Elucidating evolutionary processes in three threatened carnivores: genetic substructure, admixture, and cancer susceptibility

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Bioinformatics and Computational Biology

in the

College of Graduate Studies

University of Idaho

by

Sarah Anne Hendricks

Major Professor: Paul Hohenlohe, Ph.D.

Committee Members: Lisette Waits, Ph.D.; Sam Hunter, Ph.D.; Christine Parent, Ph.D.

Department Administrator: David Tank, Ph. D.

May 2019

Authorization to Submit Dissertation

This dissertation of Sarah Anne Hendricks, submitted for the degree of Doctor of Philosophy with a major in Bioinformatics and Computational Biology and titled "Elucidating evolutionary processes in three threatened carnivores: genetic substructure, admixture, and cancer susceptibility," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:	_____	_____
	Paul Hohenlohe, Ph.D.	Date
Committee Members:	_____	_____
	Lisette Waits, Ph.D.	Date
	_____	_____
	Sam Hunter, Ph.D.	Date
	_____	_____
	Christine Parent, Ph.D.	Date
Department Administrator:	_____	_____
	David Tank, Ph.D.	Date

Abstract

Advanced genomic techniques can be used to understand threats to global biodiversity and to promote ecosystem conservation. Population structure, migration, admixture, and the genetic basis of adaptive variation can be inferred in small, isolated, and declining populations of interest, thereby adding knowledge necessary for appropriate conservation planning. Here, I present three case examples to illustrate genomics' potential to influence the management of wild populations. First, Tasmanian devils (*Sarcophilus harrisii*) face a combination of threats to persistence, including devil facial tumor disease (DFTD), an epidemic transmissible cancer. I used RAD sequencing to investigate genome-wide patterns of genetic diversity and geographic population structure. Our results refine the geographic extent of the zone of mixed ancestry and substructure within it. DFTD has spread across all genetic clusters, but recent evidence points to a genomic response to selection imposed by DFTD. Any allelic variation for resistance to DFTD may be able to spread across the devil population, and/or be present as standing variation in both genetic regions. This can inform the management of genetic variation that existed in pre-diseased populations of the species. Second, wolves (*Canis lupus*) have naturally reestablished in the Pacific northwest region of North America. I used targeted capture sequencing to acquire SNP data from which I inferred population structure and ancestry. The wolves in Washington state represent an admixed population between the inland Northern Rocky Mountain wolf and the coastal rainforest wolf. Given this admixture, conservation status and management could be impacted if other coastal wolf populations continue to decline in size. Third, Catalina Island (SCA) foxes (*Urocyon littoralis catalinae*) have a high prevalence of ear canal (ceruminous gland) tumors that appear to be associated with inflammation from chronic ear mite (*Otodectes*) infections. I tested the hypothesis that the remarkably high incidence of tumors in SCA foxes is the result of genetic variants for cancer susceptibility, which have increased in frequency due to isolation and small population size. Using whole-genome sequencing

to identify and genotype polymorphic loci, I found that this trait is likely due to many loci across the genome with small effect. These genetic markers may be used for both monitoring and management in this threatened subspecies. Lastly, the study of the genomics of wildlife cancer is undocumented in a concise way. Therefore, I reviewed and synthesized published literature to record recent studies regarding the genomics of cancer, particularly in endangered taxa. Overall, this review, as well as the results of these three case studies, documents ways in which the genomic consequences of a small population size should be considered when managing species of conservation concern.

Acknowledgements

I would like to express my sincere gratitude to my advisor Paul Hohenlohe. Thank you for your continuous support of my Ph.D research and related projects. Thank you for your patience, motivation, and immense knowledge. I might not have been the easiest first graduate student to have, but at least you're now well trained on how to navigate when your student experience some rough waters. I'd also like to thank Lisette Waits, Sam Hunter, and Christine Parent for being amazing committee members. You have been my cheerleaders and have provided invaluable guidance throughout this process. Thank you for teaching me a great deal about scientific research and life in general.

I could not have made it through this program without the support of many University of Idaho (UI) graduate students including Kenetta Nunn, Katie Peterson, David Streett, Demin Jochimsen, Amanda Stahlke, Sarah Jacobs, Katie (Shine) Isaacs, Daniel Caetano, Austin Patton, Kristen Petersen, Brenda Hanley, Sarah Brooker, Megan Ruffley, Stacey Nerkowski, Lauren Parker, and Hannah Marx. Additionally, I learned so much from UI and Washington State University (WSU) postdoc fellows Anahi Espindola, Steve Spear, Kim Andrews, Brendan Epstein, Amy Worthington, Andy Kraemer, Sarah Wagner, Eliot Miller and Maribeth Latvis. I'd like to thank to thoughtfulness and guidance of the UI and WSU faculty and staff including Lisha Abendroth, Jack Sullivan, Dave Tank, Luke Harmon, Pat Carter, Donna Holmes, Celeste Brown and the Randall Women in Science Committee. I am fortunate to have many colleagues from University of California, Los Angeles that have been tremendously supportive during the last eight years (Sarah Wenner, Rena Schweizer, Annabel Beichman, Jacqueline Robinson, Clare Marsden, Ryan Harrigan, Bob Wayne, Sergio Nigenda-Morales, Pam Thompson, Genna Mountt, Klaus-Peter Koepfli, Pauline Charruau, and Alice Mouton). I could not have made it to this point without the support of my previous principal investigator, Greg Spicer, and lab mates, Tara Roth and Corrie Moreau, from San Francisco State University. It has been a pleasure to work with collaborators from

around the world - Louis Bernatchez, Marty Kardos, Gordon Luikart, Eric Anderson, Arun Sethuraman, Brenna Forester, Juan Pablo Torres and JP Zegarra.

I especially want to thank my friends and family. I cannot possibly express how grateful I am for your love, laughter and shoulders to cry on over the last 5 (plus) years. Stacy Isenbarger - thank you for always dancing with me. Barbara Rose - thank you for being my Moscow mom. Kim Kester - where to begin - thanks for always having my back. Shannon Stirdivant - thank you for never calling me crazy when I may have been a little crazy. Christina (Matthews) Vitale - thank you for being the sister I've never had. Ashley (Wright) Brower - thank you for always being sweet, kind, and generous. Phil West - thank you for bringing me to Disney when I needed to believe in miracles again. Tess Capen - thank you for encouraging me to pursue this path and sharing your love for the planet with me. George Skandalos - thank you for making me laugh (a lot) so that I spit out my water all over the table (or was that you?). KittyCat - you're my guiding light and the reason I got through all this. Mom - thank you for not giving up during the last 5 years. I still need you and always will. Brian - thank you for being by my side when our worlds flipped upside down and for laughing with me at the most inappropriate of times. Dad - thank you for instilling in me a love for animals and working hard. I love and miss you always.

Dedication

To my dad

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

We should preserve every scrap of biodiversity as priceless while we learn to use it and come to understand what it means to humanity.

E. O. Wilson, *The Diversity of Life*

There is a fundamental difference of opinion regarding humanity's role as stewards of the earth. While most agree that we are currently in the 6th great extinction, also known as the Anthropocene extinction, the need for resolution and the solution itself remain beyond the reach of general consensus. This mass extinction differs from past events in that this is the only extinction caused by a single species (humans). Between 1970 and 2012, there has been an estimated 58% decline in vertebrate population sizes (McRae et al. 2017) and, on average, approximately 13% of local species diversity has been lost since 1500 (Newbold et al. 2015). We collectively accomplished this through activities such as habitat conversion or degradation, land fragmentation, introduction of exotic species (including diseases), changing climate, ocean acidification, and human consumption of natural resources.

Initiatives like the Endangered Species Act (ESA), Convention on Biological Diversity (CBD), and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) have made great strides to protect imperiled biodiversity worldwide. Most of these initiative has focused on two paradigms to frame the biodiversity problem (Ouborg et al. 2006). These two paradigms are not mutually exclusive, but present alternative interpretations of the issues facing biodiversity. Firstly, the 'habitat quality' paradigm emphasizes that biodiversity problems are the consequence of habitat quality changes, leading to loss of species or populations that are not able to adapt to these changes. The solution lies in management strategies to preserve high quality habitats where it is still present or to restore to this quality where the habitat has deteriorated. Secondly, the 'conservation genetics' paradigm contextualized the biodiversity problem in the light of characteristics of

populations instead of abiotic habitat characteristics. The problem is a result of populations having small size and high degree of isolation from conspecific populations. Within this paradigm, solutions to biodiversity problems are defined in terms of increasing the number of individuals in small populations or by reducing isolation such as establishing corridors between existing populations. However, this is often not straightforward and an understanding of how evolution effects the dynamics of genetic variation within small, isolation populations is needed. Thus, the field of conservation biology, born from the pivotal book Conservation and Evolution (Frankel & Soulé 1981), aims to inform initiative such as ESA and CITES to halt and reverse biodiversity loss.

Currently, conservation geneticists can employ the power of genomic tools to answer questions in conservation that could not be answered using traditional genetics approaches (Allendorf et al. 2010; Steiner et al. 2013; McMahon et al. 2014; Harrisson et al. 2014; Shafer et al. 2015a; b; Garner et al. 2016; Bernatchez et al. 2017). Technological and analytical advances now allow us to use many thousands of loci, gene expression, or epigenetics to address basic questions of relevance for conservation. These questions include defining management units, estimating effective population size and inbreeding levels, estimating connectivity between these units, detecting and dating of admixture and hybridization events, and identifying loci associated with disease susceptibility and with local adaptation or adaptive potential in species face changing environments (Harrisson et al. 2014; Hoffmann et al. 2015; Jensen et al. 2016; Wade et al. 2016; Bernatchez 2016; Hoban et al. 2016; Flanagan et al. 2017). Genomics allows for quantification of genomic erosion (ez-del-Molino et al. 2018) and genetic load (Robinson et al. 2018). Genomic techniques have also been utilized in historical range delineation (Hendricks et al. 2016; 2017a). Further, genomics can also increase biosecurity, for example, by identifying escapees from fish farms (Wringe et al. 2018). Tools derived from genomics can also improve our ability to monitor biological environmental variation or exploited species, for instance, as provided by the development of environmental DNA (eDNA) and metabarcoding methods (reviewed in Rees et al. 2014;

Thomsen & Willerslev 2015; Barnes & Turner 2015; Deiner et al. 2017). It is vital that these types of genetic information be included in the political decision-making processes to solve the biodiversity problem at both national and global scales.

In this dissertation, I aim to illustrate three examples of utilizing traditional and/or next-generation technologies to address ecological and evolutionary questions rooted in solving issues faced by small or declining, isolated populations at a variety of spatial and temporal scales. First, I estimated pre-disease diversity and population dynamics in an endangered marsupial, the Tasmanian Devil (Hendricks et al. 2017b). Second, I conducted a pre-emptive assessment of contemporary population dynamics in anticipation of population decline in the North American Gray wolves of the Pacific Northwest (Hendricks et al. 2018). Third, I reviewed literature of genomics of cancer susceptibility in wild populations. Fourth, I explored the genomic architecture of the Catalina Island Fox to better assess cancer susceptibility as a wildlife and ecosystem threat. This dissertation adds to body of knowledge showing the power of genomic tools to answer questions in conservation.

1.1 References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature Reviews Genetics, 11, 697-709.

Barnes MA, Turner CR (2015) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics, 17, 1-17.

Bernatchez L (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. Journal of Fish Biology, 89, 2519-2556.

Bernatchez L, Wellenreuther M, Araneda C et al. (2017) Harnessing the Power of Genomics to Secure the Future of Seafood. Trends in Ecology & Evolution, 32, 665-680.

Deiner K, Bik HM, Mächler E et al. (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26, 5872-5895.

ez-del-Molino DD, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L (2018) Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology & Evolution, 33, 176-185.

Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban S (2017) Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evolutionary Applications, 119, 1267-18.

Frankel O, Soulé ME (1981) Conservation and evolution.

Garner BA, Hand BK, Amish SJ et al. (2016) Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. Trends in Ecology & Evolution, 31, 81-83.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. Evolutionary Applications, 7, 1008-1025.

Hendricks SA, Clee PRS, Harrigan RJ et al. (2016) Re-defining historical geographic range in species with sparse records: Implications for the Mexican wolf reintroduction program. Biological Conservation, 194, 48-57.

Hendricks SA, KoblmÃ¼ller S, Harrigan RJ et al. (2017a) Defense of an expanded historical range for the Mexican wolf: A comment on Heffelfinger et al. The Journal of Wildlife Management, 81, 1331-1333.

Hendricks SA, Schweizer RM, Harrigan RJ et al. (2018) Natural re-colonization and admixture of wolves (Canis lupus) in the US Pacific Northwest: challenges for the protection and management of rare and endangered taxa. Heredity, 1-17.

Hendricks S, Epstein B, Schonfeld B et al. (2017b) Conservation implications of limited genetic diversity and population structure in Tasmanian devils (Sarcophilus harrisii). Conservation Genetics, 18, 977-982.

Hoban S, Kelley JL, Lotterhos KE et al. (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist, 188, 379-397.

Hoffmann A, Griffin P, Dillon S et al. (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. Climate Change Responses, 2, 1.

Jensen JD, Foll M, Bernatchez L (2016) The past, present and future of genomic scans for selection. Molecular Ecology, 25, 1-4.

McMahon BJ, Teeling EC, Höglund J (2014) How and why should we implement genomics into conservation? Evolutionary Applications, 7, 999-1007.

McRae L, Deinet S, Freeman R (2017) The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator. (H Doi, Ed,). PLoS ONE, 12, e0169156.

Newbold T, Hudson LN, Hill SLL et al. (2015) Global effects of land use on local terrestrial biodiversity. Nature, 520, 45-50.

Ouborg NJ, Vergeer P, Mix C (2006) The rough edges of the conservation genetics paradigm for plants. Journal of Ecology, 94, 1233-1248.

Rees HC, Ben C Maddison, Middleditch DJ, Patmore JRM, Gough KC (2014) Review: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology (E Crispo, Ed,). Journal of Applied Ecology, 51, 1450-1459.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK (2018) Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology, 28, 3487-3494.e4.

Shafer ABA, Wolf JBW, Alves PC et al. (2015a) Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution, 30, 78-87.

Shafer ABA, Wolf JBW, Alves PC et al. (2015b) Reply to Garner et al. Trends in Ecology & Evolution, 1-2.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. Annual Review of Animal Bioscience, 1, 261-281.

Thomsen PF, Willerslev E (2015) Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183, 4-18.

Wade AA, Hand BK, Kovach RP et al. (2016) Accounting for adaptive capacity and uncertainty in assessments of species' climate-change vulnerability. Conservation Biology, 31, 136-149.

Wringe BF, Jeffery NW, Stanley RRE et al. (2018) Extensive hybridization following a large escape of domesticated Atlantic salmon in the Northwest Atlantic. Communications Biology, 1, 108.

CHAPTER 2

Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*)

2.1 Abstract

Tasmanian devils face a combination of threats to persistence, including devil facial tumor disease (DFTD), an epidemic transmissible cancer. We used RAD sequencing to investigate genome-wide patterns of genetic diversity and geographic population structure. Consistent with previous results, we found very low genetic diversity in the species as a whole, and we detected two broad genetic clusters occupying the northwestern portion of the range, and the central and eastern portions. However, these two groups overlap across a broad geographic area, and differentiation between them is modest (F_{ST}=0.1081). Our results refine the geographic extent of the zone of mixed ancestry and substructure within it, potentially informing management of genetic variation that existed in pre-diseased populations of the species. DFTD has spread across both genetic clusters, but recent evidence points to a genomic response to selection imposed by DFTD. Any allelic variation for resistance to DFTD may be able to spread across the devil population under selection by DFTD, and/or be present as standing variation in both genetic regions.

2.2 Introduction

Species of conservation concern often face a combination of threats, such as loss of habitat, population fragmentation, environmental change, and disease. Genetic diversity is critical to persistence in the face of these threats to maintain population fitness and adaptive potential, including the ability to evolve resistance to emerging infectious diseases (Frankham 2005). Tasmanian devils (*Sarcophilus harrisii*), marsupial carnivores whose geographic range is restricted to the island of Tasmania, exhibit exceedingly low levels of species-wide genetic

diversity as a result of several historical factors (Jones et al. 2004; Miller et al. 2011; Brüniche-Olsen et al. 2014). These factors include bottlenecks in the devil population at the end of the last glacial maximum and during the Mid-Holocene, which likely resulted from climatic fluctuations that limited food availability, and European settlement and bounty hunting (Brüniche-Olsen et al. 2014).

More recently, devils have experienced a dramatic population decline due to an emerging infectious disease, devil facial tumor disease (DFTD). This clonally transmissible cancer has led to a decrease in the total devil population of more than 85% in the last two decades, including local declines of more than 90% (Lachish et al. 2007; McCallum et al. 2007; Jones et al. 2008). First recorded in the northeastern corner of Tasmania in 1996, the disease has spread across most of the island with only a few known populations yet unaffected along the northwestern and southwestern coasts. Extinction has been predicted as a possible outcome (McCallum et al. 2009); however, most local populations have not completely disappeared. In fact, some populations have recently shown an increase in numbers for the first time since DFTD arrival, and there is evidence for a genomic response by the devil to selection imposed by the disease (Epstein et al. 2016). These results suggest the presence of genetic variation in devils for tolerance or resistance to DFTD, despite overall low levels of genetic diversity.

Understanding patterns of genetic diversity and identifying barriers to gene ow in devils are important for several reasons. Genetic patterns may inform epidemiological models of disease spread among devil populations, and may help predict how disease resistance alleles could respond to selection across the species range. An understanding of phylogeographic patterns of genetic diversity may also inform management actions, such as translocations or reintroductions, which may be warranted in response to DFTD, in addition to management of captive populations. Genetic diversity and population connectivity will strongly influence the ability of devils to adapt to other threats such as environmental change and anthropogenic disturbances.

Previous characterizations of genetic diversity and population structure in devils have

focused on relatively few (<12) microsatellite loci (Jones et al. 2004; Brüniche-Olsen et al. 2014), MHC loci (Siddle et al. 2010), mitochondrial genomes (Miller et al. 2011), SNPs (Miller et al. 2011; Wright et al. 2015; Morris et al. 2015), or whole-genome sequencing of two individuals (Miller et al. 2011). Here we use restriction-site-associated DNA sequencing (RAD-seq) (Andrews et al. 2016) to identify and genotype a large number of polymorphic nuclear loci at 38 localities across the species range to assess genome-wide patterns of genetic diversity and population structure. Our findings largely agree with previous results showing low overall genetic diversity and two broad-scale genetic clusters, but we further refine the geographic patterns of genetic diversity within the broad zone of overlap between these clusters.

2.3 Materials and methods

Ear tissue biopsies were collected at 38 locations (1-2 individuals per site) between 1998 and 2009 (Figure 2.1, Table A.1). IRB approval was obtained for tissue collection (Washington State University Institutional Animal Care and Use Committee protocol ASAF 04392; see Hawkins et al. (2006) for trapping protocols). With the exception of a few sites close to the first DFTD appearance in the northeast, sites were free of disease or no longer than 4 years after disease discovery at that location at the time of sampling. DNA was extracted using Qiagen DNA extraction kits. We constructed single-digest RAD-seq libraries (Baird et al. 2008; Etter et al. 2011) for 72 individuals, with pstI as the restriction enzyme to target a relatively large number of loci. Twenty-four individuals were barcoded and multiplexed per lane, in three lanes of paired-end 150 bp reads using an Illumina HiSeq2500. We de-multiplexed and filtered the reads with process_radtags from Stacks v1.20 (Catchen et al. 2013) using the default settings of the -q and -r options, which filter by read quality using a sliding window and rescue barcodes with up to two errors, respectively. We removed PCR duplicates with clone_filter, also from Stacks. Read pairs were aligned to the reference genome (Murchison et al. 2012) using bowtie2 (Langmead and Salzberg 2012) with the

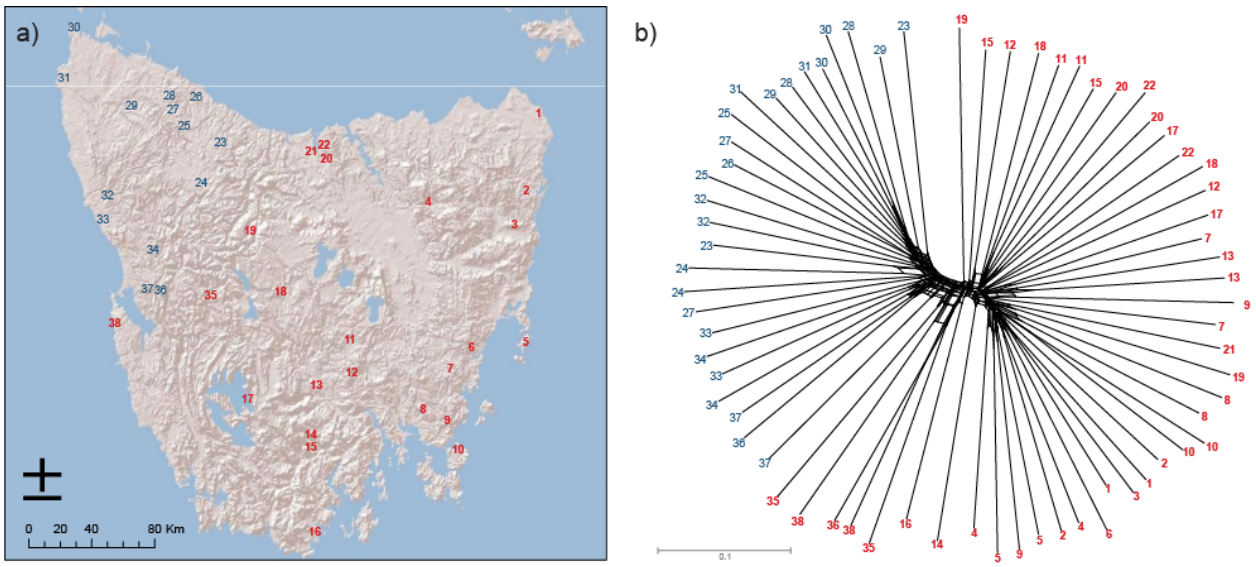

Figure 2.1: a Map of Tasmania with localities of sampling sites. Each number represents a sampling site for one or two individuals. b Neighbor-net consensus tree. Cluster assignment per individual based on STRUCTURE results is indicated by blue for the northwestern cluster and red for the central plateau and eastern coast cluster. Scale bar represents uncorrected P distance. Numbers correspond to the map (Figure 2.1a)

sensitive, end-to-end, and -X 900 options. Alignment files were processed using samtools (Langmead and Salzberg 2012) and we removed reads with a mapping quality <40. To minimize variance in sequence coverage, from this point we retained only the forward reads (containing the restriction enzyme cut sites). We identified and genotyped single-nucleotide polymorphisms (SNPs) using the Stacks reference-aligned pipeline (pstacks, cstacks, sstacks, and populations). The default settings were used except for a minimum stack depth of 3 and the bounded error model with an upper bound of 0.1, to increase sensitivity to minor alleles when PCR duplicates have been removed (Catchen et al. 2013). We dropped two individuals with more than 95% missing data, and we further removed SNPs on the X chromosome, those with an observed heterozygosity greater than 0.5 (to eliminate confounded paralogs), those genotyped in less than one-half of the samples (35 samples), or those with alleles present in only one or two copies. We kept only one SNP per RAD locus to reduce linkage disequilibrium.

For individuals that passed quality filtering (N=70), we calculated the number of segregating sites, mean coverage, Watterson's θ (Watterson 1975) and Tajima's D (Tajima 1989)

using a custom python script. F-statistics were calculated using Genepop (Raymond and Rousset 1995). We tested for isolation by distance (IBD) with a Mantel test in the Vegan package (v.2.0-10; Oksanen et al. 2015) with 1000 permutations (alpha=0.05), using log10-transformed geographic distances and genetic distances, which were calculated in Genepop.

To identify genetic clusters, we used three different analyses. First, we inferred a phylogenetic network with the neighbor-net method in SplitsTree4 (v4.13.1; Huson and Bryant 2006), using uncorrected P distance. Second, we conducted a Bayesian model-based clustering using STRUCTURE v2.3.4 (Pritchard et al. 2000). We used the general admixture model without a priori assumptions about sample locations, and tested K = 1-10. Each K was run with 20 independent iterations, each with 500,000 Markov chain Monte Carlo (MCMC) repetitions following a burn-in of 50,000. We estimated LnP(K) and ΔK using the Evanno method (Evanno et al. 2005) implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012) to obtain the most likely value of K. To combine the multiple iterations for each K, we used CLUMPAK (Kopelman et al. 2015). Third, we incorporated spatial information to inform individual ancestry estimates using TESS3 (Caye et al. 2016), which is useful in determining genetic barriers or genetic discontinuities in continuous populations. The default values of the program were implemented and each run was replicated five times. The optimal value of K corresponded to the minimum of the cross-entropy criterion, across the range K = 1-38.

2.4 Results

There were 175,274 putative RAD loci in the final filtered set, which resulted in 523,386 unfiltered putative SNPs. After all quality filters, RAD sequencing provided 6,362 SNPs. These SNPs were randomly distributed across the genome (Table A.2), with 1,311 SNPs in genic regions based on the annotated reference genome (Ensembl Devil_ref v7.0). Mean coverage across individuals at genotyped loci was ~4.7X. Mean Watterson's θ was 0.00013. The dataset did not deviate significantly from neutrality with a mean Tajima's D value of 0.00019

across loci. Differentiation among all populations was relatively low, with F_{ST}=0.1313 (95% CI: 0.1212-0.1392). We did not find a significant correlation between geographic and genetic distance across all individuals (Figure A.5; Mantel statistic r:0.002719; p=0.48052). Our

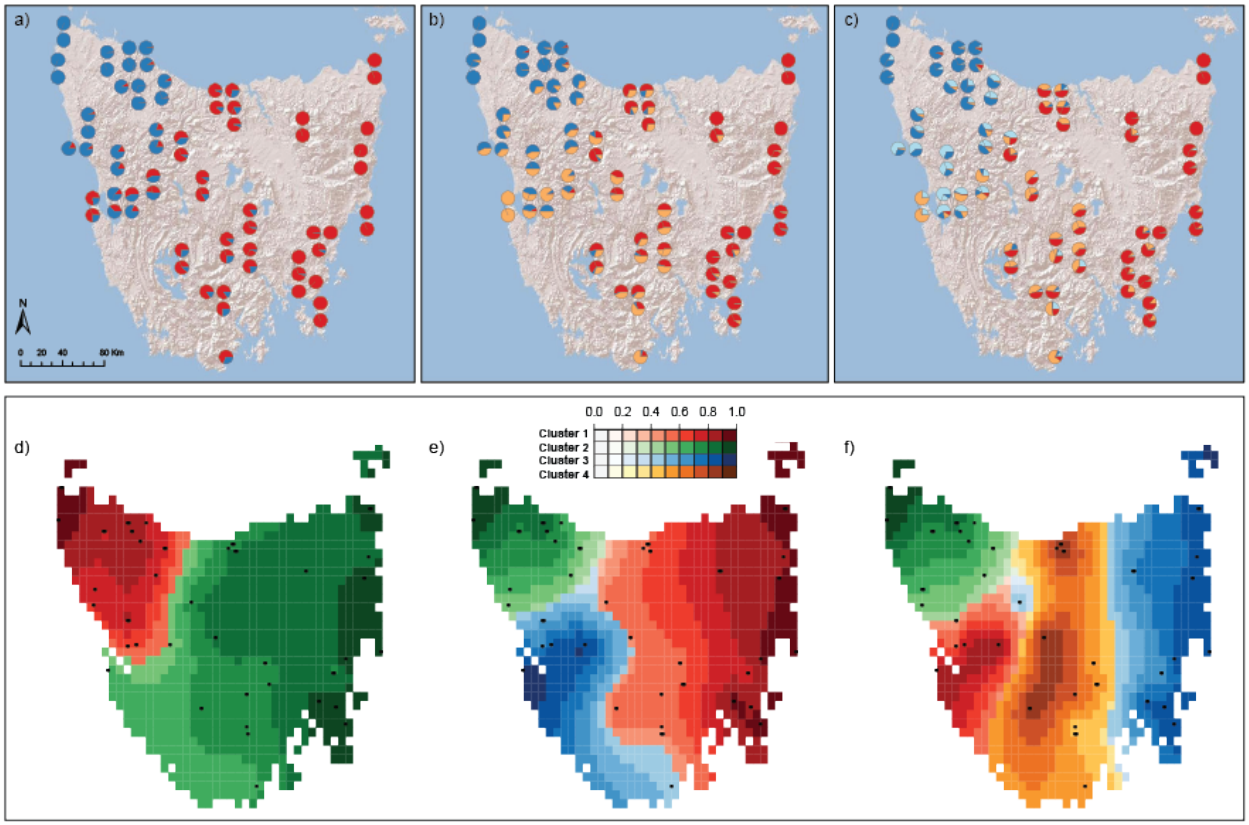

Figure 2.2: Population structure estimated using a-c STRUCTURE or d-f TESS3, based on 70 Tasmanian devil individuals across 38 sites with one to two individuals per site. a-c Pie charts indicate proportion of ancestry per individual; sites with two individuals appear as vertical pairs. d-f Colors indicate ancestry coefficients, and black dots show sampling localities. STRUCTURE showed maximum support for K=2, while TESS3 did not support a best value of K, so here we show a,d K=2, b,e K=3, and c,f K=4

analyses were generally consistent with two major genetic clusters of Tasmanian devils, but with a broad geographically structured zone of admixture between them. The neighbor-net consensus tree grouped the northwestern populations together, and the Central Plateau and east coast populations together (Figure 2.1b). However, there appear to be several sites intermediate to these two clusters, including those in the central-west region around Macquarie Harbor (sites 35-38 in Figure 2.1), as well as broad-scale geographic structuring within each region. Similarly, STRUCTURE analysis showed populations in the northwest corner of the

island are differentiated from the Central Plateau and east coast populations (Figure 2.2a). K=2 received the strongest support (Ln P(K)=-220483.4, ΔK=1910.5; Table A.3), and differentiation between these two clusters was $F_{ST} = 0.1081$ (95% CI: 0.1074-0.1082). However, there is a large degree of uncertainty in determining the optimal value of K (Pritchard et al. 2000; Evanno et al. 2005), and different values of K may reflect different demographic processes, so considering all K values with a biological interpretation is recommended (Meirmans 2015). Here, K=3 and K=4 (Figures 2.2b-c) revealed a broad zone of admixture between the two groups. The TESS3 analysis did not support a best value of K, indicating fine population structure (the cross-entropy curve generated did not exhibit a clear plateau or change in curvature; Caye et al. 2016). The results show similar clustering as the STRUCTURE results when K=2 with populations in the northwestern corner differentiated from all other populations (Figure 2.2d). At higher values of K, south and central-west populations cluster together in K=3 (Figure 2.2e) and an additional Central Plateau cluster forms at K = 4 (Figure 2.2f).

2.5 Discussion

We found low genetic diversity throughout the devil geographic range, based on our estimate of Watterson's θ, consistent with previous results (Jones et al. 2004; Lachish et al. 2010; Siddle et al. 2010; Miller et al. 2011; Brüniche- Olsen et al. 2014; Morris et al. 2015). Major periods of loss of genetic diversity resulting from population declines occurred at the last glacial maximum (~20 k years before present (YBP)) and during El Niño-Southern Oscillation climate cycles during the mid-Holocene (3-5 k YBP). Low genetic diversity predates the isolation of Tasmania from mainland Australia ~13 k YBP (Brüniche-Olsen et al. 2014). Low genetic diversity may be a contributing factor to the rapid spread of DFTD and decline of devil populations (Morris et al. 2015).

Here, we confirmed previous evidence that northwestern populations are differentiated from the eastern populations (Miller et al. 2011; Brüniche-Olsenn et al. 2014), potentially

due to limited dispersal across unsuitable habitat, such as tall wet forest and alpine regions that separate the two areas. Miller et al. (2011) found differentiation between northwestern, central, and eastern coastal populations based on mitochondrial sequence data. Contrary to this clustering, Brüniche-Olsen et al. (2014) found that individuals from Macquarie Heads (western coast) differentiated from northwest populations with Central Plateau and east coast populations being admixed. Similarly, our analysis identified the Macquarie Heads individuals as the only pure representatives of a third population cluster when K=3(Fig.2b).

While our data are consistent with two major genetic clusters, and K=2 was best supported in STRUCTURE, our results from both STRUCTURE and the neighbor-net and TESS3 analyses also highlight the presence of finer-scale genetic structure that may be biologically important (Meirmans 2015). Finer-scale patterns of differentiation and admixture may reflect historical patterns of gene flow. Long distance dispersal (~110 km) has been recorded in devils (Lachish et al. 2010), and our results suggest dispersal that has produced broad areas of admixture among genetic clusters. For instance, one individual from Lake Rowallan (site 19) genetically clusters with the eastern population (Figures 2.1, 2.2), suggesting the possibility of recent long-distance migration. Nonetheless, areas of admixture appear to be geographically well-defined, which would not be expected from frequent long-distance dispersal. It is worth noting that while the central-west and Central Plateau populations may be viewed as zones of admixture given our STRUCTURE results, that is not to say that they do not contain genetic variation that may be unique and important to adaptation (Carvalho et al. 2010).

2.6 Implications for conservation and management

Assessing levels of genetic differentiation in future studies will continue to be important in managing both wild and captive populations. Despite genetic differentiation between eastern and western populations, DFTD has spread across this genetic boundary, indicating widespread genetic susceptibility. However, given decreased population density and disper-

sal distances due to disease (Lachish et al. 2010), population structure may increase in the future due to neutral processes. Our data largely reflect patterns of genetic diversity before the effects of disease on population genetic structure could occur, so they are informative to historic patterns of gene flow and pre-disease distribution of standing genetic variation. A recent study suggests that devils are experiencing a rapid evolutionary response to strong selection imposed by DFTD, likely acting on pre-disease standing variation, with the potential for evolution of resistance to the disease (Epstein et al. 2016). If variation for resistance is not lost due to decreased density and dispersal, resistance alleles may either already occur or spread throughout the population, allowing for range-wide evolution of resistance of DFTD. However, if disease resistance is linked to phylogeographic clusters and cannot spread across the range, artificial movements of animals among genetic clusters should be considered carefully to increase disease resistance, as well as overall genetic diversity across all loci (Frankham 2016). Currently, the Tasmanian devil captive insurance population has individuals from the western side of the island and the north central coast (Hogg et al. 2015). To preserve all potential pre-bottleneck diversity detected in this study, future management practices may consider including individuals from the admixture zones of the central-west and Central Plateau (Figure 2.2c, f). Given that our dataset provides a random sample of the genome, it integrates both neutral and adaptive variation. Including individuals from the admixture zones may maximize genetic diversity and preserve adaptive potential of devils to endure climate fluctuations or future disease epidemics, such as the recently discovered DFT2, a second appearance of transmissible cancer in devils (Pye et al. 2016). Such adaptive potential may rely not only on unique genetic variants, but also on unique multi-locus genotypes that can occur in admixed populations. As recent population declines from DFTD reduce overall genetic diversity and potentially affect population structure, they will likely maintain, if not amplify, the genetic differentiation among geographic areas we have identified here.

2.7 Data availability

The sequence data has been deposited at NCBI under Bio-Project PRJNA306495 and BioSamples SRP076873. The genotype data has been deposited as doi:10.5061/dryad.r60sv in Dryad.

2.8 Acknowledgements

Funding for our work was provided by NSF grant DEB-1316549, NIH grant P30GM03324, and an Australian Research Council Future Fellowship to MJ (FT100100250). We thank the University of Idaho Institute for Bioinformatics and Evolutionary Studies for technical support and resources in the Genomics and Computational Resources Cores; Tamara Max, Mike Miller, Sean O'Rourke, Daryl Trumbo, and Doug Turnbull for assistance with sequencing. We are grateful to Amanda Stalhke, Lisette Waits, and three anonymous reviewers for comments on the manuscript.

2.9 References

Andrews KR, Good JM, Miller MR, et al. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016; 17:81-92. DOI: 10.1038/nrg.2015.28 [PubMed: 26729255]

Baird NA, Etter PD, Atwood TS, et al. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE. 2008; 3:e3376.doi: 10.1371/journal.pone.0003376 [PubMed: 18852878]

Brünich-Olsen A, Jones ME, Austin JJ, et al. Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. Biol Letters. 2014; 10:20140619-20140619. DOI: 10.1098/rsbl.2014.0619

Catchen J, Hohenlohe PA, Bassham S, et al. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013; 22:3124-3140. DOI: 10.1111/mec.12354 [PubMed: 23701397]

Caye K, Deist TM, Martins H, et al. TESS3: fast inference of spatial population structure and genome scans for selection. Mol Ecol Resour. 2016; 16:540-548. DOI: 10.1111/1755-0998.12471 [PubMed: 26417651]

De Carvalho D, Ingvarsson PK, Joseph J, et al. Admixture facilitates adaptation from standing variation in the European aspen (Populus tremula L.), a widespread forest tree. Mol Ecol Resour. 2010; 19:1638-1650.

Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources. 2012; 4:359-361. DOI: 10.1007/s12686-011-9548-7

Epstein B, Jones ME, Hamede R, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications. 2016; 7:12684.doi: 10.1038/ncomms12684

Etter PD, Preston JL, Bassham S, et al. Local de novo assembly of RAD paired-end contigs using short sequencing reads. PLoS ONE. 2011; 6:e18561.doi: 10.1371/journal.pone.0018561 [PubMed: 21541009]

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005; 14:2611-2620. DOI: 10.1111/j.1365-294X.2005.02553.x [PubMed: 15969739]

Frankham R. Genetics and extinction. Biol Conserv. 2005; 126:131-140.

Frankham R. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. Biol Conserv. 2016; 195:33-36. DOI: 10.1016/j.biocon.2015.12.038

Hawkins CE, Baars C, Hesterman H, et al. Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. Biol Conserv. 2006; 131:307-324. DOI: 10.1016/j.biocon.2006.04.010

Hogg CJ, Ivy JA, Srb C, et al. Influence of genetic provenance and birth origin on productivity of the Tasmanian devil insurance population. Conserv Genet. 2015; 16:1465-1473. DOI: 10.1007/s10592-015-0754-9

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23:254-267. DOI: 10.1093/molbev/msj030 [PubMed: 16221896]

Jones ME, Cockburn A, Hamede R, et al. Life-history change in disease-ravaged Tasmanian devil populations. P NATL ACAD SCI. 2008; 105:10023-10027.

Jones ME, Paetkau D, Geffen E, Moritz C. Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. Mol Ecol. 2004; 13:2197-2209. DOI: 10.1111/j.1365-294X.2004.02239.x [PubMed: 15245394]

Kopelman NM, Mayzel J, Jakobsson M, et al. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015; 15:1179-1191. DOI: 10.1111/1755-0998.12387 [PubMed: 25684545]

Lachish S, Jones ME, McCallum H. The impact of disease on the survival and population growth rate of the Tasmanian devil. J Anim Ecology. 2007; 76:926-936.

Lachish S, Miller KJ, Storfer A, et al. Evidence that disease-induced population decline changes genetic structure and alters dispersal patterns in the Tasmanian devil. Heredity. 2010; 106:172-182. DOI: 10.1038/hdy.2010.17 [PubMed: 20216571]

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357-359. DOI: 10.1038/nmeth.1923 [PubMed: 22388286]

McCallum H, Jones ME, Hawkins C, et al. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology. 2009; 90:3379-3392. [PubMed: 20120807]

McCallum H, Tompkins DM, Jones ME, et al. Distribution and Impacts of Tasmanian Devil Facial Tumor Disease. EcoHealth. 2007; 4:318-325. DOI: 10.1007/s10393-007-0118-0

Meirmans PG. Seven common mistakes in population genetics and how to avoid them. Mol Ecol. 2015; 24:3223-3231. DOI: 10.1111/mec.13243 [PubMed: 25974103]

Miller W, Hayes VM, Ratan A, et al. Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). P NATL ACAD SCI USA. 2011; 108:12348-12353. DOI: 10.1073/pnas

Morris KM, Wright B, Grueber CE, et al. Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (Sarcophilus harrisii). Mol Ecol. 2015; 24:3860-3872. DOI: 10.1111/mec.13291 [PubMed: 26119928]

Murchison EP, Schulz-Trieglaff OB, Ning Z, et al. Genome Sequencing and Analysis of the Tasmanian Devil and Its Transmissible Cancer. Cell. 2012; 148:780-791. DOI: 10.1016/j.cell.2011.11.065 [PubMed: 22341448]

Oksanen J, Blanchet FG, Kindt R, et al. Package 'vegan'. Community ecology package, version 2-2. 2015

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945-959. [PubMed: 10835412]

Pye RJ, Pemberton D, Tovar C, et al. A second transmissible cancer in Tasmanian devils. P Natl Acad Sci. 2016; 113:374-379.

Raymond M, Rousset F. An Exact Test for Population Differentiation. Evolution. 1995; 49:1280-1283. DOI: 10.2307/2410454 [PubMed: 28568523]

Siddle HV, Marzec J, Cheng Y, et al. MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. P R Soc B-Biol Sci. 2010; 277:2001-2006. DOI: 10.1098/rspb.2009.2362

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585-595. [PubMed: 2513255]

Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975; 7:256-276. [PubMed: 1145509]

Wright B, Morris K, Grueber CE, et al. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. BMC Genomics. 2015; 16:791-11. DOI: 10.1186/s12864-015-2020-4 [PubMed: 26467759]

CHAPTER 3

Natural re-colonization and admixture of wolves (*Canis lupus*) in the US Pacific Northwest: challenges for the protection and management of rare and endangered taxa

3.1 Abstract

Admixture resulting from natural dispersal processes can potentially generate novel phenotypic variation that may facilitate persistence in changing environments or result in the loss of population-specific adaptations. Yet, under the US Endangered Species Act, policy is limited for management of individuals whose ancestry includes a protected taxon; therefore, they are generally not protected under the Act. This issue is exemplified by the recently re-established grey wolves of the Pacific Northwest states of Washington and Oregon, USA. This population was likely founded by two phenotypically and genetically distinct wolf ecotypes: Northern Rocky Mountain (NRM) forest and coastal rainforest. The latter is considered potentially threatened in southeast Alaska and thus the source of migrants may affect plans for their protection. To assess the genetic source of the re-established population, we sequenced a ~ 300 bp portion of the mitochondrial control region and ~ 5 Mbp of the nuclear genome. Genetic analysis revealed that the Washington wolves share ancestry with both wolf ecotypes, whereas the Oregon population shares ancestry with NRM forest wolves only. Using ecological niche modelling, we found that the Pacific Northwest states contain environments suitable for each ecotype, with wolf packs established in both environmental types. Continued migration from coastal rainforest and NRM forest source populations may increase the genetic diversity of the Pacific Northwest population. However, this admixed population challenges traditional management regimes given that admixture occurs between an adaptively distinct ecotype and a more abundant reintroduced interior form. Our results emphasize the need for a more precise US policy to address the general problem of admixture in the management of endangered species, subspecies, and distinct population segments.

3.2 Introduction

A complication for the conservation of rare and endangered species is the level of protection for admixed populations. Current policy and management protocols favour the biological species concept, where speciation is defined as descent with modification in a reproductively isolated lineage (Dobzhansky 1935; Mayr 1947), and lack guide- lines with regards to admixed populations (reviewed in Jackiw et al. 2015). Consequently, adequate protection may not be granted to taxa that experience a high frequency of gene flow and introgression over the course of their evolutionary histories (Rhymer and Simberloff 1996). Yet, gene flow across species and populations is a natural phenomenon that happens often and may be critical for evolutionary processes (reviewed in Slatkin 1987; Twyford and Ennos 2012; Abbott et al. 2013). Natural admixture allows for the preservation of the historical genetic connectivity between populations. Although admixture may result in the loss of population specific adaptations (e.g. Rhymer and Simberloff 1996; Muhlfeld et al. 2009), it may also provide the genetic variation on which selection can act (e.g., Smith et al. 1997b; Hedrick 2013). If admixture provides genetic variation, it may be as important to maintaining adaptive potential, the capacity for future evolutionary change, as evolution by natural selection or genetic drift within reproductively isolated lineages (Arnold 2016; vonHoldt et al. 2017). This concept of species as fixed entities with complete reproductive isolation is currently used in many cases of species management. However, the "web-of-life" (WOL) framework is a more realistic concept that acknowledges that horizontal gene transfer through hybridization, introgression and reticulate evolution is prevalent in some systems (Arnold and Fogarty 2009). Therefore, this framework includes adaptive potential by protecting genomic and phenotypic diversity, ecological function and resilience and does not preclude an individual, population or species from protection due to natural admixture (Arnold 2016; vonHoldt et al. 2017).

Appropriate management and protection of natural re- colonized and admixed popu-

lations involving an endangered source is a critical challenge that requires case-by- case solutions (Allendorf et al. 2001; vonHoldt et al. 2017). Each context requires the initial detection and identification of hybrid individuals, which is not a trivial task (Allendorf et al. 2001). Recent advances in high-throughput genomic resources have greatly increased our ability to detect and characterize admixture in hybrid populations (Allendorf et al. 2010; Hohenlohe et al. 2011; vonHoldt et al. 2013; Abbott et al. 2016; Wayne and Shaffer 2016) and provide evidence of natural movement of genetic variation across populations and species (the WOL framework; vonHoldt et al. 2017). Once the occurrence of natural admixture is established, two questions should be answered before management decisions are made, as suggested by Wayne and Shaffer (2016). First, do admixed individuals perform ecosystem functions and services that are similar to those performed by the endangered entity? Second, would habitat restoration for the native endangered entity enable natural selection to increase the proportion of genetic variants characteristic of the native endangered entity (Wayne and Shaffer 2016)? These questions as well as the WOL framework establish an inclusive foundation for the management of admixed populations in the light of evolutionary and ecological principles.

One case of natural admixture that may highlight the importance of the WOL framework is that of the North American grey wolf (*Canis lupus*). The natural re-colonization of previously extirpated populations of large carnivores is rare but has been documented in wolves in North America and Europe (Chapron et al. 2014). Historically, wolves were common in the Pacific Northwest (PNW) region of North America (Bailey 1936; Young and Goldman 1944) but were extirpated in the US portion by the mid-1930s (Bailey 1936; Verts and Carraway 1998). Wolves naturally re-colonized northwest Montana (MT), USA, from Alberta (AB) and British Columbia (BC), Canada, in the mid-1980s (Boyd et al. 1995). More recently, wolves naturally re-colonized the states of Oregon (OR) and Washington (WA), USA, and as of 2015, 13 and 18 packs inhabit those states, respectively. Given that individuals disperse an average of 50-100km or up to several hundred kilometres before

establishing territories (Mech 1970; Fritts 1983; Merrill and Mech 2000; Jimenez et al. 2017), these re-established wolves in OR and WA are suspected to be migrants from adjacent wolf populations that consist of two ecotypes, the coastal (referred to as rainforest wolves in some previous studies) and the Northern Rocky Mountain forest (NRM) populations. These two ecotypes represent locally adapted and specialized wolves with respect to den-site use, foraging habits, physiology and prey specialization (Fritts et al. 1995; Mladenoff et al. 1995; Paquet et al. 1996; Mladenoff et al. 1997; Mladenoff and Sickley 1998; Haight et al. 1998; Mladenoff et al. 1999; Callaghan 2002) and exhibit environmentally driven genetic differences between coastal and NRM populations (Weckworth et al. 2005; Muñoz Fuentes et al. 2009; Weckworth et al. 2011; Schweizer et al. 2016a, b). Previous studies aimed at identifying ecotype-specific selection in NRM and coastal wolves found multiple signals of selection on genes related to dentition, diet, metabolism, musculature, organismal system, skeletal morphology and vision (Schweizer et al. 2016a, b). Furthermore, the coastal wolf population often has high allelic differentiation from all other populations, especially for candidate single-nucleotide polymorphisms (SNPs) under selection (Schweizer et al. 2016a, b).

The origins of some of the WA and OR population are from geographically proximate NRM wolves that in turn derive from wolves reintroduced to Idaho (ID) from northern, interior BC and AB as well as the naturally re-established wolves from MT (Jimenez et al. 2017). However, it is unknown if immigration, territory establishment and subsequent breeding of coastal wolves in the PNW has also occurred. At least two important questions currently remain unresolved: (1) is genetic admixture between divergent lineages (ecotypes) occurring in the re-established PNW population?; and (2) can the habitat of PNW region support these two ecotypes in a potential admixture zone?

Any potential admixture may have conservation implications for the wolves of the PNW region. Coastal wolves comprise genetically contiguous populations in coastal BC and the Alexander Archipelago in southeast Alaska (AK, USA; Weckworth et al. 2010, 2011). The

Alexander Archipelago wolves were considered for protection under the USA Endangered Species Act (ESA) due to human-mediated habitat alteration that resulted in a 60% decline in the population over 1 year (Toppenberg et al. 2015). Although this wolf population was ultimately not listed, it is still impacted by legal and illegal hunting and deserves special consideration as a unique ecotype not found outside this area (Muñoz Fuentes et al. 2009; Schweizer et al. 2016a, b). Furthermore, admixture is likely recent and therefore not yet in equilibrium, thus warranting protection as the populations stabilize. If Alexander Archipelago wolves attain protected status under ESA, then admixture, if found, should influence how the wolves of the PNW are managed. Protection status should be granted given several conditions: (a) admixture is a result of natural patterns of wolf dispersal; (b) historical genetic connectivity is preserved (Hendricks et al. 2015); and (c) adaptive potential is maintained, following the WOL framework (Wayne and Shaffer 2016; vonHoldt et al. 2017).

We used three complementary approaches to assess the source populations of PNW wolves and their suitability to areas of reintroduction. First, we sequenced a portion of the mtDNA control region in wolves from WA, OR, and surrounding populations to establish maternal lineages. Second, we used SNPs obtained through targeted DNA capture and sequencing to estimate local population structure, ancestry and relatedness among individuals. Third, we assessed habitat preference of re-established wolf packs in the PNW region. To do so, we used ecological niche models (ENMs) based on climate predictors to identify appropriate habitat for the NRM and coastal wolf ecotypes and then mapped centroid locations of existing WA and OR packs (as of 2015) to assess potential genetic barriers associated with environmental differences. Results from these approaches establish the source populations for naturally re-established wolf populations and better inform the conservation and management of the wolf populations in WA and OR.

3.3 Methods and materials

3.3.1 Sample collection

Tissue and/or blood samples from grey wolves were collected in WA by the Washington Department of Fish and Wildlife (WDFW) and in OR by the Oregon Department of Fish and Wildlife (ODFW) during radio collaring efforts of live animals or from carcasses. We analysed 32 samples from OR wolves (collected by ODFW 2009-2013) and 22 samples from WA (collected by WDFW 2008-2012). Scat or bone and tissue samples from wolf carcasses in coastal BC were collected as a part of a long-term study (Darimont et al. 2008). Appropriate permits were granted to researchers for sample collection. Additional samples from interior Yellowstone National Park (YNP), ID, MT, BC, AB, and coastal BC and Alaska were selected from a set of samples of grey wolves used in previous studies (vonHoldt et al. 2010, 2011). DNA was extracted from tissue, blood, bone and scat samples using the standard commercial kit protocols (Qiagen DNA QiaAmp minikit and EZNA Stool Kit from Omega Biotek).

3.3.2 Species determination

To confirm the species and subspecies status of samples, a panel of 24 species-diagnostic markers was used to resolve the contributions of ancestry from the grey wolf (*C. lupus*), domestic dog (*C. lupus familiaris*) and coyote (*C. latrans*) and resolve first- and second-generation dog-grey wolf hybrids (vonHoldt et al. 2013). These markers (20 resolving wolf vs. dog, two resolving dog vs. coyote and two resolving coyote vs. wolf) were identified and confirmed against a panel of 832 dogs, 180 grey wolves and 53 coyotes analysed on the Affymetrix Canine SNP v2. microarray (vonHoldt et al. 2013). The markers were assayed using a quantitative polymerase chain reaction (qPCR) High-Resolution Melting (HRM) assay and Roche LightCycler 480 instrument (Indianapolis, IN). Two dog, two western coyote

and two western Canadian grey wolf samples were used as non-admixed references for allele calls. The qPCR reaction mixes and thermocycler conditions followed vonHoldt et al. (2013). HRM melt temperatures were analysed using the Roche LightCycler 480 Software v1.5.0.

3.3.3 Mitochondrial haplotype determination

To further confirm the species or subspecies status of samples, eliminate possible coyote/wolf or dog/wolf hybrid individuals and assess the distribution of mtDNA haplotypes, a 318-bp portion of mitochondrial control region (MT-CR) was amplified using two sets of overlapping primer pairs: (i) Thr-L (Vilà et al. 1999) and ddl5R (Leonard et al. 2002); and (ii) ddl1F and ddl2R following the protocol in Leonard et al. (2002). Amplified PCR products were sequenced in both directions using BigDye on an ABI3730XL capillary sequencer (Applied Biosystems, Inc). Sequences were visualized and aligned in Geneious 6.0.5 (Biomatters). A local BLAST search (Altschul et al. 1997) was performed on individual consensus sequences against Genbank partial MT-CR haplotypes (n = 75 globally distributed grey wolves; n = 125 coyotes; n = 1 red wolf (*C. rufus*); n = 30 domestic dogs). Putative assigned haplotype matches were confirmed by BLAST searches against the Genbank reference nucleotide database.

3.3.4 Capture array library preparation

To determine the source populations of individuals from PNW, 96 individuals (coastal BC-8; inland BC-5; MT-23; ID-17; YNP-6; WA-16; OR-21) were sequenced using a custom capture array designed to target sequences from 1040 candidate genes and 5073 1kb neutral regions from the dog reference genome (CanFam3.1; Schweizer et al. 2016a). Putatively neutral regions were identified using methods described by Freedman et al. (2014) and for which there exists a precedent in humans (Wall et al. 2008) and wolves (Schweizer et al. 2016a). First, genic regions from the dog reference genome (CanFam3.1) were identified using annotations from the union of refGene, Ensembl and SeqGene annotation databases.

All annotated transcripts had proper start and stop codons and contained no internal stop codons. Second, 1kb neutral regions were chosen using the following characteristics: (1) minimally 100kb from any known or predicted genes (based on observed levels of linkage disequilibrium (LD) in wolves (Gray et al. 2009); (2) not located within highly repetitive regions of the dog genome; (3) uniquely mapping regions of the genome as computed by TALLYMER (Kurtz et al. 2008); (4) phastCons scores <0.5 (Siepel et al. 2005); and (5) GC content within two standard deviations of the mean dog genome GC content.

DNA quantity and quality were assessed with the Qubit Fluorometer High Sensitivity Kit and visualization after electrophoresis on a 2% agarose gel, respectively. Samples with at least 600-1000 ng of dsDNA and a molecular weight of >1kb were sheared to ~ 300-500 bp using a Bioruptor NGS Sonication System (Diagenode). Sequencing libraries for each individual were prepared using a with-bead library preparation protocol (Faircloth et al. 2013) that included labelling with a unique 6-bp index (Faircloth and Glenn 2012). Two individual libraries were pooled and allowed to hybridize to the array for 24 hours. Each pool was target-enriched and PCR-amplified according to the MYbaits protocol (MYcroarray), with modifications as in Schweizer et al. (2016a). Before sequencing, 24 individuals (12 capture libraries) were pooled and enriched libraries were run on two lanes with 100-bp paired-end sequencing on an Illumina HiSeq 2000.

3.3.5 Sequence alignment and processing

The Broad Institute GATK v2.6-4 "Best Practices" pipeline was used for sequence alignment and processing. Demultiplexed fastq reads that passed the Illumina filter using *fastq_illumina_filter 0.1* were trimmed for adapter sequences and a minimum base quality of 20 using *trim_galore 0.3.1*. Aligned forward and reverse reads were mapped to the reference dog genome (CanFam3.1) using *bwa aln* (seed length of 28) and *bwa sampe* (insert size of 1000 bp; (Li and Durbin 2009). Once duplicates were removed using *samtools rmdup*, a local realignment was completed using GATK 2.6-4 (DePristo et al. 2011). Mate information

was fixed with picard tools. GATK *Base Quality Score Recalibration (BQSR)* was performed using the *-knownSites* flag with the final SNP set from Schweizer et al. (2016a). After adding in 38 samples from coastal BC, coastal AK, AB and YNP that were previously enriched and sequenced using the same protocols (see Supplementary Table S1; Schweizer et al. 2016a, 2018), SNPs were called using the GATK *Haplotype Genotyper* algorithm.

3.3.6 Array variant filtering and final sample set

Variant filtration was completed using ten filter expressions recommended by the GATK "Best Practices" pipeline. Variants with a depth of coverage >10 and minimum genotype quality >30 were kept in the final data set. The VCFtools package (Danecek et al. 2011) was used to assess the quality of filtered and aligned reads. Sites called in <95% of individuals were subsequently removed from further analysis. The number of segregating sites and mean coverage per individual was calculated using VCFtools. Data sets were LD-pruned using PLINK (*-indep-pairwise 50 5 0.5*; Purcell et al. 2007). We used the programs KING v1.4 (Manichaikul et al. 2010) and PRIMUS v0.5 (Staples et al. 2013) to calculate relatedness and then removed one individual per related pair with a pairwise identity-by-state greater than or equal to 0.5, calculated from the LD-pruned data set. We used four data sets for subsequent analyses: (1) all individuals, all loci; (2) all individuals, putatively neutral loci; (3) unrelated individuals, all loci; and (4) unrelated individuals, putatively neutral loci.

3.3.7 Population structure, individual assignment and gene flow estimates

To verify genetic differentiation between ecotypes, we used VCFtools to calculate Weir and Cockerham's (1984) θ, an estimator of F_{ST} (Wright 1951). To genetically assign and determine ancestry of WA and OR wolves to YNP, ID, MT, AB, interior BC or coastal BC populations, we applied two clustering methods to both SNP data sets with unrelated

individuals. First, principal component analysis (PCA) was performed using SMARTPCA within EIGENSTRAT v3.0 (Price et al. 2006). Second, using the default settings, we applied the program ADMIXTURE (Alexander et al. 2009) to partition and classify individuals into K = 1 through K = 10 clusters. Inbreeding coefficients (F_{IS}) were also calculated using VCFtools.

We used two methods to identify potential related individuals across state boundaries and estimate levels of gene flow. We used the data set with all individuals and putatively neutral, LD-pruned loci. First, pairwise relatedness values >0.1 were used to identify individuals that may have shared ancestry with individuals from other geographic locations sampled in this study. Second, we used the program BayesAss v3.0.4 (Wilson and Rannala 2003), which is a Bayesian assignment test that estimates individual ancestry. The program was run for 10 million Markov chain Monte Carlo (MCMC) iterations with the first 1 million iteration discarded as burnin and sampling every 100 iterations. Adjusted mixing parameters for migration rate (m = 0.2), allele frequency (a = 0.5) and inbreeding coefficients (f = 0.002) were used to optimize the acceptance rate.

3.3.8 Ecological niche modelling

Preparation of ecotype occurrence data

Given the genetic differentiation between ecotypes and adaptive distinction of the coastal wolves (Mu/ noz Fuentes et al. 2009; Schweizer et al. 2016a, b), we predicted the optimal environmental niches of the coastal and NRM forest ecotypes to assess whether environmental differences occur within the PNW region and whether habitat differences between coastal and NRM forest act as barriers to gene flow for the re-established packs in WA and OR. Non-duplicate localities for coastal wolves (coastal BC and southeast Alaska; n = 20) and NRM forest wolves (interior BC, AB, MT, WY, ID; n = 119) were compiled using data from the Global Biodiversity Information Facility portal (www.gbif. org) and voucher museum

specimens (Supplementary Figures S7 and S8). These localities do not include occurrence points of individuals sampled for DNA. When exact geographic coordinates for a specimen were not available, we used the provided location name to estimate the geo-reference for the individual. Localities for which geo-referencing could not be defined more precisely than the level of county or similar administrative unit were excluded.

Preparation of environmental data

Bioclimatic variables from WorldClim v1.4 (Hijmans et al. 2005) were selected according to their roles in determining the physiological limits of species (e.g., variation in annual means, extremes and seasonality of temperature and precipitation). These bioclimatic variables (n = 19) are at a 1km resolution and metrics are derived from monthly interpolated temperature and rainfall climatologies spanning the years 1950 to 2000 (Hijmans et al. 2005). For each wolf ecotype, clusters of highly correlated variables were identified and removed to trim variables that were not contributing to the model (Harrigan et al. 2014). The nine variables used in both models were: annual mean temperature, annual precipitation, precipitation seasonality, precipitation of the warmest quarter, precipitation of the coldest quarter, mean diurnal range, temperature seasonality, maximum temperature of the warmest quarter, and maximum temperature of the coldest quarter. Elevation (SRTM) was also used in both models. Current vegetation data were not included in this analysis because vegetation patterns are more intensively influenced by anthropogenic activities (e.g. deforestation, land cover conversion, urban development and road network intensification) and therefore could change quickly. Additionally, because samples spanned multiple years, accurate vegetation data could not be obtained.

MaxEnt modelling

We ran MaxEnt v3.3.3k (Phillips et al. 2006), which uses a probabilistic framework, to model the environmental niches from occurrence data (described above). Its main assump-

tion is that the incomplete empirical probability distribution (which is based on the species occurrences) can be approximated by a probability distribution of maximum entropy (the MaxEnt distribution) subject to certain environmental constraints and that this distribution approximates a taxon's potential geographic distribution (Phillips et al. 2006). The use of MaxEnt is advantageous in the study of endangered taxa, for which locality data may be sparse, as it performs well with only a small number of point localities (Jordan and Ng 2002; Hernandez et al. 2006; Wisz et al. 2008). Unlike many other algorithms, it requires only presence data to assign spatially explicit probabilities of occurrence (Phillips et al. 2006) and it consistently ranks high in inter-model comparisons (Elith et al. 2006; Diniz-Filho et al. 2009; Harrigan et al. 2014). Further, several recent studies have shown that MaxEnt performs successfully in modelling the distribution of motile species (Rodrìguez Soto et al. 2011; Lv et al. 2011; Blair et al. 2013).

In this study, we used the MaxEnt default settings of convergence threshold (10^{-5}) and 100 cross-validated replicates. This cross-validation replicate process involved the random splitting of occurrence data into a number of equal-sized groups, known as "folds", where models were created leaving out one fold for each run. For each replicate, the excluded fold is used to evaluate the model (Phillips et al. 2006). The study area over which the potential distribution is computed, and from which the MaxEnt algorithm samples "background" points to train the model, are substantially larger than the known ranges of the ecotypes. We verified that modelling results were insensitive to the choice of study area size by building models with progressively larger study areas, increased at an increment of 5°latitude and longitude (data not shown). Regularization attempts to balance model fit and complexity, with the default setting multiplying each automatic regularization parameter by one. Additional multiplication of these parameters tends to smooth (make the model more generalized) at the expense of model fit (Elith et al. 2011). For comparisons of models, we chose to leave regularization parameters the same across all runs (r = 1), particularly because default settings represent a conservative approach to estimating species distributions

based on occurrences. Using ENMeval (Muscarella et al. 2014), we tested for the effects of spatial autocorrelation and model complexity by running each ecotype occurrence data set through several different partitions as recommended by the authors.

MaxEnt produces a continuous prediction with values ranging from 0 to 1 (in units of probability of occurrence) indicating least suitable to most suitable conditions for the taxa under consideration (Phillips et al. 2006). To convert this continuous output into a binary prediction that approximates the potential distribution, we used a prob- ability threshold equivalent to the minimum predicted probability of occurrence at actual occurrence localities used to train the model (Phillips et al. 2006).

Testing model performance

Model performance was evaluated by the area under the curve (AUC), which is often used to measure model performance (Rödder et al. 2009; Harrigan et al. 2010; Fourcade et al. 2014; Sesink Clee et al. 2015). AUC values were calculated by comparing model performance to a random model of associations between presence localities and environmental predicting factors (DeLong et al. 1988). AUC values range from 0.5 to 1.0; with values close to 0.5 corresponding to a model that is no better at predicting an ecological niche than a random model and a value of 1.0 corresponding to a model with a perfect fit.

Probability of occurrence

As of 2015, 31 wolf packs inhabited the PNW states of WA (n = 18) and OR (n = 13). The centroid location of each pack was used as the proxy for area used by each pack. The resulting ENMs were used to calculate the probability of each PNW wolf pack's occurrence in coastal or interior environments.

3.4 Results

3.4.1 Species determination and mitochondrial haplotype determination

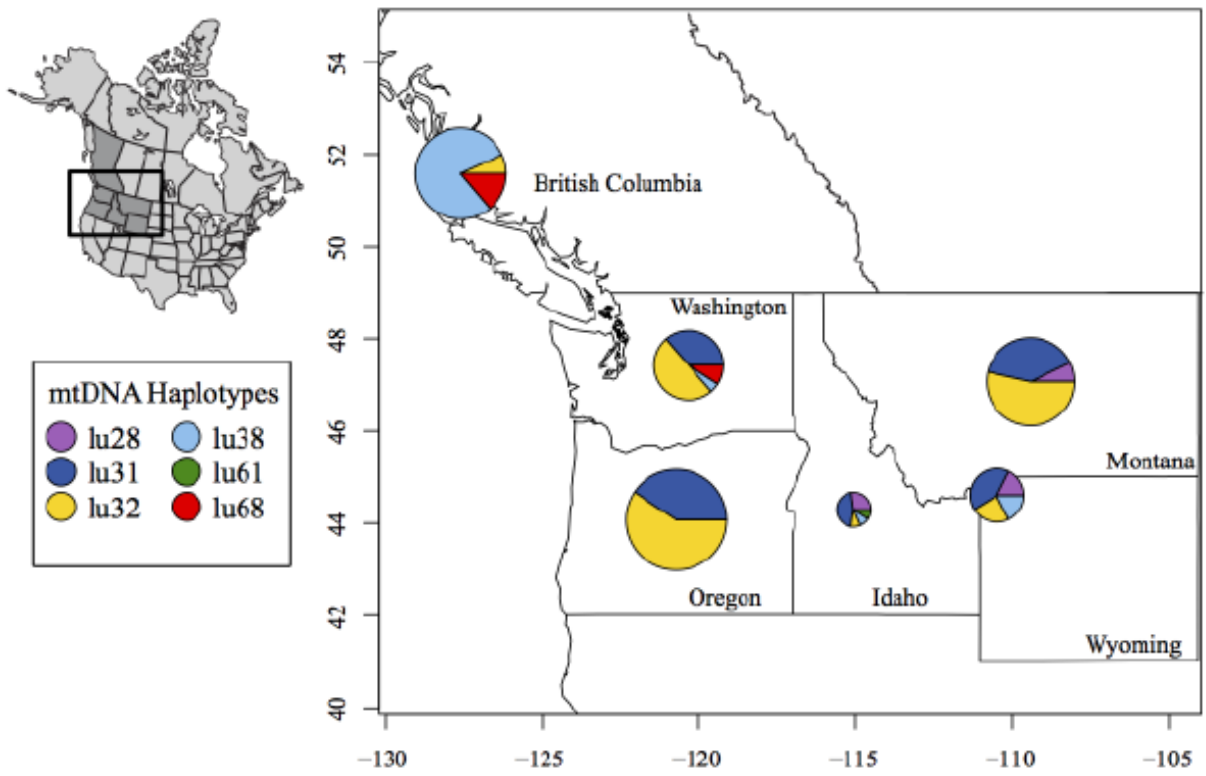

Figure 3.1: Distribution of mtDNA control region sequence haplotypes found in each population among wolves of the Pacific Northwest. Sizes of pie charts are proportional to the number of samples per location, and colours in pie charts represent one of the six mtDNA haplotypes (see key; Tables B.1 and B.2)

Genotypes from the panel of 24 species-diagnostic markers identified all modern samples as pure grey wolf, with no evidence of recent dog or coyote ancestry. Therefore, we sequenced 139 wolf samples for a 318 bp fragment of the mitochondrial control region to determine haplotype distributions (Figure 3.1, Table B.1 and B.2). Relative to other North American populations, the diversity of maternal lineages in PNW wolves was slightly lower than average (H = 4; Figure 3.1 and Table B.2, average in NA = 4.5; Table 5 in Chambers

et al. 2012). We have not included the Great Lakes or eastern Ontario populations in calculations due to taxonomical conflicts and admixture events with coyotes. Interestingly, two WA individuals had haplotype lu68 (Figure 3.1 and Tables B.1 and B.2), which is otherwise known only from populations in coastal BC (current study: N = 4 of 29; Figure 3.1 and Tables B.1 and B.2) and previous studies (Muñoz Fuentes et al. 2009, 2010; Weckworth et al. 2010). Consequently, the mitochondrial haplotype analysis confirms our diagnosis that the samples are of grey wolf ancestry and suggests some gene flow from the coastal ecotype (coastal BC and southeast AK) to the population in WA with the principal influence from NRM wolves.

3.4.2 Capture array sequences

We obtained high-quality sequence reads with a per individual average unfiltered yield of 2254.62 ±954.12 Mb, 92.92 ±3.74% raw reads passing Illumina filters and a mean quality of 36.90 ±1.84. After processing and removing low quality reads, 80 ±16.3% of raw reads mapped uniquely to the dog reference genome (i.e. after PCR duplicate removal). After genotyping, quality filtering and removing low coverage (<10x) individuals (n = 9), the mean depth of coverage over all regions on the capture array was 89.79 ±35.13 (Figure B.1), with a mean depth of coverage over neutral regions of 137.95 ±52.33 (Figure B.1). Mean depth of coverage for scat samples from coastal BC over all regions was 22.37 ±26.94 and ranged from 0.99 to 59.89 (Figure B.2), with a mean neutral depth of coverage of 34.23 ±40.87 that ranged from 1.52 to 92.64 (Figure B.2). We excluded data from half (n = 3) of our original faecal samples due to low coverage (<10x) that may be a result of complications with DNA extraction and library preparation (see Discussion in Appendix B.3).

After filtering genotypes, we separated data into two sets consisting of all variant loci (92,296 SNPs) and variant loci within neutral regions (41,735 SNPs). The transition to transversion ratio for all regions was 2.31 and for neutral regions was 2.23, which is similar to previously reported values in wolves (Freedman et al. 2014; Zhang et al. 2014; Schweizer

et al. 2016a). After LD-pruning, there were 35,406 variable positions, of which 18,508 were within neutral regions. After removal of one individual per related pair (parent-offspring or full siblings; N = 43 individuals), the remaining set of 83 individuals included 9 ID, 13 MT, 28 YNP, 7 AB, 9 coastal BC, 4 interior BC, 5 WA and 8 OR wolves.

3.4.3 Population structure, individual assignment and gene flow estimates

We calculated pairwise F_{ST} between population pairs to verify that that there is differentiation between ecotypes using our data set. Mean among-population pairwise F_{ST} of the 18,508 LD-pruned neutral SNPs was moderate. The coastal wolves were the most differentiated by this measurement with pairwise F_{ST} values ranging from 0.104 between coastal and WA to 0.170 between coastal and MT populations (Table 3.1). MT was the second most differentiated population with F_{ST} values that ranged from 0.034 (YNP) to 0.072 (AB). Excluding coastal wolves, F_{ST} values ranged from 0.001 between interior BC and WA to 0.072 (between AB and MT, Table 3.1). Inbreeding coefficients (F_{IS}) were near zero for all populations (F_{IS} = -0.0115 to 0.0425; Table B.3), except the coastal BC population (F_{IS} = 0.1116; Table B.3).

Table 3.1: Mean pairwise F_{ST} (above diagonal) and weighted pairwise F_{ST} (below diagonal) for 18,508 LD-pruned SNPs within neutral regions in 87 unrelated individuals. Population abbreviations as follows: interior British Columbia: iBC; coastal British Columbia: cBC; Idaho: ID; Montana: MT; Oregon: OR; Washington: WA; Yellowstone National Park: YNP

Populations	Alberta	iBC	cBC	ID	MT	OR	WA	YNP
Alberta	-	0.0060	0.0824	0.0190	0.0565	0.0273	0.0152	0.0460
iBC	0.0194	-	0.0782	0.0061	0.0366	0.0121	-0.0055	0.0297
cBC	0.1274	0.1137	-	0.0811	0.1087	0.0856	0.0704	0.0929
ID	0.0261	0.0141	0.1282	-	0.0384	0.0045	0.0136	0.0086
MT	0.0725	0.0418	0.1704	0.0526	-	0.0344	0.0137	0.0497
OR	0.0393	0.0191	0.1318	0.0093	0.0453	-	0.0123	0.0275
WA	0.0293	0.0010	0.1045	0.0240	0.0149	0.0201	-	0.0356
YNP	0.0454	0.0315	0.1387	0.0088	0.0697	0.0338	0.0401	-

To assess the genetic partitions based on SNP data, we used complementary analyses of

genetic clustering patterns. Plots of the first two components from PCA show several distinct clusters that correspond with sampling location (Figure 3.2). On both PC1 and PC2, the individuals of the coastal BC and AK population cluster distinctly, with no other samples occurring within the 95% confidence interval (CI) (Figure 3.2). The NRM populations (BC, AB, ID, YNP, MT and OR) are more continuous in PC space with less distinct clustering than the coastal population. The OR samples cluster together and are spatially between the YNP and MT clusters. MT individuals form a 95% CI cluster that includes one known migrant from ID (vonHoldt et al. 2010), one individual from interior BC and most of the WA individuals. Three WA individuals fall intermediate to the coastal population and the NRM populations. PCA results were similar between all loci and neutral loci for both data sets (all 126 and 83 unrelated individuals; Figure 3.2; Figure B.3).

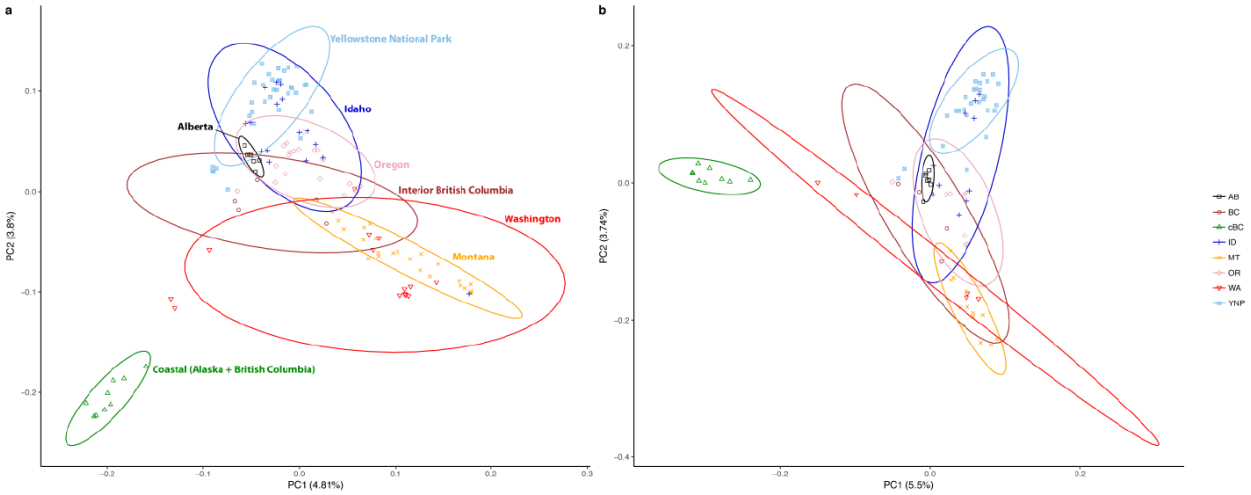

Figure 3.2: a. PCA plot of all 126 wolves for LD-pruned data set (18,508 SNPs) with 95% confidence intervals. b. PCA plot of 83 unrelated wolves for same LD-pruned data set with 95% confidence intervals. Population abbreviations as follows: AB Alberta, Canada; BC interior British Columbia, Canada; cBC coastal British Columbia, Canada and Alaska, USA; ID Idaho, USA; MT Montana, USA; OR Oregon, USA; WA Washington, USA; YNP Yellowstone National Park, USA

Our second approach to assess genetic clustering used ADMIXTURE, which showed the best-supported number of clusters equal to 3 as evident by the lowest cross-validation error rate (Figures 3.3, B.4, B.4). These three clusters represent the naturally re-established MT population, the reintroduced YNP and ID population and the coastal population (Figure B.5). However, the AB individuals form an additional, biologically relevant cluster at K = 4

(Table B.4). Therefore, we present assignment proportions from the results of K = 4. Two WA individuals have high assignment (>49%) to the AB cluster, with signatures of ancestry to the coastal population. The other three WA individuals assign to the MT cluster (>90%). All OR individuals have ancestry from NRM populations. Of the four YNP individuals that strongly assign to the AB cluster, three are founders (assign to the AB cluster with >67% ancestry) and were reintroduced from northern AB and BC populations (vonHoldt et al. 2010). Four ID individuals assign to the AB cluster (>50%), three of which are founders and assign to the AB cluster with >88.5% ancestry. Given our data set included samples spanning multiple time points occurring during wolf introduction, we wanted to test for temporal effects on the number of genetic clusters. ADMIXTURE analysis was re-run under the same parameters after removing known, unrelated founders from YNP (n = 3) and ID (n = 3). The same four clusters are observed after removing the founders (Figure 3.3), which indicates that there is little detectable temporal effect on allele frequencies per population. The cross-validation error values for ADMIXTURE runs of unrelated individuals for all data or only neutral data indicate an optimal K = 3, as described above (Figure B.4). When the coastal individuals were excluded from ADMIXTURE analysis (Figure B.6), K = 2 had the lowest CV error for both sequence data sets. This result further suggests that the coastal ecotype is a valid cluster.

To identify individuals that might be related across state boundaries, we filtered the pairwise relatedness data to include pairs related above 0.1 and from different populations (Table B.5). Individuals from AB and coastal populations (including coastal BC and AK wolves) did not have pairs from differing populations with relatedness values above 0.1. All other populations were included in this filtered subset. At the relatedness level of ~0.5 (full sibling or parent-offspring), one WA individual (WA010817_WA) was related to three OR individuals (OR11_OR, OR16_OR, OR10_OR). Individuals from MT, ID and YNP were found to have levels of relatedness from 0.1 to 0.5. This finding supports previous work highlighting gene flow among wolf populations of the Northern Rocky Mountains (vonHoldt

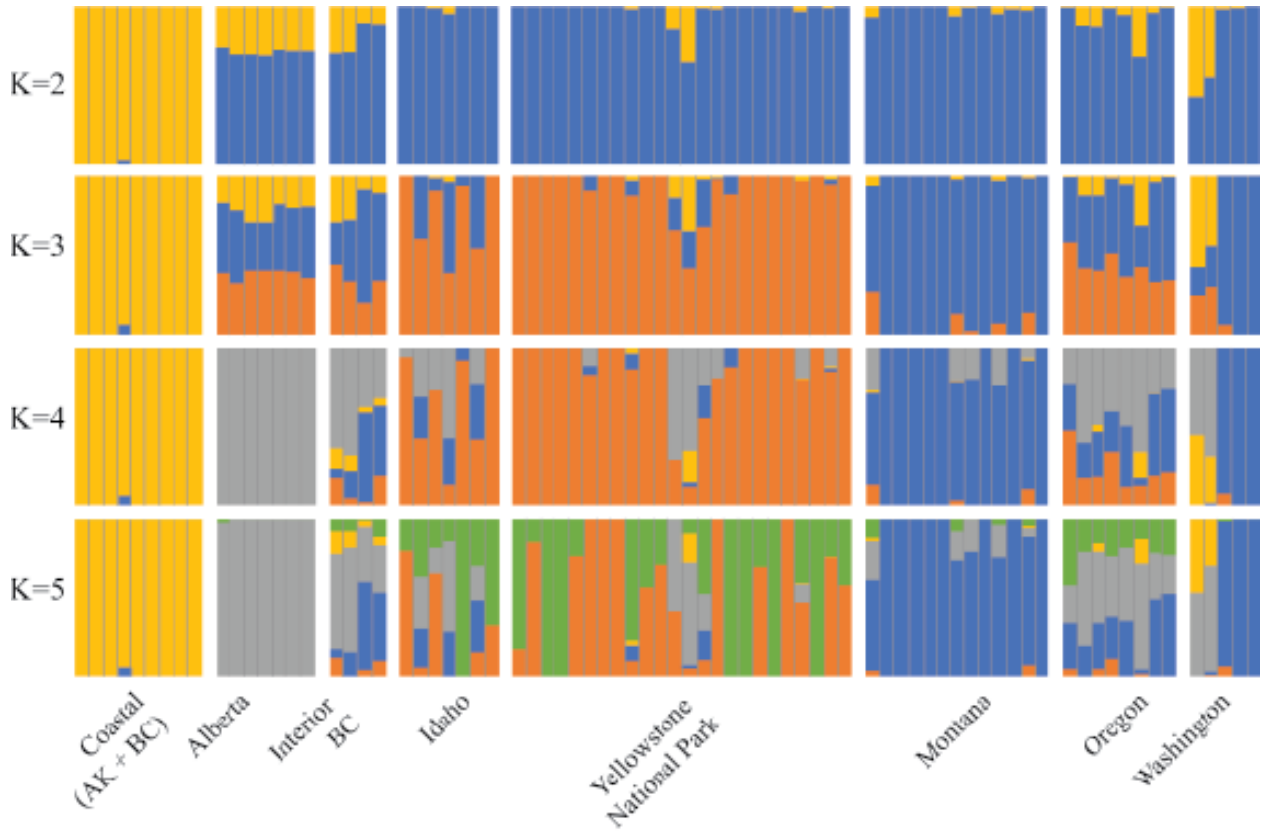

Figure 3.3: Population assignment at K=2 to K=5 for 75 unrelated individuals (after removing founder individuals from YNP and ID), as determined by running ADMIXTURE on a set of 18,508 LD-pruned SNPs within neutral regions. The lowest cross-validation error rate occurred at K = 3, which shows the naturally re-established MT population, the reintroduced YNP and ID population and the coastal population. Higher values of K are also biologically meaningful and therefore shown

et al. 2010). Interestingly, one OR wolf (OR1_OR) has relatedness of 0.13 and 0.17 with two YNP founders (033F and 040F, respectively). These two YNP founders are themselves unrelated and from different packs, but owing to limited sampling of additional related YNP wolves, the exact ancestry of OR1_OR cannot be determined.

Results for the Bayesian estimation for non-symmetrical rates of gene flow found the proportion of individuals per generation originating from within each identified cluster varied from 82.19 to 92.36%, with the highest value found in the MT cluster (Table B.6). Each independent run of BayesAss converged towards similar values of logProb despite different starting seeds. Moreover, visualization of the MCMC trace output confirmed convergence and the posterior probability values of migration suggests strong isolation for all the inferred

clusters. BayesAss estimated that 16.93% of individuals in interior AB and BC migrated from the reintroduced YNP/ID population (Table B.6). MT contributes 6.22%, 6.67% and 5.50% migrants to OR, WA and BC/AB clusters, respectively. OR received 7.13% migrants from the YNP/ID (reintroduced) cluster. WA received 5.02% migrants from the coastal cluster. There is little migration among other clusters (<2% of the population migrating per generation; Supplementary Table S6).

3.4.4 Ecological niche modelling

Aggregate ENMs were produced by averaging values from 100 replicate iterations of the data for the coastal and NRM forest wolf ecotypes niche models (Figures B.7-B.8, respectively). Training and test AUC values for both models were as high as 0.99 (model AUCs ranged from 0.75 to 0.99, depending on partition scheme used, see Figures B.9, B.10), which suggests that the models were highly informative and describe climatically suitable areas that correspond well with the environmental conditions of localities with known ecotypes. The composite model revealed complete optimal environmental niche divergence for each of these ecotypes with very little geographic overlap between ecotypes and concordance in model probability of occurrence regardless of data partitioning scheme (Figures 3.4, B.9, B.10). Suitable habitat analysis would additionally consider prey and human population densities, percentage of forest cover and forest composition, as well as interference from roads and urbanized land cover, which is out of the scope of the current study.

Using centroid pack locations and the aggregate ENMs, the likelihood that a pack occurs in the coastal environment or the interior environment was calculated (referred to as probability of presence throughout; see Elith et al. 2011). Of the 18 WA wolf packs, 17 packs have a greater probability of presence in interior environment than in coastal environment indicating more association of wolves with the interior environment based on our models (Table B.7; Figure 3.4). However, the Teanaway pack, the most western pack currently in WA, has a greater probability of presence in the coastal habitat than the interior habitat (Figure

3.4). No DNA samples were obtained from the Teanaway Pack and we do not currently know the genetic ancestry of this pack. The Lookout pack in WA was on the boundary of interior and coastal habitat and contained a wolf with mtDNA evidence for ancestry to the coastal population and admixed nuclear ancestry of 45% AB and 49% coastal wolf (Sample: RKW4318; Supplementary Table S4). The Wedge pack has a greater probability of presence in the interior habitat (Figure 3.4), yet contained an individual (Sample: WAWedge8) with coastal mtDNA ancestry and admixed nuclear ancestry of 53% AB, 35% coastal and 11% MT (Table B.4). Of the 13 OR wolf packs, all but one, the Rogue pack, have a higher interior probability of presence than coastal habitat. The Rogue pack has a very low (0.0247-0.0476) probability of presence in both habitats with a slightly higher probability of presence in coastal habitat (Figure 3.4). Data from GPS-radio collar tracking devise indicate that this pack was established from a male disperser from the Imnaha pack (NE Oregon) and mated with a female likely from the Snake River or Minam packs (NE Oregon). Unfortunately, we were not able to obtain DNA samples for genetic ancestry analysis of any individuals from the Rogue pack.

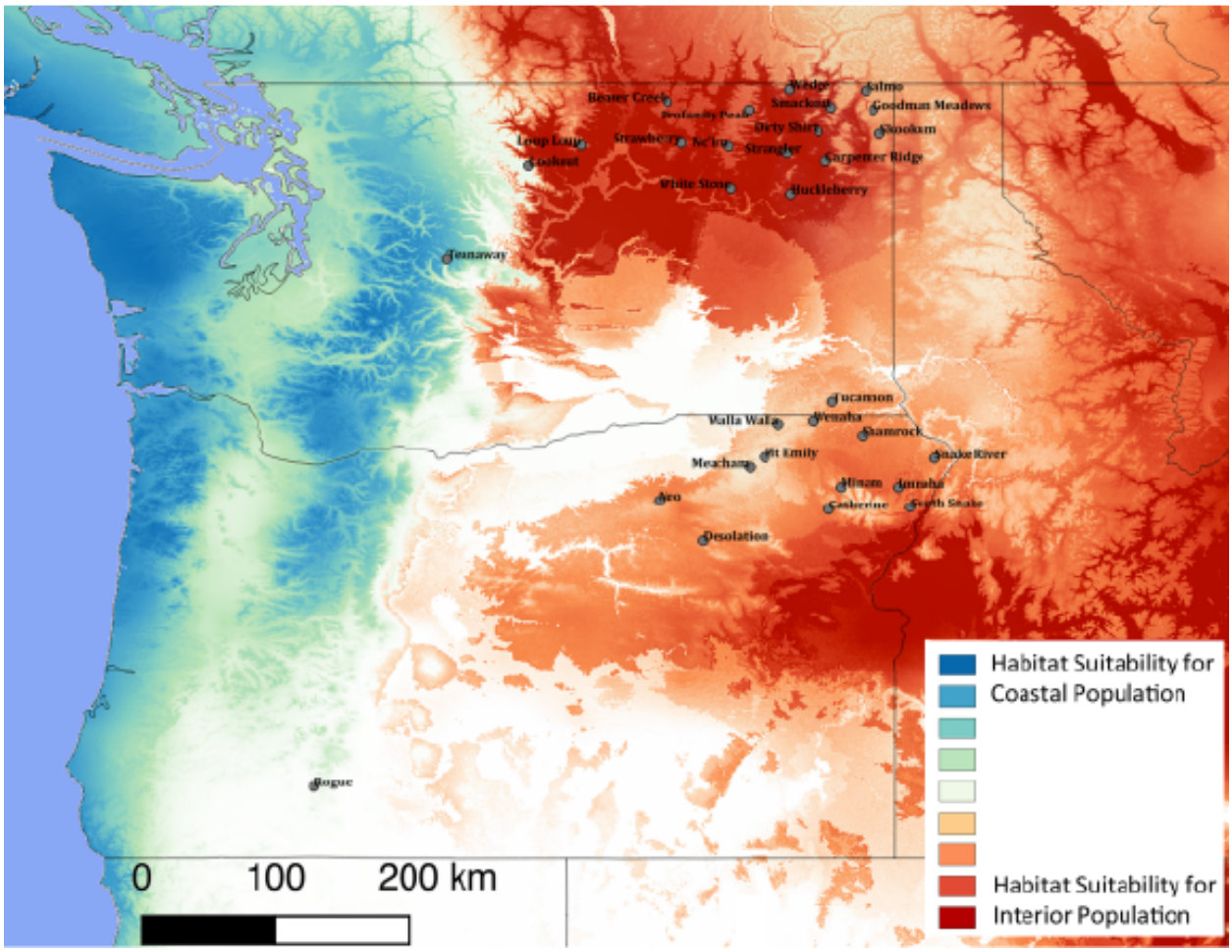

Figure 3.4: Composite MaxEnt distribution model for coastal and interior wolves within the area of the natural re-colonization and potential admixture zone. Warmer colours correspond to the most suitable environment for interior wolves and cooler colours correspond to most suitable environment for coastal wolves. As of 2015, 31 wolf packs inhabited the PNW states of Washington (n = 18) and Oregon (n = 13). Centroid location and pack name of these packs are plotted to show re-colonization of these states but were not used to inform the models. Wolves have been observed in the more coastal areas on the western side of WA but have not established packs as of the end of 2017 (https://wdfw.wa.gov/conservation/gray_wolf/reporting/ sightings.html). Full MaxEnt distribution models for coastal and interior wolves are available in Figures B.7 and B.8

3.5 Discussion

Our results confirm prior work on population structuring of wolves in western North America (Carmichael et al. 2007; vonHoldt et al. 2010, 2011; Schweizer et al. 2016b) and identify the first case of admixture between coastal and NRM wolves in the contiguous US. Wolves from Alaska cluster closely with those from coastal BC (Figures 3.2 and 3.3), which supports previous findings (Weckworth et al. 2005, 2010, 2011; Stronen et al. 2014; Weckworth et al. 2015; Schweizer et al. 2016b; but see Cronin et al. 2014). Our detection of limited differentiation among NRM populations reflects similar findings in vonHoldt et al. (2010). Although to a lesser extent than the coastal/NRM genetic partition, the MT population is distinguishable from the reintroduced populations in ID and YNP and from interior BC and AB. Consequently, the principal genetic partition in the PNW region derives from the coastal and NRM populations.

We assessed the genetic relationships of naturally re-established wolves in WA and OR to potential source populations. Once wolf ancestry was verified using species diagnostic markers, we used evidence from maternal and nuclear markers to identify the source populations' contributions to the current PNW wolf gene pool. Based on our analyses, the founding WA and OR wolves are migrants from a naturally re-established population in MT, from reintroduced populations in ID and YNP, and for the WA wolves only, from the genetically continuous population in coastal BC and southeast AK (Weckworth et al. 2005; Muñoz Fuentes et al. 2009; Weckworth et al. 2011; Schweizer et al. 2016a, b). Wolves from these source populations may have subsequently admixed within the PNW. An alternative scenario is that founding WA wolves were individuals from previous admixture events of coastal BC and NRM wolves (ID, YNP, MT) that migrated into the state. We find that OR individuals are of NRM ancestry only and find evidence for migrants derived from the YNP/ ID cluster in OR. WA individuals have more complex ancestry with some individuals of MT ancestry only and several other individuals with admixed ancestry. These patterns

are evident from population assignments within ADMIXTURE and from the presence of several mitochondrial lineages including the lu68 haplotype (Figure 3.1, Table B.1 and B.2), which is otherwise known only to exist in coastal wolves. The presence of this haplotype indicates that these individuals are direct migrants either from the coastal population or are offspring of a female wolf with coastal ancestry that dispersed into WA.Migration rates from coastal ecotype into the WA population were estimated to be high as 5% as suggested by results from the BayesAss analysis. However, given that the PCA and ADMIXTURE analyses find mixed nuclear ancestry for these individuals with traces of coastal and NRM wolf ancestry (Figures 3.2 and 3.3, and Table B.4), it is unlikely they are direct migrants from the coastal population. Limited sampling and high relatedness among some individuals may have reduced our ability to detect migrants and therefore could have led to an underestimate of gene flow occurring between these adjacent populations. Despite these limitations, this study reports the first cases of admixture between coastal and NRM wolves in the contiguous US and illustrates the complex dynamics of admixed populations of conservation concern.

The PNW likely represents an admixture zone between distinct ecotypes for several reasons. First, niche modelling of NRM and coastal wolf distributions indicates that the PNW is an intermediate landscape with environments suitable for both ecotypes in the states of WA and OR (Figure 3.4). These results confirm previous findings that the coastal wolf may have extended to southwestern OR or northern California, as supported by the presence of haplotype lu68 as far south as southern OR (Hendricks et al. 2015). Further, as proposed by Young and Goldman (1944), the distribution of *C. l. fuscus* (the coastal subspecies) extends into these states. Second, wolf packs might create territories in areas that were deemed less suitable environment by the models for both the coastal and NRM populations. Admixed individuals might be well suited to establish in these areas as evident by the Lookout pack in WA. Third, previous research suggests that admixture of wolf subspecies and/or ecotypes can take place over large geographic areas (Schweizer et al. 2016b). Our analyses support this idea, as individuals with coastal ancestry can occupy interior habitat as well as coastal

habitat. Fourth, there was a previous absence of wolves in the PNW and there are multiple sources of immigrants in nearby areas. Consequently, admixture between ecotypes in the PNW, as opposed to admixture outside of the PNW with subsequent migration into the PNW, is likely given the diversity of habitats present in the region and the presence of ecotypes in adjacent populations that can provide migrants.

3.5.1 Implications for conservation

The dynamic ancestry of PNW in the future will depend in part on wolf management in western states and the trajectory of population growth in coastal populations. For example, if extreme levels of legalized hunting are practiced in the western US, where the population can be reduced to as few as 150 wolves in each of three western source states (MT, ID, WY; Wayne and Hedrick 2011) and the coastal BC population size remains constant through ongoing protection of the Great Bear Rainforest (BC; Thomson 2016), then the PNW population may continuously receive dispersers with coastal ancestry. On the other hand, if coastal wolves (especially those in the high human impact areas of BC's south coast and Alaska's Alexander Archipelago) decline in the future, WA wolves may become a southern refugium that helps safeguard the diversity found in the coastal wolf ecotype.

If genetic influence from the coastal ecotype continues over time, the resulting increase in genetic diversity may allow the population to avoid inbreeding that could lead to the expression of deleterious recessive alleles and cause inbreeding depression as occurred in Scandinavian and Isle Royale wolves (Liberg et al. 2005; Fredrickson et al. 2007; Räikkönen et al. 2009). Although thorough research has yet to be completed, the wolves of the PNW do not show evidence of high levels of inbreeding (here, meaning loss of diversity from a population as measured with the F_{IS} inbreeding coefficient; Table B.3) or presumed inbreeding depression. Several studies have shown that canids are capable of avoiding mating with close relatives and pack members (Smith et al. 1997a; vonHoldt et al. 2008) through several behavioural mechanisms including absolute avoidance of breeding with related pack mem-

bers, male-biased dispersal to packs where they breed with nonrelatives and female-biased subordinate breeding. Immigration from other populations will increase the pool of unrelated individuals that can occupy breeding positions or territories. Further, the possible presence of reproductively successful migrants in WA may have influenced genetic diversity. Therefore, the close demographic and genetic monitoring of the population should continue to assess potential inbreeding and inbreeding depression in the PNW populations. Additionally, future projections of the population at carrying capacity should be conducted to determine whether significant inbreeding depression will occur if connectivity and migratory exchange with other populations were to cease (e.g. vonHoldt et al. 2008).

In addition to human-caused mortality, climate change has the potential to negatively affect wolf dynamics and genetic diversity. Theoretical projections suggest that burn areas in WA may increase dramatically (Littell et al. 2010), likely resulting in temporary displacement of prey and, as a result, wolf packs. Further, shifting and reduced habitat of ungulates due to climate change will likely affect the movement of wolves under these scenarios. Although this habitat change may not affect wolf density, it has been shown that disruptions such as human harvest do affect wolf social structure leading to an increase in adoption of unrelated individuals into packs (Rutledge et al. 2010).

Wolf protection and management has led to top-down effects on ecosystem health and function (Berger et al. 2008; Ripple et al. 2015). For example, in YNP, the reintroduction of wolves enhanced restoration of riparian areas, species biodiversity and community complexity (Ripple et al. 2015). Further, wolves often provide other ecosystem and human services such as regulating prey abundance, creating carrion for other species and increasing ecotourism that benefits local economies (Smith et al. 2003; Licht et al. 2010; Ripple et al. 2015; Hendricks et al. 2017).

3.5.2 Complexities of admixture in conservation

Although wolf-coyote hybridization is not common in western North America, introgression of these two species has been found to occur in the American south and Great Lakes area when wolf densities are low and finding a conspecific mate may be difficult (Wayne and Jenks 1991; Lehman et al. 1991; Roy et al. 1994; Koblmüller et al. 2009; vonHoldt et al. 2011, 2016). Given the presence of coyotes in the PNW, individual dispersing wolves or low-density wolf populations, such as those found in western WA, may provide opportunity for coyote-wolf hybridization (see vonHoldt et al. 2011). Even if the coastal ecotype were to become legally protected, wolf-coyote hybrids would not receive protection status due to human influence causing low wolf density resulting in hybridization. Keeping high wolf density and intact pack structure may guard against this possibility and the possibility of wolf-dog hybridization.

While coastal wolf-coyote hybrids would not qualify for protection, coastal wolf-NRM wolf admixed individuals would qualify for protection according to the decision tree criteria presented by Wayne and Shaffer (2016). First, the admixture has resulted between two native populations resulting from natural patterns of wolf dispersal. Second, these admixed individuals are likely ecological surrogates for the coastal wolves and provide similar community interactions and ecosystem functionality. Third, healthy coastal habitats may enhance the proportion of alleles unique to coastal wolves and decrease the fraction of genomic contribution from the NRM (non-endangered) wolf (Wayne and Shaffer 2016). Given their unique evolutionary heritage and adaptations, packs with a dominant coastal ancestry should be considered a priority for conservation.

By providing additional genetic influx to the PNW population, the coastal BC wolf population may enhance adaptation to coastal habitats and enable persistence of wolf populations along the coastal areas. For example, wolves of the coastal ecotype are smaller and focus on salmon and deer as prey rather than larger prey such as elk in NRM populations. They have

a unique hunting behaviour for this prey base, including selective eating of salmon parts to avoid parasites and swimming as a means of expanding the deer prey base (Darimont and Paquet 2002; Darimont et al. 2003; Paquet et al. 2006). Currently, there are no established packs within the more coastal areas of the PNW (Figure 3.4). Further, allowing for admixture among ecotypes in regions of intermediate habitat may facilitate the process of adaptation and improve the genetic base for selection to act upon (e.g., Hailer and Leonard 2008). As a result, gene flow between coastal BC wolves and NRM populations, such as WA, could potentially help preserve adaptations of the coastal ecotype in an appropriate habitat, enhance the possibility for wolf persistence in coastal habitats of the PNW and enable the evolutionary process of adaptation in intermediate and disturbed habitats. Consequently, we recommend efforts that maintain gene flow and coastal wolf density such as improving and maintaining corridors of immigration and preserving suitable coastal habitat.

Here we provide an example of how managers can use genomic resources to identify ancestry of re-colonized individuals and potential migrants from distinct genetic lineages. Genome-wide analyses are now allowing us to detect signatures of hybridization at a finer scale such as various classes of hybridization such as wolf-dog/wolf-coyote or ecotype-ecotype hybridization, thus advancing our understanding of introgression and divergence. Further, genomic resources (such as the sequence capture methods used here) can be used to inform management decisions as to the most appropriate conservation strategy for a given species (e.g. the distribution of individuals with diagnostic ecotype profiles and their relationship to current and projected habitats). Beyond this study, genomic approaches could be used to identify adaptive potential and further our understanding of preservation of diversity under future climate scenarios (Shafer et al. 2015; Hoffmann et al. 2017).

3.5.3 Policy and management conclusions

Using a multidimensional approach (i.e., combining genomic and ENM analyses to assess admixture during natural re-colonization and the resulting distribution of genetic

variation) may offer conservation biologists a methodological approach to discern ecotype admixture zones. These zones, which are often characterized by environmental gradients, provide selective pressure that can contribute to evolutionary change. While in many cases the evolutionary legacy of isolated populations should be preserved, admixture between once-extirpated taxa that has resulted in distinct adaptations should also be considered for protection. Legal protection and conservation guidelines differ depending on the governing body, but many assessments of endangered species policies have recognized the importance of extending some protection to admixed and hybrid populations (Jackiw et al. 2015; vonHoldt et al. 2017). This study, as well as several others (e.g., Weeks et al. 2016; Love Stowell et al. 2017; Frankham et al. 2017), challenges the historical view that admixture and hybridization threaten biodiversity. As advocated by vonHoldt et al. (2017) and Wayne and Shaffer (2016), case-by-case protection should be considered when colonization is a natural process within the integrated WOL framework and when admixed individuals represent effective ecological surrogates that might eventually restore endangered entities to their historical distribution.

3.6 Summary

Here we assess admixture during natural re-colonization and the resulting distribution of genetic variation based on mitochondrial haplotypes and 18,508 neutral nuclear SNPs. We utilize niche modelling to define ecotype boundaries and find little correspondence with genetic partitions that may reflect recent colonization from multiple sources. The PNW population is admixed, with coastal influences apparent in WA wolves. This admixture is desirable to enhance adaptation to coastal environments and, in general, enable the evolutionary process for adaptation. Admixed individuals may receive special protection if conditions are such that the historical genetic composition of coastal wolves might be restored and if the hybrids are ecological surrogates providing similar ecosystem functionality and community interactions as the endangered taxon (in this case Alexander Archipelago wolves; see arguments in Wayne and Shaffer (2016)). Determining ecological surrogates may

be possible through inferred patterns of selection across the genome, observational studies and/or reciprocal transplant experiments. Further research is needed to establish accurate migration rates and model the potential effects of changing predator/prey dynamics and climate on wolf populations. However, efforts to enhance the density and distribution of coastal wolves in the PNW should be considered as a hedge against population decline in coastal Alaskan or south coastal BC wolves. This effort will aid in the preservation of adaptations for the coastal environment and decrease the likelihood of hybridization with coyotes. To preserve this southern genetic refugium for coastal BC wolves, restore ecological processes and permit contemporary evolution, natural expansion and protection of the coastal wolves in the contiguous US should be an emphasis of wolf management in the PNW.

3.7 Data availability

Sequence reads and mapping files are archived at the NCBI SRA under SRP145376. The filtered variant call file for all individuals as well as a bed file of the neutral regions are available through Dryad Digital Repository under accession number doi:10.5061/dryad.np7t1p2.

3.8 Acknowledgements

We are grateful to Kristin Mansfield and Roblyn Brown for samples and insight into Washington and Oregon wolves, respectively. Laboratory and data processing assistance was completed by Sarah Wenner, Daniel Greenfield and Jacqueline Robinson. We thank Brendan Epstein for helpful discussions concerning methods implementation and Kim Andrews for discussions regarding data interpretation within this study. This work was supported by the Wilburforce Foundation and the National Science Foundation (NSF; DEB-1021397, OPP-0733033). Additional support was provided to SAH by the National Institute of Health (P30GM103324), NSF (DEB-1316549) and the Bioinformatics and Computational Biology program at the University of Idaho; to RMS by the NSF (DGE-1144087, DGE-

0707424, 1612859); and to RJH by the NSF (PD-08-1269). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH (S10RR029668, S10RR027303).

3.9 References

Abbott RJ, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. (2013). Hybridization and speciation. J Evol Biol 26: 229-246.

Abbott RJ, Barton NH, Good JM (2016). Genomics of hybridization and its evolutionary consequences (RJ Abbott, NH Barton, and JM Good, Eds.). Molecular Ecology 25: 2325-2332.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19: 1655-1664.

Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics. Nature Reviews Genetics 11: 697-709.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001). The problems with hybrids: setting conservation guidelines. Trends in Ecology & Evolution 16: 613-622.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Arnold ML (2016). Divergence with Genetic Exchange. Oxford University Press.

Bailey V (1936). The Mammals and Life Zones of Oregon. North American Fauna: 272-275.

Berger KM, Gese EM, Berger J (2008). Indirect effects and traditional trophic cascades: A test involving wolves, coyotes, and pronghorn. Ecology 89: 818-828.

Blair ME, Sterling EJ, Dusch M, Raxworthy CJ, Pearson RG (2013). Ecological divergence and speciation between lemur (Eulemur) sister species in Madagascar. J Evol Biol 26: 1790-1801.

Boyd DK, Paquet PC, Donelon S, Ream RR, Pletscher DH, White CC (1995). Ecology and Conservation of Wolves in a Changing World. Canadian Circumpolar Institute.

Callaghan C (2002). The ecology of gray wolf (Canis lupus) habitat use, survival, and persistence in the Central Rocky Mountains, Canada.

Carmichael LE, Krizan J, Nagy JA, Fuglei E, Dumond M, Johnson D, et al. (2007). Historical and ecological determinants of genetic structure in arctic canids. Molecular Ecology 16: 3466-3483.

Chambers SM, Fain SR, Fazio B, Amaral M (2012). An Account of the Taxonomy of North American Wolves From Morphological and Genetic Analyses. North American Fauna 77: 1-67.

Chapron G, Kaczensky P, Linnell JDC, Arx von M, Huber D, Andrén H, et al. (2014). Recovery of large carnivores in Europe's modern human-dominated landscapes. Science 346: 1517-1519.

Cronin MA, Canovas A, Bannasch DL, Oberbauer AM, Medrano JF (2014). Single Nucleotide Polymorphism (SNP) Variation of Wolves (Canis lupus) in Southeast Alaska and Comparison with Wolves, Dogs, and Coyotes in North America. Journal of Heredity 106: 26-36.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. (2011). The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

Darimont CT, Paquet PC (2002). Gray Wolves, of British Columbia's Central and North Coast: Distribution and Conservation Assessment. Canadian Field Naturalist.

Darimont CT, Reimchen TE, Paquet PC (2003). Foraging behaviour by gray wolves on salmon streams in coastal British Columbia. Canadian Journal of Zoology 81: 349-353.

Darimont CT, Paquet PC, Reimchen TE (2008). Spawning salmon disrupt tight trophic coupling between wolves and ungulate prey in coastal British Columbia. BMC Ecology 8: 14.

DeLong ER, DeLong DM, Clarke-Pearson DL (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 44: 837.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491-498.

Diniz-Filho JAF, Mauricio Bini L, Fernando Rangel T, Loyola RD, Hof C, Nogués-Bravo D, et al. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography 32: 897-906.

Dobzhansky T (1935). A critique of the species concept in biology. Philosophy of Science 2: 344-355.

Elith J, H Graham C, P Anderson R, Dudík M, Ferrier S, Guisan A, et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43-57.

Faircloth BC, Glenn TC (2012). Not All Sequence Tags Are Created Equal: Designing and Validating Sequence Identification Tags Robust to Indels. PLoS ONE 7: e42543.

Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013). A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs) (CS Moreau, Ed.). PLoS ONE 8: e65923.

Fourcade Y, Engler JO, Rödder D, Secondi J (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLoS ONE 9: e97122.

Frankham R, Ballou JD, Ralls K, Dubash MR, Fenster CB, Sunnucks P (2017). Genetic Management of Fragmented Animal and Plant Populations. Oxford University Press.

Fredrickson RJ, Siminski P, Woolf M, Hedrick PW (2007). Genetic rescue and inbreeding depression in Mexican wolves. Proceedings of the Royal Society B: Biological Sciences 274: 2365-2371.

Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, et al. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs (L Andersson, Ed.). PLoS Genet 10: e1004016-12.

Fritts SH (1983). Record dispersal by a wolf from Minnesota. Journal of Mammalogy 64: 166-167.

Fritts SH, Bangs EE, Fontaine JA, Brewster WG (1995). Restoring wolves to the northern Rocky Mountains of the United States. In: Ecology and conservation of wolves in a changing world: Edmonton, Alberta, pp 107-125.

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, et al. (2009). Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. Genetics 181: 1493-1505.

Haight RG, Mladenoff DJ, Wydeven AP (1998). Modeling Disjunct Gray Wolf Populations in Semi-Wild Landscapes. Conserv Biol 12: 879-888.

Hailer F, Leonard JA (2008). Hybridization among Three Native North American Canis Species in a Region of Natural Sympatry (H Harpending, Ed.). PLoS ONE 3: e3333.

Harrigan RJ, Thomassen HA, Buermann W, Smith TB (2014). A continental risk assessment of West Nile virus under climate change. Glob Change Biol 20: 2417-2425.

Harrigan RJ, Thomassen HA, Buermann W, Cummings RF, Kahn ME, Smith TB (2010). Economic conditions predict prevalence of West Nile virus. PLoS ONE 5: e15437.

Hedrick PW (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22: 4606-4618.

Hendricks SA, Charruau PC, Pollinger JP, Callas R, Figura PJ, Wayne RK (2015). Polyphyletic ancestry of historic gray wolves inhabiting U.S. Pacific states. Conservation Genetics 16: 759-764.

Hendricks S, Epstein B, Schonfeld B, Wiench C, Hamede R, Jones ME, et al. (2017). Conservation implications of limited genetic diversity and population structure in Tasmanian devils (Sarcophilus harrisii). Conservation Genetics 18: 977-982.

Hernandez PA, Graham CH, Master LL, Albert DL (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773-785.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour 11: 117-122.

Jackiw RN, Mandil G, Hager HA (2015). A framework to guide the conservation of species hybrids based on ethical and ecological considerations. Conserv Biol 29: 1040-1051.

Jewell DM, Ashe M, Haskett M (2015). Re: Petition to List on an Emergency Basis the Alexander Archipelago Wolf (Canis Lupus Ligoni) as Threatened or Endangered Under the Endangered Species Act.

Jimenez MD, Bangs EE, Boyd DK, Smith DW, Becker SA, Ausband DE, et al. (2017). Wolf dispersal in the Rocky Mountains, Western United States: 1993-2008. Jour Wild Mgmt 81: 581-592.

Jordan MI, Ng AY (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems.

Koblmüller S, Nord M, Wayne RK, Leonard JA (2009). Origin and status of the Great Lakes wolf. Molecular Ecology 18: 2313-2326.

Kurtz S, Narechania A, Stein JC, Ware D (2008). A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics 9: 517.

Lehman N, Eisenhawer A, Hansen K, Mech LD, Peterson RO, Gogan PJP, et al. (1991). Introgression of Coyote Mitochondrial DNA Into Sympatric North American Gray Wolf Populations. Evolution 45: 104.

Leonard JA, Vilá C, Wayne RK (2005). Legacy lost: genetic variability and population size of extirpated US grey wolves (Canis lupus). Molecular Ecology 14: 9-17.

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilá C (2002). Ancient DNA evidence for Old World origin of New World dogs. Science 298: 1613-1616.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

Liberg O, Andrén H, Pedersen H-C, Sand H, Sejberg D, Wabakken P, et al. (2005). Severe inbreeding depression in a wild wolf (Canis lupus) population. Biology Letters 1: 17-20.

Licht DS, Millspaugh JJ, Kunkel KE, Kochanny CO, Peterson RO (2010). Using Small Populations of Wolves for Ecosystem Restoration and Stewardship. BioScience 60: 147-153.

Littell JS, Oneil EE, McKenzie D, Hicke JA, Lutz JA (2010). Forest ecosystems, disturbance, and climatic change in Washington State, USA. Climatic change.

Lv W, Li Z, Wu X, Ni W, Qv W (2011). Maximum Entropy Niche-Based Modeling (Maxent) of Potential Geographical Distributions of Lobesia Botrana (Lepidoptera: Tortricidae) in China. In: Computer and Computing Technologies in Agriculture V, IFIP Advances in Information and Communication Technology. Springer, Berlin, Heidelberg: Berlin, Heidelberg Vol 370, pp 239-246.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010). Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867-2873.

Mayr E (1947). Systematics and the origin of species. Columbia University Press.

Mech LD (1970). The Wolf. American Museum of Natural History by the Natural History Press.

Merrill SB, Mech LD (2000). Details of Extensive Movements by Minnesota Wolves (Canis lupus). USGS Northern Prairie Wildlife Research Center Paper 76.

Mladenoff DJ, Sickley TA (1998). Assessing Potential Gray Wolf Restoration in the Northeastern United States: A Spatial Prediction of Favorable Habitat and Potential Population Levels. Jour Wild Mgmt 62: 1.

Mladenoff DJ, Haight RG, Sickley TA, Wydeven AP (1997). Causes and Implications of Species Restoration in Altered Ecosystems. BioScience 47: 21-31.

Mladenoff DJ, Sickley TA, Wydeven AP (1999). Predicting Gray wolf landscape recolonization: logistic regression models vs. new field data. Ecological Applications 9: 37-44.

Mladenoff DJ, Sickley TA, Haight RG, Wydeven AP (1995). A Regional Landscape Analysis and Prediction of Favorable Gray Wolf Habitat in the Northern Great Lakes Region. Conserv Biol 9: 279-294.

Muhlfeld CC, Kalinowski ST, McMahon TE, Taper ML, Painter S, Leary RF, et al. (2009). Hybridization rapidly reduces fitness of a native trout in the wild. Biology Letters 5: 328-331.

Muñoz Fuentes V, Darimont CT, Paquet PC, Leonard JA (2010). The genetic legacy of extirpation and re-colonization in Vancouver Island wolves. Conservation Genetics 11: 547-556.

Muñoz Fuentes V, Darimont CT, Wayne RK, Paquet PC, Leonard JA (2009). Ecological factors drive differentiation in wolves from British Columbia. J Biogeogr 36: 1516-1531.

Muscarella R, Galante PJ, Soley Guardia M, Boria RA, Kass JM, Uriarte M, et al. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models (J McPherson, Ed.). Methods Ecol Evol 5: 1198-1205.

Paquet PC, Alexander SM, Swan PL, Darimont CT (2006). The Influence of Natural Landscape Fragmentation and Resource Availability on Connectivity and Distribution of Marine Gray Wolf (Canis lupus) Populations on Central Coast, BC. In:Crooks K, Sanjayan MA (eds) Connectivity Conservation. p 726.

Paquet PC, Wierczhowski J, Callaghan C (1996). Summary report on the effects of human activity on gray wolves in the Bow River Valley, Banff National Park, Alberta.

Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81: 559-575.

Räikkönen J, Vucetich JA, Peterson RO, Nelson MP (2009). Congenital bone deformities and the inbred wolves (Canis lupus) of Isle Royale. Biological Conservation 142: 1025-1031.

Rhymer J, Simberloff D (1996). Extinction by hybridization and introgression. Annu Rev Ecol Syst 27: 83-109.

Ripple WJ, Beschta RL, Painter LE (2015). Trophic cascades from wolves to alders in Yellowstone. Forest Ecology and Management 354: 254-260.

Rodríguez Soto C, Monroy Vilchis O, Maiorano L, Boitani L, Faller JC, Briones MA, et al. (2011). Predicting potential distribution of the jaguar (Panthera onca) in Mexico:

identification of priority areas for conservation. Diversity and Distributions 17: 350-361.

Roy MS, Geffen E, Smith D, Ostrander EA, Wayne RK (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. Molecular Biology and Evolution 11: 553-570.

Rödder D, Schmidtlein S, Veith M, Lötters S (2009). Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? (R Thompson, Ed.). PLoS ONE 4: e7843.

Rutledge LY, Patterson BR, Mills KJ, Loveless KM, Murray DL, White BN (2010). Protection from harvesting restores the natural social structure of eastern wolf packs. Biological Conservation 143: 332-339.

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, Thalmann O, Smith D, Randi E, Ostrander EA, Lohmueller K, Green RE, Novembre J, Wayne RK. (2018). Natural selection and origin of a melanistic allele in North American gray wolves. Molecular Biology and Evolution. https://doi.org/10.1093/molbev/msy031

Schweizer RM, Robinson JA, Harrigan RJ, Silva P, Galverni M, Musiani M, et al. (2016a). Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Molecular Ecology 25: 357-379.

Schweizer RM, vonHoldt BM, Harrigan RJ, Knowles JC, Musiani M, Coltman D, et al. (2016b). Genetic subdivision and candidate genes under selection in North American grey wolves. Molecular Ecology 25: 380-402.

Sesink Clee PR, Abwe EE, Ambahe RD, Anthony NM, Fotso R, Locatelli S, et al. (2015). Chimpanzee population structure in Cameroon and Nigeria is associated with habitat variation that may be lost under climate change. BMC Evol Biol 15: 2.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research 15: 1034-1050.

Slatkin M (1987). Gene Flow and the Geographic Structure of Natural Populations. Science 236: 787-792.

Smith DW, Peterson RO, Houston D (2003). Yellowstone after Wolves. BioScience 53: 330-340.

Smith D, Meier T, Geffen E, Mech LD, Burch JW, Adams LG, et al. (1997). Is incest common in gray wolf packs? Behavioral Ecology 8: 384-391.

Smith TB, Wayne RK, Girman DJ, Bruford MW (1997). A Role for Ecotones in Generating Rainforest Biodiversity. Science 276: 1855-1857.

Staples J, Nickerson DA, Below JE (2013). Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. Genet Epidemiol 37: 136-141.

Stowell SML, Pinzone CA, Martin AP (2017). Overcoming barriers to active interventions for genetic diversity. Biodivers Conserv 26: 1753-1765.

Stronen AV, Navid EL, Quinn MS, Paquet PC, Bryan HM, Darimont CT (2014). Population genetic structure of gray wolves (Canis lupus) in a marine archipelago suggests island-mainland differentiation consistent with dietary niche. BMC Ecology 14: 11.

Thomson HS (2016). Bill 2 - 2016- Great Bear Rainforest (Forest Management) Act.

Twyford AD, Ennos RA (2012). Next-generation hybridization and introgression. Heredity 108: 179-189.

Verts BJ, Carraway LN (1998). Land Mammals of Oregon. Univ of California Press.

Vilá C, Amorim IR, Leonard JA, Posada D, Castroviejo J, Petrucci Fonseca F, et al. (1999). Mitochondrial DNA phylogeography and population history of the grey wolf (Canis lupus). Molecular Ecology 8: 2089-2103.

vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY (2017). Redefining the Role of Admixture and Genomics in Species Conservation. Conservation Letters 16: 613.

vonHoldt, B. M., Cahill, J. A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J. P., et al. (2016). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. Science Advances, 2:e1501714-e1501714.

vonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, et al. (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Research 21: 1294-1305.

vonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK (2013). Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. Mamm Genome 24: 80-88.

vonHoldt BM, Stahler DR, Bangs EE, Smith DW, Jimenez MD, Mack CM, et al. (2010). A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. Molecular Ecology 19: 4412-4427.

vonHoldt BM, Stahler DR, Smith DW, Earl DA, Pollinger JP, Wayne RK (2008). The genealogy and genetic viability of reintroduced Yellowstone grey wolves. Molecular Ecology 17: 252-274.

Wayne RK, Hedrick PW (2011). Genetics and wolf conservation in the American West: lessons and challenges. Heredity 107: 16-19.

Wayne RK, Jenks SM (1991). Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf (Canis rufus). Nature 351: 565-568.

Wayne RK, Shaffer HB (2016). Hybridization and endangered species protection in the molecular era (RJ Abbott, NH Barton, and JM Good, Eds.). Molecular Ecology 25: 2680-2689.

Weckworth BV, Dawson NG, Talbot SL, Cook JA (2015). Genetic Distinctiveness of Alexander Archipelago Wolves (Canis lupus ligoni). Journal of Heredity 106: 412-414.

Weckworth BV, Dawson NG, Talbot SL, Flamme MJ, Cook JA (2011). Going Coastal: Shared Evolutionary History between Coastal British Columbia and Southeast Alaska Wolves (Canis lupus) (RC Fleischer, Ed.). PLoS ONE 6: e19582-8.

Weckworth BV, Talbot SL, Cook JA (2010). Phylogeography of wolves (Canis lupus) in the Pacific Northwest. Journal of Mammalogy 91: 363-375.

Weckworth BV, Talbot S, Sage GK, Person DK, Cook J (2005). A Signal for Independent Coastal and Continental histories among North American wolves. Molecular Ecology 14: 917-931.

Weeks AR, Stoklosa J, Hoffmann AA (2016). Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. Frontiers in Zoology 2016 13:1-13:31.

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

Wilson GA, Rannala B (2003). Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. Genetics 163: 1177-1191.

Wisz MS, Hijmans RJ, Li J, Peterson AT (2008). Effects of sample size on the performance of species distribution models. Diversity and Distributions 14: 763-773.

Wright S (1951) Genetical structure of populations. Annual Eugenics 166: 247-249.

Young S, Goldman EA (1944). The Wolves of North America. 2 vols. Washington: American Wildlife Institute.

Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, et al. (2014). Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau (JM Akey, Ed.). PLoS Genet 10: e1004466-13.

CHAPTER 4

Genomics of wildlife cancer, with transmissible cancer in Tasmanian devils as a case study

4.1 Abstract

Studies of cancer in wildlife species present unique challenges, but research is beginning to uncover examples of cancer and its impact on wildlife populations. Causes of cancer in wildlife include environmental carcinogens, viruses and other pathogens, hereditary factors, and direct transmission of tumor cells. Population genomic tools are increasingly being used to investigate wildlife cancers, addressing issues such as the genetic variation for susceptibility within populations, comparative genomics of tumor suppressor genes, and evolutionary response to cancers. Here we review progress and potential for population genomics to address these issues. As an illustrative case study, we focus on the unique case of a transmissible cancer, devil facial tumor disease (DFTD), which has had a dramatic impact on Tasmanian devils (*Sarcophilus harrisii*). Total species abundance has declined by about 80 % since DFTD appeared in the mid 1990s, and the disease has imposed dramatic shifts in life history and demography. Recent genomic research has revealed genetic variation for DFTD-related phenotypes and signatures of rapid evolution at candidate loci associated with cancer and immune function. The DFTD system illustrates how genomic tools can be applied to an epidemic cancer in a wildlife population, providing insights into basic cancer biology as well as implications for conservation strategies.

4.2 Introduction: Cancer in wildlife

Over the past few decades wildlife health monitoring has increased, and we are now gaining an improved, and occasionally alarming, perspective about the presence and impact of cancer in wildlife, particularly some endangered species (Pesavento et al 2018). Cancer, a collection

of diseases characterized by abnormal cell growth with the potential to metastasize, affects nearly every known multicellular organism. Once viewed as an accumulation of genomic aberrations, recent studies have emphasized that cancers are heterogeneous collections of cells (Campbell et al. 2008; Maley et al. 2006; Merlo and Maley 2010; Park et al. 2010) that evolve in tumor microenvironments with complex ecologies (Bissell and Radisky 2001). Cancer can affect wild populations by reducing reproductive success, altering population dynamics, or directly or indirectly leading to population declines. Anthropogenic influences, including direct impacts such as pollution and the reduction of genetic diversity in natural populations, could increase the prevalence of cancer in wildlife (Giraudeau et al. 2018; Pesavento et al. 2018). In addition to presenting a major conservation concern, cancers in wildlife species may provide new biological models for understanding the often complex causes of cancer, with the potential for biomedical benefits.

Studying cancer in wild systems is especially challenging owing to their inaccessibility, and ethical, logistical and legal limits on experimentation. Further, cancer in wildlife goes largely undetected and determining cancer prevalence (or prevalence of any disease in wildlife) is not often achieved. Relatively few studies (n=31 vertebrate species) have recorded cancer prevalence in wild populations (Madsen et al. 2017). From the limited available data, estimates of cancer prevalence in mammal populations range from 2% (sea otter (*Enhydra lutris*); Williams and Pulley 1981) to 64% (Baltic gray seal (*Halichoerus grypus*); Bäcklin et al. 2016). Low-prevalence cancers are likely to go undetected in many wildlife populations, and could be much more pervasive than observed. In addition, estimates of prevalence in natural populations may be down-biased due to several factors. Substantially less is known about wildlife tumors than their human counterparts and diagnostic resources are often limited in these cases. Mortality is likely to be increased by the cancer itself, by secondary parasite or pathogen infections, as well as by an increased level of predation as these factors decrease individuals' ability to avoid predation (Vittecoq et al. 2013). Despite these difficulties, researchers have determined general factors leading to the development of

tumors.

4.3 Causes

4.3.1 Environmental carcinogens

Environmental factors are often associated with human and wildlife cancer incidence. UV radiation exposure, smoking, and ingestion of certain foods influence cancer in humans (Irigaray et al. 2007; Soto and Sonnenschein 2010). Moreover, there is some evidence that reproductive biology (Aktipis and Nesse 2013) as well as stresses and trauma (Antoni et al. 2006; Reiche et al. 2004) influence cancer risk. While some of these factors are not relevant to wildlife cases, some such as stress may play an important role. These factors are often difficult to measure, but studies have shown that environmental pollution is often associated with elevated risk factors for wild populations. Benthic fish communities and fauna within the Chernobyl area, found with both higher than average levels pollution and radioactive contamination, are affected by tumors and cancer (Brown et al. 1973; Mousseau and Møller 2015; Yablokov 2009). Similarly, the beluga whale (*Delphinapterus leucas*) population in the St. Laurence River estuary (Canada) has a higher rate of cancer than other populations (Black and Baumann 1991; Martineau et al. 2002). This population has also been found to be heavily contaminated by agricultural and industrial chemicals such as polycyclic aromatic hydrocarbons, polychlorinated biphenyls, dichlorodiphenyltrichloroethane, and their metabolites (Letcher et al. 2000; Martineau et al. 1987; Martineau et al. 1988; Wagemann et al. 1990). The finding of the same signature in tumors of these wildlife taxa and humans would strongly support the etiologic role of contaminants in carcinogenesis (Perera 1998; Perera and Dickey 1997).

4.3.2 Viruses

Infectious cancers fall within two categories: infection of a host by an oncogenic pathogen or by a cancer cell that is acquired from a different host. Parasite-induced cancers, cancers caused by subcellular, unicellular, or multicellular parasites, are mainly attributed to viruses as compared to other pathogens (McAloose and Newton 2009; McCallum and Jones 2012; Pesavento et al. 2018; Ujvari et al. 2017). Evaluating the effects of pathogens on cancer development is challenging partially because of potential delays between the presence of the parasite and cancer detection (Ewald and Swain Ewald 2015). Most studies about virus-induced cancer use human or domestic animal models and have shown that viruses disrupt a variety of cellular barriers to oncogenesis (Ewald and Swain Ewald 2012). Infected cells may lose the ability to control the total number of cellular divisions, apoptosis, adhesive properties to other cells, and/or cellular arrest (Ewald and Swain Ewald 2013; Ewald and Swain Ewald 2012). Otarine herpesvirus-1 is one virus highly associated with and the putative cause of genital carcinoma in mature California sea lions (Buckles et al. 2006; King et al. 2002; Lipscomb et al. 2016). Attwater's prairie chickens, western barred bandicoots, and sea turtles are other examples of systems with virus-associated cancers (Ewald and Ewald 2017).

4.3.3 Transmissible Cancers

Transmissible cancers are rarer than virus-associated cancers. Transmissible tumors are, by definition, clonal in origin and spread directly by transfer of cells between individuals. This suggests a singular event from which all tumors evolved. Transmission occurs with direct contact during mating, biting, or feeding, or tumor cells may be acquired through the environment in marine systems (Metzger and Goff 2016; Ostrander et al. 2016). The most-well studied transmissible tumor is canine transmissible venereal tumor (CTVT), which affects dogs (*Canis lupus domesticus*) and is believed to have originated thousands of year ago, making it perhaps the 'oldest continuously propagated cell lineage' (Murchison et al.

2014; Murgia et al. 2006). Devil facial tumor disease (DFTD) and DFT2 are two recent independent origins of transmissible cancer that infect Tasmanian devils (*Sarcophilus harrisii*; Pearse and Swift 2006; Pye et al 2016); we discuss this case in detail below. Another transmissible cancer produces leukemia-like neoplasias called disseminated neoplasia or hemic neoplasia, reported in at least 15 different bivalve species (Barber 2004; Carballal et al. 2015). Analysis of the neoplastic cells in soft-shell clams revealed a dramatic amplification in the copy number of a retrotransposon (Arriagada et al. 2014), with identical integration sites in neoplastic cells from multiple animals. These data, along with analysis of microsatellites and mitochondrial DNA SNPs, showed that the etiologic agent of this disease is the neoplastic cell itself, as with CTVT and DFTD (Metzger et al. 2015).

4.3.4 Hereditary factors

Over 200 hereditary cancer susceptibility syndromes have been described in humans, the majority of which are inherited in an autosomal dominant manner (Nagy et al. 2004). The following characteristics designate an inherited cancer susceptibility: "two or more relatives with the same type of cancer on the same side of the family; several generations affected; earlier ages of cancer diagnosis than what is typically seen for that cancer type; individuals with multiple primary cancers; the occurrence of cancers in one family, which are known to be genetically related; and the occurrence of nonmalignant conditions and cancer in the same person and/or family" (Nagy et al. 2004). Many of these are rare syndromes, but collectively hereditary cancers amount to a substantial burden of morbidity and mortality in the human population as they are estimated to account for at least 1-10% of all cancers (Fearon 1997; Nagy et al. 2004). The information gained through the discovery and characterization of genes and the cellular signaling pathways involved in some hereditary cancers have begun to provide insights into the pathogenesis of both inherited and sporadic forms of cancer (Fearon 1997). This research has led to the development of targeted molecular-based interventions and has changed the way these families are counseled and their health concerns are managed.

Because we typically research wildlife systems on a population rather than an individual level, we know little about hereditary cancer in wildlife.

4.4 Genetics and evolution of cancer in wildlife

4.4.1 Evolution of cancer resistance

Early-onset cancers tend to be more attributable to a specific cause than late-onset cancers, and an evolutionary perspective can help explain this difference (Frank 2004). Resolving this issue is important for understanding the occurrence of cancers in the young, for predicting the incidence of cancers in our increasingly aging population, for developing efficient strategies for detecting the genes that inhibit our cancers and, ultimately, for preventing cancer. Natural selection causes cancer to be relatively rare; thus, genomes are derived disproportionally from individuals with effective mechanisms for suppressing cancer. In the context of a multicellular animal, conflict arises because a single cell can derive short-term success from its clonal proliferation within the individual, a success guaranteed given the abundance of resources available to a selfish cell surrounded by altruistic brethren. The result is cancer, a cellular strategy that is successful in the short term even though it ultimately dooms any chance of genetic transmission to future generations by killing the parent organism. For multi-cellularity to be successful, such antisocial acts had to be inhibited by suppression and/or policing (Michod 2000; Szathmáry and Smith 1995). Selection against genetic variants that cause early-onset cancer is expected to be stronger than selection against later-acting variants, because of the costs on reproductive fitness (Leroi et al. 2003).

While organisms have evolved mechanisms to prevent cancer, these defenses may have high costs for wound healing, growth, reproduction, and aging. For example, a key tradeoff for any organism is between limiting uncontrolled cell division while maintaining the capacity to repair tissues. Fast and effective wound healing requires cell movement and

proliferation (Guo and DiPietro 2010), capacities that leave an organism more vulnerable to cancer (Hofman and Vouret-Craviari 2012). Fast growth may come at the cost of somatic maintenance, leading to cancer vulnerability (De Stavola et al. 2004). Likewise, there may be trade-offs between somatic maintenance and reproductive effort (i.e., acquiring mates, and making and caring for offspring) (Stearns 2000). For example, competitiveness in males may lead to higher susceptibility to prostate cancer (Alvarado 2013). Early menarche comes at the cost of higher susceptibility to breast cancer in females (Hsieh et al. 1990). Women with BRCA mutations have greater susceptibility to breast cancer, but also higher fertility (Smith et al. 2012). Additionally, there are constraints on the immune system's ability to detect cancer cells (Mapara and Sykes 2016) because cancer cells are derived from normal cells. The capacity for inflammation is crucial not only for defending against infection but also for dealing with rogue cells. However, inflammation also damages tissues and makes them more vulnerable to cancer (Coussens and Werb 2002; de Visser et al. 2006).

4.4.2 Genetics of population susceptibility

Certain wildlife populations may have increased susceptibility to particular types of cancer. Population susceptibility may be associated with the increase of rare germline variants or homozygosity after genetic bottlenecks, selective breeding, or founder effects. Accordingly, the emergence and persistence of some cancers in animals is associated with low genetic diversity (Vickers et al. 2015). When populations are small, both inbreeding and stochastic processes, such as genetic drift, can increase the probability of extinction by increasing homozygosity at loci with deleterious alleles and reducing fitness (Frankham 2005a; Frankham 2005b; Gomulkiewicz and Holt 1995; O'Grady et al. 2006). As a population loses variation via drift and inbreeding, genetic load increases. The genetic load is defined as the difference in the mean fitness of a population and the fitness of an optimal genotype that does not carry deleterious mutations (Glémin et al. 2003; Kirkpatrick and Jarne 2015). Small effective population sizes due to bottlenecks, founder effects, and domestication have been

found to increase the genetic load in Mountain gorillas (*Gorilla beringei beringei*), Channel Island foxes (*Urocyon littoralis*), and domestic dogs (*Canis lupus familiaris*) (Marsden et al. 2015; Robinson et al. 2016; Xue et al. 2015). Genetic load can be characterized by numerous weakly deleterious mutations, which can be particularly difficult to eliminate from bottlenecked populations (Fu et al. 2014; Henn et al. 2015). It is unknown the extent to which cancer plays a role in the reduced fitness of genetically depauperate populations across wildlife taxa.

One example of population susceptibility to cancer is found in Channel Island foxes, which are endemic to individual islands off the coast of California (Figure 4.1a). These populations have undergone severe genetic bottlenecks resulting in the accumulation of deleterious mutations (Robinson et al. 2016) and adapted to different environmental conditions (Funk et al 2016), thereby creating a set of replicate inbred populations. These populations differ markedly in the incidence of cancer: on one island (Santa Catalina Island), foxes have a high prevalence of ear canal (ceruminous gland) carcinoma and adenoma (collectively tumors) that appear to be associated with inflammation from chronic ear mite (*Otodectes*) infections (Figure 4.1b; Vickers et al. 2015). Ceruminous gland tumors have not been documented on other islands (San Clemente Island; SCL and San Nicolas Island; SNI) despite similar levels of chronic mite infection, or in the three island fox populations that do not have ear mites. Thus, mites and genetic effects may explain why cancers are restricted to Santa Catalina Island. This variation in tumor prevalence allows for the unique opportunity to investigate the role of inflammation in cancer development in a set of replicate natural populations.

4.4.3 Genomic studies of wildlife cancer

The genomics of cancer in wildlife, particularly species of conservation or management concern, is informed by general principles in conservation genomics. The list of examples of wildlife genomic studies specifically focused on cancer is short and a more complete understanding of factors that drive tumorigenesis across a wide range of wildlife species is

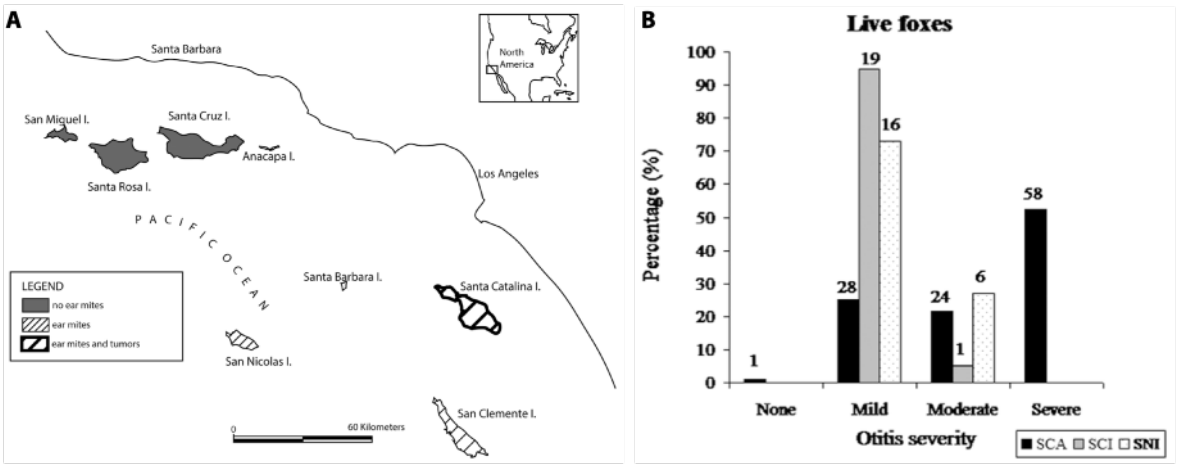

Figure 4.1: Elevated levels of inflammation due to ear mite infection in Catalina Island foxes. A. Map of the Channel Islands (California, USA), which shows the distribution of fox populations with presence/absence of ear mite infections as well presence/absence of ceruminous gland tumors. Island foxes do not reside on Santa Barbara or Anacapa islands. B. The frequency distribution of inflammation (otitis) severity scores in biopsy-sampled live adult foxes from Santa Catalina (SCA), San Clemente (SCI), and San Nicolas (SNI) islands. The number above each bar equals the sampling size for each subgroup. Figures are reprinted from (Vickers et al. 2015).

lacking, but necessary. In particular, the large case-control or genome-wide association studies that have been critical to understanding the genetic basis of cancer in humans or other model organisms are often not feasible in wildlife species. Nonetheless, a few genetic studies have identified polymorphisms possibly associated with cancer susceptibility (Browning et al. 2014; Epstein et al. 2016). For example, in a case-control study of California sea lions, urogenital carcinoma was significantly associated with homozygosity of a microsatellite loci within an intron of the heparanase 2 gene (HPSE2; Browning et al. 2014), which has been implicated in several human carcinomas. Several studies have focused on comparative genomics of copy number variants of specific tumor suppressor and genome maintenance genes (Abegglen et al. 2015; Caulin et al. 2015; Gorbunova et al. 2014; MacRae et al. 2015; Sulak et al. 2016; Tollis et al. 2017).

Some of these genomic studies aim to explain 'Peto's Paradox' or the lack of correlation between body size, life-span, and cancer risk (Figure 4.2; Abegglen et al. 2015). The expectation is that larger body size results in more somatic cell divisions, greater number of somatic cells, and more opportunity for somatic mutations leading to cancer; however,

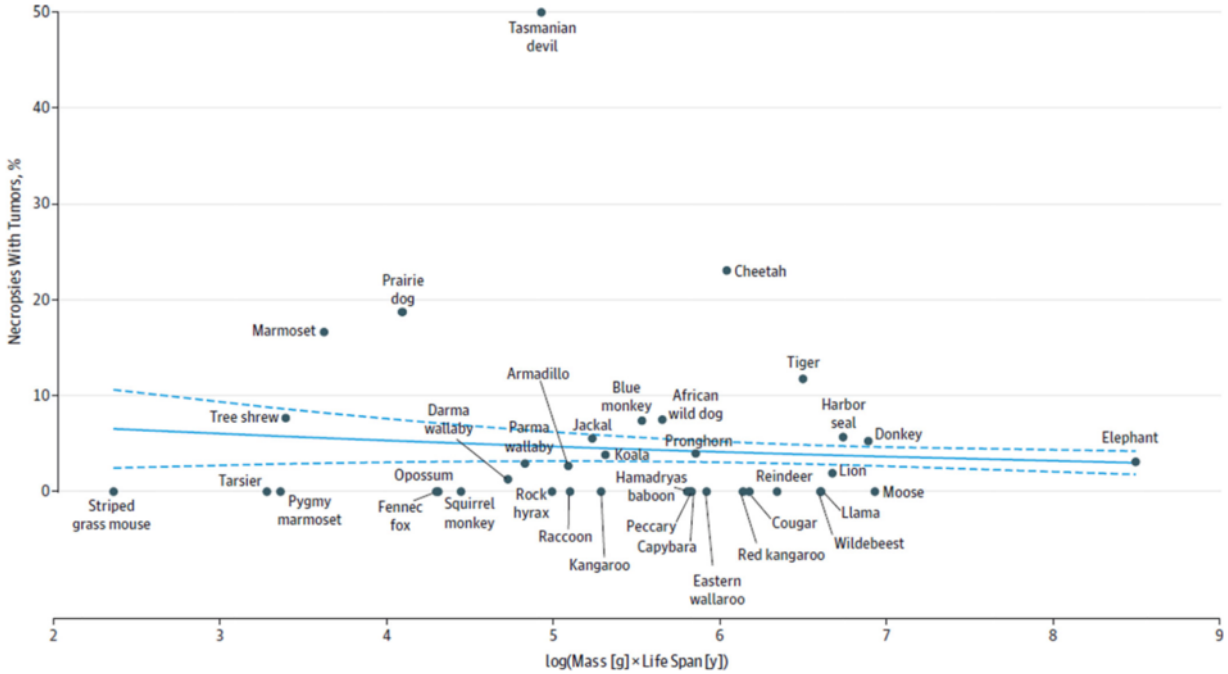

Figure 4.2: Cancer Incidence Across Species by Body Size and Life Span. Cancer incidence is not associated with mass and life span, as shown by the logistic regression (model fit shown as blue line; 95% CIs shown as dashed lines). Each data point in the graph is supported by a minimum of 10 necropsies for the included mammals (San Diego Zoo) and 644 annotated deaths for elephants (Elephant Encyclopedia database). The risk of cancer depends on both the number of cells in the body and the number of years over which those cells can accumulate mutations; therefore, cancer incidence is plotted as a function of mass Ã life span. This figure is reprinted from (Abegglen et al. 2015).

this is not the case. A recent study found that elephants, which have low cancer mortality, possess multiple copies of the TP53 (p53) tumor suppressor gene as compared with 61 other vertebrate species (Figure 4.3; Sulak et al. 2016). Elephant cells, as compared to human cells, demonstrate an increased p53-mediated apoptotic response following DNA damage (Abegglen et al. 2015), which may be due to the transcription and likely translation of several of the TP53 retrogenes (Sulak et al. 2016). Further, upon analyzing copy numbers of TP53 in other Proboscideans, Sulak et al. (2016) found that as species evolved larger body sizes they also evolved more TP53 retrogenes (Figure 4.3). In another recent study on mice and rat species, telomere maintenance strategies were found to differ depending on body mass and differential cancer risks (Tian et al. 2018). Larger species evolved repression of somatic telomerase activity and replicative senescence while smaller species evolved telomere-independent anti-cancer mechanisms that act to slow down cell proliferation and prevent pre-malignant hyperplasia. Overall, these studies largely aim to expand our knowledge of basic biological principles and identify cancer suppression mechanisms with potential therapeutic relevance.

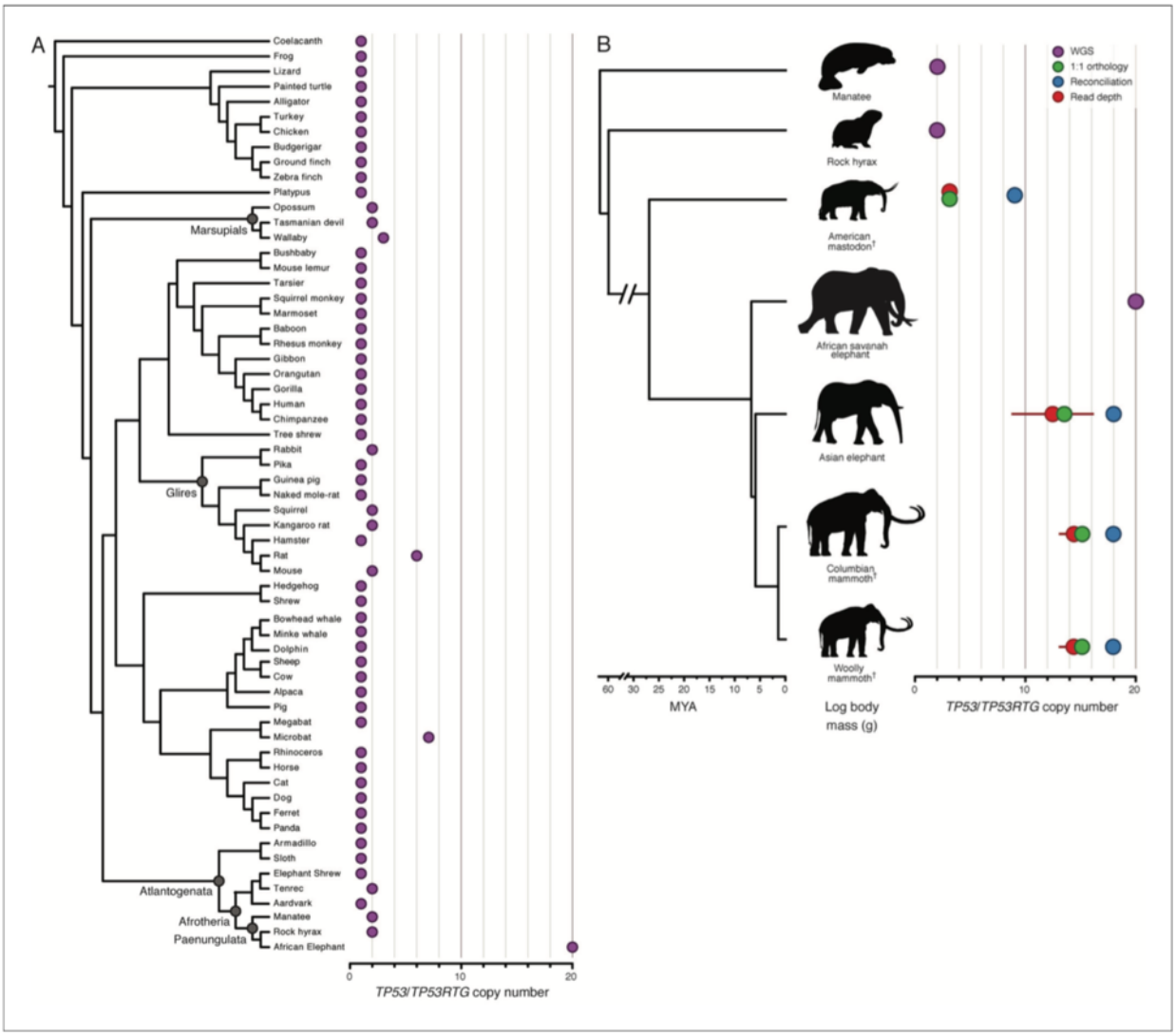

Figure 4.3: Expansion of the TP53RTG gene repertoire in Proboscideans. (A) TP53 copy number in 61 Sarcopterygian (Lobe-finned fish) genomes. Clade names are shown for lineages in which the genome encodes more than one TP53 gene or pseudogene. (B) Estimated TP53/TP53RTG copy number inferred from complete genome sequencing data (WGS, purple), 1:1 orthology (green), gene tree reconciliation (blue), and normalized read depth from genome sequencing data (red). Whiskers on normalized read depth copy number estimates show the 95% confidence interval of the estimate. This figure is reprinted from (Sulak et al. 2016).

4.5 Tasmanian devils and DFTD as a case example

4.5.1 An epidemic transmissible cancer

Tasmanian devils (*Sarcophilus harrisii*) suffer from devil facial tumor disease (DFTD), one of only a handful of transmissible cancers known in the animal kingdom (Jones et al. 2019; Metzger and Goff 2016; Russell et al. 2018). DFTD was first observed in 1996 by a wildlife photographer (Hawkins et al. 2006). Since then, the disease has spread across most of the island with only a few populations unaffected in the far western and northwestern coast. The wild devil population has been decimated by 80% due to these metastatic tumors that typically result in mortality within six months to one year of transmission (Hamede et al. 2012; Hamede et al. 2015; Jones et al. 2019; Lazenby et al. 2018). DFTD cells are undifferentiated neoplasms with highly pleomorphic and anaplastic cells (Pyecroft et al. 2007). Tumors result in ulcerating proliferative masses that tend to occur around the face and jaw, and masses within the oral cavity can prevent feeding and are prone to secondary infection (Hawkins et al. 2006). Live cancer cells are the infectious agent and are transmitted to new hosts during biting behavior while feeding and during the breeding season (Hamilton et al. 2019; Pearse and Swift 2006; Pyecroft et al. 2007). Uninfected, aggressive biters become infected after biting the tumors of infected, less aggressive bite recipients; therefore, more socially dominant devils are more likely to get cancer (Wells et al. 2017). Thus far, there is no evidence of vertical transmission from mothers to their offspring (McCallum et al. 2007), and low levels of prevalence in juveniles could be associated with dramatic changes in immune capacity at sexual maturity (Cheng et al. 2017), but downward changes in age structure in affected populations have been observed (Hamede et al. 2012; Lachish et al. 2009). Changes in life history strategies have also been observed; for instance, as the first year of breeding has shifted from age 3 to age 2, and to age 1 in some cases (Jones et al. 2008; Lachish et al. 2009). Models have predicted that the population could possibly go extinct (McCallum et al. 2009); however, no local populations have yet completely disappeared

(Lazenby et al. 2018; Storfer et al. 2018).

The etiology of DFTD and characterization of the cell of origin was largely determined through molecular cytogenetic, immunogenetic, and genomic methods. Clonality of DFTD was initially established by karyotypic data, which showed that tumors from different individuals contain the same complex chromosomal rearrangements (Deakin et al. 2012; Pearse and Swift 2006). Microsatellite and MHC analysis indicating a lack of diversity across tumors, consistent with clonal transmissibility (Siddle et al. 2007). Further, DFTD tumors share similar microsatellite genotypes across all loci tested, regardless of location, sex, or age of devil (Murchison et al. 2010). Tumors were found to express diagnostic neuron-specific markers indicating that the ancestral cell type of DFTD was neuroendocrine origin (Loh et al. 2016). Antibody staining indicated that tumor cells produce a Schwann cell-specific protein, periaxin (Murchison et al. 2010), which is now considered a sensitive and specific diagnostic for DFTD tumors (Tovar et al. 2011). DFTD is derived from a precursor neural crest cell given that Schwann and neuroendocrine cells are both derived from the neural crest and overlap in gene expression.

Two hypotheses have been considered regarding host evasion leading to DFTD's rapid spread and near-universal susceptibility to the disease. First, irregular tumor MHC expression and downregulation of host MHC by DFTD allow the tumor to escape host surveillance (Siddle et al. 2013). During the initial neoplastic transformation, epigenetic downregulation of multiple aspects within the antigen-presenting system occurs (Siddle et al. 2013). This leads to DFTD's inability to display functional MHC class I molecules, *in vivo* or *in vitro* and, thus, avoids recognition by host immune cells, specifically T cells. Second, devils may lack enough MHC diversity to recognize and destroy aberrant tumor cells (Siddle et al. 2007). The tumor, which is an allograft, should be recognizable by the host immune system and subsequently rejected. However, Siddle et al. (2017) did not detect lymphocyte response when lymphocytes from devils were tested against each other as well as lymphocytes isolated from other parts of the island. Therefore, the immune system is incapable of recognizing

foreign lymphocytes as non-self, much less as a tumor. The study further suggests the lack of MHC diversity in the population (Siddle et al. 2007). Alternatively, both hypotheses may account for the rapid spread of DFTD.

Remarkably, a second transmissible cancer has arisen recently in Tasmanian devils, called DFT2, with multiple lines of evidence supporting the view that it originated independently from the first DFTD (i.e. DFT1; Pye et al. 2015). DFT2 appeared in a geographically distinct area (southern Tasmania, as opposed to northeast Tasmania for DFTD), and cytogenetic evidence suggests that DFT2 originated in a male devil, in contrast to a female devil for DFTD. While similar in cell type origin, mode of transmission, and gross appearance, these two transmissible cancers differ in histology, in the specific mutations characteristic to each, and in the way in which changes in MHC expression facilitate evasion of the host immune system (Caldwell et al. 2018; Pye et al. 2015; Stammnitz et al. 2018). The independent origin of two transmissible cancers in Tasmanian devils within just two decades raises the hypothesis that devils are uniquely susceptible to this type of disease, and similarities among them may point toward the specific mechanisms that allow transmissible cancers in this species (Stammnitz et al. 2018). It may also be the case that transmissible cancers are more widespread across the animal kingdom than previously recognized.

4.5.2 Devil genomics

An understanding of phylogeographic patterns of genetic diversity may inform management actions, such as translocations or reintroductions, which may be warranted in response to DFTD. Genetic diversity and population connectivity will strongly influence the ability of devils to adapt to other threats such as environmental change and anthropogenic disturbances (Hendricks et al 2017). Previous studies have revealed devils have low genetic diversity and limited population structure. These studies focused on relatively few (<12) microsatellite loci (Brüniche-Olsen et al. 2014; Jones et al. 2004; Storfer 2017), MHC loci (Siddle et al. 2010), mitochondrial genomes (Miller et al. 2011), SNPs (Hendricks et al.

2017; Miller et al. 2011; Morris et al. 2015; Wright et al. 2015), or whole-genome sequencing (Miller et al. 2011; Murchison et al. 2012). Low genetic diversity in Tasmanian devils is the result of historical fluctuations in population size, and extinction of the species on mainland Australia and its restriction to the island of Tasmania (Guiler 1978; Hawkins et al. 2006; Olsen et al. 2018).

Despite the overall low genetic diversity of the species, several lines of evidence suggest the potential for an evolutionary response to the strong selection imposed by DFTD (Hohenlohe et al. 2019). First, three independent populations were found to show a parallel, rapid (4-6 generations) evolutionary response to the disease (Epstein et al. 2016). This study scanned across 90K SNP loci for signatures of selection and found two genomic regions (Figure 4.4), which contained genes with immunological and oncogenic functions (Epstein et al. 2016). Second, using the data from Epstein et al. (2016), another study used a maximum likelihood approach and improved functional annotations to find more signatures of selection in the devil genome (Hubert et al. 2018). In total, 97 genomic regions were found to putatively be under selection, most of which were population-specific with one region common to all three populations. These regions harbored 148 protein-coding genes (or human orthologues), nearly all of which have a link with cancer. Third, a genome-wide association study (GWAS) of ~600 individuals found that phenotypic variation in female survivorship (length of time after infection) could be explained by a few loci of large effect (~5 SNPs explained about >61% of the total variance; (Margres et al. 2018a). Further, Margres et al (2018a) found that female infection rates (female case-control) could be explained by more SNPs of smaller effect (~56 SNPs explained about >23% of the total variance). Given that DFTD has spread across multiple genetic clusters in the devil population (Hendricks et al. 2017), any allelic variation for resistance to DFTD may be able to spread across the devil population and increase in frequency because of selection.

Genomics is beginning to reveal mechanisms leading to spontaneous tumor regression or even complete recovery from the disease in a few devils (fewer than 20; Pye et al. 2016;

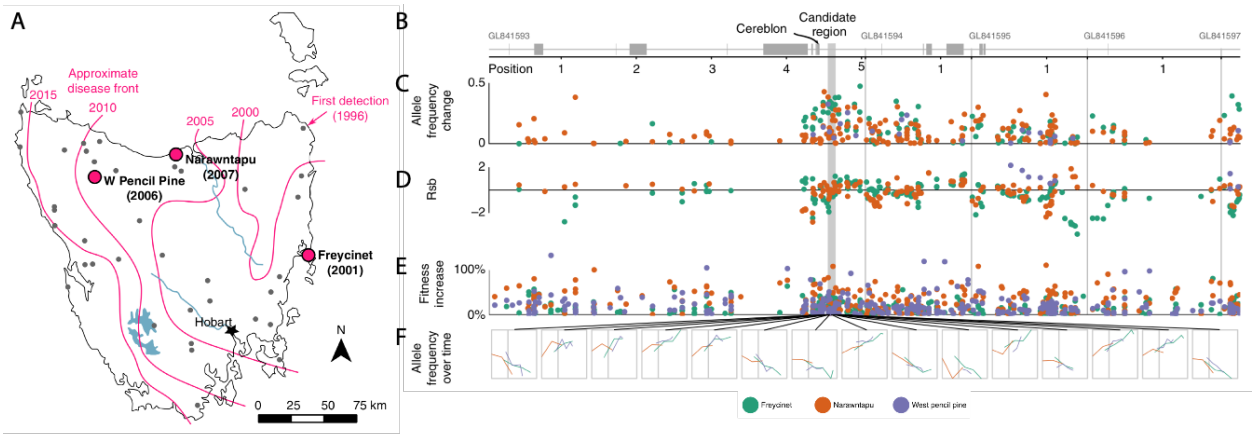

Figure 4.4: Sample collection sites of Tasmanian Devils and selection test statistics of each SNP in one candidate region (chromosome 2) and approximately 4 Mb on either side. A. The three focal populations are labelled and marked with large magenta circles. The magenta lines indicate the approximate location of the disease front (Hamede et al. 2015; Pye et al. 2015) in 2000, 2005, 2010 and 2015. B. The scaffolds, positions and genes (grey boxes) within the candidate region and surrounding genomic area. The positions are given in Mb from the start of each scaffold, which are marked with light grey vertical lines and a label (GL841593, and so on). Values of three statistics are shown for each filter-passing SNP: (C) allele frequency change; (D) rsb; and (E) point estimates of the fitness advantage of the increasing allele. Panel F show the trajectory of allele frequency change over time; for clarity, we only show SNPs with relatively high genotyping rates and the x axis is time since detection of DFTD (first detection of DFTD is marked with a vertical line). SNPs are colour-coded by population, the candidate region is marked with a dark grey box and the names of candidate genes are labelled. Due to multiple steps of data filtering, each population has a different set of SNPs. This figure is reprinted from (Epstein et al. 2016).

Wright et al. 2017). Using a comparative case-control genomic approach, two key genomic regions were identified to putatively be associated with tumor regression and, therefore, the ability to survive DFTD (Wright et al. 2017). Using targeted genotyping in additional samples, the authors were able to confirm that three genes may be involved in slowing tumor growth and allowing additional time for the effected individual to mount an immune response (Wright et al. 2017b). Another comparative genomic study found a different set of three highly differentiated regions, which contained several genes with immunological or oncogenetic functions (Margres et al. 2018b). Putative regulatory variation in candidate genes suggests that changes in gene expression may drive natural tumor regression. Despite the small number of animals that have recovered from the disease, strong selection pressure from the disease may cause the frequency of these variants to increase over time.

4.5.3 Tumor genomics

A number of different karyotypic strains have been discovered (McCallum et al. 2007). These strains resemble the original DFTD karyotype reported by (Pearse and Swift 2006; designated strain 1), but are characterized by additional cytogenetic rearrangements consistent with ongoing tumor evolution as the disease continues to spread through the population (Deakin et al. 2012). It appears from both cytogenetic and sequencing analysis that DFTD strains are continuing to accumulate karyotypic, copy number, and sequence variants, but compared with most human cancers, DFTD strains are remarkably stable (Deakin et al. 2012; Murchison et al. 2012). Selection may be working to maintain the tumorigenic properties of the DFTD genome, while permitting genomic instability and sequence substitutions in regions not critical for the survival of the DFTD cell (Deakin et al. 2012).

The number of somatic point mutations varies widely in humans, yet the mutation rate in DFTD is likely to be less than some human cancers, such as lung or skin cancer (Martincorena and Campbell 2015). As compared to the reference devil genome, the two DFTD genomes sequenced indicate that approximately 17,000 somatic mutations are present in the tumor

(Murchison et al. 2012). The presence of transversion mutations in DFTD is consistent with an endogenous process, such as a DNA repair defect, although an exogenous etiology cannot be excluded (Murchison et al. 2012). The nonsynonymous to synonymous (NS/S) ratios for unique variants in the two tumor genomes were within the typical range of somatic variants in human cancers. The majority of the copy number variants identified between the two tumors were common to both lineages. However, some copy number variants occurred in only one of the two tumors and are likely to have been somatically acquired since the divergence of the two tumor lineages indicating DFTD is evolving. Distinguishing somatic mutations from those found in the original or transient host is important for understanding what drives tumor growth and how the tumor evades immune detection by accumulating mutations in pathways related to recognition of self versus non-self. Genomic approaches, such as those used in canids (Decker et al. 2015), involve including large catalogs of variation found in modern devils, which are critical for identifying these somatic mutations.

4.5.4 Conservation of Tasmanian devils

The potential for adaptation to DFTD in devils, illustrated by the genetic variation for disease phenotypes and evidence of a rapid response to selection described above, has consequences for conservation and management of devils. For instance, supplementing wild populations with devils from captive populations that have not be exposed to the disease could increase the severity of the disease by increasing transmission rates and population-level susceptibility (Hohenlohe et al. 2019). Thus attempts at demographic rescue - increasing population size with supplementation in areas where the disease has greatly reduced devil density - could be counter-productive because of the effects on disease dynamics. Modeling of the devil-DFTD system can predict future outcomes to help guide conservation strategies, particularly if genetic variation and evolution can be explicitly included in the models (Wells et al. 2019). Additionally, the discovery of DFT2 favors the view that conservation strategies for devils consider not just genetic variation relevant to DFTD, but also genetic variation

relevant to immune function and cancer in general that could provide adaptive potential for the future (Hohenlohe et al. 2019).

4.6 Future directions in the genomics of wildlife cancer

4.6.1 Surveillance

There are several key elements for improving early recognition of cancer epizootics in wildlife and utilizing wild populations for biomedical cancer research. Strategic activities include identifying, coordinating and expanding existing surveillance networks, and properly collecting samples from diseased individuals and populations during surveillance. With limited understanding of the rate of occurrence and types of cancers in most wildlife species, surveys of cancer reports from zoological data, field work, and primary literature would provide further information on which species and their genomes to study. Increased funding for multi-disciplinary scientific research and training to develop the capacity for disease diagnostics and epidemiology would promote surveillance capabilities. This has the potential to drive timely environmental mitigation and influence environmental policy to reduce environmental contamination that may cause cancer and therefore benefit both wild and human populations.

4.6.2 Research

Methodological advances in high-throughput sequencing and genomics will benefit the study of cancer in wildlife. High levels of diversity and gene duplication (Nei et al. 1997; Temperley et al. 2008) that make immunity highly adaptable, also make immune-gene regions challenging to assemble. Therefore, it is difficult to determine how many copy-number variants of genes exist in a species or individual genome (Alcaide et al. 2014; Cheng et al. 2012). Longer sequencing reads are available with the emergence of new technologies developed by Oxford NanoPore and Pacific BioSciences (PacBio). Additionally,

continued development and assessment of computational approaches [e.g. Chin et al. 2013; Putnam et al. 2016; Salzberg et al. 2012], may aid in resolving the challenges presented by gene duplications. Further resolving this as well as runs of homozygosity, a measure of inbreeding, will help identify candidate loci involved in inbreeding depression in addition to causal polymorphisms in disease susceptibility.

A more detailed investigation of shared cancers using a multi-species approach will highlight genes associated with carcinogenesis in the context of risk related to both genetics and environmental exposure. Important insights can also be gained from studying lineages that have a high prevalence for cancer. Marine mammals (Schiffman and Breen 2015), Catalina Island Foxes (Vickers et al. 2015), and Tasmanian devils (discussed above) have all been found to have elevated risks of cancer development. The increasing number of genomes available from different species, including those with higher than average risks of cancer, could reveal the link between genotypes and phenotypes and functional consequences of many mutations in cancer (Kumar et al. 2011).

As the taxonomic scope of genomic data continues to increase, comparative genomics approaches can also lead to new understanding of the genetic basis of cancer susceptibility and mechanisms of cancer development in wildlife (Tollis et al. 2017; Caulin and Maley 2011). For example, the long lives, slow developmental rates, probable low cancer rates, and the rapid development of genomic resources for large reptiles (Tollis et al. 2015) will provide ample opportunity to study genomic mechanisms of cancer suppression in these ectothermic amniotes. Further, birds were found to have a lower incidence of cancer than mammals at the San Diego Zoo (Effron et al. 1977), which suggests that the numerous avian genomes available (Zhang et al. 2014) could provide more information regarding cancer suppression. Further, future studies should include functional experiments to understand evolution's mechanisms and strategies for cancer resistance, which could benefit both wildlife and humans through the practice of evolutionary medicine and precision medicine (Whilde et al. 2017).

Human cancer research increasingly recognizes the role of genetic variation and evolution within tumor cell populations. Diversity in the premalignant biopsies can predict the risk of progression to cancer (Maley et al. 2006; Merlo et al. 2010), just as the extent of genetic variation predicts the response to selection in any evolutionary system. Gatenby's adaptive therapy algorithm (Gatenby et al. 2009) shifts the focus from eliminating every cancer cell as used by many traditional methods, to controlling cancer by manipulating selection forces within the tumor. For example, microinflammation around tumor sites that may spur neoplastic progression (Hochberg et al. 2013) are being reduced through the development of novel therapeutic approaches and therefore incidentally reduce the accumulation of DNA damage. Such an evolutionary perspective may inform cancer research in wildlife species as well, particularly in the case of transmissible cancers in which the evolving tumor cell population is spread across many hosts and outlives any individual infected host.

4.6.3 Captive breeding programs

Genomics can support the identification of candidate loci responsible for heritable disorders, which can inform breeding decisions in captive populations of wildlife species. Genome-wide association studies (GWAS) have found unprecedented numbers of variants associated with complex human traits and diseases such as cancer. From the results of GWAS, genetic panels have been developed for preventive and personalized medicine (Vazquez et al. 2012). GWAS is increasingly being applied in wildlife species, including studies of cancer-related traits (Margres et al. 2018a). Once causal variants have been identified in a captive wildlife population, genetic information on these loci can be combined with pedigree information and used for strategic breeding. For example, this method was used in the case of the critically endangered California Condor (*Gymnogyps californianus*) and the lethal disease, chondrodystrophy (Grueber 2015; Romanov et al. 2009). Through the pedigrees obtained in the captive breeding program, researchers found this disease to show Mendelian segregation (Ralls et al. 2000). Genomic resources were developed to identify causal polymorphism

linked to the disease (Romanov et al. 2009; Walters et al. 2010) with the aim of informing the captive breeding protocols to reduce the frequency of chondrodystropy while maintaining genetic diversity at other loci. Overall, this resource would help to safeguard against inbreeding to avoid further decreases in individual fitness (Frankham 2010).

The Tasmanian devil insurance program incorporates molecular data (SNPs) with studbook management to provide a reference for captive breeding programs to ensure Tasmanian devil diversity is not further depleted (Wright et al. 2015). Genetic assays will continue to be used to monitor the genome-wide genetic diversity of the insurance populations with the aim of reducing inbreeding and maintaining variation. However, the captive devil population is not currently managed for variation at any specific cancer-related loci, as these loci are continuing to be identified as described above.

4.6.4 Genomics for monitoring and conservation of natural populations

With the expanding set of tools for genotyping panels of genetic variants in wildlife species, cancer-related marker panels could be informative for conservation and management. While genetics-informed, individual-level treatment may be rare outside of highly valuable captive populations, marker panels that could predict population-level susceptibility could be applied to natural populations. Particularly in wildlife populations with high prevalence of a specific hereditary or environmental cancer, or in the case of transmissible cancers, the disease may have a substantial impact on population fitness and viability. A marker panel could be used in monitoring to track any evolutionary response to the disease, and to allow predictions about future population outcomes. It could also provide critical information for evaluating translocations or attempts at genetic rescue, by quantifying variation at cancer-related loci in potential source populations.

4.6.5 Vaccinations and immunotherapy

The use of various genomic datasets may provide a foundation for identifying avenues for vaccination and treatment of cancers. Several studies have explored immune-stimulatory agents and vaccines against DFTD (Patchett et al. 2017; Tovar et al. 2018; Tovar et al. 2017). For example, heat shock proteins (HSPs) derived from tumor cells have been used as a source of antigens for cancer immunotherapy in humans (Murshid et al. 2008). A recent study by (Tovar et al. 2018) found that DFTD cancer cells express inducible HSP, which supports that a HSP-based vaccine against DFTD could be developed. Advancements in oral vaccine development and delivery for infectious diseases have led to successful infectious disease control as seen in the case of sylvatic plague affecting prairie dogs (*Cynomys spp.*) and the endangered black-footed ferret (*Mustela nigripes*; see Salkeld 2017). Similar methods against transmissible and viral-associated cancers could be implemented once vaccine targets are discovered.

4.7 Key mutation types in cancer

Cancer suppression involves two (sometimes overlapping) components: oncogenes and tumor suppressor genes. Oncogenes or the "gatekeepers" are the genes directly involved in preventing unregulated cell division (Kinzler and Vogelstein 1997). Tumor suppressor genes or the "caretakers" (Shields and Harris 2000) are involved in error-free DNA replication, effective DNA repair, and the maintenance of appropriate epigenetic patterning (Sarkies and Sale 2012) and chromosomal structure (Stoler et al. 1999). The multistage theory of carcinogenesis proposes that cancers proceed through successive stages, corresponding to one or more mutations (or epigenetic alterations) activating oncogenes and disabling tumor suppressor genes (Armitage and Doll 1954; Nordling 1953). The accumulation of these small genetic changes over time lead to large effects in phenotype (Hanahan and Weinberg 2011; Hanahan and Weinberg 2000). Both the number and the nature of the genes involved in

cancer suppression is at least somewhat tissue-specific.

While there is no fixed cancer genome, with a majority of mutations differing between tissue types and even cells within the one tumor, there are few common mutations across various cancer types. These driver mutations of a cell confer a selective advantage over those of surrounding cells that leads to growth advantage. Passenger mutations are incidental and have no to little effect on growth (Gerlinger et al. 2014; Stratton et al. 2009). Currently, the full picture of the cancer genome landscape is being explored through the advancement of genome sequencing (reviewed in Heng 2017). Known key mutation types are as follows:

1. Single nucleotide polymorphism (SNPs) make up approximately 95% of mutations from cancer genomes (Heng 2017). These mutations can result in nonsynonymous changes in proteins or other functional consequences such as changes in regulatory binding sites and micro-RNA loci.

2. Copy number variation (CNV) is defined as the amplification or deletion of DNA fragments >50bp (Girirajan et al. 2011). Somatic copy number alterations (SCNAs) are common in cancer; however, distinguishing driver SNCAs from numerous SCNAs that randomly accumulate during tumorigenesis is not straightforward (Heng 2017). Further, many known SCNAs are not directly related to cancer genes (Zack et al. 2013).

3. Complex structural chromosomal abnormalities such as translocations are extremely common for many cancer types and can have large effects on gene function and expression (Horne et al. 2013; Stephens et al. 2009). Chromothripsis is defined by a single, localized event within genomic regions in one or few chromosomes characterized by thousands of clustered chromosomal rearrangements. Similarly, chromoplexy is characterized by chromosomal rearrangements that involve segments of DNA from multiple chromosomes (e.g., five or more). These abnormalities have been implicated in cancer phenotypes, particularly metastasis and drug resistance (Heng et al 2016a).

4. Deregulation of telomere dynamics are involved in many cancers. Progressive shortening of telomeres typically induces cellular senescence. However, telomere fusions can result

when the shortened ends of the chromosome merge with another shortened telomere, which causes genome destabilization. Further, telomerase activity can lead to cancer-promoting phenotypes, such as over-proliferation and cell survival (Artandi and DePinho 2009).

5. Epigenetic factors, heritable changes in gene expression that are not accompanied by changes in DNA sequence, can determine cancer phenotype by altering DNA accessibly (Jones and Baylin 2007). Abnormalities in methylation, histone modification, nuclear topology, noncoding RNA have been implicated in the silencing of key tumor suppressor, regulatory, and repair genes resulting in cancer (reviewed in (Grunau 2017).

4.8 Management and Conservation using genomic data

Cancer may be one of many factors creating concern for conservation of wildlife populations, and genomics can provide powerful tools for assessing their impact. High-throughput genomic technologies have increased our ability to assess inbreeding coefficients (Kardos et al. 2015; Kardos et al. 2016), gene flow, demography including effective population size (Barbato et al. 2015), adaptive potential (Flanagan et al. 2017; Hoelzel et al. 2019), and wildlife epidemiology (Blanchong et al. 2016), important issues for population viability. We also have an increased ability to discover loci associated with adaptive variation or with reduced fitness (Hohenlohe et al. 2018; Robinson et al. 2018). When populations are small, both inbreeding and genetic drift can increase homozygosity at loci with deleterious alleles, reducing fitness and contributing significantly to extinction risk (Frankham, 2005a, 2005b; O'Grady et al., 2006). The difference in the mean fitness of a population and the fitness of an optimal genotype that does not carry deleterious mutations is called the genetic load (Glémin, Ronfort, & Bataillon, 2003; Kirkpatrick & Jarne, 2000). As a population loses variation via drift and inbreeding, genetic load increases. Hereditary cancer susceptibility due to the accumulation of oncogenic mutations could be a source of genetic load in wildlife populations.

If a population suffers from genetic load or inbreeding, genetic rescue through mediated

migration, translocation, and reintroduction via captive breeding programs can increase population fitness due to an increase in heterozygosity, which can mask deleterious mutations, and facilitate adaptive evolution (Frankham 2016; Hedrick and Garcia-Dorado 2016; Hufbauer et al. 2015; Tallmon et al. 2004; Weeks et al. 2011; Whiteley et al. 2015). Genomic tools can be used to inform genetic rescue, for instance by identifying source populations or assessing the risk of outbreeding depression (Fitzpatrick and Funk 2019). Alternatively, evolutionary rescue, evolution from standing genetic variation without migration (Hufbauer et al. 2015), may be possible particularly when there is evidence that a population is able to purge strong deleterious mutations, such as those that cause inherited cancer syndromes. Many wildlife populations have been fragmented into smaller populations that are subject to genetic drift, with the potential for increased cancer susceptibility. To the extent that hereditary cancer is caused by relatively rare, deleterious variants in these fragmented populations, genetic rescue may be highly effective in reducing cancer susceptibility.

4.9 Acknowledgements

Support was provided to SAH by the National Institute of Health grant P30GM103324; NIH grant R01GM126563 and National Science Foundation grant DEB-1316549 as part of the joint NIH-NSF-USDA Ecology and Evolution of Infectious Diseases Program; and the Bioinformatics and Computational Biology program at the University of Idaho.

4.10 References

Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, et al. Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA. 2015 Nov 3;314(17):1850-22.

Aktipis CA, Nesse RM. Evolutionary foundations for cancer biology. Evol Appl. 9 ed. 2013 Jan 21;6(1):144-59.

Alcaide M, Muñoz J, Martínez-de la Puente J, Soriguer R, Figuerola J. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot Fulica atra (Aves: Rallidae). Ecol Evol. 2014 Mar;4(6):688-98.

Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Current Opinion in Genetics & Development. 2014 Feb 1;24:52-60.

Alvarado LC. Do evolutionary life-history trade-offs influence prostate cancer risk? a review of population variation in testosterone levels and prostate cancer disparities. Evol Appl. 4 ed. 2013 Jan 1;6(1):117-33.

Antoni MH, Lutgendorf SK, Cole SW, Dhabhar FS, Sephton SE, McDonald PG, et al. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. Nat Rev Cancer. 2006 Mar;6(3):240-8.

Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. British Journal of Cancer. Nature Publishing Group; 1954 Mar;8(1):1-12.

Arriagada G, Metzger MJ, Muttray AF, Sherry J, Reinisch C, Street C, et al. Activation of transcription and retrotransposition of a novel retroelement, Steamer, in neoplastic hemocytes of the mollusk Mya arenaria. Proc. Natl. Acad. Sci. U.S.A. National Acad Sciences; 2014 Sep 30;111(39):14175-80.

Artandi SE, DePinho RA. Telomeres and telomerase in cancer. Carcinogenesis. 2009.

Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 2015 Mar 20;6:1294-6.

Barber BJ. Neoplastic diseases of commercially important marine bivalves. Aquatic Living Resources. EDP Sciences; 2004 Oct 1;17(4):449-66.

Bäcklin BM, Eriksson L, Olovsson M. Histology of Uterine Leiomyoma and Occurrence in Relation to Reproductive Activity in the Baltic Gray Seal (*Halichoerus grypus*). Veterinary Pathology. 4 ed. SAGE PublicationsSage CA: Los Angeles, CA; 2016 Jun 26;40(2):175-80.

Bissell MJ, Radisky D. Putting tumours in context. Nat Rev Cancer. Nature Publishing Group; 2001 Oct 1;1(1):46-54.

Black JJ, Baumann PC. Carcinogens and cancers in freshwater fishes. Environ. Health Perspect. National Institute of Environmental Health Science; 1991 Jan;90:27-33.

Blanchong JA, Robinson SJ, Samuel MD, Foster JT. Application of genetics and genomics to wildlife epidemiology. Jour. Wild. Mgmt. 2016 Apr 2;80(4):593-608.

Brown ER, Hazdra JJ, Keith L, Greenspan I, Kwapinski JB, Beamer P. Frequency of fish tumors found in a polluted watershed as compared to nonpolluted Canadian waters. Cancer Res. American Association for Cancer Research; 1973 Feb;33(2):189-98.

Browning HM, Acevedo-Whitehouse K, Gulland FMD, Hall AJ, Finlayson J, Dagleish MP, et al. Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. Proc. Biol. Sci. 2014 Dec 7;281(1796):20140240-0.

Brüniche-Olsen A, Jones ME, Austin JJ, Burridge CP, Holland BR. Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. Biology Letters. 2014 Nov;10(11):20140619-9.

Buckles EL, Lowenstine LJ, Funke C, Vittore RK, Wong HN, St Leger JA, et al. Otarine Herpesvirus-1, not Papillomavirus, is Associated with Endemic Tumours in California Sea Lions (*Zalophus californianus*). Journal of Comparative Pathology. 2006 Nov;135(4):183-9.

Caldwell A, Coleby R, Tovar C, Stammnitz MR, Kwon YM, Owen RS, et al. The newly-arisen Devil facial tumour disease 2 (DFT2) reveals a mechanism for the emergence of a contagious cancer. eLife. 2018;7.

Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc. Natl. Acad. Sci. U.S.A. National Acad Sciences; 2008 Sep 2;105(35):13081-6.

Carballal MJ, Barber BJ, Iglesias D, Villalba A. Neoplastic diseases of marine bivalves. Journal of Invertebrate Pathology. 2015 Oct;131:83-106.

Caulin AF, Graham TA, Wang L-S, Maley CC. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. Philos. Trans. R. Soc. Lond., B, Biol. Sci. The Royal Society; 2015 Jul 19;370(1673):20140222-10.

Caulin AF, Maley CC. Peto's Paradox: evolution's prescription for cancer prevention. Trends in Ecology & Evolution. 2011 Apr;26(4):175-82.

Cheng Y, Heasman K, Peck S, Peel E, Gooley RM, Papenfuss AT, et al. Significant decline in anticancer immune capacity during puberty in the Tasmanian devil. Scientific Reports. Nature Publishing Group; 2017 Mar 16;7(1):44716.

Cheng Y, Stuart A, Morris K, Taylor R, Siddle H, Deakin J, et al. Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. BMC Genomics. BioMed Central; 2012 Mar 12;13(1):87.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods. 2013 Jun;10(6):563-9.

Coussens LM, Werb Z. Inflammation and cancer. Nature. Nature Publishing Group; 2002 Dec 19;420(6917):860-7.

De Stavola BL, Santos Silva dos I, McCormack V et al. (2004) Childhood Growth and Breast Cancer. American Journal of Epidemiology, 159, 671-682.

de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer. Nature Publishing Group; 2006 Jan 1;6(1):24-37.

Deakin JE, Bender HS, Pearse A-M, Rens W, O'Brien PCM, Ferguson-Smith MA, et al. Genomic Restructuring in the Tasmanian Devil Facial Tumour: Chromosome Painting and Gene Mapping Provide Clues to Evolution of a Transmissible Tumour. O'Brien SJ, editor. PLoS Genet. Public Library of Science; 2012 Feb 16;8(2):e1002483.

Decker B, Davis BW, Rimbault M, Long AH, Karlins E, Jagannathan V, et al. Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. Genome Research. 2015 Nov 2;25(11):1646-55.

Effron M, Griner L, Benirschke K. Nature and rate of neoplasia found in captive wild mammals, birds, and reptiles at necropsy. J. Natl. Cancer Inst. 1977 Jul;59(1):185-98.

Epstein B, Jones ME, Hamede R, Hendricks S, McCallum H, Murchison EP, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications. The Author(s); 2016 Aug 30;7:12684.

Ewald PW, Swain Ewald HA. Infection, mutation, and cancer evolution. Journal of Molecular Medicine. 2012.

Ewald PW, Swain Ewald HA. Toward a general evolutionary theory of oncogenesis. Evol Appl. 2013 Jan 1;6(1):70-81.

Ewald PW, Swain Ewald HA. Infection and cancer in multicellular organisms. Philos. Trans. R. Soc. Lond., B, Biol. Sci. The Royal Society; 2015 Jul 19;370(1673):20140224.

Ewald PW, Swain Ewald HA. Infection and Cancer in Nature. In: Ujvari B, Roche B, Thomas F, editors. Ecology and Evolution of Cancer. 2017. pp. 47-56.

Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. Science. 1997 Nov 7;278(5340):1043-50.

Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol Appl. 2017 Dec 2;119(1):1267-18.

Fitzpatrick SW, Funk WC. Genomics for Genetic Rescue. 2019. Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol Appl. 2017 Dec 2;119(1):1267-18.

Frank SA. Genetic predisposition to cancer - insights from population genetics. Nature Reviews Genetics. Nature Publishing Group; 2004 Oct 1;5(10):764-72.

Frankham R. Genetics and extinction. Biological Conservation. Elsevier; 2005a Nov 1;126(2):131-40.

Frankham R. Stress and adaptation in conservation genetics. J. Evol. Biol. Blackwell Science Ltd; 2005b Jul 1;18(4):750-5.

Frankham R. Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation. 2010 Sep;143(9):1919-27.

Frankham R. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. Biological Conservation. 2016 Mar;195:33-6.

Fu W, Gittelman RM, Bamshad MJ, Akey JM. Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations. The American Journal of Human Genetics. 2014 Oct;95(4):421-36.

Garraway LA, Lander ES. Lessons from the Cancer Genome. Cell. 2013 Mar 28;153(1):17-37.

Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive Therapy. Cancer Res. American Association for Cancer Research; 2009 Jun 1;69(11):4894-903.

Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer: Evolution Within a Lifetime. Annual Reviews; 2014 Nov 24;48(1):215-36.

Giraudeau M, Sepp T, Ujvari B, Ewald PW, Thomas F. Human activities might influence oncogenic processes in wild animal populations. Nat Ecol Evol. Nature Publishing Group; 2018 Jul 1;2(7):1065-70.

Girirajan S, Campbell CD, Eichler EE. Human Copy Number Variation and Complex Genetic Disease. Annu. Rev. Genet. 2011 Dec 15;45(1):203-26.

Glémin S, Ronfort J, Bataillon T. Patterns of Inbreeding Depression and Architecture of the Load in Subdivided Populations. Genetics. Genetics; 2003 Dec 1;165(4):2193-212.

Gomulkiewicz R, Holt RD. When does Evolution by Natural Selection Prevent Extinction? Evolution. 1995 Feb;49(1):201.

Gorbunova V, Seluanov A, Zhang Z, Gladyshev VN, Vijg J. Comparative genetics of longevity and cancer: insights from long-lived rodents. Nature Reviews Genetics. 2014 Jul 1;15(8):531-40.

Grueber CE. Comparative genomics for biodiversity conservation. CSBJ. Elsevier B.V; 2015;13(C):370-5.

Grunau C. Chapter 6 The Epigenetic Component in Cancer Evolution. Ecology and Evolution of Cancer. Elsevier; 2017. pp. 87-98.

Guiler ER. Observations on the Tasmanian Devil, Sarcophilus harrisi (*Dasyuridae: fvlarsupiala*) at Granville Harbour, 1966-75. Papers and Proceedings of the Royal Society of Tasmania.

Guo S, DiPietro LA. Factors Affecting Wound Healing. Journal of Dental Research. SAGE Publications; 2010 Feb 5;89(3):219-29.

Hamede R, Lachish S, Belov K, Woods G, Kreiss A, Pearse A-M, et al. Reduced effect of Tasmanian devil facial tumor disease at the disease front. Conservation Biology. Blackwell Publishing Inc; 2012 Feb;26(1):124-34.

Hamede RK, Pearse A-M, Swift K, Barmuta LA, Murchison EP, Jones ME. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. Proc. Biol. Sci. The Royal Society; 2015 Sep 7;282(1814):20151468.

Hamilton DG, Jones ME, Cameron EZ, McCallum H, Storfer A, Hohenlohe PA, et al. Rate of inter-sex interactions affects injury likelihood in Tasmanian devil contact networks. Behavioral Ecology. 2019.

Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000 Jan 7;100(1):57-70.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-74.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, et al. Emerging disease and population decline of an island endemic, the Tasmanian devil (*Sarcophilus harrisii*). Biological Conservation. 2006 Aug;131(2):307-24.

Hedrick PW, Garcia-Dorado A. Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends in Ecology & Evolution. 2016 Dec;31(12):940-52.

Hendricks S, Epstein B, Schonfeld B, Wiench C, Hamede R, Jones ME, et al. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). Conservation Genetics. Springer Netherlands; 2017 Feb 7;18(4):977-82.

Heng HH. Chapter 5 The Genomic Landscape of Cancers. Ecology and Evolution of Cancer. Elsevier; 2017. pp. 69-86.

Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating Mutation Load in Human Genomes. Nature Reviews Genetics. NIH Public Access; 2015 Jun 1;16(6):333-43.

Hochberg ME, Thomas F, Assenat E, Hibner U. Preventive Evolutionary Medicine of Cancers. Evol Appl. 2013 Jan 1;6(1):134-43.

Hoelzel AR, Bruford MW, Fleischer RC. Conservation of adaptive potential and functional diversity. Conservation Genetics. Springer Netherlands; 2019;20(1):1-5.

Hofman P, Vouret-Craviari V. Microbes-induced EMT at the crossroad of inflammation and cancer. Gut Microbes. Taylor & Francis; 2012 May 1;3(3):176-85.

Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population Genomics Provides Key Insights in Ecology and Evolution. Population Genomics. 2nd ed. Cham: Springer, Cham; 2018. pp. 483-510.

Hohenlohe PA, McCallum HI, Jones ME, Lawrance MF, Hamede RK, Storfer A. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Conservation Genetics. Springer Netherlands; 2019 Feb 14;20(1):81-7.

Horne SD, Abdallah BY, Stevens JB, Liu G, Ye KJ, Bremer SW, et al. z. Systems Biology in Reproductive Medicine. FASEB: Nuclear Structure and Cancer. 2013 May 2;59(3):124-30.

Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: Associations and interactions in an international case-control study. International Journal of Cancer. Wiley Subscription Services, Inc., A Wiley Company; 1990 Nov 15;46(5):796-800.

Hufbauer RA, Szűcs M, Kasyon E, Youngberg C, Koontz MJ, Richards C, et al. Three types of rescue can avert extinction in a changing environment. Proc. Natl. Acad. Sci. U.S.A. National Acad Sciences; 2015 Aug 18;112(33):10557-62.

Hubert J-N, Zerjal T, Hospital F. Cancer- and behavior-related genes are targeted by selection in the Tasmanian devil (Sarcophilus harrisii). Crowther MS, editor. PLoS ONE. Public Library of Science; 2018 Aug 13;13(8):e0201838.

Irigaray P, Newby JA, Clapp R, Hardell L, Howard V, Montagnier L, et al. Lifestyle-related factors and environmental agents causing cancer: an overview. Biomed. Pharmacother. 2007 Dec;61(10):640-58.

Jones ME, Cockburn A, Hamede R, Hawkins C, Hesterman H, Lachish S, et al. Life-history change in disease-ravaged Tasmanian devil populations. Proceedings of the National Academy of Sciences. 2008 Jul 16;105(29):10023-7.

Jones ME, Hamede RK, Storfer A, Hohenlohe PA, Murchison E, McCallum HI. Tasmanian devil facial tumour disease: ecology and evolution of an uncommon enemy. 2019.

Jones ME, Paetkau D, Geffen E, Moritz C. Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. Molecular Ecology. 2004 Jul 9;13(8):2197-209.

Jones PA, Baylin SB. The Epigenomics of Cancer. Cell. 2007 Feb;128(4):683-92.

Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. Nature Publishing Group; 2015 Mar 18;115:63-72.

Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2nd ed. 2016 Oct 23;9(10):1205-18.

King DP, Hure MC, Goldstein T, Aldridge BM, Gulland FMD, Saliki JT, et al. Otarine herpesvirus-1: a novel gammaherpesvirus associated with urogenital carcinoma in California sea lions (*Zalophus californianus*). Vet. Microbiol. 2002 Apr 22;86(1-2):131-7.

Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature. Nature Publishing Group; 1997 Apr 24;386(6627):761-3.

Kirkpatrick M, Jarne P. The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. The American Naturalist. The University of Chicago Press; 2015 Jul 17;155(2):154-67.

Kumar S, Dudley JT, Filipski A, Liu L. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. Trends in Genetics. 2011 Sep;27(9):377-86.

Lachish S, McCallum H, Jones ME. Demography, disease and the devil: life-history changes in a disease-affected population of Tasmanian devils (*Sarcophilus harrisii*). J Anim Ecology. Blackwell Publishing Ltd; 2009 Mar 1;78(2):427-36.

Lazenby BT, Tobler MW, Brown WE, Hawkins CE, Hocking GJ, Hume F, et al. Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. Stanley M, editor. Journal of Applied Ecology. 6 ed. John Wiley & Sons, Ltd; 2018 Feb 5;55(3):1368-79.

Leroi AM, Koufopanou V, Burt A. Cancer selection. Nat Rev Cancer. Nature Publishing Group; 2003 Mar 1;3(3):226-31.

Letcher RJ, Norstrom RJ, Muir D, Sandau K. Methylsulfone PCB and DDE metabolites in beluga whale (*Delphinapterus leucas*) from the St. Lawrence River estuary and western Hudson Bay. Environ Toxicol Chem; 2000.

Lipscomb TP, Scott DP, Garber RL, Krafft AE, Tsai MM, Lichy JH, et al. Common Metastatic Carcinoma of California Sea Lions (*Zalophus californianus*): Evidence of Genital Origin and Association with Novel Gammaherpesvirus. Veterinary Pathology. 18 ed. SAGE PublicationsSage CA: Los Angeles, CA; 2016 Jun 26;37(6):609-17.

Loh R, Hayes D, Mahjoor A, O'Hara A, Pyecroft S, Raidal S. The Immunohistochemical Characterization of Devil Facial Tumor Disease (DFTD) in the Tasmanian Devil (*Sarcophilus harrisii*). Veterinary Pathology. SAGE PublicationsSage CA: Los Angeles, CA; 2016 Jun 26;43(6):896-903.

MacRae SL, Zhang Q, Lemetre C, Seim I, Calder RB, Hoeijmakers J, et al. Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. Aging Cell. 2015 Jan 28;14(2):288-91.

Madsen T, Arnal A, Vittecoq M, Bernex F. Cancer Prevalence and Etiology in Wild and Captive Animals. Ecology and Evolution of Cancer. 2017.

Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet. Nature Publishing Group; 2006 Apr 1;38(4):468-73.

Mapara MY, Sykes M. Tolerance and Cancer: Mechanisms of Tumor Evasion and Strategies for Breaking Tolerance. Journal of Clinical Oncology. American Society of Clinical Oncology; 2016 Sep 21;22(6):1136-51.

Margres MJ, Jones ME, Epstein B, Kerlin DH, Comte S, Fox S, et al. Large-effect loci affect survival in Tasmanian devils (Sarcophilus harrisii) infected with a transmissible cancer. Molecular Ecology. John Wiley & Sons, Ltd (10.1111); 2018a Nov 1;27(21):4189-99.

Margres MJ, Ruiz-Aravena M, Hamede R, Jones ME, Lawrance MF, Hendricks SA, et al. The Genomic Basis of Tumor Regression in Tasmanian Devils (Sarcophilus harrisii). Genome Biol Evol. 2018b Nov 1;10(11):3012-25.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences. 2015 Dec 22;:201512501-76.

Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. American Association for the Advancement of Science; 2015 Sep 25;349(6255):1483-9.

Martineau D, Beland P, Desjardins C, Lagace A. Levels of organochlorine chemicals in tissues of beluga whales (*Delphinapterus leucas*) from the St. Lawrence Estuary, Québec, Canada. Arch. Environ. Contam. Toxicol. 2nd ed. Springer-Verlag; 1987;16(2):137-47.

Martineau D, Lagace A, Beland P, Higgins R, Armstrong D, Shugart LR. Pathology of stranded beluga whales (*Delphinapterus leucas*) from the St. Lawrence Estuary, Québec, Canada. Journal of Comparative Pathology. 1988 Apr;98(3):287-310.

Martineau D, Lemberger K, Dallaire A, Labelle P, Lipscomb TP, Michel P, et al. Cancer in wildlife, a case study: beluga from the St. Lawrence estuary, Québec, Canada. Environ. Health Perspect. National Institute of Environmental Health Science; 2002 Mar;110(3):285-92.

McAloose D, Newton AL. Wildlife cancer: a conservation perspective. Nature Publishing Group. 2009 Jul;9(7):517-26.

McCallum H, Jones ME. Infectious cancers in wildlife. In: Aguire AA, Ostfeld RS, Dasak P, editors. New Directions in Conservation Medicine. New York; 2012.

McCallum H, Jones ME, Hawkins C, Hamede R, Lachish S, Sinn D, et al. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology. 2009 Dec 8;90(12):3379-92.

McCallum H, Tompkins DM, Jones ME, Lachish S, Marvanek S, Lazenby B, et al. Distribution and Impacts of Tasmanian Devil Facial Tumor Disease. EcoHealth. 2007 Sep 11;4(3):318-25.

Merlo LMF, Maley CC. The role of genetic diversity in cancer. The Journal of Clinical Investigation. American Society for Clinical Investigation; 2010 Feb 1;120(2):401-3.

Merlo LMF, Shah NA, Li X, Blount PL, Vaughan TL, Reid BJ, et al. A Comprehensive Survey of Clonal Diversity Measures in Barrett's Esophagus as Biomarkers of Progression to Esophageal Adenocarcinoma. Cancer Prev Res. American Association for Cancer Research; 2010 Nov 1;3(11):1388-97.

Metzger MJ, Goff SP. A Sixth Modality of Infectious Disease: Contagious Cancer from Devils to Clams and Beyond. Sheppard DC, editor. PLoS Pathog. Public Library of Science; 2016 Oct;12(10):e1005904.

Metzger MJ, Reinisch C, Sherry J, Goff SP. Horizontal Transmission of Clonal Cancer Cells Causes Leukemia in Soft-Shell Clams. Cell. 2015 Apr;161(2):255-63.

Michod RE. Darwinian dynamics: evolutionary transitions in fitness and individuality. 2000.

Miller W, Hayes VM, Ratan A. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). 2011. pp. 12348-53.

Morris KM, Wright B, Grueber CE, Hogg C, Belov K. Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*). Molecular Ecology. 2015 Jul 17;24(15):3860-72.

Mousseau TA, MÃ¸ller AP. Landscape portrait: A look at the impacts of radioactive contaminants on Chernobyl's wildlife. Bulletin of the Atomic Scientists. 2nd ed. Routledge; 2015 Nov 27;67(2):38-46.

Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, et al. Genome Sequencing and Analysis of the Tasmanian Devil and Its Transmissible Cancer. Cell. Elsevier Inc; 2012 Feb;148(4):780-91.

Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, Rebbeck CA, et al. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science. American Association for the Advancement of Science; 2010 Jan 1;327(5961):84-7.

Murchison EP, Wedge DC, Alexandrov LB, Fu B, Martincorena I, Ning Z, et al. Transmissible Dog Cancer Genome Reveals the Origin and History of an Ancient Cell Lineage. Science. American Association for the Advancement of Science; 2014 Jan 24;343(6169):437-40.

Murgia C, Pritchard JK, Kim SY, Fassati A, Weiss RA. Clonal Origin and Evolution of a Transmissible Cancer. Cell. 2006 Aug;126(3):477-87.

Murshid A, Gong J, Calderwood SK. Heat-shock proteins in cancer vaccines: agents of antigen cross-presentation. Expert Rev Vaccines. 2008 Sep;7(7):1019-30.

Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. Oncogene. 2004 Aug 23;23(38):6445-70.

Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proceedings of the National Academy of Sciences. National Academy of Sciences; 1997 Jul 22;94(15):7799-806.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell. 2012 May;149(5):979-93.

Nordling CO. A new theory on cancer-inducing mechanism. British Journal of Cancer. Nature Publishing Group; 1953 Mar;7(1):68-72.

Olsen AB, Jones ME, Burridge CP, Murchison EP, Holland BR, Austin JJ. Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. J. Biogeogr. 2nd ed. John Wiley & Sons, Ltd (10.1111); 2018 May 1;45(5):963-76.

Ostrander EA, Davis BW, Ostrander GK. Transmissible Tumors: Breaking the Cancer Paradigm. Trends in Genetics. 2016 Jan;32(1):1-15.

O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. Biological Conservation. 2006 Nov;133(1):42-51.

Park SY, Gönen M, Kim HJ, Michor F, Polyak K. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. The Journal of Clinical Investigation. American Society for Clinical Investigation; 2010 Feb 1;120(2):636-44.

Patchett AL, Tovar C, Corcoran LM, Lyons AB, Woods GM. The toll-like receptor ligands Hiltonol (polyICLC) and imiquimod effectively activate antigen-specific immune responses in Tasmanian devils (*Sarcophilus harrisii*). Developmental and Comparative Immunology. Elsevier Ltd; 2017 Nov 1;76:352-60.

Pearse AM, Swift K. Allograft theory: Transmission of devil facial-tumour disease. Nature. Nature Publishing Group; 2006 Feb 2;439(7076):549-9.

Perera FP. Molecular epidemiology of environmental carcinogenesis. In: Schwab M, Rabes H, Munk K, Hofschneider P, editors. Recent Results in Cancer Reseach. 1998. pp. 39-46.

Perera FP, Dickey C. Molecular Epidemiology and Occupational Health. Ann. N. Y. Acad. Sci. Blackwell Publishing Ltd; 1997 Dec 1;837(1):353-9.

Pesavento PA, Agnew D, Keel MK, Woolard KD. Cancer in wildlife: patterns of emergence. Nat Rev Cancer. Springer US; 2018 Aug 15;:1-16.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Research. 2016 Mar 1;26(3):342-50.

Pye R, Hamede R, Siddle HV, Caldwell A, Knowles GW, Swift K, et al. Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. Biology Letters. 2016 Oct 18;12(10):20160553-5.

Pye RJ, Pemberton D, Tovar C, Tubio JMC, Dun KA, Fox S, et al. A second transmissible cancer in Tasmanian devils. Proceedings of the National Academy of Sciences. 2015 Dec 28;:201519691-6.

Pyecroft SB, Pearse A-M, Loh R, Swift K, Belov K, Fox N, et al. Towards a Case Definition for Devil Facial Tumour Disease: What Is It? EcoHealth. Springer-Verlag; 2007;4(3):346-51.

Ralls K, Ballou JD, Rideout B, Frankham R. Genetic management of chondrodystrophy in California condors. 2000;3:145-53.

Reiche EMV, Nunes SOV, Morimoto HK. Stress, depression, the immune system, and cancer. Lancet Oncol. 2004 Oct;5(10):617-25.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in Island Foxes. Current Biology. Cell Press; 2018 Nov;28(21):3487-94.

Robinson JA, Vecchyo DO-D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, et al. Genomic Flatlining in the Endangered Island Fox. Current Biology. Elsevier Ltd; 2016 Apr 18;:1-25.

Romanov MN, Tuttle EM, Houck ML, Modi WS, Chemnick LG, Korody ML, et al. The value of avian genomics to the conservation of wildlife. BMC Genomics. BioMed Central; 2009 Jul 14;10 Suppl 2(Suppl 2):S10.

Russell T, Madsen T, Thomas F, Raven N, Hamede R, Ujvari B. Oncogenesis as a Selective Force: Adaptive Evolution in the Face of a Transmissible Cancer. BioEssays. John Wiley & Sons, Ltd; 2018 Mar;40(3):1700146.

Salkeld DJ. Vaccines for Conservation: Plague, Prairie Dogs & Black-Footed Ferrets as a Case Study. EcoHealth. Springer US; 2017 Sep 6;14(3):432-7.

Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, et al. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Research. Cold Spring Harbor Lab; 2012 Mar;22(3):557-67.

Sarkies P, Sale JE. Cellular epigenetic stability and cancer. Trends in Genetics. 2012 Mar;28(3):118-27.

Schiffman JD, Breen M. Comparative oncology: what dogs and other species can teach us about humans with cancer. Phil. Trans. R. Soc. B. The Royal Society; 2015 Jul 19;370(1673):20140231.

Shields PG, Harris CC. Cancer Risk and Low-Penetrance Susceptibility Genes in Gene-Environment Interactions. Journal of Clinical Oncology. 2000 Jun 11;18(11):2309-15.

Siddle HV, Kreiss A, Eldridge MDB, Noonan E, Clarke CJ, Pyecroft S, et al. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proceedings of the National Academy of Sciences. National Acad Sciences; 2007 Oct 9;104(41):16221-6.

Siddle HV, Kreiss A, Tovar C. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. 2013. pp. 5103-8.

Siddle HV, Marzec J, Cheng Y, Jones ME, Belov K. MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. Proceedings of the Royal Society B: Biological Sciences. 2010 May 26;277(1690):2001-6.

Smith KR, Hanson HA, Mineau GP, Buys SS. Effects of BRCA1 and BRCA2 mutations on female fertility. Proceedings of the Royal Society B: Biological Sciences. The Royal Society; 2012 Apr 7;279(1732):1389-95.

Soto AM, Sonnenschein C. Environmental causes of cancer: endocrine disruptors as carcinogens. Nat Rev Endocrinol. 2010 Jul;6(7):363-70.

Stearns SC. Life history evolution: successes, limitations, and prospects. Naturwissenschaften. Springer-Verlag; 2000;87(11):476-86.

Stephens PJ, McBride DJ, Lin M-L, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. Europe PMC Funders; 2009 Dec 24;462(7276):1005-10.

Stoler DL, Chen N, Basik M, Kahlenberg MS, Rodriguez-Bigas MA, Petrelli NJ, et al. The onset and extent of genomic instability in sporadic colorectal tumor progression. Proceedings of the National Academy of Sciences. National Acad Sciences; 1999 Dec 21;96(26):15121-6.

Storfer A. Landscape genetics of the Tasmanian devil: implications for spread of an infectious cancer. Conservation Genetics. Springer Netherlands; 2017 May 26;0(0):0-0.

Storfer A, Hohenlohe PA, Margres MJ, Patton A, Fraik AK, Lawrance M, et al. The devil is in the details: Genomics of transmissible cancers in Tasmanian devils. Spindler KR, editor. PLoS Pathog. Public Library of Science; 2018 Aug 2;14(8):e1007098.

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. Europe PMC Funders; 2009 Apr 9;458(7239):719-24.

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, et al. TP53copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife. eLife Sciences Publications Limited; 2016 Sep 19;5:1850-30.

Szathmáry E, Smith JM. The major transitions in evolution. 1995.

Tallmon D, Luikart G, Waples RS. The alluring simplicity and complex reality of genetic rescue. Trends in Ecology & Evolution. 2004 Sep;19(9):489-96.

Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW. Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss. BMC Genomics. BioMed Central; 2008 Feb 1;9(1):62.

Tian X, Doerig K, Park R, Can Ran Qin A, Hwang C, Neary A, et al. Evolution of telomere maintenance and tumour suppressor mechanisms across mammals. Phil. Trans. R. Soc. B. 2018 Jan 15;373(1741):20160443-10.

Tollis M, Hutchins ED, Kusumi K. Reptile genomes open the frontier for comparative analysis of amniote development and regeneration. Int. J. Dev. Biol. UPV/EHU Press; 2015 Jul 2;58(10-11-12):863-71.

Tollis M, Schiffman JD, Boddy AM. Evolution of cancer suppression as revealed by mammalian comparative genomics. Current Opinion in Genetics & Development. Elsevier Ltd; 2017 Feb 1;42:40-7.

Tovar C, Obendorf D, Murchison EP, Papenfuss AT, Kreiss A, Woods GM. Tumor-Specific Diagnostic Marker for Transmissible Facial Tumors of Tasmanian Devils. Veterinary Pathology. SAGE PublicationsSage CA: Los Angeles, CA; 2011 Mar 7;48(6):1195-203.

Tovar C, Patchett AL, Kim V, Wilson R, Darby J, Lyons AB, et al. Heat shock proteins expressed in the marsupial Tasmanian devil are potential antigenic candidates in a vaccine against devil facial tumour disease. Multhoff G, editor. PLoS ONE. Public Library of Science; 2018;13(4):e0196469.

Tovar C, Pye RJ, Kreiss A, Cheng Y, Brown GK, Darby J, et al. Regression of devil facial tumour disease following immunotherapy in immunised Tasmanian devils. Nature Publishing Group. Nature Publishing Group; 2017 Feb 21;:1-14.

Ujvari B, Roche B, Thomas F, editors. Ecology and Evolution of Cancer. 2017.

Vazquez AI, de los Campos G, Klimentidis YC, Rosa GJM, Gianola D, Yi N, et al. A comprehensive genetic approach for improving prediction of skin cancer risk in humans. Genetics. Genetics Society of America; 2012 Dec;192(4):1493-502.

Vickers TW, Clifford DL, Garcelon DK, King JL, Duncan CL, Gaffney PM, et al. Pathology and Epidemiology of Ceruminous Gland Tumors among Endangered Santa Catalina Island Foxes (Urocyon littoralis catalinae) in the Channel Islands, USA. PLoS ONE. 2015 Nov 30;10(11):e0143211-8.

Vittecoq M, Roche B, Daoust SP, Ducasse H, MissÃ© D, Abadie J, et al. Cancer: a missing link in ecosystem functioning? Trends in Ecology & Evolution. 2013 Nov;28(11):628-35.

Wagemann R, Stewart R, Beland P, Desjardins C. Heavy metals and selenium in tissues of beluga whales, Delphinapterus leucas, from the Canadian Arctic and the St. Lawrence Estuary. Canadian bulletin of fisheries; 1990.

Walters JR, Derrickson SR, Michael Fry D, Haig SM, Marzluff JM, Wunderle JM Jr. Status of the California Condor (*Gymnogyps californianus*) and Efforts to Achieve Its Recovery. The Auk. 2010 Oct;127(4):969-1001.

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, et al. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evol Appl. Blackwell Publishing Ltd; 2011 Nov 1;4(6):709-25.

Wells K, Hamede RK, Kerlin DH, Storfer A, Hohenlohe PA, Jones ME, et al. Infection of the fittest: devil facial tumour disease has greatest effect on individuals with highest reproductive output. Ezenwa V, editor. Ecol Lett. 2017 May 10;20(6):770-8.

Whilde J, Martindale MQ, Duffy DJ. Precision wildlife medicine: applications of the human-centred precision medicine revolution to species conservation. Glob Change Biol. 2017 May;23(5):1792-805.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA. Genetic rescue to the rescue. Trends in Ecology & Evolution. 2015 Jan;30(1):42-9.

Williams TD, Pulley LT. Leiomyomas in two sea otters, *Enhydra lutris*. J. Wildl. Dis. 1981 Jul;17(3):401-4.

Wright B, Morris K, Grueber CE, Willet CE, Gooley R, Hogg CJ, et al. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. BMC Genomics. 2015;16(1):791-11.

Wright B, Willet CE, Hamede R, Jones ME, Belov K, Wade CM. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. Scientific Reports. Springer US; 2017 Mar 18;:1-6.

Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. American Association for the Advancement of Science; 2015 Apr 10;348(6231):242-5.

Yablokov AV. 10. Chernobyl's radioactive impact on fauna. Ann. N. Y. Acad. Sci. Blackwell Publishing Inc; 2009 Nov;1181(1):255-80.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. Nature Research; 2013 Oct 1;45(10):1134-40.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014 Dec 12;346(6215):1311-20.

CHAPTER 5

Population bottleneck and the rise of a maladaptive, polygenic trait: cancer susceptibility in the threatened Catalina Island Fox (*Urocyon litteralis catalinae*)

5.1 Abstract

Small effective population sizes due to founder effects and bottlenecks have been found to increase the probability of extinction, by increasing frequency of deleterious alleles and reducing fitness. Although it is unknown the extent to which cancer plays a role in the reduced fitness of genetically depauperate wildlife populations, Santa Catalina island (SCA) foxes (*Urocyon littoralis catalinae*) infections and may be due to the accumulation of deleterious alleles. Here, we present one of the first genomic assessments of pre-disposition to cancer in a wild population. We used whole-genome sequencing of case and control individuals to test for signatures of selection at these loci, reduced nucleotide diversity in cases, and functional consequences of outlier loci. We identified a relatively large number of significant candidate loci showing some association with cancer, suggesting that cancer susceptibility is a polygenic trait, with implications for conservation of this taxon. Due to the efforts of a recovery program and weak selection caused by the disease, the population size has increased, which may allow selection to be more effective in removing these slightly deleterious alleles. Long-term monitoring of the disease alleles as well as overall genetic diversity will provide a crucial index for the long-term persistence of this threatened population.

5.2 Introduction

The "extinction vortex" hypothesis predicts that small, geographically restricted populations will experience loss of genetic diversity and an increase in deleterious variation, which could lead to higher disease susceptibility and further population declines (Gilpin & Soulé

1986). As threats to wild populations increase, understanding the genetic consequences affecting small, isolated populations can inform conservation practices to mitigate future loss of genetic variation and maintain biodiversity to prevent the extinction vortex. When populations are small, both inbreeding and stochastic processes, such as genetic drift, can increase the probability of extinction by increasing the frequency of homozygous loci with deleterious alleles and reducing fitness (Gomulkiewicz & Holt 1995; Frankham 2005a; b; O'Grady et al. 2006). As a population loses variation via drift and inbreeding, genetic load increases. The genetic load is defined as the difference between the mean fitness of a population and the fitness of an optimal genotype that does not carry deleterious mutations (Glémin et al. 2003; Kirkpatrick & Jarne 2015). Small effective population sizes due to bottlenecks, founder effects, and domestication have been found to increase proportion of genetic load in Mountain gorillas (*Gorilla beringei beringei*), Channel Island foxes (*Urocyon littoralis*), and domestic dogs (*Canis lupus familiaris*) (Xue et al. 2015; Marsden et al. 2015; Robinson et al. 2016). This can be accounted for by numerous weakly deleterious mutations (Fu et al. 2014; Henn et al. 2015), which are particularly difficult to eliminate from bottlenecked populations. It is unknown the extent to which cancer plays a role in the reduced fitness of genetically depauperate populations across wildlife taxa (Ujvari et al. 2018).

Over the past few decades wildlife health monitoring has increased, and we are now gaining an improved, and occasionally alarming, perspective about the presence and impact of cancer in wildlife, particularly among some endangered species (McAloose & Newton 2009). Population genomic tools are increasingly being used to investigate wildlife cancers, addressing issues such as the genetic variation for susceptibility within populations, comparative genomics of tumor suppressor genes, and evolutionary response to cancers (reviewed in Hendricks et al. in review). Studies of cancer in wildlife species present unique challenges, but research is beginning to uncover examples of cancer and its impact on wildlife populations (reviewed in (McAloose & Newton 2009; Madsen et al. 2017; Pesavento et al. 2018). Cancer can affect wild populations by reducing reproductive success, altering population dynamics,

or directly or indirectly leading to population declines (McAloose & Newton 2009; Pesavento et al. 2018). In addition to presenting a major conservation concern, naturally occurring cancers in wildlife species may provide new biological models for understanding the often complex causes of cancer, with the potential for biomedical benefits (Hendricks et al. in review). Known causes of cancer in wildlife include environmental carcinogens (e.g. polycyclic aromatic hydrocarbons), viruses and other pathogens, direct transmission of tumor cells, and hereditary factors (Hendricks et al. in review, Pesavento et al. 2018).

Familial and hereditary cancer syndromes are relatively common in humans, yet understanding of factors, such as heritability, that drive tumorigenesis across a wide range of wildlife species is still lacking, but necessary. Familial cancer syndromes are likely to be due to a combination of shared environmental factors and inherited genetic variants (Nagy et al. 2004). These variants can include sometimes rare, large effect (i.e. high penetrance) mutations or many low-penetrant variants acting together to alter disease susceptibility. A few wildlife genetic studies have begun to identify polymorphisms possibly associated with cancer susceptibility (Browning et al. 2014; Margres et al. 2018). For example, in a case-control study of California sea lions, urogenital carcinoma was significantly associated with homozygosity of a microsatellite locus within an intron of the heparanase 2 gene (HPSE2; Browning et al. 2014), which has been implicated in several human carcinomas. Studying the genetics of cancer in wild systems is especially challenging given difficulties in sampling and ethical, logistical and legal limits on experimentation.

The island fox (*Urocyon littoralis*), limited to the Channel Islands of California, provides a natural experiment in which these small, isolated populations may be entering into the "extinction vortex". This natural experiment may help elucidate deleterious mutations that underlie the relationship between immune response, inflammation, and tumor development. Populations of the island fox, endemic to individual islands, have recently undergone severe genetic bottlenecks. These bottlenecks took place in the late 1990's due to golden eagle predation on three northern islands (Santa Rosa, Santa Cruz, and San Miguel) and to a

canine distemper epidemic on Santa Catalina Island (SCA) (Coonan et al. 2010; US Fish and Wildlife Service 2016). Subsequently, these populations have rebounded due to human management and have the fastest recovery of any mammal under the Endangered Species Act to date (Coonan et al. 2010; US Fish and Wildlife Service 2016). These six island populations have adapted to different environmental conditions, creating a set of replicate inbred populations. Island foxes are genetically less variable at non-coding loci than coding regions, which can be attributed to the process of island colonization with subsequent inbreeding (Wayne et al. 1991; Aguilar et al. 2004; Robinson et al. 2016). Inbreeding can increase the presence of recessive deleterious mutations in populations resulting in potential fitness reductions (Charlesworth & Charlesworth 1999), and increased disease susceptibility (Spielman et al. 2004). However, high levels of genetic variability were discovered for genes of the major histocompatibility complex (MHC) that are involved in innate immunity and whose variability is likely maintained by balancing selection (Aguilar et al. 2004). Additionally, genomic evidence suggests that island fox populations do not suffer from inbreeding depression (Robinson et al. 2016; 2018) and exhibit adaptive differentiation among islands (Funk et al. 2016), despite very low levels of overall genetic diversity.

Nonetheless, one island fox population is strongly affected by disease, which may be associated with accumulation of deleterious alleles. SCA foxes have a high prevalence (≈50 percent in individuals 4 years or older sampled in 2007-2008) of ear canal (ceruminous gland) carcinoma and adenoma (collectively tumors) that appear to be associated with inflammation from chronic ear mite (*Otodectes*) infections (Vickers et al. 2015). Ceruminous gland tumors have not been documented on two geographically close islands (San Clemente Island; SCL and San Nicolas Island; SNI) despite similar levels of chronic mite infection, and they are also unobserved in the three island fox populations that have no ear mites (Santa Cruz, Santa Rosa, San Miguel). This variation in tumor prevalence allows for the unique opportunity to investigate the role of genetics in cancer susceptibility in a set of replicate natural populations.

Here, we test the hypothesis that the remarkably high incidence of ceruminous gland

tumors in SCA island foxes is the result of genetic variants for cancer susceptibility, which have increased in frequency due to isolation, drift during bottleneck, and small population size. To identify and genotype these loci, we used whole-genome sequencing of case and control individuals from Santa Catalina Island, and control individuals from a second island (SCLE) to test for signatures of selection at these loci, reduced nucleotide diversity in cases, and functional consequences of outlier SNPs.

5.3 Materials and Methods

5.3.1 Library construction and genotyping

We sequenced the genomes of 45 island foxes from Santa Catalina Island and from San Clemente Island. We obtained existing tissue and blood samples collected by one of the authors, W. Vickers (University of California, Davis; Vickers et al. 2015), and collaborator W. Andelt (Colorado State University). High-quality DNA was extracted from these samples using a standard commercial kit protocols (Qiagen DNA QiaAmp minikit). Library preparation was completed at the Vincent J. Coates Genomic Sequencing Laboratory at the University of California, Berkeley and libraries were run on 16 lanes with 150bp paired-end sequencing on an Illumina HiSeq4000.

Reads were trimmed and filtered for quality using Picard 2.9.0. Cleaned reads were aligned to the domestic dog genome (*Canis lupus domesticus*; canFam3.1) using Burrows-Wheeler Aligner version 0.7.12 (MEM; Li et al. 2009). Duplicate reads were removed using Picard. Aligned reads that were properly paired, mapped uniquely, and had high quality (Phred score >= 30) were used as input for base quality score recalibration (Genome Analysis Toolkit v3.7 (GATK; McKenna et al. 2010; Danecek et al. 2011). To obtain a set of "known variants" for recalibration, raw variant genotypes were called using default parameters and a minimum base quality Phred score of 20 with GATK UnifiedGenotyper, which was followed by GATK BaseRecalibrator and GATK PrintReads. This process was repeated three

times to reach convergence between reported and empirical quality scores, which we assessed with Qualimap (http://qualimap.bioinfo.cipf.es/; García-Alcalde et al. 2012; Okonechnikov et al. 2016). Variant calling for the recalibrated BAM files was performed with GATK HaplotypeCaller (Phred score >= 20) and the whole cohort was genotyped using GATK GenotypeGVCFs. We used GATK to remove alternate alleles not present in any genotypes.

The final dataset included individuals with mean coverage greater than 7x (n = 32) and consisted of individuals that qualified as cases (individuals with carcinoma; n = 12), control individuals from SCA (individuals >= 6 years of age that did not develop tumors (see Vickers et al. 2015) for aging methods; n = 11), and control individuals from SCLE (n = 9). Two of the SCLE individuals were previously published data (Robinson et al. 2016; 2018). SCLE individuals act as an alternative control group in that these fox populations are not found to have any incidence of ceruminous gland tumors.

We applied several variant filters to ensure high quality of the data. The program Qualimap (GarcÃa-Alcalde et al. 2012; Okonechnikov et al. 2016) was used to assess the quality of filtered and aligned reads. Chromosome X and the mitochondrial genome were excluded, as were all indels. Variants with a depth of coverage less than 6, more than 25 (representing possible paralogy or copy number variation), and minimum genotype quality less than 20 were excluded from the final dataset. Sites called in fewer than 95% of individuals were subsequently removed from further analysis. Singletons were then removed from the dataset as well as alleles with minor allele frequency less than 0.01. The number of segregating sites and mean coverage per individual was calculated using VCFtools. Datasets were LD-pruned using PLINK v1.9 (-indep-pairwise 50 5 0.2; (Purcell et al. 2007). We used the programs KING v1.4 (Manichaikul et al. 2010) and PRIMUS v0.5 (Staples et al. 2013) to calculate relatedness from the LD-pruned datasets. Individuals were assessed for a pairwise identity-by-state greater than 0.375 (equivalent to 3/4 siblings or sibling-cousins). No individuals were removed from subsequent analyses due to relatedness. Functional regions were annotated using snpEff v4.3 (Cingolani et al. 2012) based on the domestic dog genome

(canFam3.1) in Ensembl v91. We annotated variants within coding regions with respect to their effect on the amino acid sequence and polarized alleles as ancestral or derived using the domestic dog as an outgroup.

5.3.2 Demographic estimates

To assess genetic clustering, we applied two methods to the dataset of putatively neutral, LD-pruned loci. The dataset was pruned for strong LD (removing any SNPs having a multiple r^2>0.90 with all other SNPs in a 50 SNP window) using PLINK, which reduced false autozygosity calls by removing redundant markers in SNP-dense regions and making SNP coverage more uniform. Putatively neutral regions were identified using methods described by Freedman et al. (2014). First, principal components analysis (PCA) was performed using PLINK v1.9. Second, we inferred a phylogenetic network with the neighbor-net method in SplitsTree4 v4.13.1(Huson & Bryant 2006), using uncorrected P distance.

We estimated historical demography using SMC++ (Terhorst et al. 2017), which jointly estimates population histories and divergence times without phased data. We assumed a mutation rate of 2.0 x 10^{-8}/site/generation (Marsden et al. 2015; Robinson et al. 2016). We also tested if the mutation rate affected divergence times by using a mutation rate of 1.0 x 10^{-8}/site/generation (Freedman et al. 2014). Island foxes breed by then end of their first year and only a small percentage of females 6 years or older are in breeding condition (Laughrin 1977), therefore, we assumed generation time of 2 years to convert the coalescent scaling to calendar time.

The R programming environment (R Core Team 2015) was used for data manipulation, summary statistics, and plotting of runs of homozygosity (ROH) using the library "detectRUNS". ROHs were detected using a sliding-window based method using a 50-SNP window to scan across the genome of each individual using the LD pruned dataset. The proportion of overlapping homozygous windows to call a ROH was 0.05 with a maximum of 2 missing and one heterozygous SNP allowed within each ROH. To minimize the number of false positive

ROH, the minimum number of SNPs to call a ROH was set to 2 and the minimum length of a ROH was set to 100 Kb. A minimum density of one SNP per 50 Kb and a maximum of 100 Kb gap between consecutive SNPs. The inbreeding coefficient, F, was estimated for each individual using two methods, F_{ROH} using the R library "detectRUNS".

5.3.3 Identification of putative causal polymorphisms

Genome-wide Tests of Neutrality

We calculated several population genetic statistics to test for deviation from neutral expectation. Tajima's D and nucleotide diversity per population as well as Weir's F_{ST} for each comparison pair (case/control, control/SCLE, case/SCLE) were calculated using a sliding window approach in VCFtools. Z-score transformations were performed using R (R Core Team 2015) and using SciPy library. We calculated each statistic with a window size of 100 Kb, step size of 10 Kb and at a finer scale assessment using 50 Kb windows and 1 Kb steps.

The cross-population composite likelihood ratio test (XP-CLR; Chen et al. 2010) was also employed to compare allele frequency differentiation between comparison pairs to detect selective sweeps. XP-CLR scores were estimated using the following parameters: -w1 0.0005 600 50000 -p1 0.95. A set of grid points as the putative selected allele positions were positioned along each chromosome with a spacing of 50 Kb. The sliding window around each grid point was set to 0.05 cM with a maximum number of SNPs within each window set to 600. The correlation level (high LD) from which the SNPs contributed to XP-CLR scores was down-weighted to 0.95. Third, XP-CLR scores were normalized with z-scores using a custom python script using the SciPy library.

Δ π outliers

We calculated the difference in π (50 Kb windows) between cases and SCA controls using R. The top 0.1% outlier loci for Δ π between cases and SCA controls were selected across the genome. We expect that cases would have lower π values than the SCA controls in genomic regions in LD with haplotypes associated with cancer susceptibility. Therefore, we removed any 50 Kb windows where π values were lower in controls. The genes within or closest to these 50 Kb windows were used for gene ontology (GO) enrichment analysis using g:profiler (Reimand et al. 2016) with the domestic dog annotation. The gene list was tested for significant enrichment of GO terms (FDR<0.05) while correcting for multiple testing and taking into account the non-independence of GO terms.

We calculated the difference in π (50 Kb windows) between cases and both control populations (SCA and SCLE) using R and selected the top 0.1% outlier loci for Δ π between cases and both control populations (SCA and SCLE). The intersection of the top 0.1% for case/SCA control and case/both controls (resulting in 8 outlier loci) were assessed for allele frequencies at each SNP per group. The difference in allele frequencies per group was then calculated.

Intersection of F_{ST}, Δ π, and Δ Tajima's D as outliers

We calculated the difference in F_{ST} (50 Kb windows) between cases/SCA controls and cases/both controls (SCA and SCLE) using R. Similarly, we calculated the difference in Tajima's D for both comparisons. Outlier peaks for each comparison for each test were selected if they were above the ninety-ninth percentile of normalized values. The intersection of the outlier peaks for F_{ST}, Δ Tajima's D, and Δ π were chosen as possible candidate regions. The resulting candidate regions were examined further for differentiation by assessing allelic F_{ST} and difference in allele frequency calculated in VCFtools. SNPs within these regions with an allele frequency difference greater than 0.45 (n = 38) were selected. The closest gene

to these SNPs were used as input for gene ontology (GO) enrichment analysis using g:profiler (Reimand et al. 2016). Additionally, we assessed if these mutations were within transcription factor binding sites (TFBS) following the methods of Freedman et al. (2014; Section S7.3). Briefly, these regions were found within the promoter regions using the profiles in the JASPAR PHYLOFACTS database (http://jaspar.genereg.net/collection/phylofacts/), which contains count matrices of conserved motifs in human, mouse, rat and dog, originally identified by (Xie et al. 2005). After converting these motifs to probability weight matrices, the motif finding program FIMO (Grant et al. 2011) was used to find matching occurrences in the promoter regions of the dog genome.

ROH outliers

After we detected ROH using a sliding-window based method using a 50-SNP window (see above), we used "detectRuns" to calculate the proportion of times (individuals per population) each SNP falls within a ROH and plotted each proportion against the SNP position along the dog genome. We defined candidate ROHs as SNPs that were within ROHs in more than 70% of cases and less than 70% in the two control groups.

5.4 Results

5.4.1 Sequencing and genotyping

We obtained high-quality sequence reads with a per-individual average unfiltered yield of 46,553 ± 11,194 Mb (Supporting Information Table C.1). After processing and removing low-quality reads and PCR duplicates, the mean sequence depth was 11.83 ± 2.36 with an average of 87% of the genome covered by at least 6 reads (Supporting Information Table C.1). Mean depth per population was 12.72 for cases, 11.81 for SCA controls, and 10.67 for SCLE controls. After filtering genotypes, we classified data into three sets consisting of all autosome variant loci (3,945,582 SNPs), variant loci within genic regions (1,525,612

SNPs), and variant loci (24,331 SNPs). The transition to transversion ratio for all autosomes was 2.34, which is similar to previously reported values in wolves (Freedman et al. 2014; Zhang et al. 2014; Schweizer et al. 2016). After LD-pruning, there were 35,406 variable positions, of which 6,600 were within neutral regions. No individuals were removed due to high relatedness (parent-offspring or full siblings).

5.4.2 Demographic estimates

To assess LD patterns, we estimated the physical distance at which the pair-wise genotypic association (r^2) across all autosomal SNPs decays below a threshold of 0.2. We found that populations of SCA (n = 23) and SCLE (n = 9) had moderate levels of LD ($r^2_{0.2}$ = ≈40 Kb; Supporting Information Figure C.1). These LD levels are lower than that of other small, declining populations such as the Tasmanian Devil (*Sarcophilus harrisii*, ≈200 Kb; (Epstein et al. 2016), and Iberian Lynx (*Lynx pardinus*, ≈185 Kb for Andúljar population; Abascal et al. 2016), but $r^2_{0.2}$ for SCA and SCLE foxes are closer to estimates of the recently expanded Northeastern coyote population (*Canis latrans*; ≈80 Kb; (vonHoldt et al. 2011). LD estimates may be sensitive to sample size (Teare et al. 2002; England et al. 2006) and may explain inconsistencies at longer genetic distances (275 Kb). Additionally, any structural differences between the reference domestic dog genome and the island fox genome would result in an underestimate of the extent of LD.

To assess the genetic partitions based on our neutral SNP dataset, we used complementary analyses of genetic clustering patterns. The first two component axes from PCA show two distinct clusters that correspond with sampling location (Figure 5.1A). On PC1, the SCLE control individuals cluster separately from the individuals from SCA. Cases and SCA controls do not cluster separately based on any principal component (Supporting Information Figure C.2). Neighbor-net analysis shows a similar observable pattern as PC1 vs PC2 (Figure 5.1B). Geographic clustering was not present between cases and SCA controls and therefore did not have to be accounted for when identifying putative causal loci (below).

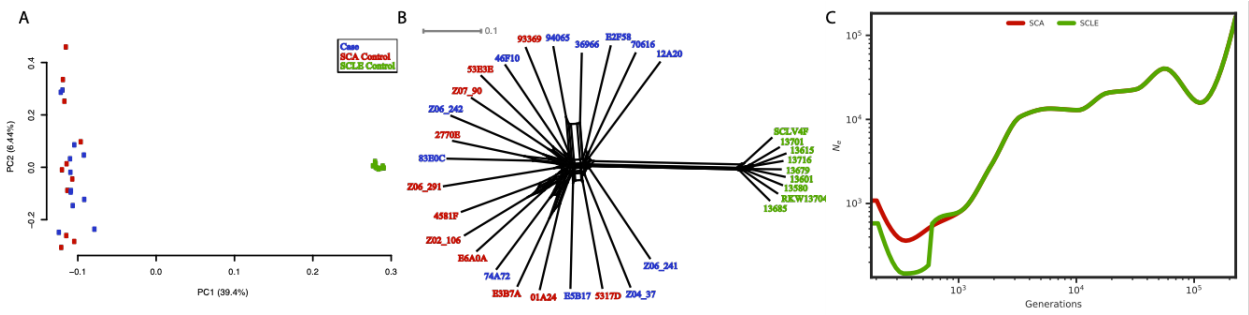

Figure 5.1: Patterns of genetic differentiation and divergence between populations for the 6.6K SNP data set using (A) PCA, and (B) Neighbor-net analysis. Case individuals are shown in blue, SCA controls in red, and SCLE controls in green. (C) Effective population size (N_e) calculated from SNP genotypes within neutral regions. A 2 year generation time and a mutation rate of 2.0 x 10^{-8}/site/generation were assumed (Marsden et al., 2015; Robinson et al., 2016). SCA individuals are shown in red and SCLE individuals in green.

We used SMC++ to estimate historical population sizes and the divergence time of these two populations (Figure 5.1C). The SCA and SCLE populations follow nearly identical reconstructions of N_e until about 1,000 generations ago, possibly representing the history of the original source population for island colonization. After this point, both populations exhibit a bottleneck, which is more severe in SCLE, followed by population recovery toward the present. The SCLE population underwent a sharp population decline 556 to 597 generations ago with a decrease from 577 individuals to 185 individuals. This decline continues with the SCLE population at its smallest 339 generations ago with 147 individuals. Although SMC++ is less accurate when estimating recent population history (Terhorst et al. 2017), the SCLE population remained roughly half that of the SCA population during population recovery starting around 300 generations ago. These results differed slightly based on mutation rate (see Supporting Information Figure C.3).

A total of 55,317 ROH were identified with an average of 1,728.7 (±787.3) ROH per individual with a minimum of 90 and maximum of 3283 ROH per individual (Supporting Information Table C.2). The average number of ROH was 1682.6 (±699.7) for cases, 1997.2 (±157.2) for SCA controls, and 1461.9 (±1231.1) for SCLE controls. There were no statistically significant differences between group means as determined by one-way ANOVA ($F(2,29) = 1.1914$, $p = 0.3182$). The average length of ROH was 0.2582 Mb for cases, 0.2527

Mb for SCA controls, and 0.2096 Mb for SCLE controls.

Using F_{ROH} as calculated using "detectRUNS" in R, the inbreeding coefficient ranged from 0.04 to 0.56 (Supporting Information Table C.2) with an average per population of f = 0.32 (±0.12; case); f = 0.37 (±0.02, SCA control); f = 0.28 (±0.19, SCLE control).

5.4.3 Identification of putative causal polymorphisms

Genome-wide Tests of Neutrality

Sliding window statistics for Weir's F_{ST}, Δ π, and Δ Tajima's D at a window size of 100 Kb (step size of 10 Kb; Figure 5.2) resulted in similar outlier peaks to those at a finer scale assessment using 50 Kb windows (step size of 1 Kb; Supporting Information Figures C.4, C.5, and C.6). We applied XP-CLR comparisons between all combinations of group comparisons (Supporting Information Figures C.7). The strongest XP-CLR score in the comparison of cases and SCA controls was on chromosome 19.

Across the genome of all 32 individuals there were 645 (0.014%) high impact variants, 17,193 (0.375%) moderate impact variants, and 22,758 (0.496%) low impact variants (Supporting Information Figure C.8).

Δ π outliers

When assessing the top 0.1% Δ π after removing loci where controls have lower π than cases, 17,667 SNPs remained with 6 of these in TFBSs. The average change in allele frequency was 0.17 with a maximum of 0.47. The mean F_{ST} estimate was 0.09 with a maximum of 0.39. After annotation and GO analysis, no high impact variants or significant GO terms were identified. With a difference in allele frequency >0.45, only 4 SNPs remained (Supporting Information Table C.3). Of these 4 SNPs, 2 were intergenic SNPs in between ENSCAFG00000033119 (LincRNA) and zinc finger protein 536 (ZNF536). ZNF536 has been found to have an important roles in neuronal differentiation (Qin et al. 2009) and

many zinc proteins have potential roles in cancer progression (Jen & Wang 2016). Another was an intronic SNP in Dipeptidyl Peptidase Like 10 (DPP10). DPP10 may play a role in disease progression of colorectal cancer and loss of DPP10 expression in primary colorectal cancer is significantly associated with poor survival outcomes (Park et al. 2013). The last of the 4 SNPs was an intronic SNP in ENSCAFG00000039265 (LincRNA). The top 0.1% of outlier loci for $\Delta\ \pi$ between cases and both control populations (SCA and SCLE) exhibited annotation impacts categorized as high (stop-gained; n =1) and moderate (missense; n = 72; Supporting Information Table C.4)). After removing SCLE individuals, the top 0.1% outlier loci for $\Delta\ \pi$ between cases and SCA controls all of which were categorized with moderate impacts (missense; n = 37; Supporting Information Table C.5). Of these 37 outlier loci, 8 outlier loci were found within both sets of outlier loci. One of these mutations was within transcription binding factor sites on chromosome 4 (position: 6705047). These 8 loci were accessed for allele frequencies per group (Supporting Information Table C.6). Several of these loci have large differences in minor allele frequency between SCA and SCLE, but there is not strong evidence for differences between cases and controls.

Intersection of F_{ST}, $\Delta\ \pi$, and Δ Tajima's D as outliers

Outlier peaks were selected if they were above the ninety-ninth percentile of normalized values from the 50 Kb window scans (F_{ST}: n = 22005; $\Delta\ \pi$: n = 9773; Δ Tajima's D: n = 440; Supporting Information Figure C.9). The intersection of these outlier peaks for F_{ST}, $\Delta\ \pi$, and Δ Tajima's D resulted in 16 candidate regions. Distribution plots of F_{ST}, $\Delta\ \pi$, Δ Tajima's D as well as XP-CLR scores suggests strong selection within these 16 candidate regions (Figure 5.3).

To further investigate these regions, we examined allelic F_{ST} and the difference in allele frequency between cases and SCA controls within $\pm$150 Kb of these regions (Supporting Information Figure C.10). No fixed SNPs were identified and highest F_{ST} was 0.53 with an average value of 0.10. The highest value of difference in allele frequency was 0.59 with an

average value of 0.22. SNPs with an allele frequency difference above 0.45 (n = 38) were annotated resulting in 17 genes (Supporting Information Table C.7). Of these 38 SNPs, two SNPs located within the intron of the Receptor-type tyrosine-protein phosphatase mu (PTPRM) gene. Another two of variants are within the TFBS of this gene. PTPRM has a role in signal transduction, growth control, and oncogenic transformation as well as cell-cell communication and adhesion (reviewed in Zhang 2003). One SNP downstream of Reticulon 1 (RTN1) was found. RTN1 is involved in neuroendocrine secretion or in membrane trafficking in neuroendocrine cells; marker for neurological diseases and cancer. We found one intergenic SNP of Keratin 8 (ENSCAFG00000002184 (KRT8)), a gene that is involved in maintaining cellular structural integrity as well as signal transduction and cellular differentiation. An intergenic SNP of Protein Kinase C Delta (PKCδ) was also highly differentiated between cases and controls as well as 3 intronic SNPs in the TFBS. PKCδ is a tumor suppressor gene that is involved in the positive regulation of cell cycle progression and can positively or negatively regulate apoptosis (Griner & Kazanietz 2007). One intergenic and 1 upstream gene variant of TFBS of Cilia And Flagella Associated Protein 45 (CFAP45; also known as Coiled-Coil Domain Containing 19 (CCDC19)) had higher differences in allele frequency and this gene is associated with lung and nasopharyngeal carcinoma (Liu et al. 2011; Wang et al. 2018). Upstream SNPs (n = 2) of SLAM Family Member 8 (ENSCAFG00000011795 (SLAMF8)) were highly differentiated between case and controls. This gene is responsible for lymphocyte activation. Three SNPs (2 intronic and 1 intergenic) where found in Fc Receptor Like 6 (FCRL6), which is involved in protein phosphatase binding. Three SNPs (1 downstream; 2 intergenic) of Dual Specificity Phosphatase 23 (ENSCAFG00000031934 (DUSP23)), which is involved in tyrosine and serine/threonine phosphatase activity (Wu et al. 2004), had a difference in allele frequency above 0.45 between cases and controls.

These seventeen genes were used for gene ontology (GO) enrichment analysis. We found 2 significantly enriched GO terms, both of which were related to the regulation of superoxide anion generation (GO:0032928 and GO:0032930). The two genes that define the significant

GO terms are Decapping MRNA 1A (DCP1A) and C-Reactive Protein (CRP). Two intronic with allele frequency <0.45 and 2 upstream variants within the TFBS of the DCP1A gene. Six variants were found downstream of the CRP gene, which is an inflammatory biomarker that recognizes and initiates the elimination of foreign pathogens and damaged cells (Gabay & Kushner 1999).

ROH outliers

Genome-wide ROHs are spread across the genome (Supporting Information Figure C.11). Figure 5.4 shows the proportion of times each SNP falls within a ROH plotted against the position along the dog genome for cases (Figure 5.4A), SCA controls (Figure 5.4B), and SCLE controls (Figure 5.4C). We identified a total of 268 ROHs with SNPs that were within ROHs in more than 70% of the individuals per population (cases: n = 18, SCA: n = 151, and SCLE: n = 99). Of the 18 ROHs in cases, 13 of these were not found in the control populations and were considered as candidate ROHs. Two of these peak regions, on chromosome 4 (Figure 5.4D) and chromosome 21 (Figure 5.4E), show differentiation between cases and control populations. For the peak in chromosome 4, 10 of 12 individuals (83.3%) of cases have the same ROH, whereas 7 of 11 of SCA control individuals (63.6%) and 4 of 9 (44.4%) of SCLE controls had the same ROH.

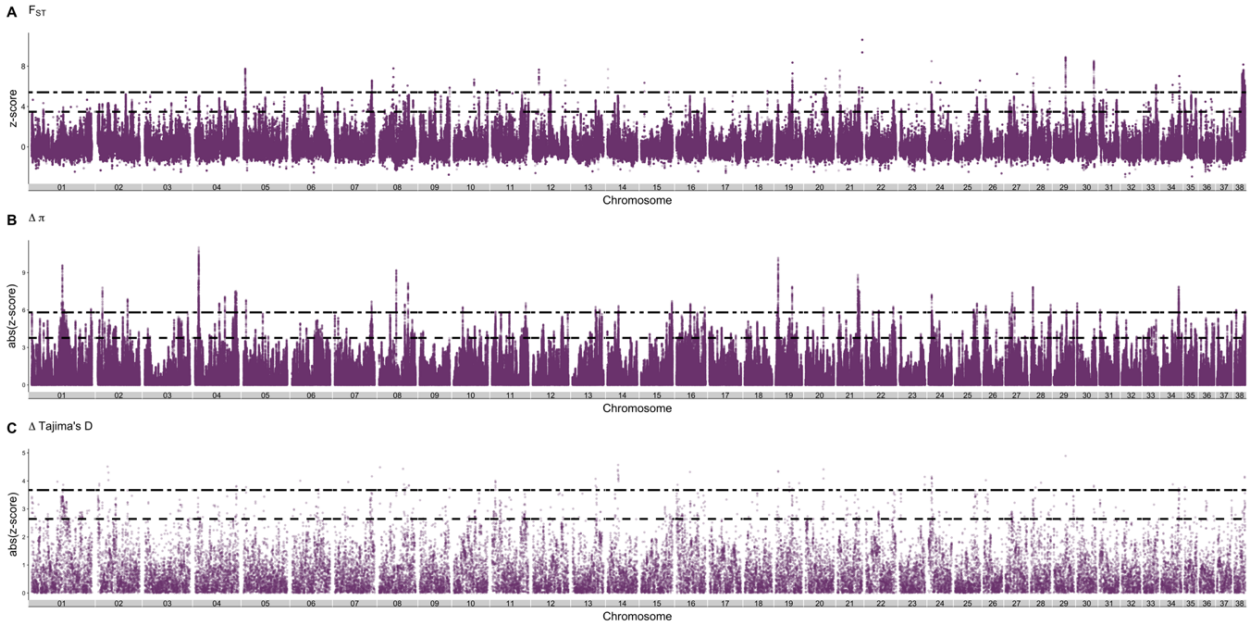

Figure 5.2: Summary statistics between case (n = 12) and SCA control (n = 11) individuals using the 35K SNP data set, which are calculated from overlapping 50 Kb windows in 1 Kb steps. (A) Manhattan plot of F_{ST} values. Dashed line indicated top 1% (z-score = 3.48). Double dashed line indicated top 0.1% (z-score of 5.41). (B) Distribution plots of $\Delta\ \pi$ values. Dashed line indicated top 1% (z-score = 3.77). Double dashed line indicated top 0.1% (z-score of 5.81) (C) Distribution plots of ΔTajimaâs D values. Dashed line indicated top 1% (z-score = 2.64). Double dashed line indicated top 0.1% (z-score of 3.67).

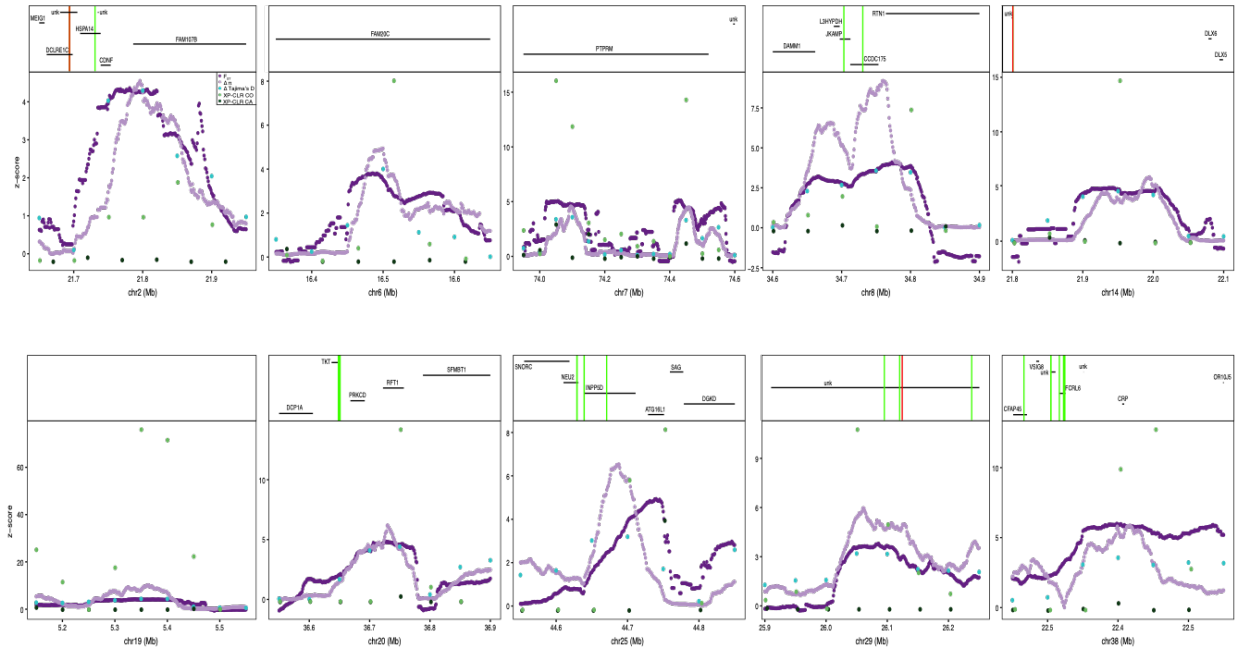

Figure 5.3: Z-transformed selection scan statistics (bottom) and gene annotations (top) plotted across the top 16 ranked candidate regions highly differentiated between case and SCA controls. F_{ST} (purple), $\Delta\ \pi$ (lavender), ΔTajimaâs D (turquoise), and XP-CLR (CO: control (light green); CA: case (dark green)). Gene annotation show genes as black horizontal bars with missense mutations as green vertical lines and stop-gained mutations as red vertical lines.

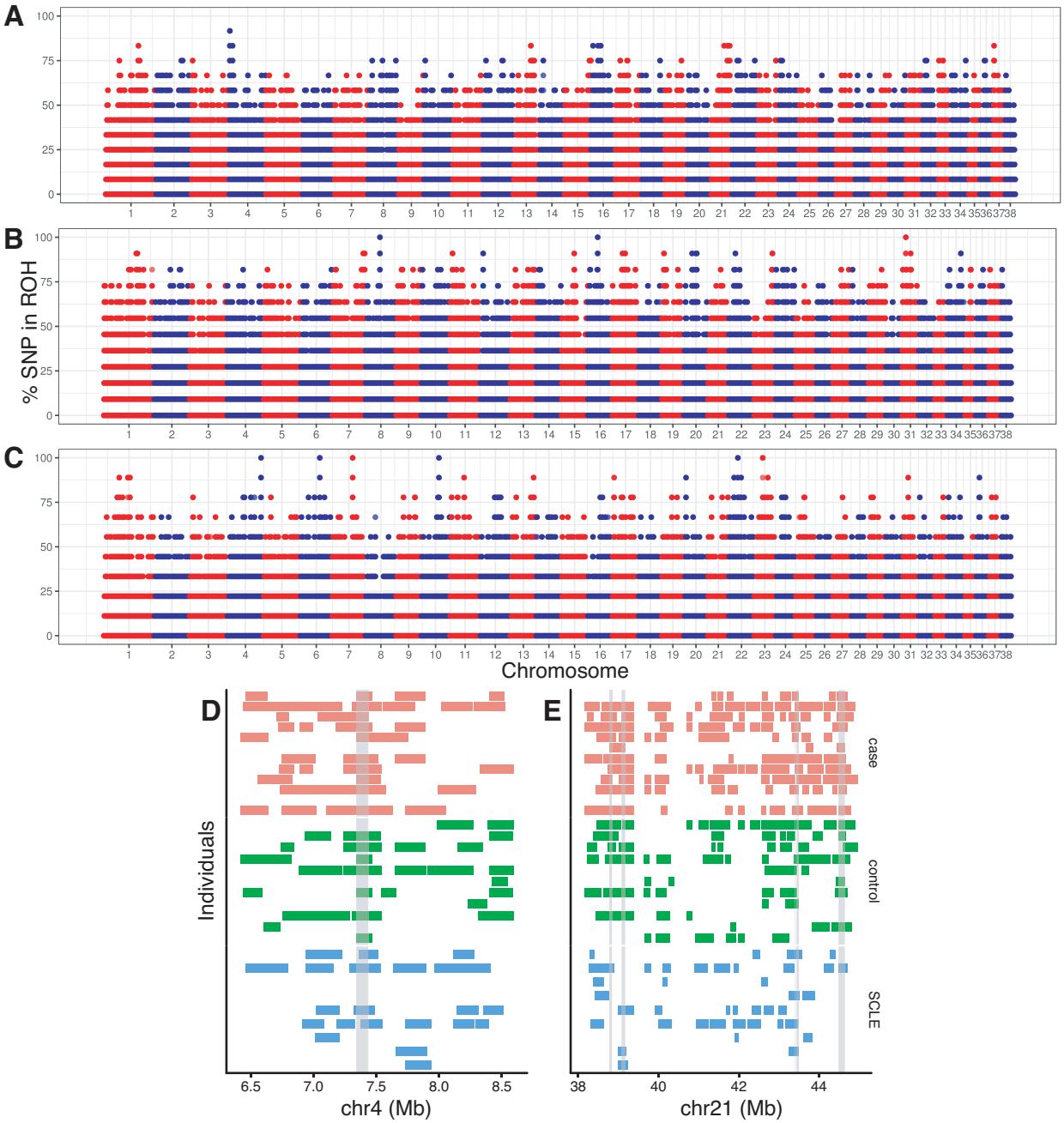

Figure 5.4: Manhattan plot of the proportion of times each SNP falls within an ROH in the (A) cases (B) SCA controls and (C) SCLE controls. Chromosomes 1-38 are arranged left to right, with alternating red and blue representing different chromosomes. (D) ROH detected in each individual on a part of chromosome 4 and (E) chromosome 21. Gray bars represent regions where 75% of cases (shown in salmon) have overlapping ROH and less than 75% of SCA controls (shown in green) and SCLE controls (shown in light blue). See supplemental information for genome-wide ROH for each individual.

5.5 Discussion

5.5.1 Demographic estimates

Our results estimate that SCLE and SCA foxes diverged roughly 2,000 years ago. However, Hofman et al. (2015) estimated that southern island lineages diverged from each other ≈5500 to 5700 years ago before present based on Bayesian phylogeny of mitogenomes and Accelerator Mass Spectrometry (AMS)^{14}C dating (Hofman et al. 2015). Other archaeological records show that foxes first appear in the on Santa Catalina Island sometime between 3,880 to 800 years B.P., which corresponds with the length of time that the Little Harbor Site was inhabited by humans (Meighan 2017). As for SCLE foxes, radiocarbon dated bones suggest that island foxes were introduced to this island by Indians sometime after 4,300 B.P. but before 3,400 B.P., also corresponding to the length of time that the island is believed to have been occupied (Collins 1991). Sluggish recovery from founder effect/bottlenecks and drift as well as inbreeding could have led to decreasing N_e after divergence from other populations. Our low estimate of 2,000 years may be due to incorrect mutation rate or may be attributed to lag in genetic divergence once the populations established on each island.

5.5.2 Candidate genes for cancer susceptibility: multiple loci with small effect

Overall, we found no evidence for a locus of major effect. In contrast, we identified a relatively large number of significant candidate loci showing some association with cancer, which is consistent with most population-attributable cancer heritability (Ponder 2001). Additionally, we found 17 genes with putative functional relationships to inflammation-induced cancer in island foxes. Our limited sample size of cases and controls reduces our ability to quantify the relative effect sizes or proportion of phenotypic variance explained by these loci. Nonetheless our results suggest that inflammation-induced cancer susceptibility

in SCA is a polygenic trait, the outcome of accumulation of multiple slightly deleterious alleles under genetic drift.

With the different island populations acting as a natural experiment, we have good evidence that susceptibility is strongly heritable; however, given that none of the SNPs within these 16 regions of high differentiation between cases and controls had high F_{ST} or difference in allele frequency, this may be an example of missing heritability (Maher 2008; Manolio et al. 2009; Eichler et al. 2010). Many genome-wide association studies (GWAS) have found that most of the hundreds of variants found in of complex diseases only explain a modest amount of the observed heritability (Manolio et al. 2009). More recent GWAS have developed methods to overcome missing heritability, but these still require thousands to tens of thousands of individuals (Yang et al. 2010; 2015). Many factors may contribute to missing heritability with inadequacies in genotyping and phenotyping being a main factor. For example, highly repetitive structural and sequence variants have remained inaccessible to large-scale genotyping using many GWAS methods. Calling large-scale structural variants was out of the scope of this study, but should be considered when assessing the susceptibility of cancer in SCA foxes. Furthermore, while using whole genome data assays every variant, regardless of allele frequency, low coverage data, such as those presented here, can result in rare SNPs being called less often and with higher error rates than common SNPs. Additionally, the rare variants responsible for complex disease are particularly difficult to identify and to associate with phenotype using statistics.

Several studies have found that the wide interindividual differences in the sensitivity to cancer-inducing or cancer-promoting compounds may be due to allelic differences in low-penetrance genes (Smith et al 1995; Caporaso et al. 1991). The increased cancer risk for the individual carrying a variant in one of these genes is estimated to be small. Variable penetrance may explain why unaffected individuals can carry potentially pathogenic variants without displaying disease characteristics and therefore, these diseases are occasionally transmitted through unaffected parents. However, the high frequency in the population of

some of these variants suggests that the population attributable risk can be high (Meerman & de Vries 2001). Very few consistent gene-disease associations have been identified in humans due to association studies lacking statistical power required to detect the weak associations that low-penetrance alleles are likely to present (Houlston & Peto 2004). This is consistent with multiple small-effect alleles in our assessment of SCA foxes. Genomic studies, such as high-throughput re-sequencing analysis, will need much larger samples sizes of well-characterized cases and controls, in order to identify rare pathogenic variants that constitute the majority of low-penetrance cancer susceptibility alleles (Houlston & Peto 2004; Chung & Chanock 2011).

Heritable, and/or familial, cancer syndromes may be caused by the interaction of low-penetrance genes, gene-environmental interactions, or both (Nagy et al. 2004). Little is known regarding island fox ear canal microbiome, mite infection, and inflammation. Previous studies have not found any bacteria or fungus associated with elevated otitis in SCA foxes; however, other approaches, such as metagenomics, have not been applied to examine the microbiome associations with mite infection or host genetics. Interestingly, KRT8 (ENSCAFG00000002184, an ortholog of Keratin, type II cytoskeletal 8), a gene containing SNPS identified by our selection scans as differentiated between cases and controls, maintains gut microbiota homeostasis as well as reduces colonic permeability, which is important in protecting against inflammation leading to colitis and colitis-associated tumorigenesis (Liu et al. 2017). This gene has been found to be downregulated in cancers such as breast and colorectal carcinomas (Woelfle et al. 2004; Knösel et al. 2006). It may be possible that a similar mechanism is present in the development of ceruminous gland carcinoma in SCA foxes.

5.5.3 Candidate genes for cancer susceptibility

Both candidate gene (Supporting Information Table C.7) and GO analyses suggested that selection on immunological and tumorigenesis pathways has occurred in SCA foxes, as

we predicted. Given the potential problem of over analyzing candidate genes (Pavlidis et al. 2012), we keep the discussions brief (see Supporting Information for additional comments on candidate genes). We have chosen especially convincing candidates that have high support as outliers from multiple independent selection tests each with their own unique assumption. Furthermore, most of our candidate SNPs were found within intergenic and intronic regions and some within TFBS's. Many variants associated with cancer risk are located in intergenic and intronic regions with unknown functions (Hindorff et al. 2011). This can make interpretation difficult, but may suggest that modification of gene regulatory regions may contribute disproportionately to the modulation of risk.

5.5.4 Candidate genes for cancer susceptibility: Oxidative Stress

The two significant genes involved in superoxide anion generation are Decapping MRNA 1A (DCP1A) and C-Reactive Protein (CRP). The DCP1A gene is part of the PIWI-piRNA pathway and functions to degradation of normal mRNA turnover and nonsense-mediated mRNA as well as TGF-beta signaling pathway. Genetic mutations within this gene have been found to predict survival in melanoma patients (Zhang et al. 2016). The second gene, CRP encodes for a protein that is produced mainly in the liver in response to cytokines during infection, trauma, advanced cancer, and chronic inflammatory conditions (Gabay & Kushner 1999). This protein is often measured in order to predict disease such as cardiovascular disease in women (Ridker et al. 2000) and various types of cancer (Heikkilä et al. 2007; Allin et al. 2009). Several studies have shown genetic variants in the CRP gene affect blood concentrations of CRP and colorectal cancer risk and survival (Erlinger et al. 2004; Nimptsch et al. 2015).

The two significant GO terms both involve the (positive) regulation of superoxide anion generation. Several lines of evidence suggest that an excessive generation of the reactive oxygen species superoxide anion ($O_2^{\cdot-}$) is a key event in cancer development and survival (reviewed in (Lázaro 2007; Kumari et al. 2018). An increase in the cellular production

of $O_2^{\cdot-}$ may activate glycolysis and glycolysis activation has been associated with invasion, metastasis, angiogenesis and cell proliferation (Gatenby & Gillies 2004). Simultaneously, $O_2^{\cdot-}$ generation activates Hypoxia-inducible factor 1 (HIF-1; Wang et al. 2004a; López-Lázaro 2006). Activation of HIF-1 can induce apoptosis resistance, invasion, metastasis and angiogenesis (Semenza 2003; Yeo et al. 2004) and may also produce cellular immortalization via activation of telomerase (Nishi et al. 2004). Furthermore, A high concentration of reactive oxygen species, such as $O_2^{\cdot-}$, has been shown to produce oxidative damage of DNA, causing mutations which eventually lead to cancer (reviewed in Kumari et al. 2018). Therefore, the genes (DCP1A and CRP) and surrounding intergenic regions may be good candidates for a subsequent fine-mapping study using more individuals to identify potential causal loci underlying cancer risk in the SCA population.

Several candidate genes, including DCP1A and CRP, interact with phosphate (Soelter & Uhlenbruck 1986; Blumenthal et al. 2009) and potentially are involved in phosphocholine metabolism, a marker for oxidative stress (Frey et al. 2000; Mateos et al. 2008). Some studies have shown that elevated concentrations of phosphate increased cell proliferation and expression of protumorigenic genes (Camalier et al. 2010; 2013). Two of these genes, both members of the Protein Tyrosine Phosphatase (PTP) family, were found with differentiated SNPs between cases and controls. These two genes are Dual Specificity Phosphatase 23 (ENSCAFG00000031934 (DUSP23) and Protein tyrosine phosphatase μ (PTPRM; also named PTPμ). In general, PTPs are susceptible to oxidative stress and loss-of-function mutations involving PTPs are frequently observed in various types of cancers (Wang et al. 2004b; Chen et al. 2006; Flavell et al. 2008; Novellino et al. 2008; Cheung et al. 2008; Chan & Heguy 2009). PTPRM is a member of the type IIb subfamily of receptor PTPs (Lamprianou & Harroch 2006) and is involved in cell-cell adhesion (Brady-Kalnay & Tonks 1993). Cleavage, by multiple proteases, of PTPRM regulates cell migration in glioblastoma, an aggressive form of brain cancer (Burgoyne et al. 2009; Phillips-Mason et al. 2014). Promotor hypermethylation result in the loss of function of PTPRM and has been shown in

some acute lymphocytic leukemia (Stevenson et al. 2014) and can lead to colony formation in colon cancer (Sudhir et al. 2015). Additionally, decreased expression of PTPRM is associated with poor prognosis in breast cancer (Sun et al. 2012). Atypical dual specificity phosphatases have been implicated in various types of cancer (Cain & Beeser 2013). VHZ protein, which is encoded for by DUSP23, was overexpressed in breast cancers and various other tumor types (Tang et al. 2010). Further investigation of these two phosphatases will help to increase understanding of the molecular mechanism responsible for its role in promoting cell migration and cell proliferation.

5.5.5 Candidate genes for cancer susceptibility: lincRNA

Long non-coding RNAs (>200 nts; lncRNAs) play critical roles in gene transcription, translation, and chromatin modification. Most of large intergenic noncoding RNAs (lincRNAs), a subclass lncRNAs, are enriched in evolutionarily conserved sequences and therefore likely functional (Khalil et al. 2009); yet are rarely functionally annotated (Quek et al. 2015). lncRNAs compete with proteins in terms of their diversity and regulatory potential through a wide range of mechanisms such as chromatin and methylation modification and activation, direct effect on stability of protein and protein complexes, or by acting as a sponge for miRNA inhibition (Quinn & Chang 2016; Bartonicek et al. 2016). lncRNAs can modify the phosphorylation state of proteins by masking phosphorylation motifs (Liu et al. 2015). lncRNAs as regulatory molecules have been implicated in the majority of these hallmarks of cancer (reviewed in (Gutschner & Diederichs 2012; Bartonicek et al. 2016)). The hallmarks of cancer or the six properties required for cell transformation in tumorigenesis are self-sustained growth signaling, insensitivity to growth inhibition, apoptosis avoidance, uncontrolled proliferation, angiogenesis and metastasis (Hanahan & Weinberg 2000; 2011). It may be plausible that the three lincRNA's (ENSCAFG00000039008, ENSCAFG00000039447, and ENSCAFG00000039265) we found with downstream and intronic SNPs may have a role in cancer susceptibility in SCA foxes. Additionally, ENSCAFG00000030255 (DLX6 antisense

RNA 1 (DLX6-AS1)), is a developmentally-regulated long non-coding RNA that was also contained a highly differentiated SNP between cases and controls. High DLX6-AS1 expression was noticed in lung adenocarcinoma and associated with histological differentiation and TNM stage (Li et al. 2015). DLX6-AS1 was also up-regulated in hepatocellular carcinoma tissue and correlated with clinical prognosis (Li et al. 2017). Future studies of these lncRNA's should be explored to better elucidate the pathways involved in cancer susceptibility in SCA foxes.

5.5.6 Complementary causes of disease

Once viewed as an accumulation of genomic aberrations, recent studies have emphasized that cancers are heterogeneous collections of cells (Maley et al. 2006; Campbell et al. 2008; Merlo & Maley 2010; Park et al. 2010) that evolve in tumor microenvironments with complex ecologies (Bissell & Radisky 2001). Tumor microenvironments may be a vital aspect of tumorigenesis (Wang et al. 2017). When SCA foxes were treated for mite infections, not only were mite loads greatly reduced, but treated foxes has significantly reduced hyperplasia compared with untreated controls (Moriarty et al. 2015). The researchers proposed that the long-term presence of mites is associated with epithelial hyperplasia, and in support of causality, removal of the parasite burden resulted in reversal of tumor development. In addition to the inherited increased risk in cancer as we have proposed here, it may be that the mites, undetected toxins, or microbial community alterations due to mite infection may affect the tumor environment. Further exploration of the tumor microenvironments should take place to gain a better understanding of the mechanisms involved in cancer development.

5.5.7 Conservation Implications

Substantial morbidity and mortality in many wildlife species may be due to cancer (reviewed in (McAloose & Newton 2009; Pesavento et al. 2018). In domestic animals, anthropogenically-induced population bottlenecks and selective breeding appear to con-

tribute to oncogenic processes (Vail & MacEwen 2000; Dobson 2013). This, as well as other anthropogenic activities such as polluting, may also influence cancer development in wild populations (Ujvari et al. 2018; Giraudeau et al. 2018). In the case of Santa Catalina Island foxes, a human-induced viral epidemic led to a bottleneck. The reduced effective population size, and increased effect of drift may have contributed to an increase in cancer susceptibility. However, hidden heritability of this polygenic trait is likely to impede the genetic monitoring of this population over time. Ultimately, the development of a susceptibility panel of genetic loci, based on identified genes and genetic pathways, would provide an efficient way to genetically assess an individual fox's probability of developing cancer as a hyper-immune response to mite infection. Because this trait is likely due to the effects of many, slightly deleterious mutations, the number of loci required for a susceptibility genetic marker panel would need to be greater than if a large effect allele were the causal polymorphism, although the precise number is difficult to estimate. Little is known on how to genetically monitor quantitative maladaptive traits using cost effective methods. However, novel method development involving the combination of outlier analysis, genome mapping, and machine learning may reveal more about the molecular basis of this polygenic trait (Brieuc et al. 2015) and allow for more accurate prediction of individuals with elevated risk of cancer. Long-term monitoring of the disease alleles as well as overall genetic diversity will provide a crucial index for their long-term persistence and management.

While it is well supported in applied conservation that there are positive effects of population augmentation with divergent immigrants (reviewed in Whiteley et al. 2015), it is often difficult to predict if this type of genetic rescue outweighs the potential risks of outbreeding depression or if natural purging of deleterious alleles would have resulted in the desired outcome without the extensive resources often required of genetic rescue. Recent genomic assessments of Island foxes have shown that this species purges strong deleterious alleles reducing the risk associated with inbreeding depression (Robinson et al. 2016, 2018). In fact, morphological assessment indicates that island foxes do not exhibit canonical signs

of inbreeding depression (Robinson et al. 2018). Despite elevated genetic diversity in SCA as compared with other island fox populations, small population size may have led to an increased effect of drift of deleterious mutations of small effect. Perhaps as the population size increased, this subspecies has been able to tolerate inbreeding depression and time will allow more purifying selection to take place.

Here we found evidence that cancer susceptibility in SCA foxes results from deleterious alleles at multiple loci. Although selection against these weakly deleterious mutations may take place slower than at fewer large-effect alleles, augmentation of captively breed individuals from parents without cancer may contribute to overall genetic robustness and could decrease the prevalence at a faster rate than naturally occurring selection. However, it appears unlikely that the SCA population completely lacks genetic variation for susceptibility. We found no evidence of candidate loci fixed for alternative alleles between groups, suggesting that the population maintains allelic variation that can allow for an adaptive response to the disease, even if selection is relatively weak on each locus. This implies that attempts to increase adaptive potential by translocating individuals from other islands are unlikely to have a large benefit. We suggest further field and genetic monitoring to illuminate trends in cancer prevalence as well as simulations of purifying selection - drift balance to predict the future course of population-level susceptibility under different scenarios.

An increasing body of knowledge is emerging with regards to conservation and management of adaptive potential (Hoelzel et al. 2019); yet, theory and applied studies of genetic management to confront situations where reduced population fitness is caused by many small-effect loci, rather than large-effect loci, remain scarce. Due to the complexity of polygenic diseases, embracing recent advances in other fields, such as quantitative genetics, human genetics, epigenetics, and expression profiling, should result in better informed applications of conservation genomics than have been previously possible for most wild organisms (Harrisson et al. 2014). Further studies are needed to decipher the underpinnings of cancer risk, inbreeding, and infection/inflammation particularly in a changing environment.

5.6 Acknowledgements

We thank the Catalina Island Conservancy, Megan Moriarty, Patty Gaffney, William Andelt, and Chris Funk for sample acquisition, characterization, and sharing for genomic sequencing. We thank Rena Schweizer, Clare Marsden, Amanda Stahlke, and Annabel Beichman for helpful discussions concerning methods' implementation. This study could not have been completed without Dr. Linda Munson, 1948-2010. She was one of the initiators and contributors of the island fox carcinoma studies. She was committed to investigating cancer and other diseases in island foxes as well as many other wildlife populations. This work used the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. Support was provided by the Morris Animal Foundation and the Institute for Wildlife Studies. Support was provided to SAH by the National Institute of Health (P30GM103324), National Science Foundation (DEB-1316549) and the Bioinformatics and Computational Biology program at the University of Idaho.

5.7 References

Abascal F, Corvelo A, Cruz F et al. (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology, 1-20.

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, 101, 3490-3494.

Allin KH, Bojesen SE, Nordestgaard BG (2009) Baseline C-reactive protein is associated with incident cancer and survival in patients with cancer. Journal of Clinical Oncology, 27, 2217-2224.

Bartonicek N, Maag JLV, Dinger ME (2016) Long noncoding RNAs in cancer: mechanisms of action and technological advancements. Molecular Cancer, 15, 43.

Bissell MJ, Radisky D (2001) Putting tumours in context. Nature Reviews Cancer, 1, 46-54.

Blumenthal J, Behar L, Elliott E, Ginzburg I (2009) Dcp1a phosphorylation along neuronal development and stress. FEBS letters, 583, 197-201.

Brady-Kalnay SM, Tonks NK (1993) Purification and characterization of the human protein tyrosine phosphatase, PTP mu, from a baculovirus expression system. Molecular and cellular biochemistry, 127-128, 131-141.

Brieuc MSO, Ono K, Drinan DP, Naish KA (2015) Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology, 24, 2729-2746.

Browning HM, Acevedo-Whitehouse K, Gulland FMD et al. (2014) Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. Proceedings of The Royal Society B: Biological sciences, 281, 20140240-20140240.

Burgoyne AM, Phillips-Mason PJ, Burden-Gulley SM et al. (2009) Proteolytic cleavage of protein tyrosine phosphatase mu regulates glioblastoma cell migration. Cancer Research, 69, 6960-6968.

Camalier CE, Yi M, Yu LR et al. (2013) An integrated understanding of the physiological response to elevated extracellular phosphate. Journal of cellular physiology, 228, 1536-1550.

Camalier CE, Young MR, Bobe G et al. (2010) Elevated Phosphate Activates N-ras and Promotes Cell Transformation and Skin Tumorigenesis. Cancer Prevention Research, 3, 359-370.

Campbell PJ, Pleasance ED, Stephens PJ et al. (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proceedings of the National Academy of Sciences of the United States of America, 105, 13081-13086.

Caporaso N, Landi MT, Vineis P (1991) Relevance of metabolic polymorphisms to human carcinogenesis: evaluation of epidemiologic evidence. Pharmacogenetics, 1, 4-19.

Chan TA, Heguy A (2009) The protein tyrosine phosphatase receptor D, a broadly inactivated tumor suppressor regulating STAT function. Cell cycle (Georgetown, Tex.), 8, 3063-3064.

Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. Genetical research, 74, 329-340.

Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. Genome Research, 20, 393-402.

Chen L, Juszczynski P, Takeyama K, Aguiar RCT, Shipp MA (2006) Protein tyrosine phosphatase receptor-type O truncated (PTPROt) regulates SYK phosphorylation, proximal B-cell-receptor signaling, and cellular proliferation. Blood, 108, 3428-3433.

Cheung AKL, Lung HL, Hung SC et al. (2008) Functional analysis of a cell cycle-associated, tumor-suppressive gene, protein tyrosine phosphatase receptor type G, in nasopharyngeal carcinoma. Cancer Research, 68, 8137-8145.

Chung CC, Chanock SJ (2011) Current status of genome-wide association studies in cancer. Human Genetics, 130, 59-78.

Cingolani P, Platts A, Wang LL et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6, 80-92.

Collins PW (1991) Interaction between island foxes (*Urocyon littoralis*) and Indians on islands off the coast of southern California: I. Morphologic and archaeological evidence of of human assisted dispersal. Journal of Ethnobiology.

Coonan TJ, Schwemm CA, Garcelon DK (2010) Decline and Recovery of the Island Fox: A Case Study for Population Recovery. Cambridge University Press.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156-2158.

Eichler EE, Flint J, Gibson G et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics, 11, 446-450.

England PR, Cornuet J-M, Berthier P, Tallmon DA, Luikart G (2006) Estimating effective population size from linkage disequilibrium: severe bias in small samples. Conservation Genetics, 7, 303-308.

Epstein B, Jones ME, Hamede R et al. (2016) Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications, 7, 12684.

Erlinger TP, Platz EA, Rifai N, Helzlsouer KJ (2004) C-reactive protein and the risk of incident colorectal cancer. JAMA, 291, 585-590.

Flavell JR, Baumforth KRN, Wood VHJ et al. (2008) Down-regulation of the TGF-beta target gene, PTPRK, by the Epstein-Barr virus encoded EBNA1 contributes to the growth and survival of Hodgkin lymphoma cells. Blood, 111, 292-301.

Frankham R (2005a) Stress and adaptation in conservation genetics. Journal of Evolutionary Biology, 18, 750-755.

Frankham R (2005b) Genetics and extinction. Biological Conservation, 126, 131-140.

Freedman AH, Gronau I, Schweizer RM et al. (2014) Genome sequencing highlights the dynamic early history of dogs. (L Andersson, Ed,). PLoS Genetics, 10, e1004016-12.

Frey B, Haupt R, Alms S et al. (2000) Increase in fragmented phosphatidylcholine in blood plasma by oxidative stress. Journal of Lipid Research, 41, 1145-1153.

Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations. The American Journal of Human Genetics, 95, 421-436.

Funk WC, Lovich RE, Hohenlohe PA et al. (2016) Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Molecular Ecology, 25, 2176-2194.

Gabay C, Kushner I (1999) Acute-Phase Proteins and Other Systemic Responses to Inflammation (FH Epstein, Ed,). The New England Journal of Medicine, 340, 448-454.

García-Alcalde F, Okonechnikov K, Carbonell J et al. (2012) Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics, 28, 2678-2679.

Gatenby RA, Gillies RJ (2004) Why do cancers have high aerobic glycolysis? Nature Reviews Cancer, 4, 891-899.

Gilpin ME, Soulé ME (1986) Minimal viable populations: processes of species extinction. In: Conservation Biology The Science of Scarcity and Diversity (ed Soulé ME), pp. 19-34, Sunderland, Mass.

Giraudeau M, Sepp T, Ujvari B, Ewald PW, Thomas F (2018) Human activities might influence oncogenic processes in wild animal populations. Nature ecology & evolution, 2, 1065-1070.

Glémin S, Ronfort J, Bataillon T (2003) Patterns of Inbreeding Depression and Architecture of the Load in Subdivided Populations. Genetics, 165, 2193-2212.

Gomulkiewicz R, Holt RD (1995) When does Evolution by Natural Selection Prevent Extinction? Evolution; international journal of organic evolution, 49, 201.

Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinformatics, 27, 1017-1018.

Griner EM, Kazanietz MG (2007) Protein kinase C and other diacylglycerol effectors in cancer. Nature Reviews Cancer, 7, 281-294.

Gutschner T, Diederichs S (2012) The hallmarks of cancer. RNA Biology, 9, 703-719.

Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. Cell, 100, 57-70.

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell, 144, 646-674.

Harrisson KA, Pavlova A, Scott MT, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. Evolutionary Applications, 7, 1008-1025.

Heikkilä K, Ebrahim S, Lawlor DA (2007) A systematic review of the association between circulating concentrations of C reactive protein and cancer. Journal of Epidemiology and Community Health, 61, 824-833.

Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015) Estimating Mutation Load in Human Genomes. Nature Reviews Genetics, 16, 333-343.

Hindorff LA, Gillanders EM, Manolio TA (2011) Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis, 32, 945-954.

Hoelzel AR, Bruford MW, Fleischer RC (2019) Conservation of adaptive potential and functional diversity. Conservation Genetics, 20, 1-5.

Hofman CA, Rick TC, Hawkins MTR et al. (2015) Mitochondrial genomes suggest rapid evolution of dwarf California Channel Islands foxes (*Urocyon littoralis*). PLoS ONE, 10, e0118240.

Houlston RS, Peto J (2004) The search for low-penetrance cancer susceptibility alleles. 23, 6471-6476.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 23, 254-267.

Jen J, Wang Y-C (2016) Zinc finger proteins in cancer progression. Journal of Biomedical Science, 23, 53.

Khalil AM, Guttman M, Huarte M et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences, 106, 11667-11672.

Kirkpatrick M, Jarne P (2015) The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. The American Naturalist, 155, 154-167.

Knösel T, Emde V, Schlüns K et al. (2006) Cytokeratin profiles identify diagnostic signatures in colorectal cancer using multiplex analysis of tissue microarrays. Cellular Oncology, 28, 167-175.

Kumari S, Badana AK, G MM, G S, Malla R (2018) Reactive Oxygen Species: A Key Constituent in Cancer Survival. Biomarker Insights, 13, 117727191875539.

Lamprianou S, Harroch S (2006) Receptor protein tyrosine phosphatase from stem cells to mature glial cells of the central nervous system. Journal of molecular neuroscience : MN, 29, 241-255.

Laughrin LL (1977) The island fox: a field study of its behavior and ecology.

Lázaro ML (2007) Excessive superoxide anion generation plays a key role in carcinogenesis. International Journal of Cancer, 120, 1378-1380.

Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 2078-2079.

Li J, Li P, Zhao W et al. (2015) Expression of long non-coding RNA DLX6-AS1 in lung adenocarcinoma. Cancer cell international, 15, 48.

Li L, Hou A, Gao X et al. (2017) Lentivirus-mediated miR-23a overexpression induces trophoblast cell apoptosis through inhibiting X-linked inhibitor of apoptosis. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie, 94, 412-417.

Liu B, Sun L, Liu Q et al. (2015) A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. Cancer cell, 27, 370-381.

Liu C, Liu E-D, Meng Y-X et al. (2017) Keratin 8 reduces colonic permeability and maintains gut microbiota homeostasis, protecting against colitis and colitis-associated tumorigenesis. Oncotarget, 8, 96774-96790.

Liu Z, Li X, He X et al. (2011) Decreased expression of updated NESG1 in nasopharyngeal carcinoma: Its potential role and preliminarily functional mechanism. International Journal of Cancer, 128, 2562-2571.

López-Lázaro M (2006) HIF-1: hypoxia-inducible factor or dysoxia-inducible factor? The FASEB Journal, 20, 828-832.

Madsen T, Arnal A, Vittecoq M et al. (2017) Cancer prevalence and etiology in wild and captive animals. In: Ecology and Evolution of Cancer (eds Ujvari B, Roche B, Thomas F). Elsevier.

Maher B (2008) Personal genomes: The case of the missing heritability. Nature, 456, 18-21.

Maley CC, Galipeau PC, Finley JC et al. (2006) Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nature Genetics, 38, 468-473.

Manichaikul A, Mychaleckyj JC, Rich SS et al. (2010) Robust relationship inference in genome-wide association studies. Bioinformatics, 26, 2867-2873.

Manolio TA, Collins FS, Cox NJ et al. (2009) Finding the missing heritability of complex diseases. Nature Reviews Genetics, 461, 747-753.

Margres MJ, Jones ME, Epstein B et al. (2018) Large-effect loci affect survival in Tasmanian devils (Sarcophilus harrisii) infected with a transmissible cancer. Molecular Ecology, 27, 4189-4199.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP et al. (2015) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences, 201512501-76.

Mateos MV, Uranga RM, Salvador GA, Giusto NM (2008) Activation of phosphatidylcholine signalling during oxidative stress in synaptic endings. Neurochemistry international, 53, 199-206.

McAloose D, Newton AL (2009) Wildlife cancer: a conservation perspective. Nature Reviews Genetics, 9, 517-526.

McKenna A, Hanna M, Banks E et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297-1303.

Meerman te GJ, de Vries E (2001) Relevance of high and low penetrance. The Lancet, 358.

Meighan CW (2017) The Little Harbor Site, Catalina Island: An Example of Ecological Interpretation in Archaeology. American Antiquity, 24, 383-405.

Merlo LMF, Maley CC (2010) The role of genetic diversity in cancer. The Journal of Clinical Investigation, 120, 401-403.

Moriarty ME, Vickers TW, Clifford DL et al. (2015) Ear Mite Removal in the Santa Catalina Island Fox (Urocyon littoralis catalinae): Controlling Risk Factors for Cancer Development. PLoS ONE, 10, e0144271-15.

Nagy R, Sweet K, Eng C (2004) Highly penetrant hereditary cancer syndromes. 23, 6445-6470.

Nimptsch K, Aleksandrova K, Boeing H et al. (2015) Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk. International Journal of Cancer, 136, 1181-1192.

Nishi H, Nakada T, Kyo S et al. (2004) Hypoxia-inducible factor 1 mediates upregulation of telomerase (hTERT). Molecular and Cellular Biology, 24, 6076-6083.

Novellino L, De Filippo A, Deho P et al. (2008) PTPRK negatively regulates transcriptional activity of wild type and mutated oncogenic beta-catenin and affects membrane distribution of beta-catenin/E-cadherin complexes in cancer cells. Cellular signalling, 20, 872-883.

Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics, 32, 292-294.

O'Grady JJ, Brook BW, Reed DH et al. (2006) Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. Biological Conservation, 133, 42-51.

Park HS, Yeo HY, Chang HJ et al. (2013) Dipeptidyl peptidase 10, a novel prognostic marker in colorectal cancer. Yonsei Medical Journal, 54, 1362-1369.

Park SY, Gönen M, Kim HJ, Michor F, Polyak K (2010) Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. The Journal of Clinical Investigation, 120, 636-644.

Pesavento PA, Agnew D, Keel MK, Woolard KD (2018) Cancer in wildlife: patterns of emergence. Nature Reviews Cancer, 1-16.

Phillips-Mason PJ, Craig SEL, Brady-Kalnay SM (2014) A protease storm cleaves a cell-cell adhesion molecule in cancer: multiple proteases converge to regulate PTPmu in glioma cells. Journal of cellular biochemistry, 115, 1609-1623.

Ponder BAJ (2001) Cancer genetics. Nature Reviews Genetics, 411, 336-341.

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics, 81, 559-575.

Qin Z, Ren F, Xu X et al. (2009) ZNF536, a Novel Zinc Finger Protein Specifically Expressed in the Brain, Negatively Regulates Neuron Differentiation by Repressing Retinoic Acid-Induced Gene Transcription. Molecular and Cellular Biology, 29, 3633-3643.

Quek XC, Thomson DW, Maag JLV et al. (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic acids research, 43, D168-73.

Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nature Reviews Genetics, 17, 47-62.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Reimand J, Arak T, Adler P et al. (2016) g:Profiler - a web server for functional interpretation of gene lists (2016 update). Nucleic acids research, 44, W83-W89.

Ridker PM, Hennekens CH, Buring JE, Rifai N (2000) C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. The New England Journal of Medicine, 342, 836-843.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK (2018) Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology, 28, 3487-3494.e4.

Robinson JA, Vecchyo DO-D, Fan Z et al. (2016) Genomic Flatlining in the Endangered Island Fox. Current Biology, 1-25.

Schweizer RM, Robinson JA, Harrigan RJ et al. (2016) Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Molecular Ecology, 25, 357-379.

Semenza GL (2003) Targeting HIF-1 for cancer therapy. Nature Reviews Cancer, 3, 721-732.

Soelter J, Uhlenbruck G (1986) The role of phosphate groups in the interaction of human C-reactive protein with galactan polysaccharides. Immunology, 58, 139-144.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004) Does Inbreeding and Loss of Genetic Diversity Decrease Disease Resistance? Conservation Genetics, 5, 439-448.

Staples J, Nickerson DA, Below JE (2013) Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. Genetic epidemiology, 37, 136-141.

Stevenson WS, Best OG, Przybylla A et al. (2014) DNA methylation of membrane-bound tyrosine phosphatase genes in acute lymphoblastic leukaemia. Leukemia, 28, 787-793.

Sudhir P-R, Lin S-T, Chia-Wen C et al. (2015) Loss of PTPRM Associates with the Pathogenic Development of Colorectal Adenoma-Carcinoma Sequence. Scientific Reports, 5, 759.

Sun P-H, Ye L, Mason MD, Jiang WG (2012) Protein tyrosine phosphatase μ (PTP μ or PTPRM), a negative regulator of proliferation and invasion of breast cancer cells, is associated with disease prognosis. (VK Poojary, Ed,). PLoS ONE, 7, e50183.

Tang JP, Tan CP, Li J et al. (2010) VHZ is a novel centrosomal phosphatase associated with cell growth and human primary cancers. Molecular Cancer, 9, 128.

Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF (2002) Sampling distribution of summary linkage disequilibrium measures. Annals of Human Genetics, 66, 223-233.

Terhorst J, Kamm JA, Song YS (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics, 49, 303-309.

Ujvari B, Klaassen M, Raven N et al. (2018) Genetic diversity, inbreeding and cancer. Proceedings of The Royal Society B: Biological sciences, 285, 20172589.

US Fish and Wildlife Service (2016) Removing the San Miguel Island fox, Santa Rosa Island fox, and Santa Cruz Island fox from the federal list of Endangered and Threatened wildlife, and reclassifying the Santa Catalina Island fox from Endangered to Threatened; final rule.

Vickers TW, Clifford DL, Garcelon DK et al. (2015) Pathology and Epidemiology of Ceruminous Gland Tumors among Endangered Santa Catalina Island Foxes (*Urocyon littoralis catalinae*) in the Channel Islands, USA. PLoS ONE, 10, e0143211-18.

vonHoldt BM, Pollinger JP, Earl DA et al. (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Research, 21, 1294-1305.

Wang F-S, Wang C-J, Chen Y-J et al. (2004a) Ras induction of superoxide activates ERK-dependent angiogenic transcription factor HIF-1alpha and VEGF-A expression in shock wave-stimulated osteoblasts. Journal of Biological Chemistry, 279, 10331-10337.

Wang M, Zhao J, Zhang L et al. (2017) Role of tumor microenvironment in tumorigenesis. Journal of Cancer, 8, 761-773.

Wang Y, Liu Z, Luo R, X Y (2018) Decreased CCDC19 is correlated with unfavorable outcome in lung squamous cell carcinoma. International Journal of Clinical and Experimental Pathology, 11, 802-807.

Wang Z, Shen D, Parsons DW et al. (2004b) Mutational analysis of the tyrosine phosphatome in colorectal cancers. Science, 304, 1164-1166.

Wayne RK, George SB, Gilbert D et al. (1991) A Morphologic and Genetic Study of the Island Fox, Urocyon littoralis. Evolution; international journal of organic evolution, 45, 1849.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. Trends in Ecology & Evolution, 30, 42-49.

Woelfle U, Sauter G, Santjer S, Brakenhoff R, Pantel K (2004) Down-regulated expression of cytokeratin 18 promotes progression of human breast cancer. Clinical cancer research, 10, 2670-2674.

Wu Q, Li Y, Gu S et al. (2004) Molecular cloning and characterization of a novel dual-specificity phosphatase 23 gene from human fetal brain. The international journal of biochemistry & cell biology, 36, 1542-1553.

Xie X, Lu J, Kulbokas EJ et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature Reviews Genetics, 434, 338-345.

Xue Y, Prado-Martinez J, Sudmant PH et al. (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science, 348, 242-245.

Yang J, Bakshi A, Zhu Z et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nature Genetics, 47, 1114-1120.

Yang J, Benyamin B, McEvoy BP et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics, 42, 565-569.

Yeo E-J, Chun Y-S, Park J-W (2004) New anticancer strategies targeting HIF-1. Biochemical Pharmacology, 68, 1061-1069.

Zhang W, Fan Z, Han E et al. (2014) Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau (JM Akey, Ed,). PLoS Genetics, 10, e1004466-13.

Zhang W, Liu H, Yin J et al. (2016) Genetic variants in the PIWI-piRNA pathway gene DCP1A predict melanoma disease-specific survival. International Journal of Cancer, 139, 2730-2737.

Zhang Z-Y (2003) Protein Tyrosine Phosphatases: Structure and Function, Substrate Specificity, and Inhibitor Development. Annual Review of Pharmacology and Toxicology, 42, 209-234.

Appendix A: Supplementary Information to Chapter 2

A.1 Sampling data

Table A.1: Data per site including the latitude and longitude of each location, the number of samples collected per site, the year samples were collected, and the year of first detection of disease (or absence of disease)

ID	LOCATION	LAT.	LON.	N	DATE COLLECTED	~ DATE OF DISEASE OUTBREAK
1	Mt William	-40.886	148.190	2	2008	<1996
2	St Helens	-41.340	148.170	1	2006	<1996
3	St Marys	-41.544	148.117	2	2006	<1996
4	Blessington	-41.473	147.457	2	2005	2003
5	Freycinet	-42.220	148.310	2	1999	2001
6	Little Swanport	-42.290	147.910	1	1999	1999
7	Buckland	-42.430	147.772	2	2006	1999
8	Pawleena	-42.690	147.610	2	1999	2003
9	Kellevie	-42.740	147.800	2	1998	2003
10	Forestier	-42.897	147.906	2	2004	2006
11	Bothwell	-42.330	147.000	2	2008	2004
12	Elderslie	-42.526	147.042	2	2008	2004
13	Fentonbury	-42.621	146.790	2	2004	2004

Table A.1 continued from previous page

14	Lonnavale	-42.928	146.792	1	2004	2004
15	Glen Huon	-42.988	146.802	2	2008	2004
16	Hastings Caves	-43.479	146.905	1	2005	2006
17	Reedy Marsh	-42.750	146.287	2	2007	2004
18	Bronte Park	-42.096	146.441	2	2004	2004
19	Lake Rowallan	-41.766	146.166	2	2004	2004
20	Wisedale	-41.294	146.671	2	2006	2004
21	Harford	-41.260	146.590	1	2006	2003
22	Narawntapu	-41.220	146.640	2	1999	2007
23	Mt Housetop	-41.268	145.874	2	2007	2005
24	West Pencil Pine	-41.517	145.767	2	2006 & 2007	2006
25	Takone	-41.195	145.592	2	2007	2010
26	Mt Hicks	-41.019	145.658	1	2007	2011
27	Flowerdale	-41.107	145.499	2	2009	2012
28	Peegra Dip Falls	-41.030	145.460	2	2009	2013
29	Milkshake Hills	-41.109	145.188	2	2004	2013
30	Woolnorth	-40.693	144.710	2	2006	No disease
31	Arthur River	-40.992	144.672	2	2007	No disease

Table A.1 continued from previous page

32	Savage River	-41.651	145.079	2	2007	No disease
33	Granville Harbour	-41.794	145.066	2	2004	2015
34	Queenstown North	-41.941	145.459	2	2007	2014
35	Mt Arrowsmith	-42.165	145.927	2	2004 & 2005	2005
36	Queenstown South	-42.166	145.547	2	2007	No disease
37	Strahan	-41.170	145.450	2	2008	No disease
38	Macquarie Heads	-42.393	145.232	2	2006	No disease

A.2 SNP data

Table A.2: The number of SNPs discovered per chromosome

Chromosome	# of SNPs
1	1419
2	1485
3	1228
4	924
5	636
6	666
Unknown	4

A.3 STRUCTURE Harvester results

Table A.3: The harvested results of STRUCTURE analyses. Using the Evanno method, K = 2 (in bold) was the optimal number of clusters

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	—Ln"(K)—	Delta K
1	20	-234254.8500	13.6578	NA	NA	NA
2	20	-220483.4000	4.6169	13771.450000	8820.500000	1910.479989
3	20	-215532.4500	3.0266	4950.950000	2853.470000	942.785941
4	20	-213434.9700	235.0678	2097.480000	15949.040000	67.848677
5	20	-227286.5300	26674.4181	-13851.560000	20719.135000	0.776742
6	20	-220418.9550	23029.3639	6867.575000	3436.580000	0.149226
7	20	-210114.8000	166.3816	10304.155000	9570.520000	57.521497
8	20	-209381.1650	154.7725	733.635000	304.065000	1.964594
9	20	-208951.5950	296.2365	429.570000	864.250000	2.917433
10	20	-209386.2750	940.3062	-434.680000	NA	NA

A.4 Isolation-by-distance results

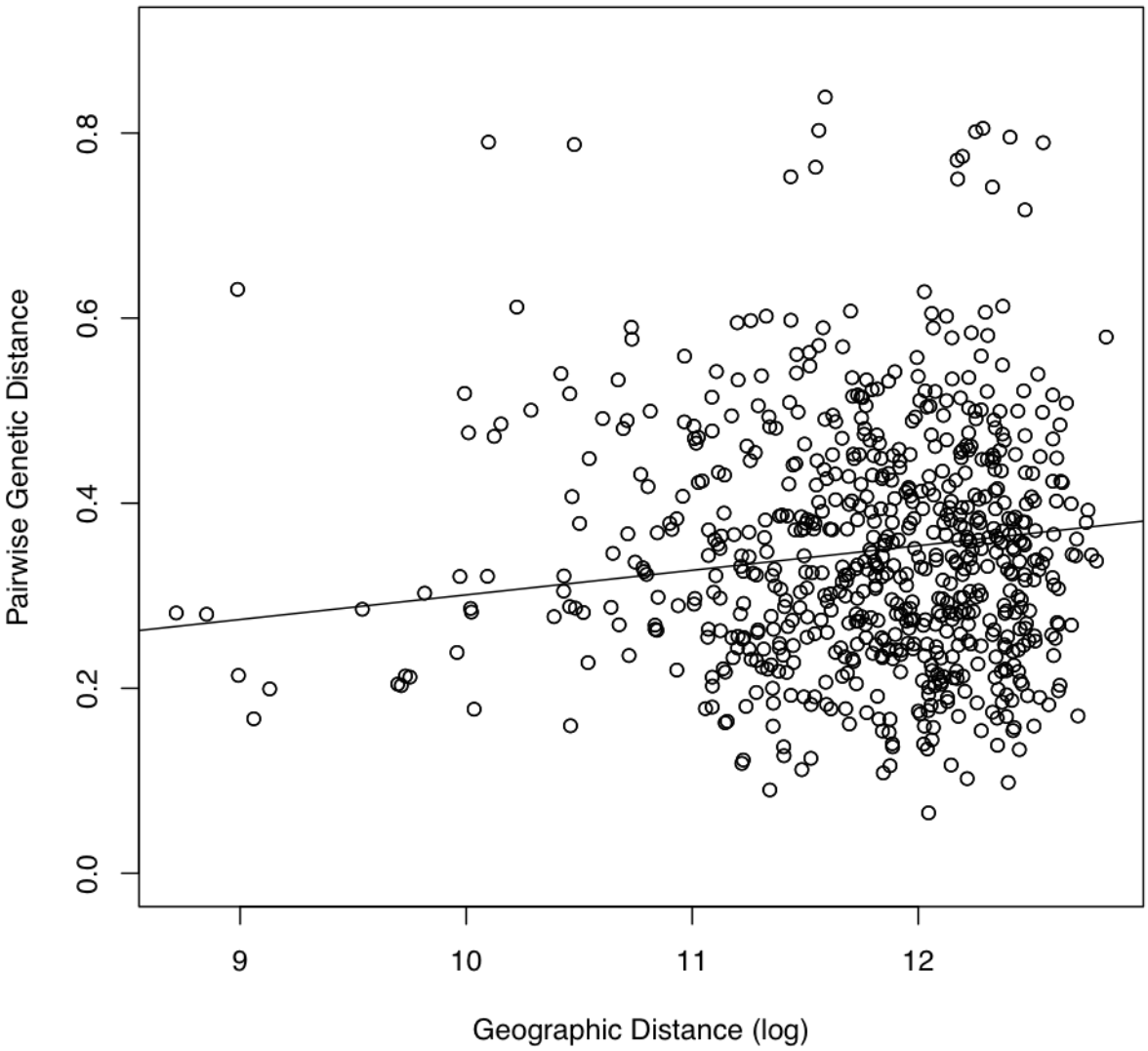

Figure A.5: Genetic distance (Rousett's) across all SNPs vs. log10-transformed geographic distance, showing the Mantel test best-fit line. There was no significant relationship between geographic and genetic distance (Mantel r = 0.002179; p = 0.48)

Appendix B: Supplementary Information to Chapter 3

B.1 Haplotype data

Table B.1: Pack name, haplotypes, and sequence capture information of each individual per population.

Locality	Sample	Pack Name	mtDNA Haplotype	Completed	Note about sequencing
OR	OR 130716WW	Walla Walla	lu32	N/A	
OR	OR 140422WE6	Wenaha	lu32	N/A	
OR	OR W42	Wenaha	lu32	current study	
OR	OR W45	Umatilla River	lu31	N/A	
OR	OR1	N/A	lu32	current study	
OR	OR14	Umatilla River	lu31	current study	
OR	OR2	Imnaha	lu31	current study	
OR	OR3	Imnaha	lu31	N/A	
OR	OR4	Imnaha	lu32	current study	
OR	OR5	Imnaha	lu31	current study	
OR	OR5957	Wenaha	lu32	N/A	
OR	OR6	Imnaha	lu32	N/A	
OR	OR6155	Wenaha	lu32	N/A	

Table B.1 continued from previous page

OR	OR7	Imnaha	lu31	N/A	
OR	OR10	Walla Walla	lu32	current study	
OR	OR11	Walla Walla	lu32	current study	
OR	OR12	Wenaha	lu31	current study	removed from analysis - low coverage
OR	OR15	Snake River	lu32	current study	
OR	OR16	Walla Walla	lu32	current study	
OR	OR17	Imnaha	lu31	current study	
OR	OR18	Snake River	lu32	current study	
OR	OR19	Wenaha	lu32	current study	
OR	OR20	Minam	lu32	current study	
OR	OR21	Wenaha	lu32	current study	
OR	OR22	Umatilla River	lu31	current study	
OR	OR23	Umatilla River	lu31	current study	
OR	OR24	Snake River	lu32	N/A	
OR	OR25	Imnaha	lu31	N/A	
OR	OR26	N/A	lu31	current study	
OR	OR27	Minam	lu32	current study	
OR	OR28	Mt Emily	lu32	current study	removed from analysis - low coverage

Table B.1 continued from previous page

OR	OR9	Imnaha	lu31	N/A	
WA	WAWedge3	Wedge	lu31	N/A	
WA	WAWedge4	Wedge	lu31	N/A	
WA	WAWedge5	Wedge	lu31	N/A	
WA	WAWedge6	Wedge	lu31	N/A	
WA	WAWedge7	Wedge	lu31	N/A	
WA	WAWedge8	Wedge	lu68	current study	
WA	RKW3912	N/A	lu31	current study	
WA	RKW4317	Lookout	lu32	N/A	
WA	RKW4318	Lookout	lu68	current study	
WA	RKW9506	N/A	lu32	current study	
WA	RKW9507	N/A	lu32	current study	
WA	RKW9508	N/A	lu32	current study	
WA	RKW9708	N/A	lu32	N/A	
WA	WA017	N/A	lu31	current study	
WA	WA018	Smackout	lu31	current study	
WA	WA019F	N/A	lu32	current study	removed from analysis - low coverage
WA	WA100	Diamond	lu38	current study	removed from analysis - low coverage

Table B.1 continued from previous page

WA	WA377	Diamond/N/A	lu32	current study
WA	WA378	Diamond/N/A	lu32	current study
WA	WA380	Diamond	lu32	current study
WA	WA382	Diamond	lu32	current study
WA	WA-012F	Lookout/N/A	lu32	current study
WA	WA010817	N/A	N/A	current study
AK	RKW7682	N/A	lu38	Schweizer et al. (2016)
AK	RKW7687	N/A	lu32	N/A
AK	RKW7688	N/A	lu38	Schweizer et al. (2016)
AK	RKW7690	N/A	lu38	N/A
AK	RKW7692	N/A	lu38	N/A
AK	RKW7693	N/A	lu38	N/A
AK	RKW7694	N/A	lu38	N/A
AK	RKW7684	N/A	N/A	Schweizer et al. (2016)

Table B.1 continued from previous page

AK	RKW7686	N/A	N/A	Schweizer et al. (2016)	
cBC	52310	N/A	lu38	N/A	
cBC	52413	N/A	lu38	N/A	
cBC	52813	N/A	lu68	N/A	
cBC	60509	N/A	lu38	N/A	
cBC	6182013	N/A	lu68	N/A	
cBC	RKW3755	N/A	lu68	current study	
cBC	RKW3756	N/A	lu38	N/A	
cBC	RKW3757	N/A	lu38	Schweizer et al. (2016)	
cBC	BCscat1	N/A	lu38	N/A	
cBC	BCscat10	N/A	lu38	current study	
cBC	BCscat12	N/A	lu38	current study	removed from analysis - low coverage (See Fig. S2)
cBC	BCscat13	N/A	lu38	current study	
cBC	BCscat14	N/A	lu38	current study	removed from analysis - low coverage (See Fig. S2)

Table B.1 continued from previous page

cBC	BCscat18	N/A	lu38	N/A	
cBC	BCscat2	N/A	lu38	current study	
cBC	BCscat4	N/A	lu38	N/A	
cBC	BCscat5	N/A	lu68	current study	removed from analysis - low coverage (See Fig. S2)
cBC	BCscat9	N/A	lu32	N/A	
cBC	Ye01	N/A	lu38	N/A	
cBC	Ye09	N/A	lu38	N/A	
cBC	Ye10	N/A	lu38	N/A	
cBC	Ye102	N/A	lu38	N/A	
cBC	bone	N/A	N/A	current study	removed from analysis - low coverage
AB	RKW2030	N/A	N/A	Schweizer et al. (2016)	
AB	RKW3116	N/A	N/A	Schweizer et al. (2016)	
AB	RKW2040	N/A	N/A	Schweizer et al. (2016)	

Table B.1 continued from previous page

AB	RKW2057	N/A	N/A	Schweizer et al. (2016)
AB	RKW2026	N/A	N/A	Schweizer et al. (2016)
AB	RKW2070	N/A	N/A	Schweizer et al. (2016)
AB	RKW2075	N/A	N/A	Schweizer et al. (2016)
BC	BC08301	N/A	N/A	current study
BC	BC0803	N/A	N/A	current study
BC	RKW5271	N/A	N/A	current study
BC	RKW5270	N/A	N/A	current study
BC	RKW5269	N/A	N/A	current study
ID	RKW1126	N/A	lu31	current study
ID	RKW1127	N/A	lu31	current study
ID	RKW1134	N/A	lu28	N/A
ID	RKW1135	N/A	lu28	current study

Table B.1 continued from previous page

ID	RKW1139	N/A	lu31	current study	
ID	RKW1140	N/A	lu32	current study	
ID	RKW1215	N/A	lu61	N/A	
ID	RKW1601	N/A	lu38	N/A	
ID	RKW1604	N/A	lu31	N/A	
ID	RKW1605	N/A	lu31	N/A	
ID	RKW1607	N/A	lu28	current study	
ID	WL1317	N/A	N/A	current study	
ID	IFG1	N/A	N/A	current study	
ID	RKW5299	N/A	N/A	current study	
ID	WL13N1	N/A	N/A	current study	
ID	RKW5297	N/A	N/A	current study	
ID	WL1121	N/A	N/A	current study	
ID	WL1281	N/A	N/A	current study	
ID	B400	N/A	N/A	current study	
ID	RKW5298	N/A	N/A	current study	
ID	RKW5294	N/A	N/A	current study	
ID	RKW5300	N/A	N/A	current study	removed from analysis - low coverage

Table B.1 continued from previous page

MT	RKW1002	N/A	lu32	N/A
MT	RKW1005	N/A	lu32	current study
MT	RKW1006	N/A	lu32	N/A
MT	RKW1007	N/A	lu32	current study
MT	RKW1008	N/A	lu32	N/A
MT	RKW1009	N/A	lu32	N/A
MT	RKW1010	N/A	lu32	N/A
MT	RKW1013	N/A	lu31	current study
MT	RKW1014	N/A	lu32	N/A
MT	RKW1015	N/A	lu28	N/A
MT	RKW1020	N/A	lu32	N/A
MT	RKW1021	N/A	lu32	N/A
MT	RKW1022	N/A	lu32	N/A
MT	RKW1033	N/A	lu31	N/A
MT	RKW1041	N/A	lu32	current study
MT	RKW1047	N/A	lu31	current study
MT	RKW1051	N/A	lu31	N/A
MT	RKW1058	N/A	lu31	N/A

Table B.1 continued from previous page

MT	RKW1059	N/A	lu31	current study
MT	RKW1062	N/A	lu32	N/A
MT	RKW1064	N/A	lu31	N/A
MT	RKW1071	N/A	lu31	N/A
MT	RKW1094	N/A	lu32	N/A
MT	RKW1097	N/A	lu31	N/A
MT	RKW1643	N/A	lu28	N/A
MT	RKW1651	N/A	lu31	N/A
MT	RKW1653	N/A	lu32	current study
MT	RKW1658	N/A	lu31	current study
MT	RKW1028	N/A	N/A	current study
MT	RKW1049	N/A	N/A	current study
MT	RKW1004	N/A	N/A	current study
MT	RKW1080	N/A	N/A	current study
MT	RKW1034	N/A	N/A	current study
MT	RKW1086	N/A	N/A	current study
MT	RKW1654	N/A	N/A	current study
MT	RKW1649	N/A	N/A	current study

Table B.1 continued from previous page

MT	RKW1030	N/A	N/A	current study
MT	RKW1087	N/A	N/A	current study
MT	RKW1012	N/A	N/A	current study
MT	RKW1035	N/A	N/A	current study
MT	RKW1031	N/A	N/A	current study
MT	RKW1096	N/A	N/A	current study
MT	RKW1060	N/A	N/A	current study
YNP	RKW1263	N/A	lu31	current study
YNP	RKW14281	N/A	lu32	N/A
YNP	RKW14287	N/A	lu31	N/A
YNP	RKW14290	N/A	lu31	N/A
YNP	RKW14296	N/A	lu32	N/A
YNP	RKW14303	N/A	lu31	N/A
YNP	RKW14308	N/A	lu38	N/A
YNP	RKW14309	N/A	lu38	N/A
YNP	RKW14310	N/A	lu31	N/A
YNP	RKW14311	N/A	lu31	N/A
YNP	RKW14313	N/A	lu32	N/A

Table B.1 continued from previous page

YNP	RKW14330	N/A	lu31	N/A
YNP	RKW852	N/A	lu38	current study
YNP	RKW968	N/A	lu32	N/A
YNP	RKW1267	Nez Perce	lu28	N/A
YNP	RKW861	Nez Perce	lu28	N/A
YNP	RKW862	Nez Perce	lu28	N/A
YNP	RKW874	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW875	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW667	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW899	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW1479	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW1552	N/A	N/A	Schweizer et al. (in prep)

Table B.1 continued from previous page

YNP	RKW965	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW958	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW13498	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW12538	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW1107	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW12532	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW12535	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW13497	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW13499	N/A	N/A	Schweizer et al. (in prep)

Table B.1 continued from previous page

YNP	RKW1547	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW3622	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW642	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW893	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW917	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW921	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW922	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW951	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW962	N/A	N/A	Schweizer et al. (in prep)

Table B.1 continued from previous page

YNP	RKW977	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW780	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW7515	N/A	N/A	Schweizer et al. (in prep)
YNP	RKW1474	N/A	N/A	current study
YNP	RKW1471	N/A	N/A	current study
YNP	RKW639	N/A	N/A	current study
YNP	RKW969	N/A	N/A	current study

Table B.2: Haplotypes and number of individuals per haplotype found in each population. Distribution of mtDNA control region sequence haplotypes among wolves of North America from previous studies were included as a comparison. Population abbreviations as follows: British Columbia: BC; Idaho: ID; Montana: MT; Oregon: OR; Washington: WA; Yellowstone National Park: YNP; Alaska: AK; Southeastern Alaska: SEAK; Yukon: YUK; Northwest Territories: NWT; Alberta: ALTA; Saskatchewan: SASK; Manitoba: MAN; Great lake States: GLS; Eastern Ontario: EONT.

GenBank Accession	Haplotype	cBC	ID	MT	OR	WA	YNP	Total	Haplotype Localities
FM201759	lu28	0	3	2	0	0	3	8	Coastal AK; Interior AK, YUK, NWT, BC, MAN, GLS, EONT
AF005312	lu31	0	5	11	13	8	7	44	Interior AK, Yukon, NWT, BC, ALTA
AF005309	lu32	2	1	15	19	11	4	52	Coastal BC; Interior NWUS, AK, NWT, BC, ALTA, SASK, MAN, GLS, EONT
AF812731	lu38	23	1	0	0	1	3	28	Coastal SEAK and BC; Interior AK, NWT, BC
AY812741	lu61	0	1	0	0	0	0	1	Coastal AK; Interior AK, NWT, SASK, MAN
FN298179	lu68	4	0	0	0	2	0	6	Coastal SEAK and BC
	Total	29	11	28	32	22	17	139	

B.2 Depth of coverage for capture sequencing

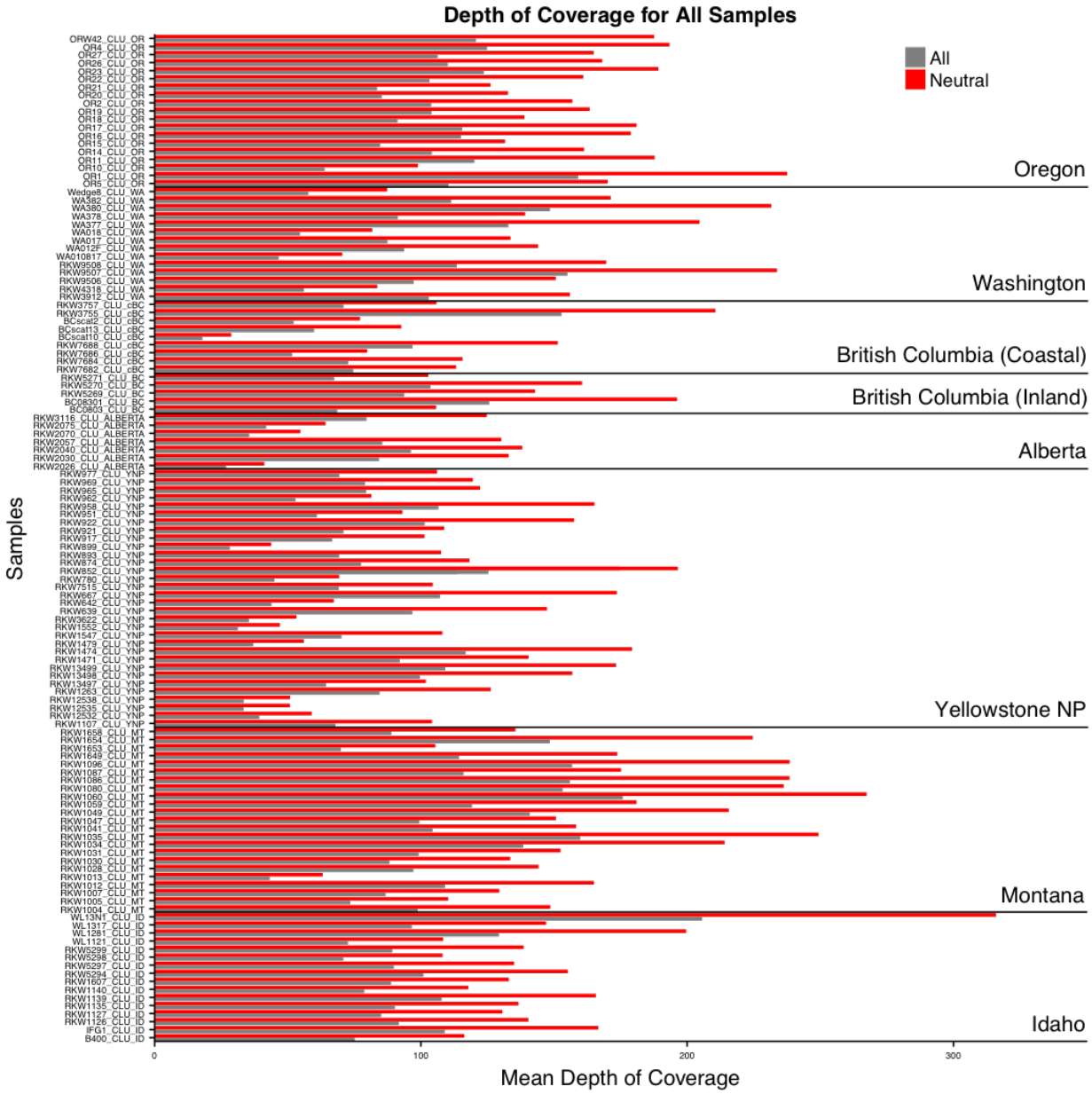

Figure B.1: Mean depth of coverage of neutral (red) and all (gray) capture regions for 126 wolves from the Pacific Northwest region of the U.S. included in clustering analyses.

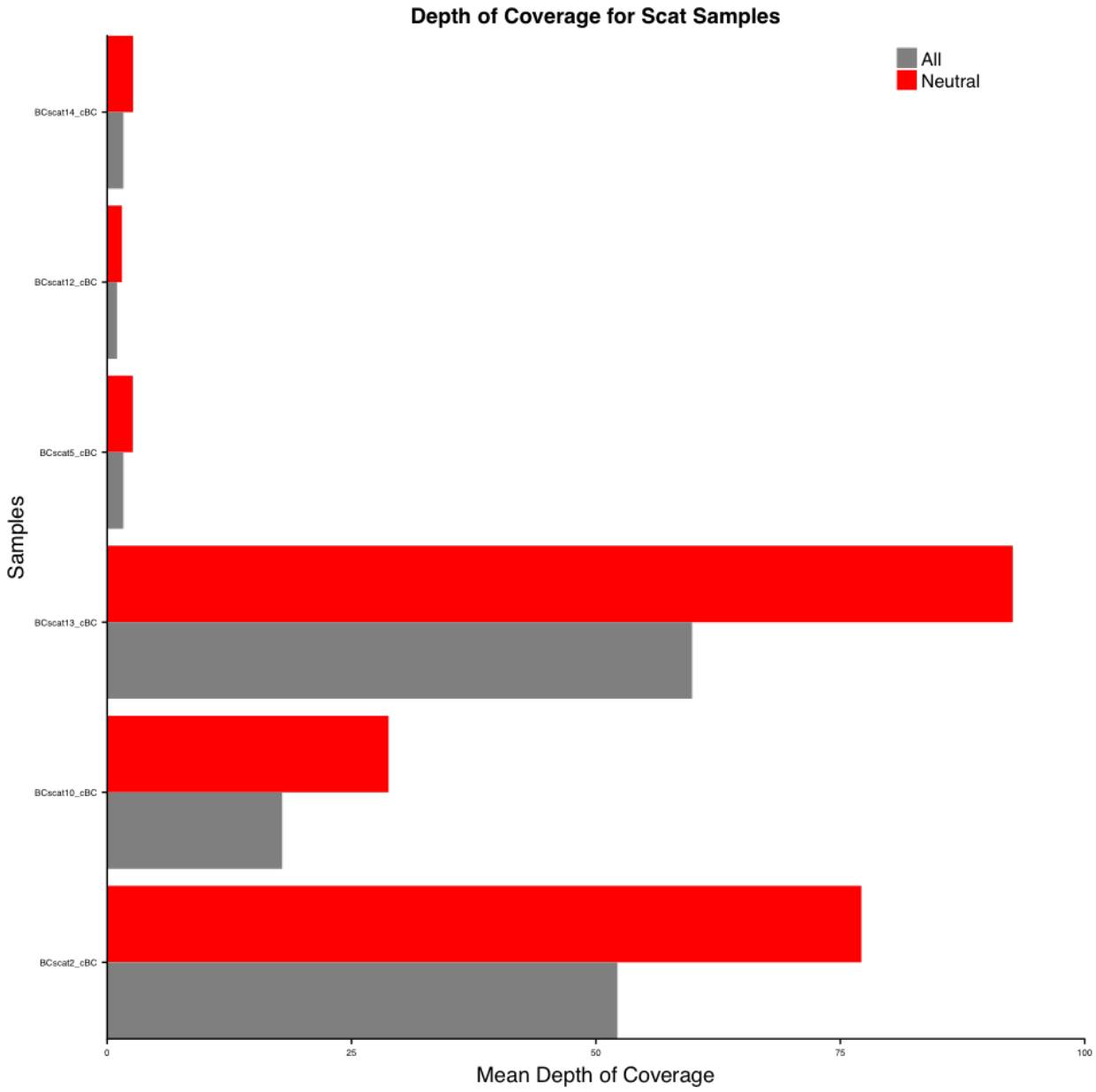

Figure B.2: Mean depth of coverage of neutral (red) and all (gray) capture regions for 6 wolves from the coastal British Columbia using fecal samples.

B.3 Use of scat samples

We successfully used DNA extracted from non-invasively collected faecal samples in our study. However, we chose to remove three low-quality samples from further analyses. These samples with low coverage (Figure S2) may have been dominated by DNA from exogenous sources, particularly gut bacteria. Additionally, DNA extracted from scat may contain chemicals that inhibit PCR (Kohn and Wayne, 1997; Nechvatal et al., 2008). DNA extraction and library preparation using a modified protocol (e.g. Perry et al., 2010) may improve sequencing coverage of faecal samples. Methodological improvements may to be required for many population level questions, particularly of threatened, endangered, or elusive species, due to the necessity of non-invasive sampling. For example, owing to respect of local Indigenous protocols, cBC rainforest wolves are protected against capture and invasive sampling and can only be studied via non-invasive sampling; thus, improvement of DNA extraction of endogenous cells from scat as well as modified library preparation would enhance high-throughput sequencing capabilities.

B.4 Inbreeding coefficients

Table B.3: Inbreeding coefficients (F_{IS}) for 18,508 LD-pruned SNPs within neutral regions in 87 unrelated individuals. Population abbreviations as follows: British Columbia: BC; Idaho: ID; Montana: MT; Oregon: OR; Washington: WA; Yellowstone National Park: YNP.

Population	F (stdev)
Alberta	-0.0021 (0.0118)
interior BC	-0.0015 (0.0348)
coastal BC	0.1116 (0.1557)
ID	0.0081 (0.0535)
MT	0.0153 (0.1272)
OR	-0.0115 (0.0472)
WA	0.0425 (0.0896)
YNP	0.0139 (0.0893)

B.5 Population structure using Principal components analysis

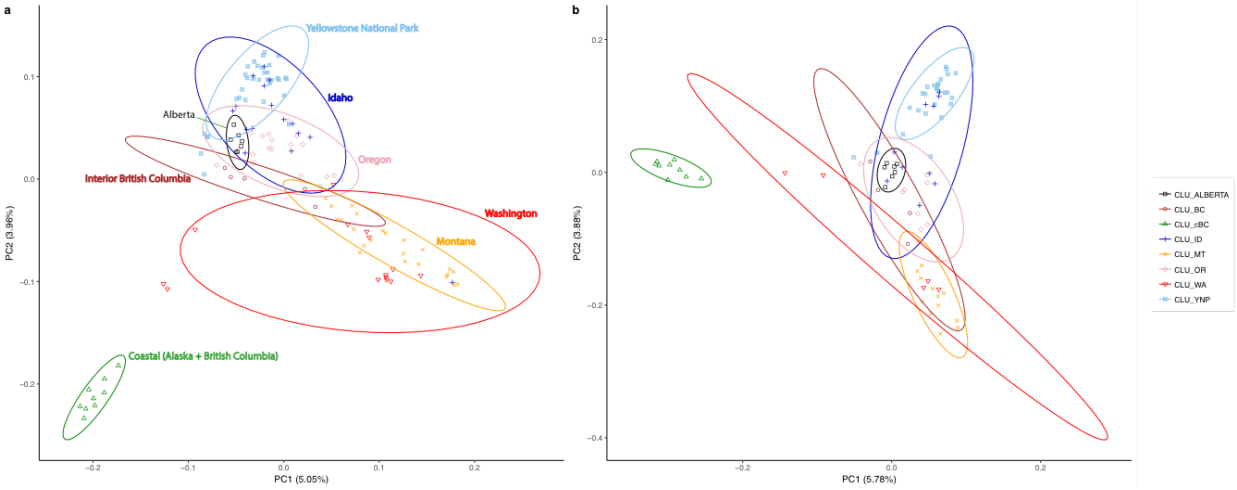

Figure B.3: a) PCA plot of all 126 wolves for LD-pruned dataset within neutral regions with 95% confidence intervals. b) PCA plot of 83 unrelated wolves for LD-pruned dataset within neutral regions with 95% confidence intervals. Population abbreviations as follows: Interior British Columbia: BC; coastal British Columbia: cBC; Idaho: ID; Montana: MT; Oregon: OR; Washington: WA; Yellowstone National Park: YNP.

B.6 Population structure using ADMIXTURE

Table B.4: Pack name, haplotype, and admixture assignments per population.

Locality	Sample	Pack Name	Haplotype	Admixture assignment (K=4)				Founder Status
OR	OR10	Walla Walla	lu32	0.602561	0.153906	0.243523	0.00001	
OR	OR27	Minam	lu32	0.590774	0.215957	0.193259	0.00001	
OR	OR2	Imnaha	lu31	0.473206	0.04935	0.477433	0.00001	
OR	OR4	Imnaha	lu32	0.137707	0.144992	0.476168	0.241132	
OR	OR19	Wenaha	lu32	0.248426	0.250237	0.480052	0.021285	
OR	OR1	N/A	lu32	0.247454	0.132934	0.619602	0.00001	
OR	OR18	Snake River	lu32	0.359847	0.358424	0.281719	0.00001	
OR	OR14	Umatilla River	lu31	0.258428	0.24341	0.498152	0.00001	
WA	RKW3912	N/A	lu31	0.99997	0.00001	0.00001	0.00001	
WA	RKW9506	N/A	lu32	0.99997	0.00001	0.00001	0.00001	
WA	WA017	N/A	lu31	0.951354	0.048626	0.00001	0.00001	
WA	RKW4318	Lookout	lu68	0.05421	0.00001	0.452044	0.493737	
WA	WAWedge8	Wedge	lu68	0.109309	0.00001	0.538764	0.351917	

Table B.4 continued from previous page

AK	RKW7688	N/A	lu38	0.00001	0.00001	0.00001	0.99997
AK	RKW7684	N/A	N/A	0.00001	0.00001	0.00001	0.99997
AK	RKW7686	N/A	N/A	0.00001	0.00001	0.00001	0.99997
AK	RKW7682	N/A	lu38	0.00001	0.00001	0.00001	0.99997
cBC	RKW3757	N/A	lu38	0.00001	0.00001	0.00001	0.99997
cBC	BCscat10	N/A	lu38	0.00001	0.00001	0.00001	0.99997
cBC	BCscat2	N/A	lu38	0.00001	0.00001	0.00001	0.99997
cBC	BCscat13	N/A	lu38	0.00001	0.00001	0.00001	0.99997
cBC	RKW3755	N/A	lu68	0.00001	0.00001	0.00001	0.99997
AB	RKW2030	N/A	N/A	0.016194	0.00001	0.983786	0.00001
AB	RKW3116	N/A	N/A	0.00001	0.00001	0.99997	0.00001
AB	RKW2040	N/A	N/A	0.014861	0.00001	0.985119	0.00001
AB	RKW2057	N/A	N/A	0.00001	0.00001	0.99997	0.00001
AB	RKW2026	N/A	N/A	0.00001	0.00001	0.99997	0.00001
AB	RKW2070	N/A	N/A	0.00001	0.00001	0.99997	0.00001
AB	RKW2075	N/A	N/A	0.00001	0.00001	0.99997	0.00001

Table B.4 continued from previous page

BC	BC08301	N/A	N/A	0.56104	0.00001	0.403048	0.035901	
BC	BC0803	N/A	N/A	0.458043	0.132384	0.365361	0.044212	
BC	RKW5271	N/A	N/A	0.133118	0.009911	0.789093	0.067877	
BC	RKW5270	N/A	N/A	0.118274	0.080279	0.624321	0.177125	
ID	WL1317	N/A	N/A	0.331508	0.144623	0.523858	0.00001	
ID	RKW1126	N/A	lu31	0.114843	0.00001	0.885137	0.00001	Founder
ID	RKW1135	N/A	lu28	0.05579	0.00001	0.94419	0.00001	Founder
ID	IFG1	N/A	N/A	0.360076	0.288705	0.351209	0.00001	
ID	RKW1127	N/A	lu31	0.00001	0.00001	0.99997	0.00001	Founder
ID	RKW5299	N/A	N/A	0.00001	0.590525	0.409455	0.00001	
ID	WL13N1	N/A	N/A	0.057314	0.818366	0.124311	0.00001	
ID	RKW5297	N/A	N/A	0.00001	0.739657	0.260323	0.00001	
ID	WL1121	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
ID	WL1281	N/A	N/A	0.298443	0.383523	0.318024	0.00001	
MT	RKW1007	N/A	lu32	0.99997	0.00001	0.00001	0.00001	

Table B.4 continued from previous page

MT	RKW1028	N/A	N/A	0.99997	0.00001	0.00001	0.00001	
MT	RKW1013	N/A	lu31	0.99997	0.00001	0.00001	0.00001	
MT	RKW1049	N/A	N/A	0.99997	0.00001	0.00001	0.00001	
MT	RKW1653	N/A	lu32	0.99997	0.00001	0.00001	0.00001	
MT	RKW1004	N/A	N/A	0.99997	0.00001	0.00001	0.00001	
MT	RKW1080	N/A	N/A	0.99997	0.00001	0.00001	0.00001	
MT	RKW1034	N/A	N/A	0.99997	0.00001	0.00001	0.00001	
MT	RKW1059	N/A	lu31	0.99997	0.00001	0.00001	0.00001	
MT	RKW1086	N/A	N/A	0.817028	0.00001	0.182952	0.00001	
MT	RKW1654	N/A	N/A	0.773256	0.029789	0.17958	0.017375	
MT	RKW1649	N/A	N/A	0.716509	0.008057	0.255262	0.020173	
MT	RKW1030	N/A	N/A	0.569627	0.082211	0.316299	0.031863	
YNP	RKW852	N/A	lu38	0.01205	0.00001	0.870492	0.117448	Founder
YNP	RKW874	N/A	N/A	0.263089	0.493429	0.243472	0.00001	
YNP	RKW667	N/A	N/A	0.00001	0.095502	0.904478	0.00001	
YNP	RKW899	N/A	N/A	0.132947	0.867033	0.00001	0.00001	
YNP	RKW1479	N/A	N/A	0.00001	0.307759	0.692221	0.00001	Founder

Table B.4 continued from previous page

YNP	RKW969	N/A	N/A	0.00001	0.32853	0.67145	0.00001	Founder
YNP	RKW1552	N/A	N/A	0.068363	0.907378	0.00001	0.024249	
YNP	RKW965	N/A	N/A	0.00001	0.616178	0.383802	0.00001	
YNP	RKW958	N/A	N/A	0.00001	0.630553	0.369427	0.00001	
YNP	RKW13498	N/A	N/A	0.00001	0.73251	0.26747	0.00001	
YNP	RKW12538	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW1107	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW12532	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW12535	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW13497	N/A	N/A	0.00001	0.82146	0.17852	0.00001	
YNP	RKW13499	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW1547	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW3622	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW642	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW893	N/A	N/A	0.00001	0.627405	0.372575	0.00001	
YNP	RKW917	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW921	N/A	N/A	0.00001	0.99997	0.00001	0.00001	
YNP	RKW922	N/A	N/A	0.00001	0.99997	0.00001	0.00001	

Table B.4 continued from previous page

YNP	RKW951	N/A	N/A	0.00001	0.99997	0.00001	0.00001
YNP	RKW962	N/A	N/A	0.00001	0.99997	0.00001	0.00001
YNP	RKW977	N/A	N/A	0.00001	0.99997	0.00001	0.00001
YNP	RKW780	N/A	N/A	0.037012	0.00001	0.811201	0.151777

Table B.5: Pairwise relatedness values for individuals from different geographic locations. Relatedness was calculated from the LD-pruned SNP set, and pairs were filtered to have relatedness above 0.1 and be from different populations.

Individual 1	Pack 1	Population 1	Individual 2	Pack 2	Population 2	SNPs	Relatedness
RKW1096_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26255	0.571
RKW639_YNP	Chief Joseph	CLU_YNP	RKW1139_ID	0	CLU_ID	26250	0.5606
OR11_OR	0	CLU_OR	WA010817_WA	0	CLU_WA	26098	0.5434
RKW639_YNP	Chief Joseph	CLU_YNP	RKW1127_ID	0	CLU_ID	26231	0.5324
OR16_OR	0	CLU_OR	WA010817_WA	0	CLU_WA	26093	0.4518
OR10_OR	Walla Walla	CLU_OR	WA010817_WA	0	CLU_WA	25918	0.4294
RKW1012_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26239	0.395
RKW1087_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26246	0.388
RKW1031_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26249	0.383
RKW1060_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26254	0.3328
RKW1087_MT	0	CLU_MT	WA018_WA	0	CLU_WA	26209	0.3016
RKW1013_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26173	0.2824
B400_ID	0	CLU_ID	OR2_OR	Imnaha	CLU_OR	26153	0.2564
RKW1047_MT	0	CLU_MT	RKW1140_ID	0	CLU_ID	26249	0.2456
RKW1096_MT	0	CLU_MT	RKW9506_WA	0	CLU_WA	26217	0.2298
OR1_OR	0	CLU_OR	RKW1139_ID	0	CLU_ID	26259	0.21

Table B.5 continued from previous page

RKW667_CLU_YNP	Swan Lake	CLU_YNP	RKW1139_ID	0	CLU_ID	26195	0.203
RKW1552_CLU_YNP	Gibbon	CLU_YNP	WL1121_ID	0	CLU_ID	25937	0.1888
IFG1_ID	0	CLU_ID	OR16_OR	0	CLU_OR	26259	0.1854
OR1_OR	0	CLU_OR	RKW1607_ID	0	CLU_ID	26258	0.1824
RKW1086_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26247	0.179
RKW1658_MT	0	CLU_MT	RKW3912_WA	0	CLU_WA	26233	0.177
RKW1140_ID	0	CLU_ID	RKW1653_MT	0	CLU_MT	26237	0.1746
OR16_OR	0	CLU_OR	RKW5294_ID	0	CLU_ID	26255	0.1726
OR1_OR	0	CLU_OR	RKW969_YNP	Druid	CLU_YNP	26246	0.1712
RKW965_CLU_YNP	Geode	CLU_YNP	RKW5297_ID	0	CLU_ID	26226	0.169
RKW958_CLU_YNP	Leopold	CLU_YNP	RKW1127_ID	0	CLU_ID	26225	0.1626
RKW1140_ID	0	CLU_ID	RKW9506_WA	0	CLU_WA	26209	0.1618
OR16_OR	0	CLU_OR	RKW1653_MT	0	CLU_MT	26243	0.1606
OR1_OR	0	CLU_OR	RKW5294_ID	0	CLU_ID	26257	0.1596
RKW1012_MT	0	CLU_MT	RKW9506_WA	0	CLU_WA	26202	0.1586
IFG1_ID	0	CLU_ID	OR1_OR	0	CLU_OR	26261	0.1584
OR16_OR	0	CLU_OR	RKW1028_MT	0	CLU_MT	26259	0.1548
RKW921_CLU_YNP	Rose Crk II	CLU_YNP	RKW5297_ID	0	CLU_ID	26232	0.154

Table B.5 continued from previous page

RKW667_CLU_YNP	Swan Lake	CLU_YNP	RKW1127_ID	0	CLU_ID	26178	0.1524
RKW1004_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26241	0.1524
RKW1654_MT	0	CLU_MT	RKW3912_WA	0	CLU_WA	26244	0.1516
RKW958_CLU_YNP	Leopold	CLU_YNP	RKW1139_ID	0	CLU_ID	26242	0.1492
RKW969_YNP	Druid	CLU_YNP	RKW5294_ID	0	CLU_ID	26241	0.1492
ORW42_OR	0	CLU_OR	RKW1126_ID	0	CLU_ID	26235	0.1394
IFG1_ID	0	CLU_ID	WA010817_WA	0	CLU_WA	26094	0.1372
RKW965_CLU_YNP	Geode	CLU_YNP	RKW1139_ID	0	CLU_ID	26227	0.1368
OR1_OR	0	CLU_OR	RKW1126_ID	0	CLU_ID	26251	0.1354
OR16_OR	0	CLU_OR	RKW1086_MT	0	CLU_MT	26264	0.1342
OR1_OR	0	CLU_OR	RKW639_YNP	Chief Joseph	CLU_YNP	26250	0.1314
RKW893_CLU_YNP	Druid	CLU_YNP	RKW5297_ID	0	CLU_ID	26189	0.1306
BC08301_BC	0	CLU_BC	WA018_WA	0	CLU_WA	26186	0.13
RKW969_YNP	Druid	CLU_YNP	RKW5297_ID	0	CLU_ID	26242	0.1286
RKW893_CLU_YNP	Druid	CLU_YNP	RKW1139_ID	0	CLU_ID	26191	0.127
RKW1653_MT	0	CLU_MT	WA377_WA	0	CLU_WA	26241	0.1256
B400_ID	0	CLU_ID	OR26_OR	0	CLU_OR	26163	0.1238
RKW1004_MT	0	CLU_MT	RKW3912_WA	0	CLU_WA	26240	0.1238

Table B.5 continued from previous page

OR18_OR	Snake River	CLU_OR	RKW1126_ID	0	CLU_ID	26241	0.1218
OR10_OR	Walla Walla	CLU_OR	RKW1653_MT	0	CLU_MT	26059	0.1212
RKW1004_MT	0	CLU_MT	WA377_WA	0	CLU_WA	26256	0.1176
IFG1_ID	0	CLU_ID	OR11_OR	0	CLU_OR	26264	0.1172
OR23_OR	0	CLU_OR	RKW1139_ID	0	CLU_ID	26260	0.1172
OR1_OR	0	CLU_OR	RKW1127_ID	0	CLU_ID	26242	0.117
RKW1653_MT	0	CLU_MT	WA010817_WA	0	CLU_WA	26082	0.1168
IFG1_ID	0	CLU_ID	OR10_OR	Walla Walla	CLU_OR	26073	0.116
RKW1474_YNP	Nez Perce	CLU_YNP	RKW5298_ID	0	CLU_ID	26235	0.1156
RKW1031_MT	0	CLU_MT	RKW9506_WA	0	CLU_WA	26212	0.1138
OR26_OR	0	CLU_OR	RKW5294_ID	0	CLU_ID	26262	0.1118
OR1_OR	0	CLU_OR	RKW1135_ID	0	CLU_ID	26253	0.1116
RKW1012_MT	0	CLU_MT	WA377_WA	0	CLU_WA	26249	0.1098
RKW1012_MT	0	CLU_MT	WA378_WA	0	CLU_WA	26237	0.109
IFG1_ID	0	CLU_ID	RKW921_CLU_YNP	Rose Crk II	CLU_YNP	26235	0.1076
RKW958_CLU_YNP	Leopold	CLU_YNP	RKW5299_ID	0	CLU_ID	26221	0.1066
RKW1005_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26225	0.1064
OR2_OR	Imnaha	CLU_OR	RKW1126_ID	0	CLU_ID	26245	0.1062

Table B.5 continued from previous page

RKW1653_MT	0	CLU_MT	RKW9506_WA	0	CLU_WA	26201	0.1062
RKW893_CLU_YNP	Druid	CLU_YNP	RKW5299_ID	0	CLU_ID	26168	0.1056
RKW1012_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26240	0.1056
RKW1653_MT	0	CLU_MT	WA017_WA	0	CLU_WA	26226	0.1022
RKW965_CLU_YNP	Geode	CLU_YNP	RKW5299_ID	0	CLU_ID	26206	0.1004
OR26_OR	0	CLU_OR	RKW1086_MT	0	CLU_MT	26271	0.1002

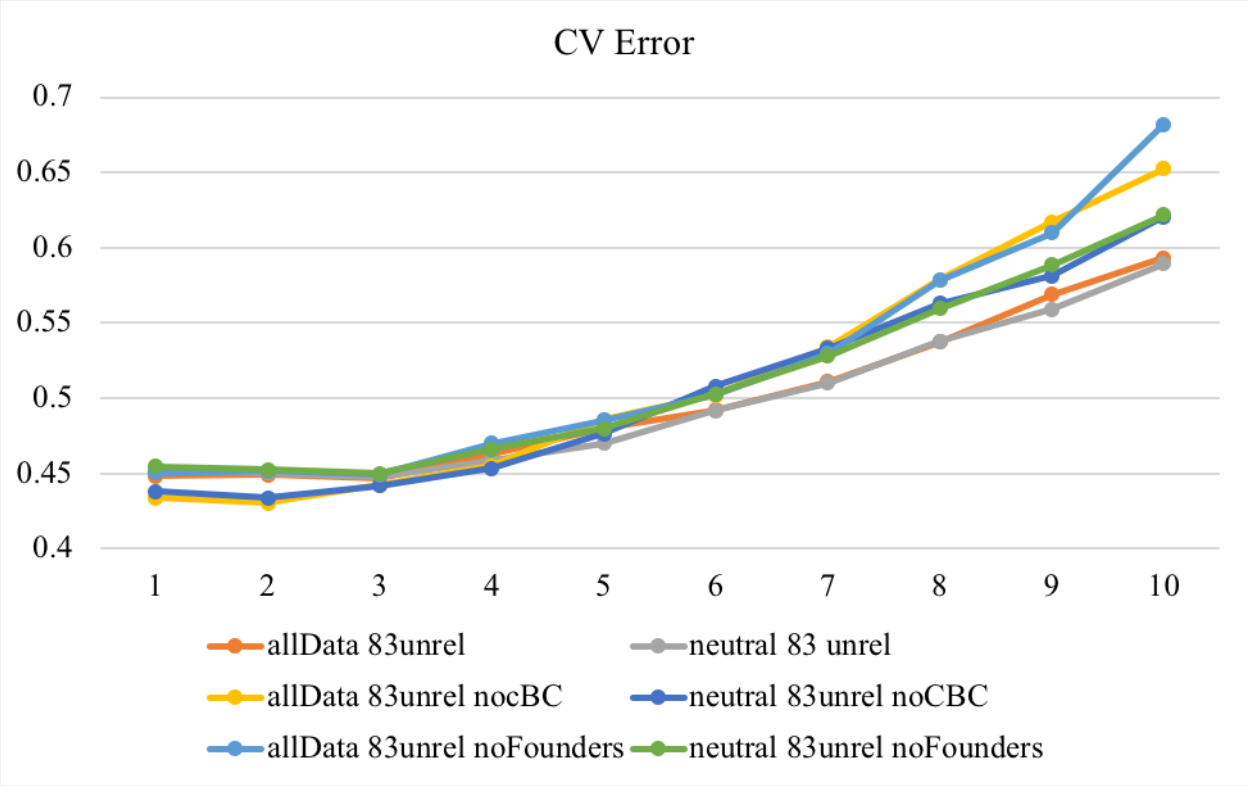

Figure B.4: Cross validation error values from ADMIXTURE results for all data, 83 unrelated individuals (orange); neutral data, 83 unrelated individuals (gray); all data, unrelated individuals excluding cBC (yellow); neutral data, unrelated individuals excluding cBC (dark blue); all data, unrelated individuals excluding founders (light blue); neutral data, unrelated individuals excluding founders (green). Excluding the cBC population reduced the optimal K value suggesting the ecotype is a valid cluster., while excluding founders did not affect the optimal K value.

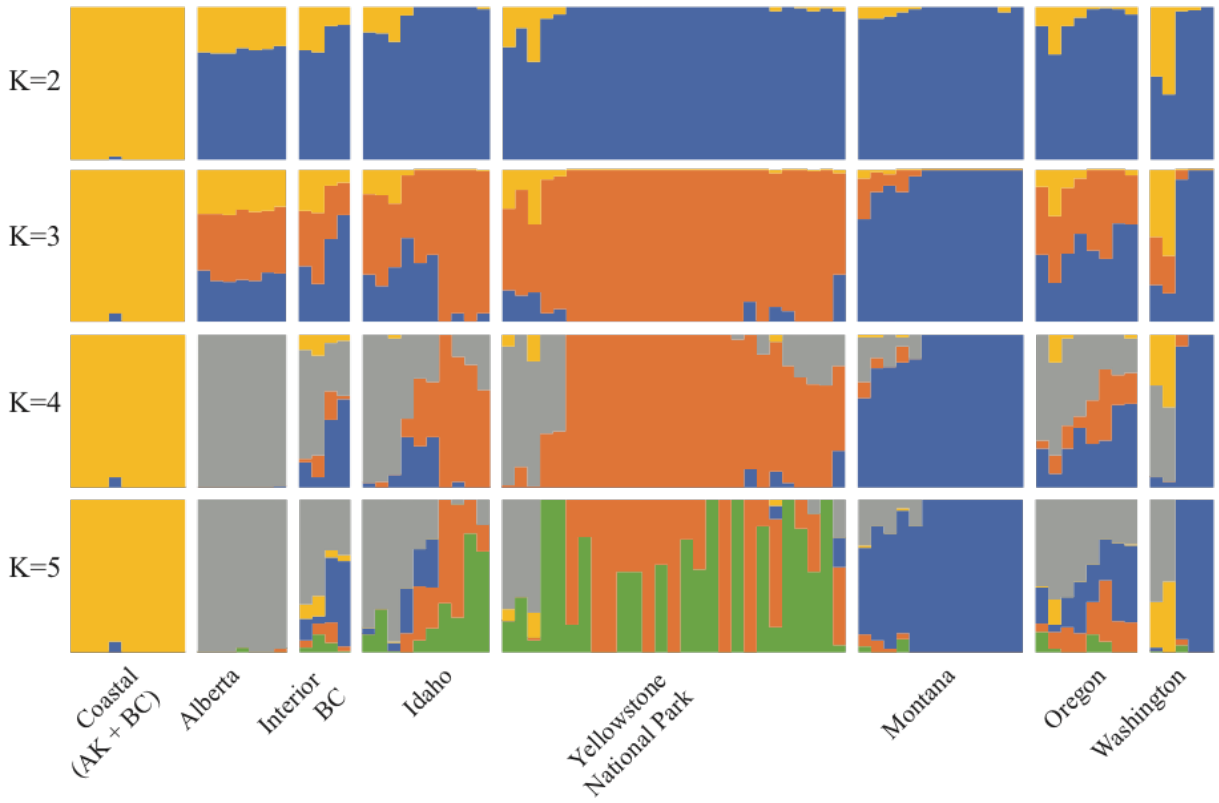

Figure B.5: Population assignment at K=2 to K=5 for 87 unrelated individuals, as determined by running Admixture on a set of 18,508 LD-pruned SNPs within neutral regions.

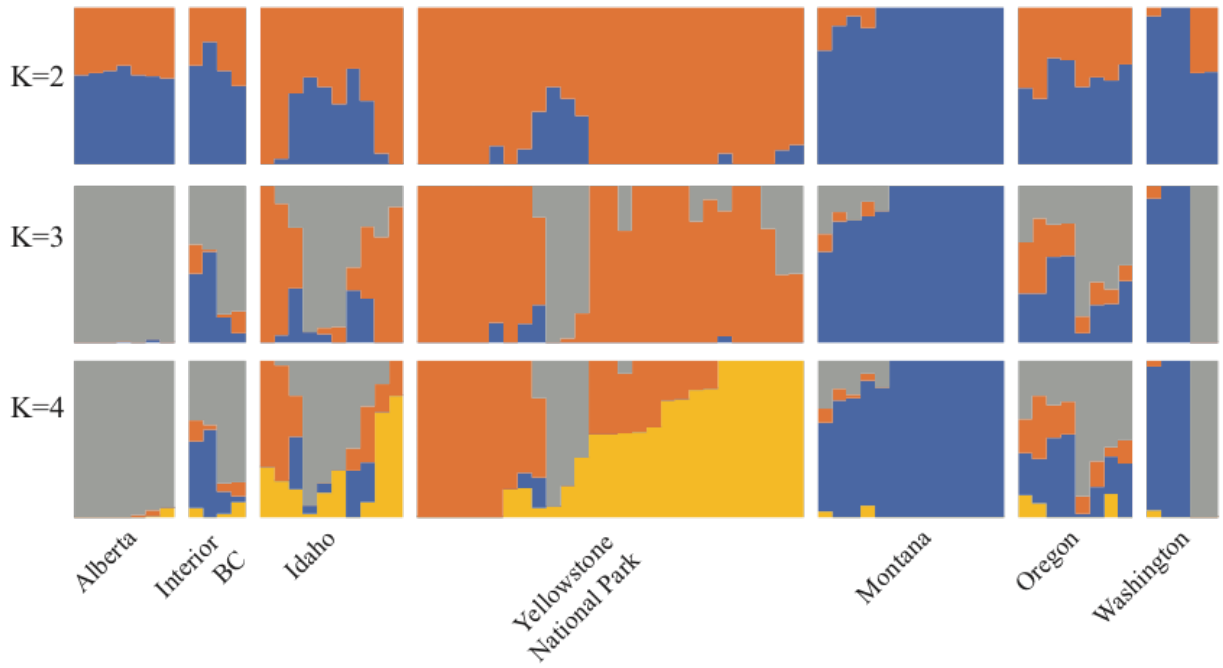

Figure B.6: Population assignment at K=2 to K=4 for 74 unrelated individuals (after removing coastal wolves), as determined by running Admixture on a set of 18,508 LD-pruned SNPs within neutral regions.

B.7 Migration rates using BayesAss

Table B.6: Non-symmetrial gene flow estimates from BayesAss analysis are shown. Grouping are based on the ADMIXTURE results of K=4. The rows list the populations to which each individual migrated and the columns list the populations from which the individuals migrated. Means of the posterior distributions of migration rate are shown. Standard deviation for all distributions were <0.07 and are shown in parentheses. Population abbreviations as follows: coastal: British Columbia and Alaska; interior: Alberta and British Columbia; reintroduced: Yellowstone National Park and Idaho; Montana: MT; Oregon: OR; Washington: WA.

		From:					
		coastal	interior	reintroduced	MT	OR	WA
To:	coastal	0.8887 (0.0393)	0.0223 (0.0208)	0.0224 (0.0209)	0.0222 (0.0208)	0.0221 (0.0208)	0.0224 (0.0210)
	interior	0.0185 (0.0175)	0.7200 (0.0686)	0.1693 (0.0734)	0.0550 (0.0283)	0.0185 (0.0175)	0.0186 (0.0177)
	reintroduced	0.0079 (0.0077)	0.0159 (0.0109)	0.9209 (0.0217)	0.0236 (0.0132)	0.0238 (0.0130)	0.0079 (0.0078)
	MT	0.0115 (0.0112)	0.0116 (0.0112)	0.0304 (0.0183)	0.9236 (0.0262)	0.0114 (0.0111)	0.0114 (0.0110)
	OR	0.0133 (0.0128)	0.0134 (0.0128)	0.0713 (0.0276)	0.0622 (0.0262)	0.8219 (0.0332)	0.0179 (0.0161)
	WA	0.0502 (0.0261)	0.0333 (0.0218)	0.0168 (0.0160)	0.0667 (0.0290)	0.0332 (0.0216)	0.7997 (0.0355)

B.8 Ecological Niche Modelling using MaxEnt

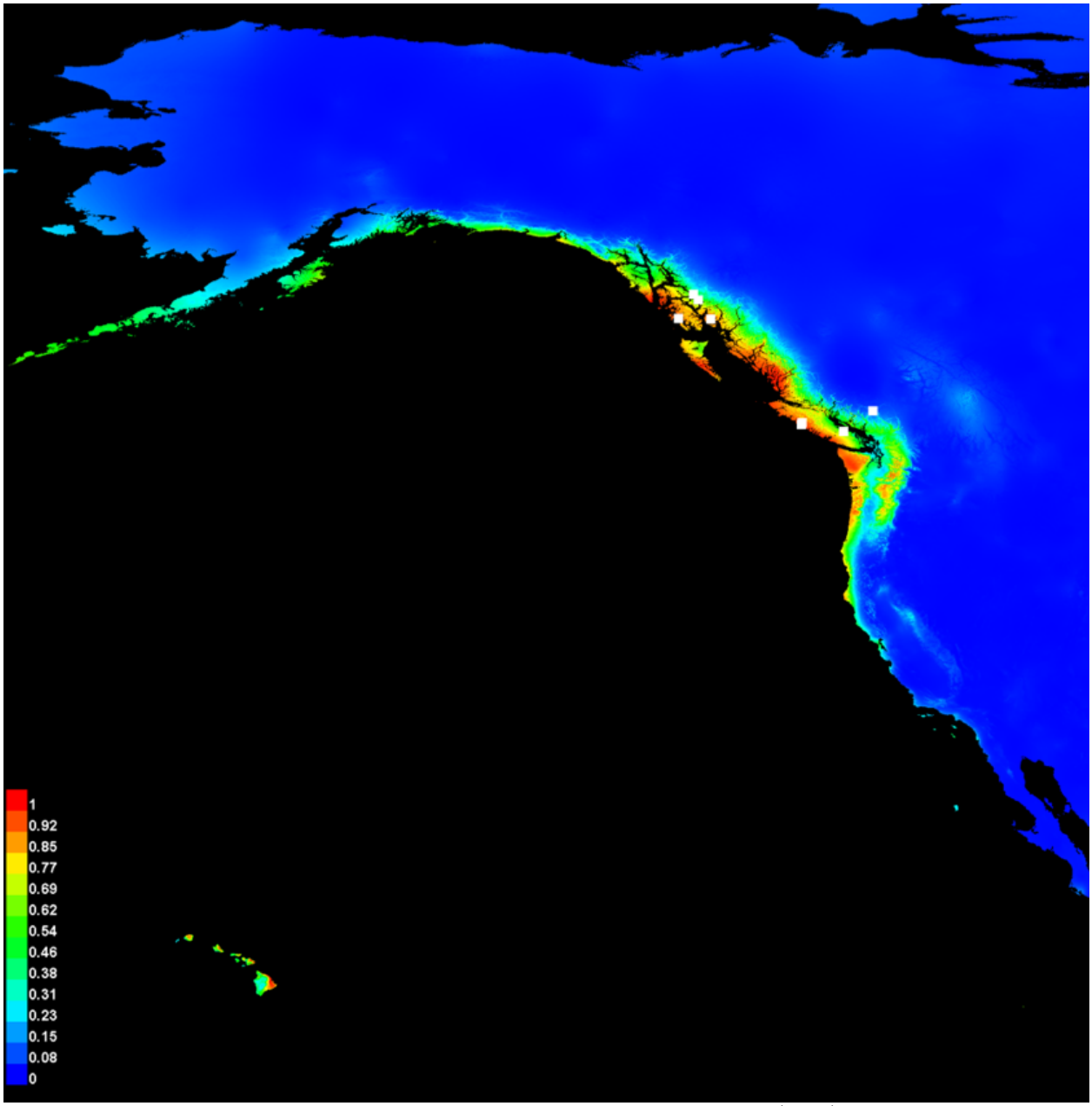

Figure B.7: MaxEnt distribution model for coastal Rainforest wolves (n=8). A logarithmic scale is shown and ranges from 0, corresponding to unsuitable climate (cooler colors), to 1, corresponding to most suitable climate (warmer colors).

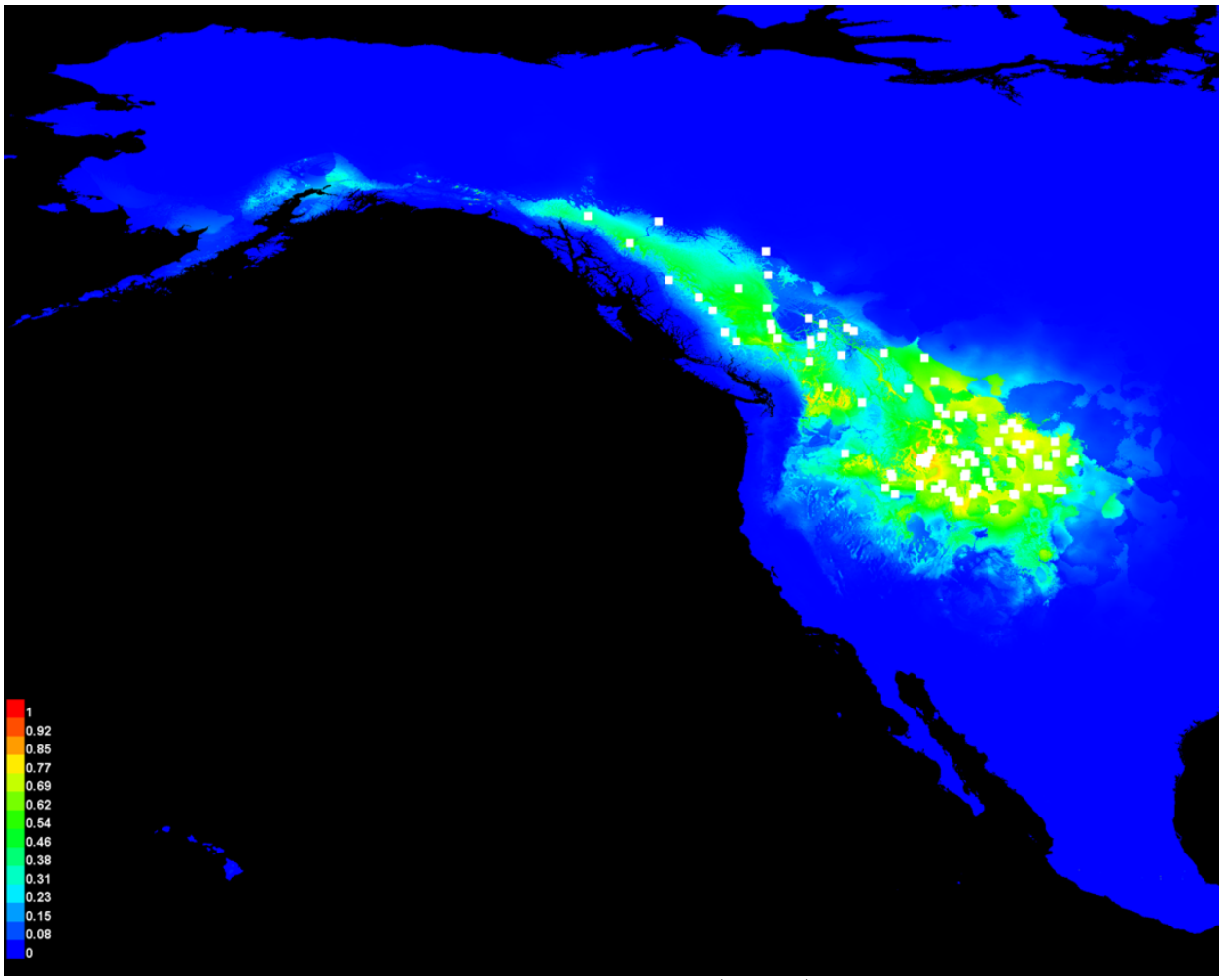

Figure B.8: MaxEnt distribution model for NRM wolves (n=108). A logarithmic scale is shown and ranges from 0, corresponding to unsuitable climate (cooler colors), to 1, corresponding to most suitable climate (warmer colors).

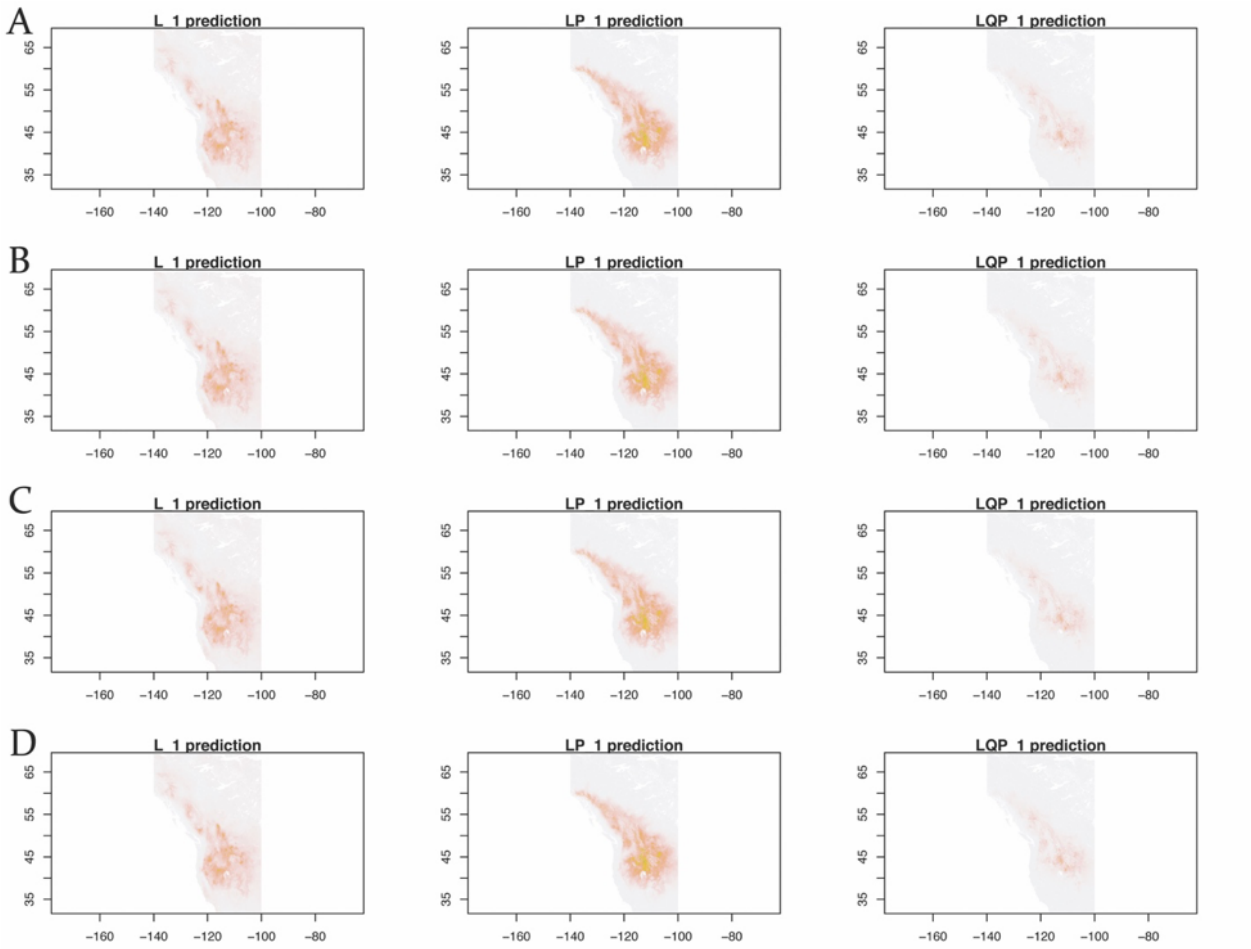

Figure B.9: Variation in MaxEnt model output variation for Northern Rocky Mountain wolves. Models were run using the following occurrence permutations; A) Checkerboard 2 (mean AUC = 0.89), B) Checkerboard 1 (mean AUC = 0.88), C) Jackknife (mean AUC = 0.89), and D) Random K fold (mean AUC = 0.89). For each occurrence permutation, we ran factors as follows; Linear (L), Linear Quadratic (LQ) and Linear Quadratic Product (LQP).

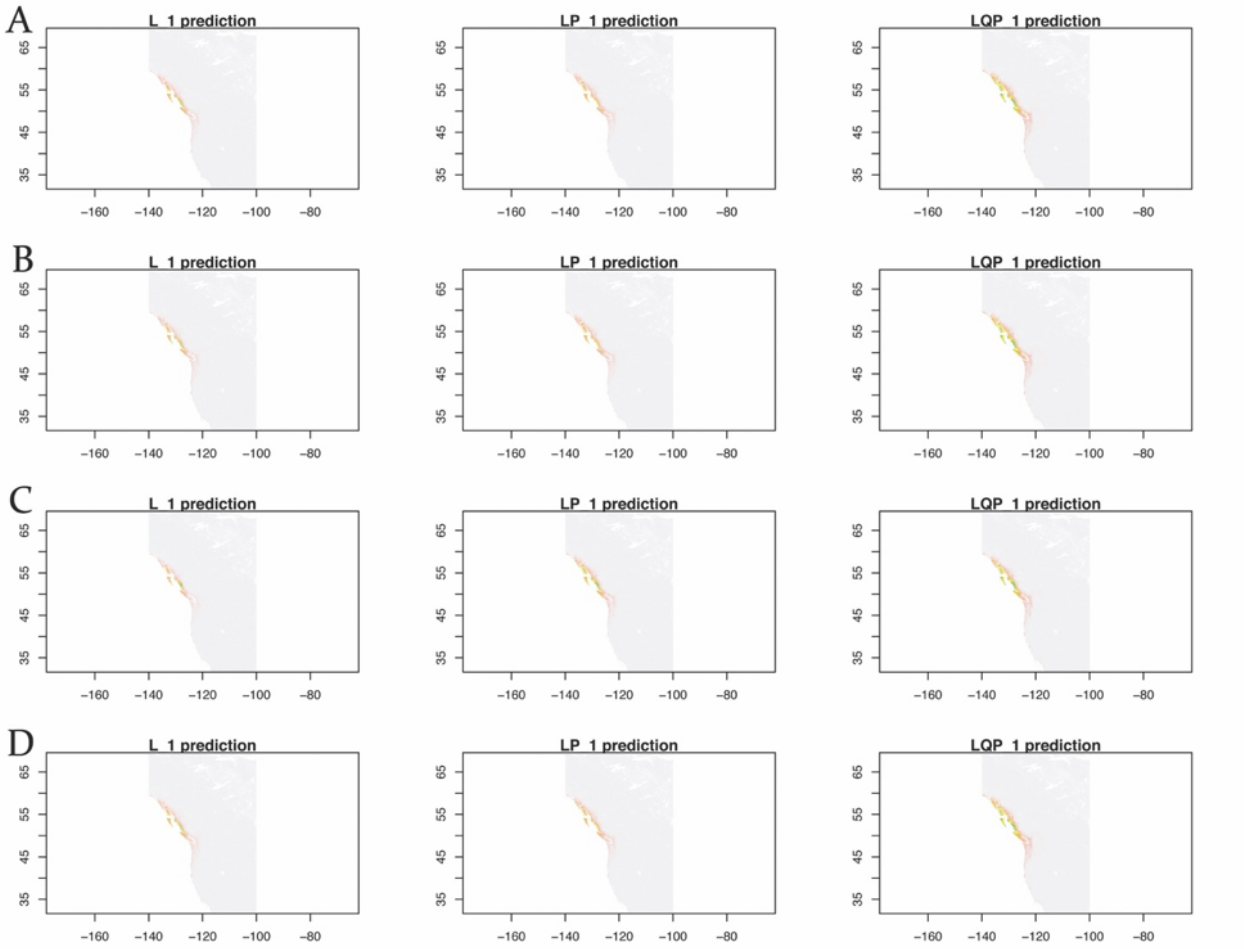

Figure B.10: Variation in MaxEnt model output variation for coastal rainforest wolves. Models were run using the following occurrence permutations; A) Checkerboard 2 (mean AUC = 0.99), B) Checkerboard 1 (mean AUC = 0.75), C) Jackknife (mean AUC = 0.99), and D) Random K fold (mean AUC = 0.98). For each occurrence permutation, we ran factors as follows; Linear (L), Linear Quadratic (LQ) and Linear Quadratic Product (LQP).

Table B.7: Centroid location of wolf pack for Washington (WA) and Oregon and probability of presence per pack to interior and coastal habitats. Location was determined as approximate central point of pack range. The habitat of the Washington Teanaway pack, was more similar to the habitat of coastal BC wolves than to the habitat of interior wolves of the region.

Location	Pack Name	Longitude	Latitude	Inland Probability of Presence	Coastal Probability of Presence
WA	Salmo	-117.291892	48.934843	0.4583	0.0551
WA	Smackout	-117.599509	48.790292	0.5513	0.0557
WA	Wedge	-117.981692	48.950232	0.6175	0.0448
WA	Dirty Shirt	-117.726721	48.562771	0.6287	0.0418
WA	Carpenter Ridge	-117.65531	48.304001	0.685	0.0316
WA	Huckleberry	-117.970706	47.998874	0.719	0.0306
WA	Nc'icn	-118.533755	48.433521	0.4699	0.0275
WA	Strawberry	-118.962222	48.466312	0.4641	0.0255
WA	Lookout	-120.34648	48.26566	0.141	0.094
WA	Teanaway	-121.096316	47.417627	0.0223	0.2887
WA	Tucannon	-117.59216	46.11894	0.2867	0.0315
WA	White Stone	-118.50952	48.05054	0.667	0.0176
WA	Loup Loup	-119.87182	48.4456	0.3924	0.038
WA	Beaver Creek	-119.08903	48.83949	0.4856	0.025
WA	Profanity Peak	-118.34747	48.76524	0.5497	0.0294
WA	Goodman Meadows	-117.22412	48.77067	0.296	0.071

Table B.7 continued from previous page

WA	Skookum	-117.16094	48.5457	0.3127	0.0578
WA	Strangler	-118.00689	48.37632	0.7025	0.0275
OR	Imnaha Pack	-116.99821	45.34829	0.366	0.0108
OR	Wenaha Pack	-117.77223	45.94805	0.2796	0.0269
OR	Walla Walla Pack	-118.07968	45.91867	0.23	0.0453
OR	Snake River Pack	-116.66244	45.60636	0.2826	0.0134
OR	Minam Pack	-117.518	45.356	0.2919	0.0172
OR	Mt Emily Pack	-118.21563	45.61885	0.201	0.0491
OR	Meacham Pack	-118.34609	45.52656	0.2193	0.0489
OR	Rogue Pack	-122.29705	42.6481	0.0247	0.0476
OR	Desolation Pair	-118.77044	44.87241	0.2995	0.0112
OR	Shamrock Pack	-117.32849	45.82115	0.3141	0.0145
OR	Catherine Pack	-117.62237	45.15493	0.3159	0.0109
OR	South Snake Pack	-116.89109	45.16949	0.4225	0.0106
OR	Unnamed Pack	-119.17557	45.23329	0.2949	0.0167

Appendix C: Supplementary Information to Chapter 5

C.1 Sequencing data

Table C.1: Sample, sequence, alignment, and filtering statistics.

status	Population	Seq ID	Sample ID	Sex	Col Date	Base Pairs (Mbp)	# of reads aligned, post-filtering (x 10^6)	Mean depth of coverage (X)	95%-ile of coverage (X)	% of reference covered by at least 6 reads
case	Santa Catalina	PHSH004A	94065	F	2006	53,698	241	15.3227	25	91.99
case	Santa Catalina	PHSH004B	83E0C	F	2006	38,688	170	10.7951	18	87.53
case	Santa Catalina	PHSH004C	74A72	F	2010	53,055	237	15.0565	24	91.9
case	Santa Catalina	PHSH005A	46F1D	M	2010	52,311	209	13.273	22	90.02
case	Santa Catalina	PHSH005B	E5B17	F	2007	47,081	192	12.2243	20	89.92
case	Santa Catalina	PHSH005C	Z02-106	F	2002	52,473	224	14.2091	23	91.68
case	Santa Catalina	PHSH006A	12A20	M	2007	49,288	199	12.6484	21	89.39
case	Santa Catalina	PHSH006B	Z06-241	F	2006	61,321	216	13.6927	22	90.29
case	Santa Catalina	PHSH006C	Z04-37	M	2004	47,903	179	11.3438	19	88.98

Table C.1 continued from previous page

case	Santa Catalina	PHSH007H	Z06-242	M	2006	51,987	247	15.6107	26	91.11
case	Santa Catalina	PHSH12A/008I	36966	F	2010	39,455	141	8.7725	41	79.45
case	Santa Catalina	PHSH12B	E2F58	M	2011	30,932	156	9.6872	49	81.96
control	Santa Catalina	PHSH001A	E3B7A	M	2009	47,682	174	11.0228	19	87.29
control	Santa Catalina	PHSH001B	5317D	M	2010	54,310	194	12.3341	21	89.08
control	Santa Catalina	PHSH001C	Z06-291	M	2006	51,623	195	12.3518	21	89.06
control	Santa Catalina	PHSH002A	E6A0A	F	2007	48,538	171	10.8626	18	88.12
control	Santa Catalina	PHSH002B	01A24	F	2009	53,052	186	11.783	20	89.6
control	Santa Catalina	PHSH002C	53E3E	M	2007	51,224	185	11.7151	20	88.2
control	Santa Catalina	PHSH003A	4581F	M	2007	51,430	181	11.4769	19	88.12
control	Santa Catalina	PHSH003B	2770E	F	2007	42,449	159	10.1297	17	86.41
control	Santa Catalina	PHSH003C	93369	M	2007	57,306	222	14.0851	23	90.38
control	Santa Catalina	PHSH007A	Z07-90	M	2007	48,236	207	13.0961	22	88.58
control	Santa Catalina	PHSH12C	70616	F	2011	34,453	177	11.0005	50	87.24
control	San Clemente	PHSH007K	RKW13580	F	2006	27,973	129	8.1548	15	76.47
control	San Clemente	PHSH010A	RKW13701	F	2009	32,939	150	9.4755	14	83.11
control	San Clemente	PHSH010B	RKW13601	M	2006	34,185	156	9.8598	18	83.21
control	San Clemente	PHSH010C	RKW13615	M	2007	34,438	154	9.7457	17	82.8

Table C.1 continued from previous page

control	San Clemente	PHSH011A	RKW13679	M	2008	37,548	174	11.0213	19	86.22
control	San Clemente	PHSH011B	RKW13716	M	2008	27,982	131	8.3208	15	75.63
control	San Clemente	PHSH011C	RKW13685	M	2009	33,645	158	10.0065	18	83.64
control	San Clemente	JARW010	RKW13704	F	2009	69,500	446	18.99	33	90.13
control	San Clemente	SRR5198013	RKW4045	F	1988	73,000	124	10.5003	51	86.42

C.2 Extent of LD (r^2) decay

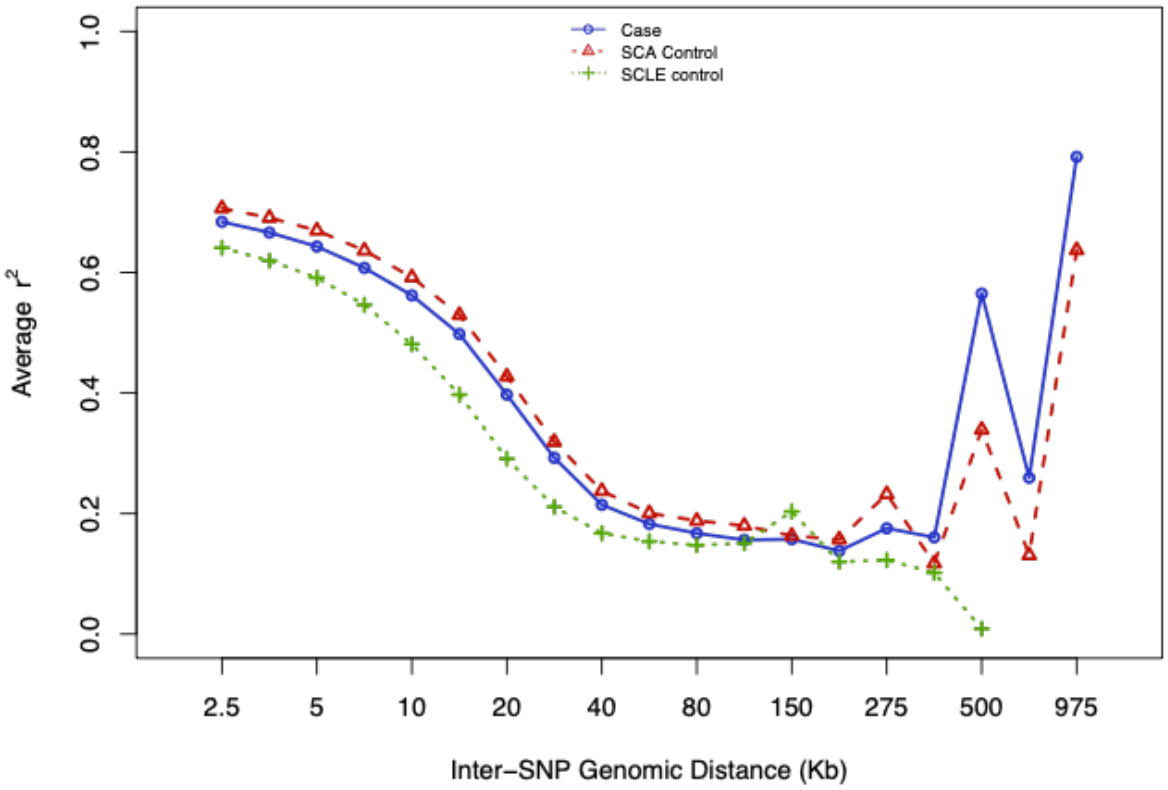

Figure C.1: Extent of LD (r^2) as a function of inter-SNP distance (Kb) for SCA cases (shown in blue), SCA controls (shown in red), and SCLE controls (shown in green).

C.3 Principal Components Analyses of 32 individuals

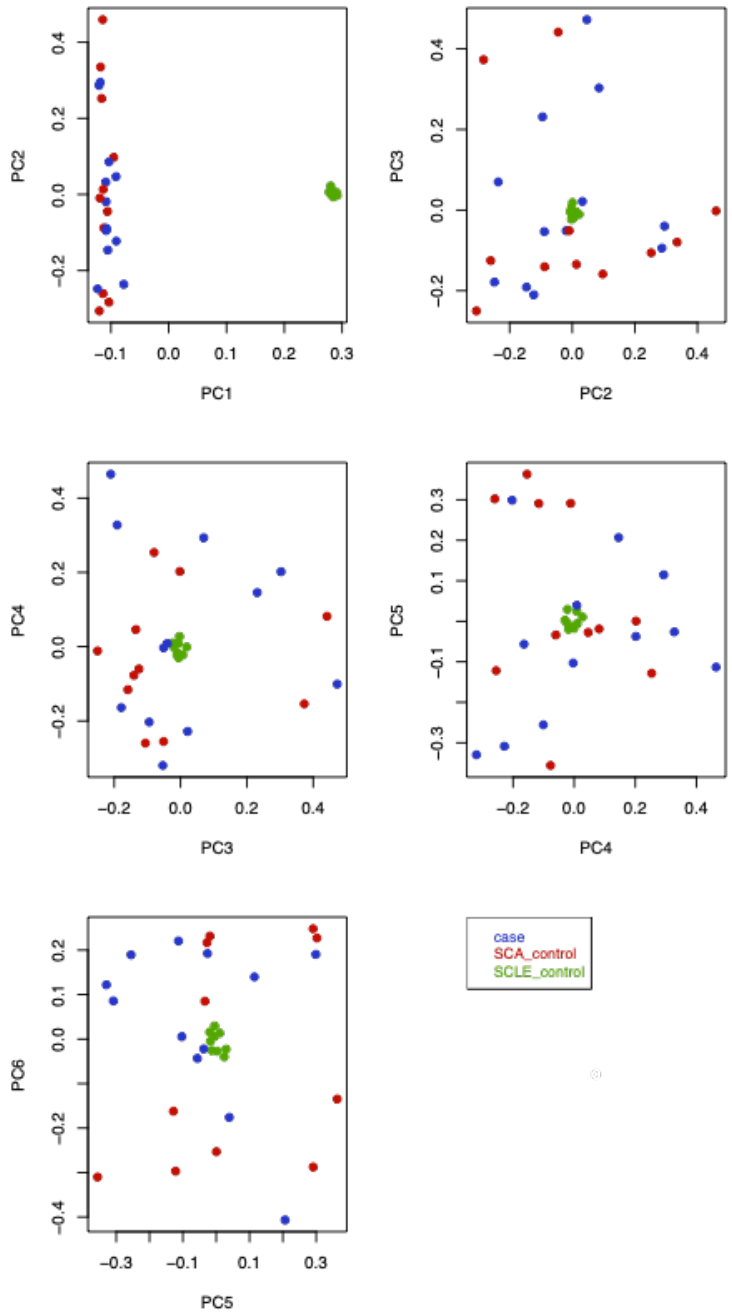

Figure C.2: PCA of 32 individuals using 6.6K SNP data set. Case individuals are shown in blue; SCA controls in red, and SCLE controls in green.

C.4 Effective population size

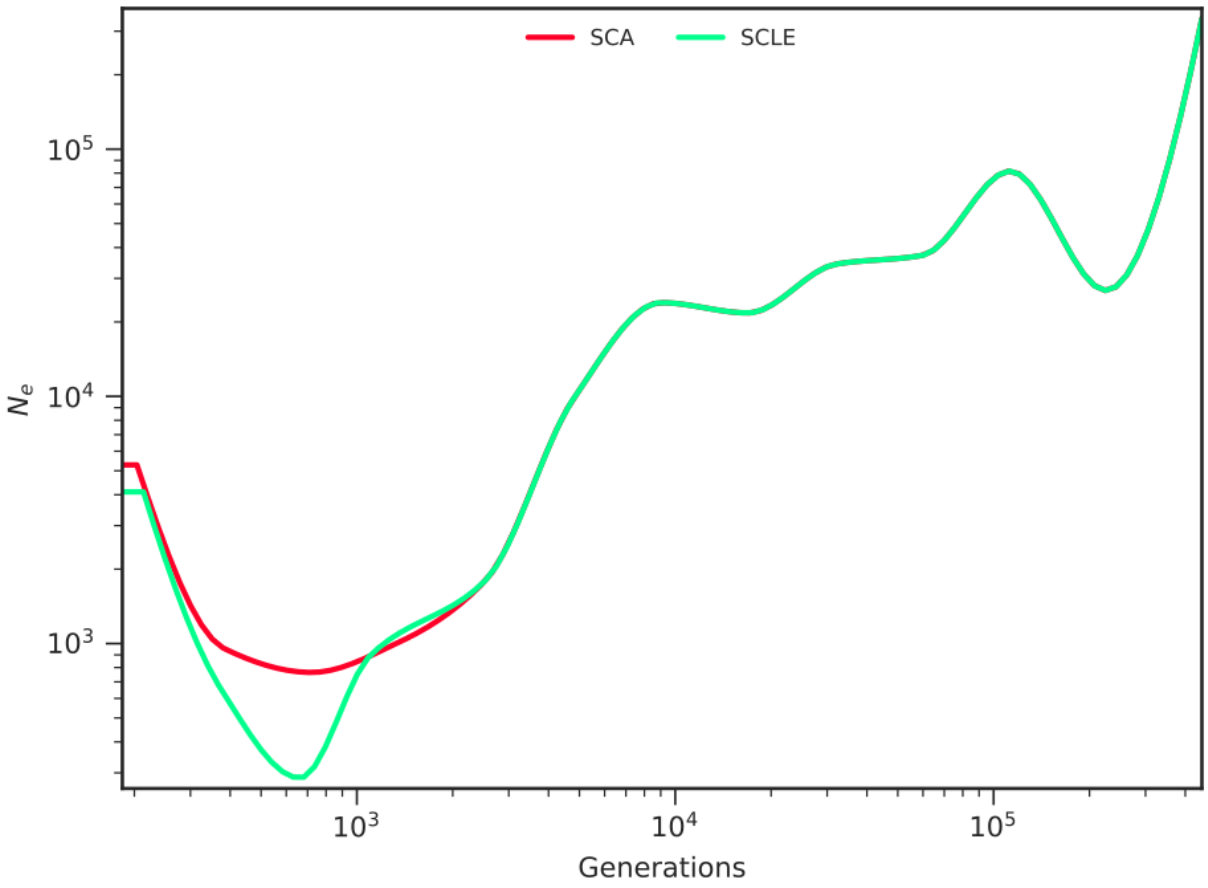

Figure C.3: Effective population size (Ne) calculated from SNP genotypes within neutral regions. A mutation rate of 1.0 x 10-8/site/generation was assumed (Freedman et al. 2014). SCA individuals in red, and SCLE individuals in green..

C.5 Inbreeding coefficients

Table C.2: Inbreeding coefficients calculated in "detectRUNS", which estimates F(ROH) as the proportion of ROH to total number of loci across the genome.

Sample ID	Status	Total # of ROH	Sum of bp in ROH	F(ROH)
12A20	case	1586	667991211	0.30316264
36966	case	90	83157560	0.03774041
46F10	case	1818	761179387	0.34545537
74A72	case	1819	753340597	0.3418978
83E0C	case	2119	872384237	0.39592483
94065	case	1906	794023852	0.36036158
E2F58	case	537	331820929	0.15059436
E5B17	case	1870	785358381	0.35642882
Z02-106	case	1760	734816196	0.33349064
Z04-37	case	2107	838816642	0.38069044
Z06-241	case	2672	1047276801	0.47529847
Z06-242	case	1907	784979413	0.35625683
01A24	control	1943	805198758	0.36543322
2770E	control	1751	741738457	0.33663226
4581F	control	1880	786046606	0.35674117
5317D	control	2111	826548134	0.37512247
53E3E	control	2066	846358848	0.38411341
70616	control	1745	730895423	0.33171123
93369	control	2210	888304705	0.40315021
E3B7A	control	1966	801499092	0.36375416
E6A0A	control	2035	817478917	0.37100648
Z06-291	control	2081	828492551	0.37600493

Table C.2 continued from previous page

Z07-90	control	2181	852745169	0.38701179
13580	SCLE	157	108027652	0.04902751
13601	SCLE	247	150608118	0.06835233
13615	SCLE	539	312245667	0.14171028
13679	SCLE	3282	1227963205	0.55730159
13685	SCLE	1486	674528561	0.30612956
13701	SCLE	860	423618268	0.19225587
13716	SCLE	852	438136079	0.19884467
RKW13704	SCLE	3089	1156055528	0.52466685
SCLV4F	SCLE	2645	991156244	0.44982859

C.6 Manhattan plots of F_{ST}

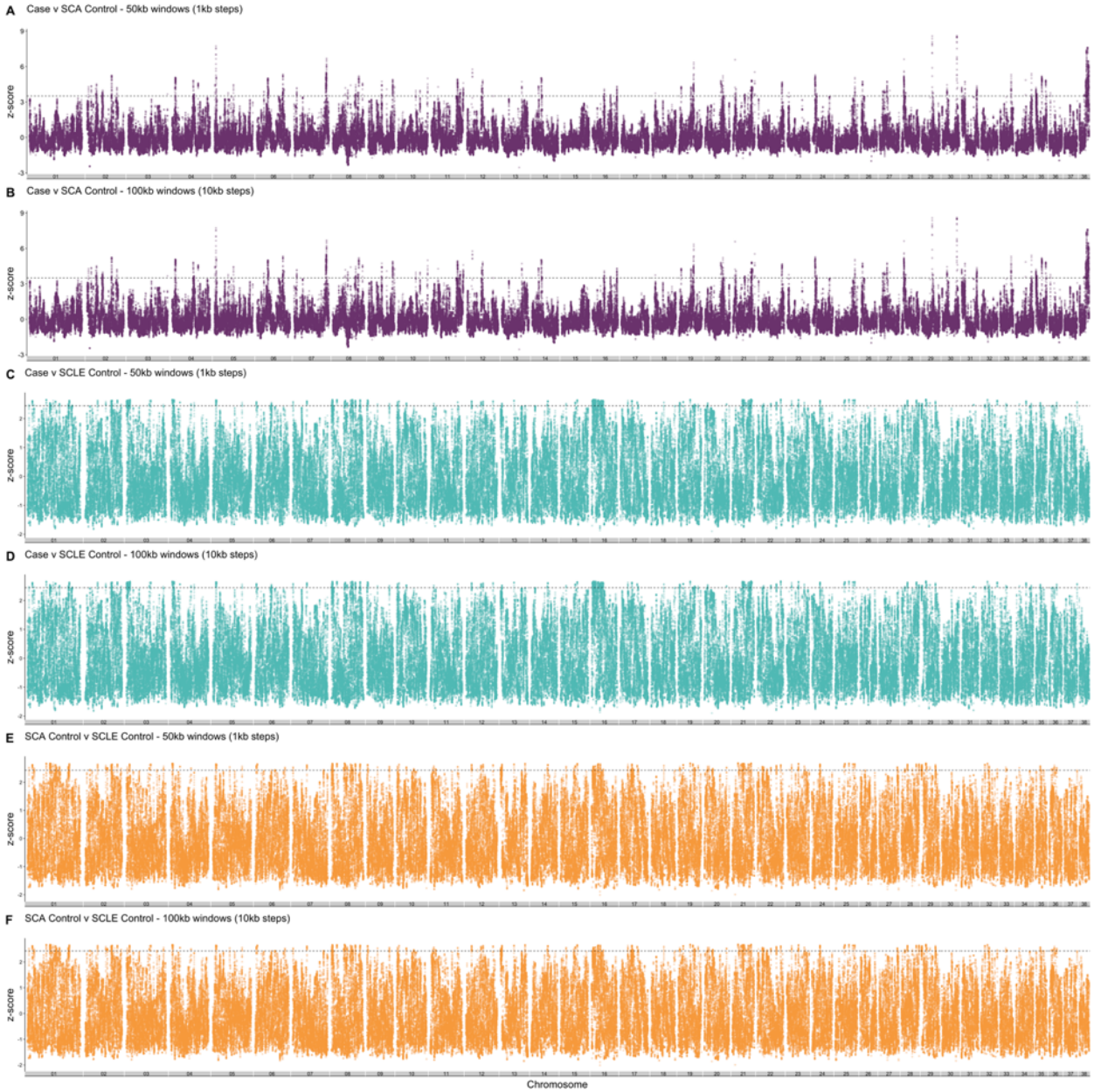

Figure C.4: Manhattan plot of F_{ST} values calculated from overlapping 100 Kb windows in 10 Kb steps using the 35K SNP data set between (A & B) case (n = 12) and SCA control (n = 11) individuals, (C & D) case and SCLE control (n = 9) individuals, (E & F) SCA control and SCLE control individuals. Dashed lines indicated top 1%.

C.7 Distribution plots of Δ π values

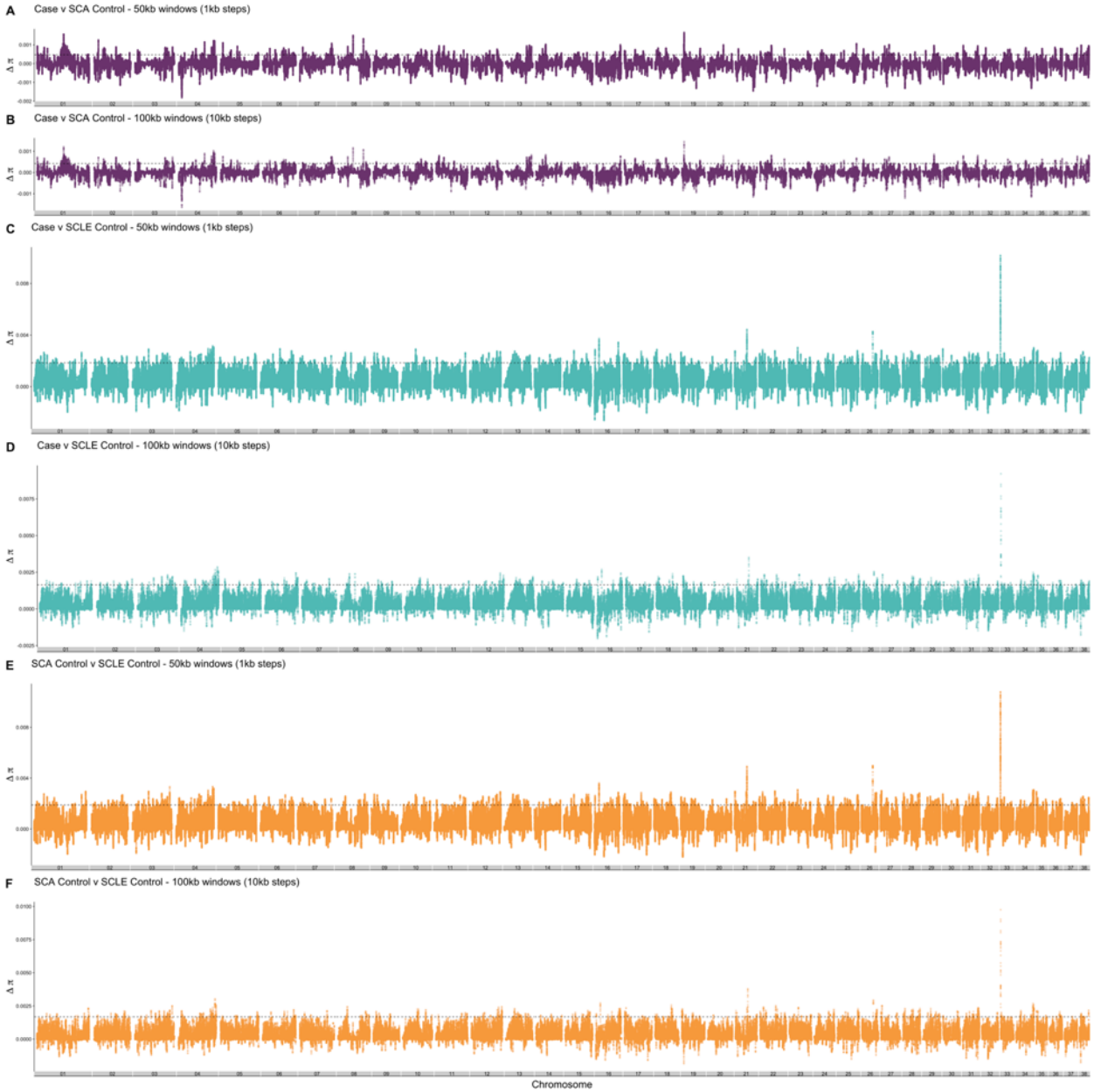

Figure C.5: Distribution plots of Δ π values calculated from 50 Kb (1 Kb steps) and 100 Kb (10 Kb steps) windows using the 35K SNP data set between (A & B) case (n = 12) and SCA control (n = 11) individuals, (C & D) case and SCLE control (n = 9) individuals, (E & F) SCA control and SCLE control individuals. Dashed lines indicated top 1%.

C.8 Distribution plots of Δ Tajima's D values

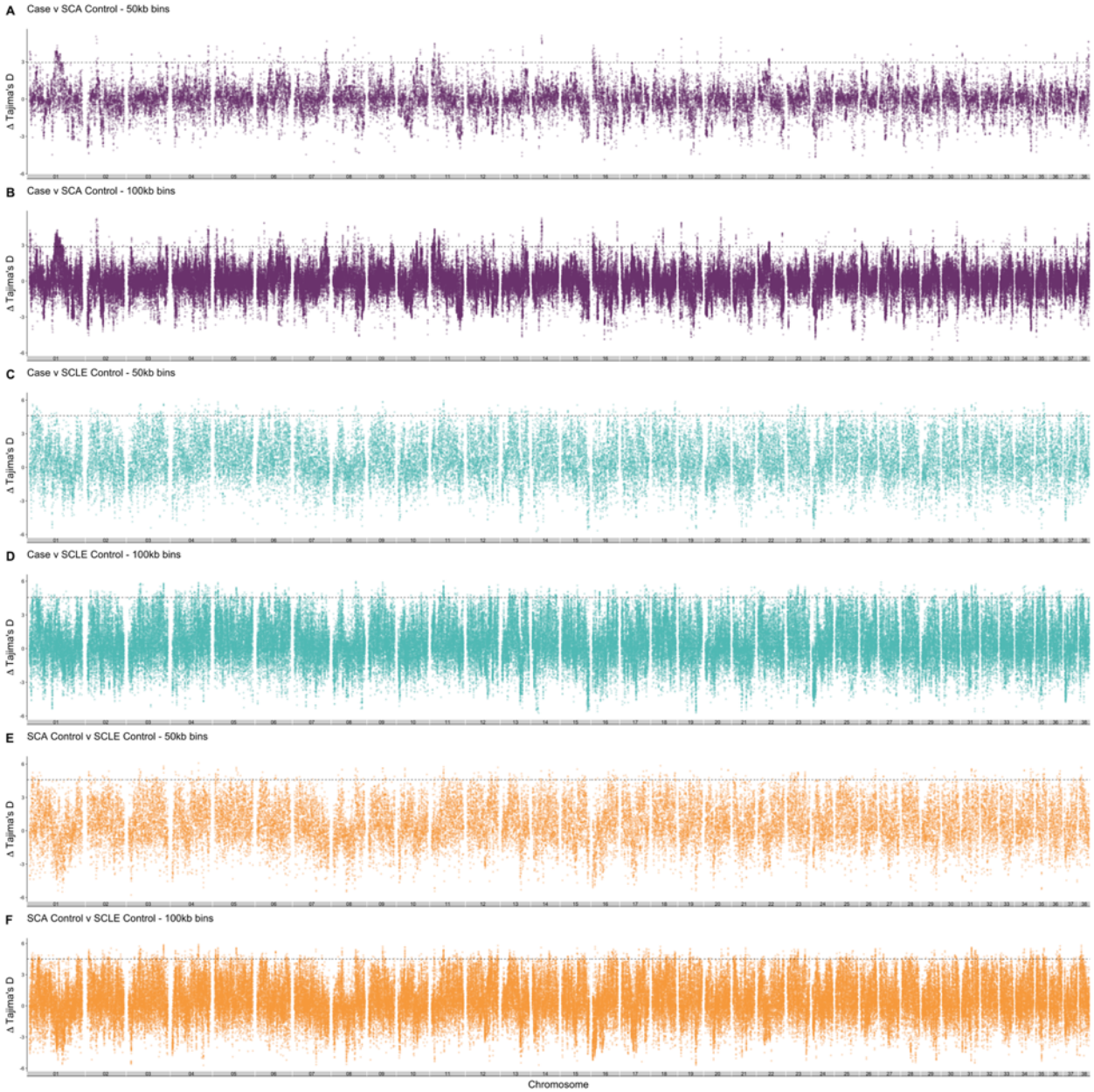

Figure C.6: Distribution plots of Δ Tajima's D values calculated in 50 Kb and 100 Kb bins using the 35K SNP data set between (A & B) case and SCA control individuals, (C & D) case and SCLE control individuals, (E & F) SCA control and SCLE control individuals. Dashed lines indicated top 1%.

C.9 Distribution plots of XP-CLR scores

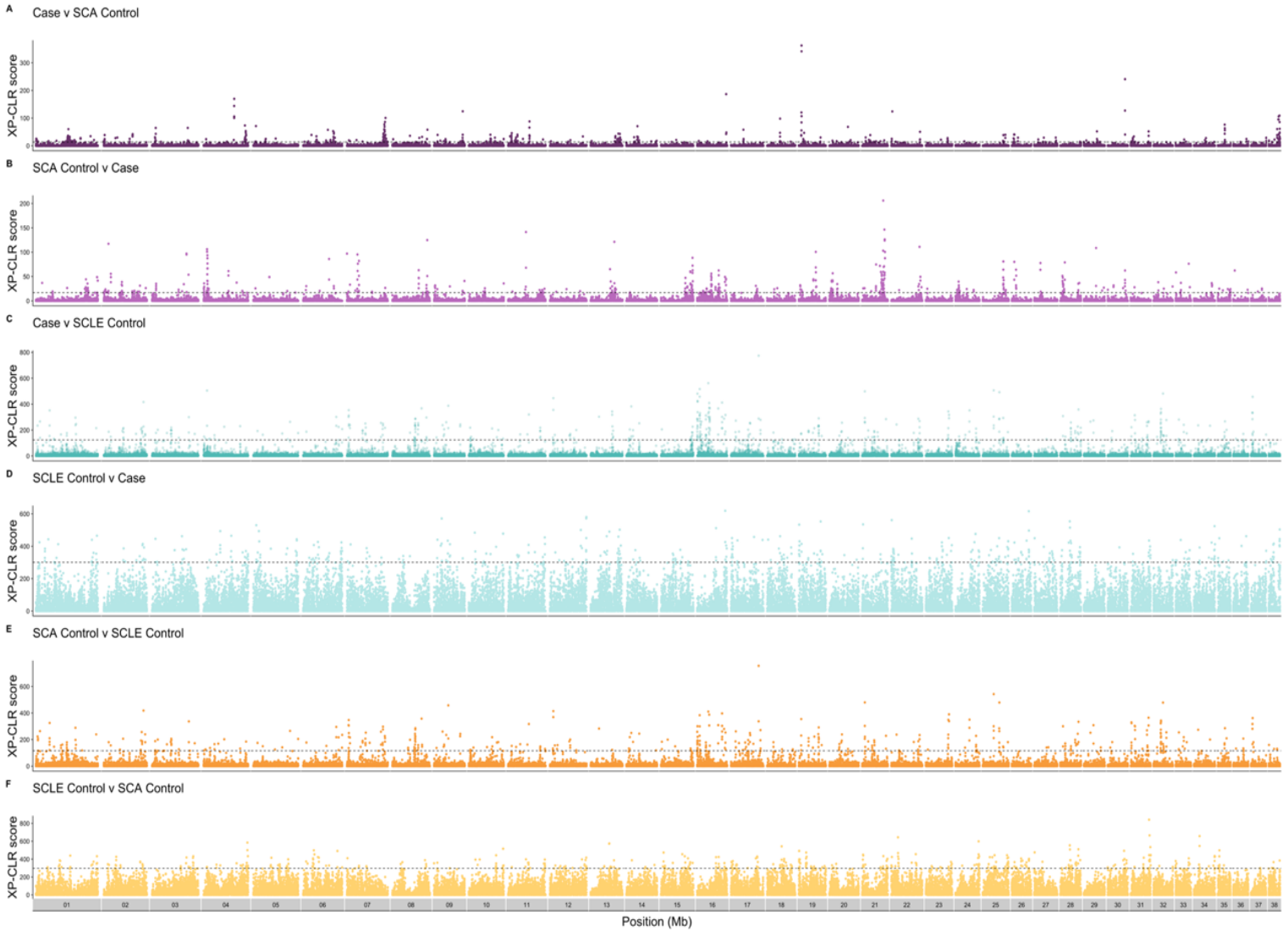

Figure C.7: Plot of XP-CLR scores along each autosome in a comparison of all groups. Dashed line indicated top 1%.

C.10 Genome-wide mutations classified by annotation type

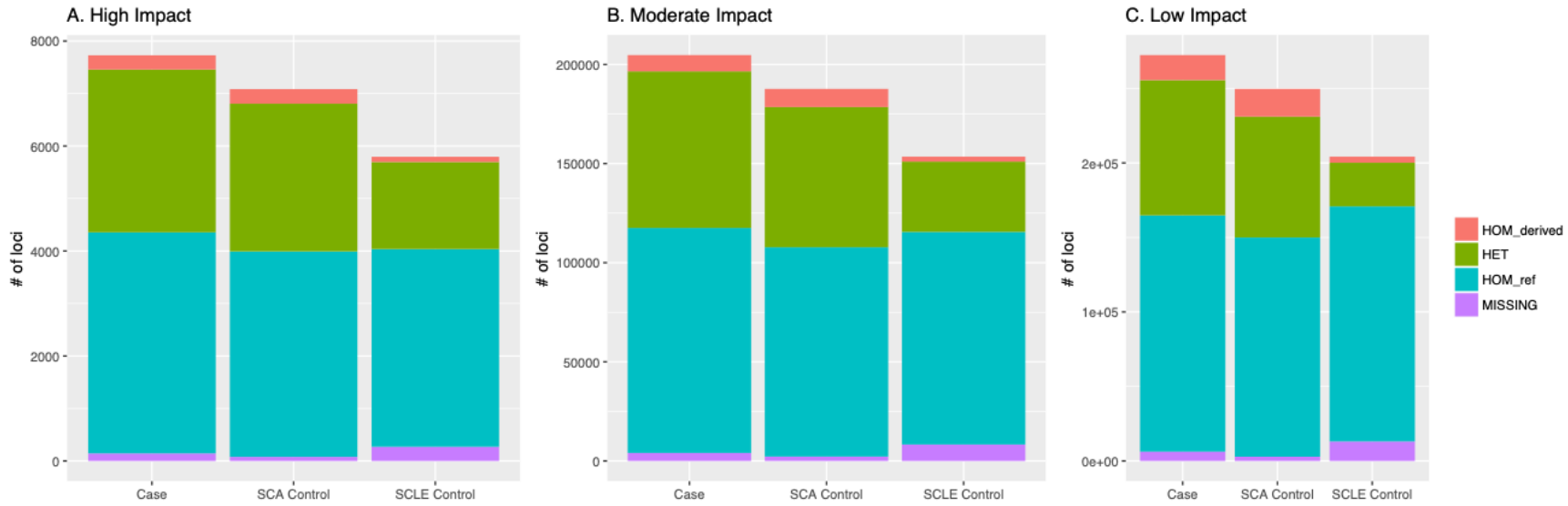

Figure C.8: Genome-wide mutations are classified by annotation type as determined by SNPeff per population. (A) High impact mutations are those that encode premature stop codons. (B) Moderate impact mutations are missense mutations. (C) Low impact mutations are classified as synonymous changes are mutations that do not change the encoded amino acid and are presumed to be neutral.

Table C.3: Candidate SNPs and minor allele frequency (MAF) of the top 1% of delta pi with delta allele frequency greater than 0.45 where cases have lower pi values than controls and difference in allele frequency is 0.45 or greater.

TFBS	Chrom #	Position	Ref. allele	Alt. allele	Quality score	Case MAF	SCA Control MAF	SCLE Control MAF	Annotation	Gene name(s)
-	1	120796783	C	A	4263.49	0.82	0.35	0	intergenic	ENSCAFG00000033119-ZNF536
-	1	120810036	C	A	4791.85	0.9	0.45	0	intergenic	ENSCAFG00000033119-ZNF536
-	19	33846660	A	G	6171.15	0.86	0.41	1	intron	DPP10
-	24	6929554	G	A	6914.41	0.95	0.5	0.25	intron	ENSCAFG00000039265

C.11 The top 0.1% of outlier loci for $\Delta\ \pi$

Table C.4: The top 0.1% of outlier loci for $\Delta\ \pi$ between cases and both control populations (SCA and SCLE).

TFBS	Chrom #	Position	Bin start	Bin end	Putative impact	Annotation	Gene name (ortholog)
-	4	5382219	5337001	5427000	MODERATE	missense_variant	TARBP1
-	4	5407342	5358001	5427000	MODERATE	missense_variant	TARBP1
TFBS	4	6705047	6700001	6755000	MODERATE	missense_variant	ENSCAFG00000011612 (NTPCR)
-	8	2496720	2474001	2542000	MODERATE	missense_variant	TRAV21
-	8	34755349	34738001	34799000	MODERATE	missense_variant	CCDC175
-	16	6945885	620001	681000	HIGH	start_lost	TRBV3-1
-	16	6900739	6885001	6950000	MODERATE	missense_variant	ENSCAFG00000003812 (TRBV13)
-	16	6900898	6885001	6950000	MODERATE	missense_variant	ENSCAFG00000003812 (TRBV13)
-	16	6944443	6895001	6963000	MODERATE	missense_variant	ENSCAFG00000030524 (TRBV4-2)

Table C.4 continued from previous page

-	16	6945648	6896001	6963000	MODERATE	missense_variant	ENSCAFG00000030524 (TRBV4-2)
-	16	6945875	6896001	6963000	MODERATE	missense_variant	ENSCAFG00000030524 (TRBV4-2)
-	16	6949242	6900001	6963000	MODERATE	missense_variant	TRBV3-1
-	16	6949347	6900001	6963000	MODERATE	missense_variant	TRBV3-1
-	18	11757556	11744001	11807000	MODERATE	missense_variant	STARD3NL
-	21	27507872	27490001	27558000	MODERATE	missense_variant	ENSCAFG00000031799 (OR52J3)
-	21	27539941	27490001	27572000	MODERATE	missense_variant	OR52E2
-	21	27539954	27490001	27572000	MODERATE	missense_variant	OR52E2
-	21	27540166	27491001	27572000	MODERATE	missense_variant	OR52E2
-	21	27571379	27522001	27619000	MODERATE	missense_variant	ENSCAFG00000029703 (OR52E2)
-	21	27614512	27565001	27619000	MODERATE	missense_variant	ENSCAFG00000025564 (OR52R1)
-	21	27614591	27565001	27619000	MODERATE	missense_variant	ENSCAFG00000025564 (OR52R1)

Table C.4 continued from previous page

-	21	27846473	27824001	27891000	MODERATE	missense_variant	ENSCAFG00000012321 (Olfr609)
-	21	27855891	27824001	27891000	MODERATE	missense_variant	ENSCAFG00000024022 (Olfr610)
-	21	28281555	28279001	28330000	MODERATE	missense_variant	OR51B6
-	21	28281699	28279001	28330000	MODERATE	missense_variant	OR51B6
-	21	28362320	28337001	28390000	MODERATE	missense_variant	OR51Q1
-	21	28362468	28337001	28390000	MODERATE	missense_variant	OR51Q1
-	21	28362896	28337001	28390000	MODERATE	missense_variant	OR51Q1
-	23	27521535	27481001	27560000	MODERATE	missense_variant	ENSCAFG00000024259 (Col6a4)
-	25	32112720	32072001	32123000	MODERATE	missense_variant	DOCK5
-	25	45379280	45351001	45429000	MODERATE	missense_variant	SPP2
-	25	45385328	45351001	45435000	MODERATE	missense_variant	SPP2
-	25	45386146	45351001	45436000	MODERATE	missense_variant	SPP2
-	26	34571069	34562001	34621000	MODERATE	missense_variant	PCDH15
-	26	34571630	34562001	34621000	MODERATE	missense_variant	PCDH15
-	28	39453901	39404001	39461000	MODERATE	missense_variant	TCERG1L

Table C.4 continued from previous page

-	28	40582200	40579001	40632000	MODERATE	missense_variant	CFAP46
-	28	40587027	40579001	40637000	MODERATE	missense_variant	CFAP46
-	28	40593740	40579001	40643000	MODERATE	missense_variant	CFAP46
-	28	40593766	40579001	40666000	MODERATE	missense_variant	CFAP46
-	28	40626144	40579001	40676000	MODERATE	missense_variant	CFAP46
-	28	40646035	40597001	40678000	MODERATE	missense_variant	CFAP46
-	28	40655600	40606001	40678000	MODERATE	missense_variant	CFAP46
-	28	40657361	40608001	40678000	MODERATE	missense_variant	CFAP46
-	28	40671212	40622001	40678000	MODERATE	missense_variant	CFAP46
-	32	15203073	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203113	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203173	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203194	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)

Table C.4 continued from previous page

-	32	15203327	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203416	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203427	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	32	15203790	15154001	15207000	MODERATE	missense_variant	ENSCAFG00000009965 (CCNB1)
-	33	282099	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282168	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282210	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282237	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282279	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)

Table C.4 continued from previous page

-	33	282337	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282435	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282546	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282598	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282676	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282700	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282706	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	33	282731	260001	332000	MODERATE	missense_variant	ENSCAFG00000025407 (CSNKA2IP)
-	34	36044827	36019001	36076000	MODERATE	missense_variant	PLD1
-	35	25490111	25481001	25537000	MODERATE	missense_variant	ENSCAFG00000024969

Table C.4 continued from previous page

-	35	25527520	25481001	25547000	MODERATE	missense_variant	ZSCAN12
-	35	25533128	25484001	25547000	MODERATE	missense_variant	ZSCAN12
-	35	25533162	25484001	25547000	MODERATE	missense_variant	ZSCAN12
-	35	25533210	25484001	25547000	MODERATE	missense_variant	ZSCAN12
-	35	25533343	25484001	25547000	MODERATE	missense_variant	ZSCAN12

Table C.5: The top 0.1% of outlier loci for delta pi between cases and SCA control populations .

TFBS	Chrom #	Position	Putative impact	Annotation	Gene name
-	chr01	62494037	MODERATE	missense_variant	TRDN
-	chr01	66204179	MODERATE	missense_variant	ENSCAFG00000001071
-	chr01	66204461	MODERATE	missense_variant	ENSCAFG00000001071
-	chr01	66204467	MODERATE	missense_variant	ENSCAFG00000001071
TFBS	chr04	6705047	MODERATE	missense_variant	ENSCAFG00000011612
-	chr04	7553233	MODERATE	missense_variant	DISC1
-	chr04	49209739	MODERATE	missense_variant	GABRA6
-	chr04	60561295	MODERATE	missense_variant	GZMK
-	chr04	60568637	MODERATE	missense_variant	GZMK
-	chr04	60596763	MODERATE	missense_variant	ESM1
-	chr07	127002	MODERATE	missense_variant	PPFIA4
-	chr07	127110	MODERATE	missense_variant	PPFIA4
-	chr07	74708177	MODERATE	missense_variant	ENSCAFG00000018677
-	chr07	74708257	MODERATE	missense_variant	ENSCAFG00000018677
-	chr08	34710237	MODERATE	missense_variant	JKAMP
-	chr08	34710331	MODERATE	missense_variant	JKAMP
-	chr08	34755349	MODERATE	missense_variant	CCDC175
-	chr14	22718195	MODERATE	missense_variant	ASNS
-	chr14	22718261	MODERATE	missense_variant	ASNS
-	chr14	22718264	MODERATE	missense_variant	ASNS
-	chr14	22718275	MODERATE	missense_variant	ASNS
-	chr28	4592611	MODERATE	missense_variant	KIF20B
-	chr28	4592634	MODERATE	missense_variant	KIF20B
-	chr28	4599491	MODERATE	missense_variant	KIF20B

Table C.5 continued from previous page

-	chr28	4603488	MODERATE	missense_variant	KIF20B
-	chr28	4608874	MODERATE	missense_variant	KIF20B
-	chr28	4613750	MODERATE	missense_variant	KIF20B
-	chr28	4659348	MODERATE	missense_variant	KIF20B
-	chr28	4659369	MODERATE	missense_variant	KIF20B
-	chr28	4703449	MODERATE	missense_variant	ENSCAFG00000030727
-	chr28	40616066	MODERATE	missense_variant	CFAP46
-	chr28	40626144	MODERATE	missense_variant	CFAP46
-	chr28	40646035	MODERATE	missense_variant	CFAP46
-	chr28	40655600	MODERATE	missense_variant	CFAP46
-	chr28	40657361	MODERATE	missense_variant	CFAP46
-	chr34	36044827	MODERATE	missense_variant	PLD1
-	chr34	36612933	MODERATE	missense_variant	FNDC3B

Table C.6: Minor Allele Frequencies of the SNPs resulting from the intersection of the top 1% of delta pi between cases and both control populations (SCA and SCLE) and the top 1% of delta pi between cases and SCA controls..

Chrom	Position	Case MAF	SCA Control MAF	SCLE Control MAF
chr04	6705047	0.0416667	0.227273	1
chr08	34755349	0.541667	0.818182	1
chr16	6945885	0.625	0.636364	0.625
chr28	40616066	0.95	0.85	0.1875
chr28	40626144	0.958333	0.85	0.1875
chr28	40646035	1	1	0.5
chr28	40655600	0.95	0.863636	0.0625
chr28	40657361	0.818182	0.95	0.0555556
chr34	36044827	0	0	1

C.12 Venn diagram

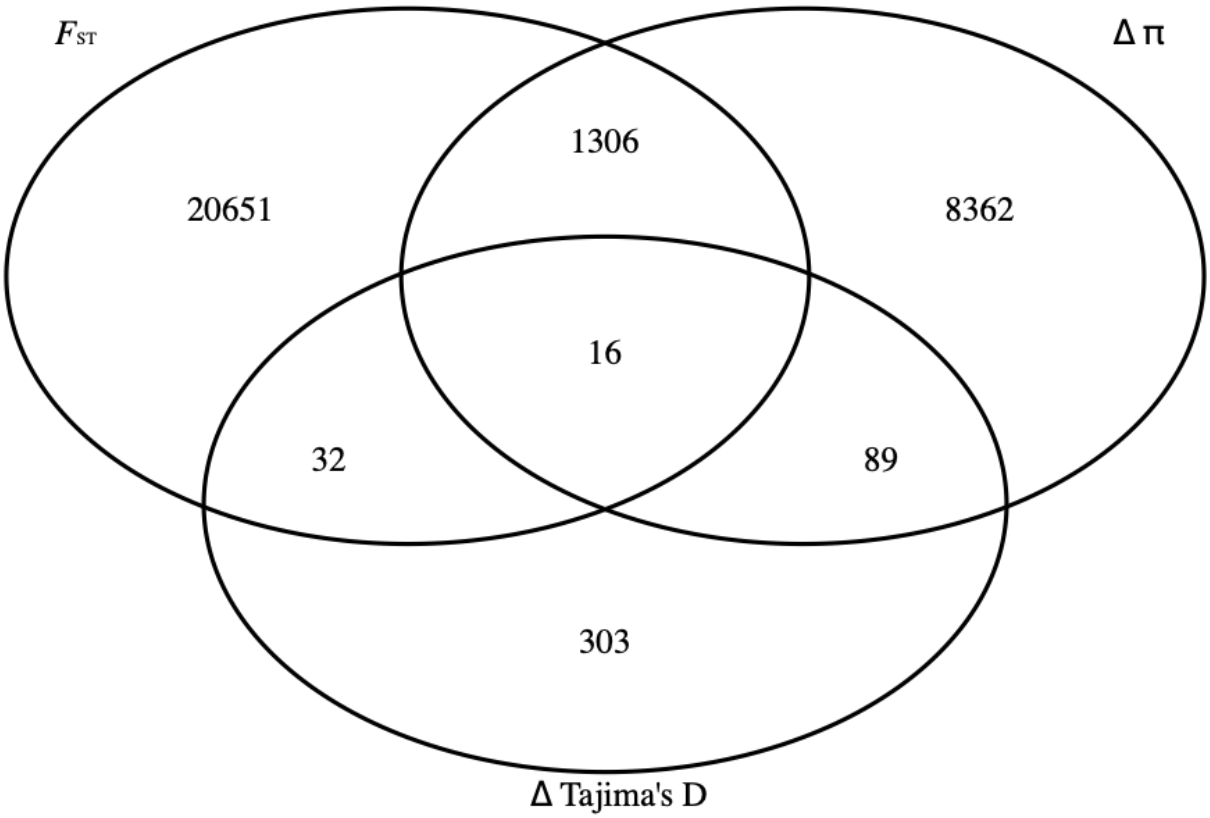

Figure C.9: Venn diagram of top 1% of F_{ST}, $\Delta\ \pi$, and Δ Tajima's D to identify regions under selection. .

C.13 F_{ST} values and absolute values of the difference in allele frequency within the top 16 candidate regions

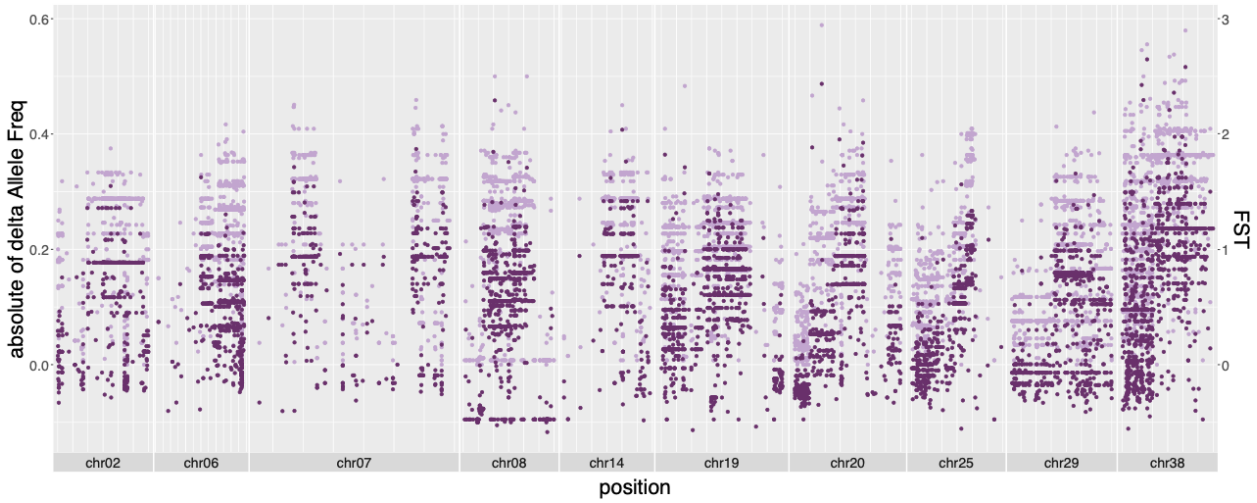

Figure C.10: F_{ST} values (dark purple) and absolute values of the difference in allele frequency (light purple) within the top 16 ranked candidate regions under selection between case and SCA controls.

C.14 Candidate genes identified from SNPs within 16 candidate regions

Table C.7: Candidate genes identified from SNPs within 16 candidate regions with an allele frequency difference above 0.45 (n = 38). (*) designates significant GO genes.

Chrom #	Genes start	Genes end	Genes name	Genes ID (Ortholog)	# & annotation
7	73815576	74597564	Protein Tyrosine Phosphatase, Receptor Type M	PTPRM	2 intronic
8	34686245	34700199	Trans-L-3-Hydroxyproline Dehydratase	L3HYPDH	1 upstream
8	34725466	34792685	Coiled-Coil Domain Containing 175α	CCDC175	1 intronic
8	34810342	35033387	Reticulon 1	RTN1	1 downstream
14	21807022	21808512	Keratin 8	ENSCAFG00000002184 (KRT8)	1 intergenic
14	22026498	22026678	unknown	ENSCAFG00000030255	-
19	5167424	5262738	unknown	ENSCAFG00000039008	1 intronic
20	36597454	36657503	Decapping MRNA 1A	*DCP1A	2 intronic
20	36724062	36749485	Protein Kinase C Delta	PRKCD	1 intergenic
20	36781858	36818854	Requiring Fifty Three 1 Homolog	RFT1	-

Table C.7 continued from previous page

38	22250764	22269238	Cilia And Flagella Associated Protein 45 (Coiled-Coil Domain Containing 19)	CFAP45 (CCDC19)	1 intergenic
38	22281625	22285509	V-Set And Immunoglobulin Domain Containing 8	VSIG8	-
38	22300140	22307357	SLAM Family Member 8	ENSCAFG00000011795 (SLAMF8)	2 upstream
38	22312491	22320700	Fc Receptor Like 6	FCRL6	2 intronic; 1 intergenic
38	22344722	22344910	Dual Specificity Phosphatase 23	ENSCAFG00000031934 (DUSP23)	1 downstream; 2 intergenic
38	22396604	22399244	C-Reactive Protein	*CRP	6 downstream
38	22456044	22457102	unknown	ENSCAFG00000039447	3 downstream

C.15 Runs of homozygosity (ROH)

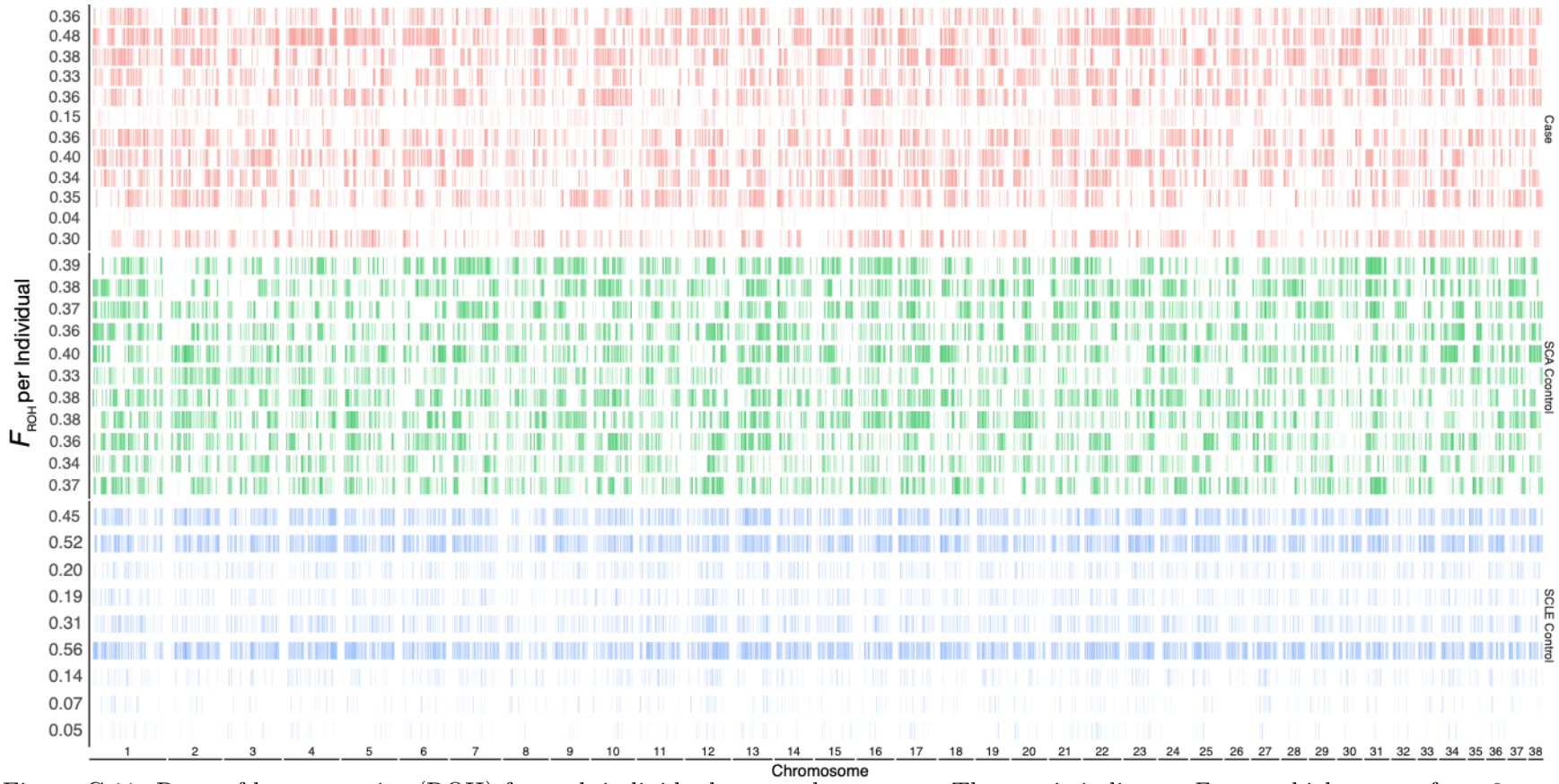

Figure C.11: Runs of homozygosity (ROH) for each individual across the genome. The y-axis indicates F_{ROH}, which ranges from 0 to 1, for each individual. The individuals are separated by population with cases shown in salmon, SCA controls in green, and SCLE controls in light blue.

C.16 Discussion of Candidate Genes

Many of the putative candidate genes have previously been shown to be involved in tumorgenesis. For example, Protein kinase C (PKC) is a family of isozymes that play major roles in the control of signaling pathways associated with proliferation, migration, invasion, tumorigenesis, and metastasis (Cooke et al. 2017). PKC represents one of the most extensively studied kinases, with >60,000 citations in PubMed and >10,000 citations associated with cancer (Cooke et al. 2017). PKCδ mainly functions as an anti-proliferative kinase that negatively regulates cell cycle progression (Griner & Kazanietz 2007). Furthermore, overexpression of PKCδ in the epidermis of transgenic mice protects against tumorigenesis when exposed to the tumor promoting phorbol esters (Reddig et al. 1999). Contrary to this, PKCδ has tumor promoting activity in mammary gland and pancreatic tumorigenesis (Mauro et al. 2010; Allen-Petersen et al. 2014).

Reticulon 1 (RTN1) proteins are members of highly conserved reticulons, which are localized in the endoplasmic reticulum (ER). Reticulons show pro-apoptotic activity via the induction of ER stress (Kuang et al. 2005; Di Sano et al. 2007). Reduced expression of RTN1 in colon adenocarcinomas (Lemire et al. 2015) and highly expressed in small cell and non-small cell lung cancers (van de Velde et al. 1994; Senden et al. 1996; 1997a; b). DPP10 may play a role in disease progression of colorectal cancer and loss of DPP10 expression in primary CRC is significantly associated with poor survival outcomes (Park et al. 2013). Another gene, Cilia And Flagella Associated Protein 45 (CFAP45; Coiled-Coil Domain Containing 19 (CCDC19)), is highly associated with both pharynx, nasopharyngeal, and lung carcinoma in humans (Liu et al. 2011) (Wang et al. 2018). This gene has been found to regulate miR-184 and thereby suppress cell proliferation, invasion and migration (Liu et al. 2014). Decreased expression of CCDC19 is correlated with poor prognosis in lung cancer patients (Wang et al. 2018). With SNPs within the TFBS's and strong evidence for selection across these regions, these genes are likely involved in tumor suppression/promotion in SCA

foxes.

Immunological surveillance hypothesis states that in large, long-lived animals, heritable genetic changes are common in somatic cells and represent progression to malignancy. Further, there must be a mechanism, likely immunological in character, for eliminating potentially dangerous mutant cells. In the case of SCA and ceruminous gland carcinoma, two genes, FCRL6 and ENSCAFG00000011795 (Ortholog of SLAMF8), were both in regions of the genome highly differentiated between cases and controls, have immunological functions, and disruption in these genes have been found to lead to tumorigenesis. For example, FCRL6 distinguishes mature cytotoxic lymphocytes and has been found to be upregulated in patients with B-cell chronic lymphocytic leukemia (Schreeder et al. 2008). Slam family member antigens are lymphocyte activation proteins involved in the pathogenesis of immunological disorders, and may also contribute to the activation of cancer cells (Furukawa et al. 2010). Exploring expression of these genes may be informative to better understand cancer susceptibility in SCA foxes.

C.17 Discussion of Candidate Genes references

Allen-Petersen BL, Carter CJ, Ohm AM, Reyland ME (2014) Protein kinase Cδ is required for Erbβ2-driven mammary gland tumorigenesis and negatively correlates with prognosis in human breast cancer. 33, 1306-1315.

Cooke M, Magimaidas A, Casado-Medrano V, Kazanietz MG (2017) Protein kinase C in cancer: The top five unanswered questions. Molecular carcinogenesis, 56, 1531-1542.

Di Sano F, Fazi B, Tufi R, Nardacci R, Piacentini M (2007) Reticulon-1C acts as a molecular switch between endoplasmic reticulum stress and genotoxic cell death pathway in human neuroblastoma cells. Journal of neurochemistry, 102, 345-353.

Freedman AH, Gronau I, Schweizer RM et al. (2014) Genome sequencing highlights the dynamic early history of dogs. (L Andersson, Ed,). PLoS Genetics, 10, e1004016-12.

Furukawa H, Tohma S, Kitazawa H et al. (2010) Role of SLAM-associated protein in the pathogenesis of autoimmune diseases and immunological disorders. Archivum immunologiae et therapiae experimentalis, 58, 37-44.

Griner EM, Kazanietz MG (2007) Protein kinase C and other diacylglycerol effectors in cancer. Nature Reviews Cancer, 7, 281-294.

Kuang E, Wan Q, Li X et al. (2005) ER Ca2+ depletion triggers apoptotic signals for endoplasmic reticulum (ER) overload response induced by overexpressed reticulon 3 (RTN3/HAP). Journal of cellular physiology, 204, 549-559.

Lemire M, Qu C, Loo LWM et al. (2015) A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. Human Genetics, 134, 1249-1262.

Liu Z, Li X, He X et al. (2011) Decreased expression of updated NESG1 in nasopharyngeal carcinoma: Its potential role and preliminarily functional mechanism. International Journal of Cancer, 128, 2562-2571.

Liu Z, Mai C, Yang H et al. (2014) Candidate tumour suppressor CCDC19 regulates miR-184 direct targeting of C-Myc thereby suppressing cell growth in non-small cell lung cancers. Journal of Cellular and Molecular Medicine, 18, 1667-1679.

Mauro LV, Grossoni VC, Urtreger AJ et al. (2010) PKC Delta (PKCδ) Promotes Tumoral Progression of Human Ductal Pancreatic Cancer. Pancreas, 39, e31-e41.

Park HS, Yeo HY, Chang HJ et al. (2013) Dipeptidyl peptidase 10, a novel prognostic marker in colorectal cancer. Yonsei Medical Journal, 54, 1362-1369.

Reddig PJ, Dreckschimdt NE, Ahrens H et al. (1999) Transgenic Mice Overexpressing Protein Kinase Cδ in the Epidermis Are Resistant to Skin Tumor Promotion by 12-O-Tetradecanoylphorbol-13-acetate. Cancer Research, 59, 5710-5718.

Schreeder DM, Pan J, Li FJ, Vivier E, Davis RS (2008) FCRL6 distinguishes mature cytotoxic lymphocytes and is upregulated in patients with B-cell chronic lymphocytic leukemia. European Journal of Immunology, 38, 3159-3166.

Senden NHM, Timmer EDJ, BruÃ¯ne A de et al. (1997a) A Comparison of NSP-Reticulons with Conventioal Neuroendocrine Markers in Immunophenotyping of Lung Cancer. The Journal of Pathology, 182, 13-21.

Senden NH, Timmer ED, Boers JE et al. (1996) Neuroendocrine-specific protein C (NSP-C): subcellular localization and differential expression in relation to NSP-A. European journal of cell biology, 69, 197-213.

Senden N, Linnoila I, Timmer E et al. (1997b) Neuroendocrine-specific protein (NSP)-reticulons as independent markers for non-small cell lung cancer with neuroendocrine differentiation. Histochemistry and Cell Biology, 108, 155-165.

van de Velde HJ, Senden NH, Roskams TA et al. (1994) NSP-encoded reticulons are neuroendocrine markers of a novel category in human lung cancer diagnosis. Cancer Research, 54, 4769-4776.

Wang Y, Liu Z, Luo R, X Y (2018) Decreased CCDC19 is correlated with unfavorable outcome in lung squamous cell carcinoma. International Journal of Clinical and Experimental Pathology, 11, 802-807.