysis during and after the workshop.

I want to acknowledge a few people who encouraged me to go to graduate school, as they believed in my abilities as a scientist. Boripat Siriaroonrat and Astrith Rubiano are great friends who I worked with at the Conservation and Research Center (now SCBI) in Front Royal, VA, and told me I had what it takes to do this. Steve Monfort was my supervisor at the CRC and gave me great advice about my career and encouraged me to do a PhD. I am sincerely grateful to these people for their support.

This research, my degree, and the presentation of our results have been supported by numerous financial sources. I am very grateful for every one of them, including Althouse and Meade, Inc., US Department of the Interior – Bureau of Land Management, SI Pre-doctoral Fellowship, SI Next-Generation Sequencing Small Grant, GMU Provost Office – President's Fellowship, GMU ESP small research grant, GMU Graduate Student Travel Fund, GMU ESP Graduate Student Associate Travel Fund, American Society of Mammalogy Travel Fund, NSF GK-12 Sunrise Fellowship, and NSF-AGA Scholarship for Workshop: Evolutionary Genomics of Non-Model Species: 2011 Next-Gen Sequencing, Data Management, & Hypothesis Testing.

Last but definitely not least, I am extremely grateful to my family and to Vidit, the love of my life. Dad, you are my constant supporter, fan, and comfort, even if you are still trying to understand what I'm doing. I love you for this and I don't know how I would have made it through this PhD without you. Vidit, you believe in me more than I do myself. Thank you for always knowing how to push me to do my best and also knowing how to just be there when I have a rough day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS OR SYMBOLS

A	total number of variants
ABS	antigen-binding sites
A_i	number of variants per individual
Aime	*Ailuropoda melanoleuca* (giant panda)
AMOVA	analysis of molecular variance
A_R	allelic richness
BRSA	biological resource study area
BSA	bovine serum albumin
BYM	Bayesian algorithm with spatial autoregression (admixture)
C-P	Ciervo-Panoche core area
Calu	*Canis lupus* (grey wolf)
CAR	conditional autoregression (admixture)
CCEG	Center for Conservation and Evolutionary Genetics
CDV	canine distemper virus
CH	Ciervo Hills
CPNM	Carrizo Plain National Monument
CPV	canine parvovirus
CR-Car-Lok	region of Camp Roberts, Carrizo Plain, and Lokern
D	Tajima's D
DIC	deviance information criterion
DLA	*Canis lupus familiaris* (dog)
d_N	rate of nonsynonymous substitutions
DNA	deoxyriboneucleic acid
DNTP	deoxynucleotide triphosphates
DQA	α-chain of the DQ gene (MHC)
DQB	β-chain of the DQ gene (MHC)
DRB	β-chain of the DR gene (MHC)
d_S	rate of synonymous substitutions
ECM	equal capture probability model
emPCR	emulsion PCR
F_{ij}	frequency of each variant per individual (Nij/Ni)
F_{IS}	inbreeding coefficient
GD	genotypic distance
GPS	Global Positioning System
Gugu	*Gulo gulo* (wolverine)
H_E	expected heterozygosity

HMRF	Hidden Markov Random Field (no admixture)
H_O	observed heterozygosity
HWE	Hardy Weinberg equilibrium
i	individual
IBD	isolation by distance
Indel	insertion or deletion
j	variant
K	number of genetic clusters
ln(P|D)	likelihood of the model given the data
LnGGD	linearized geographic distance
LRM	mean Lynch and Ritland estimator
M	mutation-scaled migration rates
m	contemporary migration rates
m_v	potential number of variants (gene copy number*ploidy)
MAFFT	Multiple Alignment using Fast Fourier Transform
MBAC	Microbiome Analysis Center
MCMC	Markov chain Monte Carlo
MEGA	Molecular Evolutionary Genetics Analysis
$MgCl_2$	magnesium chloride
MHC	major histocompatability complex
mRNA	messenger RNA
mtDNA	mitochondrial DNA
MTPS	multi-tagged pyrosequencing
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
n	number of samples
N	total number of reads
Na	number of alleles
NaCl	sodium chloride
Nc	census population size
Ne	effective number of alleles
Ne	effective population size
N_i	number of sequences per individual
N_{ij}	number of sequences for each variant
N_j	number of sequences for each variant
NPRC	Naval Petroleum Reserves in California
Pati	*Panthera tigris* (tiger)
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PEG	polyethylene glycol
P_{ID}	probability of identity
PV	Panoche Valley
QGM	mean Queller and Goodnight estimator
qPCR	quantitative PCR
r	autocorrelation coefficient

RNA	ribonucleic acid
SAA	spatial autocorrelation analysis
SESAME	SEquence Sorter & AMplicon Explorer
SJKF	San Joaquin kit fox, *Vulpes macrotis mutica*
SJT	Little San Juan Ranch and Thorup properties
SJV	San Joaquin Valley
SNP	single nucleotide polymorphism
SQL	Structured Query Language
S_T	total number of amino acid sites
S_V	number of variable amino acid sites
T_1	1st sequence number threshold – per individual
T_2	2nd sequence number threshold – per variant
TIRM	two innate rates model
TSF	Topaz Solar Farms
Uram	*Ursus americanus* (American black bear)
Urth	*Ursus thibetanus* (Asian black bear)
Vula	*Vulpes lagopus* (Arctic fox)
Vuma	*Vulpes macrotis mutica* (San Joaquin kit fox)
Z	Z-statistic
ZF	zinc finger
Θ	mutation-scaled effective population size
μ	mutation rate
$\chi 2$	Chi-squared statistic

ABSTRACT

PATTERNS AND PROCESSES OF GENETIC DIVERSITY IN THE ENDANGERED SAN JOAQUIN KIT FOX

Tammy R Wilbert, Ph.D.

George Mason University, 2013

Dissertation Director: Dr. Patrick M. Gillevet

Survival of the endangered San Joaquin kit fox (SJKF; *Vulpes macrotis mutica*) is challenged by reduced and fragmented habitat resulting from anthropogenic land uses in the Central Valley of California. Because the SJKF is adapted to arid landscapes and utilizes large home ranges to prey on kangaroo rats, genetic techniques are very useful to study these elusive animals. We used annual noninvasive genetic surveys from 2009 to 2011 to document the kit foxes on the Topaz Solar Farms (TSF) in San Luis Obispo Co., California, prior to the construction of the solar facility. We identified 45 individuals from 351 fresh scat samples and found kit foxes predominantly used the annual grasslands and rarely used agricultural lands. Second, we conducted 3 years of systematic transect surveys for scat throughout the Ciervo-Panoche area. We identified 93 kit foxes, with half of the individuals carrying a mtDNA haplotype with a 16bp deletion that had not been previously identified. Next, we used 795 scat and tissue

samples collected over a period of 20 years to ascertain landscape level patterns of connectivity and population dynamics of kit foxes in the San Joaquin Valley. We found 3 geographically and genetically distinct groups that are highly disconnected and have undergone a 76% reduction from their estimated historic population size. Finally, we used pyrosequencing to characterize three major histocompatibility (MHC) genes that have been shown to play a role in immunity. We selected a subset of individuals from two relatively stable natural environments and compared their patterns of MHC variation to kit foxes adapted to the urban environment in Bakersfield, CA. Every gene (DQA, DQB, and DRB) was polygenic with varying degrees of polymorphism, although DRB alleles translated into a greater number of unique antigen binding sites, and thus greater immune system diversity. Urban and wild populations had similar allelic diversity, and each carried unique alleles. Each of these studies resulted from long-term collaborative efforts from a team of researchers that are working closely with federal, state and private agencies responsible for the protection of the endangered SJKF. The results of our research on patterns of genetic diversity and connectivity are already playing an important role in the design of SJKF conservation and management plans.

CHAPTER ONE: INTRODUCTION

Endangered species garner the attention of many people due to their inherent status of being low in number and facing possible extinction. Conservation biologists study them to learn more about what brought them to this point, what they need to survive, and how to better protect them. Geneticists learn different types of information from the DNA of endangered species to address questions as broad as phylogenetics and historical distributions or as narrow as mating pairs and individual dispersal distances. Ecologists spend a great deal of time in the field studying endangered species to learn about their behaviors and interactions with the environment. Land managers are keenly aware of endangered species on their property, whether it is public or private land, and the policies that they must follow for short and long-term management. The individual interests of all of these groups can initiate many scientific questions worth studying. More dynamic research can be initiated when people from these different perspectives pool their knowledge and resources.

Conservation geneticists have many tools they can use to address questions about endangered species. Non-coding regions of the genome are free to allow mutations to accumulate and provide information on relatedness from the individual level all the way out to evolutionary relationships and histories. Microsatellite loci are highly repetitive tandem repeats typically found in nuclear DNA with the fastest rate of change due to

amplification errors that are passed through generations. Thus, a panel of polymorphic microsatellite loci can be used to identify the number of individuals in an area, determine paternity and family groups, or measure levels of connectivity between groups. A single nucleotide polymorphism (SNPs) is the most basic difference between two sequences, with each SNP being one nucleotide substitution. While these can be easy to detect with sequencing, many SNPs may be needed in order to provide enough statistical power for the analysis at hand. However, if SNPs are within a protein-coding region, even one SNP could be significant because it could translate into an amino acid substitution and thus the creation of a different protein. Depending on the function of a gene, these coding parts of the genome can be evolving neutrally or under different types of selection, such as negative selection to maintain their functions (e.g. cell regulating genes) or positive selection to generate different alleles (e.g. immune system genes in the major histocompatibility complex). Additionally, the introns of genes are not under selection and can be examined for mutations that accumulate over time and may help to explain evolutionary relationships. Thus we can utilize different parts of the genome to address different questions in endangered species research.

This dissertation includes four highly collaborative projects involving the endangered San Joaquin kit fox (SJKF; *Vulpes macrotis mutica*), followed by a conclusion. Each project has been written in preparation for submission to peer-reviewed scientific journals. The first project (Chapter 2) focuses on kit foxes on the Topaz Solar Farm (TSF) project in the northern end of the Carrizo Plains. This project was initiated by the TSF project through the environmental management group, Althouse and Meade,

Inc. It has involved incorporation of information from project managers, conservation detection dog-handler teams, and experts in SJKF biology to inform the genetic analyses presented here. This work represents three years of data collection before solar panel installation, and data collection will continue for another 4-5 years. Chapter 3 focuses on the status of San Joaquin kit foxes in the Ciervo-Panoche core area. This is the largest area of suitable habitat in the northern and central portions of their historic range. This is the first extensive study of the area and cannot come at a more crucial time. A large portion of the Panoche Valley was just leased to a solar company (June 2013) but environmental planning and management is still in progress. While the manuscript is purposely written without any prejudiced perspective, this scientific information will be very important for conservation management in the area. Chapter 4 considers the historical and current connectivity of kit foxes across the entire San Joaquin Valley (SJV). Samples from over 20 years of surveys allowed for a metapopulation level analysis of the SJKF, looking at what processes may have generated the patterns of genetic diversity we find in kit foxes today. In addition we are able to estimate historic population size and changes in migration rates between genetically distinct regions of kit foxes. Chapter 5 changes our focus from patterns detected in non-coding DNA to those found in the major histocompatibility complex (MHC), which codes for immune function. By looking for genetic diversity in the MHC, we can look for differences that may be due natural selection. One population of SJKF has adapted to a city, Bakersfield, which has different disease threats due to urban disease vectors (feral dogs, feral cats, raccoons, and skunks). Recent advances in genetic techniques over the last 5 years made

this research possible, such that we could amplify MHC loci that are highly polymorphic and polygenic (multiple gene copies). This manuscript characterizes the methods and diversity we found for individuals in both urban and natural landscapes. Finally, Chapter 6 brings all of this information together to think about current conservation issues for this endangered species and the next steps for research on San Joaquin kit foxes.

CHAPTER TWO: NON-INVASIVE GENETIC MONITORING OF ENDANGERED SAN JOAQUIN KIT FOXES ON A FUTURE SOLAR FARM

A manuscript to be submitted to *Endangered Species Research*

Collaborators: Deborah A. (Smith) Woollett[2], Alice Whitelaw[2], Jason Dart[3], Joseph R. Hoyt[1], Spencer Galen[1], Katherine Ralls[1], Daniel E Meade[3], and Jesus E. Maldonado[1]

[1]Smithsonian Conservation Biology Institute

[2]Working Dogs for Conservation

[3]Althouse and Meade, Inc.

Introduction

Increasing the production of solar energy will have long-term benefits for the environment including the reduction of greenhouse gas emissions. However, large-scale solar facilities require many acres of relatively inexpensive land in areas where solar energy potential is high. Such lands are usually located far from human population centers in areas where threatened and endangered species occur, which is creating a conflict between solar energy development and the preservation of biodiversity throughout the arid southwestern United States (Cameron *et al.* 2012; Lovich& Ennen 2011; Stoms *et al.* 2013).

The Topaz Solar Farms (TSF) project in California is one of the world's largest solar farms under construction. It is located in the northern end of the Carrizo Plain, which is separated from the San Joaquin Valley by the Temblor Range. The Carrizo Plain is an important area for species adapted to the arid scrub- and grassland habitats of the San Joaquin Valley, because the majority of these habitat types are now extremely rare within the Valley itself due to the conversion of land to agricultural production and urban areas (Kelly *et al.* 2005). The federally endangered and state threatened San Joaquin kit fox (*Vulpes macrotis mutica*; SJKF) is one of 13 endangered species that occur in the Carrizo Plain and requires the largest amount of habitat. By protecting the habitat required by the SJKF, it is likely that habitat required by some of the other species in the ecosystem will also be preserved (U.S. Fish and Wildlife Service 1998).

In order to reduce impacts on endangered species in the area (Cameron *et al.* 2012; Lovich& Ennen 2011; Stoms *et al.* 2013), TSF will be constructed on a large proportion of degraded habitat in the form of agricultural lands. A larger study area with potential TSF project sites was identified to investigate variation across its four predominant habitat types: active cropland, fallowed cropland, grazing land dominated by non-grass species, and annual grassland largely composed of introduced grass species. The variation in habitat quality provided an opportunity to learn about patterns of habitat utilization by SJKF on the site before, during, and after construction of the solar facility. Annual grassland, if not excessively grazed, provides good habitat for SJKF (Germano *et al.* 2012; U.S. Fish and Wildlife Service 1998). However, SJKF have limited ability to use agricultural lands, due to low prey availability and lack of den sites (Warrick *et al.*

2007). Although SJKF may range into agricultural lands at night (Warrick *et al.* 2007), they typically travel on natural or man-made paths and den on nearby less-disturbed, more natural lands. Farming activities on the TSF site ended in 2011, so it is expected that these former farmlands will gradually revert to better SJKF habitat.

We monitored SJKF foxes on the TSF study area prior to construction of the solar facility annually from 2009 to 2011. We used well-established methods for conducting non-invasive genetic surveys (Bozarth *et al.* 2011; De Barba *et al.* 2010; Dutta *et al.* 2013; Schwartz *et al.* 2010) that were developed specifically for SJKF (Ralls *et al.* 2010; Smith *et al.* 2006b; Smith *et al.* 2003). In brief, conservation detection dog-handler teams search transects for fresh SJKF scats and DNA from the scats is then analyzed using molecular genetic methods to identify species, sex, and unique genotypes of individuals (Bozarth *et al.* 2010; Ortega *et al.* 2004; Smith *et al.* 2006a). The primary objectives of our study were to document the presence of SJKF on the site and nearby properties, estimate the number of individuals present and the extent to which they used different habitats, and compare levels of genetic diversity in the foxes on the site to that found in foxes in the Carrizo Plain National Monument to the south. We used our results to suggest placement of solar panels on the TSF site and establish a pre-construction baseline for SJKF. We will conduct similar surveys in future years to document the combined effects of the construction of the solar farm and changing habitat conditions on SJKF use of the site.

Study Area

The Topaz Solar Farm project is located approximately 100 km west of Bakersfield, in the northern end of the Carrizo Valley, San Luis Obispo County, California (Figure 2-1). It is northwest of the Carrizo Plains National Monument, which contains over 150,000 acres of scrub and grassland habitat suitable for kit fox and the largest remaining kit fox population (2010). A project Biological Resource Study Area (BRSA) of ±9,700 acres was initially identified as a potential site for investigation that included habitats ranging from good to poor in quality (Figure 2-4). The northwest portions contained poor quality habitat that was farmed for about 70 years. Cropland was plowed yearly and grazed by cattle after harvest, leading to low numbers of small mammals (Collins 2010) and few underground SJKF dens (*pers*. *com*. Althouse and Meade, Inc.). The southern and eastern portions have relatively better habitat than the northern section, but contain a mixture of habitat types ranging from recently fallowed cropland to annual grassland (Figure 2-4). Cattle grazed these sections either once a year or once every 3 years, producing varied habitat suitability for kit foxes and their prey. The final footprint of the TSF project is a subset of the BRSA with a maximum size of ±3,510 acres. Grassland vegetation grows in the alleys and under the solar panels, and the remaining area outside of the solar farms will gradually revert to annual grassland.

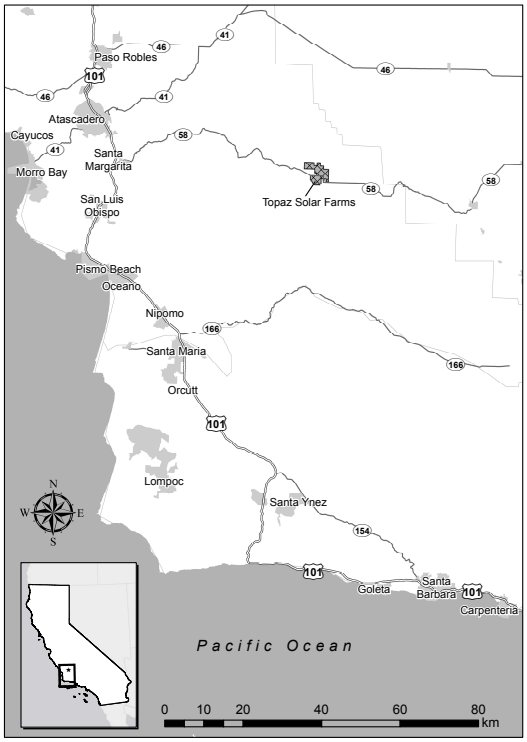


Figure 2-1. Location Map. The survey area is approximately 9,700 acres in size, located in the Carrizo Plain, San Luis Obispo County. The Topaz Solar Farms will be built on 3,510 acres.

Methods

Sampling

Transects. - During 18 August 2009 to 01 September 2009, and 11 November 2009 to 21 November 2009, search routes were established along 18 transects (approximately 103.31 km) running throughout the entire BRSA (Figure 2-2). We designed transects to bisect hypothetical kit fox home ranges multiple times using previous estimates of an average home range size of approximately 4-11 km^2 for kit foxes in the Carrizo Plain and similar habitats (Spiegel& Bradbury 1992; White& Ralls 1993; Zoellick *et al.* 2002); Cypher *pers. comm.*). Transect routes utilized unpaved roads,

trails, fence-lines, and vegetation in the study area. During 02 November 2010 to 17 November 2010, and 01 November 2011 to 16 November 2011, search routes were slightly adjusted after a final project footprint was further refined. A total of 17 transects (approximately 108.04 km) were surveyed throughout the newer study area boundary (Figure 2-2). We also surveyed two private parcels east of the Topaz study area, the Little San Juan Ranch and Thorup properties (Figure 2-2), for SJKF in November 2010.

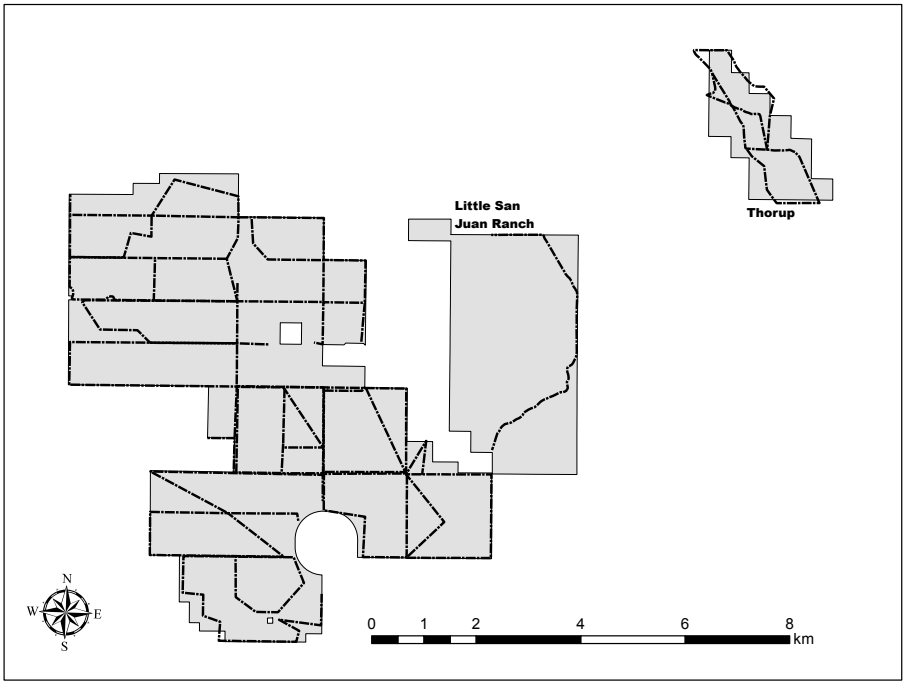


Figure 2-2. Study area with survey transects shown in dashed black lines.

Confirmation of presence of SJKF on these properties would confirm that they could be purchased and set aside as protected lands to partially offset the impacts of developing the solar farm. All surveys from 2009 to 2011 included an initial and a repeat

survey session that were one week apart. To keep track of each sampling group and carry out appropriate comparisons, we separated the data into sampling groups: Aug 2009 TSF, Nov 2009 TSF, and 2009 All; Nov 2010 TSF, Nov 2010 surveys from the additional Little San Juan Ranch and Thorup properties ("Nov 2010 SJT"), and 2010 All; Nov 2011; all 3 years during November on TSF ("3yrs Nov TSF"); and All Combined (all times, all locations).

Scat samples. - Professional conservation detection dog-handler teams (Working Dogs for Conservation, Three Forks, MT) searched for kit fox scats. Only fresh scats were collected for DNA analysis as determined by a freshness rating method based on their physical characteristics (Smith *et al.* 2003). Scats determined to be more than 8 days old were not collected but locations for all scats detected by the dog-handler teams were geo-referenced and recorded using Global Positioning System (GPS) units (Garmin GPS III+). Fresh scats were stored in plastic bags with silica gel for desiccation (Fisher Scientific, Pittsburgh, PA) and shipped to the Center for Conservation and Evolutionary Genetics laboratory.

Mapping. - X-Y coordinates of each transect and the GPS location of each scat were entered into ArcGIS (ESRI Geographical Information System), and plotted over an aerial photograph with project boundary lines. Scat samples were also mapped with reference to genetic results (i.e. by species or individual identification).

Molecular Methods

DNA Extraction. - DNA was extracted using the QIAamp DNA stool mini kit (QIAGEN®) using modifications from the manufacturer's protocol as in Eggert et al.

(2005) and an extended overnight incubation in lysis buffer and proteinase K at 56°C on a shaker. Extractions were carried out in a separate room dedicated to DNA extractions to prevent contamination. Negative controls (no scat) accompanied each set of extractions and were used to check for contamination.

Species Identification. – To confirm species, we conducted fast and reliable molecular identification of every scat sample following previously established protocols (Bozarth *et al.* 2010). This protocol requires the amplification of a short fragment of the mtDNA control region that is a different length in each of the potential canid species in the study area. PCR reactions were set up as follows: a 20µL total volume consisting of 9.3µL of PCR water, 2.0µL of 10x PCR buffer (No $MgCl_2$), 1.5µL of 10µM DNTP (2.5µM each), 1.0µL of primer KFSPID-F and 1.0µL of primer KFSPID-R, 2.0µL of $MgCl_2$ (25mM), 0.2µL of AmpliTaq Gold, and 3.0µL of substrate DNA. Reactions were denatured at 95°C for 10 minutes; then 35 cycles of 95°C denaturing for 30 sec, 58°C annealing for 30 seconds, and a 72°C extension for 2 minutes; then a final extension of 72°C for 30 minutes; and stored at 4°C. Samples with poor amplification were diluted (1:15 up to 1:45) to minimize interference from PCR inhibitors in scat samples, and replicated as needed.

Molecular Sexing – We determined the sex of the animal that deposited each scat using protocols developed in our laboratory with primers that have good specificity for canids, minimize the chances of errors due to prey item contamination, and increase amplification success (Ortega *et al.* 2004; Ralls *et al.* 2010). We amplified a small fragment (195 bp) of the zinc finger (ZF) protein gene, found in both X- and Y-

chromosomes, and digested the PCR product with a Taq I restriction enzyme yielding a clear pair of fragments for males and a single uncut fragment for females (Ortega *et al.* 2004; Ralls *et al.* 2010).

Genotyping and Identification of Individual kit foxes – After samples were positively identified as kit foxes, we genotyped them for individual identification using six microsatellite tetranucleotide repeat loci that were developed from domestic dogs (Francisco *et al.* 1996) and proven to reliably work for individual identification of kit foxes in our lab (FH2137, FH2140, FH2226, FH2535, FH2561, Pez19; (Smith *et al.* 2006a). We assessed our ability to differentiate individuals by estimating the probability of identity (P_{ID}) (i.e. the probability of different individuals sharing an identical genotype at random) and the P_{ID} between siblings (Mills *et al.* 2000; Waits *et al.* 2001).

For each DNA extract, we performed at least 5 independent PCR amplifications of each locus for homozygous individuals to verify allele size and detect allelic drop out. We ran heterozygotes a minimum of three times to confirm both alleles. We amplified microsatellites in 10 µl volumes using: 4.35µL of PCR water, 1.0µL of 10x PCR buffer, 1.0µL of 10µM DNTP (2.5µM each), 0.25µL of forward primer and 0.25µL of reverse primer, 1.0µL of $MgCl_2$ (25mM), 0.15µL of AmpliTaq Gold, and 1.25µL of substrate DNA. The PCR conditions for scat extracts, as well as extract and PCR negative controls, included an initial hot start, 35 cycles of the following profile: 94° C for 1 min, 58° C for 1 min, and 72° for 1.5 min, and a final extension at 72°C for 30 minutes. We used up to 45 additional cycles to re-amplify samples with poor amplification success, usually from low starting DNA concentrations or dilutions to avoid PCR inhibitors found in scat.

We used fluorescently labeled forward primers (TET, HEX or FAM) in all of the PCR reactions (for species ID, sex ID, and microsatellite loci). We combined PCR product (1.0-2.5uL) with 9.0μL of a 5:100 mix of Gene Scan ROX-500 (Applied Biosystems) and Hi-Di Formamide (Applied Biosystems) to visualize our fragment sizes on an ABI PRISM* 3130 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA), which allowed for a plate of the 384 PCR reactions to be loaded at once. We analyzed samples using Genemapper® software to determine the size of each fragment.

We used the Excel Microsatellite Toolkit (Park 2001) to compare genotypes and defined individuals by unique genotypes and samples with matching alleles at all loci. We checked genotypes that differed at only 1 or 2 loci for accuracy of genotype and data entry. We also compared genotypes between samples collected in 2009, 2010, and 2011 to determine if any individuals had been recaptured between survey sessions. We assigned recaptured individuals the same number previously used. In addition, we used sex to differentiate between closely related individuals that shared microsatellite genotypes.

Analysis

Genetic variability. – To obtain a more robust estimate of genetic diversity, we genotyped individuals at 5 additional microsatellite loci (AHTh171, FH2054, FH2328, FH2848, and Ren162). Since we had previously determined that we had enough power to distinguish individuals using the original 6 microsatellite loci (P_{ID} =2.03 10^{-6}; see below), we selected 1 representative scat sample that amplified most reliably for each individual and genotyped these samples for additional loci using the same protocols and

conditions described above. A summary of the workflow for scat sample processing is shown in Figure 2-3. We then compared genetic variability of SJKF in the study area to that of kit foxes in the Carrizo Plain National Monument (CPNM) using tissue samples collected during two previous studies: 32 kit foxes trapped near Soda Lake Rd in 1988-

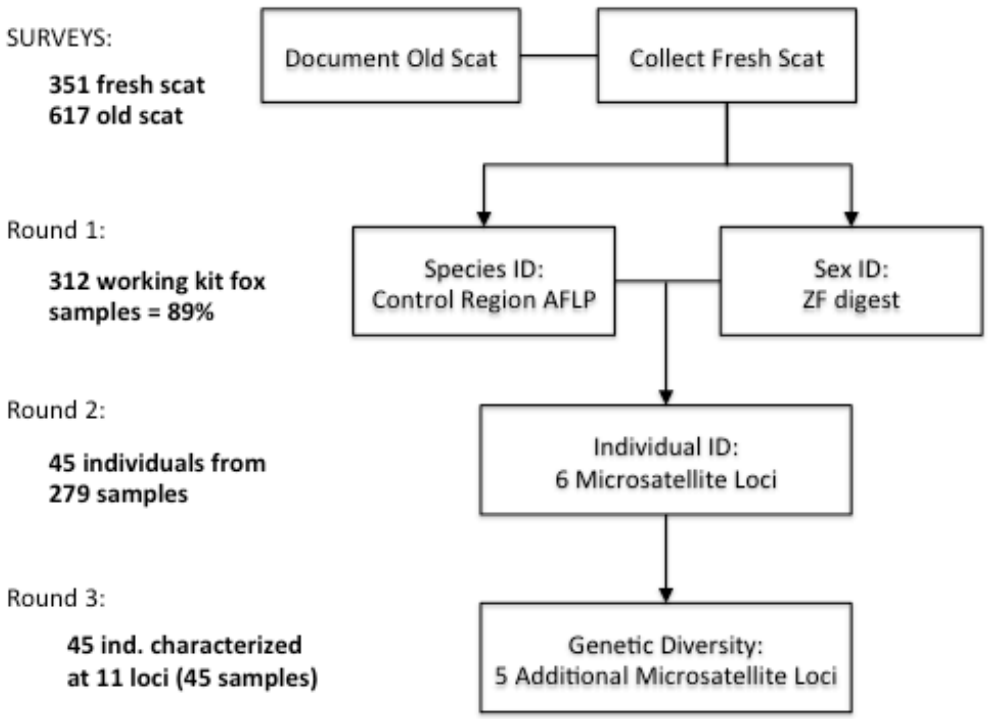


Figure 2-3. Schematic of workflow. Numbers represent totals for all 3 years of the study.

1991 (White & Ralls 1993) and 29 individuals trapped along Elkhorn Rd in 1998 (Bean 2002). We genotyped samples for the 11 microsatellites in the same manner as described for the TSF samples, except that, because tissue extractions yielded high quality/quantity

DNA, fewer PCR replicates (2-3) per sample were required to obtain reliable genotypes. We tested all three groups (TSF, Soda Lake Rd, and Elkhorn Rd) for deviation from Hardy-Weinberg expectations and for linkage disequilibrium between loci using GenePop (Raymond& Rousset 1995). We also used GenePop to determine allelic diversity and expected and observed heterozygosity values at each locus.

Demographic parameters and Capwire estimates. - To obtain estimates of the number of SJFK using the study area, we analyzed samples collected during the annual November collection periods using the program *Capwire* (Miller *et al.* 2005). *Capwire* performs as well if not better than other methods such as the Chao or jackknife estimators, or the Eggert's rarefaction estimator as calculated in GIMLET (Miller *et al.* 2005). *Capwire* uses the number of samples per individual (as identified by genetics) as an estimate of the number of times we captured an individual and then infers probabilities of detection in the population. *Capwire* does urn simulations using 2 models, the equal capture probability model (ECM) and the two innate rates model (TIRM), to determine which model best fits the data. The appropriate model is chosen using a likelihood-ratio test, and TIRM is used when capture rate is heterogeneous between individuals.

Results

We collected 351 fresh scat samples over the three years (Table 2-1, Figure 2-4). The dogs also identified 617 older scats during the same time period. Using our mtDNA protocol, we were able to determine the species for 312 (89%) of the fresh scats. Of those 312 scats, 279 (89.4%) were San Joaquin kit fox, 27 (8.7%) were red fox (*Vulpes vulpes*), and 6 (1.9%) were coyote (*Canis latrans*; Table 2-1). All but 1 kit fox sample

had the 252bp mtDNA haplotype; scat DT1014 had the shorter 236bp kit fox haplotype first reported in Bozarth et al. (2010). We identified kit fox scats predominantly on the southern and eastern portions of the TSF (Figure 2-4).

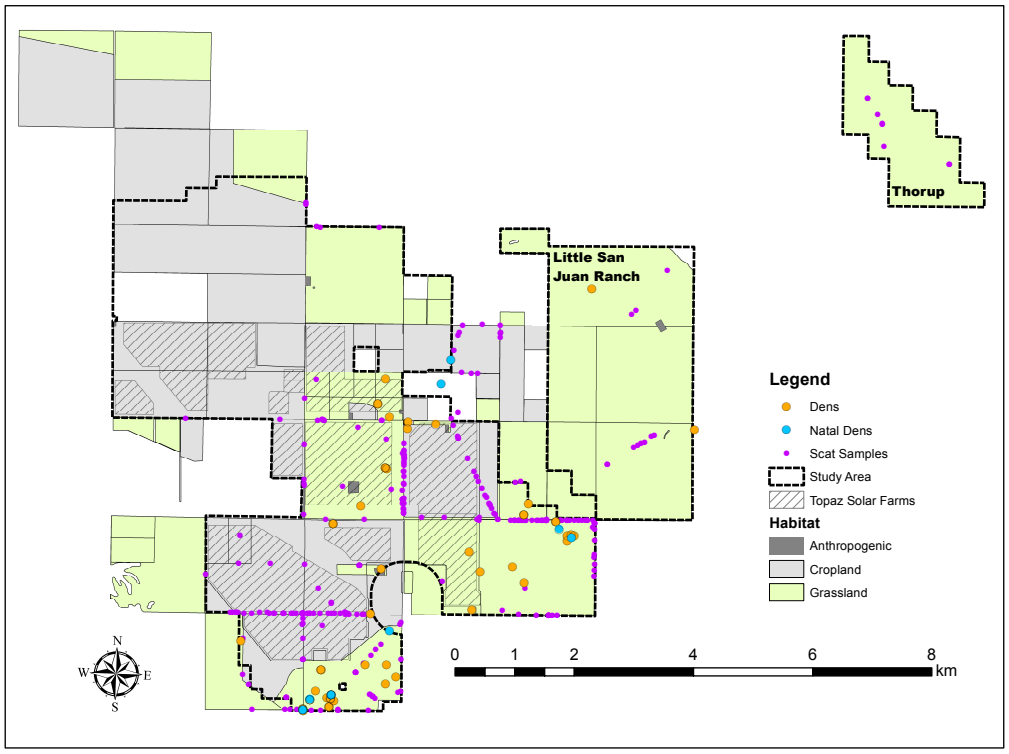


Figure 2-4. Location of the 279 San Joaquin kit fox scat samples identified by year, out of 351 fresh scat samples collected from 2009 – 2011, shown with habitat types on the Topaz Solar Farms survey area and additional properties. In addition all dens (active and natal) were identified on the map.

Genetic variation – The probability of a random match between unrelated individuals for all multilocus genotypes was 9.8 x 10^{-6} (P_{ID} unbiased), and the probability of a random match between siblings for all multilocus genotypes was 9.7 x 10^{-3} (P_{ID} sibs). These low P_{ID} values indicated that our 6 microsatellites were adequate to differentiate between

individual foxes, including relatives. We assigned 258 of the 279 kit fox samples to an individual genotype (Table 2-1); the remaining samples could not be assigned due to incomplete genotypes. The scat sample with the 236bp mtDNA haplotype also had a unique microsatellite genotype. Sex was identified for all but 1 individual (TZ33). Our microsatellite genotyping confirmed the presence of 45 individuals over the 3 years, including 18 males, 26 females, and 1 individual of unknown sex, for an overall 0.7:1 sex

Table 2-1. Species, sex, and individual identification of scat samples from all surveys. The numbers of individuals recaptured are listed after the colon. The number of samples completely genotyped and assigned to an individual is listed in parenthesis. *not identified to individual
Example: # detected: # recaptured (# associated scats)

Survey		Scat Samples	Total KF IDs	San Joaquin Kit Foxes		Non-Targets	
Time	Location			Female	Male	Red Fox	Coyote
Aug 2009	TSF	91	72	7 (37)	9 (35)	1 (1)	0
Nov 2009	TSF	62	48	3:1 (12)	8:7 (36)	1:1 (2)	0
Nov 2010	TSF	69	49	4:1 (22)	3:1 (17)	2 (16)	1 (3)
Nov 2010	SJT	16	16	3 (7)	3:1 (9)	0	0
Nov 2011	TSF	113	94	10:1 (49)	6 (33)	(8*)	(3*)
				+1 ind. of undetermined sex in 2011			

ratio of males to females. However sex ratios in the TSF area significantly fluctuated over the years (1:0.6 in 2009, 0.9:1 in 2010, 0.6:1 in 2011; Fisher's exact test, p=0.046). Number of scat samples per individual varied from 1-18, with an average of 5.3 samples for each individual detected in a given year. Scats from the same individual were found

in close proximity to each other, on the same or adjacent transects, and generally < 3.5 km apart. Three individuals detected in 2009 were found in 2010, one of which was found on the TSF property in 2009 and on the Little San Juan Ranch property in 2010. One female (TZ21) was found in the same area of the TSF (southeastern corner) in 2 years and was the only individual identified in 2010 that was found again in 2011. Active and natal dens were also identified in the same area during both of these years.

The allelic diversity in the 11 microsatellite loci screened in individuals from the TSF ranged from 3 - 10 alleles per locus, with a mean number of 6.55 alleles per locus, which was very similar to the results for the Soda Lake and Elkhorn samples (Table 2-2). The most polymorphic locus was FH2137 with 10 alleles, and the least polymorphic was Pez19 with only 3 alleles (Supp. Table 3). The individuals at TSF also carried the highest number of private alleles (7 for Soda Lake, 5 for Elkhorn, and 12 for TSF). We also graphed the distribution of allele frequencies from all loci using methods described in Luikart et al. (1998; allele frequency classes from 0-0.1, 0.1-0.2, etc.) and found the expected L-shaped curve for all three locations (Supp. Figure 4). Analysis in GENEPOP revealed that all loci were under Hardy-Weinberg Equilibrium and none showed evidence of linkage disequilibrium. Allelic diversity and heterozygosity were not significantly different between individuals found on the TSF, Soda Lake Rd, or Elkhorn Rd (Table 2-2, Figure 2-5). The observed heterozygosities were also not significantly different than the expected values, nor did they differ significantly between populations or survey years on the TSF.

Table 2-2. Genetic diversity statistics for kit fox detected on the Topaz Solar Farm, Soda Lake Road, and Elkhorn Road, in the Carrizo Plain National Monument, California. n = number of individuals; H_E = expected heterozygosity; H_O = observed heterozygosity; N_A = average number of alleles; N_{A5} = average number of alleles ≥ 5%; N_E = average number of effective alleles; N_P = average number of private alleles. SE is one standard error from the mean.

Location	*n*	H_E (SE)	H_O (SE)	N_A (SE)	N_{A5}	N_E	N_P
Soda Lake	32	0.649 (0.052)	0.647 (0.026)	6.46 (3.17)	4.091	3.510	0.636
Elkhorn	29	0.674 (0.046)	0.684 (0.027)	6.55 (3.33)	4.364	3.735	0.455
Topaz	45	0.670 (0.040)	0.627 (0.022)	6.55 (2.21)	3.909	3.364	1.091
Overall	106	0.686 (0.044)	0.649 (0.014)	8.73 (4.00)	4.088	3.537	0.780

Estimates of kit fox abundance – We used samples from the TSF area that were assigned to an individual kit fox during November 2009 (n=48), November 2010 (n=39), and November 2011 (n=83) to estimate kit fox abundance on the TSF. This totaled 170 samples collected over 3 years that led to a detection of 32 individuals with unique genotypes, including 16 males, 15 females, and 1 individual of unknown sex. We used the program *Capwire* to estimate the number of kit fox on the TSF for each November survey (Figure 2-6). *Capwire* chose the two innate rates model for each estimate,

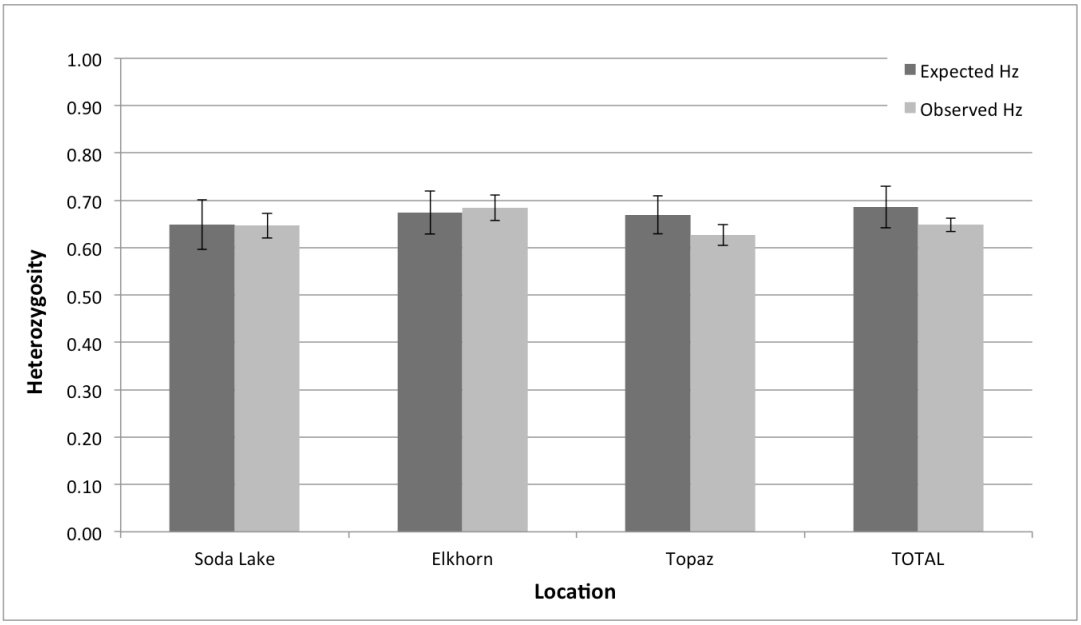


Figure 2-5. Expected and observed heterozygosity values (mean +/- standard deviation) for the Topaz Solar Farm in comparison to the Soda Lake and Elkhorn populations and overall diversity detected in the Carrizo Plain.

indicating heterogeneity in capture probability of individuals. *Capwire* estimated 11±0 individuals for Nov. 2009, 7±0 ind. for Nov. 2010, and 17+1 ind. for Nov. 2011. The number of individuals identified based on unique genotypes was similar or identical to the *Capwire* estimates (Figure 2-6). Combining all surveys conducted on TSF in 2009 and all surveys conducted in 2010 yielded higher numbers of individuals for each year than the November surveys alone (Figure 2-6). Combining the November TSF samples for all three years and all samples collected on TSF during the study produced Capwire estimates of 33±1 and 46±2 individuals, respectively (Figure 2-6). The latter two combined estimates are not estimates of individuals using the TSF site at any point in time, but represent an estimate of the total number of individuals that were exposed to sampling

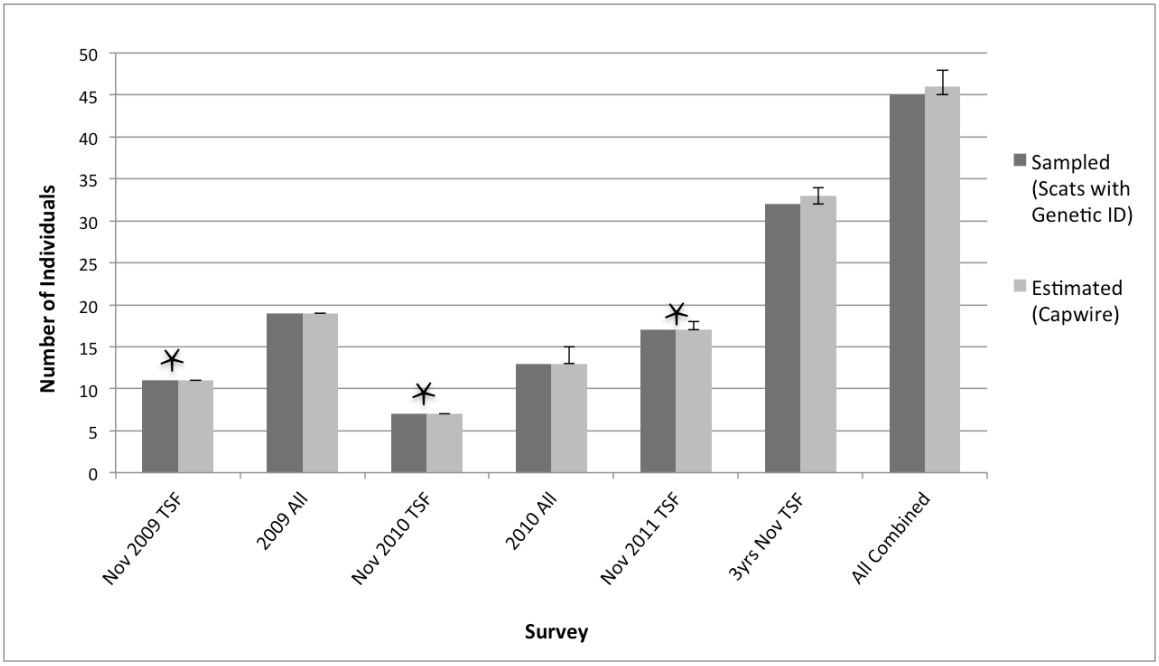


Figure 2-6. Minimum number of kit fox individuals throughout the surveys. Dark grey bars are the number of individual genotypes from scat surveys and genetic analysis. Light grey bars are the corresponding *Capwire* estimates (error bars represent the *Capwire* minimum and maximum values). Time points with an asterisk are numbers of individuals from each November survey on the Topaz Solar Farm.

during the sampling periods that were combined to produce the estimates, a measure that is called a super-population estimate (Schwarz & Arnason 1996). It is important to view these superpopulation estimates as likely referring to all individuals using the landscape rather than a population restricted to the TSF core area (Boulanger & McLellan 2001).

Discussion

Molecular analysis of scat samples was highly successful as compared to other studies (Kohn *et al.* 1999; Taberlet *et al.* 1997; Woods *et al.* 1999; 88.9% species identification and 92.5% genotype assignment). We attributed our success in PCR

amplification of problematic scat samples, particularly those that had a large proportion of insect exoskeleton remains, to minor modifications in our PCR protocols such as diluting samples and increasing the number of PCR cycles. In addition, we looked at sample locations for matching genotypes as described in Smith et al. (2006a) to increase our confidence in assigning samples to individuals. We found an average of 5.3 scat samples per individual per year overall, and samples from individuals were clumped throughout the survey area (see Supp. Figures 1, 2, and 3). We collected a similar number of scat samples from males and females over all three years (Table 2-2), indicating that both sexes were equally detectable at this time of year, in contrast to the drastic reduction in female scat samples found during reproductive denning in the spring (Ralls *et al.* 2010). We detected 45 individuals, 40 individuals that used the TSF survey area at some time during the 3 years and 6 that used the Little San Juan Ranch and Thorup properties (1 individual used TSF and Little San Juan Ranch).

Our expanded analysis with 11 microsatellite loci showed that the genetic characteristics of the foxes in the TSF area are similar to those of foxes in the Carrizo Plain National Monument to the south. The three localities had similar levels of heterozygosity and did not deviate from Hardy Weinberg Equilibrium or show signatures of linkage disequilibrium. Furthermore, levels of allelic richness did not differ significantly among sampling sites. Thus, there appear to be no major barriers to kit fox dispersal between the CPNM and the TSF area.

However, we found a surprisingly high number of private alleles at each of the localities, which suggests that there is some degree of population substructure, even

though the TSF area and the two localities sampled in the CPNM are only 15-30 km apart, and are within the dispersal distance capabilities of a kit fox. Distributions of dispersal distances in vertebrates are typically skewed towards short distances with an extended tail of longer distances (Koenig *et al.* 1996). Kit foxes follow this pattern as Scriver et al. (1987) found that dispersing kit foxes moved a median of 4.5 km and an average of 8 km, although some moved long distances and one moved more than 120 km. The genetic substructure we found may reflect the fact that the average dispersing fox moves only a short distance, so that same-sex foxes that live on adjacent home ranges tend to be more closely related than same-sex foxes that did not live on adjacent home ranges, even though mated pairs are unrelated (Ralls *et al.* 2001).

We detected one individual (male TZ15 – scat DT1014) with a unique mitochondrial haplotype (236bp) and a unique microsatellite genotype. This shorter length mtDNA haplotype was recently discovered in the Ciervo-Panoche area and had not been previously detected in any other area (Bozarth *et al.* 2010). This individual provides evidence that there may be some connectivity between kit foxes on the northern end of the Carrizo Plain and those farther north in the hills of the Diablo mountain range.

We found that kit foxes primarily used the southern and eastern portions of the TSF site (Figure 2-4), which contains a high proportion of suitable kit fox habitat. Den locations over all three years paralleled scat density and location of kit foxes in these suitable areas. Kit foxes seldom used the northern and western portions of the site, which were mainly agricultural lands. These results are concordant with evidence from several radio-tracking studies. All 2,231 nocturnal locations taken from 47 radio-collared kit

foxes in the Lokern area were located in natural habitats and none were in the adjacent annual crops or almond orchards (Nelson *et al.* 2007). Radiotelemetry studies in the city of Bakersfield showed that kit foxes did not use the active annual croplands bordering the urban areas that they were using (B. Cypher, personal communication). Studies of kit foxes inhabiting the Semitropic region in northern Kern County and the Bena landfill near Bakersfield also showed that they avoided agricultural lands (B. Cypher, personal communication). Although we consistently found foxes on active cropland in the southern part of the TSF site, these samples were found mostly along the access road through the fields rather than within the fields. In addition, this cropland is surrounded on three sides by annual grasslands, which are good habitat for kit foxes (U.S. Fish and Wildlife Service 1998), even with managed grazing (Germano *et al.* 2012). Agricultural lands are unattractive to kit foxes due to a lack of underground den sites (Warrick *et al.* 2007), low plant diversity and decreased prey availability (Collins 2010). Attacks by larger predators, particularly coyotes, are the primary source of mortality for kit foxes (Cypher *et al.* 2000; Nelson *et al.* 2007; Ralls& White 1995b). As kit foxes maintain multiple dens in their home range (Moehrenschlager *et al.* 2004) and attempt to escape from danger by running to the nearest den (Ralls& White 1995b), they are likely more vulnerable to larger predators when traveling across or near agricultural lands with few dens.

Although we did not target them for monitoring, there was evidence of coyotes and non-native red foxes, which are also known to kill kit foxes (Ralls& White 1995b), in the TSF area. Coyotes and/or large coyote scats were observed and coyote howls were

heard by handlers on every transect throughout the TSF as well as on the Little San Juan Ranch and Thorup properties. Over all three years, only 10.6% (n= 33) of the samples we collected were from these non-target species: 6 coyote scats and 27 red fox scats. A handler may collect non-target scat when a dog correctly locates a latrine containing fresh scats from multiple canids (i.e. fox/coyote; (Ralls & Smith 2004) and the handler unwittingly gathers scat from the non-target species. A dog may err in scent discrimination by keying on a similar (yet incorrect) target, or by selecting an incorrect target when a few target scats are present in order to receive the reward (Schoon 1996; Smith *et al.* 2003).

We had few recaptures across years: 3 within TSF and 1 detected first on TSF and subsequently on the Little San Juan Ranch property. We detected a female kit fox (TZ21) in 2010 that was later recaptured in 2011 in the same part of the TSF. In both years, active and natal dens were documented in the same area where she left multiple scat samples. This suggests that she may have been a resident of the area and this transect lies within her home range (Figure 2-4, Supp. Figures 2 and 3). Many dens were also documented in the southern end of TSF but we cannot infer any correlated individuals without more data (genetic or otherwise). Several factors may have contributed to our low recapture rate. First, kit foxes have high annual mortality rates. Annual adult mortality rates ranged from 0.44 to 0.61 in various studies (Moehrenschlager *et al.* 2004) and mortality of young dispersing foxes is extremely high. Koopman et al. (2000) found that 65% of dispersing juveniles died within 10 days of leaving their natal home range. Second, it is possible that some of the young foxes were dispersing late in the season

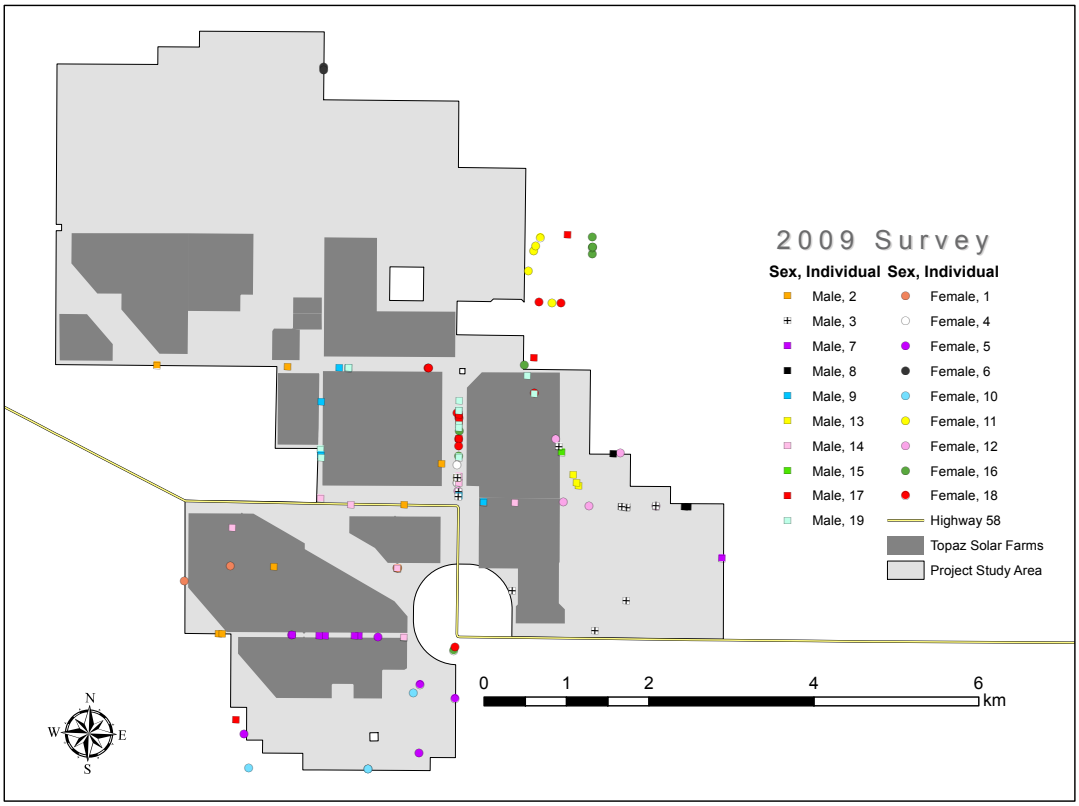
(Koopman *et al.* 2000), and mortality or movement across the landscape prevented the subsequent detection of them. Third, some individuals in the area may have been using or traveling across the TSF site on an intermittent basis and so were not available for sampling during every survey. Finally, we may have failed to detect some of the foxes using the site during each survey. However, the fact that the *Capwire* abundance estimate for each sampling period closely matched the number of unique genotypes found in the area suggests that we detected a high proportion of the individuals present during each survey.

Our finding that kit foxes were present on the Little San Juan Ranch and Thorup properties showed that these properties were suitable as "mitigation" properties to offset any impacts of developing the TSF. We found the same individual on TSF and Little San Juan Ranch, demonstrating that kit foxes could move between them. The Little San Juan Ranch and Thorup properties were both purchased by the developer of the TSF and the Little San Juan Ranch property has been transferred to the California Department of Fish and Wildlife for management as a protected area. Conservation of these properties to the east of TSF prevents future development and farming of habitat that is suitable for kit fox and is connected to the CPNM. Agricultural activities on the TSF (apart from carefully managed grazing) were discontinued in 2011, so habitat conditions for kit foxes on the northern parts of the site are expected to gradually improve as these lands revert to a more natural state. The southern parcels of cropland will be converted into solar panels and the remaining habitat will gradually revert to a more natural condition. Protection of

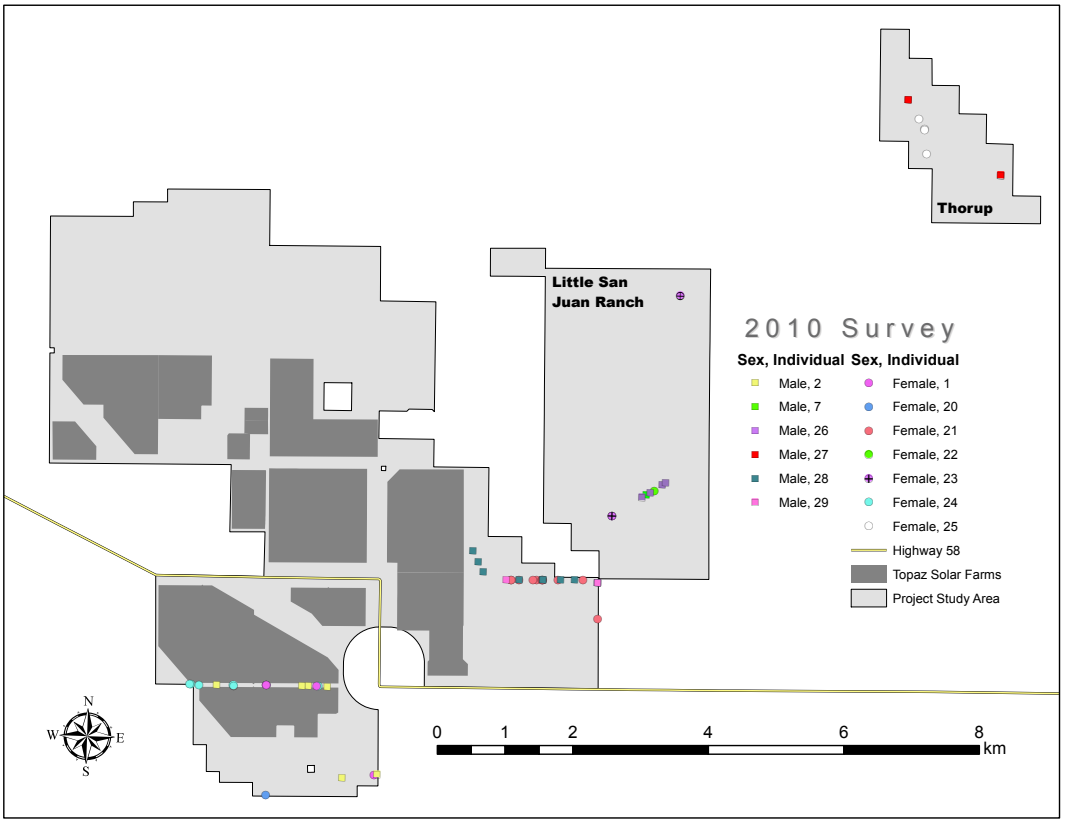
the grasslands surrounding the study area as a part of the TSF conservation program will support an area where we have consistently found kit foxes.

Results from our 2009-2011 surveys were used to suggest placement of the solar panels on the TSF site and establish a pre-construction baseline for SJKF use of the site. Construction of the solar farm began in 2012 and is expected to be finished by 2015. Kit foxes have already demonstrated that they will occupy the completed array areas on TSF. Kit foxes are able to move within the project areas, which contain aboveground panels and no other barriers to movement. The TSF plans to use fencing that allows kit foxes and other medium and small sized mammals to enter and leave the site while deterring coyotes. We plan additional surveys in future years to document the combined effects of constructing the solar plant, exclusion of coyotes, and changing habitat conditions on SJKF use of the area.

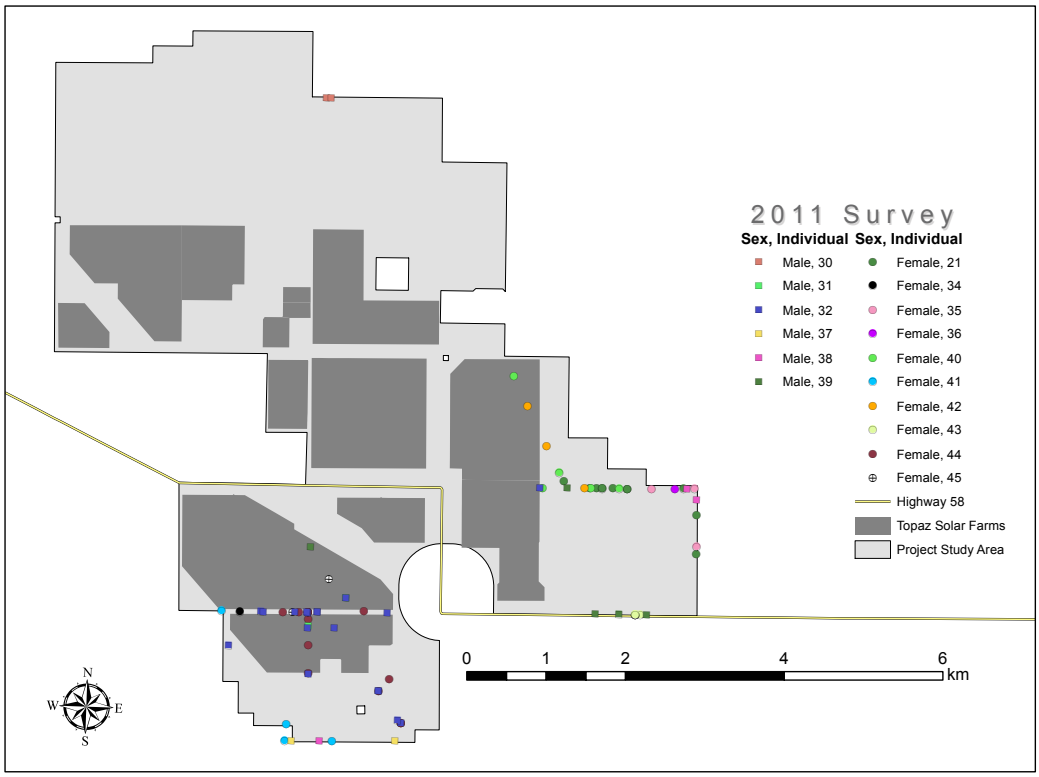
Supplementary Materials



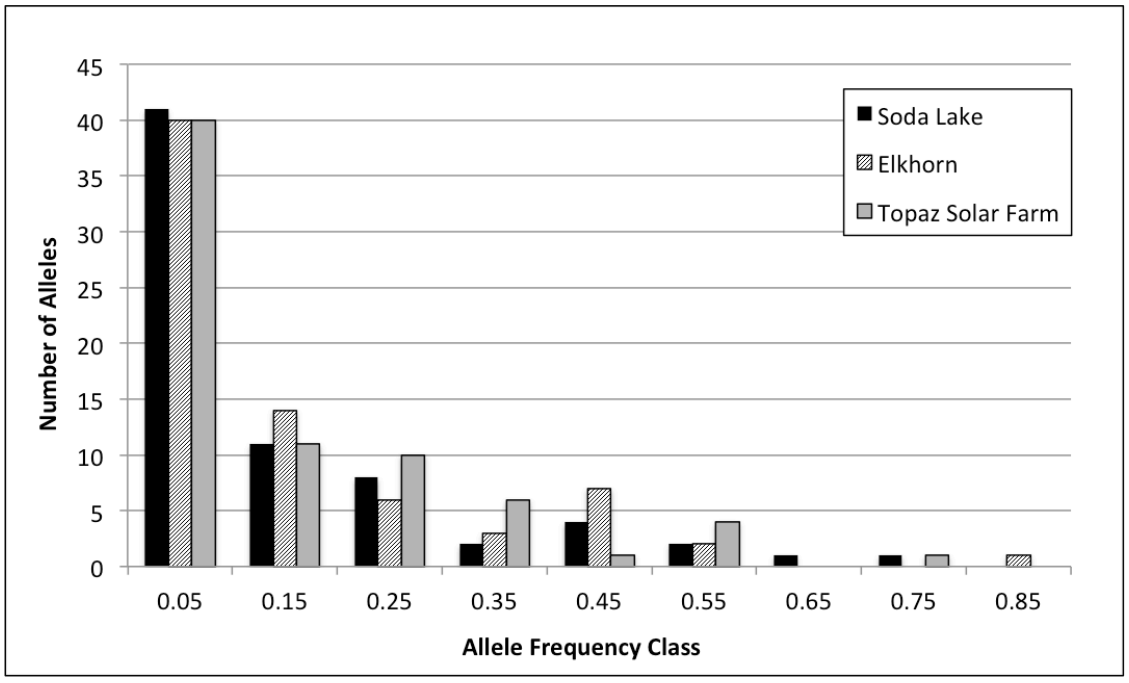
Supp. Figure 1. Locations of individual kit foxes detected in 2009 on TSF (August = 16, November = 11).



Supp. Figure 2. Locations of individual kit foxes detected in Nov 2010 (TSF = 7, KT properties = 6).



Supp. Figure 3. Locations of individual kit foxes detected in Nov 2011 on TSF (17).



Supp. Figure 4. Allele frequency distributions of TSF, Soda Lake, and Elkhorn populations using methods described in Luikart et al. (1998). Allele frequency classes are represented by the median value; ie. 0.05 for allele frequencies from 0-0.1, 0.15 for 0.1-0.2, etc. While a few allele frequencies were as high as 0.75-0.85 in the populations, most alleles were very rare.

Supp. Table 1. Allele frequencies for each microsatellite locus for kit fox detected on the Topaz Solar Farm, Soda Lake Road, and Elkhorn Road, in the Carrizo Plain National Monument, California. Number of alleles varies due to dropout.

Locus	Allele Size	SodaLake	Elkhorn	Topaz
AHTh171	**N**	25	21	42
	216	0.000	0.000	0.012
	220	0.000	0.000	0.012
	222	0.260	0.476	0.560
	224	0.000	0.024	0.012
	226	0.420	0.286	0.333
	228	0.320	0.214	0.060
	230	0.000	0.000	0.012
FH2054	**N**	29	25	42
	163	0.017	0.020	0.000
	167	0.000	0.020	0.000
	171	0.190	0.160	0.298
	175	0.621	0.400	0.274
	179	0.069	0.180	0.262
	183	0.052	0.080	0.083
	187	0.052	0.140	0.083
FH2328	**N**	31	27	45
	139	0.161	0.241	0.200
	143	0.484	0.574	0.578
	147	0.016	0.037	0.078
	151	0.161	0.074	0.033
	159	0.000	0.019	0.022
	163	0.081	0.037	0.011
	167	0.016	0.019	0.078
	171	0.081	0.000	0.000
FH2848	**N**	29	27	40
	226	0.017	0.000	0.013
	232	0.034	0.019	0.075
	234	0.000	0.019	0.063
	236	0.000	0.000	0.038
	238	0.207	0.074	0.338
	240	0.500	0.407	0.225
	242	0.241	0.426	0.238
	244	0.000	0.019	0.013
	246	0.000	0.037	0.000
REN162	**N**	32	26	45
	189	0.000	0.000	0.011
	191	0.406	0.385	0.344
	193	0.016	0.000	0.000
	195	0.250	0.154	0.256
	197	0.281	0.423	0.344
	203	0.000	0.019	0.044
	205	0.047	0.019	0.000
FH2137	**N**	32	29	45
	175	0.000	0.017	0.033
	179	0.000	0.000	0.044
	183	0.047	0.052	0.011
	191	0.188	0.103	0.167
	195	0.094	0.103	0.467
	199	0.047	0.190	0.044
	203	0.156	0.052	0.156
	207	0.047	0.086	0.044
	211	0.031	0.017	0.000
	215	0.063	0.034	0.000
	219	0.031	0.052	0.000
	223	0.031	0.121	0.022
	227	0.125	0.052	0.011
	231	0.000	0.052	0.000
	239	0.125	0.069	0.000
	243	0.016	0.000	0.000

Locus	Allele Size	SodaLake	Elkhorn	Topaz
FH2140	**N**	32	29	45
	107	0.000	0.000	0.011
	109	0.047	0.000	0.089
	129	0.031	0.155	0.022
	139	0.500	0.586	0.322
	144	0.234	0.190	0.211
	149	0.078	0.069	0.289
	154	0.031	0.000	0.000
	159	0.078	0.000	0.056
FH2226	**N**	32	29	45
	127	0.766	0.431	0.544
	139	0.000	0.017	0.000
	159	0.016	0.121	0.000
	163	0.125	0.086	0.167
	167	0.016	0.017	0.000
	171	0.063	0.328	0.167
	175	0.016	0.000	0.000
	179	0.000	0.000	0.111
	195	0.000	0.000	0.011
FH2535	**N**	32	29	45
	126	0.000	0.000	0.011
	142	0.000	0.000	0.011
	146	0.109	0.138	0.144
	150	0.828	0.810	0.789
	154	0.016	0.017	0.022
	174	0.047	0.034	0.022
FH2561	**N**	30	29	45
	204	0.167	0.017	0.000
	208	0.050	0.017	0.000
	210	0.000	0.000	0.044
	216	0.000	0.017	0.056
	224	0.100	0.121	0.022
	228	0.083	0.000	0.000
	238	0.050	0.103	0.167
	244	0.033	0.086	0.256
	248	0.317	0.224	0.189
	252	0.033	0.276	0.111
	256	0.067	0.069	0.133
	262	0.033	0.052	0.022
	264	0.000	0.017	0.000
	266	0.067	0.000	0.000
pez19	**N**	31	29	44
	198	0.290	0.345	0.545
	202	0.258	0.414	0.330
	206	0.452	0.241	0.125

CHAPTER THREE: DISTRIBUTION AND CONNECTIVITY OF SAN JOAQUIN KIT FOXES IN THE PANOCHE VALLEY, CALIFORNIA

A paper to be submitted to *Journal of Mammalogy*

Collaborators: Deborah A. (Smith) Woollett[2], Michael F. Westphal[3], Alice Whitelaw[2], Katherine Ralls[1], and Jesus E. Maldonado[1]

[1]Smithsonian Conservation Biology Institute

[2]Working Dogs for Conservation

[3]Bureau of Land Management, Hollister Field Office

Introduction

Assessing the distribution and abundance of an endangered species is central to understanding past and current influences, population structure, and conservation issues. Carnivores have large home ranges that make surveys of populations very difficult because they require a large amount of resources with low detection rates (Gompper *et al.* 2006). Since the 1990s, non-invasive genetic sampling has been successfully used to study wildlife biology without disturbing or seeing the targeted species (Taberlet *et al.* 1999; Waits& Paetkau 2005). In addition, the use of conservation detection dog-handler teams has increased the number of scat samples that can be collected and the size of the area that can be surveyed in shorter amounts of time. Dogs can survey a greater distance

in a shorter period of time with a higher scat-detection rate as compared to humans (Homan *et al.* 2001; Nussear *et al.* 2008; Smith *et al.* 2003). In sum, all of these things contribute to obtaining larger sample sizes and more data to understand the species of interest across the landscape.

Gathering this information for endangered species can be difficult, especially for species like the small, nocturnal San Joaquin kit fox (SJKF; *Vulpes macrotis mutica*), which dens underground and requires large home ranges for foraging (4-11km^2, (Cypher *et al.* 2000). Although the distribution of the SJKF was not well known when it was listed as endangered in 1967, its range was believed to be reduced mainly to the western and southern ends of the San Joaquin valley with a few satellite populations remaining in the northern end and adjacent Coastal Range valleys (Federal Register 32: 4001, 11 March 1967; (U.S. Fish and Wildlife Service 2010). In 1998, a recovery plan identified three locations with different environmental regimes as core areas that should be protected: the Carrizo Plains, Lokern, and Ciervo-Panoche (U.S. Fish and Wildlife Service 1998).

Most research on SJKF has been conducted in the southern end of the range where good habitat remains and fox densities are high (U.S. Fish and Wildlife Service 2010). Documentation of kit fox in low density can be difficult because coyote pups can easily be mistaken for SJKFs, but the few available studies have shown very low kit fox numbers. In 1970 the California Department of Fish and Game started quarterly spotlight surveys (1 night) on seven survey routes, with only two, Ortigalita and Panoche, in the northern or central part of the historic range (Bell *et al.* 1994). Analysis of 20 years of

surveys showed typical summer counts of 5 kit foxes on the Ortigalita and Panoche routes, with an occasional high of 7-10 kit fox per survey (Bell *et al.* 1994). More extensive surveys of the Ciervo-Panoche area in the 1995 detected 25 kit foxes during 6 survey nights, 18 kit foxes trapped for radiotelemetry, and a mark-recapture estimate totaling 23 (95% CI: 20-31; (Williams *et al.* 1996). Non-invasive scat surveys completed in 2003 on public lands distributed across the entire range of the SJKF revealed that kit foxes were absent in the northern portion and occurred at low densities (2-3 scats/km) in central region of the range (Smith *et al.* 2006b). While this more recent scat survey may confirm kit fox presence, the status of SJKF in low-density areas is unclear and low numbers bring up questions about population stability and whether this area is a source or sink.

The abundance and distribution of SJKF in the Ciervo-Panoche (C-P) core area is poorly known, although this area is thought to contain one of the only substantial kit fox populations left with large portions of habitat with high and moderate suitability for SJKF (Cypher *et al.* 2013). The C-P is located in a rural area that has been used mainly for grazing and contains very few man-made structures. We surveyed the C-P to obtain current information on the status of SJKF in the area. We used well-established methods for conducting non-invasive genetic surveys (Bozarth *et al.* 2011; De Barba *et al.* 2010; Dutta *et al.* 2013; Schwartz *et al.* 2010) that were developed specifically for SJKF (Ralls *et al.* 2010; Smith *et al.* 2006b; Smith *et al.* 2003). We used conservation detection dog-handler teams to survey the landscape for scat samples (Ortega *et al.* 2004; Smith *et al.* 2006a). We documented the distribution of SJKF in the area and identified individuals to

estimate abundance. We assessed genetic diversity in the area and investigated connectivity within the C-P and factors influencing population structure. Using this information we made inferences about the stability and relative genetic value of the C-P population for SJKF conservation.

Methods

Sampling

Study area. – The C-P core area comprises ~450 km^2 of medium to highly suitable SJKF habitat. The C-P includes suitable kit fox habitat in the Ciervo, Tuney, and Panoche Hills, the narrow stretch of habitat between I-5 and the foothills, and the large continuous stretch of relatively undisturbed land in the Panoche Valley (Figure 3-1). The Panoche Valley (PV) was the basin of a Pleistocene lake (Barrows& Ingersoll 1893), resulting in a fertile valley supplied by the Panoche Creek and full of native grasslands. Like other areas in the San Joaquin Valley, parts of the PV have been used for grazing and cultivation of crops since the mid-1800s (Frusetta 1991). However, the PV has not been further developed, with ranchers living outside of the valley and using the land modestly for ranching over the past centuries (Frusetta 1991). In addition, a large portion of the PV is federally owned and protected habitat will remain undeveloped. Thus, the PV has highly suitable habitat for SJKF and is the largest area of land in the northern or central range with low levels of human impact.

Surveys. – Professional conservation detection dog and handler teams (Working Dogs for Conservation, Three Forks, MT) were used to locate scats of kit fox within specific portions of the Ciervo-Panoche (C-P) core area over 3 years. Sampling

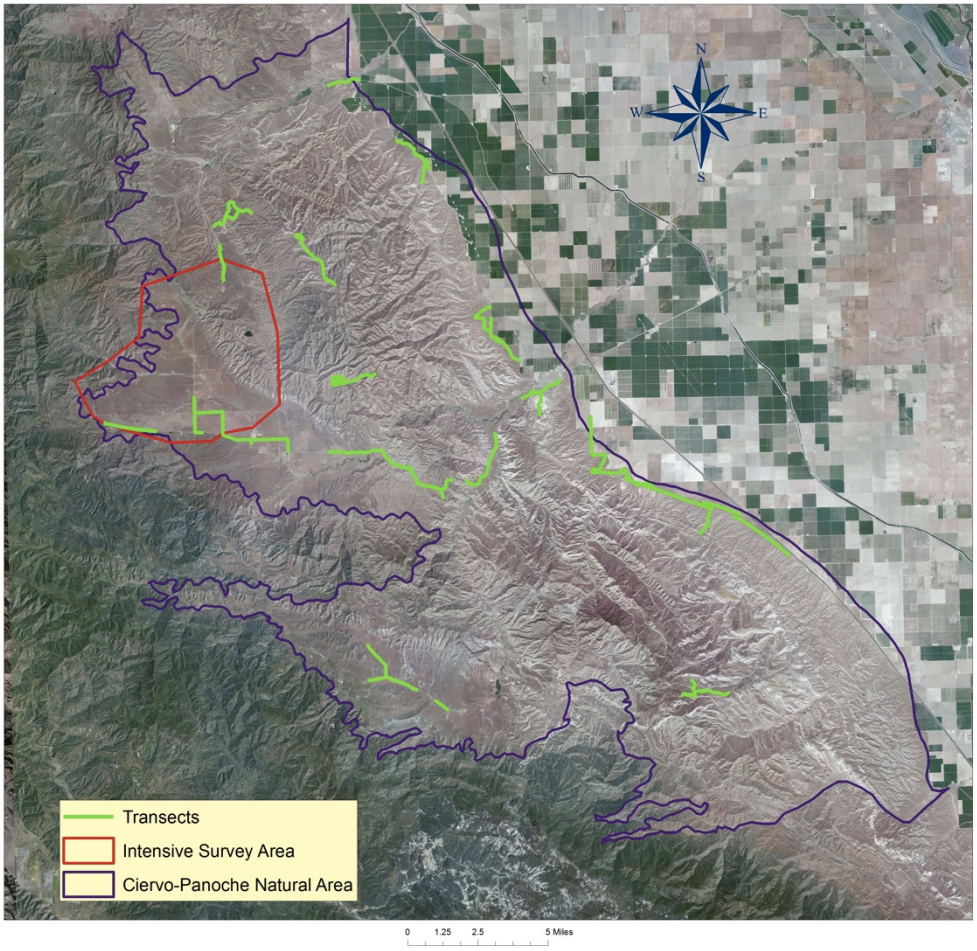


Figure 3-1. Survey areas and transects are demarcated on the habitat suitability map of the Ciervo-Panoche Natural Area.

methodology was similar to that of Smith et al. (Smith *et al.* 2006b): transects were established in areas with suitable kit fox habitat as determined by prior site visits, and through use of previously generated habitat suitability maps (Cypher *et al.* 2013). The majority of transects were on public properties and right-of-ways, though private lands were also sampled when access permission was obtained. Transect routes utilized unpaved roads, where kit foxes frequently deposit a high number of scats (Smith *et al.*

2005). Additionally, some transects were in vegetation, and in one distinct section of the study area alongside of a prominent, intersecting paved road.

Locations of all scats detected by the dog and handler teams were geo-referenced and recorded using Global Positioning System (GPS) units (Garmin GPS III+). As determined by a freshness rating method based on their physical characteristics (Smith *et al.* 2003), both fresh and old scats were collected for DNA analysis, so that recent past as well as current presence in an area could be determined (Smith *et al.* 2006b). Scats were stored in plastic bags with silica gel for desiccation (Fisher Scientific, Pittsburgh, PA) and shipped to the Center for Conservation and Evolutionary Genetics laboratory.

All scat samples collected for genetic analysis from 2009 to 2011 were combined with scats from a separate SJKF monitoring study using detection dogs that we (TRM, DAW, AW, JEM) conducted in the PV study area in 2010. Here, efforts on two adjacent sites included an initial and a repeat survey session that were one week apart, and for which only fresh scats were collected for DNA analysis to determine presence and estimate the number of individual SJKF. By combining data from these two studies we can gain a more comprehensive understanding of SJKF in the C-P.

Molecular techniques

DNA was extracted using the QIAamp DNA stool mini kit and completed species identification, molecular sexing, and microsatellite genotyping as described in Wilbert et al. (*in prep*). Briefly, we determined the species of a scat sample by amplifying a fragment of the mitochondrial control region, which is species-specific by size (Bozarth *et al.* 2010). During this step, we diluted samples with poor amplification (1:15 up to

1:45) to minimize interference from PCR inhibitors in scat samples, and replicated as needed. These dilutions served as an optimization step for the following molecular steps because amplification issues were usually related to the scat sample itself. We then used a canid specific primer to amplify a small fragment of the zinc finger gene, which is found on both X- and Y-chromosomes, and performed a restriction digest, which created two bands for males and one for females (Ortega *et al.* 2004; Ralls *et al.* 2010). Only samples that amplified successfully for the mtDNA species id and ZF sex primers were then characterized for six microsatellites (FH2137, FH2140, FH2226, FH2535, FH2561, Pez19) that have been used reliably for individual identification of kit foxes in our lab (Smith *et al.* 2006a). We previously assessed our ability to differentiate individuals by estimating the probability of identity (P_{ID}) (i.e. the probability of different individuals sharing an identical genotype at random) and the P_{ID} between siblings (Mills *et al.* 2000; Waits *et al.* 2001). Both P_{ID} unbiased and P_{ID} sibs are low enough to suggest that we can differentiate between individuals, including relatives (P_{ID} unbiased = 2.03 x 10^{-6}, P_{ID} = sibs 7.95 x 10^{-3}, Wilbert Topaz *in prep*). Because DNA samples extracted from scats are prone to genotypic error and contamination, we used a modification of the multi-tube approach (Taberlet *et al.* 1997). Briefly, each DNA extract was subject to a minimum of 5 independent PCR amplifications for each locus for homozygous individuals and heterozygotes a minimum of 3 times to verify allele size and to detect allelic drop out. We used fluorescently labeled forward primers in all PCRs to visualize them on an ABI PRISM* 3130 Genetic Analyzer (Applied Biosystems Inc. Foster City, CA).

We used the Excel Microsatellite Toolkit (Park 2001) to compare genotypes and defined individuals by unique genotypes and samples with matching alleles at all loci. We checked all genotypes carefully but paid particular attention to those that differed at only 1 or 2 loci for accuracy of genotype and data entry. We also compared genotypes between samples collected in 2009, 2010, and 2011 to determine if any individuals had been recaptured between survey sessions and assigned them the same number.

Because we identified an abundant number of kit foxes that would allow us to further investigate patterns of genetic diversity in this population, we genotyped individuals at 5 additional microsatellites (AHTh171, FH2054, FH2328, FH2848, and Ren162) for more robust statistical power. Since we had previously determined that we had enough power to distinguish individuals using the original 6 microsatellite loci (P_{ID} $=2.03\ 10^{-6}$; see below), we selected 1 representative scat sample that amplified most reliably for each individual and genotyped these samples for additional loci using the same protocols and conditions described above. We then used the final genotype of 11 microsatellites to assess genetic diversity and population structure.

Analysis

Genetic diversity. – We used GenAlEx to calculate expected (H_E) and observed (H_O) heterozygosity, and SPAGeDi to calculate allelic richness (A_R) as defined by Nielsen (2003), effective number of alleles (Ne), and global estimates of F_{IS}. We tested for departure from Hardy-Weinberg equilibrium with a global test of heterozygote deficiency using the Markov chain method in GenePop 3.4 (Raymond& Rousset 1995). We also used GenePop to test for linkage disequilibrium between pairs of loci. We

calculated several measures of relatedness: the Ritland estimator (RI; (RIRitland 1996), the mean Lynch and Ritland estimator (LRM; (Lynch& Ritland 1999), and the mean Queller and Goodnight estimator (QGM; (Queller& Goodnight 1989) in GenAlEx. We did these calculations for the C-P area overall, as well as the defined populations within C-P.

Genetic patterns of population structure. – We investigated genetic differentiation and structure within the Panoche Valley through Bayesian analyses and statistical measures of variance. We used 3 Bayesian clustering approaches to detect population structure in individuals in the C-P, 1 non-spatial (STRUCTURE version 2.3.2; Pritchard *et al.* 2000) and 2 spatially sensitive methods (TESS - Chen *et al.* 2007; GENELAND - Guedj& Guillot 2011). STRUCTURE assigns genotypes to genetic clusters that maximize Hardy-Weinberg and linkage equilibrium. We investigated the likelihood of the number of genetic clusters (K) from 1-8 with 10 independent replicates, admixture ancestry and correlated allele frequency models, and Markov chain Monte Carlo resampling for 400,000 generations following a burn-in of 100,000. We visualized the STRUCTURE results using STRUCTURE Harvester (Earl& vonHoldt 2011), and determined the number of clusters using the Evanno method (2005). Results from the ten replicates of the chosen K were averaged using CLUMPP 1.1.2 (Jakobsson& Rosenberg 2007) and the appropriate "ind" file created from Harvester. We assigned individuals to a cluster if they had an ancestry assignment of at least 0.70, as done by Corander and Marttinen (2006) and Marsden et al. (2012).

The spatially sensitive analyses, TESS and GENELAND, use the same principles as STRUCTURE except that they also incorporate the geographic coordinates for each individual. Because we recaptured many individuals and currently no programs can incorporate more than one geographic location per individual for this type of analysis, we chose one true GPS point that was located in the middle of each individual's sample distribution. In the case where an individual had only two scat locations, we chose the point that was closer to other individuals so that the analysis was as conservative as possible in detecting structure that was influenced by geographic information. In TESS, we analyzed our data using the default geospatial weighting of 0.6, which ranges from 0 (no geographic information used and acts like STRUCTURE) to 1.0 where location is heavily weighted over genetic information. We ran values of K from 1 to 10, with 10,000 burn-in sampling and 50,000 recorded steps. Each potential population was run 10 times for 3 models: no admixture (HMRF model), the CAR admixture model, and the BYM admixture model. The most appropriate number of populations was chosen using the highest 30% DIC scores for all runs and then looking at two indicators: 1) determining the lowest value, and 2) graphing the average DIC scores for each model as the value of K increased. In GENELAND, we ran the correlated frequency model for 500,000 steps with every 100 step recorded and a burnin of 100,000 steps. This model was run for 10 replicates of 1-10 populations.

We performed a hierarchical analysis of molecular variance (AMOVA) using GenAlEx to test for the proportion of genetic variation between populations previously identified using STRUCTURE. The AMOVA calculates the percentage of variance

between and within populations and we tested the significance of these values using an F-test. With 1 degree of freedom for the numerator (number of groups, K, – 1) and 180 degrees of freedom for the denominator (total sample size equals 2 alleles per individual – 2), we calculated the F ratio and probability of variance in Excel. We visualized genetic differentiation with a principal coordinate analysis (PCoA) as conducted in GenAlEx. We also calculated the mean Eigen values for principal coordinates 1 and 2 for both populations and plotted this mean on the PCoA to display the population separation. Using GenAlEx, we also calculated two measures of genetic differentiation: F'st as described by Slatkin (Slatkin 1995) and D_{est} (Jost 2008) as calculated by Meirmans and Hedrick's equation 2 (Meirmans& Hedrick 2011).

To examine alternative causes for genetic structure, we tested for isolation by distance (IBD) and kin clustering. We generated matrices of genotypic distance (GD) and linearized geographic distance (LnGGD). We executed Mantel tests between GD and LnGGD with 10,000 permutations to test for significant correlations of IBD. We created matrices of pairwise relatedness using the mean Queller and Goodnight estimator (QGM; (Queller& Goodnight 1989) and tested for correlation with LnGGD using the same Mantel test parameters. Both mantel tests were performed for the entire region as well the defined populations using GenAlEx. We also performed a spatial autocorrelation analysis (Koressaar& Remm) to see if individuals next to each other were more related, which would be expected with family groups, and at what spatial scale this kin clustering might have been occurring. SAA is a multivariate analysis that measures the genetic similarity of individuals within a distance class. A distance class of 2km was

chosen because of the ability to detect fine scale structure relating to a home range of 4-11km^2. The auto correlation coefficient (r) ranges from -1 to +1 and is closely related to Moran's I, a measure of genetic relatedness. We ran SAA in GenAlEx using 10,000 bootstraps around r and 10,000 permutations to calculate the 95% upper and lower confidence limits of random sampling of the data assuming no spatial structure. Heterogeneity of spatial structure (omega) was used to test significance of the correlogram.

Population estimate. – In order to obtain an estimate of kit fox abundance in the Panoche Valley, we analyzed our data for all three years with *Capwire* (Miller *et al.* 2005). *Capwire* uses the number of samples per individual to estimate the detection probability and then runs urn simulations to estimate population size using 2 capture probability models (equal capture probability, ECM, or two innate rates model, TIRM). The appropriate model is chosen using a likelihood-ratio test, whereby TIRM is used when capture rate is heterogeneous between individuals. We calculated a population estimate using *Capwire* in two ways: 1) all of our data together for an overall estimate and 2) by population as designated by previous analyses.

Results

We collected a total of 597 scat samples during the 4 surveys; 159 scats in July 2009, 250 in July 2010, 94 in September 2010, and 94 in May 2011. Using mitochondrial DNA alone, we were able to identify for 314 scats as kit fox (41, 96, 90, and 87 respectively). We found a new mtDNA haplotype that was 16bp shorter than the typical fragment length (236bp vs. 252bp; Table 3-1), which has been recently described

(Bozarth *et al.* 2010). After adding the sex identification and the microsatellite genotypes based on 6 loci (both from nuclear DNA), we identified an additional 84 samples as SJKF for a total of 398, with a very high amplification rate in the last two surveys (July 2009 = 26%, July 2010 = 70%, Sept 2010 = 96% and May 2011 = 93%; also see Figure 3-2). Using 6 microsatellites the C-P area, the unbiased P_{ID} was 1.4 x 10^{-5} and P_{ID} sibs was 9.9 x 10^{-3}. We assigned 332 samples to unique genotypes, identifying 93 individuals, 56 males and 37 females for an overall 1:0.66 sex ratio.

Table 3-1. Number of mitochondrial species ID sizes found in each subpopulation. Both of the populations have both of the haplotypes despite the differentiation shown in the nuclear markers.

	West		East	
Size	#	%	#	%
236bp	29	45%	11	69%
252bp	36	55%	5	31%
Unsure	3		7	

We collected an average of 3.6 scats per individual, although for a few kit foxes we found 10-20 scat samples per individual. We recaptured 4 individuals in 2010 and 5 in 2011, including 1 female (P011) found every year despite a low number of scats (6). SJKF scat samples were only found in certain survey area, specifically the central Panoche Valley, the western side of the Panoche Hills, Silver Creek Ranch, the eastern

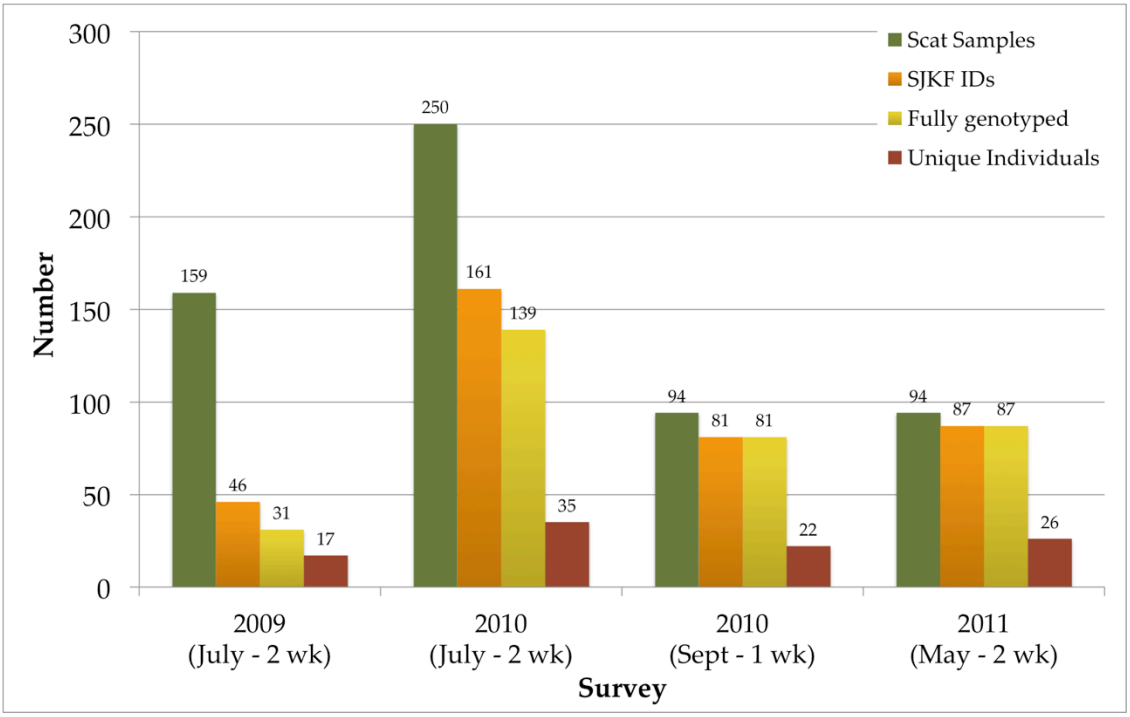


Figure 3-2. Survey effort and genetic productivity.

foothills of the Ciervo Hills up to Interstate-5, and Vallecitos area (Figure 3-3). No SJKF were identified in Little Panoche Reservoir, Tumey Hills, the eastern foothills of the Panoche Hills, or the Ciervo Hills.

We only included individuals for the genetic analysis that we could confidently genotype for at least 8 of the 11 microsatellite loci. We could not use two individuals for further analysis due to poor genotyping success (each with only one scat sample identified), but had sufficient data for a majority of the individuals detected (total n=91; n_8=3, n_9=5, n_{10}=10, n_{11}=74). We found moderate levels of heterozygosity (H_O=0.576±0.055, unbiased H_E=0.641±0.048; Table 3-2). Microsatellite loci had 4 - 13

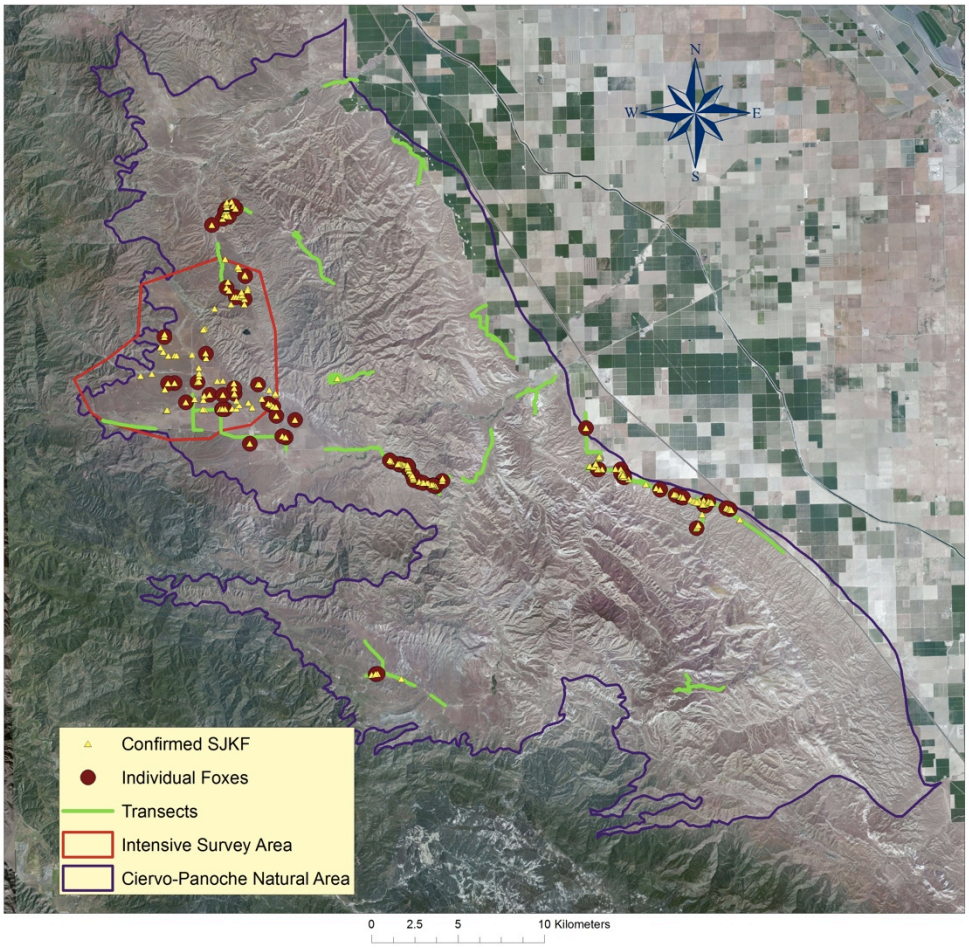


Figure 3-3. Scats of SJKF (genetically identified; n=392) over the 3 years are mapped on the Ciervo-Panoche survey area. The red dots represent the central location of each individual (n=91), chosen as the central scat location from all samples.

alleles per locus with allelic richness of 7.79 and 3.18 effective number of alleles. The global estimate of F_{IS} for the overall population was 0.100. The C-P population was not in Hardy Weinberg (HW) equilibrium due to heterozygote deficiency (p = 0.000).

Table 3-2. Measures of genetic diversity in the Panoche Valley; A_R = allelic richness, Ne = number of effective alleles, He = unbiased heterozygosity, Ho= observed heterozygosity, and F = fixation index, RI = the Ritland estimator, LRM = the mean Lynch and Ritland estimator, QGM = the mean Queller and Goodnight estimator.

Pop	A_R	Ne	He	Ho	F_{IS}	RI	LRM	QGM
Panoche Valley	6.72	3.065	0.612	0.564	0.074	-0.009	-0.016	-0.017
Ciervo Hills	5.50	2.764	0.617	0.612	0.011	-0.023	-0.045	-0.043
Total	7.79	3.136	0.641	0.576	0.100	-0.006	-0.012	-0.012

Bayesian clustering analysis in STRUCTURE indicated strong population subdivision, with the ΔK statistic (Evanno *et al.* 2005) showing a peak at K = 2 (Figure 3-4). Cluster 1 was comprised of individuals predominantly located in the main Panoche Valley west of the Panoche and Ciervo Hills) as well as individuals detected along Little Panoche Road in the north and into the western plateau of the Panoche Valley. Cluster 2 included individuals on the eastern slope of the Ciervo Hills between the foothills and I-5. Using the ancestry assignments of at least 70% with 10 replicates, 56 individuals were assigned to cluster 1 and 22 were assigned to cluster 2. Two of these individuals assigned to cluster 1 were located to the east of the Ciervo Hills and 2 individuals from cluster 2 were located in the Panoche Valley. These 4 individuals were identified as migrants. We also identified individuals with mixed ancestry (between 30-70%) that were assigned to the cluster corresponding to the location where they were found. We found 13 individuals with mixed ancestry, including 12 located in the valley and 1 to the east of the Ciervo Pass. These putative migrants and individuals of mixed ancestry are delineated in Figure 3-4a by a dashed line separating the individuals with 70% ancestry. TESS also detected 2

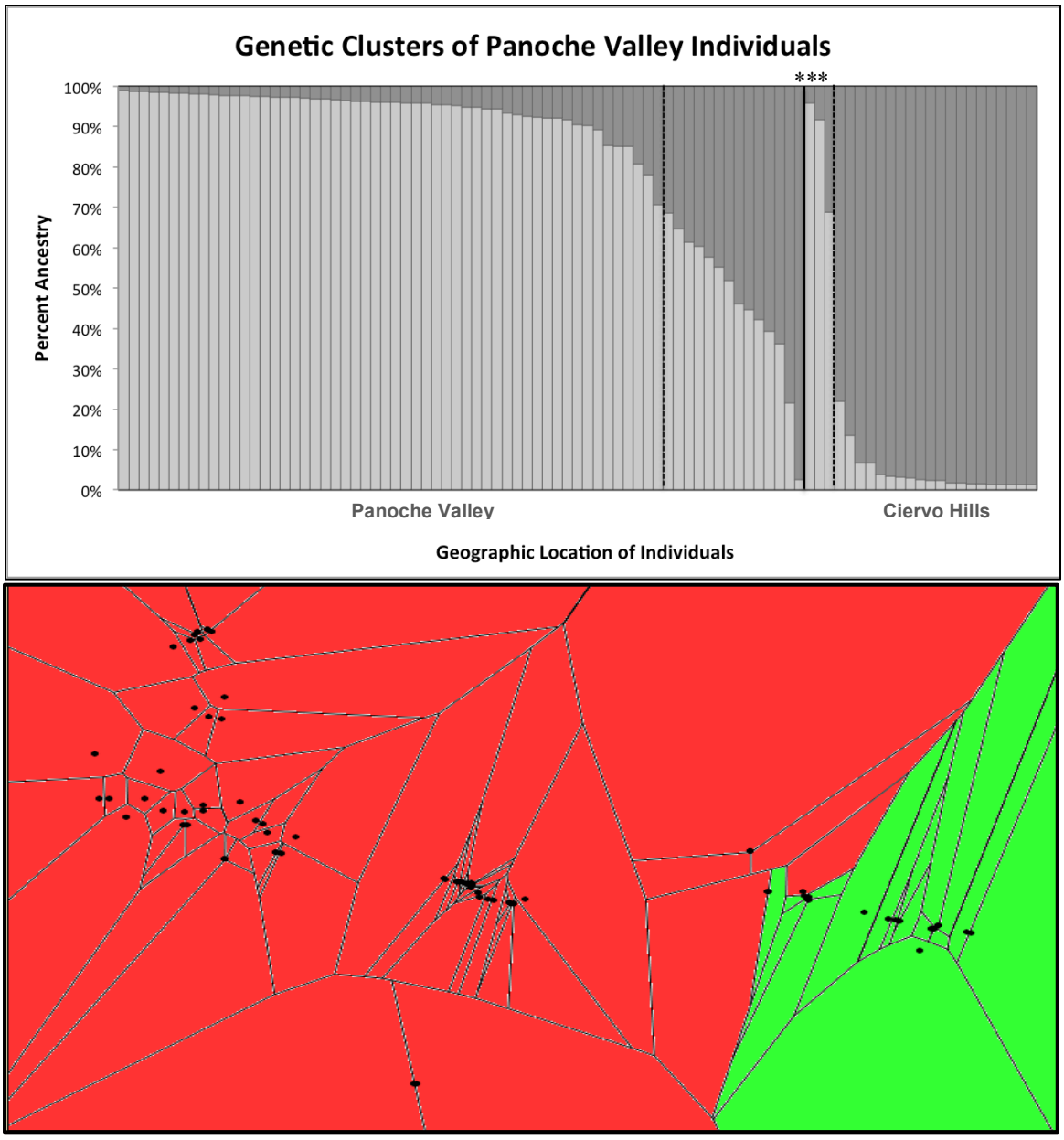


Figure 3-4. a) Percent ancestry of each individual (n=91) identified by Structure; Cluster 1 in light grey and Cluster 2 in dark grey. Individuals to left and right of the dark line represent individuals found to the West and East of the Ciervo Hills. The dotted lines represent a 70% cutoff, where by individuals are either hybrids or migrants (asterisks) if they are located to the right of the dotted line in the Western group or to the left of the dotted line in the Eastern group. b) Clustering of individuals (n=91) from TESS using geospatial weighting and the clear division that aligns with the eastern ridge of the Panoche Valley.

populations within the samples. TESS found 20 individuals in cluster 2 and 71 in cluster 1, placing all individuals detected in the Panoche Valley and the two migrant individuals with cluster 2 ancestry into one group. GENELAND supported the same structuring found in TESS, assigning higher posterior probabilities for migrants and mixed ancestry to the location of sampling (results not shown). For future analyses, we will compare the two geographically related populations – the Panoche Valley (PV, n=68) and the Ciervo Hills (CH, n=23), with mixed ancestry and migrant individuals assigned to their population of origin.

Separation between individuals in the PV and CH was also supported by statistical measures of variance. The hierarchical AMOVA found 10% variance between populations and 90% within populations. The one-tailed F-test of variance of the AMOVA detected significant differentiation between populations ($F = 8.892$, $F_{1,180}<0.003$). The first two coordinates of the PCoA represented 47.8% of the total variation (coordinate 1 = 31.0%, coordinate 2 = 16.8%; Figure 3-5). Measures of genetic differentiation also showed significant separation between the two populations: $F'st = 0.278$ and Jost's $D = 0.190$ (all $p<0.0001$).

Mantel tests showed a weak but significant correlation between genetic diversity and geographic distance, even within populations (C-P: $r = 0.200$, $P < 0.0001$; PV: $r = 0.111$, $P < 0.013$; CH: $r = 0.225$, $P < 0.003$). Mantel tests showed a significant negative correlation between relatedness and geographic distance (C-P: $r = -0.314$, $P < 0.0001$, PV: $r = -0.180$, $P < 0.0001$; CH: $r = -0.325$, $P < 0.0001$). Spatial autocorrelation analysis of all SJKF in the C-P area indicated that individuals within 6km of each other are

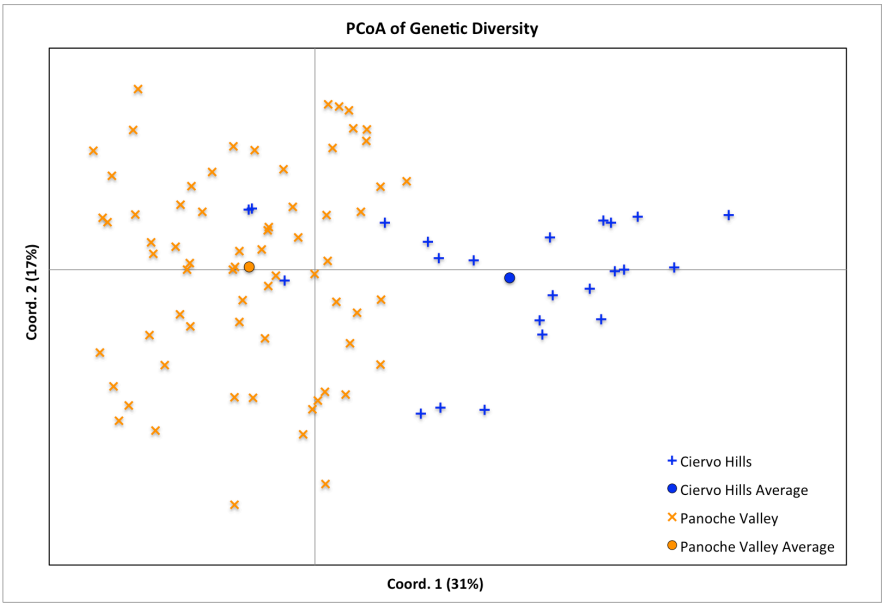


Figure 3-5. Principal coordinate analysis of shows a clear separation between the genetic diversity of individuals detected in the Panoche Valley (orange) and those found in the Ciervo Hills (blue).

significantly more related than average ($p<0.0001$; Figure 3-6). All estimators for relatedness for the C-P area and populations were similarly low; for example, the Queller & Goodnight relatedness estimate for the C-P was -0.012 (QGM), CH = -0.043 and PV = -0.017 (Table 3-1).

Genetic diversity was similar between the two populations. The PV population had 3 - 11 alleles per locus with A_R = 6.72 and 3.12 effective number of alleles. The CH population had 3 - 9 alleles per locus with A_R = 5.50 and 2.87 effective number of alleles. Expected and observed heterozygosities were lower for each population as compared to the overall C-P values (see Table 3-1). The F_{IS} value for PV was higher (0.08) than CH (0.01), but both are lower than overall C-P F_{IS} = 0.103. Tests in GenePop showed the CH

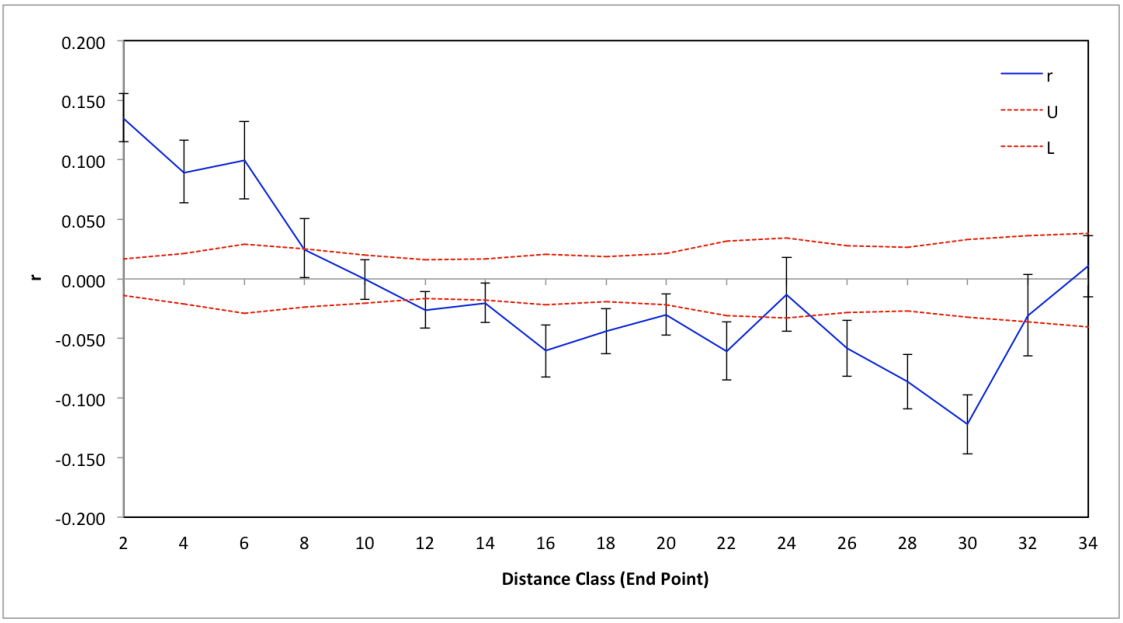


Figure 3-6. Spatial autocorrelation analysis of SJKF in the C-P area shows that individuals within 6km of each other are significantly more related than average ($p<0.0001$). The autocorrelation coefficient (r), shown in blue, represents the genetic similarity of individuals within a distance class and is similar to Moran's I. The dotted red lines signify the upper (U) and lower (L) bounds of the 95% confidence interval of a random distribution of the data.

population is in HW and linkage equilibrium. However the PV population is not in HW equilibrium. The PV population also had 3/55 pairs of loci showing linkage disequilibrium. However we could not see a trend indicating that a locus was more often involved in linkage and this number of significant test is expected by chance at the 0.05 level (0.05 X 55 = 3).

We used the number of times that an individual was resampled over the three year period to estimate the population size in the Panoche Valley with the program *Capwire*. The TIRM model indicating capture heterogeneity was selected by the program as the most likely model of capture probability. We estimated a population size of 114 (96-

125). Running the same program for each population, we estimated 25 individuals (23-28) in the CH population and 90 (75-109) in the PV population.

Discussion

This study confirmed the presence of SJKF in the C-P natural area and explicitly defined kit fox distributions within the area. Using this type of sampling is notoriously difficult due to low quality and low quantity DNA that leads to low amplification rates (Beja-Pereira *et al.* 2009; Taberlet *et al.* 1999; Waits& Paetkau 2005). Despite a broad range in quality of scat samples used in this study, we were able to identify 66.7% of all scat and 96.8% of fresh scat as SJKF. Fresh scat had higher success downstream in the genetic analysis, which further reinforces the importance of collecting fresh samples. We were conservative in identifying scat samples to the species and to the individual level unless we had clear and consistent data that supported this conclusion, and yet we identified 398 SJKF scats and 93 individuals.

Species identification showed that SJKF in this area had two mtDNA control region haplotypes of different fragment sizes, with half of the individuals carrying each type. The smaller 236bp haplotype has only been found in one individual outside of the Panoche Valley (Wilbert et al. in prep), suggesting that this haplotype could be more restricted to the C-P area.

We identified two male SJKF individuals in the Vallecitos area, one in 2010 and one in 2011. Detecting the presence of these two males is significant because the last official sighting of kit fox in the area occurred in the 1970s. Two other samples from that same location had mixed genetic signatures of SJKF and coyote but we could not

confidently determine if this was due to a cross-contamination in the field, evidence of a coyote preying on a kit fox, or a kit fox scavenging on a coyote. Regardless, our two positive SJKF identifications provide evidence of the presence of kit foxes in the Vallecitos area.

Multiple analyses of the C-P area consistently split our sampled individuals into two groups despite the small geographic scale of this study. Although genetic diversity increased with distance and relatedness decreased with distance, the SAA showed that individuals are significantly related only within 6km of each other. This indicates that family groups were influencing the high correlations with genetic distance and relatedness overall. The small distance over which individuals were related was the result of family groups present in the C-P area. However, we also found genetic structure at a larger geographic scale, indicating the presence of 2 distinct populations with fine-scale family group relatedness. On a larger scale the dissimilarity between the two populations, PV and CH, was confirmed with 10% of variance between populations (AMOVA, $p<0.003$, Figure 3-6) and Jost's $D = 0.190$ ($p<0.0001$). These values are both large and highly significant given the small number of individuals in close proximity.

Population structuring differed slightly between STRUCTURE and the spatially explicit programs, TESS and GENELAND. This may be expected due to the nature of their algorithms and the biological importance of geographic isolation. When TESS used spatial information to assign individuals to clusters, it tended to assign individuals to the locations where they were sampled, and was biased to assign individuals to cluster 1. Therefore, more rigorous analyses would have been needed if the focus were to identify

the source population of these migrants. In our case, this had no bearing on the conclusion of population structure in the PV. Similar to other studies (Pelletier *et al.* 2012), the exact cutoff values, such as 60% by (Coulon *et al.* 2008) or 80% by Li et al. (2009), did not change the assignment of an individual to a cluster or as containing mixed ancestry, nor the specific individuals identified, thus confirming our results. While only a handful of migrants were identified, admixed individuals were predominantly found in the valley population. This suggests that dispersing individuals may have greater ability to move westerly or to get settled and survive in the valley. The eastern slopes of the Ciervo Hills are very limited in the amount of habitat and kit foxes may experience increased mortality by agriculture, cars or humans.

We found that the 2 SJKF populations generally correspond to the central PV and the eastern side of the CH. While it may seem that the PV and CH are separated by a geographic barrier to kit fox dispersal because of the high elevation and unsuitable habitat, there is a pass at the base of Tumey and Panoche Hills that connects the two areas at a low elevation with habitat that appears to be suitable for kit foxes. No SJKF scat samples were found in this connecting habitat. Coyotes (*Canis latrans*) and red foxes (*Vulpes vulpes*) were detected during our surveys, and gray foxes (*Urocyon cinereoargenteus*) and bobcats (*Lynx rufus*) are known to be present in the area. These carnivores could be using this area and displacing or killing kit foxes that move into or through the area (Briden *et al.* 1992; Cypher& Scrivner 1992; Cypher& Spencer 1998; Cypher *et al.* 2000; Nelson *et al.* 2007; Ralls& White 1995a; Spiegel& Dao 1997). Thus, the presence of larger predators could not only have immediate effects on the survival of

an individual, but would also be responsible for a reduction in the number of migrants that would result in the patterns of population substructure that we detected within the C-P area.

The C-P natural area is of conservation concern because it is unique in that it is geographically isolated from the other core populations of SJKF and is facing threats due to anthropogenic development. Therefore, despite our surveys not being specifically designed to estimate population sizes, we nevertheless attempted to obtain these estimates because we deem this information to be useful for conservation planning. First, our estimates based on genetic identification of scat samples suggest that a minimum of 93 individuals occupied the C-P area over the 3 years of surveys. Next, we calculated super-population estimates using *Capwire* because of the variety of locations and times of our surveys. A super-population represents an estimate of the total number of individuals that were exposed to sampling during the sampling periods that were combined to produce the estimates (Schwarz& Arnason 1996). The overall C-P estimate of 114±26 individuals related closely to the 2 population estimates, whereby combining the 25 individuals in CH (23-28) and the 90 individuals in PV (75-109) would equal a total estimate of 115 individuals (98-137). These estimates correspond to previous estimates of fluctuating population sizes from 100-300 individuals, and are a reasonable number of individuals given the amount of habitat in the C-P natural area. In optimal habitat, Nelson et al. (2007) estimated that 6 km^2 are required for each kit fox family group, meaning that this area could support 225 kit foxes at most. Our estimates are much lower than this number, but we have detected family groups within a 6 km distance, which may

translate into home ranges that are 2-3 times larger. In addition, capture rates were heterogeneous (as indicated by the TIRM model), which may relate to changing numbers of kit fox at different times of the year and at different locations. Kit fox are known to have fluctuating population sizes, with high reproductive rates leading to high numbers in spring and early summer and then high juveniles mortality rates during dispersal in late summer and fall (Koopman *et al.* 2000). This population fluctuation has been documented in long-term spotlighting data (Ralls& Eberhardt 1997). In addition to these diurnal fluctuations, precipitation cycles that change between years influence kit fox reproduction through a trophic cascade of resource availability (Cypher *et al.* 2000; Standley& McCue 1992; White& Ralls 1993). With these seasonal and annual changes, population estimation is difficult without long-term monitoring. Therefore, we have established a baseline of a minimum number of 93 individuals in the C-P natural area and a superpopulation estimate of 114±26 individuals. These estimates can be a reference for future surveys.

CHAPTER FOUR: HISTORICAL AND CURRENT PATTERNS OF CONNECTIVITY IN THE SAN JOAQUIN KIT FOX METAPOPULATION

A paper to be submitted to *PLoS ONE*

Collaborators: Katherine Ralls[1], Brian Cypher[2], Deborah A. (Smith) Woollett[3], Christine Van Horn Job[2], Michael F. Westphal[4], Samantha Bremner-Harrison[5], Joseph R. Hoyt[1,6], and Jesus E. Maldonado[1]

[1]Smithsonian Conservation Biology Institute

[2]Endangered Species Recovery Program, CSU-Stainslaus

[3]Working Dogs for Conservation

[4]Bureau of Land Management, Hollister Field Office

[5]Nottingham Trent University

[6]University of California Santa Cruz

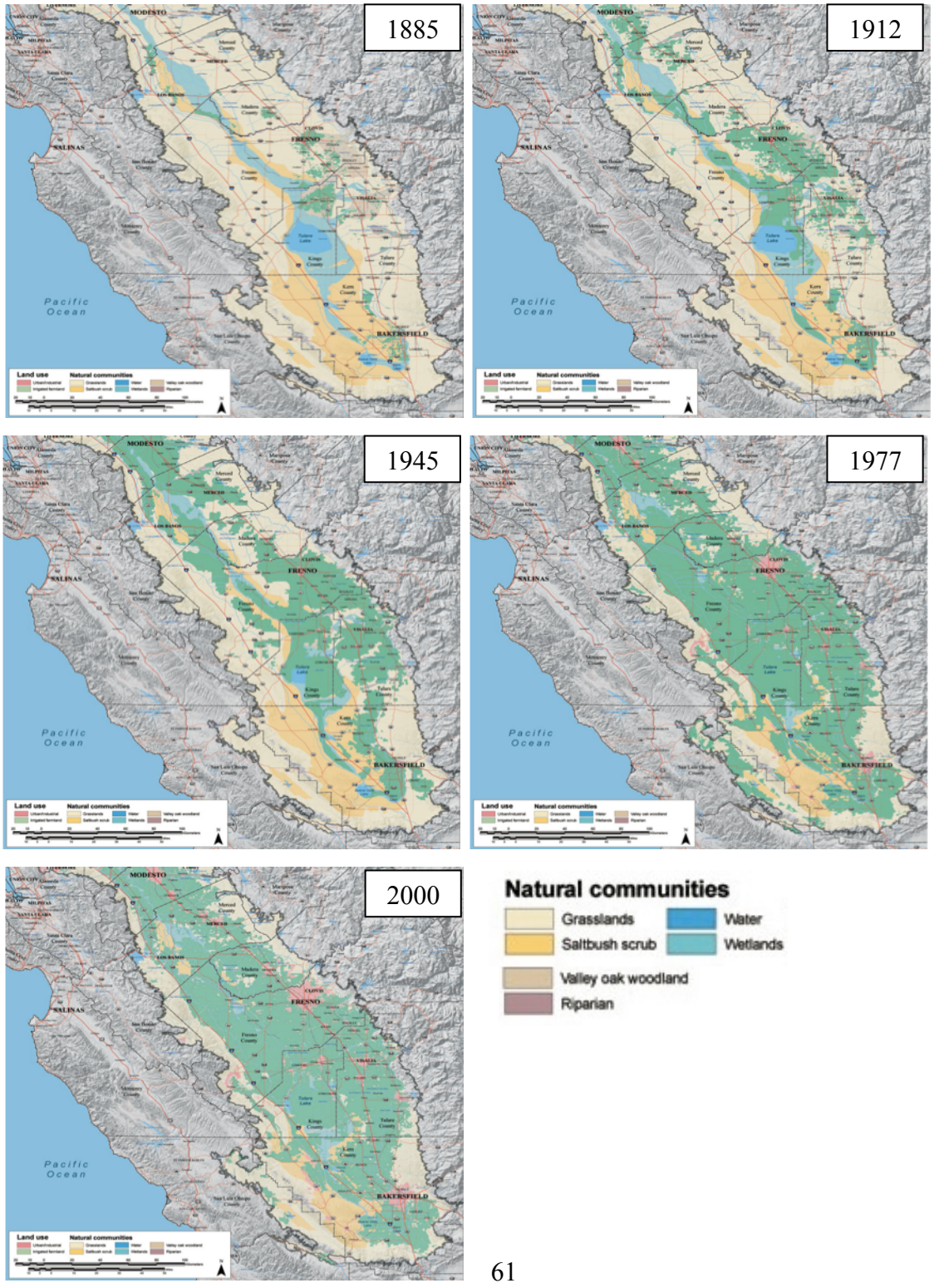
Introduction

San Joaquin kit foxes (SJKF; *Vulpes macrotis mutica*) were historically abundant throughout the San Joaquin Valley but have dramatically declined over the last 150 years with many anthropogenic impacts including fur trapping, pesticides, road kill, habitat loss and fragmentation (Bjurlin *et al.* 2005). Biologists currently estimate approximately 3000 individuals throughout the remaining arid scrub and grassland habitat in the San Joaquin

Valley and surrounding hillsides (Cypher *pers. comm.*). The 40,000 km^2 valley has changed from a mosaic primarily of grasslands, arid scrub, and wetlands, to a landscape predominated by irrigated farmland and urban development with patchy remnants of grassland or scrub (Kelly *et al.* 2005). Even as early as 1937, Grinell and Swarth noted that most of the valley was cultivated or used for grazing, and kit foxes had been heavily impacted by fur trappers (as noted in (Kelly *et al.* 2005) The following decades brought even greater habitat change as irrigation technology expanded the capabilities of intensive agriculture in the valley. Recent habitat modeling of suitable habitat throughout the species' range estimated that at most 4,267 km^2 of high suitability habitat remains (Cypher *et al.* 2013). Patches of suitable habitat vary in size and degrees of connectivity, with some areas too small to support any kit foxes (Cypher *et al.* 2013).

SJKF have been thought to persist as a metapopulation, with 3 core populations and several other smaller populations (Cypher *et al.* 2000; U.S. Fish and Wildlife Service 1998). If populations are truly behaving in this way, they should be linked by immigration and emigration and recolonization would occur via dispersal after extinction events. However, the degree to which kit foxes can disperse between populations in this highly fragmented landscape is currently unknown. Although previous surveys at the Naval Petroleum Reserves in California (NPRC) suggest that kit foxes do not have large home ranges and only disperse an average of 8 km from their home range, one individual was recorded moving over 120 km (Scrivner *et al.* 1987). However, dispersal mortality rates can be as high as 65% (Koopman *et al.* 2000), showing that dispersal can be very costly even before an individual attempts to find a mate. This has resulted in several

Figure 4-1. Habitat changes in the San Joaquin Valley over the last 120 years. Maps were generated using the Software: Arc Info 9 (ESRI) Data Sources: California Department of Water Resources; California Gap Analysis Program; California State University, Stanislaus, Endangered Species Recovery Program; and Water Resources Center Archives at UC Berkeley.



small populations becoming isolated and vulnerable to stochastic demographic events or negative genetic consequences such as inbreeding or a genetic bottleneck that affect the long-term viability of a population (Frankham 1995; Frankham& Ralls 1998). For instance, a SJKF population at Camp Roberts is thought to have either gone extinct or if there are a few remaining individuals, they may be experiencing Allee effects (U.S. Fish and Wildlife Service 2010).

In the last 40 years, the largest SJKF populations have been well studied using more traditional methods (e.g. (Cypher *et al.* 2000; Ralls& Eberhardt 1997; Warrick& Harris 2001) (Briden *et al.* 1992) while the smaller or more sparse populations are typically surveyed using non-invasive methods, either conservation detection dog-handler teams (Smith *et al.* 2006a; Smith *et al.* 2006b) or a combination of spotlighting, camera trapping, and track plates (Bidlack 2007; Clark 2007). To date, the only study that has used genetics to assess connectivity between SJKF populations determined that they had low genetic population differentiation with moderate to high levels of migration between them and a weak correlation of isolation by distance (Schwartz *et al.* 2005). However, this study was unable to untangle whether the observed pattern was due to historical or contemporary gene flow (Schwartz *et al.* 2005). This was attributed to low sample sizes available from the central to northern range of the SJKF and to the low levels of polymorphism in the microsatellite loci that were utilized. In this study, we aimed to further explore the question of connectivity between populations in the San Joaquin Valley by utilizing genetic data from larger sample sizes and more locations of the Carrizo Plain and the Ciervo-Panoche areas, and by using a panel of 11 highly

polymorphic tetranucleotide loci. We used a Bayesian approach to examine population genetic structure, derived temporal effective population sizes, and measured the magnitude and directionality of historical and contemporary levels of migration in this metapopulation across a human-altered landscape.

Methods

Samples

We collected tissue and non-invasive samples from a total of 795 individuals during numerous projects from across the current range of the San Joaquin kit fox in the Central Valley of California. We selected 91 individuals from representative scat samples detected during non-invasive surveys conducted by conservation detection dog handler teams in 2009-2011 in the Ciervo-Panoche core area, San Benito County (Wilbert et al. *in prep*). We also collected tissue samples from 32 individuals trapped in 1988-1991 on Soda Lake Road (White& Ralls 1993) and 29 individuals trapped in 1998 on Elkhorn Road (Bean 2002) in the Carrizo Plains National Monument. Representative scat samples were also used from 45 individuals identified on the Topaz Solar Farm project during 2009-2011 (Wilbert et al. *in prep*). We collected ear tissue or hair samples during surveys conducted from 1989-2009 in Bakersfield, Camp Roberts Army National Guard Training Site (Camp Roberts), and Lokern. In total, the study includes 453 individuals from Bakersfield, 8 from Camp Roberts, 105 from the Carrizo Plain, 138 from the Lokern area, and 91 from the Ciervo-Panoche area (hereafter referred to as Panoche).

Figure 4-2. Map showing the sampling localities for San Joaquin kit foxes (*Vulpes macrotis mutica*).

Molecular Methods

DNA Extraction – We extracted DNA from scat samples using the QIAamp DNA stool mini kit (QIAGEN) with modifications as described in Wilbert et al (*in prep*). We also extracted DNA from ear punches that were collected from animals captured for ear-tagging during monitoring surveys. We used approximately ½ of a 3mm diameter circle of tissue for each DNA extraction with a QIAGEN tissue kit. Occasionally, we used a clump of hairs that were also taken during animal trapping. We placed approximately 5-10 hairs consisting of various qualities (texture, length, and follicle presence) in a 2mL tube with lysis buffer overnight and extracted DNA the following day. All extractions

were carried out in a separate room under quasi-clean conditions to prevent contamination. Negative controls (no DNA material added to the extraction) accompanied each set of extractions and were used to check for contamination.

Genotyping – We genotyped DNA samples using eleven microsatellite tetranucleotide repeat loci (FH2137, FH2140, FH2226, FH2535, FH2561, Pez19; Smith et al. 2006; AHTh171, FH2054, FH2328, FH2848, and Ren162) that have been developed from domestic dogs (Francisco *et al.* 1996) and proven to work reliably for kit foxes (Schwartz *et al.* 2005; Smith *et al.* 2006a) and to be highly polymorphic in other canids (Spiering *et al.* 2010). For tissue and hair samples, we optimized PCR conditions (reagent amounts and cycling) so that two or three loci could be multiplexed during amplification and still maintain reliable microsatellite profiles for kit fox samples (Table 4-1). We amplified microsatellites in 10 µl volumes as follows: 1.0µL of 10x PCR buffer, 1.0µL of 10µM dNTPs (2.5µM each; Invitrogen, Carlsbad, CA, USA), 1.0µL of $MgCl_2$ (25mM), 1.0µL of BSA (25mM), 0.15µL of AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA), 1.25µL of template DNA, labeled forward primers, reverse primers, and sterile water (Table 4-1). The PCR profile for multiplex 1 and 2 included: 10 min at 95° C; 35 cycles of 95° C for 1 min, 58°C or 60°C for 1 min, and 72° for 1.5 min; and 72°C for 30 minutes. The PCR profile for multiplex C, E, and G included: 10 min at 95° C; 12 cycles of 95° C for 1 min, 64°C - 1°C/cycle for 1 min, and 72° for 1.5 min; 30 cycles of 95° C for 1 min, 54°C for 1 min, and 72° for 1.5 min; and 72°C for 30 minutes. We ran samples 2-3 times with independent PCR reactions to verify alleles. For scat samples,

microsatellites were amplified independently and rigorously checked for errors as discussed in Wilbert et al. (*in prep*).

Multiplex	Primer Sets			Annealing Temp
1	FH2226 (0.25uL)	FH2561 (0.25uL)	Pez19 (0.25uL)	58°C
2	FH2140 (0.25uL)	FH2535 (0.25uL)	FH2137 (0.25uL)	60°C
C	REN105L03 (0.20uL)	REN162C04 (0.20uL)		TD
E	FH2848 (0.20uL)	FH2328 (0.25uL)		TD
G	AHTh171 (0.30uL)	FH2054 (0.20uL)		TD

Figure 4-3. Multiplex primer combinations and conditions.

We amplified all of the PCR products using fluorescently labeled forward primers (TET, HEX or FAM). We added PCR product (1.0-2.5uL) to 9.0μL of 5:100 mix of Gene Scan ROX-500 (Applied Biosytems) and Hi-Di Formamide (Applied Biosystems). We then ran the labeled PCR products on an ABI PRISM* 3130 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA), which allows for a plate of the 384 PCR reactions to be loaded at once, and analyzed using Genemapper® software to determine the size of each fragment.

Analysis

Genetic Diversity – We used GenAlEx (Peakall& Smouse 2006, 2012) to calculate expected (H_E) and observed (H_O) heterozygosity, mean number of alleles (Na), effective number of alleles (Ne), and global estimates of F_{IS}. We tested for differences between H_E and H_O with a χ^2 test in Excel. We tested for departure from Hardy-Weinberg equilibrium with a global test of heterozygote deficiency using the Markov chain method in GenePop 3.4 (Raymond& Rousset 1995). We also used GenePop to test for linkage disequilibrium between pairs of loci. We performed a hierarchical analysis of molecular variance (AMOVA) using GenAlEx to test for the proportion of genetic variation between genetic groups identified using STRUCTURE (see below). The AMOVA calculates the percentage of variance between and within groups and we tested the significance of these values using an F-test. With 2 degree of freedom for the numerator (number of groups, K, – 1) and 1588 degrees of freedom for the denominator (total sample size equals 2 alleles per individual – 2), we calculated the F ratio and probability of variance in Excel. We also calculated three measures of genetic differentiation: the traditional Wright's Fst according to Weir and Cockerham (1984) and two standardized measures - F'st as calculated by the equation of G"st (Meirmans& Hedrick 2011; Nei& Chesser 1983); Nei 1987) and D_{est} (Jost 2008) as calculated by Meirmans and Hedrick's equation 2 (Meirmans& Hedrick 2011). F-statistics and significance levels with 1000 permutations were calculated using GenAlEx (Peakall& Smouse 2006, 2012). While Fst is provided for reference, the measure of F'st is more

useful in detecting demographic effects and Dest is valuable for measuring each population's unique genetic signature (Meirmans& Hedrick 2011)).

Population structure – We tested for signatures of population genetic structure using hierarchical Bayesian clustering analyses with STRUCTURE version 2.3.2 (Pritchard *et al.* 2000). STRUCTURE assigns genotypes to genetic clusters that maximize Hardy-Weinberg and linkage equilibrium. We investigated the likelihood of the number of genetic clusters (K) from 1-10 with 10 independent replicates using a model of admixture ancestry, correlated allele frequency models, and Markov chain Monte Carlo resampling for 400,000 generations following a burn-in of 100,000. We visualized the results using STRUCTURE Harvester (Earl& vonHoldt 2011), and determined the number of clusters by looking at the highest likelihood value (Ln(P|D); Pritchard et al. 2000), the greatest ΔK (Evanno *et al.* 2005), and the overall assignment values given to individuals as represented in the column bar graph of ancestry (Q plot). We averaged the results from the ten replicates of each K using CLUMPP 1.1.2 (Jakobsson& Rosenberg 2007) and the appropriate .ind file created from Harvester. CLUMPP combines the ancestry assignment values for individuals assigned to unique genetic clusters to produce one "averaged" summary of the STRUCTURE analyses. Due to the large dataset, we ran CLUMPP with the "Greedy" algorithm with 1000 permutations and weighting by the number of individuals in each population. For 2nd and 3rd level analyses, we repeated this process for each cluster with more than 2 populations using the same steps except that we only explored K=1-8.

Historical and contemporary estimates of gene flow and population sizes – Historical and current levels of gene flow were analyzed between the major genetic groups of SJKFs within the valley. Historical migration patterns were estimated using Bayesian inference in Migrate 2.3.4 (Beerli 2006; Beerli& Felsenstein 2001) using parallel processing on two clusters including the Smithsonian's High Performance Computing Cluster, Hydra. The full model of migration was estimated, including mutation-scaled migration rates (M) and mutation-scaled effective population size (θ). MIGRATE was run using the default parameters, a Brownian motion model of stepwise mutation for microsatellites and Metropolis-Hastings sampling every 100 steps of the MCMC chains, consisting of 1 cold sampling chain and 4 hot chains (1.0, 1.5, 3.0, and 100,000). Multiple runs were performed with priors informed on previous runs with coalescence assessed by the shape of the posterior distribution of each parameter, the Bezier approximation of log-probability given the data (ln(P|D), the effective sample sizes and the acceptance ratios of each parameter. The final parameters used in MIGRATE included priors for Theta and M (estimated from previous MIGRATE runs), estimated mutation rates for microsatellite loci, and 15 replicates of 20 x 10^6 MCMC steps with a burn-in of 4 x 10^6 MCMC steps. Following estimation of M and θ in MIGRATE, we calculated effective population size using the equation $\theta = 4Ne*\mu$ with a mutation rate of $\mu=10^{-4}$. We also calculated the number of migrants per generation (Nm) as $M=m/\mu$ and $Nm=Ne*m$.

We estimated contemporary migration rates (m) in BayesAss (Wilson& Rannala 2003). In order to achieve effective mixing with acceptance rates between 0.20 and 0.40

(Rannala 2007), we used the following parameters: deltaA = 0.10, deltaF = 0.10, and deltaM = 0.05. For each of 5 replicates, we sampled every 100^{th} step in 10^7 MCMC steps with a 10^6 burn-in using different numbers for the seed generator. We then examined the trace file for convergence and we chose the run with the highest likelihood values. Migration rates represent the proportion of individuals from one population into another. From these migration rates, we calculated the number of migrants per generation using Nm=Ne*m. We estimated current census population sizes (Nc) from surveys over the last 15 years (Cypher *pers. comm.*). Estimates of effective population sizes derived from Nc are controversial (Palstra& Fraser 2012) and opinions vary widely on which methods are the most appropriate for this calculation (Luikart *et al.* 2010). However, for the purpose of estimating migration rates in a contemporary framework, we selected a liberal ratio for Ne/Nc of 1:2, which will provide the high-end estimate for the number of migrants and allow us to evaluate the maximum number of individuals that may be dispersing between populations.

Results

Genetic variation – We found moderate to high levels of genetic variation within populations. Mean number of alleles per population ranged from 3.818 to 9.727, with the number of alleles per loci ranging from 5-27 overall (Table 4-2). All populations except for Camp Roberts, which had a small sample size, had a large number of low frequency alleles and therefore a decrease in effective number of alleles. In each population, except for Camp Roberts, some of the loci were out of HWE or in linkage disequilibrium but no loci were consistently significant across populations. Comparison over all loci showed

highly significant heterozygote deficiency ($p<0.0001$). All three measures of Fst statistics were significant ($p<0.001$) between all 5 populations (Table 4-3a). F'st and Dest were very similar in magnitude (Supp. Table 1) and this may suggest that differentiation between populations has been influenced by mutations accumulated over time resulting in unique allelic diversity (which Dest is more sensitive to) as well as a loss of demographic movement between populations (which F'st is more sensitive to). Fst values are much smaller in magnitude but show similar trends and can be compared more easily to other research.

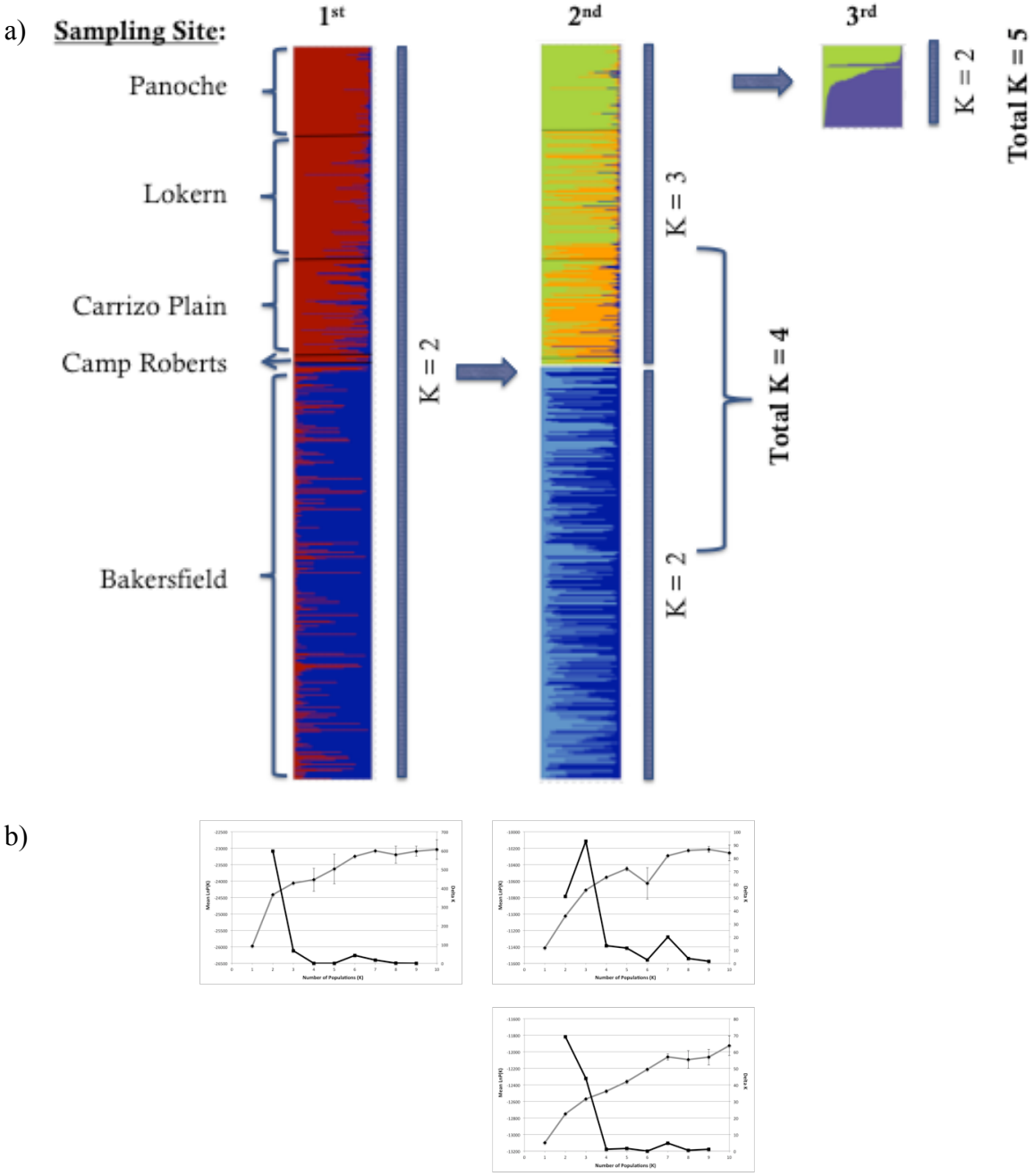
Table 4-1. Genetic diversity statistics. N is the sample size for genetic group. Na is the mean number of alleles per locus with the range in number of alleles in parentheses. Ne is the effective number of alleles per locus. Ho is the observed heterozygosity. He is the unbiased expected heterozygosity. Standard error (SE) is one standard error from the mean. *Populations out of HWE.

	N	Na	Ne (SE)	Ho (SE)	He (SE)
Bakersfield*	453	9.73 (4 - 24)	3.334 (0.544)	0.620 (0.039)	0.638 (0.041)
Camp Roberts	8	3.82 (2 - 9)	2.965 (0.497)	0.703 (0.084)	0.611 (0.067)
Carrizo Plain*	105	8.46 (3 - 16)	3.854 (0.609)	0.651 (0.040)	0.686 (0.044)
Lokern*	138	8.82 (5 - 18)	3.975 (0.615)	0.684 (0.031)	0.709 (0.030)
Panoche*	91	7.91 (4 - 13)	3.132 (0.316)	0.575 (0.055)	0.641 (0.048)
Total	795	7.75 (5 - 27)	3.132 (0.316)	0.647 (0.024)	0.657 (0.021)

Genetic clusters – Bayesian clustering analysis in STRUCTURE revealed a strong signature of subdivision within SJKF population genetic diversity. The 1st level of population structure analysis with all individuals (n=795) yielded the highest support for

2 clusters with the greatest ΔK value (596.9) (Evanno *et al.* 2005), even though overall likelihood values continued to increase with K values greater than 2 (lnP(X|2)=-24416.4 and lnP(X|7)=-23079.9). A second smaller peak in ΔK (42.6) was seen at K=6 but we chose to use the most parsimonious K value and further investigate genetic population structure using hierarchical analysis (Figure 4-3b). Therefore, we used K=2 for level 1 analysis with cluster 1 (blue in Figure 4-3a) closely corresponding to Bakersfield and cluster 2 (red in Figure 4-3a) to the remaining sample localities. We split the data set into two groups for the 2nd level of population analysis: the Bakersfield samples (n=453) and the rest of the locations (n=342). STRUCTURE analysis of Bakersfield samples supported further division into 2 clusters (light blue and dark blue in Figure 4-3a) with the highest ΔK value (69.2). The graph of mean lnP(X|K) continued to gradually increase with increasing numbers of clusters (Figure 4-3b). The analysis of the remaining locations (Camp Roberts, Carrizo Plains, Lokern, and Panoche) showed separation into 3 clusters with a sharp increase in ΔK (92.8) and a lnP(X|3)=-10708.9 (Figure 4-3b). Larger numbers of clusters had larger likelihood values but the most parsimonious solution was 3 clusters. The bar plots of assignment of each individual's ancestry (Figure 4-3a) showed that the blue cluster from the 2nd level analysis closely resembled the ancestry signature of the blue cluster of the same individuals from the 1st level analysis, leading to the conclusion that the blue cluster from this group was referencing genetic signatures from Bakersfield. Bar plots from K=3 and K=4 from the 1st level analyses confirmed this correlation (Supp. Figure 1). Therefore, our total number of clusters from the 2nd level analyses was 4. Because individuals from Panoche showed high assignment

Figure 4-4. Estimated population structure inferred from hierarchical STRUCTURE analysis. a) Bar graphs representing the average ancestry coefficient (q) of each individual for 10 replicates of K clusters. Each individual is represented as a thin horizontal line with sampling locations designated on the far left side. Lines separate the individuals from different sampling locations. b) For each level of STRUCTURE analysis, mean lnP(X|K) (gray line, left y-axis) and ΔK (black line, right y-axis) are plotted for K=1-10.



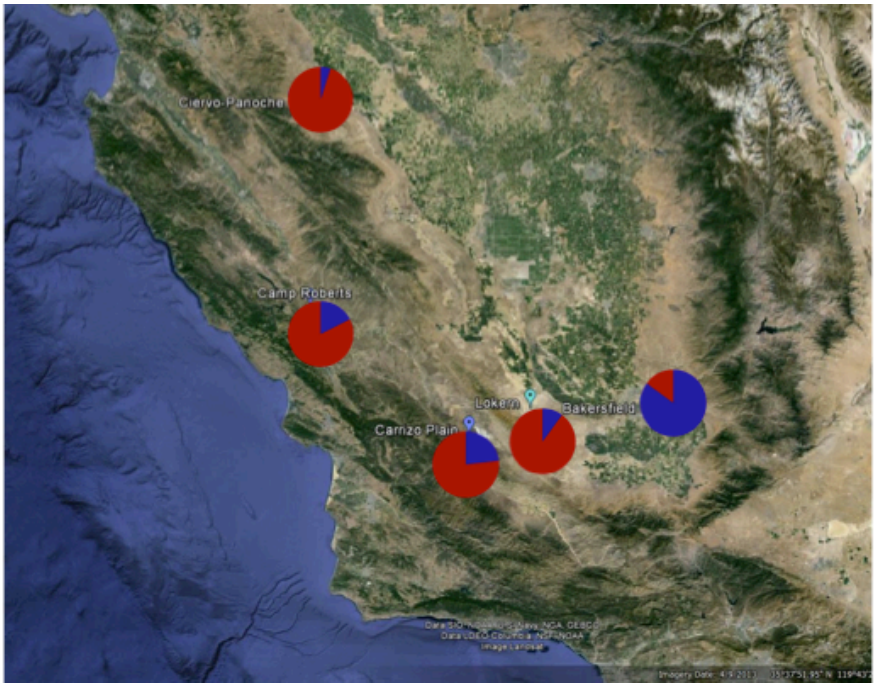
values to the one cluster from the 2nd level analysis (shown in green, Figure 4-3a), we ran a 3rd level of hierarchical STRUCTURE analysis, which further subdivided the Panoche group into two populations, which brings the total number of genetic clusters to 5 (Figure 4-3a).

We created pie charts of the ancestry assignment for individuals in each location from our 1st level STRUCTURE analysis to view the percentage of mixed ancestry in each location (Figure 4-4a). The bar plots of assignment in Figure 4-3a and these pie charts revealed high levels of genetic admixture between the Camp Roberts, Carrizo Plain, and Lokern populations. We therefore chose to conduct further analyses with these locations grouped together and designated them as the Camp Roberts, Carrizo Plain, and Lokern region (hereafter abbreviated CR-Car-Lok). In addition, for this study of regional connectivity, we used the 3 regional locations that corresponded to the 3 most parsimonious regional clusters: 1) Bakersfield, 2) CR-Car-Lok, and 3) Panoche. The pie charts of the genetic ancestry for each of the regional genetic clusters show clear differences between them (Figure 4-4b). We did not focus on the population subdivisions that we detected within the Bakersfield and Panoche areas. These questions were fully addressed in more focused studies dealing with finer scale population genetic structure within each of these two areas (Wilbert et al in prep).

Historical and contemporary gene flow – The AMOVA detected 8% of variation between and 92% within groups respectively, which was highly significant (one-tailed F-test, $F_{ratio}=38.55$, $F_{2,1587}<0.00001$). All pairwise comparisons of differentiation between groups using Fst, F'st and Dest were significant ($p<0.001$; Table 4-3b), with higher

Figure 4-5. Genetic ancestry of sampling locations. a) 1st level of STRUCTURE analysis (K=2) b) Primary difference between 3 regional genetic clusters.

a)



b)

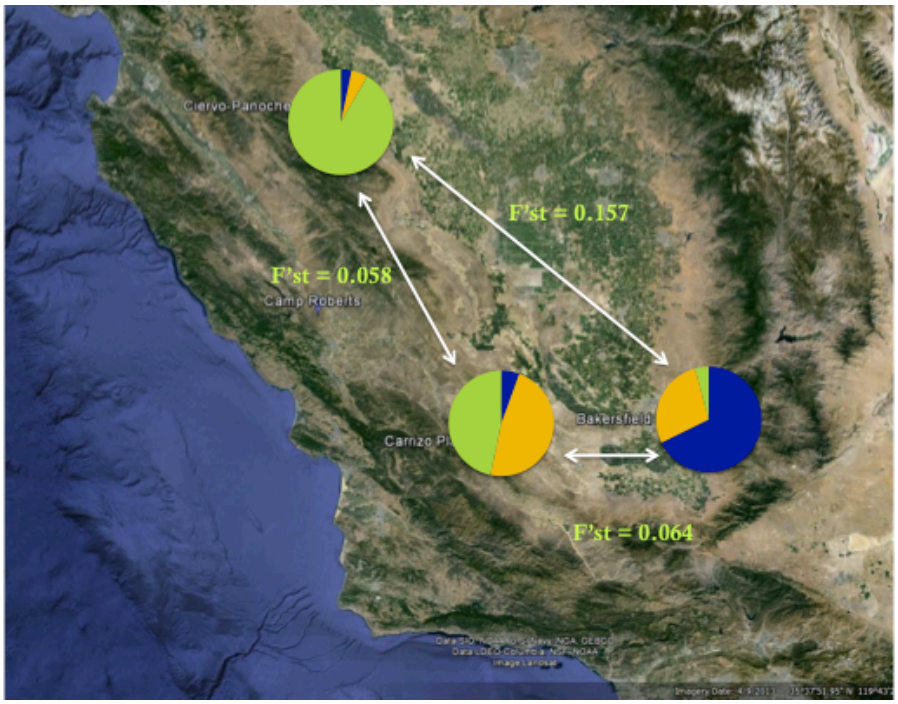


Table 4-2. Measures of population differentiation. Above the diagonal are pairwise Fst estimates and below the diagonal are pairwise Dest estimates. a) Pairwise comparisons between all 5 populations. b) Pairwise comparisons between the 3 regions most likely to represent genetic clusters. Note that the Camp Roberts, Carrizo Plain, and Lokern localities were grouped in a single region and abbreviated as CR-Car-Lok. *p<0.005.

a)

	Bakersfield	**Camp Roberts**	**Carrizo**	**Lokern**	**Panoche**
Bakersfield	-	0.140*	0.057*	0.087*	0.166*
Camp Roberts	0.243*	-	0.069*	0.060*	0.070*
Carrizo	0.108*	0.143*	-	0.022*	0.086*
Lokern	0.174*	0.135*	0.052*	-	0.061*
Panoche	0.294*	0.126*	0.171*	0.130*	-

b)

	Bakersfield	**CR-Car-Lok**	**Panoche**
Bakersfield	-	0.067*	0.166*
CR-Car-Lok	0.136*	-	0.063*
Panoche	0.294*	0.133*	-

estimates of population differentiation between Bakersfield and Panoche.

Historical effective population sizes (θ) and migration rates (M) were estimated from the 33rd of 36 runs in Migrate with acceptance ratios that ranged from 0.083-0.295 and effective sampling sizes that ranged from 212,250-992,211. The posterior distribution converged at one value with a normal distribution for each parameter and the run had the highest Bezier approximation of the log probability of the data for all of the

Migrate runs. Historical θ was 1.376 for Bakersfield (Ne=3400 ± 515), 0.5764 for CR-Car-Lok (Ne=1440 ± 415), and 0.3365 for Panoche (Ne=840 ± 380), with M varying between 3.31-19.11 (Table 4-4, Figure 4-5a). These values translated into historical numbers of migrants that ranged from 0.766 – 6.574 individuals per generation (Figure 4-5a).

Table 4-3. Historical migration rates (M) and mutation-scaled effective population sizes (θ, diagonal values) estimated using MIGRATE (Beerli 2006). Migration rates represent the proportion of individuals that emigrated into the population listed in the column from the population listed in the row. Numbers in parentheses are 95% CI. Camp Roberts, Carrizo Plain, and Lokern region abbreviated as CR-Car-Lok.

θ and M	Bakersfield	CR-Car-Lokern	Panoche
Bakersfield	**1.376 (0.7696-2.0623)**	8.31 (1.81-14.61)	11.51 (0.01-17.81)
CR-Car-Lokern	19.11 (12.01-26.21)	**0.5764 (0.1299-1.1161)**	9.11 (2.21-16.01)
Panoche	3.31 (0.01-8.81)	15.51 (7.41-23.61)	**0.3365 (0.01-0.663)**

Contemporary estimates of migration rates (m) from BayesAss ranged from 0.0024 between Bakersfield and Panoche to 0.0212 between Bakersfield and CR-Car-Lok (Table 4-4). The proportion of individuals assigned to their source population was high amongst all groups, with 97.63% from Bakersfield, 97.53% from CR-Car-Lok, and 98.82% from Panoche (Table 4-4). Population census sizes (Bakersfield Nc=400, CR-Car-Lok Nc=2000, Panoche Nc=125) and a liberal Ne/Nc ratio of 1/2 still yielded

relatively low numbers of migrants per generation, except for large numbers of individuals moving into CR-Car-Lok from Bakersfield and Panoche (Figure 4-5b).

Table 4-4. Contemporary migration rates (m) calculated using BayesAss (Wilson& Rannala 2003). Migration rates represent the proportion of individuals that emigrated in the population listed in the column from the population listed in the row. Values in bold along the diagonal are the proportion of individuals from the source population. All values are the mean of the posterior distribution ± standard deviation. Camp Roberts, Carrizo Plain, and Lokern region abbreviated as CR-Car-Lok.

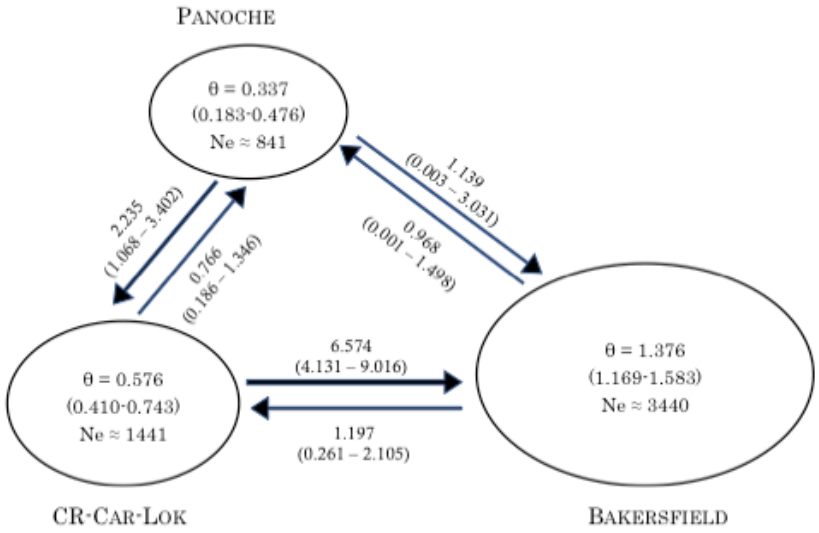
m	Bakersfield	CR-Car-Lokern	Panoche
Bakersfield	**0.9763 ± 0.0064**	0.0212 ± 0.0064	0.0024 ± 0.0019
CR-Car-Lokern	0.0201 ± 0.0096	**0.9757 ± 0.0103**	0.0042 ± 0.0039
Panoche	0.0048 ± 0.0047	0.0070 ± 0.0064	**0.9882 ± 0.0078**

Discussion

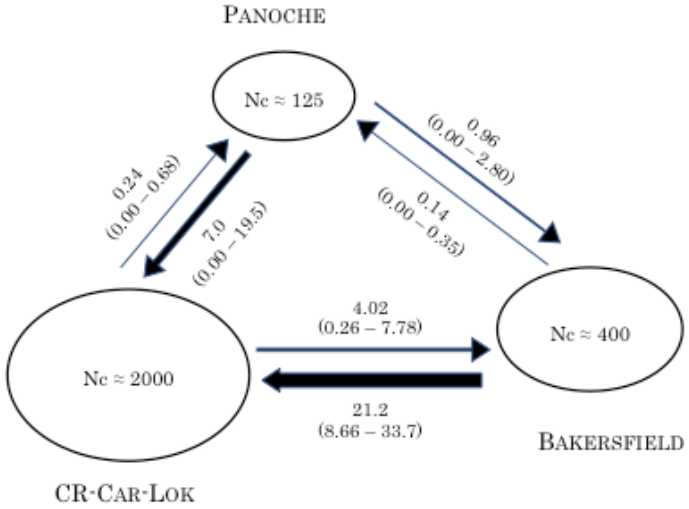
Overall, most of the populations had high levels of allelic diversity with lower numbers of effective alleles (Table 4-2), which suggests that these populations had higher levels of genetic diversity than expected if they had been subject to the effects of genetic drift during a bottleneck. The notable exception was the Camp Roberts population that had a very small sample size compared to the rest (N=8), which undoubtedly had some influence in the lower allelic diversity values that we detected. The inconsistent Hardy-Weinberg and linkage disequilibrium between loci may suggest that some of the populations are under a Wahlund effect (Wahlund, 1928 #1808), which

Figure 4-6. Number of migrants per generation and population sizes of regional groups. a) Historical effective population sizes (θ) estimated using MIGRATE (Table 4-3). Number of migrants per generation (Nm) calculated from historical migration rates using MIGRATE (M), effective population sizes (Ne) calculated from θ, and a mutation rate of $\mu=10^{-4}$. b) Contemporary numbers of migrants per generation (Nm) were calculated with migration rates (m) from BayesAss (Table 4-4), a theoretical Ne/Nc ratio of 1:2, and current census population sizes (Nc) that were based on surveys over the last 15 years. With this ratio of Ne/Nc, these estimates of Nm estimate the maximum number of individuals dispersing per generation. Camp Roberts, Carrizo Plain, and Lokern region abbreviated as CR-Car-Lok.

a)



b)



results from distinct gene pools not mating randomly, affecting allelic frequencies and leading to heterozygosity deficit (Hartl& Clark 1997). Nevertheless, we observed much higher levels of heterozygosity and allelic diversity but similar values of effective number of alleles (3-4 alleles/locus) to those detected in a previous study that used a different panel of microsatellite loci and smaller sample sizes from some of the same populations (Schwartz *et al.* 2005). The higher levels of genetic diversity and heterozygosity within and amongst populations provided a greater statistical power that allowed us to unravel fine scale patterns of genetic structure and highlights the benefit of utilizing a panel of highly polymorphic microsatellite loci and an extensive sample size that was only possible to obtain through collaborative research over a long period of time.

We used a conservative approach in defining the number of clusters using STRUCTURE and a hierarchical analysis to explore the patterns of genetic admixture of kit foxes in the San Joaquin Valley (SJV). We found the highest support for the genetic subdivision of our metapopulation into 3 genetic clusters. SJKF populations from Bakersfield and from Panoche are genetically subdivided from those in the Camp Roberts, Carrizo Plains, and Lokern (CR-Car-Lok) areas. These three localities (CR-Car-Lok) were closest in geographic proximity and also had similar patterns of genetic composition and low levels of genetic structure (Figure 4-3a, Figure 4-4a). Kit foxes in Camp Roberts are geographically isolated from the rest of the SJV by coastal mountain ranges and only a small number of kit foxes may persist in pockets of suitable habitat within these ranges. However, dispersal across these areas may be complicated by the lack of connecting suitable habitat and a high abundance of predators. The topography

between Camp Roberts and the Panoche Valley is very rugged, with multiple ridges within the Diablo Range, making it very difficult for kit foxes to move across this landscape. On the other hand, kit foxes may be able to more easily disperse from Camp Roberts into the Carrizo Plain area using the habitat provided by the interconnecting mountain ravines. Our results show that high levels of connectivity are supported by the genetic similarity between kit foxes from these two areas (Figure 4-3a, Figure 4-4a).

In addition, our results also show high levels of genetic admixture and small Fst and Dest values between kit foxes from the Carrizo Plain and Lokern. While they are in close geographic proximity, the Temblor Mountain Range separates the Lokern area from the Carrizo Plain, and there are two passes with suitable habitat that kit foxes may be able to use to move between them. One of the passes is located along Route 58 in the northeastern edge and the other at the very southern end along Route 33. Lokern is located just east of the northern pass and our high values of contemporary migration rates between Lokern and Carrizo suggest that kit foxes may be actively utilizing the shortest travel route along the northern pass on Route 58 for dispersal. In addition, Lokern is one of the few places in the SJV that still has suitable habitat and this makes it an important area for kit fox connectivity between the Carrizo Plain and other areas of the SJV. However, we should note that our samples were collected in the northern range of the Carrizo Plain and a larger sample from the southern range of the Carrizo plain may be needed to better understand movement patterns along the southern rim of the SJV. The relatively low Fst/Dest values between the Carrizo Plain and Bakersfield (Table 4-3a) coupled with a number kit fox sightings that have been reported in areas near the

southern pass (Cypher *pers. comm.*) suggest that systematic surveys of this area could help determine if the southern pass is also being utilized as a dispersal route by kit foxes from Bakersfield into the southern range of the Carrizo plain.

We grouped kit foxes from Camp Roberts, Carrizo Plain, and Lokern, and obtained estimates of migration and population sizes between the major areas of kit foxes in the SJV. Since these historical estimates are based on 4Ne generations, the estimated timing of these events can date as far back as 20,000 years ago, when the landscape along the central region of the SJV was dominated by lakes, waterways, and other riparian habitats (Matocq *et al.* 2012) that would have been a barrier to kit fox dispersal. The larger areas of natural scrub and grassland habitat surrounding the watershed within the SJV are estimated to have supported much larger kit fox populations, with higher levels of genetic connectivity between them. We estimated historic effective population sizes of 840 ± 380 individuals in Panoche, 1440 ± 415 individuals in CR-Car-Lok, and 3400 ± 515 individuals in Bakersfield. If we double these numbers to obtain census population sizes (with a Ne/Nc=1/2) our total historical is Nc=10,440 ± 2620 and the total contemporary Nc= 2525. These estimates suggest that there has been a dramatic decline of 76% of the census population. With over 90% of habitat loss, a dramatic decrease in population size is not surprising, especially in the Panoche and Bakersfield areas. Nevertheless, large population sizes have persisted in the Carrizo Plain and Lokern areas primarily due to the protection of natural grasslands and well-managed grazing. With the correlation between habitat and kit fox presence (Cypher *et al.* 2013), maintaining

connectivity of these patches will be even more important to prevent further population decline and genetic isolation.

The dramatic changes in habitat and population sizes have also impacted the movement patterns of kit foxes between these areas over time. The numbers of migrants decreased slightly or not at all in both directions between Panoche and Bakersfield (Figure 4-5). It is not surprising that migration rates were historically low and have remained low because of historic separation of the populations due to the valley watershed and more contemporary fragmentation caused by agriculture and other anthropogenic development of the land (irrigation, canals, roads, etc.). We found an increase in migrants from Panoche to the CR-Car-Lok populations over time, and a decrease in individuals moving the opposite direction. With limited habitat suitable for kit foxes connecting these areas, individuals trying to disperse northward would likely have high rates of mortality. Additionally, the limited amounts of suitable habitat would make finding a mate and settling in the Panoche area increasingly difficult. On the other hand, our results show an increase in the number of migrants from Panoche as well as from Bakersfield into the Carrizo Plain, which also contains the largest area of suitable habitat for kit foxes. We see a large number migrants moving from Bakersfield to the Car-Lok area, with 5 times fewer moving backward. Our data correspond to kit foxes moving from their historic range across the SJV into the remaining pockets of habitat where they can survive, which is Bakersfield, the small pockets of space in the foothills, or into the last large piece of remaining habitat in the Carrizo Plain and Lokern areas.

Our 3rd level hierarchical analysis showed two genetic clusters within the Panoche area. The boundary between the two clusters strongly corresponded to a ridgeline that may act as a barrier for kit fox dispersal (Wilbert et al. *in prep*). Other boundaries such as roads, perimeter, fencing, and canals have affected kit fox dispersal (Cypher *et al.* 2009). Further investigation into fine-scale patterns of genetic structure could determine whether kit fox are impeded from moving across these barriers or if kit fox survival is affected by alternate causes (e.g. soil, predators, and agricultural practices).

Our results highlight the importance of recognizing that the kit fox population that now occupies the city of Bakersfield is likely the result of a recolonization of kit foxes. Historically, kit fox distribution was across the entire valley even though populations had limited habitat connectivity due to the presence of the large watershed along the central part of the valley. Our results revealed a distinct genetic signature in individuals from Bakersfield with a historic effective population size over 3,000 individuals and no signature of a bottleneck (Figure 4-3a, Figure 4-5). It is likely that kit foxes occupied the eastern side of the valley that had scrub and grassland habitat, including the location where Bakersfield was built. When the city was established, kit fox could inhabit suitable habitat in the surrounding areas and may have been or seemed absent from the city. However, with increased agriculture over the last 100 years, kit fox recolonized the city of Bakersfield, which has been steadily growing and providing an abundance of resources that kit foxes could utilize. The founders of this population were likely residents of the much less populated historical Bakersfield area or its surrounding areas, and were likely pushed to inhabit the city due to the inhospitable anthropogenic-induced

destruction of their native habitat. In the process, kit foxes successfully adapted to an urban environment (Bjurlin *et al.* 2005; Ralls *et al.* 2007a, Gehrt et al. 2010) but they are native SJKF from the eastern side of the valley. The adaptations associated with living in an urban environment would not be revealed using neutral microsatellite markers and we have yet to look at markers that would elucidate if they have been under strong selective pressures during this period or if they have behavioral plasticity that allowed them to adapt.

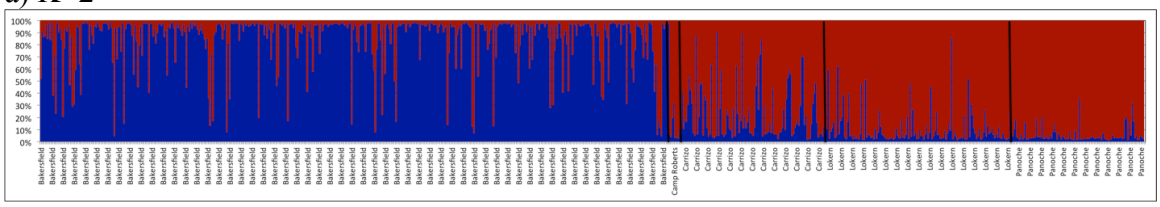
Currently conservation management of the San Joaquin kit fox has identified 3 core areas of protection: the Ciervo-Panoche area, the Carrizo Plain, and the Western Kern area (including Lokern; (U.S. Fish and Wildlife Service 2010). This study shows that the Bakersfield population contains a unique signature of genetic diversity that is absent or underrepresented in other locations and deserves further attention. Areas near Bakersfield and east of the historic watershed should be further explored for future conservation management possibilities. More genetic analysis of kit foxes that may still be present in East Porterville or Pixley National Wildlife Refuge may provide more information on the genetic diversity of kit foxes on the eastern side of the SJV. In addition to these areas, it appears that many of the smaller populations of kit foxes that could have served as intermediates in a stepping-stone population model of these regions may have already gone extinct. However, more intensive population surveys and genetic studies are critically needed to explore areas that may still have small numbers of kit foxes that to date have been undetected. Here we provide a robust genetic framework that can be used to further explore genetic relationships and possible barriers to kit fox

movement in a fragmented landscape. Maintaining connectivity between these regions will be important to prevent inbreeding in small populations and allow for recolonization of available suitable habitats should local extinctions occur.

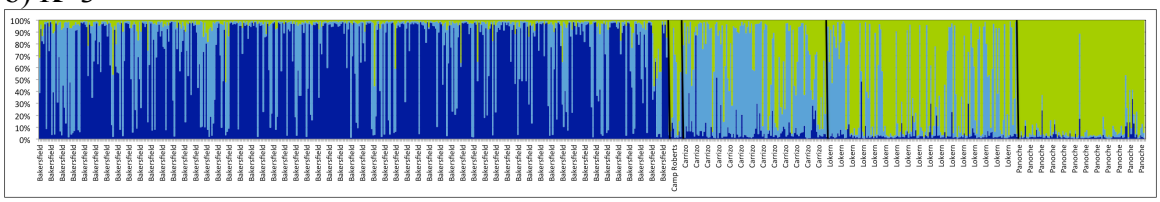
Supplementary Materials

Supp. Figure 1. Potential numbers of populations with STRUCTURE at the 1st level (n=795).

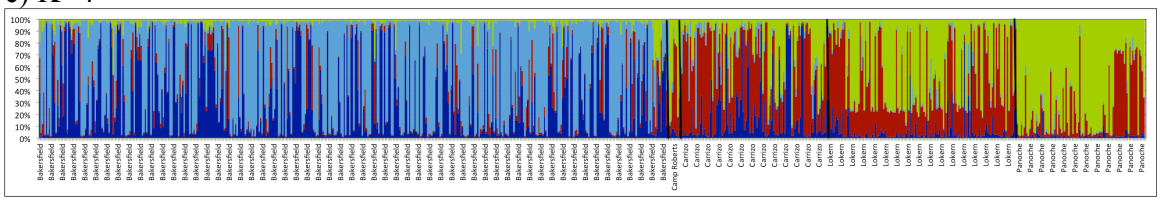
a) K=2



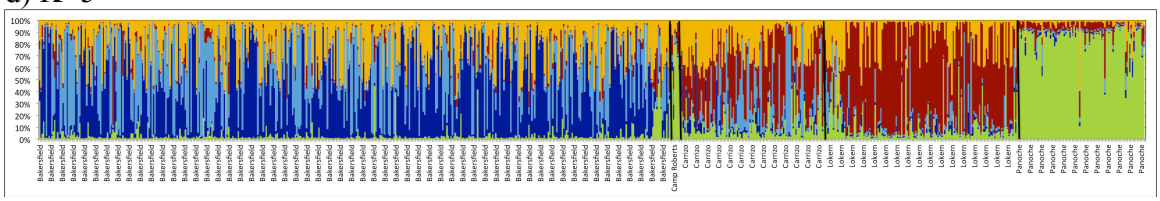
b) K=3



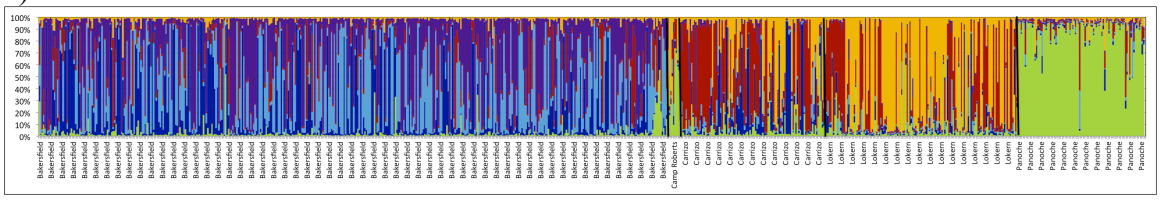
c) K=4



d) K=5

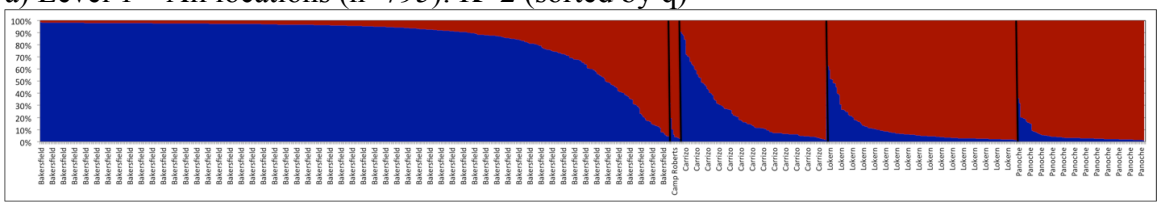


e) K=6

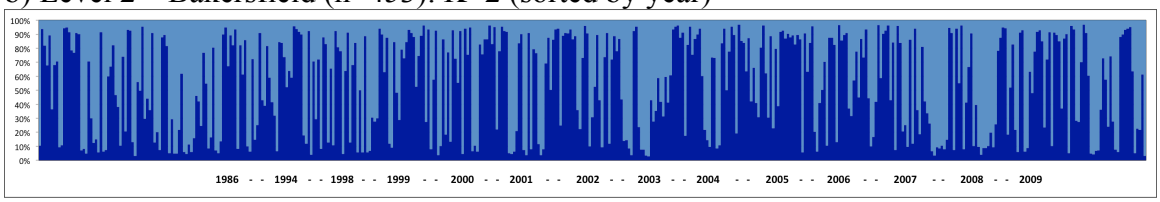


Supp. Figure 2. Overview of Hierarchical Structure Analysis

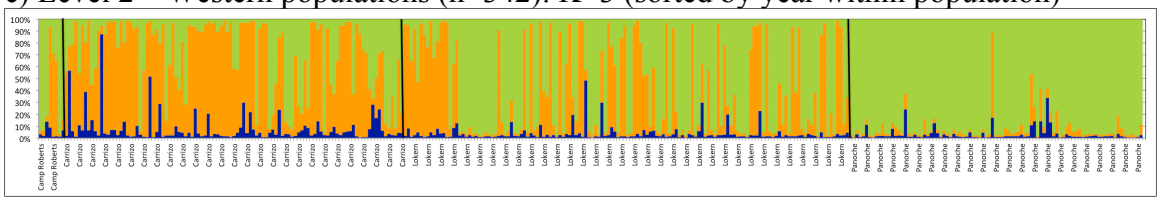
a) Level 1 – All locations (n=795): K=2 (sorted by q)



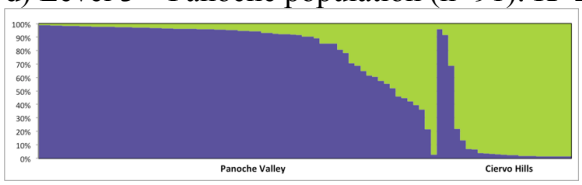
b) Level 2 – Bakersfield (n=453): K=2 (sorted by year)



c) Level 2 – Western populations (n=342): K=3 (sorted by year within population)



d) Level 3 – Panoche population (n=91): K=2 (sorted by q)



Supp. Table 1. Measures of population differentiation. Above the diagonal is pairwise **F'st** estimates and below the diagonal is pairwise Dest estimates. Camp Roberts, Carrizo Plain, and Lokern region abbreviated as CR-Car-Lok. Similar magnitude of F'st and Dest values may indicate that both temporal connectivity as well as demographic movements contributed to degrees of differentiation. *p<0.005

a) Pairwise comparisons between all 5 populations.

	Bakersfield	**Camp Roberts**	**Carrizo**	**Lokern**	**Panoche**
Bakersfield	-	0.340*	0.155*	0.239*	0.394*
Camp Roberts	0.243*	-	0.204*	0.191*	0.188*
Carrizo	0.108*	0.143*	-	0.073*	0.236*
Lokern	0.174*	0.135*	0.052*	-	0.181*
Panoche	0.294*	0.126*	0.171*	0.130*	-

b) Pairwise comparisons between the 3 regions.

	Bakersfield	**CR-Car-Lok**	**Panoche**
Bakersfield	-	0.189*	0.394*
CR-Car-Lok	0.136*	-	0.186*
Panoche	0.294*	0.133*	-

CHAPTER FIVE: MULTI-TAGGED PYROSEQUENCING REVEALS HIGHLY POLYMORPHIC MHC GENES IN THE ENDANGERED SAN JOAQUIN KIT FOX

A paper to be submitted to the *Journal of Evolutionary Biology*

Collaborators: Masoumeh Sikaroodi[1], Brian Cypher[3], Christine Van Horn Job[3], Katherine Ralls[2], Jesus E. Maldonado[2], and Patrick M. Gillevet[1]

[1]George Mason University

[2]Smithsonian Conservation Biology Institute

[3]Endangered Species Recovery Program, CSU-Stainslaus

Introduction

Wildlife is profoundly affected by increasing urbanization, with loss and fragmentation of habitat. A limited number of species have been found to use the urban landscape (Randa & Yunger 2006), and little is known about how these species behave in the urban habitat and whether they are losing genetic variation over time. Urban areas may become an alternative refuge for species, such as the endangered San Joaquin kit fox (SJKF; *Vulpes macrotis mutica*), that are losing wild habitats (Parris & Hazell 2005). The SJKF has decline to approximately 3,000 individuals with the contraction of populations associated to loss of 90% of its habitat (U.S. Fish and Wildlife Service 2010). However, kit foxes have been increasing in number over the last 30 years in the city of Bakersfield.

Kit foxes are opportunistic foragers with preferences for small mammals like rodents and rabbits but benefit from increased resources in the city such as water and anthropogenic food sources (Bjurlin *et al.* 2005; Newsome *et al.* 2010). Annual mortality in the Bakersfield population is only 30% compared to about 50% in non-urban habitats, with the primary causes of death being motor vehicles in the city and coyotes in the wild (Bjurlin *et al.* 2005). This urban population has a higher density of kit foxes than the wild populations and an increased proportion of trios (one male and two females) instead of pairs within a territory. Urban kit foxes live in long-term monogamous pairs within small territories and have smaller, more overlapping home ranges (Ralls *et al.* 2007b; Ralls *et al.* 2001). Kit foxes are adapting well to the urban environment, but the increased density may also increase disease transmission.

Domestic and wild carnivores increase the risk of infectious disease transmission to and from urban wildlife. A rapid decline of island foxes on Santa Catalina Island, California, in 1999-2000 was likely caused by canine distemper virus (CDV) from a raccoon accidentally introduced to the island or by an infected dog (Clifford *et al.* 2006; Timm *et al.* 2002). Disease outbreaks have increased in wild carnivores worldwide (Woodroffe *et al.* 2004) with domestic dogs spreading rabies to highly endangered Ethiopian wolves and African wild dogs (Laurenson *et al.* 2004). With urban kit foxes being in close contact with domestic and urban carnivores, they may be vulnerable to canine diseases such as rabies, distemper (CDV), and parvovirus (CPV), all of which have been documented in kit foxes (Miller *et al.* 2000; Standley & McCue 1992; White *et al.* 2000).

Genetic variation in the immune system has been shown to affect the survival of individual foxes, especially when pathogen evolution is occurring rapidly or in areas with high pathogen loads, such as urban environments (Hedrick *et al.* 2003). The major histocompatibility complex (MHC) is a highly variable part of the genome that codes for immune system genes. MHC polymorphism may be created through a number of mechanisms, including: pathogen-mediation, over-dominance selection – also called heterozygote advantage (Hughes & Nei 1988), negative frequency dependent selection (Tollenaere *et al.* 2008), or disassortative mating (Milinski 2006). These processes may also produce parasite resistance or inbreeding avoidance which have direct and/or indirect effects on the fitness of progeny (Landry *et al.* 2001) and could be very important for urban kit foxes in high densities. Extensive behavioral, pedigree, and population data are needed to simultaneously compare these alternative hypotheses (Piertney & Oliver 2006) and this information has already been collected for the San Joaquin kit fox.

Next-generation sequencing methods have made MHC studies much more feasible (Babik *et al.* 2009; Wegner 2009). To date, MHC genes have been characterized in only four canid genera (*Urocyon*, *Canis*, *Lycaon* and *Vulpes*; (Aguilar *et al.* 2004; Hedrick *et al.* 2002; Hedrick *et al.* 2000b; Kennedy *et al.* 2007; Marsden *et al*. 2012; Ploshnitsa *et al.* 2012; Seddon & Ellegren 2002, 2004). This study is the first to characterize patterns of MHC genetic diversity and natural selection in kit foxes (genus *Vulpes*), expanding the understanding of MHC diversity within the family Canidae and in mammals in general. Therefore we first developed methods to amplify three class II MHC genes (DQA, DQB, and DRB) in the San Joaquin kit fox with Sanger sequencing

and then multi-tagged pyrosequencing (MTPS; (Gillevet *et al.* 2010). We then explored genetic variation of these genes in the Bakersfield kit fox population and wild reference populations in the nearby Lokern and Carrizo Plain, California, that are likely experiencing different selection pressures with respect to disease. We compared levels of MHC diversity between populations and to other species as well as looking for evidence of selection.

Methods

Samples

We selected a random subset of tissue samples from Bakersfield (n=48), Lokern (n=24), and the Carrizo Plain (n=24) to optimize our protocols and characterize MHC diversity. We assessed the selected samples for high quality DNA based on high amplification success in another genetic study using microsatellites on the metapopulation of San Joaquin kit fox (SJKF).

Molecular Methods

Amplification & Sequencing – Since no MHC genes had been characterized in the SJKF, or any other foxes at the time, we optimized PCR amplification for three MHC genes (DQA, DQB, and DRB), focusing on exon 2 since it is the variable region of class II genes that translates into functional differences in pathogen recognition. The DRB gene in SJKF did not amplify with primers developed by Wagner et al. in 1996 that are used by most studies of canids (Hedrick *et al.* 2002; Kennedy *et al.* 2002; Kennedy *et al.* 1998; Seddon& Ellegren 2002, 2004; Short *et al.* 2005; Wagner *et al.* 1996) and the study of island fox (Aguilar *et al.*). However Hedrick et al. (Hedrick *et al.* 2000b) had the same

problem with amplification in Mexican wolves and designed new primers for DRB based on dog sequences from the homologous gene described in Kennedy et al. (1998). The DRB gene amplified from SJKF samples using primers Hedrick et al. designed (2000b) DM-1: 5'-AAGTCCGAGTGCTATTTCACC-3' and DM-2: 5'-TCGCCGCTGCACCGTGAAGCT-3'). Similarly studies of wolves and dogs used primers for DQB designed from dog sequences in 1996 and 1998 (Kennedy *et al.* 2007; Short *et al.* ; Wagner *et al.* 1996), whereas DQB from island fox was amplified using DQB primers designed for seals (Aguilar *et al.* 2004; Hoelzel *et al.* 1999). DQB in SJKF was amplified using the forward primer designed for seals and used for island foxes (DQB1: 5'-TCGTGTACCAGTTTAAGGGC-3') and the reverse primer designed for wolves by Kennedy et al. (Kennedy *et al.* 2002); DQBR2: 5'-CACCTCGCCGCTGCAACGTG-3'). Finally, DQA in SJKF amplified using a primer set that worked in gray wolves (Kennedy *et al.* 2007) and was developed for dogs (Wagner *et al.* 1996); DQAin1: 5'-TAAGGTTCTTTTCTCCCTCT-3'; DQAin2: 5'-GGACAGATTCAGTGAAGAGA-3'). Because pyrosequencing, or sequencing by synthesis, sequences hundreds of thousands of unique fragments at the same time, we could characterize the three genes for 96 samples in one pooled pyrosequencing run (Babik *et al.* 2009; Wegner 2009). Fusion primers have been shown to reduce chimeric sequences which can be important in a gene complex such as the MHC. Therefore, primers were modified to include a 5' addition of a 19-mer adaptor and 4bp key sequence (TCAG) required for emulsion PCR and 454 sequencing. In addition, we created 96 forward primers that included a unique 8bp barcode (individual sample tag) between the

key sequence and the primer that differed by at least 2 base pairs and did not create any homopolymer sites (Supp. Table 1,2 and 3). Fusion primers were synthesized by Operon (Louisville, KY) with the barcoded forward primer on a plate format at 10nmol concentration to avoid contamination that may be introduced in manipulating primers.

We amplified each gene in 20µl volumes as follows: 2.0µL of 10x PCR buffer, 2.0µL of 10µM dNTPs (2.5µM each; Invitrogen, Carlsbad, CA, USA), 1.75µL of $MgCl_2$ (25mM), 2.0µL of BSA (25mM), 0.25µL of AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA), 3.0µL of template DNA, 0.75µL of 10µM fusion forward primers, 0.75µL of 10µM reverse primers, and 7.5µL of sterile water. A touchdown PCR protocol was used for DRB and DQB amplifications, which consisted of an initial 10 min at 95°C, 16 touchdown cycles of 95°C for 30 s, 45 sec touchdown annealing step (starting at 63°C and reducing by 1°C each cycle), and then 72°C for 45 sec. Then 29 cycles of 95°C for 40 s, 47°C for 55 sec, and 72°C for 55 sec plus a final extension at 72°C for 10 min.

Amplification of DQA required a modified protocol because the DQA gene did not amplify with the fusion primer. Therefore we amplified DQA with the original primer set (non-fusion primers DQAin1 and DQAin2) using the above touchdown protocol with only 10 cycles at the constant annealing temperature of 47°C. Then we cleaned PCR reagents from each reaction using 2.5uL of 1:10 diluted ExoSAP-IT (Fisher Scientific). We used 3 µl of these cleaned products in new PCR with the fusion primers in the same 20µl volume recipe as above with the following conditions: an initial 10 min at 95°C, 20 cycles of 95°C for 40 sec, 47°C for 55 sec, and 72°C for 55 sec, plus a final

extension at 72°C for 10 min. Using this modified protocol, we investigated the presence of DQA amplicons, both with and without the 5' modifications (adaptor etc.), using a dissociation curve on a qPCR machine (Mx300P, Agilent Technologies). In the same dissociation curve, we confirmed our amplicons of DRB and DQB with fusion primers.

We cleaned the PCR products two times with Sera-Mag Speed-beads (ThermoScientific, Waltham, MA) in a PEG/NaCl buffer (Rohland & Reich 2012) and visualized pre- and post-cleanup PCR products on agarose gels using Gel Red (Phenix Research Products, Candler, NC). Samples (96) from each gene were pooled using gel visualization, such that 10uL of most amplicons were added to the pool except when very strong or very weak. Only 5uL was added when amplicons were very strong, and 20uL was added when amplicons were weak. Each gene pool was cleaned again using Speed-beads. The pre- and post-cleanups of gene pools were run on an agarose to confirm the presence of amplicon size and lack of PCR primers and truncated products. Gene pools were then quantified using the fluorescent PicoGreen DNA assay and pooled in equimolar concentrations for emulsion PCR. The pool was sequenced in two separate runs on the GS Junior located in the Microbiome Analysis Center at George Mason University. Each run consisted of clonal amplification in two separate emPCR kits, each with independent sequencing according to manufacturer's instructions.

Data Processing

The data generated from 454 sequencing was analyzed in a number of ways that have been suggested in the literature (Meglécz *et al.* 2011; Neiman *et al.* 2011), with a final analysis based on the stepwise procedure in Galan et al. (Galan *et al.* 2010). This

process was used to filter out as many nontarget genes and sequencing errors. Briefly, the steps are as follows:

- Step 1: Using custom PERL scripts on the MBAC Metabiome Portal, sequences less than 150bp were removed and bases with a quality score less than 9 were trimmed from the 3' end. Sequences from each gene were then separated into three files using the forward primer sequence with no base pair mismatches allowed. We used jMHC to detect variants (unique sequences that are putative alleles) within each gene pool and index the number of sequences for each variant in each sample (Stuglik *et al.* 2011). In this process jMHC uses forward and reverse primers sequences, and sample barcodes to create a cross-referenced SQL database. We discarded variants with 3 or less reads in the entire gene pool due to the chances for sequencing errors and to facilitate analysis.
- Step 2: The T_1 threshold developed by Galan et al. (2010) depends upon the number of copies of the gene, which was not known for SJKF. However almost all of our samples had such deep sequencing after filtering that there were either no reads or hundreds of reads. To have a 99.9% probability of genotyping an individual, we selected T_1=18 for m_v=2, T_1=46 for m_v =4, T_1=74 for m_v =6 (Galan *et al.* 2010). Thus, all samples with reads were kept and would pass the T_1 threshold despite the potential number of variants (m).
- Step 3: Each variant detected by jMHC was aligned in Geneious and grouped based on high similarity and disregarding homopolymer errors. Random G-nucleotide SNPs were prevalent in the dataset so these were generally ignored unless a larger

number of reads or variants identified by jMHC confirmed the SNP. We chose this conservative approach in refining the dataset because of the known error rates in 454 sequencing (Babik *et al.* 2009; Metzker 2009). Reads for all variants within a new group were summed for each individual using excel.

- Step 4: We calculated a number of statistics according to Galan et al. (2010), including: the total number of reads (N), the total number of variants (A), and the number of sequences for each variant j (N_j). We also calculated the following for each individual: number of sequences per individual i (N_i), the number of variants per individual (A_i), the number of sequences for each variant (N_{ij}), the frequency of each variant per individual (F_{ij}) which equals $N_{ij}/N_{i.}$ We created histograms of the number of sequences (N) for each variant frequency per individual (F_{ij}). We compared the number of modes in the histogram to those from rodents with different numbers of gene copies (Galan *et al.* 2010). We also used the histograms to look for the appropriate T_2 threshold for each gene since it is expected to vary based on the number of gene copies and the complexity of the system (Galan *et al.* 2010). A large number of low frequency variants can be assumed to be artifacts due to SNP and indel errors from 454 sequencing (Babik *et al.* 2009; Harismendy *et al.*). After determining the T_2 threshold for each gene, all variants with $F_{ij}<T_2$ were removed.
- Step 5: The remaining variants were aligned in MEGA v.4.1 (Kumar *et al.* 2008) as protein-coding nucleotide sequences and checked for stop codons. Variants were then ordered from the most to least abundant alleles and named using the conventional MHC naming protocol (Klein *et al.* 1990).

Analysis

We assumed that if our bioinformatics pipeline included enough quality checks to ensure that we eliminated individuals without enough sequences for complete genotyping (T_1 threshold), then we should find no correlation between the number of reads per individual (N_i) and the number of alleles per individual (A_i). We used the test statistic of a linear regression to check for a correlation between the number of reads of an individual and the number of alleles identified for that individual. We constructed a minimum spanning network between alleles of each gene using TCS 1.21 (Clement *et al.* 2000). Circle sizes roughly represent the number of individuals that carried each allele with color-coding according to the presence of that allele in individuals in urban (Bakersfield) and natural (Carrizo Plain and Lokern) environments (Figure 5-2, Figure 5-3 and Figure 5-4).

Amino acid positions involved in antigen binding were taken from an MHC study on the arctic fox (Ploshnitsa *et al.* 2012), a sister species to the kit fox. We used MEGA software v. 4.1 (Tamura et al. 2011) to calculate the following statistics: a) frequencies of non-synonymous (d_N) and synonymous (d_S) substitutions in the antigen-binding site (ABS) and non-ABS sites using the Nei and Gojobori (1986) method, with Jukes–Cantor correction (Jukes & Cantor 1969), b) a Z-test to test rejection of two null hypotheses, neutrality ($d_N=d_S$) and positive selection ($d_N>d_S$), with significance using 1000 bootstraps, c) Tajima's D statistic for ABS and non-ABS sites to test for neutrality or evidence of selection (Tajima 1989), and d) nucleotide diversity for ABS and non-ABS sites. We also compared sequences for exon 2 of DQA, DQB and DRB to other

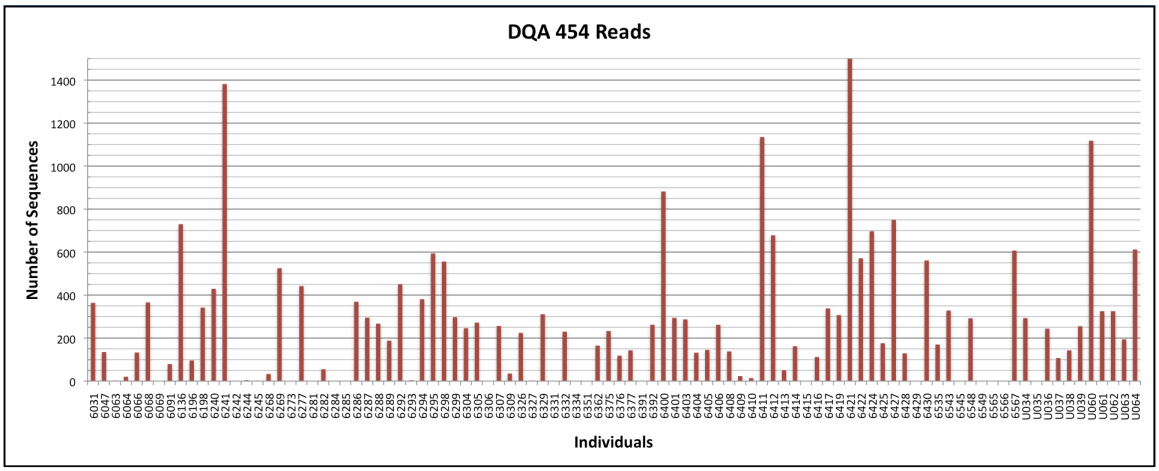
carnivores. We used Geneious to align sequences using MAFFT and create a phylogenetic tree using the neighbor-joining algorithm (Saitou & Nei 1987) with a 50% consensus of 1000 bootstrap iterations. For comparison, the following sequences were downloaded from GenBank – DQA: *Canis familiaris* (U44785.1, U44786.1, U44787.1, U44788.1, U44789.1, U44790.1), *Canis lupus* (DQ777758.1, DQ777759.1), *Ursus arctos* (AB378100.1, AB378101.1, AB378102.1), *Zalophus californianus* (AF502560.1, AF502561.1, AF502562.1), *Ailuropoda melanoleuca* (EF554077.2, EF554076.2, EF554075.2); DQB: *Ursus thibetanus japonicas* (AB763932.1), *Ailuropoda melanoleuca* (GQ496182.1, GQ496188.1), *Canis lupus* (AY126652.1, AY126653.1), *Vulpes lagopus* (HQ602710.1, HQ602709.1, HQ602708.1, HQ602707.1, HQ602706.1, HQ602705.1, HQ602704.1, HQ602703.1, HQ602702.1, HQ602701.1, HQ602700.1, HQ602699.1); and DRB: *Gulo gulo* (JX409655.1, JX409665.1, JX409656.1), *Ailuropoda melanoleuca* (AY895155.1, JF518956.1, GQ496165.1), *Panthera tigris* (DQ189261.1), *Ursus americanus* (AB490480.1), *Canis lupus baileyi* (AY009508.1, AY009509.1), *Canis lupus* (AY694183.1, AY126663.1, AY126661.1), *Vulpes lagopus* (HQ602698.1, HQ602697.1, HQ602696.1, HQ602695.1, HQ602694.1, HQ602693.1, HQ602692.1, HQ602691.1, HQ602690.1, HQ602689.1, HQ602688.1, HQ602687.1).

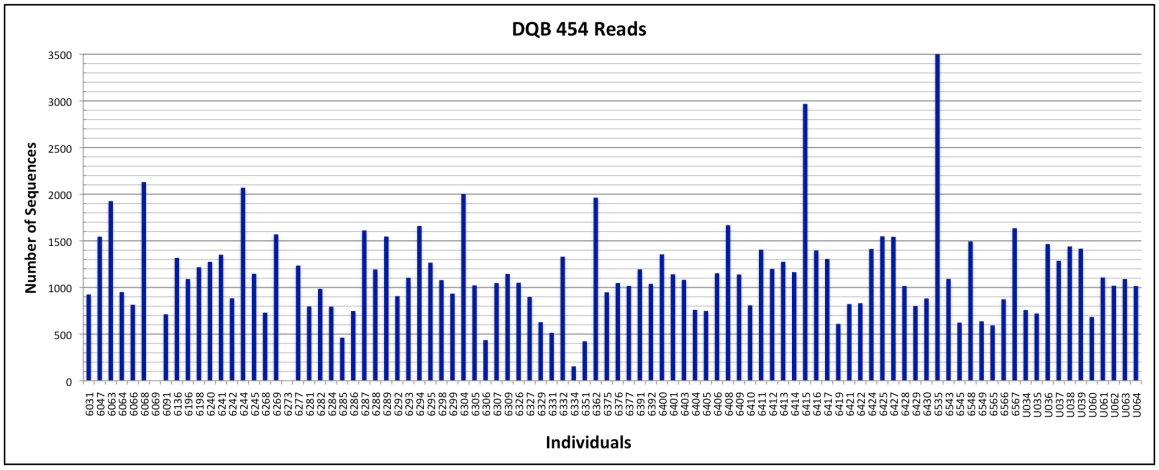
Results

Amplifications of exon 2 of each gene showed strong distinct bands on agarose gels that were absent of primer dimer bands after cleaning with Speed-Beads. Gel imaging and dissociation curves confirmed the modified protocol for DQA contained the larger sized

molecule with the adaptor on the 5' end for pyrosequencing. From both of the 454 GS Jr. runs, we obtained a total of 236,567 sequences using regular processing protocols. After filtering reads less than 150bp and separating them based on the forward primer, we obtained 22,546 reads of DQA, 76,396 reads of DQB, and 93,590 reads of DRB, which equaled a loss of 19% due to short or non-specific reads. Separation of the gene reads into individuals was successful for DQB and DRB but difficult for DQA. Upon inspection of the DQA reads in Geneious, we discovered that the full reverse primer sequence was not present. We removed the first two bases from the reverse sequence used by jMHC and were able to separate reads into individuals. For each gene, 70-80% of sequences were assigned to 75-95 samples (Table 5-1, Figure 5-1). The number of variants identified by jMHC that included a barcode was enormous (374 for DQA, 529 for DQB, and 935 for DRB), with an overall total number of unique sequences (with or without barcodes) identified as 3847, 6883, and 13130 respectively. We plotted the number of sequences based on the frequency of each variant in a sample (F_{ij}) in 10% intervals as well as in smaller increments for variants with F_{ij} less than 10% (Supp. Figure 5-1, Figure 5-2 and Figure 5-3). By looking at the histograms and the number of high frequency alleles within a sample, we used T_1 thresholds corresponding to a duplicated gene in a diploid species for DQA and DQB (46 reads) and a triploid gene in a diploid species for DRB (74 reads; Galan, 2010 #142). Therefore, we removed 10 samples from DQA, 1 from DQB and 0 from DRB, which resulted in a minor loss in reads overall (Table 5-1). Alignment of the barcoded variants revealed a large numbers of random SNPs and homopolymer mistakes that were sequencing errors. Variant

Figure 5-1. Distribution of pyrosequencing reads across genes and individuals





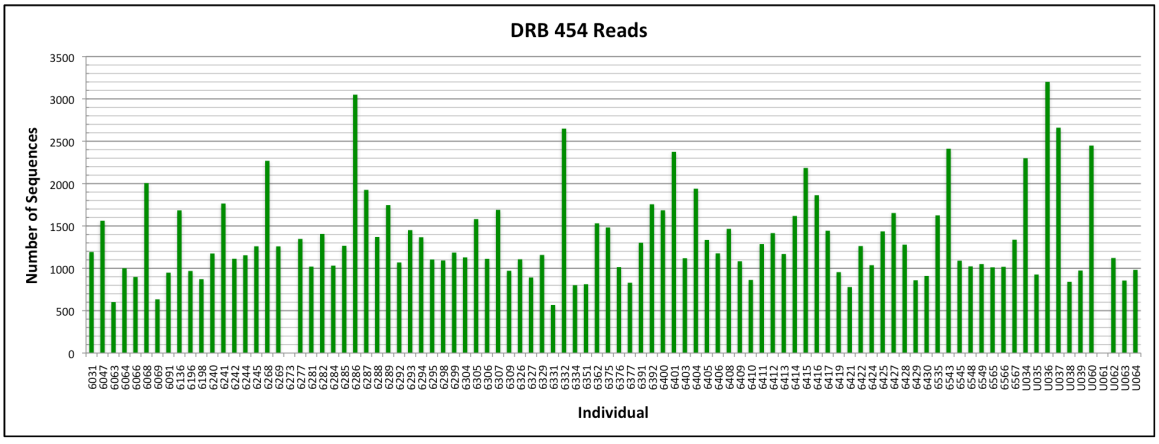


Table 5-1. Bioinformatics Pipeline and Data. Bioinformatics procedures numbered here from 1-4 match steps explained more fully in the methods. Percentages indicate the number of reads retained from that step as compared to the previous step, not the original total.

	454 GS Jr. Regular Processing	1) Galaxy Filtering: >150bp	Percentage
MTPS33	128,064	104,964	82%
MTPS34	108,503	101,885	94%
	236,567 reads	206,849 reads	87%

		DQA	DQB	DRB
1) Galaxy Gene Separation	MTPS33 Reads:	13,952	35,088	46,684
	MTPS34 Reads:	8,594	41,310	46,906
	Total:	22,546	76,398	93,590
1) jMHC Individual Separation	Reads	17,361	61,610	66,164
	% Reads	77.0%	80.6%	70.7%
	Variants (Total)	374 (3,847)	529 (6,883)	935 (13,130)
	Individuals	75	94	95
2) T1 Threshold	Reads	17,173	61,610	66,163
	% Reads	98.9%	100%	100%
	Variants	374	529	935
	Individuals	65	93	94
3) Variant Grouping in Geneious	Reads	16,351	59,368	62,150
	% Reads	95.2%	96.4%	93.9%
	Variants	214	43	56
	Individuals	65	93	94
4) T2 Threshold 5) Translation Check	Reads	13,925	57,650	50,067
	% Reads	85.2%	97.1%	80.6%
	Variants	*11*	*13*	*29*
	Individuals	*65*	*93*	*94*

alignment and grouping in Geneious reduced the number of variants greatly, and made our dataset much more reasonable. From the modal distributions shown in the frequency histograms we chose a T_2 threshold of 5% for DQA, 3% for DQB, and 6% for DRB (Supp. Figures 1, 2 and 3). We did not find stop codons in any of the alleles after being translated into amino acid sequences, and thus, we characterized 11 alleles of DQA from 13,925 reads, 13 alleles of DQB from 57,650 reads, and 29 alleles from 50,067 reads (Table 5-1). The test statistic showed no correlation between the number of reads and the number of alleles (data not shown).

We labeled the alleles *Vuma*-DQA*01-11, *Vuma*-DQB*01-13, and *Vuma*-DRB*01-29 (Table 5-2, Table 5-3 and Table 5-4). Nucleotide sequences will soon be available on GenBank. Minimum spanning networks of DQA and DQB show star shaped formations with a few common alleles shared by most if not all individuals from both urban and natural populations, as well as a few alleles unique to individuals from either urban or natural populations (Figure 5-2, Figure 5-3). DQB has 4 alleles that are quite distant from the rest of the network, with 11bp substitutions connecting them to *Vuma*-DQB*10 and another 12bp substitutions to the major allele (Figure 5-3). For both DQA and DQB, a majority of individuals carried either 1 or 2 alleles, with a handful of individuals having 3 alleles (data not presented). The network for DRB is much more complex with 1-16bp changes connecting alleles and no star formations (Figure 5-4). While individuals in urban and natural environments had many unique alleles (9 and 11, respectively; Table 5-4), four alleles, carried by 45-70 out of 94 total individuals, were found to be very common. Most individuals in each environment carried 3-4 alleles, with

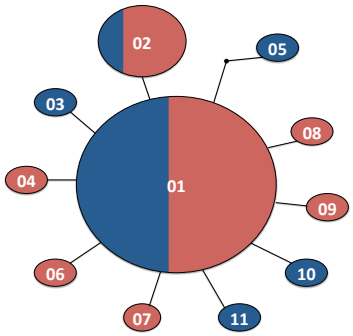


Figure 5-2. DQA Haplotype Network. Circles are generally representative of abundance, with red representing presence in urban individuals and blue in wild wild individuals.

Table 5-2. DQA Alleles. Alleles below the line were found in only 1 individual, and are shaded in red (urban) or blue (wild) to highlight the indicated population.

	BAK	CAR-LOK	OVERALL
	37	32	69
MHC allele	Urban	Wild	
Vuma-DQA*01	34	31	65
Vuma-DQA*02	17	7	24
Vuma-DQA*03	0	1	1
Vuma-DQA*04	1	0	1
Vuma-DQA*05	0	1	1
Vuma-DQA*06	1	0	1
Vuma-DQA*07	1	0	1
Vuma-DQA*08	1	0	1
Vuma-DQA*09	1	0	1
Vuma-DQA*10	0	1	1
Vuma-DQA*11	0	1	1

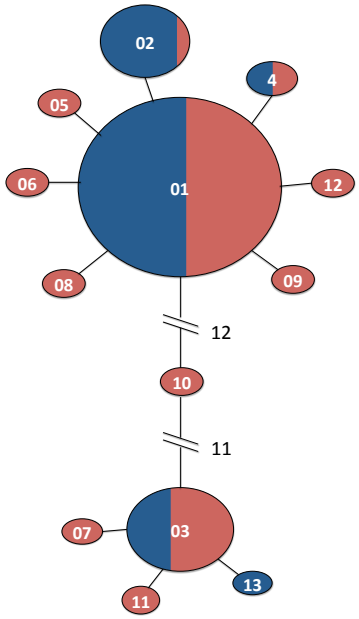


Figure 5-3. DQB Haplotype Network. Circles are generally representative of abundance, with red representing presence in urban individuals and blue in wild wild individuals.

Table 5-3. DQB Alleles. Alleles below the line were found in only 1 individual, and are shaded in red (urban) or blue (wild) to highlight the indicated population.

	BAK	CAR-LOK	OVERALL
	46	47	93
MHC allele	Urban	Wild	
Vuma-DQB*01	46	42	88
Vuma-DQB*02	4	23	27
Vuma-DQB*03	13	9	22
Vuma-DQB*04	1	1	2
Vuma-DQB*05	1	0	1
Vuma-DQB*06	1	0	1
Vuma-DQB*07	1	0	1
Vuma-DQB*08	1	0	1
Vuma-DQB*09	1	0	1
Vuma-DQB*10	1	0	1
Vuma-DQB*11	1	0	1
Vuma-DQB*12	0	1	1
Vuma-DQB*13	0	1	1

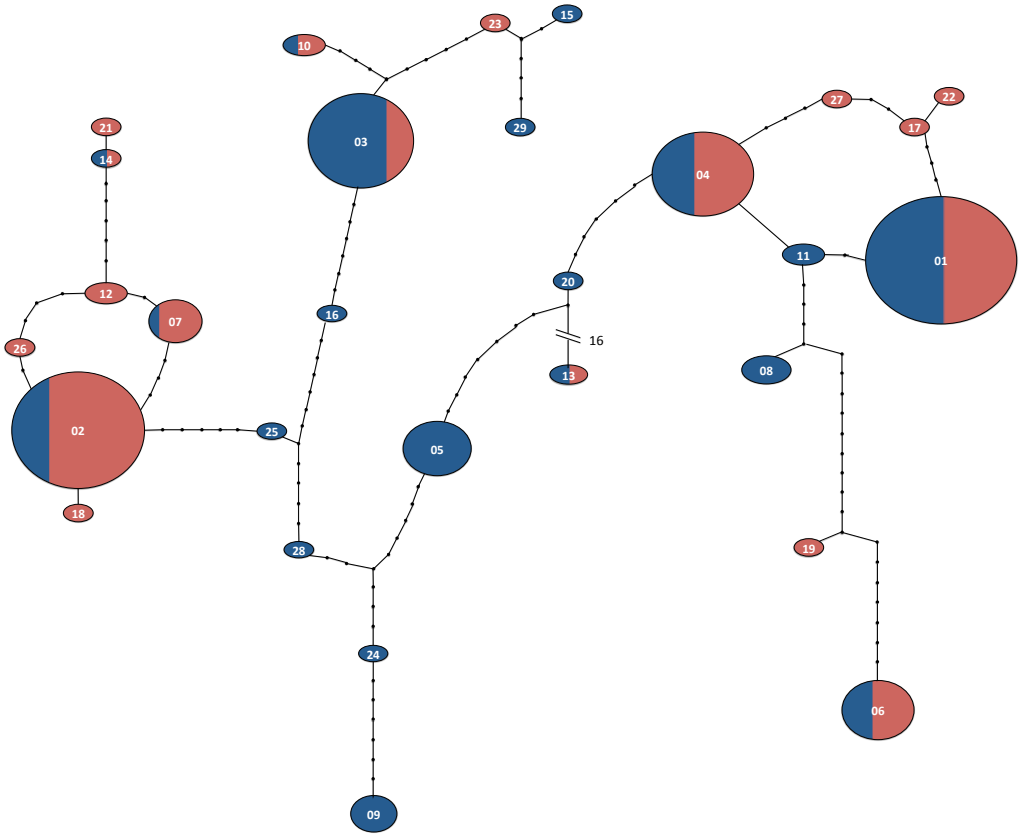


Figure 5-4. DRB Haplotype Network. Circles are generally representative of abundance, with red representing presence in urban individuals and blue in wild wild individuals.

a few carrying 1 allele and 1 individual carrying 6 alleles (data not presented here). Even if all of the rare alleles were removed, the number of DRB alleles per individual would still range from 1-6, which may suggest variation in gene copy number between individuals.

The number of variable nucleotides was 11 for *Vuma*-DQA (out of 288), 36 for *Vuma*-DQB (out of 225), and 47 for *Vuma*-DRB (out of 207). The mean number of pairwise nucleotide differences between alleles was 2.00 (range: 1-3) for *Vuma*-DQA,

Table 5-4. DRB Alleles. Alleles below the line were found in only 1 individual, and are shaded in red (urban) or blue (wild) to highlight the indicated population.

	BAK	CAR-LOK	OVERALL
	47	47	94
MHC allele	Urban	Wild	
Vuma-DRB*01	35	35	70
Vuma-DRB*02	45	17	62
Vuma-DRB*03	14	38	52
Vuma-DRB*04	27	18	45
Vuma-DRB*05	0	21	21
Vuma-DRB*06	10	7	17
Vuma-DRB*07	9	2	11
Vuma-DRB*08	0	5	5
Vuma-DRB*09	0	5	5
Vuma-DRB*10	2	1	3
Vuma-DRB*11	0	3	3
Vuma-DRB*12	3	0	3
Vuma-DRB*13	1	1	2
Vuma-DRB*14	1	1	2
Vuma-DRB*15	0	2	2
Vuma-DRB*16	0	2	2
Vuma-DRB*17	1	0	1
Vuma-DRB*18	1	0	1
Vuma-DRB*19	1	0	1
Vuma-DRB*20	0	1	1
Vuma-DRB*21	1	0	1
Vuma-DRB*22	1	0	1
Vuma-DRB*23	1	0	1
Vuma-DRB*24	0	1	1
Vuma-DRB*25	0	1	1
Vuma-DRB*26	1	0	1
Vuma-DRB*27	1	0	1
Vuma-DRB*28	0	1	1
Vuma-DRB*29	0	1	1

14.00 (range: 1-28) for *Vuma*-DQB, and 17.09 (range: 1-35) for *Vuma*-DRB. The number of variable amino acids was 4/95 for *Vuma*-DQA, 21/75 for *Vuma*-DQB, and 47

for *Vuma*-DRB (Table 5-5). However the diversity in the amino acid sequences only converts to 1 functionally different group of amino acids in the ABS for DQA, 6 for DQB and 25 for DRB, which is most relevant for immunity.

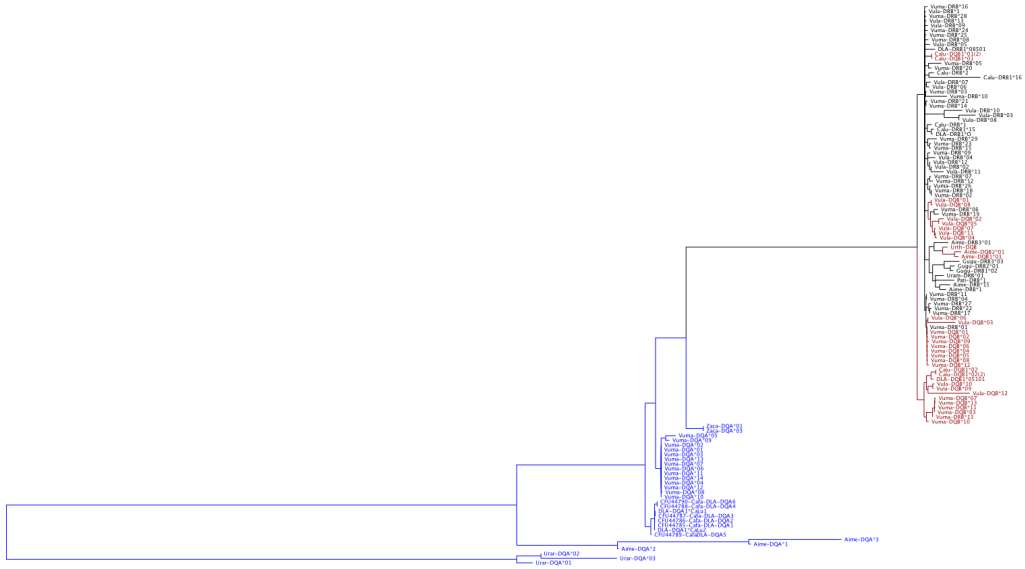
The overall d_N/d_S ratio was 0.545 for *Vuma*-DQA, 0.917 for *Vuma*-DQB, and 1.084 for *Vuma*-DRB, with d_N/d_S ratios 2 and 3 times higher in the ABS for DQB and DRB respectively (Table 5-5). The Z-test for positive selection was not significant for any of the genes, nor could we reject the hypothesis of neutrality for any gene using the Z-test (Table 5-5). We found a significantly negative value of Tajima's D for DQA and significantly positive value for DRB (Table 5-5). While Tajima's D was elevated for

Table 5-5. Rates of substitution and test of selection. S_T = total number of amino acid sites, S_V = number of variable amino acid sites, d_N = rate of nonsynonymous substitutions, d_S = rate of synonymous substitutions, Z = Z-statistic, D = Tajima's D.

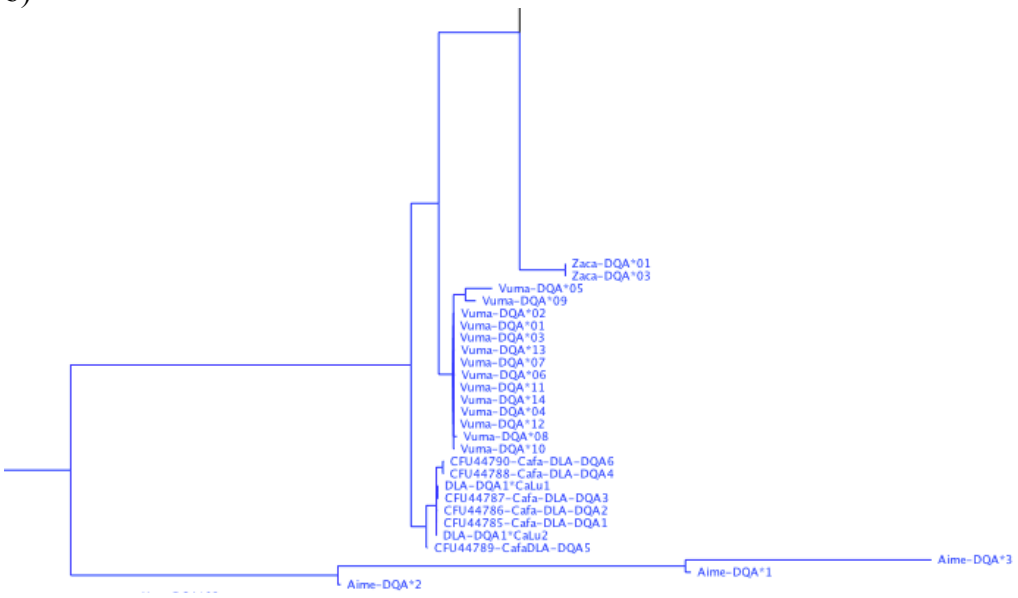
	S_T	d_N	d_S	d_N/d_S	$H_O: d_N=d_S$		$H_O: d_N>d_S$		S_V	D
					Z	p	Z	p		
DQA (n=11)										
Overall	95	0.006	0.011	0.545	-0.884	0.378	-0.882	1.000	4	-1.712
ABS	21	0.000	0.034	0.000	-1.177	0.241	-1.296	1.000	0	n/c
non-ABS	74	0.008	0.007	1.143	0.203	0.839	0.208	0.418	4	-1.712
DQB (n=13)										
Overall	75	0.066	0.072	0.917	0.783	-0.276	1	-0.283	21	1.067
ABS	21	0.165	0.128	1.289	0.550	0.599	0.282	0.578	11	1.338
non-ABS	54	0.034	0.051	0.667	0.467	-0.730	1	-0.740	10	0.630
DRB (n=29)										
Overall	69	0.090	0.083	1.084	0.766	0.299	0.380	0.307	23	1.735
ABS	21	0.275	0.162	1.698	0.161	1.412	0.067	1.508	14	2.820
non-ABS	48	0.024	0.053	0.453	0.239	-1.183	1	-1.180	9	-0.183

Figure 5-5. Phylogenetic tree of DQA, DQB, and DRB exon 2 alleles for the San Joaquin kit fox (*Vuma*) in relation to other species of Carnivora. a) entire Neighbor-Joining consensus tree - displayed for overall pattern, b) close up of the DQA clade, c) close-up of the DQB/DRB clade. DQA alleles are labeled in blue, DQB in red, and DRB in black. *Aime* = *Ailuropoda melanoleuca* (giant panda), *Calu* = *Canis lupus* (grey wolf), *DLA* = *Canis lupus familiaris* (dog), *Gugu* = *Gulo gulo* (wolverine), *Pati* = *Panthera tigris* (tiger), *Uram* = *Ursus americanus* (American black bear), *Urth* = *Ursus thibetanus* (Asian black bear), *Vula* = *Vulpes lagopus* (Arctic fox).

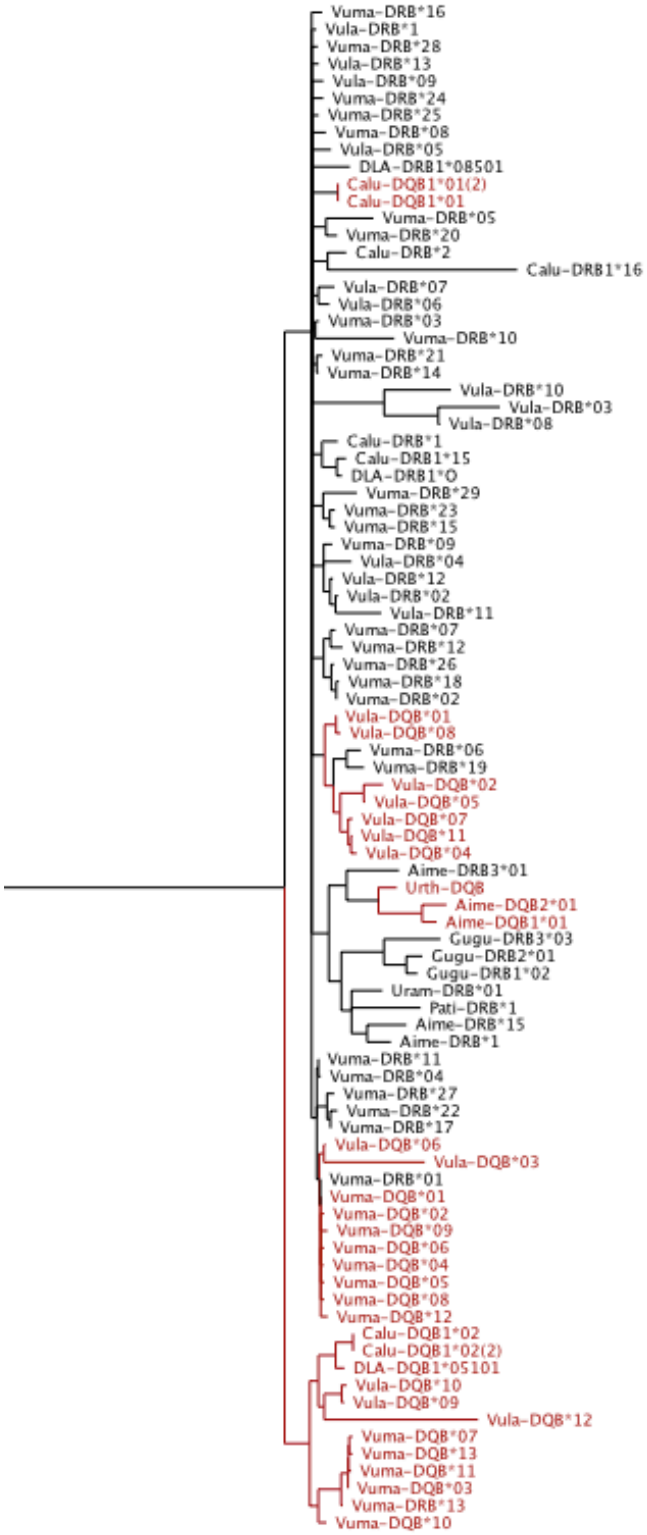
a)



b)



c)


Vuma-DRB*16
Vula-DRB*1
Vuma-DRB*28
Vula-DRB*13
Vula-DRB*09
Vuma-DRB*24
Vuma-DRB*25
Vuma-DRB*08
Vula-DRB*05
DLA-DRB1*08501
Calu-DQB1*01(2)
Calu-DQB1*01
Vuma-DRB*05
Vuma-DRB*20
Calu-DRB*2
Calu-DRB1*16
Vula-DRB*07
Vula-DRB*06
Vuma-DRB*03
Vuma-DRB*10
Vuma-DRB*21
Vuma-DRB*14
Vula-DRB*10
Vula-DRB*03
Vula-DRB*08
Calu-DRB*1
Calu-DRB1*15
DLA-DRB1*O
Vuma-DRB*29
Vuma-DRB*23
Vuma-DRB*15
Vuma-DRB*09
Vula-DRB*04
Vula-DRB*12
Vula-DRB*02
Vula-DRB*11
Vuma-DRB*07
Vuma-DRB*12
Vuma-DRB*26
Vuma-DRB*18
Vuma-DRB*02
Vula-DQB*01
Vula-DQB*08
Vuma-DRB*06
Vuma-DRB*19
Vula-DQB*02
Vula-DQB*05
Vula-DQB*07
Vula-DQB*11
Vula-DQB*04
Aime-DRB3*01
Urth-DQB
Aime-DQB2*01
Aime-DQB1*01
Gugu-DRB3*03
Gugu-DRB2*01
Gugu-DRB1*02
Uram-DRB*01
Pati-DRB*1
Aime-DRB*15
Aime-DRB*1
Vuma-DRB*11
Vuma-DRB*04
Vuma-DRB*27
Vuma-DRB*22
Vuma-DRB*17
Vula-DQB*06
Vula-DQB*03
Vuma-DRB*01
Vuma-DQB*01
Vuma-DQB*02
Vuma-DQB*09
Vuma-DQB*06
Vuma-DQB*04
Vuma-DQB*05
Vuma-DQB*08
Vuma-DQB*12
Calu-DQB1*02
Calu-DQB1*02(2)
DLA-DQB1*05101
Vula-DQB*10
Vula-DQB*09
Vula-DQB*12
Vuma-DQB*07
Vuma-DQB*13
Vuma-DQB*11
Vuma-DQB*03
Vuma-DRB*13
Vuma-DQB*10

DQB, neutrality could not be rejected with this test (Schmidt& Pool 2002). The phylogeny resolved two clades, one with alleles from DQA from all taxa and one with alleles from DQB and DRB from all taxa (Figure 5-5).

Discussion

This study presents the first characterization of MHC genes in the San Joaquin kit fox and provides the foundation for future work into understanding kit fox survival in an urban landscape. DQA and DQB were found to have duplicated genes and moderate levels of allelic diversity while DRB had three gene copies and high levels of allelic diversity as compared to other carnivores (Aguilar *et al.* 2004; Marsden *et al.* 2012; Oomen *et al.* 2013; Ploshnitsa *et al.* 2012; Seddon& Ellegren 2002, 2004). While polymorphism at exon 2 in the MHC genes increases the ability for the immune system to bind a pathogen, polygeny of MHC genes increases the number of alleles that can be expressed by an individual at one time (Janeway Jr. *et al.* 2001). Therefore both of these levels of diversity are important to the functional diversity of SJKF. Overall, SJKF carried similar numbers of alleles as compared to other wild canids. Outside of a putative DQA mRNA sequence from Arctic fox, DQA had only been characterized in European wolves, which showed moderate DQA allelic diversity (9; (Seddon& Ellegren 2002) that is similar to the diversity we found in SJKF (11 alleles). However none of the DQA alleles in SJKF had a nonsynonymous mutation within the ABS (Table 5-5), meaning that all of the alleles are functionally the same. We also found moderate levels of DQB diversity in SJKF (13 alleles) which is comparable to European wolves (10 alleles; (Seddon& Ellegren 2002) and Arctic fox (11 alleles; (Ploshnitsa *et al.* 2012). The d_N/d_S

ratio within the ABS of DQB alleles was above 1 and twice the ratio in non-ABS sites, both of which indicate positive selection, but did not statistically deviate from neutrality by Z statistics or Tajima's D. This could be the result of a low number of alleles and the short sequence within which the rates of substitution are compared. We found higher DRB allelic diversity for SJKF (29 alleles) than what had been found in other wild canids, 9 in Arctic fox (Ploshnitsa *et al.* 2012), 19 in African wild dog (Marsden *et al.* 2012), and 17 in European wolves (Seddon& Ellegren 2002). While the Z-test for positive selection was not significant (p=0.067), Tajima's D indicated a deviation from neutrality (D=2.820), suggesting balancing selection as D statistics above 2 have been shown to be correlated with an excess of common alleles produced by positive selection (Schmidt& Pool 2002). With the high degree of variation and numerous common DRB alleles, it seems likely that DRB has undergone some degree of positive selection that helped maintain this diversity.

MHC genes within canids show varying degrees of diversity that could reflect different selection pressures on SJKF in the past. Lower levels of diversity at DQA or DQB could signify a disease-related selective pressure (Hedrick *et al.* 2003) while higher levels of diversity at DRB may indicate balancing selection through over-dominance (Hughes& Nei 1988) or the divergent allele advantage (Wakeland *et al.* 1990). However since these patterns of diversity, particularly lower in DQA and higher in DRB, are somewhat consistent across carnivores in general, these patterns may also represent genic diversity of the Order Carnivora from millions of years of MHC evolution with gene conversion or recombination creating different numbers of gene copies. From MHC data

of European wolves, polymorphism within DQA (6.4±3.6 million years) was estimated to be younger than DQB (27.2±10.1) and DRB (37.0±11.3; (Seddon& Ellegren 2002). Thus more recent species history along with processes of evolution – drift, mutation, and selection – would dictate the patterns of polymorphism we find within SJKF.

Phylogenetic comparison of MHC alleles from SJKF to other carnivores only separated DQA alleles from DQB and DRB alleles, with alleles from different species completely intertwined. DQB and DRB have similar structures (Klein 1986) and have been shown to have similar nucleotide motifs that suggest intergenic recombination (Seddon& Ellegren 2002). While primer-binding sites are different enough to design specific primers, exon 2 is highly similar between these two loci and displays trans-species evolution (Figure 5-5a) as noted in Hedrick et al. (2000a), and Seddon and Ellegren (2002). However the high allelic diversity of SJKF alleles ("*Vuma*" on Figure 5-5a-c) are spread across the phylogenetic tree, which is interesting considering the small geographic distribution of the SJKF as compared to European wolves or African wild dogs. This also suggests that SJKF did not lose MHC diversity when they became isolated in the San Joaquin Valley over the last millennium. If anything, DRB shows higher allelic diversity than other wild canids and balancing selection may have occurred during the changing environments that kit foxes experienced within the valley (Matocq *et al.* 2012).

More recently SJKF have adapted to the urban environment in Bakersfield. This is the first opportunity to compare allelic diversity of a canid species with populations in

natural and urban environments. The Bakersfield population as a whole has a higher allelic diversity at all MHC loci (7 DQA, 11 DQB, 18 DRB alleles) than the wild populations (6 DQA, 6 DQB, 20 DRB alleles). Paired t-tests showed no difference in the number of MHC alleles or in the number of proteins with unique ABS between wild and urban populations ($p=0.289$). However there were some notable differences in the alleles found between urban and wild kit foxes. Both groups had unique DQA alleles found in one individual, but all of the alleles translate into the same functional ABS and are, therefore, indifferent for immune function. Urban and wild kit foxes both carried 3 DQB alleles with high allelic frequencies overall (Table 5-3). However the 13 alleles produce 6 different combinations of amino acids within the ABS, making very little difference to their immune system from DQB diversity. On the other hand, 14 of the 21 amino acid positions in the ABS were variable in DRB and the 29 DRB alleles translate into 25 unique ABS for pathogen recognition. Most of the high frequency DRB alleles were present in the urban and wild populations, except for *Vuma*-DRB*05, which is absent in Bakersfield but present in 45% of the individuals of wild populations. In addition, no individuals from Bakersfield carried *Vuma*-DRB*08 or *Vuma*-DRB*09, whereas each of these alleles were found in 11% of the wild populations. All three of these alleles translate into functionally unique ABS. It is unclear if we will find these alleles in urban kit foxes with more sampling, or if individuals that carried these alleles also lacked other alleles crucial to urban survival and did not survive. Outside of these more common DRB alleles, individuals from both wild and urban populations carried alleles that translated into unique proteins and may or may not be contributing to their immune

function. With further analysis, we will be able to assess whether these MHC differences in individuals are more important for kin recognition than for the antigen binding functions we have looked at here.

The results of this study show how parts of the MHC can be both conserved and divergent when we look within and across species. Amplification of DQA proved troublesome, as primer-binding sites were not conserved with other closely related carnivores and continued to add challenges throughout data processing. The initial bioinformatics tools only used the forward primers of each gene for recognition and reads for DQA were retained until separation by jMHC. While the incomplete reverse primer was problematic to identify, it may have contributed to the poor amplification of DQA with the fusion primers. If we had discovered this earlier, we could have redesigned the primers for better amplification of DQA instead of doing two amplifications with different primer sets. However, we do not know if new primers would solve these problems or if poor amplification is caused by other unknown factors such as tertiary structure.

Low read number of DQA made determination of true and artifactual variants difficult. While random SNPs could be generally ignored with DQB and DRB during variant grouping, with DQA the low number of reads for per individual meant that a SNP could have a high frequency per individual (F_{ij}) with a low actual number (N_{ij}). In the end we reviewed the variants that passed the T2 threshold and removed ones that appeared to be sequencing errors in homopolymer regions. While this seems to be the

best strategy thus far, other SNPs that passed as alleles could be sequencing error that were clonally amplified in emulsion PCR and are not true alleles. This same issue could exist with alleles that passed the T_2 threshold for DQB and DRB and are only found in one individual. However we were conservative in ignoring random SNPs that did not appear in more than one variant detected by jMHC or did not have a high numbers of reads overall. Therefore we likely avoided the error of confusing sequencing artifacts and true alleles. Our bioinformatic methods, particularly the T_1 and T_2 thresholds, appear to have been stringent enough to avoid these errors because the test statistic found no correlation between the number of reads and the number of alleles identified for an individual. If we had used lower thresholds we would expect more artifactual alleles to be identified and may have found a greater correlation between read and allele numbers. Our thresholds and error checking methods are similar, if not identical, to previously suggested approaches (Babik *et al.* 2009; Galan *et al.* 2010; Oomen *et al.* 2013), but clonal amplification errors from emulsion PCR are still hard to detect. We translated all alleles into amino acid sequences and confirmed that no alleles included stop codons and are therefore functional. However, it's very difficult to verify alleles even with cloning or mRNA studies, we must rely on bioinformatics standards to assess the validity of MHC alleles at this time.

Supplementary Materials

Supp. Table 1. DQA Fusion Primers

DQAin2 Reverse Fusion Primer (DQA exon 2)
CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGGACAGATTCAGTGAAGAGA
DQAin1b Forward Fusion Primers (DQA exon 2)
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCGTGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGATGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATACGAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCAGAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGATGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGTACGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGCGTCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGTAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATACAGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTATACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCACGTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTGCGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATCACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGCGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCGTCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCATGCGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGACACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGATAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATCTCGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCATAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGTAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCGACATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTAGCGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGTATCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGCTGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGATACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACGTACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACACTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTACTACTGTGTCCTGCTTCCTGCTC

CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGAGTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGATGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGCTCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACACACGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGAGTAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGCAGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGTACGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATAGAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACGAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACTCTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGATGTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGACTGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGTGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCTGACGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCAGCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACATCGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATGCAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACACTGTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATAGCTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATGATATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGATGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTACACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGTACATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCGTCTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATATATGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATACTATCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGACAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGCACTCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGCTCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCGATCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGCTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGATCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGATCAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTATGTACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCATCGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCTCGTATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACGATGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTCTGCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATGATGTGTCCTGCTTCCTGCTC

CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATCGTGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACATAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACTAGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCAGTGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTATATATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACTGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTCTGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCACTCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTACGTGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCTCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCACGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTAGCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCGCAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTGATCTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACTGATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCGCAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGTAGTGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACGCGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTCGATATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGACTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACTCAGTGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCGCATATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACTGATATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGCTCATGTGTCCTGCTTCCTGCTC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTACGCATGTGTCCTGCTTCCTGCTC

Supp. Table 2. DQB Fusion Primers

DQBR2 Reverse Fusion Primer (DQB exon 2)
CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCACCTCGCCGCTGCAACGTG
DQB1 Forward Fusion Primers (DQB exon 2)
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCGTGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGATGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATACGAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCAGAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGATGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGTACGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGCGTCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGTAGTCGTGTACCAGTTTAAGGGC

CCATCTCATCCCTGCGTGTCTCCGACTCAGCATACAGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTATACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCACGTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTGCGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATCACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGCGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCGTCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCATGCGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGACACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGATAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATCTCGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCATAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGTAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCGACATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTAGCGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGTATCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGCTGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGATACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACGTACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACACTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTACTACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGAGTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGATGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGCTCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACACACGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGAGTAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGCAGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGTACGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATAGAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACGAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACTCTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGATGTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGACTGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGTGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCTGACGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCAGCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACATCGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATGCAGTCGTGTACCAGTTTAAGGGC

CCATCTCATCCCTGCGTGTCTCCGACTCAGACACTGTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATAGCTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATGATATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGATGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTACACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGTACATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCGTCTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATATATGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATACTATCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGACAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGCACTCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGCTCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCGATCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGCTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGATCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGATCAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTATGTACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCATCGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCTCGTATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACGATGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTCTGCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATCGTGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACATAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACTAGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCAGTGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTATATATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACTGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTCTGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCACTCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTACGTGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCTCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCACGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTAGCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCGCAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTGATCTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACTGATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCGCAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGTAGTGTCGTGTACCAGTTTAAGGGC

CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACGCGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTCGATATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGACTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACTCAGTCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCGCATATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACTGATATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGCTCATCGTGTACCAGTTTAAGGGC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTACGCATCGTGTACCAGTTTAAGGGC

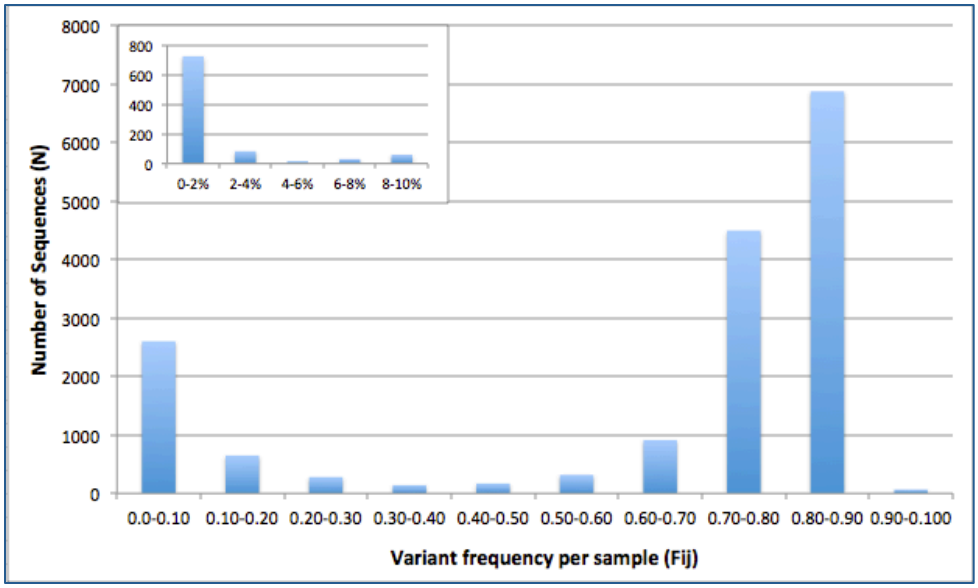
Supp. Table 3. DRB Fusion Primers

DM-2 Reverse Fusion Primers (DRB exon 2)
CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTCGCCGCTGCACCGTGAAGCT
DM-1 Forward Fusion Primers (DRB exon 2)
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGATGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATACGAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCAGAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGAGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGTACGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGCGTCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGTAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTATACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTGCGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCGCGATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATCACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATGTGATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGCTATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGTACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGCGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCATGCGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGACACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCACGCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGATGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGATAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACGCGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATCTCGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCATAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAAGTCCGAGTGCTATTTCACC

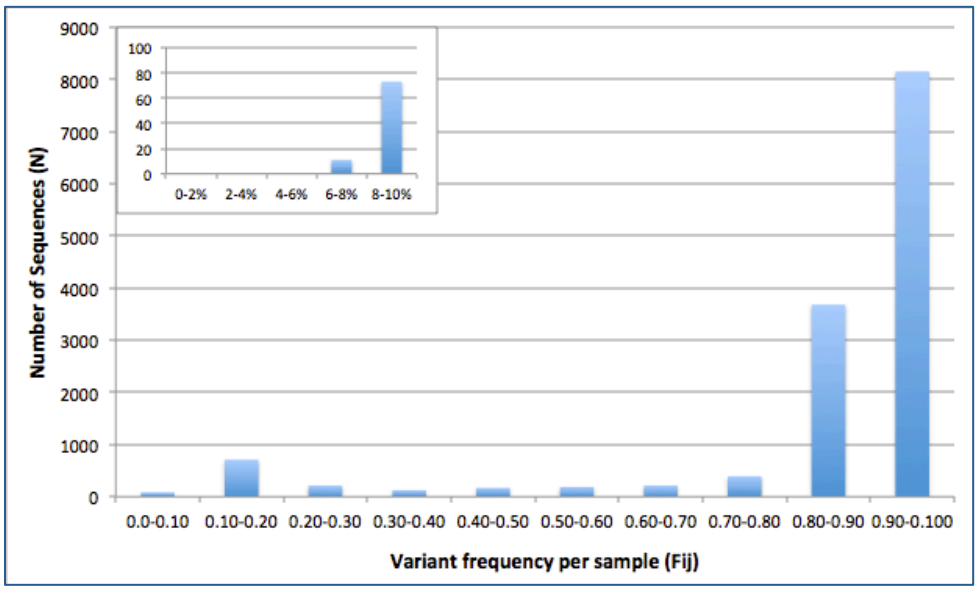
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTGCGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGTAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATACGCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTAGCGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGTATCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATAGTCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGATACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACGTACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTCGATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTACTACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGCTCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACACACGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACTATGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGAGTAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGACACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATACATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGTGTATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGTACGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATAGAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACGAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGACTGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCTGACGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACATCGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATGCAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTAGCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGATGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTACACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACGTCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCACACTATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATCACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATATATGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATACTATCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTATGATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGACAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTAGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCATCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGCACTCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGCTCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCATGACGTAAGTCCGAGTGCTATTTCACC

CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCGATCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGATCAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTATGTACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGTCATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACGATGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATGTATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGAGACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTCTATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGCGCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACATGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATCGTGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACATAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCAGTGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCGACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTACTGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTCTGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTACGTGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATACTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCACGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTAGCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCGCAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTGATCAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCGCAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGTAGTGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTATCTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACGCGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCATATAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGACAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGTGACTCAGAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGACGTAAGTCCGAGTGCTATTTCACC
CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCAGCGAAGTCCGAGTGCTATTTCACC

a) before T2 threshold

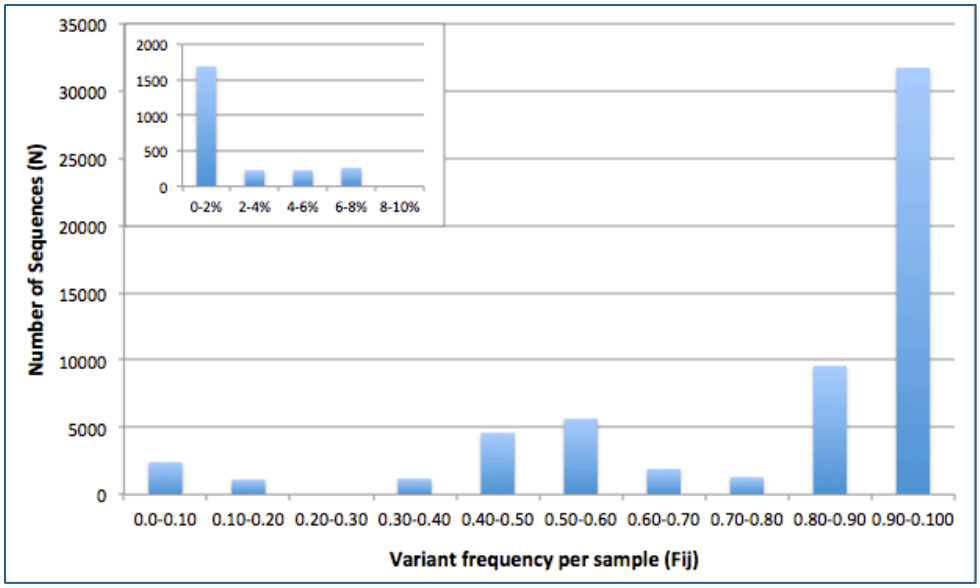


b) after removal of variants below T2 threshold (5%)

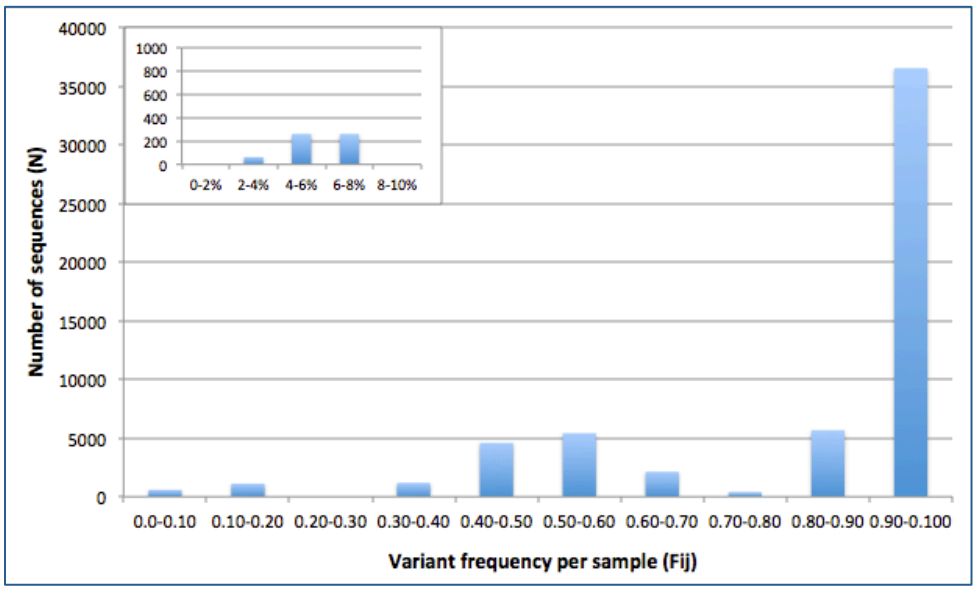


Supplemental Figure 1. Histograms for DQA showing the distributions of Fij, the frequency of each variant j within each individual sample i. Inset is an expansion of 0-10% from the main graph.

a) before T2 threshold

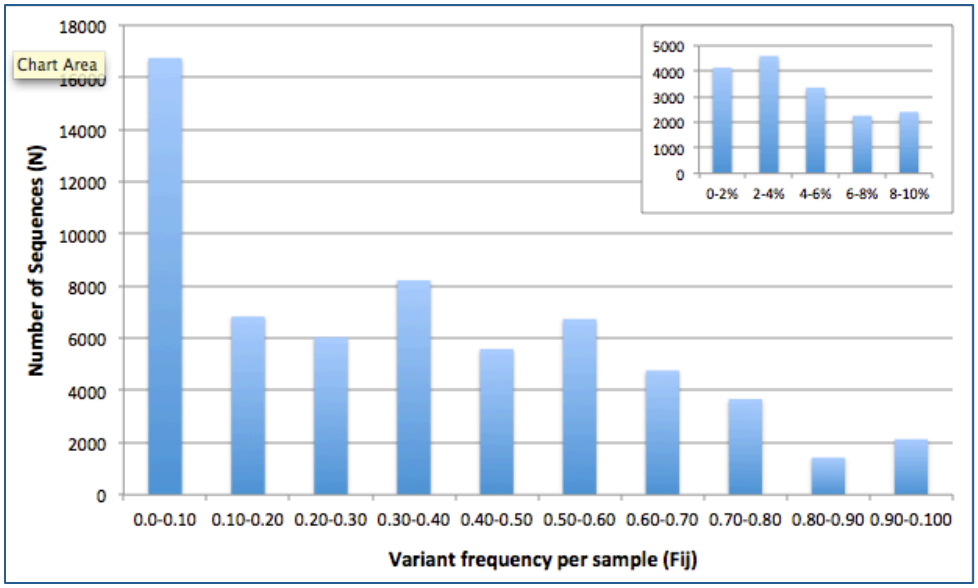


b) after removal of variants below T2 threshold (3%)

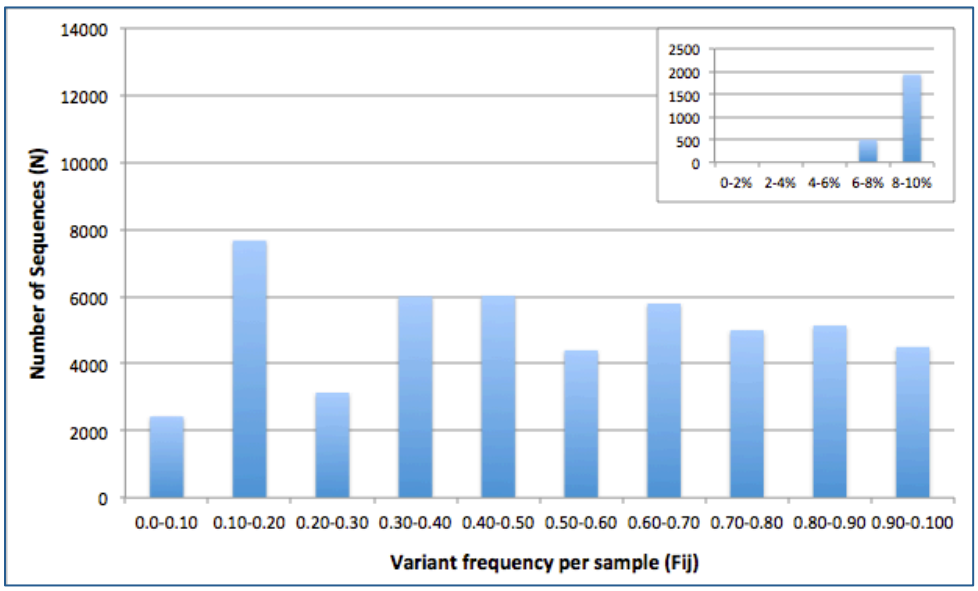


Supplemental Figure 2. Histograms for DQB showing the distributions of Fij, the frequency of each variant j within each individual sample i. Inset is an expansion of 0-10% from the main graph.

a) before T2 threshold



b) after removal of variants below T2 threshold (6%)



Supplemental Figure 3. Histograms for DRB showing the distributions of Fij, the frequency of each variant j within each individual sample i. Inset is an expansion of 0-10% from the main graph.

CHAPTER SIX: CONCLUSIONS

Through these projects we gained a lot of important information about San Joaquin kit foxes in a variety of environments. Despite protections under the Endangered Species Act, SJKF continue to decline in numbers as habitat become more fragmented with smaller patches and less connected. Our analysis of kit foxes across the San Joaquin Valley showed genetically distinct signatures in the Ciervo-Panoche core area and the Bakersfield area that are both largely absent from the populations in Lokern and Carrizo Plain. We realize that both historic and contemporary processes lead to these patterns with current decreases in population sizes and migration rates. While Lokern and the Carrizo Plain contain the largest numbers of kit foxes and have the largest amount of highly suitable habitat, the unique genetics in C-P and Bakersfield necessitate more conservation attention than they currently receive. The history of development and kit fox displacement predicts further differentiation between these groups, and this makes them more vulnerable to stochastic events. Clearly kit foxes cannot rebound or adapt to all of the challenges of an anthropogenic landscape; large numbers of kit foxes have died from pesticides, rodenticides, and road kill. However, kit foxes have also shown great potential to adapt to new environments like Bakersfield or to survive in and around human settlements throughout the San Joaquin Valley. Our new understanding of kit fox and the importance of each region can help create a plan to protect genetic diversity in addition to areas with good habitat and large populations.

In addition, the possibility of corridors between these regions could be further investigated with additional sampling of small highly suitable habitat patches in the Valley. This research has provided a foundation for understanding the levels of genetic diversity in smaller sampling efforts or in new locations. Addition of new sampling locations could also be helpful in understanding the cost of dispersal across the landscape, with many variables needing to be incorporated – agriculture, canals, roads, slope, soil, predators, prey, distance. Our research in the Ciervo-Panoche area showed us that looks can be very deceiving, with some areas looking identical to others and being located in between kit fox groups and yet having no perceivable signs of kit foxes.

Our data are abundant with questions to can be address on a finer scale. We have known family groups from the Carrizo Plain, Lokern, and Bakersfield that consist of either a pair or a trio. We plan to examine the social and genetic dynamics of the third individual – whether she is significantly related, a sister or offspring, and if she is allowed to breed. In addition, do these dynamics change between the wild and urban kit foxes where resources and densities are different? Secondly, we have found geographic barriers to dispersal and we wonder if the city of Bakersfield contains additional barriers, such as canals, major highways, and large areas with fencing. With the large numbers of samples from Bakersfield, we hope to learn about urban migration patterns and how kit foxes are avoiding inbreeding. In addition, we plan to use our data on MHC types with pedigree data to explicitly test questions of mate choice. We can then determine if MHC type correlates with kit fox behavior in urban and wild environments.

Finally, this research has led to new challenges technologically. We want to amplify MHC from scat samples from the Ciervo-Panoche area so that we can determine if those kit foxes have different immune system diversity. We plan to try a capture protocol with MHC bait for a small portion of each gene and then use the same fusion primers with 454 sequencing. In addition, we would like to identify gene copies of DRB in the SJKF. We would like to use circular consensus sequencing of the genomic region of DRB to find the adjacent sequences and identify gene copies. This would allow us to explore more questions about evolution and population genetics of the MHC in the San Joaquin kit fox. Ultimately these studies of MHC would be paired with disease risk modeling and prevalence rates of common diseases in other animals that kit foxes may come into contact, like skunks, raccoons, dogs and cats. Populations of the endangered San Joaquin kit fox may or may not be very susceptible to diseases, and conservation of this fox could be greatly improved through immune system and disease research.

REFERENCES

REFERENCES

(2010) Carrizo Plain National Monument Approved Resource Management Plan and Record of Decision; Final Rule, pp. 18547-18548, Federal Register.

Aguilar A, Roemer G, Debenham S*, et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3490-3494.

Babik W, Taberlet P, Ejsmond M, J R (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources* **9**, 713-719.

Barrows HD, Ingersoll LA (1893) *A Memorial and Biographical History of the Coast Counties of Central California* Lewis Publishing Company, Chicago.

Bean E (2002) *An evaluation of distance and mark–resight sampling to estimate abundance of San Joaquin kit fox.*, University of MInnesota.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4563-4568.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources* **9**, 1279-1301.

Bell HM, Alvarez JA, Eberhardt LL, Ralls K (1994) Distribution and abundance of San Joaquin kit fox, p. 93. California Department of Fish and Game, Nongame Bird and Mammal Section.

Bidlack A (2007) *Mesocarnivore responses to changes in habitat and resource availability in California.*, University of California, Berkeley.

Bjurlin CD, Cypher BL, Wingert CM, Van Horn Job CL (2005) Urban Roads and the Endangered San Joaquin kit fox, p. 47. The California Department of Transportation.

Boulanger J, McLellan B (2001) Closure violation in DNA-based mark-recapture estimation of grizzly bear populations. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **79**, 642-651.

Bozarth CA, Alva-Campbell YR, Ralls K*, et al.* (2010) An efficient noninvasive method for discriminating among faeces of sympatric North American canids. *Conservation Genetics Resources* **2**, 173-175.

Bozarth CA, Hailer F, Rockwood LL, Edwards CW, Maldonado JE (2011) Coyote colonization of northern Virginia and admixture with Great Lakes wolves. *Journal of Mammalogy* **92**, 1070-1080.

Briden LE, Archon M, Chesemore DL (1992) Ecology of the San Joaquin Kit Fox in Western Merced County, California. In: *Endangered and Sensitive Species of the San Joaquin Valley, California: their Biology, Management and Conservation.* (eds. Williams DF, Byrne S, Rado TA), pp. 81-87. California Energy Commission, Sacremento, California.

Cameron DR, Cohen BS, Morrison SA (2012) An Approach to Enhance the Conservation-Compatibility of Solar Energy Development. *PLoS ONE* **7**.

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

Clark HO, Jr. (2007) Marking of novel objects by kit foxes. *California Fish and Game* **93**, 103-106.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Clifford D, Mazeta J, Dubovib E, al. e (2006) Pathogen exposure in endangered island fox (Urocyon littoralis) populations: Implications for conservation management. *Biological Conservation* **131**, 230-243.

Collins PW (2010) Results of small mammal trapping on portions of the Topaz Solar Farm Project site in California Valley, Sections 4, 5, 15, 26, 28, 32-35, San Luis Obispo County, California.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15**, 2833-2843.

Coulon A, Fitzpatrick JW, Bowman R*, et al.* (2008) Congruent population structure inferred from dispersal behavior and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma coerulescens*). *Molecular Ecology* **17**, 1685-1701.

Cypher BL, Bjurlin CD, Nelson JL (2009) Effects of Roads on Endangered San Joaquin Kit Foxes. *Journal of Wildlife Management* **73**, 885-893.

Cypher BL, Phillips SE, Kelly PA (2013) Quantity and distribution of suitable habitat for endangered San Joaquin kit foxes: conservation implications. *Canid Biology & Conservation* **16**, 25-31.

Cypher BL, Scrivner JH (1992) *COYOTE CONTROL TO PROTECT ENDANGERED SAN JOAQUIN KIT FOXES AT THE NAVAL PETROLEUM RESERVES, CALIFORNIA*.

Cypher BL, Spencer KA (1998) Competitive interactions between coyotes and San Joaquin kit foxes. *Journal of Mammalogy* **79**, 204-214.

Cypher BL, Warrick GD, Otten MRM*, et al.* (2000) Population dynamics of San Joaquin kit foxes at the Naval Petroleum Reserves in California. *Wildlife Monographs*, 1-43.

De Barba M, Waits LP, Garton EO*, et al.* (2010) The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. *Molecular Ecology* **19**, 3938-3951.

Dutta T, Sharma S, Maldonado JE*, et al.* (2013) Fine-scale population genetic structure in a wide-ranging carnivore, the leopard (Panthera pardus fusca) in central India. *Diversity and Distributions* **19**, 760-771.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*.

Eggert LS, Maldonado JE, Fleischer RC (2005) Nucleic acid isolation from ecological samples - Animal scat and other associated materials. In: *Molecular Evolution: Producing the Biochemical Data, Part B* (eds. Zimmer EA, Roalson EH), pp. 73-87.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome* **7**, 359-362.

Frankham R (1995) Inbreeding and extinction - a threshold effect. *Conservation Biology* **9**, 792-799.

Frankham R, Ralls K (1998) Conservation biology: Inbreeding leads to extinction. *Nature* **392**, 441-442.

Frusetta PC (1991) *Quicksilver country: California's New Idria mining district* Peter Frusetta, Clovis, CA.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* **11**, 296-310.

Germano DJ, Rathbun GB, Saslaw LR (2012) Effects of grazing and invasive grasses on desert vertebrates in California. *Journal of Wildlife Management* **76**, 670-682.

Gillevet P, Sikaroodi M, Keshavarzian A, Mutlu EA (2010) Quantitative assessment of the human gut microbiome using multitag pyrosequencing. *Chemical Biodiversity* **7**, 1065-1075.

Gompper ME, Kays RW, Ray JC*, et al.* (2006) A comparison of noninvasive techniques to survey carnivore communities in northeastern North America. *Wildlife Society Bulletin* **34**, 1142-1151.

Guedj B, Guillot G (2011) Estimating the location and shape of hybrid zones. *Molecular Ecology Resources* **11**, 1119-1123.

Harismendy O, Ng PC, Strausberg RL*, et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**, R32.

Hartl DL, Clark AG (1997) *Principles of population genetics* Sinauer associates Sunderland.

Hedrick PW, Hedgecock D, Hamelberg S, Croci SJ (2000a) The impact of supplementation in winter-run chinook salmon on effective population size. *Journal of Heredity* **91**, 112-116.

Hedrick PW, Lee RN, Buchanan C (2003) Canine parvovirus enteritis, canine distemper, and major histocompatibility complex genetic variation in Mexican wolves. *Journal of Wildlife Diseases* **39**, 909-913.

Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Molecular Ecology* **11**, 1905-1913.

Hedrick PW, Lee RN, Parker KM (2000b) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity* **85**, 617-624.

Hoelzel AR, Stephens JC, O'Brien SJ (1999) Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds. *Molecular Biology and Evolution* **16**, 611-618.

Homan HJ, Linz G, Peer BD (2001) Dogs increase recovery of passerine carcasses in dense vegetation. *Wildlife Society Bulletin* **29**, 292-296.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167 - 170.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Janeway Jr. CA, Travers P, Walport M, Shlomchick MJ (2001) The major histocompatibility complex and its functions. In: *Immunobiology: The Immune System in Health and Disease*. Garland Science, New York.

Jost L (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-4026.

Jukes TH, Cantor CR (1969) Evolution of protein molecules.

Kelly PA, Phillips SE, Williams DF (2005) Documenting Ecological Changes in Time and Space: The San Joaquin Valley of California. In: *Mammalian Diversification: From Chromosomes to Phylogeography (A Celebration of the Career of James L. Patton)* (eds. Lacey EA, Myers P), pp. 57-78. University of California Press, Berkeley, California.

Kennedy LJ, Angles JM, Barnes A*, et al.* (2007) DLA-DRBI, DQAI, and DQBI alleles and haplotypes in North American gray wolves. *Journal of Heredity* **98**, 491-499.

Kennedy LJ, Barnes A, Happ GM*, et al.* (2002) Evidence for extensive DLA polymorphism in different dog populations. *Tissue Antigens* **60**, 43-52.

Kennedy LJ, Carter SD, Barnes A*, et al.* (1998) Nine new dog DLA-DRB1 alleles identified by sequence-based typing. *Immunogenetics* **48**, 296-301.

Klein J (1986) *Natural history of the major histocompatibility complex* Wiley New York.

Klein J, Gutknecht J, Fischer N (1990) The major histocompatibility complex and human evolution. *Trends in Genetics* **6**, 7-11.

Koenig WD, VanVuren D, Hooge PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution* **11**, 514-517.

Kohn MH, York EC, Kamradt DA*, et al.* (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society B-Biological Sciences* **266**, 657-663.

Koopman ME, Cypher BL, Scrivner JH (2000) Dispersal patterns of San Joaquin kit foxes (Vulpes macrotis mutica). *Journal of Mammalogy* **81**, 213-222.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289-1291.

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinformatics* **9**, 299-306.

Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar). *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 1279-1285.

Laurenson M, Cleaveland S, Artois M, Woodroffe R (2004) Assessing and managing infectious disease threats to canids. In: *Canids: Foxes, Wolves, Jackals and Dogs. Status Survey and Conservation Action Plan* (eds. Sillero-Zubiri C, Hoffman M, Macdonald DW), pp. 246-255. IUCN/SSC Canid Specialist Group, Gland, Switzerland, and Cambridge.

Li HF, Han W, Zhu YF*, et al.* (2009) Analysis of genetic structure and relationship among nine indigenous Chinese chicken populations by the Structure program. *Journal of Genetics* **88**, 197-203.

Lovich JE, Ennen JR (2011) Wildlife Conservation and Solar Energy Development in the Desert Southwest, United States. *BioScience* **61**, 982-992.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* **11**, 355-373.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753-1766.

Marsden CD, Woodroffe R, Mills MGL*, et al.* (2012) Spatial and temporal patterns of neutral and adaptive genetic variation in the endangered African wild dog (Lycaon pictus). *Molecular Ecology* **21**, 1379-1393.

Matocq MD, Kelly PA, Phillips SE, Maldonado JE (2012) Reconstructing the evolutionary history of an endangered subspecies across the changing landscape of the Great Central Valley of California. *Molecular Ecology* **21**, 5918-5933.

Meglécz E, Piry S, Desmarais E*, et al.* (2011) SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. *Bioinformatics* **27**, 277-278.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F-ST and related measures. *Molecular Ecology Resources* **11**, 5-18.

Metzker ML (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**, 31-46.

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics* **37**, 159-186.

Miller CR, Joyce P, Waits LP (2005) A new method for estimating the size of small populations from genetic mark-recapture data. *Molecular Ecology* **14**, 1991-2005.

Miller DS, Covell DF, McLean RG*, et al.* (2000) Serologic survey for selected infectious disease agents in swift and kit foxes from the Western United States. *Journal of Wildlife Diseases* **36**, 798-805.

Mills LS, Citta JJ, Lair KP, Schwartz MK, Tallmon DA (2000) Estimating animal abundance using noninvasive DNA sampling: Promise and pitfalls. *Ecological Applications* **10**, 283-294.

Moehrenschlager A, Cypher B, Ralls K, List R, Sovada M (2004) Comparative ecology and conservation priorities of swift and kit foxes. In: *Biology and conservation of wild canids* (eds. Macdonald DW, Sillero-Zubiri C), pp. 185-198. Oxford University Press, Oxford.

Nei M, Chesser RK (1983) ESTIMATION OF FIXATION INDEXES AND GENE DIVERSITIES. *Annals of Human Genetics* **47**, 253-259.

Neiman M, Lundin S, Savolainen P, Ahmadian A (2011) Decoding a Substantial Set of Samples in Parallel by Massive Sequencing. *PLoS ONE* **6**, e17785.

Nelson JL, Cypher BL, Bjurlin CD, Creel S (2007) Effects of habitat on competition between kit foxes and coyotes. *Journal of Wildlife Management* **71**, 1467-1475.

Newsome SD, Ralls K, Van Horn Job CL, Fogel ML, Cypher BL (2010) Stable isotopes evaluate reliance on anthropogenic food by the endangered San Joaquin kit fox (*Vulpes macrotis mutica*). *Journal of Mammalogy* **91**, 1313-1321.

Nielsen R, Tarpy DR, Reeve HK (2003) Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology* **12**, 3157-3164.

Nussear KE, Esque TC, Heaton JS*, et al.* (2008) Are wildlife detector dogs or people better at finding desert tortoises (*Gopherus agassizii*)? *Herpetological Conservation and Biology* **3**, 103-115.

Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources* **13**, 103-116.

Ortega J, Franco MD, Adams BA, Ralls K, Maldonado JE (2004) A reliable, non-invasive method for sex determination in the endangered San Joaquin kit fox (Vulpes macrotis mutica) and other canids. *Conservation Genetics* **5**, 715-718.

Palstra FP, Fraser DJ (2012) Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution* **2**, 2357-2365.

Park SDE (2001) *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*, University of Dublin.

Parris KM, Hazell DL (2005) Biotic effects of climate change in urban environments: The case of the grey-headed flying-fox (*Pteropus poliocephalus*) in Melbourne, Australia. *Biological Conservation* **124**, 267-276.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Pelletier A, Obbard ME, Mills K*, et al.* (2012) Delineating genetic groupings in continuously distributed species across largely homogeneous landscapes: a study of American black bears (Ursus americanus) in Ontario, Canada. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **90**, 999-1014.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7-21.

Ploshnitsa AI, Goltsman ME, Macdonald DW, Kennedy LJ, Sommer S (2012) Impact of historical founder effects and a recent bottleneck on MHC variability in Commander Arctic foxes (Vulpes lagopus). *Ecology and Evolution* **2**, 165-180.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data *Genetics* **155**, 945-959.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, 258-275.

Ralls K, Cypher B, Spiegel LK (2007a) Social monogamy in kit foxes: formation, association, duration, and dissolution of mated pairs. *Journal of Mammalogy* **88**, 1439-1446.

Ralls K, Cypher BL, Spiegel LK (2007b) Social monogamy in kit foxes: formation, association, duration, and dissolution of mated pairs. *Journal of Mammalogy* **88**, 1439-1446.

Ralls K, Eberhardt LL (1997) Assessment of abundance of San Joaquin kit foxes by spotlight surveys. *Journal of Mammalogy* **78**, 65-73.

Ralls K, Pilgrim KL, White PJ*, et al.* (2001) Kinship, social relationships, and den sharing in kit foxes. *Journal of Mammalogy* **82**, 858-866.

Ralls K, Sharma S, Smith DA*, et al.* (2010) Changes in Kit Fox Defecation Patterns During the Reproductive Season: Implications for Noninvasive Surveys. *Journal of Wildlife Management* **74**, 1457-1462.

Ralls K, Smith DA (2004) Latrine use by San Joaquin kit foxes (Vulpes macrotis mutica) and coyotes (Canis latrans). *Western North American Naturalist* **64**, 544-547.

Ralls K, White PJ (1995a) Predation on endangered San Joaquin kit foxes by larger canids. *Journal of Mammalogy* **276**, 723-729.

Ralls K, White PJ (1995b) PREDATION ON SAN-JOAQUIN KIT FOXES BY LARGER CANIDS. *Journal of Mammalogy* **76**, 723-729.

Randa LA, Yunger JA (2006) Carnivore occurrence along an urban-rural gradient: A landscape-level analysis. *Journal of Mammalogy* **87**, 1154-1164.

Rannala B (2007) BayesAss Edition 3.0 User's Manual, p. 12. University of California Davis, Davis, California.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research* **67**, 175-186.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**, 939-946.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Schmidt D, Pool J (2002) The effect of population history on the distribution of the Tajima's D statistic.

Schoon GAA (1996) Scent identification lineups by dogs (Canis familiaris): experimental design and forensic application. *Applied Animal Behaviour Science* **49**, 257-267.

Schwartz MK, McKelvey KS, Cushman SA, Luikart G (2010) Landscape genomics: a brief perspective. In: *Spatial complexity, informatics, and wildlife conservation*, pp. 165-174. Springer.

Schwartz MK, Ralls K, Williams DF*, et al.* (2005) Gene flow among San Joaquin kit fox populations in a severely changed ecosystem. *Conservation Genetics* **6**, 25-37.

Schwarz CJ, Arnason AN (1996) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52**, 860-873.

Scrivner JH, O'Farrell TP, Kato TT (1987) Diet of the San Joaquin kit fox, *Vulpes macrotis mutica*, on Naval Petroleum Reserve #1, Kern County, California, 1980-1984, p. 26. EG&G Energy Measurements.

Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics* **54**, 490-500.

Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral

evolution. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 2283-2291.

Short AD, Kennedy LJ, Forman O*, et al.* (2005) Canine DNA Subjected to Whole Genome Amplification is Suitable for a Wide Range of Molecular Applications. *Journal of Heredity* **96**, 829-835.

Slatkin M (1995) A MEASURE OF POPULATION SUBDIVISION BASED ON MICROSATELLITE ALLELE FREQUENCIES. *Genetics* **139**, 457-462.

Smith D, Ralls K, Hurt A*, et al.* (2006a) Assessing reliability of microsatellite genotypes from kit fox faecal samples using genetic and GIS analyses. *Molecular Ecology* **15**, 387-406.

Smith DA, Ralls K, Cypher BL*, et al.* (2006b) Relative abundance of endangered San Joaquin kit foxes (Vulpes macrotis mutica) based on scat-detection dog surveys. *Southwestern Naturalist* **51**, 210-219.

Smith DA, Ralls K, Cypher BL, Maldonado JE (2005) Assessment of scat-detection dog surveys to determine kit fox distribution. *Wildlife Society Bulletin* **33**, 897-904.

Smith DA, Ralls K, Hurt A*, et al.* (2003) Detection and accuracy rates of dogs trained to find scats of San Joaquin kit foxes (Vulpes macrotis mutica). *Animal Conservation* **6**, 339-346.

Spiegel LK, Bradbury M (1992) HOME-RANGE CHARACTERISTICS OF THE SAN-JOAQUIN KIT FOX IN WESTERN KERN COUNTY, CALIFORNIA. *Transactions of the Western Section of the Wildlife Society 1992, Vol 28* **28**, 83-92.

Spiegel LK, Dao TC (1997) The occurrence of hydrogen sulfide gas in San Joaquin kit fox dens and rodent burrows in an oil field in California. *California Fish and Game* **83**, 38-42.

Spiering PA, Somers MJ, Maldonado JE, Wildt DE, Gunther MS (2010) Reproductive sharing and proximate factors mediating cooperative breeding in the African wild dog (Lycaon pictus). *Behavioral Ecology & Sociobiology* **64**, 583-592.

Standley WG, McCue PM (1992) Blood characteristics of San Joaquin kit fox (Vulpes velox macrotis) at Camp Roberts Army National Guard Training Site, California, p. 25 pp. U.S. Department of Energy, National Technical Information Service.

Stoms DM, Dashiell SL, Davis FW (2013) Siting solar energy development to minimize biological impacts. *Renewable Energy* **57**, 289-298.

Stuglik MT, Radwan J, Babik W (2011) jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources* **11**, 739-742.

Taberlet P, Camarra JJ, Griffin S*, et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**, 869-876.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution* **14**, 323-327.

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

Timm S, Barker W, Johnson S, al. e (2002) Island fox recovery efforts on Santa Catalina Island, California, September 2000-October 2001 annual report. prepared by the Institute for Wildlife Studies for the Ecological Restoration Department, Santa Catalina Island Conservancy.

Tollenaere C, Bryja J, Galan M*, et al.* (2008) Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology* **21**, 1307-1320.

U.S. Fish and Wildlife Service (1998) Recovery Plan for upland species of the San Joaquin Valley, California (ed. Region 1 P, OR.), p. 319 pp.

U.S. Fish and Wildlife Service (2010) San Joaquin kit fox (Vulpes macrotis mutica) 5-year review: summary and evaluation (ed. Service USFaW), p. 122, Sacremento.

Wagner JL, Burnett RC, Works JD, Storb R (1996) Molecular analysis of DLA-DRBB1 polymorphism. *Tissue Antigens* **48**, 554-561.

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology* **10**, 249-256.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**, 1419-1433.

Wakeland EK, Boehme S, She JX*, et al.* (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. *Immunologic research* **9**, 115-122.

Warrick GD, Clark HO, Jr., Kelly PA, Williams DF, Cypher BL (2007) Use of agricultural lands by San Joaquin kit foxes. *Western North American Naturalist* **67**, 270-277.

Warrick GD, Harris CE (2001) Evaluation of spotlight and scent-station surveys to monitor kit fox abundance. *Wildlife Society Bulletin* **29**, 827-832.

Wegner K (2009) Massive parallel MHC genotyping: titanium that shines. *Molecular Ecology* **18**, 1818-1820.

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

White PJ, Berry WH, Eliason JJ, Hanson MT (2000) Catastrophic decrease in an isolated population of kit foxes. *The Southwestern Naturalist* **45**, 204-211.

White PJ, Ralls K (1993) REPRODUCTION AND SPACING PATTERNS OF KIT FOXES RELATIVE TO CHANGING PREY AVAILABILITY. *Journal of Wildlife Management* **57**, 861-867.

Williams D, Kelly PA, Bell HM, O'Brien L (1996) Report of kit fox activities conducted under California Department of Fish and Game Memorandum of Understanding and U.S. Fish and Wildlife Service Subpermit WILLDF-12, p. 11. San Joaquin Valley Endangered Species Recovery Planning Program, Fresno, California.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Woodroffe R, Cleaveland S, Courtenay O, al. e (2004) Infectious disease in the management of conservation of wild canids. In: *Biology and Conservation of Wild Canids* (eds. MacDonald DW, Sillero-Zubiri C), p. 450. Oxford University Press, Oxford, UK.

Woods JG, Paetkau D, Lewis D*, et al.* (1999) Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin* **27**, 616-627.

Zoellick BW, Harris CE, Kelly BT*, et al.* (2002) Movements and home ranges of San Joaquin kit foxes (Vulpes macrotis mutica) relative to oil-field development. *Western North American Naturalist* **62**, 151-159.

BIOGRAPHY

Tammy R Wilbert graduated from Emmaus High School, Emmaus, Pennsylvania, in 1998. She received her Bachelor of Science in Biochemistry and Molecular Biology from University of Maryland Baltimore County in 2002. She explored different fields of science, such as proteomics and endocrinology, before getting hooked on genetics. All of her research has always had a focus on conservation and environmental science, and she hopes to always be connected to these disciplines. Tammy enjoys using genetic techniques and creative approaches to address conservation issues. She hopes her research in conservation genetics will contribute to species and ecosystem survival, and it is her small way of giving something back to this world.