UNIVERSITY OF CALIFORNIA
SANTA CRUZ

A GENOMIC APPROACH TO STUDYING THE EVOLUTIONARY CONSEQUENCES OF POPULATION DECLINES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR ENGINEERING AND BIOINFORMATICS

by

Nedda F. Saremi

September 2020

The Dissertation of Nedda F. Saremi
is approved:

Professor Beth Shapiro, Chair

Professor Richard E. Green

Professor Giacomo Bernardi

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

A genomic approach to studying the evolutionary

consequences of population declines

by

Nedda F. Saremi

Since the release of the human genome sequence in 2003, decreasing costs of DNA sequencing and advances in laboratory techniques have allowed scientists to study demography, admixture, and evolution for a growing diversity of taxa beyond humans. The work I present in this dissertation is part of the continued expansion of genomics. In particular, I present the assembly and analysis of whole genome data of extant and extinct species to study the evolutionary processes and consequences of population reduction.

My first chapter looks at the effects of habitat loss in the puma, whose range is the largest of any felid in the Western Hemisphere. I present a genome assembly of a puma from the Santa Cruz Mountains, and resequencing data for a panel of pumas from across their current range. I learn about the genomic health of the species, uncover Central American ancestry in present-day Florida pumas, and present a new model for the demographic history of the species.

In my second chapter, I present an assembly of the genome of the sole bear species to inhabit South America, the spectacled bear. The spectacled bear is the closest living relative to the extinct giant short-faced bears that roamed North America during the last Ice Age. I use the genome of the spectacled bear as a reference for mapping the genome of the extinct giant

short-faced bear, and analyze both genomes to learn their population histories and relationships to extant ursids.

In my third chapter, I focus on another extinct bear, the California grizzly bear. Having gone extinct in the early 1900s due to overhunting, the California grizzly bear is an example of the detrimental results humans impose on their environment. Using genomic data from two preserved California grizzly bears, as well as the genomes of several modern brown bears, I characterize the genetic diversity among extant brown bears and explore the diversity of California grizzly bears.

Together, the research that comprises my dissertation presents a significant contribution to the growing body of literature for non-model organisms and provides resources relevant to the conservation of top predators.

For my mom and dad.

Acknowledgments

First, thank you to my committee members, Professor Beth Shapiro, Professor Ed Green, and Professor Giacomo Bernardi for dedicating their time to help guide my research. Beth and Ed, I'm extremely grateful to you both for letting me rotate in your lab when no one else would, and ultimately letting me join your amazing lab. Giacomo, thank you for your invaluable insight and encouragement.

I cannot begin to express my thanks to the Paleogenomics lab, versions 2014-2020. The knowledge I gained early on in the 'sharktank' is even more important and impactful now than it was then. I am truly grateful that my graduate school years were surrounded by such smart and helpful individuals.

To the local Santa Cruz soccer leagues and my teammates for providing me respite.

Thank you to my cohort, classmates, and friends. I've learned so much from you within the context of graduate school and outside of it. I am so grateful to have had such delightful people to have shared this experience with. My time and memories with you all have helped me through the ups and downs of academia to enable me to reach this point.

Thank you Robert for our talks about bears, bursting, figures and breaking a stick into three pieces to form a triangle. I love you and look forward to many more enthralling discussions.

And lastly, thank you to my mom, dad, and sister. Your encouragement, support and love mean the world to me.

The text of this dissertation includes a reprint of the following previously published material:

Saremi, N. F. et al. Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature Communications* 10, 4769 (2019).

The co-authors listed in this publication directed and supervised the research which forms the basis for chapter one of my dissertation.

Introduction

0.1 DNA and genomics

The genome encodes for all the information needed to develop and create a living organism. By studying the genomic sequence of an organism, we can learn about an organism's evolutionary history, providing insight into processes such as admixture, selection, and past demography.

A key discovery in the field of genomics lay in the advent of DNA sequencing in the 1970s, when Sanger et al. [236] reported the complete 5,375 nucleotide DNA sequence of the bacteriophage ΦX174. With this discovery and Sanger's new approach, DNA sequences could be resolved at a nucleotide resolution within a few hours, with sequence lengths reaching greater than 500 bases. Sanger sequencing represented the first generation of sequencing technologies, and is still widely used today for validation and smaller scale projects in the field.

Second generation sequencing, commonly called next-generation sequencing (NGS), presented a move to a large-scale, massively parallel sequence data output. The key to the success of NGS was the ability to sequence a multitude of different DNA template strands at the same time.

With the advent of NGS in the early 2000s, the field and study of genomics has grown at an impressive speed. The continued decrease in sequencing costs [5] has allowed for the generation of vast amounts of sequence data. In 2005, the first commercially available sequencer, 454's GS-20, generated 2.2 billion base pairs of data [233]. Today, one Illumina NovaSeq run generates up to 3,000 billion base pairs of sequence data, providing nearly 1,000 X coverage of the human genome [2]. Early on, the quick generation of large amounts of sequence data allowed researchers to assess conservation between long diverged species [42], within a species [139], and with extinct species [97].

While NGS provides a wealth of DNA sequences, one of its limitations is its sequence read length (50-150 base pairs). Third generation sequencing, also known as long-read or single molecule sequencing, addresses the short read length concern. Typical read lengths for the two prominent third generation technologies, Pacific Biosciences of California and Oxford Nanopore Technologies, at the present time average in the thousands of base pairs, while maximum read lengths have reached upwards of 100,000 base pairs [281]. However, some long read technologies suffer from error rates as high as 10 to 15%, which require correction. Today, all three technologies are used, sometimes collectively [131], to provide a well-rounded approach to genomic research.

Sequenced and assembled genomes allow us to investigate gene function, genomic variation, diseases, species relatedness, and a growing number of topics that would be largely impossible without a full genetic blueprint. The genetic blueprint, or genome assembly, represents the full DNA sequence for a living organism. For humans, this sequence is 3.2 billion nucleotide bases in length, and represents only one of the two copies of genome found in almost

every cell in the body.

The Human Genome Project was launched in 1990 by the National Institute of Health (NIH) and the Department of Energy (DOE) with the goal of sequencing and mapping the human genome, model organism genomes, and developing technology to study DNA [193]. By 2001, the International Human Genome Sequencing Consortium (IHGSC) reported on their initial analysis of the human genome generated by Sanger sequencing [149]. By that point, scientists had already sequenced the genomes of yeast, *Escherichia coli*, nematode, fruit fly, *Arabidopsis thaliana*, and mouse [93, 26, 45, 10, 276, 42]. In 2003, the IHGSC completed deciphering the DNA sequence which defined a human being [118]. Four years later, a human genome sequence from one individual (the Human Genome Project was a composite) had been completed, again using Sanger technology [155]. Second-generation sequencing methods had developed enough by 2008 that scientists could assemble a human genome using two different sequencing platforms. First, using technology created by 454 Life Sciences [297], and followed shortly by technology from Illumina, Inc [23, 290, 156]. The Human Genome Project propelled the field of DNA sequencing and genomics (Figure 0.1).

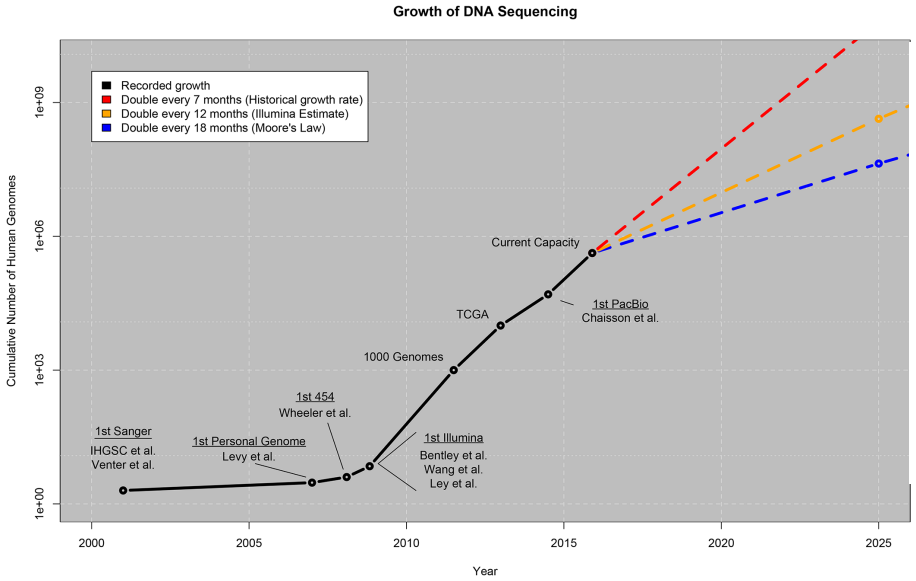

Figure 0.1: Growth of DNA sequencing measured in the total number of human genomes sequenced taken from Stephens et al.[265]. Notable milestones for each generation of sequencing technology are shown, as well as large-scale sequencing projects such as 1000 Genomes Project (a thousand human genomes) [275] and The Cancer Genome Atlas (TCGA) (several thousand tumor/normal genome pairs [41]). Values beyond the year 2015 represent projections based on three possible growth curves.

By 2010, the cost of sequencing a genome had decreased almost four orders of magnitude from the year the draft human genome sequence was announced in 2001 [117]. The decrease in sequencing costs made genome sequencing more widely accessible, and thus scientists sought to generate large catalogues of genomic data to use for research. One of the first was the 1000 Genomes Project, which sought to detail human genetic variation by sequencing the genomes of at least 1,000 anonymous participants from different populations [69].

Continued cost reduction also propelled genome assembly of non-human species. The Genome 10K project was one of the earliest such initiatives, with the proposal to sequence 10,000 species spanning the diversity of extant vertebrates [200]. The work has been so success-

ful, that subsequent phases widen the span of the species to be sequence. Today the consortium seeks to assemble reference-grade genomes for all vertebrate species. Success with broader initiatives has driven projects such as insect 5k (i5k) [230], Bat1K [274], and Bird 10,000 Genomes (B10K) initiative [303] to highlight focused lineages.

For those interested in better understanding a species, *de novo* genome assembly is now the current standard for studying its evolutionary history or genomic health, as my work in chapters 1 and 2 highlights. However, what can researchers do when the species of interest is extinct? The answer comes from ancient DNA (aDNA).

0.2 Ancient DNA

Ancient DNA is classified as DNA extracted from the remains of organisms that used to be living, sometimes tens or hundreds of thousands of years ago. As a result of its age, aDNA has a number of technical issues which make it more difficult to work with than DNA from living organisms. First, the DNA obtained from an ancient sample (museum specimen, archeological remains, etc) is fragmented. The sequence length of an ancient DNA fragment is very short, usually less than 100 base pairs [30]. Short sequences provide less information than the longer sequences that are easily obtained from living organisms. Thus, an ancient DNA sample requires more sequencing to generate the same amount of informative data as from a modern DNA sample.

Second, aDNA sequences show signs of postmortem degradation (PMD). Hydrolytic deamination is the one of the most common forms of DNA damage; the residue cytosine being

the main target [166]. Cytosine deamination effectively converts a cytosine residue into a uracil residue. When a deaminated DNA sequence is used as a template for amplification, the cytosine residues are read as thymine, resulting in a pattern of C to T substitutions [110] that increase near the 5' end of the sequence [30]. The 5' deamination coincides with a complementary G to A substitution pattern at the 3' end of the sequence due to repair [187], as seen in Figure 0.2. Interestingly, the frequency of 5' deaminations has shown a correlation with sample age, such that more ancient samples show greater levels of C to T substitutions [129].

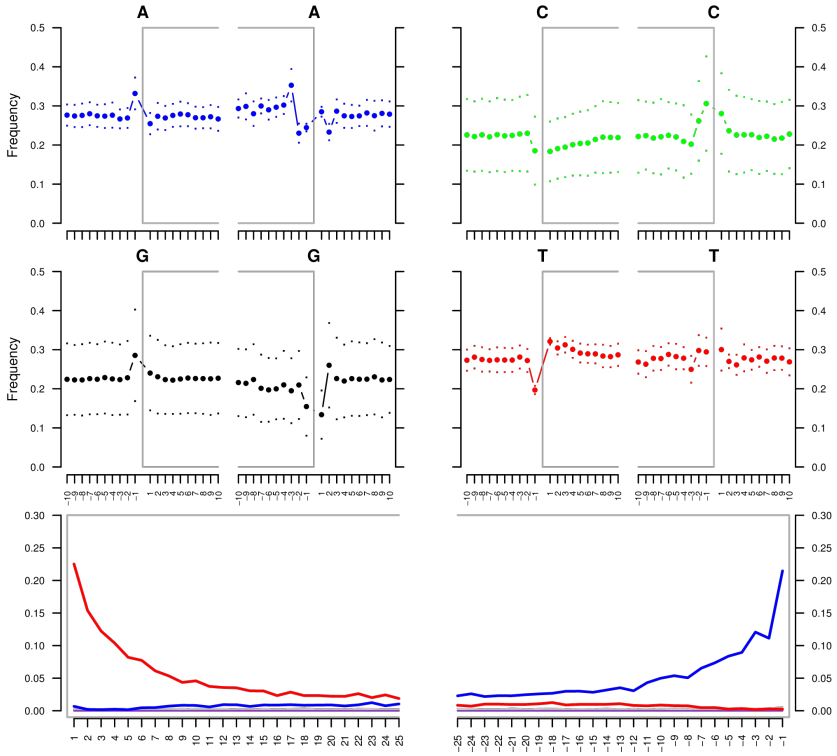

Figure 0.2: A deamination plot showing the postmortem degradation of an ancient pathogen extracted from bones and teeth [245] generated using mapDamage2 [129]. The four top plots show the according base frequency outside and in the read (depicted as the gray box). The bottom plots show the specific substitution at each position from the 5' end (left) and the 3' end (right). Colors indicate different substitution types (red: C to T, blue: G to A, gray: all other substitutions, orange: soft-clipped bases, green: deletions relative to the reference sequence, and purple: insertions relative to the reference sequence). The prevalence of C to T substitutions at the 5' end and G to A substitutions at the 3' end of sample sequences are used to validate the presence of authentic ancient DNA in a sample.

The third issue associated with ancient DNA is non-endogenous DNA, or DNA that is not from the host species. Upon the death of a species, microbes may invade the host organism, such that upon sequencing, a large portion of the generated data does not originate from the host.

Greater sequencing would be required to generate the same amount of informative data compare to a sample with little microbial contamination. Additionally, non-endogenous material may result from sample processing. Ancient DNA work is performed in a dedicated aDNA facility separate from modern DNA laboratories observing a number of guidelines to ensure decreased risk of contamination [204].

Early ancient DNA work was largely limited to mitochondrial DNA, as its high copy number made retrieval from ancient samples less difficult [99, 104, 64]. Over time, aDNA was successfully being amplified by PCR in many ancient samples [102, 114, 113], and was no longer limited to mitochondrial DNA [99]. Yet concerns regarding contamination and reproducibility, warranted based on prior studies [304], meant much time was spent discussing standards for aDNA work [292, 104].

The application of shotgun sequencing and NGS for aDNA work brought the field into a new realm of possibility. Scientists were able to directly sequence aDNA molecules, without PCR amplification [195]. Soon after, second-generation sequencing was generating millions of base pairs from aDNA samples [214, 97]. Today, scientists have generated high coverage nuclear genomes from many ancient species [187, 192, 296].

Research and methods developments have helped minimize the technical issues associated with aDNA. Limitations in analyzing NGS data from ancient DNA today are mostly computational; how to classify endogenous DNA and determine authentic nucleotide differences [218]. Some solutions include using closely related extant species to detect endogenous reads, and algorithms that account for damage seen in aDNA [244, 91]. I will elaborate on solutions for overcoming the challenges associated with aDNA in my work in chapters 2 and

3.

Regardless of technical and computational advancements, limitations of sample age still exist for the successful extraction of authentic ancient DNA. The current limits of successful authentic ancient DNA extraction are set by a sample from roughly 700 thousand years before present, a horse bone from Yukon, Canada [203]. Thus researchers are limited to samples from the Pleistocene epoch for DNA insights on the past.

0.3 Genome assembly

Advances in sequencing technologies, and reductions in the cost of sequencing have meant that genome-scale data is increasingly feasible for researchers (Figure 0.3). While data generation methods such as Restriction site Associated DNA-Sequencing (RAD-Seq), amplicon sequencing, or transcriptome sequencing can be used without a reference genome, a complete genome sequence is considered a benchmark for genomic research. Full genome sequences, as well as high quality annotations of the sequence, have the ability to reveal population histories [163], admixture events [189], and lineage-specific adaptations [221].

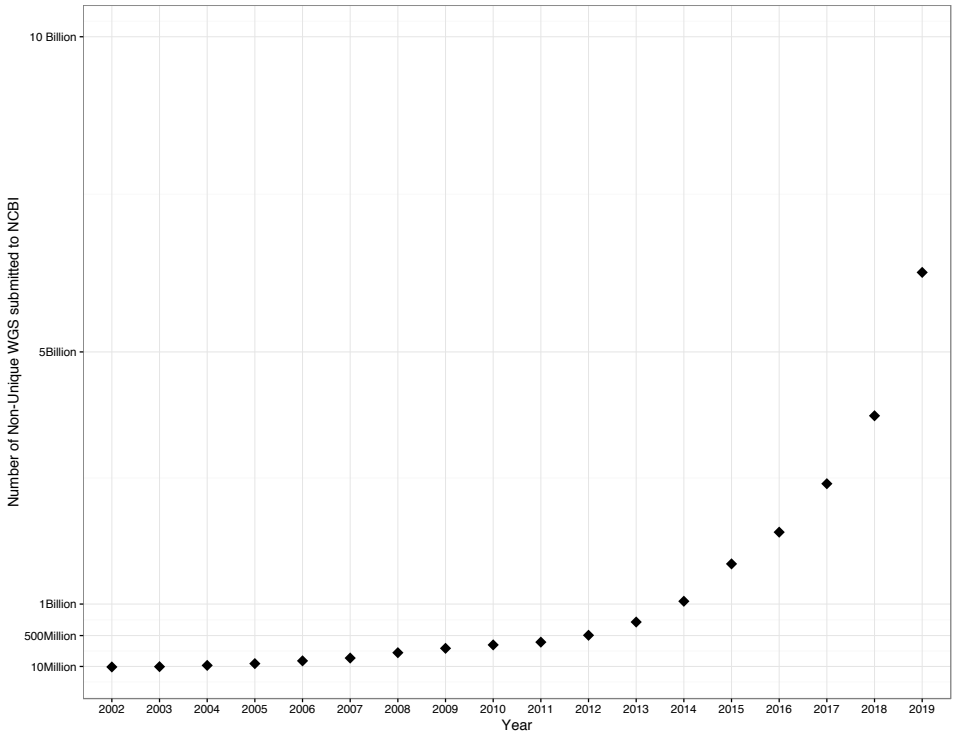

Figure 0.3: Yearly growth of submissions of Whole Genome Sequencing (WGS) projects to the NCBI database [1]. Submissions may not represent unique species.

0.3.1 Developments in *de novo* genome assembly

A *de novo* genome assembly is the generation of long contiguous DNA sequences from shorter DNA sequences without the use of any reference sequence. Early genome sequencing projects such as the human genome began by generating a map of the genome. A physical map containing several thousand sequence-tagged sites (STSs) known distances apart was proposed [201]. Short single-copy STSs could be easily amplified using Polymerase Chain Reaction (PCR) to determine proximity between fragmented regions of the genome [116]. A map served as a guide for the locations of genomic fragments of the genome that would be

sequenced. The genomic fragments themselves were initially inserted into artificial chromosomes, such as Bacterial Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs), which allowed for stable storage and easy cloning of the DNA fragment [250, 191]. This technique was termed hierarchical shotgun assembly due to the tiling of overlapping BAC clones which are sequenced as a single compartment and then merged with other compartment sequences [295].

Whole-genome shotgun (WGS) assembly is a less labor and time intensive process. The genome is fragmented and then sequenced, typically using paired-end technology to capture linkage information of known sizes. Then the entire genome is assembled into contiguous sequences of bases, contigs. There are risks associated with the WGS assembly approach, as the sheer amount of sequence information makes for computational challenges. Yet the shift from Sanger to NGS quickened the data generation process and increased the data volume, though with a reduction in read lengths.

While the lower cost of NGS data generation made it possible to sequence at high depth of coverage, assembling genomes remained a challenged. Changes in data coverage across the genome or a lack of overlapping reads in the shotgun data resulted in fragmented genome assemblies. Though NGS made high coverage data generation more accessible, it also meant a reduction in read lengths. Biological challenges, such as polyploidy, heterozygosity or repeat sequences, could not be overcome solely with greater shotgun sequencing.

Scaffolding was one method to improve the fragmented shotgun assemblies of shorter second generation reads. Some techniques included mate-pair sequencing and proximity ligation data. Mate-pair sequencing began with fragmenting the genome into long DNA of known

lengths (1-10 kilobases). Sequencing the end of the DNA fragments generates reads a known distance apart in the genome. If the mate-pair reads existed on two non connected contigs, mate-pair information could join them together into one scaffold, a known distance apart using a string of uninformative Ns, known as a gap. Proximity ligation uses long range contact information of chromosomes to detail interactions at a variety of distances [84, 220]. Models of contact frequency based on genomic distance are turn sequencing information into another means for scaffolding assembly contigs. Scaffolding methods such as these improved assembly quality such that shotgun *de novo* assemblies could generate scaffolds the size of chromosomes [32].

The implementation of scaffolding techniques improved the contiguity of genome assemblies, but did not fully overcome the problem of gaps. The initial chimpanzee genome assembly contained more than 6% of its sequence in gaps [294]. Gaps result from biological challenges, or limitations in the data or techniques used to generate an assembly. The advent of long read lengths in third generation sequencing has been the recent means of genome quality improvement. Long read lengths help overcome one of the biggest challenges in genome assembly, repeats. Repeat regions compose roughly 60% of vertebrate genomes [60]. The ability to generate sequencing reads which span the length of repeat or a gap in a scaffold results in a marked improvement in genome quality [76]. Continued improvement in the error rate of long-read sequencing methods has meant greater use of this technology for genome assembly. Researchers are even able to generate genome assemblies constructed predominantly or only from long-reads [94, 124]. Combining sequencing technologies has allowed researchers to obtain almost chromosome-level assemblies without physical mapping techniques [172].

0.3.2 Reference-guided genome assembly

While a *de novo* genome assembly represents the reference-free generation of the nucleotides sequence encoding an organism, a reference-guided genome assembly makes use of a genome from a closely related species onto which ancient DNA reads are mapped [215]. This direct comparison has meant the ability to assess divergence times [194], selection [123], or past population dynamics [66].

0.3.3 Paleogenomes

Though the cost of high-throughput sequencing has reduced, it is still prohibitively expensive to generate all the sequencing data needed for a *de novo* genome assembly. For ancient samples, the problem is exacerbated by low levels of endogenous DNA and short fragment lengths (40-170bp [238]).

However, due to advances in the extraction of ancient DNA and NGS technologies, scientists have been able to generate genomes of extinct species, paleogenomes. *De novo* assemblies of ancient DNA have been accomplished for pathogens, aided by their small genome size [29, 280]. For mammalian genomes, the vast amount of sequencing needed to generate a high coverage ancient dataset would mean destructive sampling. High coverage ancient genomes (20X), mapped to a closely related reference genome, soon after became reality [225]. High coverage ancient genomes enabled new analyses like genotyping and past demography that previously were impractical due to DNA damage signals obscuring real DNA sequence. Modern aDNA studies, using high coverage ancient datasets, have been able to uncover ancient species [187], the history of domestication [54], and the history of species [80]. Continued progress in

aDNA computational methods and laboratory techniques means a greater ability to understand evolution.

0.4 Conclusion

In the work described to follow, I present *de novo* genome assemblies for two megafaunal species, the puma and the spectacled bear, generated from second and third generation sequencing data. The puma represents a species whose North American range experienced significant habitat loss in the past 100 years, largely the result of humans. Prior research on the genetic effects of habitat loss and fragmentation on pumas were limited to mitochondrial and microsatellite data. Using a genome assembly generated from a puma from the Santa Cruz mountains, I learn about the genomic health of the species, uncover a Central American puma ancestry in the Florida pumas, and create a new model for the past demography of the species. This chapter was published in Nature Communications in October of 2019, for which I served as co-first author.

The second genome assembly I present here is for the spectacled bear, the only remaining member of the tremarctine bears. In the Pleistocene, up to seven tremarctine bear species roamed North and South America. Initially representing some of the largest land mammals at the time, the family increasingly became smaller over time, likely as a means to survive with decreasing prey availability and increasing competition for food by other large carnivores. The genome of the extant spectacled bear served as the reference used to study on of the extinct tremarctine bears, the giant short-faced bear. Using a high coverage paleogenome of the giant

short-faced bear, I assess its relationship to the spectacled bear and uncover traits that may have led to its demise during the late Pleistocene extinctions. This work is currently in preparation for submission.

My third chapter covers another bear species, the brown bear. The range of brown bear is the largest of any extant bear species. North American brown bears, also known as grizzly bears, are endangered today due mainly to depredation by humans. Their range once extended from Alaska to Mexico and from the Great Plains to California, where they were known as the California grizzly bearthe iconic symbol of the state. Very little is known about the California golden bear, including its specific adaptations and habits or its evolutionary relationships to other grizzly bears. Working with collaborators at the Museum of Vertebrate Biology (MVZ) at UC Berkeley, the California Academy of Sciences (CAS), and the the La Brea Tar Pits, I generated low coverage (5X) nuclear genomic data from two of the last California golden bears, one of whom served as the model for the bear on the California state flag. I perform population genomic analyses to answer questions about the history and uniqueness of the extinct California golden bear and its evolutionary relationships to modern brown bears. This work is also currently in preparation for submission.

Chapter 1

Puma genomes provide insights into the genomic consequences of inbreeding

Published in Nature Communications October 18, 2019

Nedda F. Saremi*[1], Megan A. Supple*[2], Ashley Byrne[3], James A. Cahill[2], Luiz Lehmann Coutinho[4], Love Dalén[5], Henrique V. Figueiró[6], Warren E. Johnson[7], Heather J. Milne[2], Stephen J. O'Brien[8], Brendan O'Connell[1], David P. Onorato[9], Seth P.D. Riley[10,11], Jeff A. Sikich[10], Daniel R. Stahler[12], Priscilla Marqui Schmidt Villela[13], Christopher Vollmers[1], Robert K. Wayne[11], Eduardo Eizirik[6], Russell B. Corbett-Detig[1], Richard E. Green[1], Christopher C. Wilmers[14], and Beth Shapiro[2,15]

*These authors contributed equally to this work.

Affiliations

[1] Department of Biomolecular Engineering and Bioinformatics, UC Santa Cruz

[2] Department of Ecology and Evolutionary Biology, UC Santa Cruz

[3] Department of Molecular, Cell, and Developmental Biology, UC Santa Cruz

[4] Laboratório de Biotecnologia Animal, Departamento de Zootecnia, ESALQ, Universidade de São Paulo

[5] Department of Bioinformatics and Genetics, Swedish Museum of Natural History

[6] Escola de Ciências, Pontifical Catholic University of Rio Grande do Sul

[7] Smithsonian Conservation Biology Institute, Smithsonian Institution

[8] Theodosius Dobzhansky Center for Genome Bioinformatics, Saint Petersburg State University

[9] Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission

[10] Santa Monica Mountains National Recreation Area

[11] Department of Ecology and Evolutionary Biology, UC Los Angeles

[12] Yellowstone Center for Resources, Yellowstone National Park

[13] EcoMol Consultoria e Projetos

[14] Environmental Studies Department, UC Santa Cruz

[15] Howard Hughes Medical Institute

1.1 Abstract

Pumas are the most widely distributed felid in the Western Hemisphere. Increasingly, however, human persecution and habitat loss are isolating puma populations. To explore the genomic consequences of this isolation, we assemble a draft puma genome and a geographically broad panel of resequenced individuals. We estimate that the lineage leading to present-day North American pumas diverged from South American lineages 300 to 100 thousand years ago. We find signatures of close inbreeding in geographically isolated North American populations, but also that tracts of homozygosity are rarely shared among these populations, suggesting that assisted gene flow would restore local genetic diversity. The genome of a Florida panther descended from translocated Central American individuals has long tracts of homozygosity despite recent outbreeding. This suggests that while translocations may introduce diversity, sustaining diversity in small and isolated populations will require either repeated translocations or restoration of landscape connectivity. Our approach provides a framework for genome-wide analyses that can be applied to the management of similarly small and isolated populations.

1.2 Background

The ancestors of the puma, *Puma concolor*, also known as the mountain lion or cougar, colonized North America approximately 6 million years ago (mya) [283, 175, 127]. Although their Pliocene fossil record is sparse and felid fossil assignments have been difficult, previous mitochondrial analyses suggested that the ancestral puma lineage diverged from the extinct North American cheetah-like cat *Miracinonyx* around 3.2 mya [17]. The geographic ori-

gin of *P. concolor* remains contested, however. At sites across North America, the oldest puma fossils date to the Rancholabrean land mammal age [21], roughly 200 thousand years ago (kya) [87]. Analyses of mitochondrial and microsatellite data, however, estimated that the common ancestor of North American pumas lived within the last 20,000 years [50, 178] and that the genetic diversity of all modern pumas traces to eastern South America [50]. The combination of genetic and fossil data were interpreted as reflecting a North American origin of the puma lineage followed by local extinction in North America during the late Pleistocene and subsequent recolonization from South America as the climate warmed after the last ice age [50, 178]. Recently, however, an unequivocal puma fossil was discovered in Argentina that dates to 1.2 to 0.8 mya [40]. This discovery pushes back the age of the puma lineage by more than 500,000 years, and suggests that the ancestor of all living pumas may have evolved in South America rather than North America. Today, pumas are among the most widely-distributed mammals in the Western hemisphere, ranging from Canada's Yukon to the southern tip of South America shown in Figure 1.1 [196, 119]. During the 19^{th} and 20^{th} centuries, bounty hunting reduced, and in some cases extirpated, puma populations across North America [267], restricting them to the North American West and the southern tip of Florida. By the middle of the 20^{th} century, hunting quotas and some outright bans [82] allowed puma populations to increase and recolonize parts of their former range. Although some puma populations today are large and well-connected [112], others are small and fragmented (e.g. Santa Ana, CA [285], Santa Monica Mountains, CA [228], and/or critically endangered (e.g. Florida [128]). Many populations are experiencing increased isolation with the expansion of highways, residential developments, and agriculture [285, 228]. The consequences of geographic isolation on puma genetic diver-

sity and fitness have been well documented, particularly in Florida, where they are a federally protected subspecies commonly called the Florida panther. By the 1990s, the canonical Florida panther population in Big Cypress National Preserve was suffering from reproductive failure and phenotypic defects associated with inbreeding [128, 18]. To rescue the Florida panthers from extinction, eight female pumas from Texas were released in South Florida in 1995, of which five successfully produced offspring. By 2008, the occurrence of phenotypic defects had significantly declined, survival measures had improved, and the population size increased almost threefold [128, 202]. All Florida panthers genotyped since 2012 show ancestry that includes admixture with the introduced lineages [284].

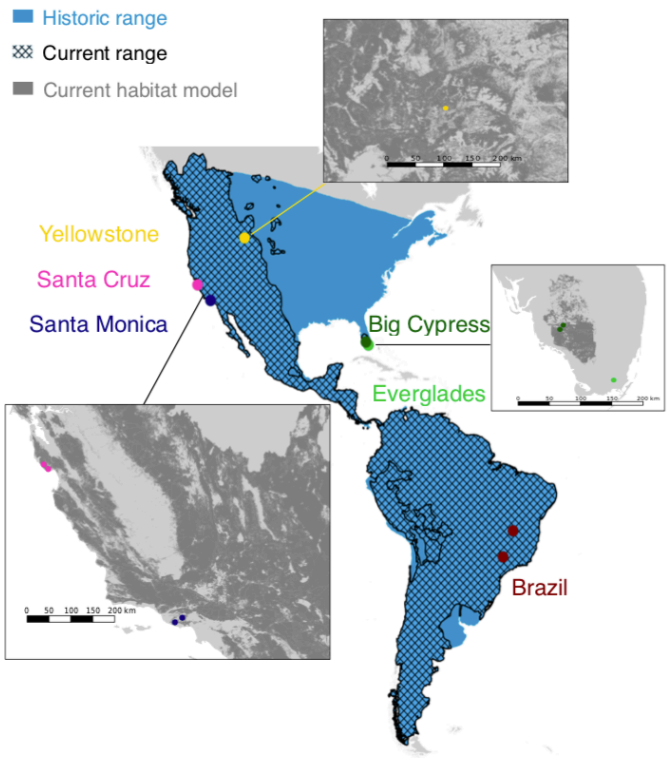

Figure 1.1: Puma range past and present. The current range of pumas (hashed) compared to their historic range (blue). Circles denote the geographic coordinates of the puma populations sampled in this study. Panels show zoom-ins of puma habitat distribution (dark gray) within the known range of the species in the contiguous United States as predicted by the USGS [269]. Current range data is from the IUCN Red List of Threatened Species [119]. Historic range data is approximated based on prior reports [196, 82].

Florida panthers in Everglades National Park are partially isolated from the core canonical population that persisted in Big Cypress National Preserve by a semipermeable barrier associated with hydrologic fluctuations of the Everglades. Intriguingly, during the 1990s, the Everglades panthers did not show the same high incidence of inbreeding-associated phenotypes as in the Big Cypress population. The absence of observed phenotypic defects in the Everglades population may be attributable to the release of captive-bred Florida panthers with

mixed Central American ancestry into Everglades National Park during the 1950s and 1960s. The introduced individuals'ancestry was unclear at the time of release, although it was known that the captive population had greater reproductive success than wild Florida panthers [197]. The admixed ancestry of the Everglades population and potential explanation for the reproductive success of the captive population was later discovered through genotyping [198]. Genetics has a long history as a tool in wildlife conservation [86]. In traditional conservation genetics, researchers sequence a small number of genetic markers across a large sampling of the species of interest. Advances in sequencing technologies have made it possible to sequence whole genomes of non-model organisms, including species of conservation concern. While the cost of sequencing continues to decrease, sequencing whole genomes will undoubtedly remain more costly than sequencing only a handful of genetic loci. This presents a choice: whole genome data sets exchange spatial resolution for finer-scale genomic resolution, allowing researchers to test different hypotheses. Each whole genome contains a multitude of largely-independent genealogies, that provides increased power to infer past events [115, 241]. In particular, the dense haplotype information provided by whole genomes is necessary to examine the very short timescales relevant to conservation efforts [133, 207]. Here, we reconstruct the last two million years of puma demographic history by generating and analyzing a draft genome from an individual sampled in the Santa Cruz Mountains (California, USA), along with nine resequenced genomes from pumas from North and South America. We confirm the recent maternal ancestry of North American pumas and describe genomic diversity in the sampled populations. We use shared tracts of homozygosity to predict the effectiveness of assisted gene flow in restoring lost genetic diversity. Finally, we analyze the genome of a Florida panther with admixed

ancestry that was collected 30 years after the first release of Central American admixed pumas into the Everglades. This genome allows us to assess the long-term efficacy of inter-population admixture as a means to rescue small and isolated populations from the deleterious effects of inbreeding.

1.3 Results

1.3.1 Genome assembly and variant calling

We assembled a de novo nuclear genome for a wild male puma (SC36) from the Santa Cruz Mountains using a combination of shotgun Illumina (47X coverage), long-range linking Illumina, and Oxford Nanopore Technology (ONT) (1.2X coverage) [125] data (see Methods). Our PumCon1.0 assembly has a BUSCO [252] score of 93.0%, a scaffold N50 of 100 Mb, and 87.6% of the genome represented on 26 autosomal scaffolds, each larger than 20 Mb (Tables 1.1, 1.2). Although our ONT coverage was only 1.2X, the use of these data for gap-filling recovered an additional 5.74% of the genome sequence, which we error-corrected by re-mapping the Illumina reads (Table 1.1). We obtained high-coverage (27X-55X) whole-genome resequencing data from nine additional pumas from locations in North and South America (Figure 1.1 and Table 1.3), and aligned the data to our reference assembly (PumCon1.0) for variant calling. We produced three final call sets: the first containing 8 million variable sites using the 10 pumas, the second decreased to 166,037 variable sites after filtering the first call set for linkage disequilibrium (LD), and the final call set containing 557,741 SNPs after LD filtering using the 10 pumas and the African cheetah (see Methods).

1.3.2 Demographic history

We reconstructed puma demographic history using both mitochondrial and nuclear genomes. Analyses of mitochondrial DNA estimate the most recent common maternal ancestor of all sampled pumas ~300 kya (Figure 1.2A). North American mitochondrial haplotypes cluster together, sharing an inferred common maternal ancestor 31-11 kya. The North American mitochondrial clade excludes the Florida Everglades puma (EVG21), which has a mitochondrial ancestry that is distinct from the rest of North America, consistent with the reported mixed ancestry of this individual [128]. The nuclear genomic data revealed a similar demographic history to that inferred from the mitochondrial data, and allowed us to estimate changes in effective population size over time. Pairwise sequentially Markovian coalescent (PSMC) modeling of the nuclear genomic data suggested that two puma lineages, one represented by the two Brazilian individuals and the other represented by all individuals sampled in North America, diverged by 300-100 kya (Figure 1.2B and Figure 1.9), similar to the age of the oldest puma fossils in North America. Populations on both continents were largest around 130 kya, during the warmest part of Marine Isotope Stage (MIS) 5, and then declined throughout the colder MIS 4-2, with populations reaching their current small sizes by the peak of the last ice age 25-20 kya. Our North American pumas show a continued increase in effective population size between 500 and 200 kya, whereas the effective population size of the Brazil pumas stabilized. This increase may reflect an increase in numbers during colonization of unoccupied habitats in Central and North America, but may also be attributable to PSMC modeling overestimating effective population size when a species has divided into subpopulations [179]. To test whether this observed

peak was the result of population structure, we modeled pseudo-diploid individuals using the X chromosomes of our male pumas (see Methods). We found evidence of a cessation of gene flow between all Brazil-North America pseudo-diploid male puma pairs by at least 100 kya, as shown by the sharp increase in the inferred effective population size (N_e), signifying no coalescent events occurred more recently than the estimated divergence time (Figure 1.11). The divergence dates obtained from the pseudo-diploid X chromosome PSMC analysis overlapped the time at which North and South American inferred N_e began to differ in the autosomal PSMC model. Thus, structure alone was not the reason behind the observed increase in N_e during this time. The spike in effective population size observed for EVG21 probably does not reflect the coalescent process within a single population, but is instead consistent with mixed ancestry comprised of two divergent lineages [35].

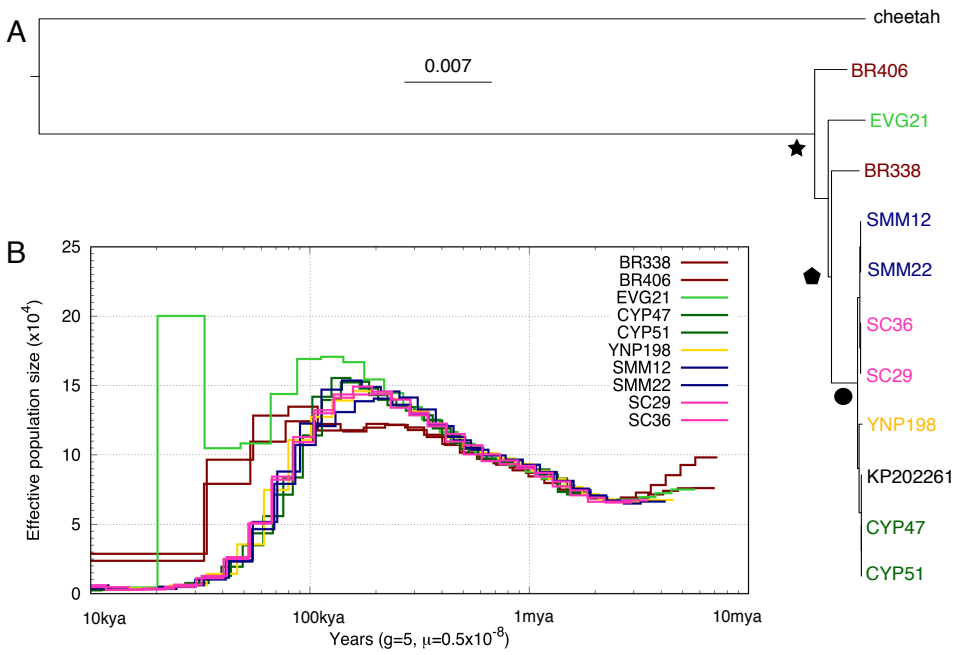

Figure 1.2: Demographic history of pumas. A) Mitochondrial maximum likelihood phylogeny of the ten pumas in this study plus an additional puma from Big Cypress (KP202261.1) and the African cheetah (KP202271.1) as the outgroup. We calculated divergence times by determining the number of pairwise divergences between sequences and used a mitochondrial divergence rate of 1.15%bp per Myr [170, 50]. We estimate a common maternal ancestor of these pumas 278,000 $\pm$ 5,639 years ago (star; 100% bootstrap support), divergence between North American and South American mitochondrial lineages 201,000 $\pm$ 1,952 years ago (pentagon; 63% bootstrap support), and a common maternal ancestor of North American pumas 21,000 $\pm$ 10,412 years ago (circle; 100% bootstrap support). B) Inferred changes in effective population size (N_e) over time using the pairwise sequentially Markovian coalescent (PSMC) model [163] for the ten pumas. We assume a generation time of 5 years and a per generation mutation rate of 0.5e-8 [43] per bp per generation. The PSMC plot for EVG21 shows a sharp increase in inferred N_e that is probably attributable to its hybrid ancestry [35].

1.3.3 Population structure

We used the nuclear genomic data to characterize genetic structure among puma populations (Figure 1.3). We performed principal component analysis (PCA) of 166,037 LD-filtered SNPs and found evidence of a geographic pattern (Figure 1.3A). The first two axes of the PCA, which explain 52% of the genetic variance, separated North and South America and revealed a gradient of relatedness from east to west across North America. The Everglades puma (EVG21) fell between the Big Cypress and Brazil populations, consistent with this individual's known history of admixture. Pumas sampled from the same population clustered together. We estimated a consensus nuclear phylogeny from 557,741 SNPs from the LD-filtered data set that included ten pumas and the African cheetah. This analysis found further evidence of structure, with the highest likelihood tree including a single migration event from a South (or Central) American lineage into the Everglades lineage (Figure 1.3B and Figures 1.12, 1.13). Finally, our cluster assignment tests based on the puma only LD-filtered SNPs also partitioned the data geographically, first separating out the two California populations at K=2, and then the Florida and Brazil populations at K=3 (Figure 1.3C and Figure 1.14). Notably, EVG21 shares ancestry with both Florida and Brazil at all K values (Figure 1.15). We note that the discrete populations identified in these analyses could simply reflect the spatial sampling of our data set. Spatially structured sampling can cause analyses to report distinct populations even when no discrete population structure exists [246]. This artifact is particularly likely to occur when geographically widespread samples are taken from well connected species where limited dispersal results in the accumulation of local genetic variants, resulting in genetic isolation by distance [300].

However, the observed geographic structure could also be the result of discrete genetic structure due to population isolation. Some puma populations have experienced persecution and degradation of their habitat, resulting in limited gene flow between populations [78, 227, 182, 128]. These isolated populations would show increased divergence over time, resulting in geographic structure.

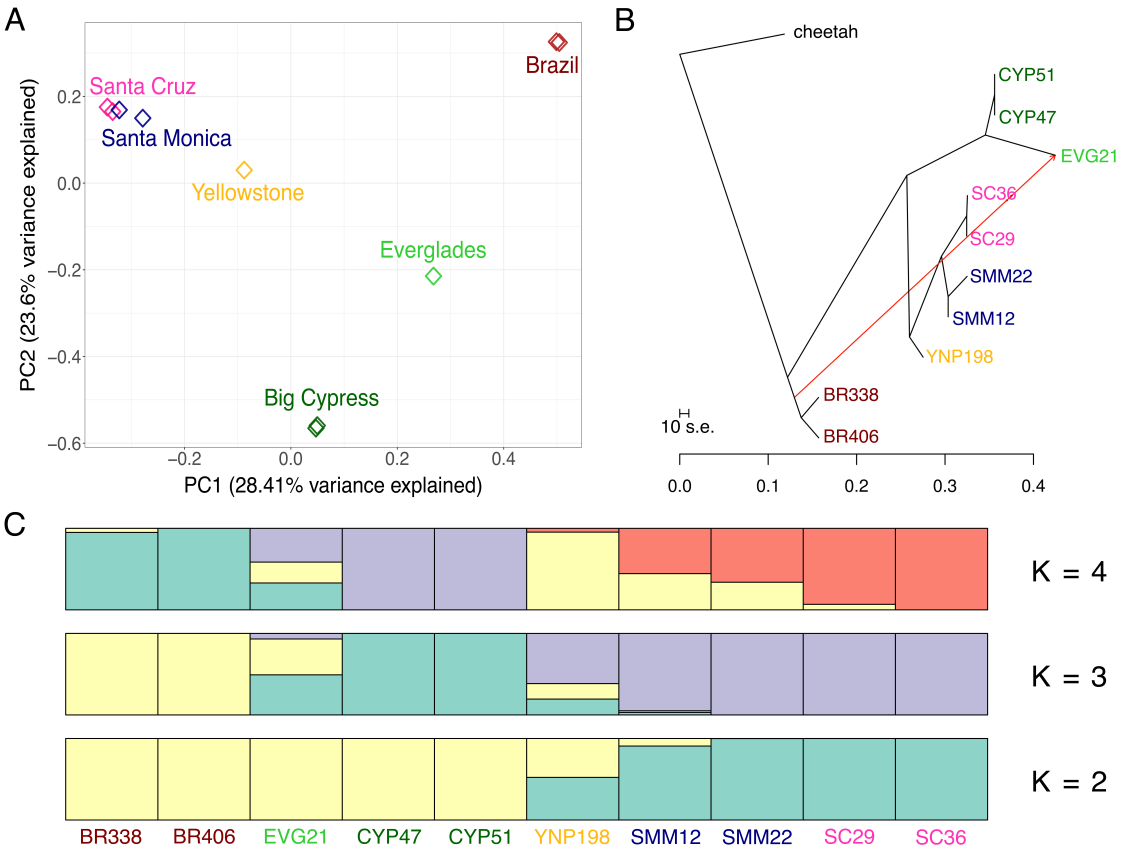

Figure 1.3: Stratification of pumas based on geographic population. A) Principal component analysis [208] of 166,037 sites separates the sampled pumas based on population. The first component primarily separates South and North American pumas, while the second component distinguishes the variation within North America. All California pumas (Santa Cruz and Santa Monica) cluster closely. B) TreeMix [213] analysis, using the African cheetah as the outgroup, indicates the best tree separates pumas based on population and includes one migration event (weight = 0.453911) from the branch of South American diversity into the admixed Everglades puma (EVG21). C) The mean of 10 permuted matrices of STRUCTURE [81] analysis for each of K=2 through 4, performed using CLUMPP [126]. Both delta K and L(K) values indicated that K=3 was the best K (Supplementary Figure 9) [79].

1.3.4 Heterozygosity and inbreeding

To examine the extent of inbreeding in our puma samples, we estimated for each individual the average genome-wide heterozygosity and identified runs of homozygosity (ROH) across the 26 largest autosomal scaffolds (Figure 1.4, see Methods). We focused our analyses on ROH >2 Mb, as we were able to call these longer tracts with high confidence. Although genetic drift is a dominant evolutionary force in small populations, the strong correlation among linked sites that is characteristic of ROH >2 Mb requires close inbreeding, and would not be observed due to genetic drift alone [37, 210]. The distribution of ROH across the genome varied among scaffolds and individuals (Figure 1.4A and Supplementary Figure 12), as did average genome-wide heterozygosity and proportion of the genome in ROH (Figure 1.4B). The two pumas from Brazil were the least inbred, with the highest heterozygosity and smallest proportions of their genomes in ROH. Conversely, the Big Cypress Florida panthers sampled prior to the 1995 genetic rescue were the most inbred, with the lowest heterozygosity and the largest proportions of their genomes in ROH, consistent with the phenotypic defects recorded in these individuals [128]. The other North American pumas fell between these two extremes. Of the two individuals from the Santa Monica Mountains, SMM12 appeared to be less inbred than SMM22, with higher heterozygosity and a lower proportion of its genome in ROH. This is consistent with their origins, as genetic analysis suggests that SMM22 was likely born in the small and more isolated Santa Monica Mountains population south of US 101 freeway, whereas SMM12 was first observed in the larger and more connected population north of US 101 and dispersed into the Santa Monica Mountains as a subadult [228].

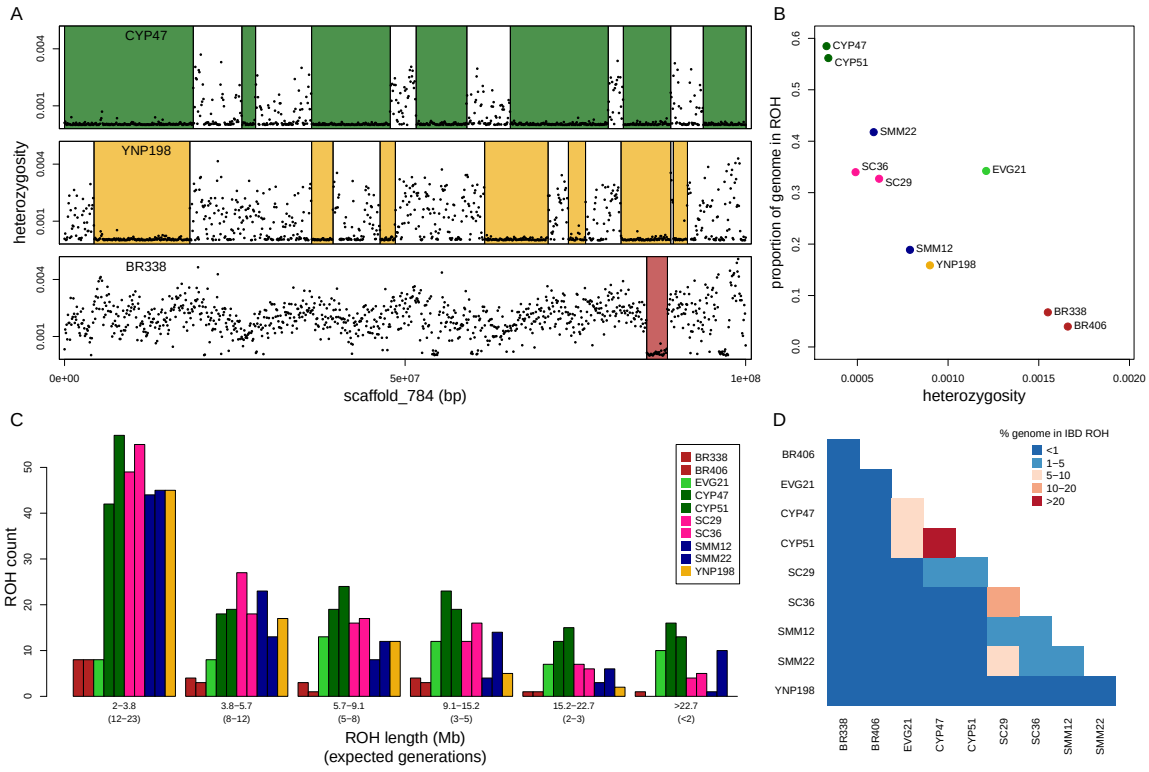

Figure 1.4: Heterozygosity and Runs of Homozygosity. A) Sliding window heterozygosity (black dots) and called ROH (colored boxes) across a single scaffold for three pumas from three different populations (Big Cypress, Yellowstone, and Brazil). Plots for all pumas are provided as Supplementary Figure 11. B) Average genome-wide heterozygosity versus the proportion of the genome in ROH for the ten pumas sequenced. C) Distribution of lengths of ROH. The length in Mb is indicated, as is the associated expected number of generations since the individual 's maternal and paternal lineages shared a common ancestor. D) Heat map showing the percent of the genomes that are in ROH that are shared IBD between pairs of pumas.

EVG21, the admixed Florida panther from the Everglades population, was an outlier in the general correlation between heterozygosity and proportion of the genome in ROH. The proportion of EVG21's genome in ROH was high relative to the expectation based on its average genome-wide heterozygosity. This is consistent with both ancestral admixture resulting in

a more diverse genetic background and close inbreeding leading to long tracts of homozygosity (Supplementary Figure 13). To better explore inbreeding history, we examined the distribution of ROH tract lengths in each puma. We correlated those lengths with the expected number of generations since the individual's maternal and paternal lineages shared a common ancestor using an estimated average recombination rate from the domestic cat of 1.1 cM per Mb [67] and the equation $g = 100/(2\ r\ L)$, where g is the time in generations, r is the recombination rate, and L is the length of the ROH tract in Mb [133, 277] (Figure 1.4C, see Methods). Long ROH (>15.2 Mb) occur due to close inbreeding (a common ancestor <3 generations ago). Short ROH (<5.7 Mb) occur due to shared ancestors further back in time (>8 generations ago). All North American pumas sampled had a large number of short ROH, indicating that these populations were small in the recent past (8-23 generations ago). The puma from Yellowstone had mostly short ROH and a small number of intermediate and long ROH, consistent with a population that was small in the recent past, but that does not suffer from a considerable amount of close inbreeding in recent generations. The pumas from the Santa Cruz and Santa Monica Mountains had patterns similar to the Yellowstone puma, except they had additional long ROH, suggesting that these populations are experiencing close inbreeding. The Big Cypress panthers each had many long ROH, which we estimated to reflect shared ancestors within the last three generations. The admixed Everglades panther, EVG21, had a small number of short ROH, similar to the Brazilian pumas, but had mostly long ROH, similar to the more inbred Florida individuals. This combination can be attributed to EVG21's complex history of admixture and inbreeding. EVG21 has historic admixture, and is the offspring of an inbreeding event- the sire of EVG21 was also EVG21's half brother [128] (Figure 1.8). The peak of the ROH length distribution for

EVG21 occurs at 5.7-9.1 Mb, indicating that EVG21's maternal and paternal lineages shared a common ancestor as far back as 5-8 generations, shortly after the admixture event that occurred 6-9 generations prior [128]. Although the sampled North American pumas all have long ROH, these tracts were generally not identical by descent (IBD) between individuals (Figure 1.4D). Long ROH that are also shared IBD between individuals are concerning because they represent regions of the genome with no genetic diversity in the four haplotypes analyzed. Of the pumas sequenced, only the two individuals from Big Cypress (CYP47 and CYP51) shared a considerable proportion (36%) of their genomes in ROH that are IBD between two individuals. The pumas from the Santa Cruz Mountains (SC29 and SC36) shared 12% of their genomes in IBD ROH, whereas the pumas that originate from different areas in and near the Santa Monica Mountains (SMM12 and SMM22) shared only 4%. Individuals from the Santa Cruz and Santa Monica Mountains shared between 3% and 5%. While most sampled North American populations show signs of close inbreeding, different populations are fixed for different variants and considerable genetic variation still exists when considering the species as a whole.

1.4 Discussion

We present a draft assembly of a puma genome, which we use to reconstruct the demographic history of the species and measure genome-wide heterozygosity and ROH, the latter of which is less practical with lower-quality or reference-guided genome assemblies. Our assembly strategy combined short-read Illumina data with long-read data from ONT to generate a scaffold N50 of 100 Mb, making this one of the most contiguous wild felid genomes assembled

to date.

Our analyses of ten complete puma genomes revealed the dynamic history of a once widespread species whose population size is now reduced across much of its range. We showed that extant North American pumas are descended from a population that dispersed northward from South America by at least 200 kya, consistent with the age of the oldest puma fossils in North America. Previously, the incomplete fossil record paired with divergence estimates based on rapidly evolving microsatellites and partial mitochondrial genomes led to the hypothesis of a North American origin of the species, followed by a late Pleistocene local extinction in North America and then a recolonization from South America within the last 20,000 years [50, 178]. Our results using complete nuclear and mitochondrial genomes are consistent with previous genetic analyses in that we show that North American pumas represent a subset of puma genetic diversity. However, the nuclear genomic data suggest that the lineage leading to North American pumas diverged from South American pumas $\sim$300-100 kya, considerably older than the 20 kya inferred previously. While we are unable to exclude the possibility of a local late Pleistocene extinction in North America followed by a recolonization from an unsampled lineage elsewhere in South or Central America, we argue that this nuclear genomic data in combination with a recently identified puma fossil in South America that dates to 1.2-0.8 mya [40] supports a simpler demographic hypothesis in which the puma lineage originates in South America, disperses into North America by 300-100 kya and persists there to the present day. We note that new fossils or genomic data from late Pleistocene aged pumas or pumas from other locations in South and Central America will be necessary to test this demographic hypothesis.

If true, the new model for puma demographic history means that pumas would have

been present in North America for at least one complete glacial/interglacial cycle, indicating that pumas were capable of surviving in a broad range of habitats and environments. This hypothesis is supported by data from living pumas, which, despite a preference in North America for mountainous habitats, are also known to occupy grassland habitats in South America, such as Patagonia [267]. Differences in habitat selection between the two continents probably reflect a long history of competition with a diverse carnivore guild on both continents. For example, jaguars are better adapted than pumas to living in habitats that flood periodically [240], and predation by wolves in North America probably precludes pumas living in open habitat without escape terrain [73].

Intriguingly, North American pumas share a common maternal ancestor around the peak cold period of the last ice age, $\sim$20 kya. This period is associated with a reduction of available habitat across the continental United States, as the coalesced Laurentide and Cordilleran glaciers covered much of present-day Canada and the Upper Midwestern United States [70]. Forests would have been reduced significantly at that time, as would available habitat for the smaller prey preferred by pumas, providing a potential mechanism for a reduction in puma population size around that time.

The recent history of pumas is marked by human persecution and encroachment on their habitat, resulting in small and isolated populations that are susceptible to loss of genetic diversity and predisposed to inbreeding. Over many generations, without the input of novel variation from migrants, isolated populations can accumulate local genetic variation while losing overall genetic diversity. Loss of genetic diversity may be a common situation for top predators, as their population densities are usually low and successful migrants are infrequent.

Consequently, even moderate levels of fragmentation will affect their genomic diversity. While pumas in South America currently experience less habitat degradation than pumas in North America, pumas in South America will likely face further habitat loss and fragmentation as rapid human population growth and land development continues on the continent [112]. The result may be small, isolated populations in South America similar to those currently seen in North America. Thus conservation efforts and findings taken from isolated populations in North America may need be applied in the future to other parts of the puma range.

In North America, pumas were hunted extensively, resulting in low population densities in many areas of their range [267]. Hunting was so severe until regulations were put in place during the mid 20^{th} century that pumas likely experienced a population bottleneck. All North American pumas sampled in this study exhibit short ROH that date to approximately the early 20^{th} century, indicative of small effective population sizes during the time when hunting was severe.

In many areas of North America, including California and Florida, large-scale hunting was followed by shrinking habitat availability, resulting in small, isolated populations [228, 291, 270]. Our sampling focused on populations in North America that are known to be isolated and, as such, our results highlight the genomic consequences of this isolation- reduced diversity and signatures of close inbreeding. Pumas in the isolated populations of Big Cypress (CYP), Santa Monica Mountains (SMM), and Santa Cruz Mountains (SC) all have many ROH of all length categories, indicating ongoing inbreeding as a result of continued small population sizes. In contrast, the Yellowstone individual had a similar number of short ROH to these more isolated populations, but fewer long ROH. This pattern is consistent with the known history of hunting

and habitat availability in the Yellowstone area: Pumas in the Yellowstone area were hunted to low densities into the mid 20th century [267], but today Yellowstone National Park is a large protected area surrounded by wildlands. This connectivity between the Park and wildlands facilitated the recovery and maintenance of genetic diversity in the local puma population once hunting pressures were reduced.

Florida panthers are among the most well-studied populations of pumas, especially with regard to the phenotypic manifestations of isolation and inbreeding. The 1995 introduction of pumas from Texas, the most geographically proximate population to the Florida panthers, is widely regarded as a successful genetic rescue via translocation. However, Florida panther genetic diversity in the Everglades population had been bolstered several decades earlier, when seven individuals were released into Everglades National Park from a captive breeding facility where pumas from Central America had been included in the breeding population. One Florida panther that we sequenced, EVG21, is admixed, having both Floridian and Central American ancestry. Her genome is a combination of regions with comparatively high heterozygosity, similar to that observed in the Brazilian pumas, and long ROH, similar to the highly inbred Florida panthers. The distribution of the lengths of ROH suggest that her maternal and paternal lineages shared a common ancestor that lived shortly after the release of the admixed pumas into the Everglades population. This suggests that the genomic consequences of inbreeding happen quickly, with much of the gains from the genetic rescue being quickly erased. EVG21's genome provides evidence that when the population is small, it is likely that an individual's parental lineages will share a very recent common ancestor, even after genetic rescue through admixture (Supplementary Figure 13). Thus, a consistent effort is required to maintain the

benefits of translocation.

In many areas of the current puma range, human land use has reduced the connectivity that is critical to recovery and maintenance of healthy populations. Despite these barriers, gene flow among neighboring populations can be facilitated by enhancing landscape connectivity through coordinated land use planning and by adding bridges or underpasses across freeways [101]. Although pumas are capable of traveling long distances, large roads are a major barrier to their movement [285, 20]. A model of population dynamics in the Santa Monica Mountains that incorporated landscape connectivity and its effects on genetic diversity predicted a high probability of extinction (99.7%) within 50 years after survival rates first began to decrease due to inbreeding, unless connectivity was increased [22]. Our genomic analyses of the samples from the Santa Monica Mountains also support the effectiveness of population connectivity. The two pumas sequenced from the region (SMM12 and SMM22) both currently reside in the small subpopulation south of US 101 freeway. However SMM12 migrated into the subpopulation from north of US 101 [228], a larger area that shows greater connectivity to surrounding regions. Migrations between these two areas are now rare, but the two subpopulations were probably part of a larger panmictic population prior to the existence of US 101. The genomic analysis of ROH highlighted that SMM22 had an increased number of large ROH relative to SMM12, consistent with SMM22 originating in a population that is smaller and more isolated. The examination of IBD ROH between SMM12 and SMM22 showed that only 4% of their genomes are in ROH that are IBD between the two individuals. In contrast, individuals that originated from the same population have a much larger proportion of their genomes in IBD ROH (e.g. 12% for SC29 and SC36). This indicates that while inbreeding has reduced diversity in a considerable proportion

of the genomes of individuals within small populations, these low diversity regions are generally not shared between populations. Thus, reconnecting the populations on either side of US 101, as currently proposed via a wildlife crossing over the freeway [229], would help restore the lost genetic diversity.

Genome-scale data sets have the potential to inform conservation planning. Our results highlight how whole genome data can provide new insights when compared to traditional conservation genetic techniques. For instance, measures of average heterozygosity are the most commonly used metrics to characterize the genetic health of a species, as estimates are relatively simple to generate and are easily comparable among organisms. However, average heterozygosity provides only a narrow insight into the health and genetic potential of a species [231]. While in some species average genome-wide heterozygosity is highly correlated with the level of inbreeding estimated using ROH [133], in systems with admixture, average heterozygosity estimates can be deceptive, as demonstrated with our admixed Everglades puma (EVG21). The heterozygosity of EVG21 is almost as high as the Brazilian pumas, but EVG21 has a large portion of her genome in ROH. We would infer two very different genetic conditions when considering each metric separately, and thus both heterozygosity and proportion of the genome in ROH should be considered in assessing genomic health. Finally, knowledge of shared ROH, an analysis which can only be done with very high density markers across the genome, is critical when designing mitigation plans, as this analysis predicts whether enhancing connectivity would restore lost genetic diversity and helps identify potential candidates for translocation. In this context, this study can serve as a template for future conservation genomic research targeting species living in small, isolated populations.

1.5 Methods

1.5.1 SC36 Sequencing Data

1.5.1.1 Sampling

We captured and drew blood from a wild, male puma (SC36) who lived in the Santa Cruz Mountains in California, USA in accordance with guidelines and regulations of local governing bodies. Capturing, handling, and sampling protocols for SC36 were captured by methods detailed previously [299] and approved by the Institutional Animal Care and Use Committee (IACUC) at the University of California, Santa Cruz (protocol #Wilmc1101) and granted by the California Department of Fish and Game.

1.5.1.2 Illumina libraries

We performed two DNA extractions using the Qiagen DNeasy Blood & Tissue kit following the manufacturer's protocol for non —nucleated erythrocytes using 100 µL of blood. We made four indexed Illumina libraries from these two extractions, following the Meyer Kircher protocol [186], targeting insert sizes of 350bp and 550bp. We created a third DNA extraction for SC36 using the Qiagen Blood and Cell Culture Mini Kit using 500 µL of blood, following the manufacturer's protocol up until the spooling step. The DNA was not concentrated enough to precipitate, so instead we centrifuged the sample, washed it with cold 70% EtOH, and further centrifuged it. We removed the supernatant and air dried the pellet. We resuspended the pellet in 50 µL of TE buffer at 55°C for 2 hours. We prepared four indexed Illumina libraries from this extraction following the Meyer Kircher protocol [186], targeting an insert size of 330bp.

We sent out the eight paired-end shotgun libraries for sequencing at UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory on an Illumina HiSeq 2500 (2x150bp, 2x100bp), at UC Santa Cruz Ancient and Degraded Processing Center on an Illumina MiSeq (v3 chemistry, 2x300bp), and at UC San Diego Institute for Genomic Medicine Genomics Center on an Illumina HiSeq 2500 (2x100bp).

1.5.1.3 CHiCago library

We extracted additional DNA from the SC36 blood samples using the Qiagen Blood and Cell Midi Kit (catalog no. 13343). We quantified DNA using a Qubit fluorometer. We assembled 1µg of in vitro chromatin using the Active Motif In Vitro Chromatin Assembly kit, according to the manufacturer's instructions. We crosslinked the chromatin in 1% formaldehyde for 15 minutes at room temperature, and quenched with 2.5M Glycine. We immobilized chromatin on SPRI beads at a SPRI-lysate ratio of 2:1 [62], and washed with 10mM Tris and 50mM NaCl to remove non-histone-associated DNA. We resuspended the bead and chromatin mixture in 49.5µL 1X NEBuffer 2, and digested with 5 units of DpnII enzyme for one hour at 37°C in a thermal-mixer. After digesting, we concentrated the beads and washed twice with the wash buffer. We resuspended the sample in a 50µL reaction containing dA-dT-and dGTP, biotinylated dCTP, and Klenow. We performed end-labeling at 25°C for 30 minutes, after which we washed the sample twice with the wash buffer. The sample was ligated overnight in a 250 µL reaction containing 1X NEB T4 ligase buffer, 0.1mg/ml BSA, 0.25% Triton X-100, and 50 units T4 DNA ligase. After ligation, we added 2.5µL 10mM dNTPs and 5.5 units T4 DNA polymerase to remove unligated biotin-dCTP. After concentrating the sample and removing the ligation

buffer, the crosslinks were reversed and the sample was deproteinated in 50 µL cross-link reversal buffer (50mM Tris pH=8.0, 1% SDS, 0.25mM CaCl2, and 0.5 mg per mL Proteinase K). We incubated the sample at 55°C to digest the histones, then increased the temperature to 68°C to reverse crosslinks. We then separated the sample from the beads and purified on fresh SPRI beads at a ratio of 2:1 [220]. DNA recovery was quantified with a Qubit fluorometer. We prepared 400ng of sample for sequencing using the NEB Ultra library preparation kit according to the manufacturer's instructions, with one exception: prior to the indexing PCR, the sample was enriched by pulldown on 30µL Invitrogen C1 Streptavidin beads, then washed to remove non-biotinylated DNA fragments. The washes were as follows: 1 wash with lithium wash buffer (1M LiCl, 0.5mM EDTA, 0.05% Tween-20), followed by 2 washes with sodium wash buffer (1M NaCl, 0.5mM EDTA, 0.05% Tween-20), and 2 washes with TE-Tween. We sent the library for sequencing at the UC San Diego Institute for Genomic Medicine Genomics Center on an Illumina HiSeq 2500 [23] (2x125bp reads).

1.5.1.4 Hi-C library

We generated a Hi-C library for SC36 by centrifuging 1mL of blood stored in EDTA at 2500 rcf for 5 minutes to pellet and washing with 1mL PBST, then resuspending in 100ul PBST. We crosslinked the cells with 1% final concentration formaldehyde for 15 minutes at room temperature and quenched the reaction with 2.8 µL 2.5M Glycine, and pelleted to exchange the buffer. We incubated the cells in 500µL lysis buffer (10mM HEPES pH-8.0, 10mM NaCl, 0.2% IGEPAL CA-630, and 1X Protease inhibitors solution (Roche)), and then lysed them further with the addition of 0.5mm garnet beads (MoBio PowerMax beads) by adding 100

μL of beads and vortexing the tube on its side for 1 minute at max speed. We decanted the lysate, and then purified the chromatin by centrifuging at 2500 rcf for 5 minutes, followed by 3 washes with 200μL wash buffer (10mM Tris ph=7.5, 50mM NaCl, 0.1% Tween-20) before resuspension in 55 μL wash buffer with 1μL 2% SDS, followed by incubation at 37C for 20 minutes [165]. We made two technical replicates with 1μg of chromatin per replicate, as determined by Qubit fluorometry. We immobilized chromatin on SPRI beads at a SPRI-lysate ration of 2:117, then washed with 10mM Tris, 50mM NaCl to remove non-histone-associated DNA. We resuspended the bead/chromatin mixture in 49.5μL 1X NEBuffer 2, and digested with 5 units of DpnII enzyme for one hour at 37°C in a thermal-mixer. After digesting, we concentrated the beads and washed them twice with the wash buffer. We resuspended the sample in a 50μL reaction containing dA-dT-and dGTP, biotinylated dCTP, and Klenow. We performed end-labeling at 25°C for 30 minutes, then washed the sample twice with the wash buffer. We ligated the sample overnight in a 200μL reaction containing 1X NEB T4 ligase buffer, 0.1mg per ml BSA, 0.25% Triton X-100, 1mM DTT, and 50 units T4 DNA ligase. After concentrating the sample and removing the ligation buffer, we reversed the crosslinks and deproteinated the sample in 50μL cross-link reversal buffer (50mM Tris pH=8.0, 1% SDS, 0.25mM CaCl2, and 0.5 mg per mL Proteinase K). We incubated the sample at 55°C to digest the histones, then increased the temperature to 68 °C to reverse the crosslinks. After cross-link reversal, we separated the sample from the beads and purified on fresh SPRI beads at a ratio of 2:1 [220]. DNA recovery was quantified by Qubit fluorometer. We prepared the sample for sequencing using the NEB Ultra library preparation kit according to the manufacturer's instructions, with one exception: prior to indexing PCR, the sample was enriched by pulldown on 30μL Invitrogen C1

Streptavidin beads, then washed to remove non-biotinylated DNA fragments as follows: 1 wash with lithium wash buffer (1M LiCl, 0.5mM EDTA, 0.05% Tween-20), followed by 2 washes with sodium wash buffer (1M NaCl, 0.5mM EDTA, 0.05% Tween-20), followed by 2 washes with TE-Tween. We sent the library for sequencing at UC San Diego Institute for Genomic Medicine Genomics Center on a HiSeq 4000 [23] (2x75bp).

1.5.1.5 Oxford Nanopore libraries

We extracted genomic DNA from a whole blood sample from SC36 with the Qiagen Blood and Cell Culture Mini Kit, using 750 µL of starting material. We followed the protocol exactly, up until the spooling step. The DNA was not concentrated enough to precipitate, so we instead centrifuged the sample at 5000g for 15 minutes, washed it with cold 70% EtOH, and centrifuged it again at 5000g for 10 minutes. We removed the supernatant and air-dried the pellet. We resuspended the pellet in 50 µL of TE buffer on a shaker at 22°C overnight. We quantified DNA with a Qubit 2.0 dsDNA HS kit at 71.5 ng per µL (3.58 µg total). We verified the DNA size distribution with a pulse-field gel using a 0.75% agarose TAE gel, run at 75V for 16 hours, with the preset 5-150 kb program on the Pippin Pulse power Supply (version 1.3.2), and estimated the DNA fragments to range in size from 20-25 kb in length.

We performed both a Rapid and a 1D2 Sequencing run (Oxford Nanopore, SQK-RAD002 and SQK-LSK308). For the Rapid library preparation, we used an input of 200 ng of DNA, as recommended by the manufacturer. We first treated the high molecular weight DNA with 2.5 µL of fragment repair mix and incubated for 1 minute at 30°C, followed by 1 minute at 75°C. We ligated the repaired DNA with Rapid Adapters using Blunt/TA Ligase (NEB). We

quantified the Rapid libraries using a Qubit prior to sequencing. The Rapid sequencing libraries were run on an R9.5 flow cell using the NC_48hr_Sequencing_FLO-MIN107_SQK_RAD002 protocol. For the 1D2 libraries, we used the recommended 1 µg of high molecular weight DNA. We end-repaired and A-tailed the DNA using NEBNext UltraII End-Repair/dA-tailing mix (NEB). We cleaned up the end-repair product using AMPure XP beads and then ligated 1D2 adapters using Blunt/TA Ligase (NEB). We performed a subsequent ligation to add the Barcoded Adapter mix (BAM) using Blunt/TA Ligase (NEB). We quantified the 1D2 libraries using a Qubit prior to sequencing. The 1D2 libraries were run on an R9.5 flow cell using the NC_48hr_Sequencing_FLO-MIN107_SQK_LSK308 protocol. We base called all Fast5 raw reads generated from both sequencing runs using the latest Albacore software (Oxford Nanopore proprietary) (version 2.0.2).

1.5.2 Nuclear genome assembly

We removed adapters from the eight Illumina libraries for SC36 with SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15) and reducing the minimum length of trimmed reads to 25 bp (-L 25) [262]. We used Trimmomatic [27] (version 0.33) to 1) remove additional Meyer Kircher IS3 adapters [186] using a seed mismatch of 2 and a simple clip threshold reduced to 5 for the shorter adapter sequence, 2) quality end trim using minimum qualities of 2 or 5 for leading and trailing ends of reads, respectively, 3) window quality trim, using a window size of 4 and a minimum quality of 15, and 4) remove reads shorter than 50 bp. We determined the expected genome assembly size by creating a histogram of kmer counts for k=45 with Jellyfish [173](version 2.2.10), that we visualized with GenomeScope

[287] to obtain a genome length estimate of 2.257Gb, that we used for calculating NG50 (Table 1.1). We used this processed shotgun data to assemble a de novo genome using the Meraculous-2D Genome Assembler [39](version 2.2.4), with diploid mode set to 1 and a kmer size of 45. The contig N50 of the shotgun assembly was 17.6 kb and the scaffold N50 was 36.6 kb. We had approximately 47X coverage of the genome from the eight Illumina shotgun libraries.

We scaffolded the Meraculous assembled genome using HiRise [220] (version 2.1.1) run in serial mode using the default parameters with the Chicago and Hi-C libraries as input. The resulting scaffold N50 was 103.8 Mb. We further visualized the HiC assembly using HiGlass [136] after filtering for duplicates and valid Hi-C reads with pairtools [3] (Figure 1.5).

Given that the puma used for the SC36 shotgun assembly was a male, we had difficulty in assembling the X and Y chromosomes with Meraculous run in diploid mode 1. We identified potential X and Y scaffolds in the SC36 HiRise assembly by using Exonerate [254] (version 2.2.0) to align to known domestic cat X and Y chromosome genes [209]. We identified two scaffolds (scaffolds 2173 and 1964) which had numerous X chromosome mapping genes, and validated these scaffolds as being X chromosome scaffolds based on coverage of the Illumina shotgun data for SC36. We removed both scaffolds from the HiRise assembly. We identified one scaffold that confidently mapped exclusively to a Y chromosome gene, but were unable to validate it based on coverage, so we did not remove it from our HiRise assembly.

To generate an assembly for the X chromosome, we used shotgun data generated from a female puma from the Santa Monica Mountains, SMM13 (Section Additional Puma Sequencing Data), using Meraculous [39] (version 2.2.4) in diploid mode 1 and a kmer size of 45. We determined sex chromosome scaffolds in the SMM13 genome assembly using Exonerate [254]

(version 2.2.0) by identifying scaffolds that align to known domestic cat X genes [209]. We identified three large scaffolds in this way, and validated them based on coverage for both SC36 and SMM13 using bedtools [222] (version 2.25.0) (Figure 1.6). For SMM13, we found that these scaffolds had approximately the same coverage as the average genome-wide coverage. For SC36, we found these scaffolds had roughly half coverage of the average genome coverage. We added these three X chromosome scaffolds (scaffolds X1, X2, and X3) to the HiRise assembly for SC36.

We performed gap filling on the HiRise scaffolded assembly using Oxford Nanopore Technologies (ONT) long reads with the tool PBJelly, part of the PBSuite [77] (version 15.8.24). PBJelly also resolved the sizes of the gaps introduced by HiRise. We used Porechop [234] (version 0.2.3) to adapter trim the ONT reads. The resulting reads provided 1.2 fold coverage of the genome. We set PBJelly to correct only intrascaffold gaps, and used the default minimum of 1 read spanning a gap. We reduced the number of gaps (represented in the assembly as a series of Ns) in the genome from 258,836 to 207,433 and reduced the total number of Ns in the genome from 154,284,192 to 132,359,239. Although our coverage from ONT was only 1.2X, the addition of these data to fill gaps between Hi-C linked scaffolds recovered over 140Mbp of sequence.

We used the genome improvement tool Pilon [289] (version 1.22) to correct sequence errors in the gaps that were filled with the high-error ONT data. We first aligned the Illumina shotgun data back to the PBJelly genome using bwa mem [159] (version 0.7.7) and marked duplicates with Picard toolkit MarkDuplicates [4] (version 1.114). We then ran Pilon with the alignment file as the " - - frags " input, the PBJelly genome as the genome file, specified the

genome as diploid, and used the default setting to fix all types of changes. We ran two iterations of alignment and consensus sequence calling.

We assessed this final version of the genome (PumCon1.0) by alignment to the domestic cat genome (GCA_000181335.4) using SyMap2 [257] (version 4.2) using the BLAT default alignment and synteny parameters, but with Min Dots set to 10, and Top N set to 1 (Figure 1.7). We also used the genome assessment tool BUSCO [252] (version 2.0.1) to evaluate genome completeness based on a set of conserved single-copy orthologous genes (human gene set; n=4104). In the PumCon1.0 genome, 93.0% of these genes are complete and present in a single copy only (Table 1.2). The number of Illumina shotgun reads mapping to each version of the genome was calculated using the command "samtools view -q 30 -c" (Table 1.1).

The final genome assembly was 2,432,985,507 bp in length and had an N50 of 100.53 Mb, with 178,994 gaps and 114,069,924 Ns. Ninety percent of the PumCon1.0 assembly is represented by 28 scaffolds, two of which are X related (scaffold_X1 and scaffold_X2). Thus 87.6% of the genome is represented on 26 autosomal scaffolds, each larger than 20 Mb.

1.5.3 Genome annotation

1.5.3.1 RNA-Seq library

We used a Trizol RNA extraction to obtain total RNA from whole blood collected from a wild puma (SC85) from the Santa Cruz Mountains. We removed unwanted globin mRNA by performing the GLOBINclear protocol (Thermo Fisher). For cDNA synthesis, we used 20 ng of input RNA, a reverse transcriptase (Clonetech), a SmartSeq2 template-switch oligo [212], and an Oligo-dT primer to enrich for poly-A+ RNA. We performed reverse tran-

scription at 42°C for 1 hour. We treated the cDNA product with 1 μL of 1:10 dilution of RNase A (Thermo Fisher) and Lambda Exonuclease (NEB) and incubated at 37 °C for 30 minutes. We amplified the cDNA for 18 cycles using KAPA Hifi Hotstart 2X readymix (KAPABiosystems) with an ISPCR primer [212]. We tagmented the amplified cDNA at 55°C for 7 minutes in a 20 μL reaction using a Tn5 enzyme to generate the RNA –Seq libraries. The Tn5 enzyme was loaded with custom oligos Tn5ME-A/R and Tn5ME-B/R75. We amplified the tagmented products for 13 cycles using KAPA Hifi (KAPA Biosystems) with Nextera primers. The amplified RNA-Seq library was size-selected targeting 300-800 bp using a 2% agarose E-gel (Thermo Fisher). The RNA-Seq library was then quantified using a Qubit and the size distribution was verified using an Agilent 2100 Bioanalyzer. We sent the RNA-Seq library for sequencing on a HiSeq 4000 at UC San Diego Institute for Genomic Medicine Genomics Center (2x100bp).

1.5.3.2 Annotation

The PumCon1.0 genome was annotated by NCBI according to the NCBI Eukaryotic Genome Annotation Pipeline [85] using our cDNA data and a publicly available dataset generated from a wild puma from Arizona (SAMN02885420, SRX633288).

1.5.4 Additional puma sequencing data

1.5.4.1 Sampling

We sampled 11 pumas for this study. Ten of these, including the individual used for the genome assembly (SC36), were used in a panel for analysis of demographic history and population structure. One was used to assemble the X chromosome. The nine additional pumas

that formed our panel were: an additional puma from the Santa Cruz Mountains (SC29), a puma from Yellowstone National Park (YNP198), two pumas from the Big Cypress National Preserve that were members of the canonical (pre-Texas admixture) Florida panther population (CYP47, CYP51), one puma from population that lived in the Everglades National Park in Florida that was the admixed descendent of a canonical Florida panther and a puma of Central American ancestry that was released into the Everglades decades prior to the Texas panther introduction (EVG21) [197], two pumas from the Santa Monica Mountains in Southern California (SMM12, SMM22), and two pumas from eastern Brazil (BR406, BR338). We also obtained a sample from a female puma from the Santa Monica Mountains for the purpose of assembling an X chromosome (SMM13) (Table 1.3).

The Yellowstone puma was captured using specially-trained hounds, immobilized using a syringe dart (ketamine/medetomidine drug combination), and blood was obtained with aseptic techniques. Approval for Yellowstone sampling was granted by the National Park Service IACUC (#IMR_YELL_Stahler_Cougar_2015.A3). Capturing, handling and sampling of pumas from the Santa Cruz pumas were captured by methods detailed in prior work [299] and approved by the IACUC at UCSC (protocol #Wilmc1101) and granted by the California Department of Fish and Wildlife. Permission of approval to capture and handle pumas from the Santa Monica Mountains pumas were obtained from the National Park Service IACUC (PWR_SAMO_Riley.Sikich_MtnLion_2017) and the California Department of Fish and Wildlife [228]. Capture and sampling of Florida panthers were made possible by a cooperative agreement between the Florida Fish and Wildlife Conservation Commission and the US Fish and Wildlife Service (TE01553-3 and TC146761-1) [128]. Brazilian samples were collected by cer-

tified veterinarians following federal legislation and animal welfare guidelines. Sample BR406 was collected under federal permit SISBIO 40079-2, and sample BR338 was provided by the Centro de Educaao Profissional (CENAP) Instituto Chico Mendes de Conservao da Biodiversidade (ICMBio) national carnivoran sample repository, having been used in a previous study of puma population genetics [178].

1.5.4.2 Illumina libraries

We aimed for high coverage ($\sim$ 30X), whole genome sequencing data for the ten pumas. We extracted DNA from blood samples using the Qiagen DNeasy Blood & Tissue kit using the same method as used for SC36. We prepared indexed Illumina libraries following the Meyer Kircher protocol [186]. We verified DNA concentrations and 350-bp insert sizes by running the libraries on an Agilent 2200 TapeStation system.

We sent samples for sequencing at the National Genomics Infrastructure of SciLife in Stockholm, Sweden on an Illumina HiSeq XTen, Laboratòrio de Biotecnologia Animal at the Universidade de Såo Paulo in Brazil on an Illumina HiSeq 1500, and UC San Diego Institute for Genomic Medicine Genomics Center on an Illumina HiSeq 4000 [23]. We also downloaded shotgun sequencing data for the African cheetah [65] (SRR2737512-SRR2737518) to use as the outgroup for our analyses. We selected cheetah data with an insert size of 170 bp and obtained a final coverage of 34X when mapped to PumCon1.0.

1.5.5 Variant calling and filtering

Prior to alignment of resequencing reads, we added the mitochondrial genome sequence for SC36 (Section Mitochondrial Genome Assemblies and Phylogeny) as a scaffold (scaffold_Mt) to the final nuclear sequence. Due to the high number of nuclear mitochondrial DNA segments (NUMTs) in felids [171], we sought to decrease erroneous mismappings of true mitochondrial DNA in the Illumina data to the NUMTs.

We removed adapters from all puma resequencing data and cheetah SRA data using SeqPrep2 [262], discarding reads shorter than 25 bp in length. We then aligned reads to the PumCon1.0 genome, including the mitochondrial scaffold, using bwa mem [161] (version 0.7.12). We filtered alignments using samtools [158] (version 1.2.1) to keep alignments with a map quality score greater than or equal to 30, remove secondary alignments, and keep reads where both reads in the read pair mapped. Within each library, we removed duplicate sequences due to PCR amplification using samtools rmdup [158] (version 0.1.18). Realignment around insertions and deletions was performed using GATK Realigner Target Creator and Indel Realignment [180] (version 3.5.0). We called variants in each sample using GATK HaplotypeCaller [180] (version 3.7.0), with a minimum base quality score of 18, and emitting all sites, including invariant positions.

We generated three sets of genotypes: two sets consisted of the ten pumas (LD filtered and nonLD filtered) and another set of the ten pumas plus the cheetah for use as the outgroup. For the puma only data set, we joint genotyped all ten puma samples using GATK GenotypeGVCFs [180] (version 3.7.0), including non-variant sites. For the data set including the African

cheetah, we again used GATK GenotypeGVCFs to perform joint genotyping for all 11 samples, but emitted only variant sites. For all variant files, we masked or removed sites that were not biallelic SNPs. We then filtered biallelic SNPs based on a number of parameters. We determined filtering thresholds by visualizing parameter distributions. We used GATK VariantFiltration and SelectVariants [180] to filter strand odds ratio (SOR $\geq$ 3.0), Fisher strand bias (FS $>$ 60.00), quality by depth (QD $<$ 2.00), mapping quality (MQ $<$ 40.00 and MQRankSum $<$ -10.00), read position (-8.00 $<$ ReadPosRankSum $>$ -8.00), and excess heterozygosity in accordance with Hardy-Weinberg equilibrium to remove potentially paralogous genomic regions (ExcessHet $>$ 10.0). We used GATK VariantFiltration and SelectVariants [180] to remove variants above a cumulative depth threshold (DP $>$ 1500 for the puma only variant file, DP $>$ 1650 for the puma and cheetah variant file). We then used vcftools [55] (version 0.1.12b) to remove singleton variants where only one individual carried one copy of a different allele (- -mac 2), sites where any individual had a depth less than 10 (- -minDP 10.0), and sites where any of the ten pumas did not have a base called (- -max-missing-count 0) We only used autosomal scaffolds for further analyses, removing mitochondrial and X chromosome related scaffolds (scaffold Mt, X1, X2, X3, 869, 1862). Scaffolds 869 and 1862 were removed due to syntenic mappings with the X chromosome of FelisCatus9.0 in SyMap [257]. For the LD filtered variant files, we performed LD filtering using PLINK [38] (version 1.90b4.4) with the command "- -indep-pairwise 100 10 0.2".

The non-LD filtered puma only variant file contained 8,212,535 SNPs. The final LD-filtered puma only variant file contained 166,037 SNPs. The puma and cheetah variant LD-filtered variant file contained 557,741 SNPs. The larger number of variants in the puma and

cheetah file is due to the high numbers of sites where the cheetah carries two of the alternate allele, while all pumas carry the reference allele.

Using the non-LD filtered SNP calls from the puma only data set, we generated a fasta file for each sample using GATK FastaAlternateReferenceMaker [180] (version 3.7.0), with heterozygous positions represented by IUPAC codes. Failed SNP sites (see filters above) and failed individual genotypes (RGQ$<$20, depth$<$10) were masked to Ns using the maskfasta function in bedtools [222].

1.5.6 Mitochondrial genome assemblies and phylogeny

We used Unicycler [298] (version 0.4.4) in hybrid assembly mode to assemble an initial mitochondrial sequence for SC36. To decrease our initial input dataset, we first identified mitochondrial mapping reads. To do this, we mapped adapter-trimmed Illumina and ONT reads from SC36 to a publicly available puma reference mitochondrial sequence (KP202261.1) [157] using bwa mem [159] (version 0.7.12) and bwa mem ont2d, respectively. We took the reads in the Illumina and ONT alignment files that mapped to the reference mitochondrial sequence and converted them into fastq format using bedtools [222] bamToFastq (version 2.25.0). We used both the Illumina and ONT mitochondrial-mapped readsets as input into Unicycler as long reads using the flag "l" and unpaired reads using the flag"-", respectively. The assembly created with Unicycler was 17,065 bp in length, circular, and had a depth of 490X. To validate this assembly, we used the Unicycler output as the reference sequence in an assembly using mapping iterative assembler (mia) [98] (version 1.0) using roughly 20 million randomly selected adapter-trimmed Illumina shotgun reads (mia flags: -i -c -C -F -k 13). The resulting mia assembly was identical

in sequence to the inputted Unicycler assembly, and thus was used as the final mitochondrial sequence for SC36.

We used adapter-trimmed Illumina shotgun data to assemble the mitochondrial genome sequences of the remaining nine pumas with mia, using the SC36 mitochondrial sequence as the reference. The coverages of these mitochondrial assemblies ranged from 35X to 138X. For each of the nine assemblies, we filtered using a consensus threshold of 90% and required 10X coverage per site. Any sites which did not meet these requirements was changed to an 'N', resulting in between 2 and 86 Ns per assembly. We removed two repetitive sequences located in the control region, RS2 (sites 16,511-16,861) and RS3 (sites 279-657), from each of the mitochondrial sequences. These regions are known to have highly variable numbers of repeats [122, 170, 253, 199], and were difficult to accurately resolve using short read data. We annotated the mitochondrial genomes using MITOS [24].

We ran PartitionFinder [150] (version 1.1.1) and jModelTest2 [56] (version 2.1.6) to determine the partitions and best substitution model. Due to the small number of substitutions seen in the pumas, PartitionFinder suggested a single mitochondrial partition. jModelTest2 recommended a GTR+GAMMA substitution model for tree building. We used muscle [71] (version 3.8.31) to align our ten assembled puma mitochondrial genomes, the publicly available puma reference mitochondrial sequence (KP202261.1), and a cheetah mitochondrial sequence (KP202271.1) as the outgroup. We used RAxML [263] (version 8.2.4) to produce a maximum likelihood phylogeny, with a GTR+GAMMA evolutionary model, running one hundred bootstrap replicates. We also ran tree inference including the RS2 and/or RS3 repetitive regions, and saw little change in the bootstrap support values and no change to the topology of the tree.

We estimated divergence times between South and North American pumas using a prior composite estimate of the feline mitochondrial divergence rate of 1.15%bp/Myr [170, 50]. Branch lengths were taken from a phylogeny that did not include RS2 or RS3, and were used to calculate divergence dates using the formula $\sigma = 2\lambda T$, where σ is percent divergence between a pair of sequences, λ is the rate of mitochondrial divergence, and T is time [170].

1.5.7 Demographic History

We used the pairwise sequentially Markovian coalescent (PSMC) model [163] to estimate the historical effective population sizes of puma populations. Of note to our sampled population histories, is that the Brazil pumas used in this analysis originate from eastern South America, which prior work has shown is the centrum of present-day puma diversity [50], and likely represents the most ancestral puma lineage. The input was a realigned alignment file of 26 autosomal scaffolds larger than 20 Mb, excluding sex chromosome associated scaffolds (Section: Variant calling and filtering). We filtered the alignment file for each puma to include sites which had between one third and twice the average coverage for that puma. We used a generation time of 5 years and a mutation rate of 0.5e-8 per bp per generation, based on previous estimates of the feline mutation rate [43]. We performed one hundred replicate bootstraps for each individual per the software instructions (Figure 1.9).

We also ran the PSMC tool on regions of the alignment file which were identified as outbred based on our ROH hidden Markov model (Section Runs of Homozygosity). We masked regions of homozygosity for each of the pumas, and re-ran the PSMC analysis to compare the demographic estimates. We saw no considerable difference between the two sets of results

(Figure 1.10).

We investigated the divergence time between our male pumas by running PSMC modeling of pseudo-diploid sequences using pairs of haploid male X chromosomes to avoid the need to phase the data. In this way, we can approximate the population split time between two populations, as the y-axis in the PSMC plot can be thought of as the inverse of the rate of coalescence at a time along the x-axis [179]. We used an alignment file containing the three largest X scaffolds (scaffold_X1, scaffold_X2, and scaffold_X3), for a total of $\sim$103 Mbp of the X chromosome sequence for analysis. We filtered the input to only include sites which had up to the average coverage of the genome for each puma. We created a pseudohaploid fasta by randomly selecting a single high quality base using samtools (mpileup -Q 40 -q 30) [159] and pu2fa (https://github.com/Paleogenomics/Chrom-Compare). We combined the pseudohaploid fasta file of each male North American puma with that of either of the two male Brazilian pumas, and ran PSMC as above (Figure 1.11). To account for the lower mutation rate on the X than the autosomes, we used a formula [163] to convert the autosomal mutation rate into the X mutation rate, using a male to female mutation ratio of 1.4 from the domestic cat [239]. This gives us a per generation mutation rate for the X chromosome of 0.472e-8 per bp per generation, using a generation time of 5 years.

1.5.8 Population structure

We used SmartPCA from the EIGENSOFT [208] (version 6.1.4) package to run principal component analysis on the LD-filtered variant file for the ten pumas, which consisted of 166,037 SNPs. We converted the variant file into eigenstrat format using PGDSpider2 [168]

(version 2.1.0.0) for input into SmartPCA.

We constructed a tree to show population splits using Treemix [213], both with and without the admixed sample EVG21. We used the LD-filtered variant file with the ten pumas and the cheetah as input, which consisted of 557,74 SNPs, and ran Treemix by grouping 5,000 SNPs into one window (k). For the dataset including EVG21, the tree with the highest log likelihood predicted one migration, and explained 99.91% of the variation. In the dataset without EVG21, the best model predicted no migrations and explained 99.91% of the variation.

We used the software STRUCTURE [81] (version 2.3.4) to infer the population structure of the pumas. We converted the LD filtered variant file with 10 pumas and 166,037 SNPs into the input format using PGDSpider2 [168] (version 2.1.0.0). We ran 10 replicates of STRUCTURE for values from one to ten for the number of populations (K), using an admixture model with a degree of admixture (alpha) value of 1 for each K, 10,000 burns ins, and running 10,000 MCMC repetitions. We used the software CLUster Matching and Permutation Program (CLUMPP) [126] (version 1.1.2) under the greedy algorithm to align the ten replicates for each K into one representative mean output matrix for plotting (Figure 1.15). We identified the best K based on mean likelihood L(K) and delta K values [79]. For our dataset, K=3 had the largest delta K value, and showed a break in slope for mean L(K) as well (Figure 1.14). Although our analyses indicate that K=3 best explains our panel, due to our sampling scheme, we are unable to determine the true number of clusters for the species overall. We do note that all population level analyses performed indicate that our samples are best described by three clusters.

1.5.9 Genome-wide heterozygosity estimates

We calculated the average coverage of each puma using samtools depth on the re-aligned alignment file, removing scaffolds for the mitochondria and X chromosome (scaffolds Mt, X1, X2, X3, 869, 1862). We calculated genome-wide heterozygosity by generating a pileup file from map quality and base quality filtered alignment files using samtools [158] (version 1.3.1) (mpileup -q 30 -Q 30) at all sites in the genome with exactly the average coverage depth for that individual. We created a histogram by binning the sites based on the number of reads representing the reference allele. We visually classified which bins were designated as homozygous or heterozygous. We calculated the genome-wide heterozygosity by summing all heterozygous bins and dividing by the total number of genome-wide sites used in the analysis.

Additionally, we examined sliding window heterozygosity using custom script that counts heterozygous positions in the IUPAC coded fasta files. We estimated average heterozygosity in 100-kb windows and used the outputted counts of the number of heterozygous positions and the number of positions with a genotype call to obtain another estimate of average genome-wide heterozygosity (Table 1.4). The slightly higher values observed for the IUPAC estimates could be due to the pileup method being a more conservative method for calling a site heterozygous than GATK FastaAlternateReferenceMaker, which was used with the IUPAC method. Lastly, we calculated genome heterozygosity in a non-reference based method so as to prevent any effects of genome quality on the estimations (Table 1.4). We used adapter-trimmed shotgun data for each of the ten pumas to obtain kmer frequency distributions for k=21 for each puma using Jellyfish [173] (version 2.2.10) and visualized the results using GenomeScope

[287].

1.5.10 Runs of homozygosity

We used a hidden Markov model (HMM) to identify ROH by identifying transitions between inbred and outbred regions of the genome [48]. We first estimated HMM model parameters from the data (Table 1.5). For each of the eight male pumas, we used the sample's heterozygosity estimate from the two large X scaffolds as an estimate of the genotyping error rate. For the two female pumas, we used the estimate from another sample that was geographically close by and had similar sequencing coverage. For each sample, we then estimated the rate of heterozygosity in outbred regions by visually selecting clearly outbred regions and determining the mean heterozygosity across those regions. We determined a single transition rate parameter (t=1e-50) by running different transition parameters and visually inspecting the results. We used these parameters and the filtered fasta files with IUPAC codes as input into the ROH HMM program. For downstream analyses, we used only the 26 largest autosomal scaffolds and discarded ROH less than 2 Mb. We converted the ROH tract lengths to generations using an estimated average recombination rate from the domestic cat of 1.1 cM per Mb [67] and the equation $g = 100/(2rL)$, where g is the time in generations, r is the recombination rate, and L is the length of the ROH tract in Mb [133].

We also used the sliding window approach in PLINK [219] (version 1.90b4.4) to identify ROH for comparison with our ROH HMM. We used the non-LD filtered puma variant file as input. We relaxed the parameters (–homozyg-window-het 20 –homozyg-window-missing 20 –homozyg-window-threshold 0.02 –homozyg-het 750 –homozyg-kb 500) to prevent sequenc-

ing errors from breaking up homozygous tracts. Even with relaxed parameters, PLINK still tended to break up long tracts (Figure 1.18). Since accurate estimates of tract lengths were key to our inbreeding analysis, we used the ROH called by the our ROH HMM program for further analyses.

The low frequency of short ROH observed in the genome of the admixed Everglades panther (EVG21) relative to the other Florida panthers led us to believe that admixture in previous generations had prevented the formation of short ROH because an individual can not have a shared maternal and paternal ancestor that dates to before the admixture event. To test our hypothesis, we used Ancestry_HMM [47, 183] to classify tracts of ancestry in the IUPAC coded fasta file of EVG21 into three types: pure Central/South American ancestry, pure Floridian ancestry, and mixed Central/South American and Floridian ancestry. We used the two Brazil samples (BR338 and BR406) as proxies for Central/South American ancestry, and the two Big Cypress samples (CYP47 and CYP51) as proxies for Floridian ancestry. Because small sample sizes preclude accurate estimates of LD and because the program is sensitive to sites in strong LD, we pruned all ancestry informative markers within 250 kb of another site. The genome of EVG21 was comprised of 21.98% Central/South American ancestry, 28.24% Floridian ancestry, and 49.58% mixed ancestry based on the HMM (Figure 1.19). Using ROH greater than 2 Mb that we identified with the ROH HMM, we classified each ROH as one of the three ancestry types. We did this by using the intersect function in bedtools [222](version 2.25.0) with the parameters '-e -f 0.90 'such that 90% of any ROH identified was only one ancestry type. The results of this analysis classified all ROH as either pure Florida or pure Central/South American ancestry. We saw no ROH that were classified as being of mixed ancestry. Thus, admixture ef-

fectively prevents the formation of mixed ancestry ROH (Figures 1.19, 1.20). ROH comprised 76% of the Central/South American ancestry regions of the genome, 63% of the Floridian ancestry regions of the genome, and 0% of the mixed ancestry regions of the genome. We tested for significant differences in the number of ROH identified per ancestry type using an exact test of the ratio in a Poisson distribution. We found that significantly more ROH were of Floridian ancestry relative to mixed ancestry (p=7.276e-12), and significantly more ROH were of Central/South American ancestry than of mixed ancestry (p=1.526e-05). Additionally, significantly more ROH were of Floridian ancestry than Central/South American ancestry (p=0.006456). The identification of a greater number of ROH of Floridian ancestry, especially ROH of shorter length (Figure 1.20), is consistent with the longer history of sustained inbreeding in the Floridian population.

We next estimated the proportion of the ROH that are shared between pairs of pumas. First, we used the intersect function in bedtools [222] to find genomic regions where ROH overlap between pairs of samples. Then we used a custom script to generate a hybrid diploid fasta file from the IUPAC coded fasta files for each pair of samples. The script first makes a pseudo-haploidized sequence for each sample by randomly selecting one of the two bases at heterozygous sites. The script then uses a pair of pseudo-haploidized sequences to generate a hybrid diploid fasta file by using IUPAC codes to represent differing bases between the two samples, while shared bases between the two samples are represented by the base itself. We ran the ROH HMM program on the generated fasta file, using the same transition rate parameter as above (t=1e-50) and the outbred heterozygosity rate and genotyping error rate averaged across all ten pumas (h=0.00135, e=0.0000765). These ROH indicate where two individuals

share regions of the genome IBD. To determine if ROH that overlap between two samples were also IBD, we used the intersect function in bedtools [222] (version 2.25.0) to find the intersection between the ROH that overlap between the two samples and the ROH generated from the hybrid diploid fasta file. Finally, from these outputs, for each pair of pumas we calculated the proportion of the genome that occurs in ROH that are IBD 1.6.

1.6 Supplementary figures and tables

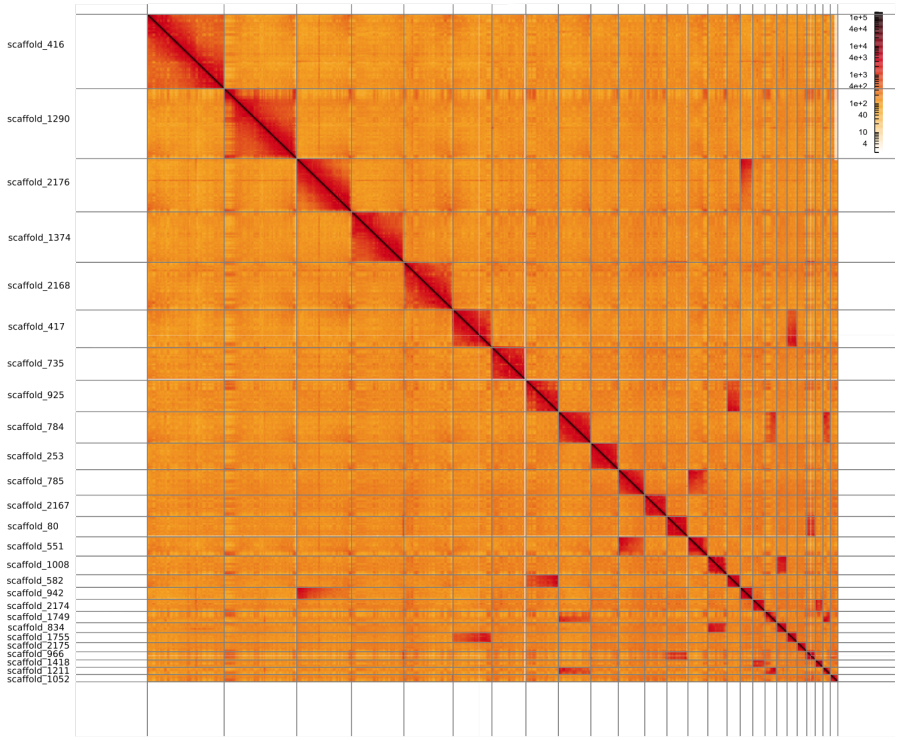

Figure 1.5: Linkage map of the HiRise genome assembly for SC36. The x and y axes mark the mapping positions of the first and second read in a read pair of the Hi-C library, respectively. Mapping results are binned by genomic regions and each bin is colored according to the number of read pairs within the bin. Scaffolds less than 20 Mb in length are not shown.

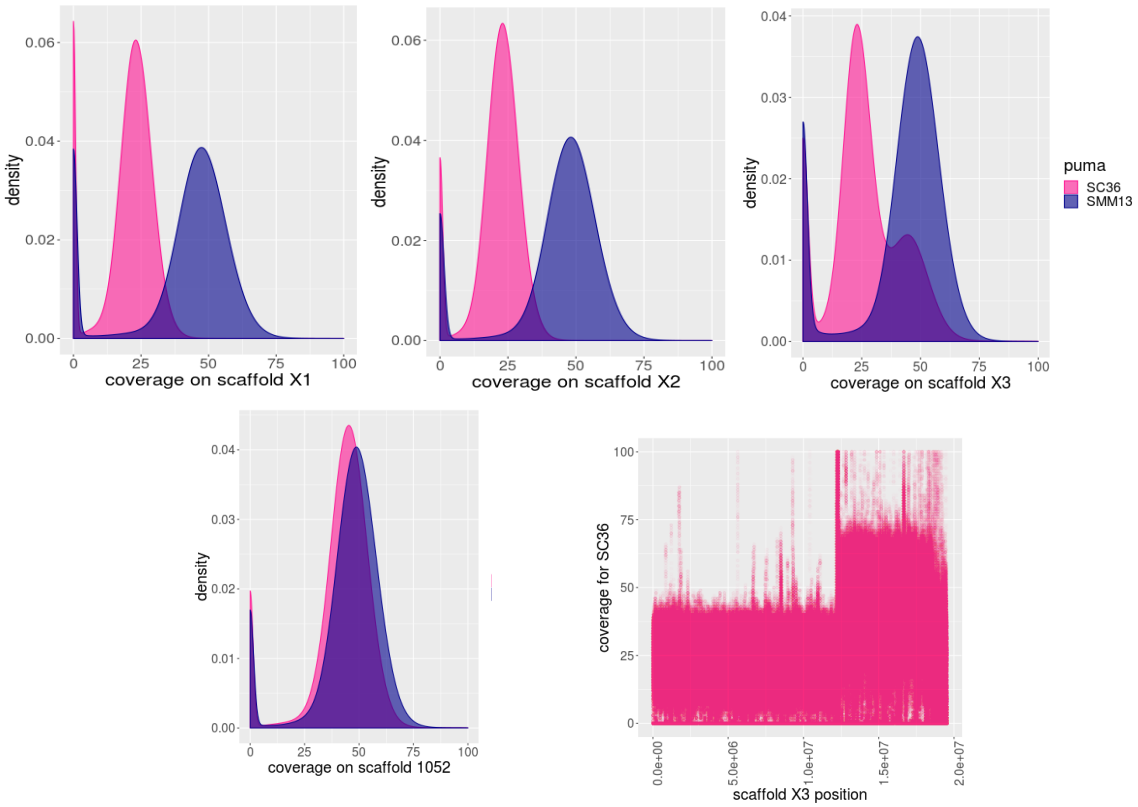

Figure 1.6: Autosomal to X chromosome coverage ratio. Density coverage plots for the three X chromosome scaffolds (top) obtained from the SMM13 genome assembly and an autosomal scaffold (bottom left) from the SC36 genome assembly. Coverage is shown for SC36 (male) and SMM13 (female) mapped to four scaffolds. For all X chromosome scaffolds, SC36 shows half the autosomal scaffold coverage. The trimodal distribution for SC36 on scaffold_X3 is due to the homology of pseudoautosomal region 1 between X and Y chromosomes resulting in twice the coverage for that region of the X in the male. Coverage across scaffold_X3 for SC36 shows a doubling in average coverage at approximately the beginning of the pseudoautosomal region 1 (bottom right).

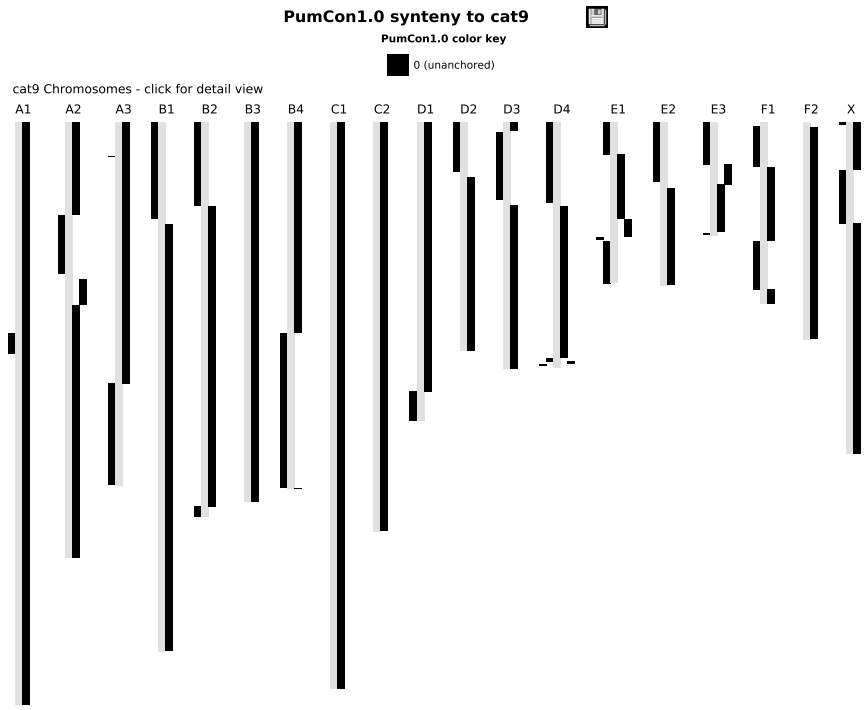

Figure 1.7: Synteny between the puma and domestic cat genomes. Alignment of syntenic regions of the Felis catus genome version 9.0 (domestic cat, gray) with our PumCon1.0 assembly (black) using SyMap2 [257].

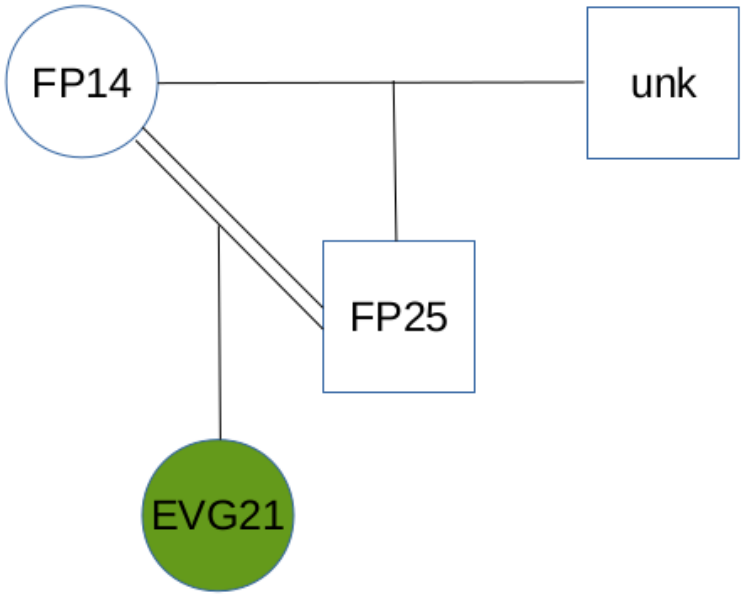

Figure 1.8: Pedigree of EVG21. Pedigree of the inbred and admixed Florida panther sequenced from Everglades National Park [128]. All panthers in the pedigree are of Everglades ancestry. The Central American admixture into this population occurred approximately 6-9 generations prior to these individuals [128]. Note: EVG21 is referred to as FP021 in the original dataset [128].

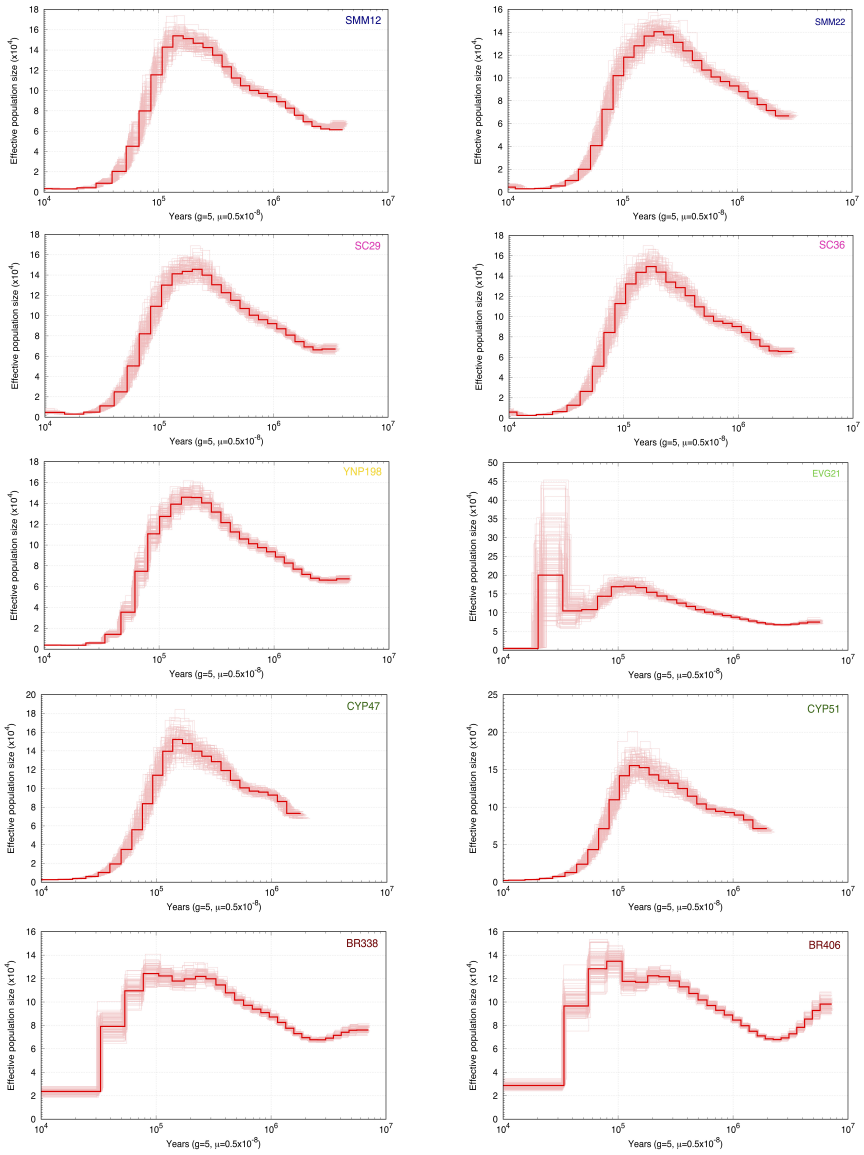

Figure 1.9: Bootstrap replicate PSMC plots for ten pumas. We ran one hundred bootstrap replicates for each of the pumas using the PSMC model [163]. We used a generation time of 5 years, and a mutation rate of 0.5e-8 [43] per bp per generation.

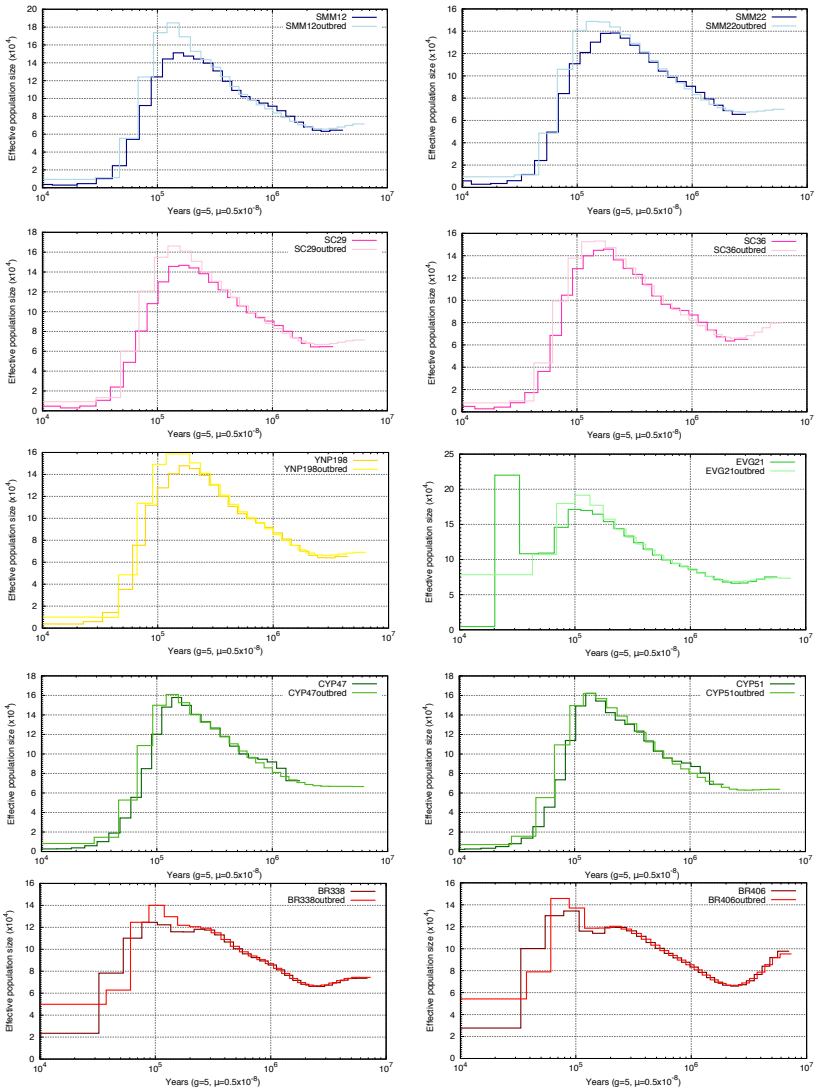

Figure 1.10: PSMC plots for outbred (non-ROH) regions and entire genomes for the ten pumas. We see no substantial differences between the models [163]. We used a generation time of 5 years, and a mutation rate of 0.5e-8 [43] per bp per generation.

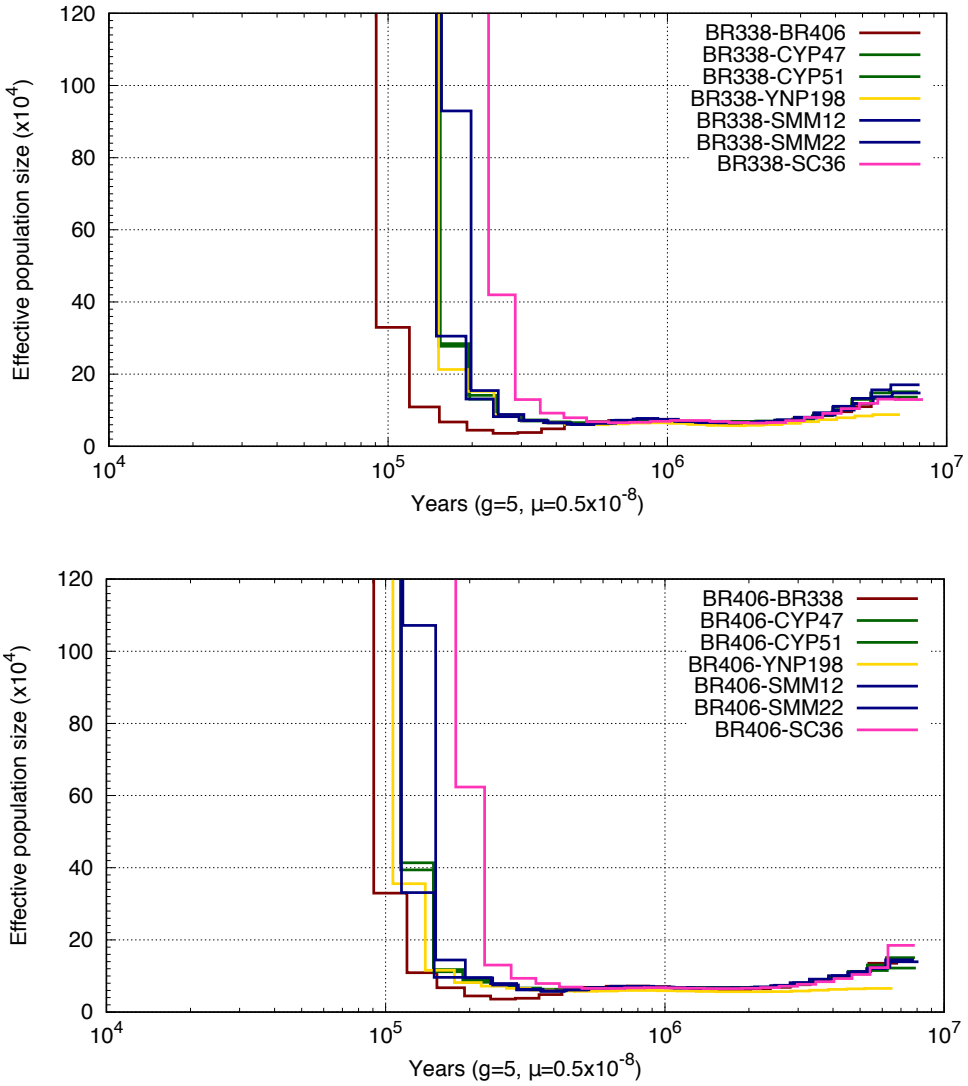

Figure 1.11: Pseudo-diploid PSMC plots for X chromosomes of male pumas. We investigated the divergence time between our pumas by running PSMC modelling on pseudo-diploid sequences from pairs of haploid male X chromosomes. We see a sharp rise in the inferred Ne, and thus an approximate divergence time between 250-100 kya for every North American male when paired with a Brazil puma (either BR338 or BR406). The Brazil-Brazil pseudo-diploid pair had the most recent divergence time, roughly 100-90 kya. Thus the observed rise in effective population size observed in our autosomal PSMC model is not solely the result of population structure in pumas. We used a generation time of 5 years, a male to female mutation ratio of 1.4 [239], and a mutation rate of 0.472e-8 per bp per generation.

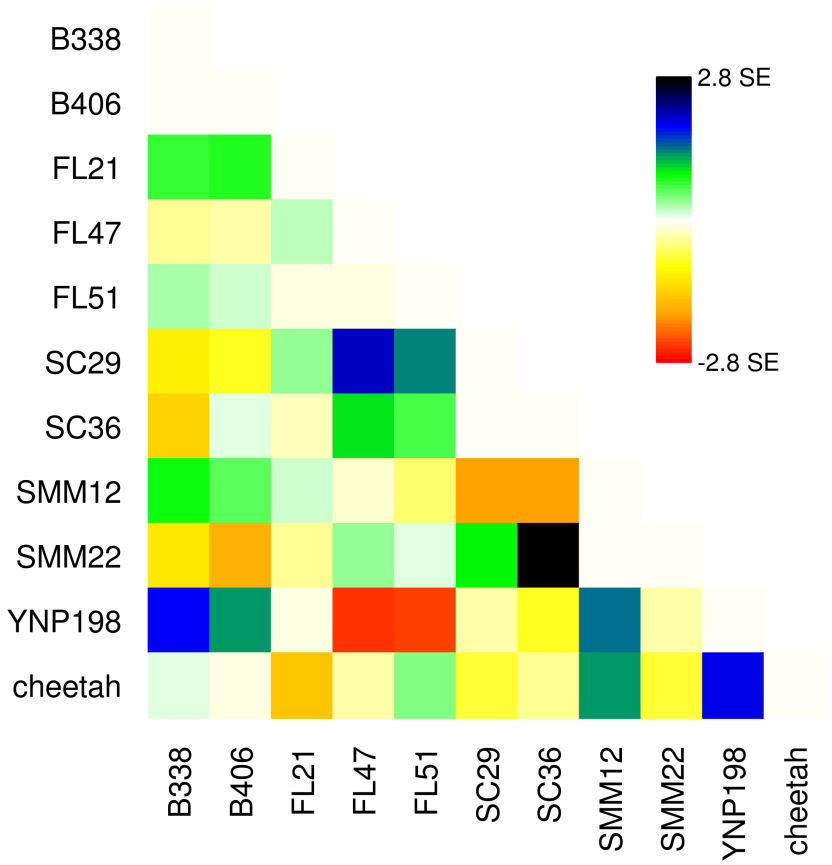

Figure 1.12: TreeMix residual fit of the model. TreeMix [213] run on the LD filtered variant file containing ten pumas and the African cheetah, including one migration, and grouped SNPs together in windows of 5,000 SNPs (k). The model explained 99.91% of the variation.

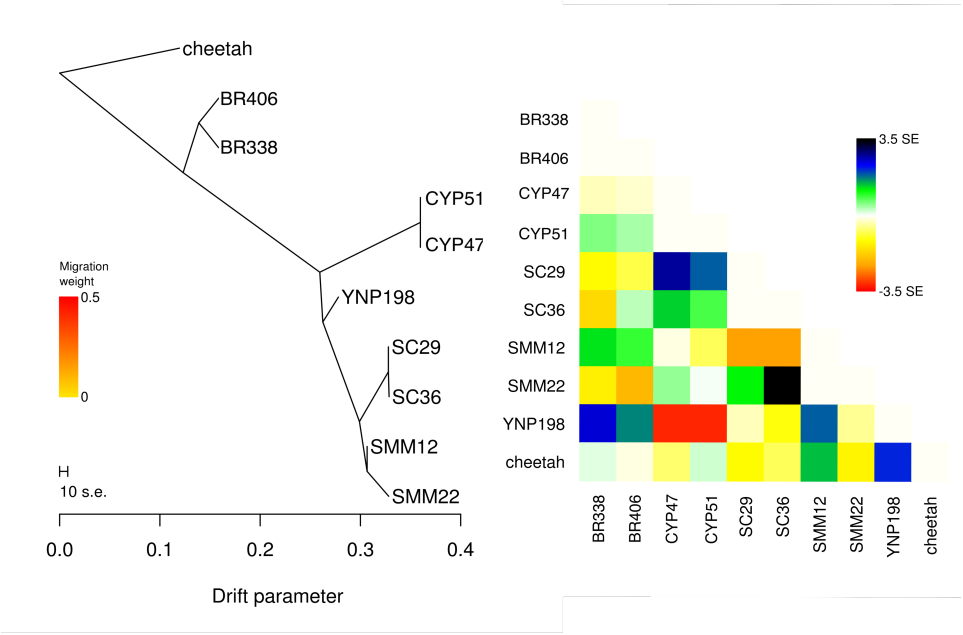

Figure 1.13: TreeMix without EVG21. The result of TreeMix [213] with the highest likelihood when run on the LD filtered puma and cheetah variant file with EVG21 removed, including no migrations, and grouped SNPs together in windows of 5,000 SNPs (k). The model explained 99.91% of the variation.

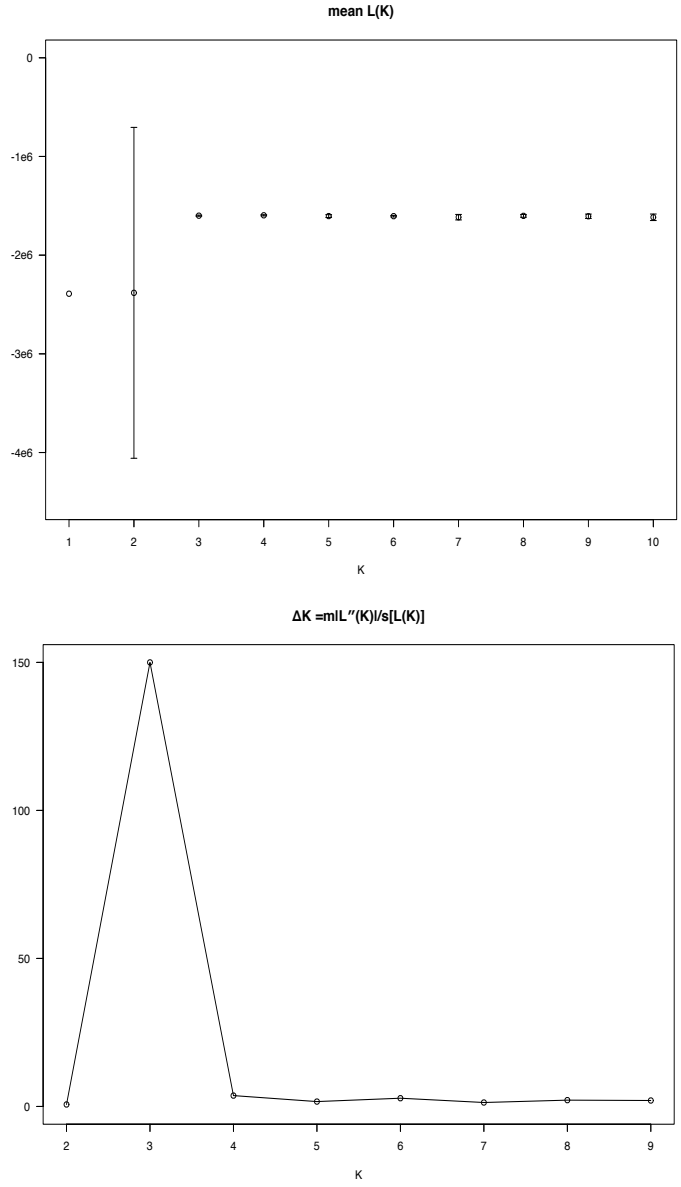

Figure 1.14: Selection of best K in STRUCTURE. We identified K=3 as best for our panel using L(K) (top) and delta K (bottom)[79]. The top panel shows the mean likelihood (L(K)) and standard deviation from the 10 replicates per K value, which plateaued at K=3. The bottom panel shows that the rate of change in the log probability of data between successive K values (delta K) is highest at K=3.

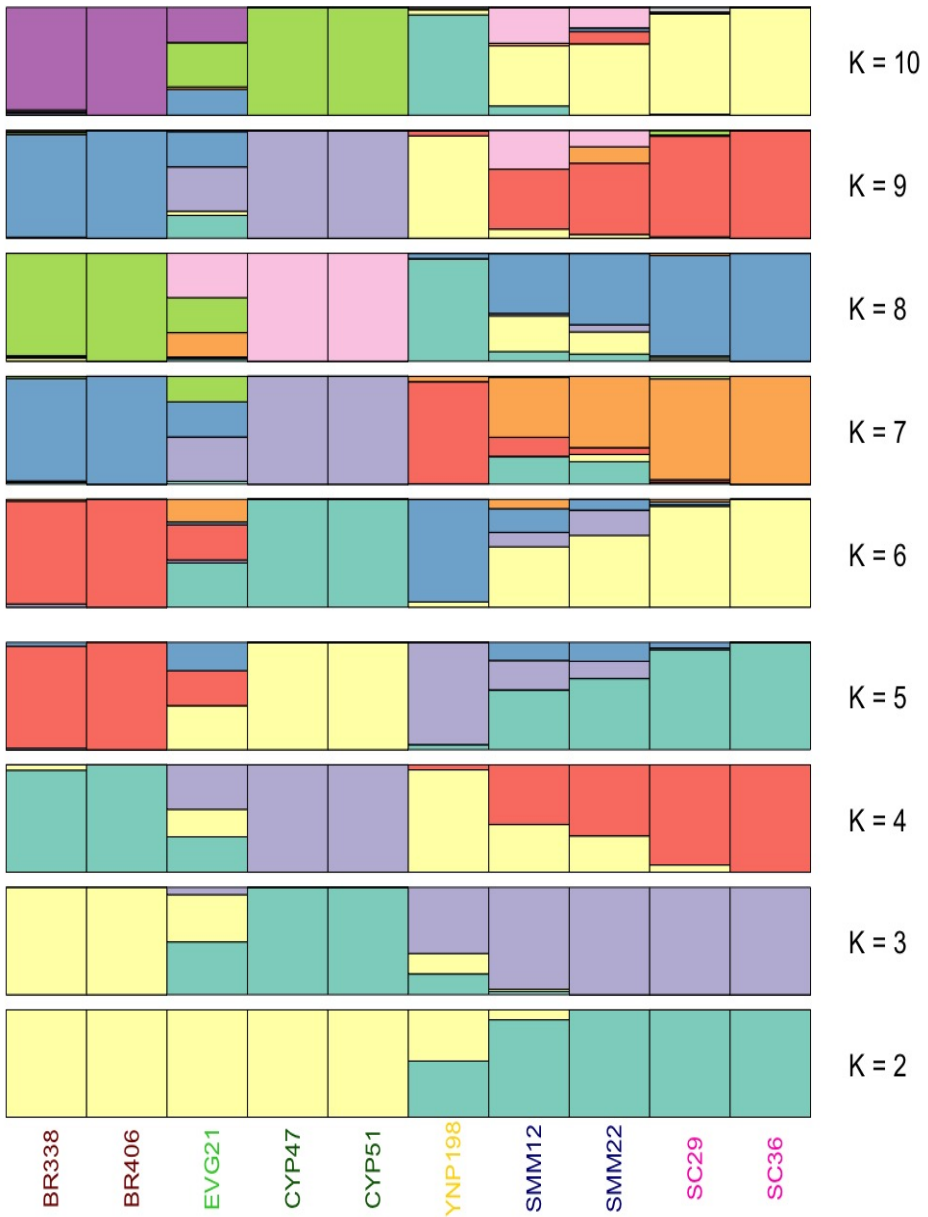

Figure 1.15: STRUCTURE plots for K=2 through K=10. The mean of 10 permuted matrices of STRUCTURE [81] analysis for each of K=2 through K=10, performed using CLUMPP [126]. Both delta K and mean L(K) values indicated that K=3 was the best K for our panel (Figure 1.14) [79].

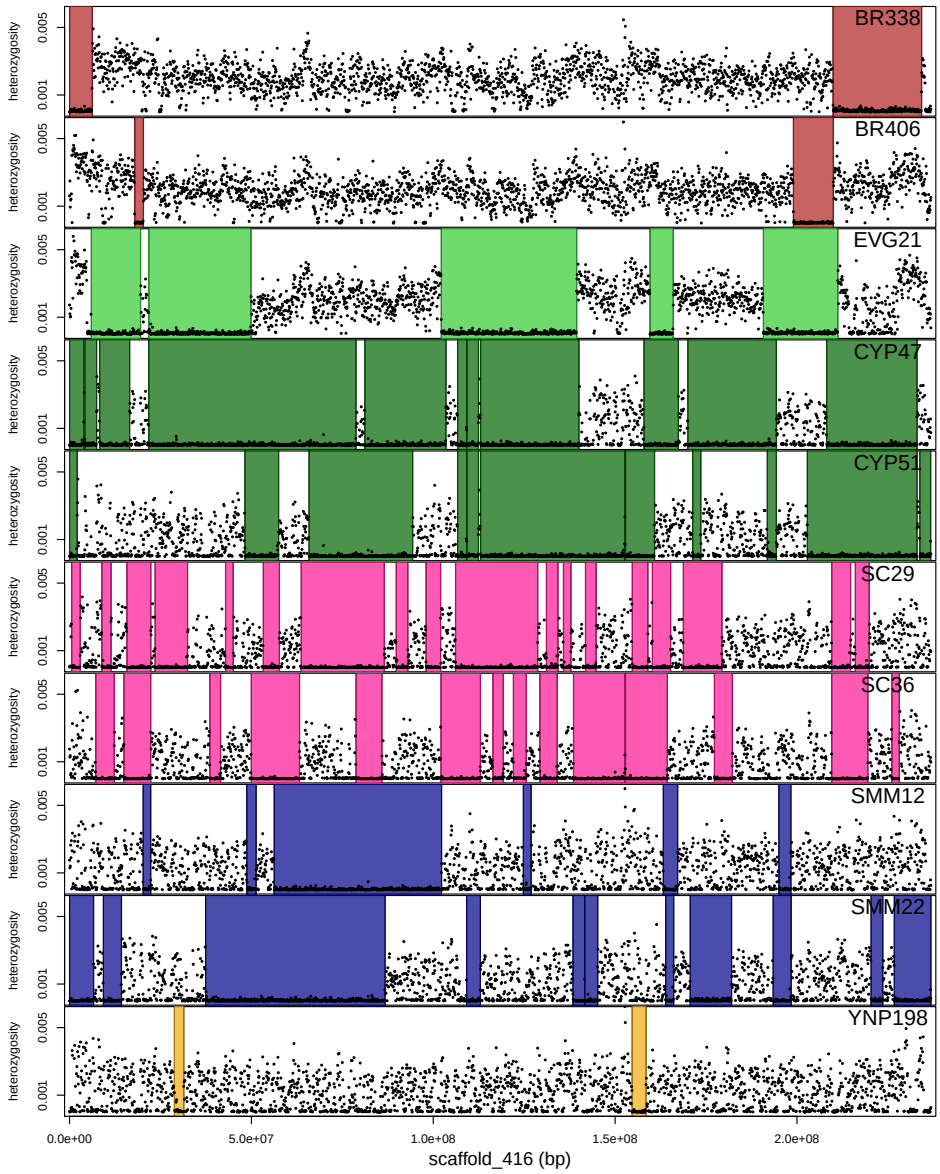

Figure 1.16: ROH for all ten pumas as called by our ROH HMM. Black dots represent average heterozygosity in 100 kb windows; colored regions represent blocks called as ROH. Scaffold_416 represents the largest scaffold in the genome assembly at 236.8 Mb.

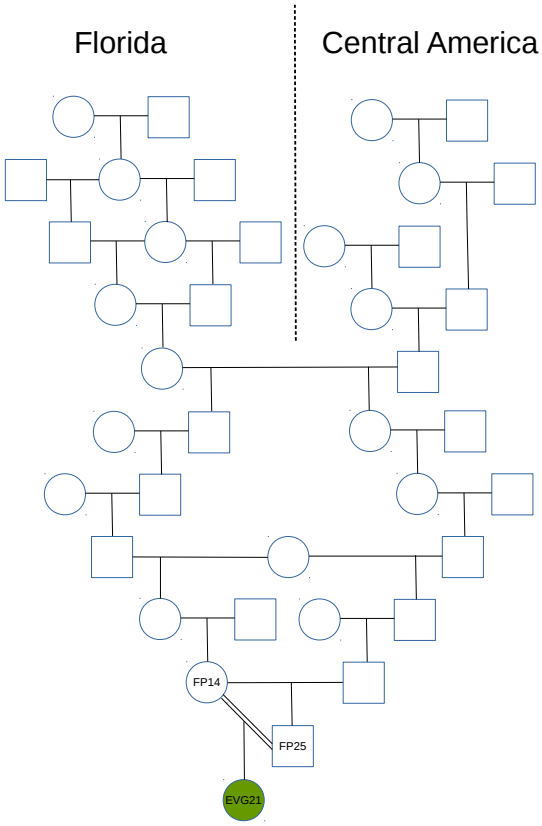

Figure 1.17: Hypothetical pedigree for EVG21. Hypothetical pedigree showing how EVG21 could have both historical admixture and be the offspring of an inbreeding event. Early generations show separate Florida and Central American populations, which come together through admixture. The ancestral Floridian population was small and experienced considerable inbreeding. The ancestral Central American population would have been much larger and therefore experienced little inbreeding. Later generations show multiple inbreeding loops, where an individual 's maternal and paternal ancestries trace back to a recent shared ancestor. Regions of the genome that are of mixed ancestry do not trace back to prior to the admixture event.

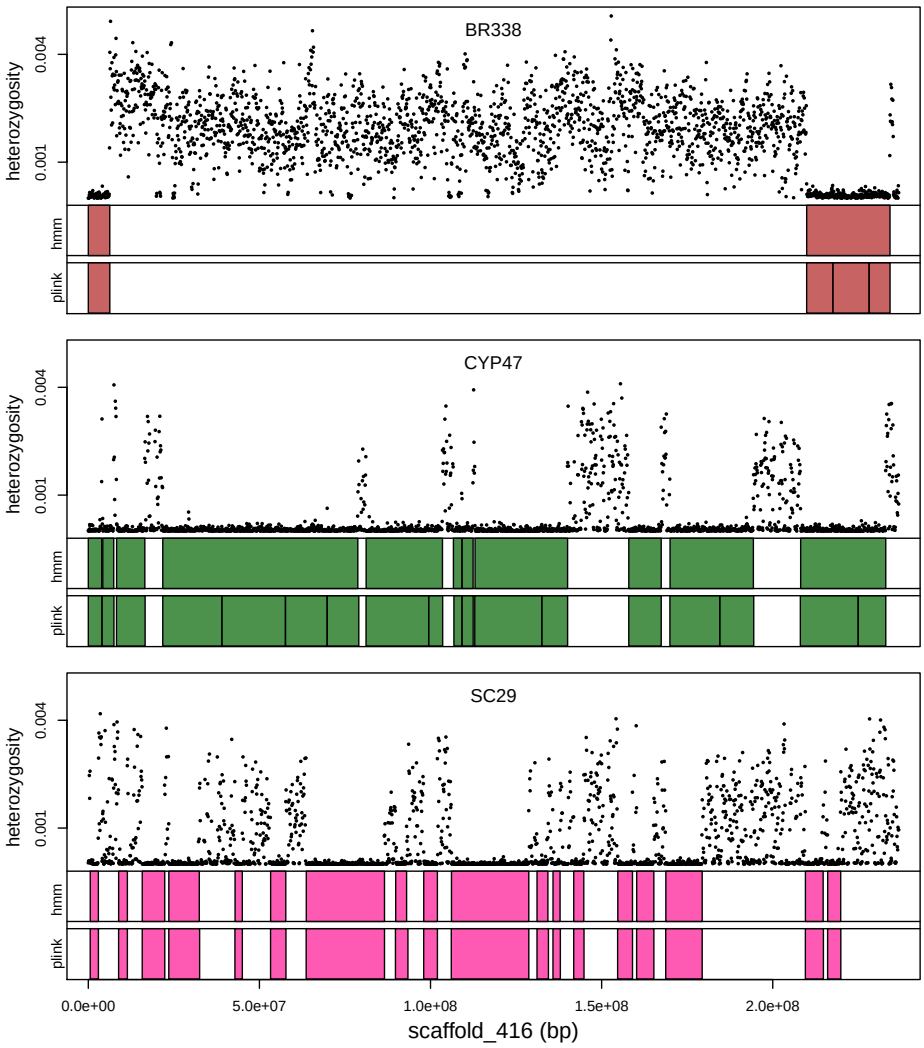

Figure 1.18: ROH calls using two different methods. Top panel for each sample shows heterozygosity in 100 kbp windows. Bottom panels for each sample shows colored boxes indicating ROH called using our ROH HMM (top), and PLINK (bottom). PLINK tended to break up long tracts of ROH. Given that we were interested in the distribution of ROH lengths, we decided to use our HMM for ROH calls for further analyses.

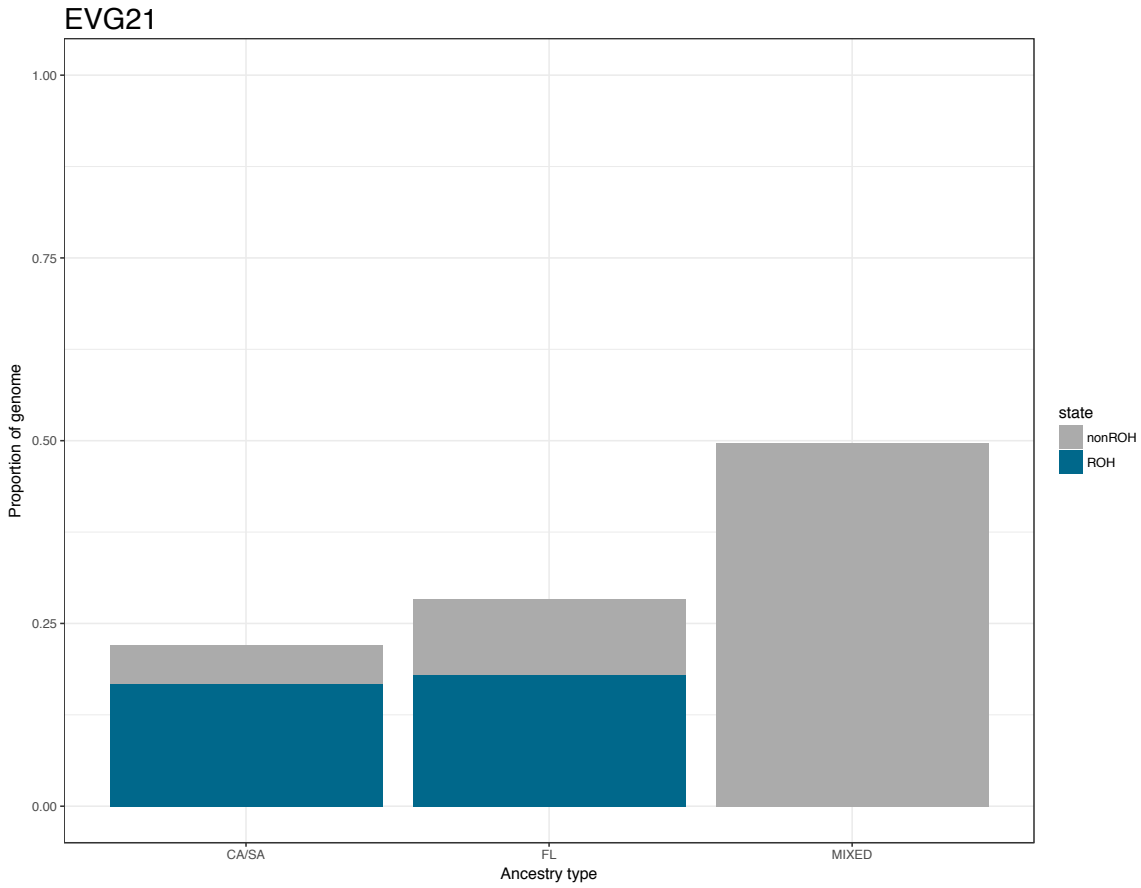

Figure 1.19: Proportion of ancestry and ROH in genome of EVG21. The proportion of the genome classified as each of three ancestry types (pure Central/South American, pure Floridian, and mixed CentralSouth American and Floridian) by Ancestry_HMM [47, 183]. The genome of EVG21 is comprised of 21.98% Central/South American ancestry, 28.24% Floridian ancestry, and 49.58% mixed ancestry based on the HMM. For all ROH greater than 2 Mb, we determined the ancestry type as classified by the HMM. While we observed no ROH greater than 2 Mb in length in mixed ancestry regions for EVG21, we find numerous ROH greater than 2 Mb for both pure CentralSouth American ancestry and pure Floridian ancestry.

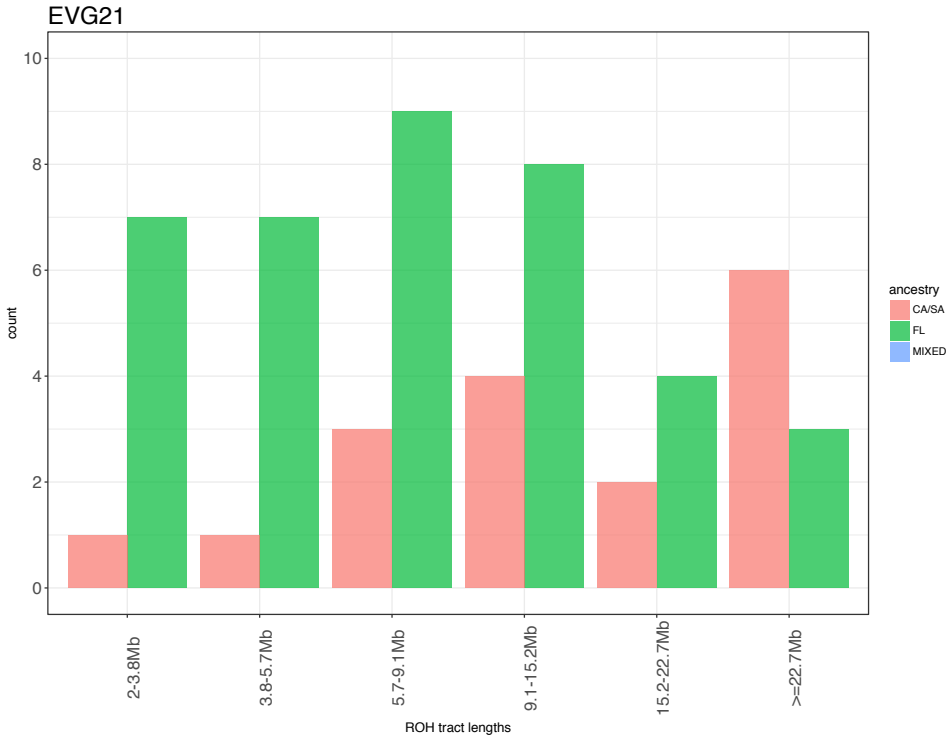

Figure 1.20: Length distribution of ROH by ancestry type in EVG21 genome. Short ROH of Floridian ancestry occur at higher numbers than short ROH of Central/South American ancestry. This is likely due to a long history of small population size in Florida, resulting in inbreeding in the population prior to admixture.

Assembly version	Meraculous	HiRise	PBJelly	Pilon iteration 1	Pilon iteration 2
Assembly step	shotgun assembly	scaffolding	gap filling	error correcting	error correcting
Input data	Illumina shotgun	Chicago + Hi-C	ONT	Illumina shotgun	Illumina shotgun
Genome length (bp)	2,181,316,782	2,293,137,739	2,433,777,904	2,433,231,347	2,432,985,507
Contig N50 (Mb)	0.01955	0.01957	0.02765	0.03101	0.03207
Scaffold N50 (Mb)	0.0366	103.78	100.51	100.54	100.53
Scaffold L50	17,135	7	8	8	8
Scaffold NG50 (Mb)	0.0351	103.78	104.57	104.60	104.60
Scaffold LG50	18,191	7	7	7	7
# of gaps	124,710	258,836	207,433	184,611	178,994
# of Ns	26,631,327	154,284,192	132,359,239	119,328,697	114,069,924
% of Ns in genome	1.22	6.73	5.44	4.90	4.69
# of shotgun reads mapping	958,095,130	979,540,843	982,339,501	985,819,801	987,305,346

Table 1.1: Genome metrics of puma assembly stages. The metrics of the puma genome at different stages of the assembly process. The final column represents the final assembly, PumCon1.0. We saw a marked improvement in N50 as a result of scaffolding with HiRise. Gap filling with PB Jelly notably decreased the numbers of Ns and strings of Ns. As a result of this conversion of Ns to useful sequence, the number of Illumina reads that mapped to the assembly increased. Due to the high error rate of ONT reads, the correction of the gap filled sequences using iterative rounds of Pilon increased the number of reads that mapped to the genome. The Meraculous assembly did not include the final versions of the X chromosome scaffolds.

Complete BUSCOs	3832 (93.4%)
Complete and single-copy BUSCOs	3815 (93.0%)
Complete and duplicated BUSCOs	17 (0.4%)
Fragmented BUSCOs	141 (3.4%)
Missing BUSCOs	131 (3.2%

Table 1.2: Benchmarking Universal Single-Copy Orthologs (BUSCO) gene completeness score. The results of running BUSCO [252] on the PumCon1.0 genome using the human gene set (n=4104).

Puma	Population	Sex, Year sampled	Coverage	Alternate IDs	SRA accession IDs
BR338	Minas Gerais state, Brazil (BR)	male, 2009	48X	bPco338 [178]	SRR7639695-6
BR406	São Paulo state, Brazil (BR)	male, 2013	27X	D406	SRR7542886-8
EVG21	Everglades National Park (EVG)	female, 1987	51X	FP021, Pco-0075 [128]	SRR7660678-9
CYP47	Big Cypress National Preserve (CYP)	male, 1992	43X	FP047, Pco-0423 [128]	SRR7664677-8
CYP51	Big Cypress National Preserve (CYP)	male, 1992	55X	FP051, Pco-0428 [128]	SRR7956993-4
YNP198	Yellowstone National Park (YNP)	male, 2015	40X	M198	SRR7610940-1
SMM12	Santa Monica Mountains (SMM)	male, 2009	46X	P12 [228]	SRR7661934-5
SMM13	Santa Monica Mountains (SMM)	female, 2009	40X	P13 [228]	SRR760239-40
SMM22	Santa Monica Mountains (SMM)	male, 2012	34X	P22 [228]	SRR7543017-8
SC29	Santa Cruz Mountains (SC)	female, 2014	35X	29F [299]	SRR7537344-5
SC36	Santa Cruz Mountains (SC)	male, 2015	47X	36M [299]	SRR7148342-54

Table 1.3: Details for the panel of pumas used in this study. *Note: SMM13 was used solely for the X chromosome scaffold assembly, and thus further metrics were not calculated.

Puma	Sex	Coverage	Heterozygosity (pileup method)	Heterozygosity (IUPAC filtered fastas)	Heterozygosity (GenomeScope, k=21*)	Proportion of genome in an ROH
BR338	male	48X	0.00155	0.00183	0.00352	0.06736
BR406	male	27X	0.00166	0.00183	0.00441	0.03970
EVG21	female	51X	0.00121	0.00128	0.00159	0.34221
CYP47	male	43X	0.00033	0.00037	0.00210	0.5485
CYP51	male	55X	0.00034	0.00040	0.00280	0.56168
YNP198	male	40X	0.00090	0.00096	0.00247	0.15873
SMM12	male	46X	0.00079	0.00086	0.00247	0.18884
SMM22	male	34X	0.00059	0.00059	0.00249	0.417637
SC29	female	35X	0.00062	0.00067	0.00165	0.32707
SC36	male	47X	0.00049	0.00060	0.00270	0.33987

Table 1.4: Coverage and heterozygosity values for the panel of pumas used in this study. *Note: For samples with lower coverage and/or lower genome-wide heterozygosity, the two peaks used by GenomeScope to estimate heterozygosity were not clearly discernible, reducing the accuracy of the estimates. Additionally, we expect that males and females will show differences due to having heterogametic or homogametic sex chromosomes for the GenomeScope values.

Sample	Genotyping error rate	Outbred Heterozygosity
BR338	0.000146	0.0019
BR406	0.000054	0.0018
EVG21	0.000071	0.0019
CYP47	0.000078	0.0011
CYP51	0.000071	0.0010
YNP198	0.000087	0.0012
SMM12	0.000027	0.0010
SMM22	0.000078	0.0012
SC29	0.000078	0.0012
SC36	0.000027	0.0010

Table 1.5: ROH HMM parameters. Parameters used an input into the HMM ROH script for each puma.

	BR338	BR406	EVG21	CYP47	CYP51	SC29	SC36	SMM12	SMM22	YNP198
BR338	6.7									
BR406	0.0	4.0								
EVG21	0.0	0.0	34.2							
CYP47	0.0	0.0	8.7	58.5						
CYP51	0.0	0.0	6.8	35.9	56.2					
YNP198	0.0	0.0	0.2	1.2	1.1	32.7				
SMM12	0.0	0.0	0.1	0.7	0.7	11.6	34.0			
SMM22	0.0	0.0	0.0	0.6	0.5	3.0	2.7	18.9		
SC29	0.0	0.0	0.1	0.9	0.8	5.1	4.1	4.4	41.8	
SC36	0.0	0.0	0.2	0.3	0.2	0.6	0.2	0.7	0.6	15.9

Table 1.6: Pairwise ROH IBD values. Percent of the genome in an IBD ROH between pairs of pumas as shown in Figure 1.4. Diagonal values are the percentage of the genome in ROH for each individual.

Chapter 2

Tremarctine bears of present and past

[Manuscript in preparation]

Nedda F. Saremi[1], James A. Cahill[2], Peter D. Heintzman[2], Joshua Kapp[2], Heather J. Milne[2], Katherine L. Moon[3], Brendan O'Connell[1], Oliver A. Ryder[4], Christopher Vollmers[1], Grant Zazula[5], Richard E. Green[1], and Beth Shapiro[1,3]

Affiliations

[1] Department of Biomolecular Engineering and Bioinformatics, UC Santa Cruz

[2] Department of Ecology and Evolutionary Biology, UC Santa Cruz

[3] Howard Hughes Medical Institute

[4] San Diego Zoo Institute for Conservation Research

[5] Government of Yukon, Department of Tourism and Culture, Yukon Palaeontology Program

2.1 Abstract

The spectacled bear is the sole extant member of the Tremarctine subfamily, and the only extant ursid species to inhabit South America. The tremarctine bears were some of the largest terrestrial carnivores, but only one species survived into the Holocene. Here, we assemble a nuclear genome for a captive female spectacled bear, and use the assembly to generate a genome for an extinct tremarctine bear, the giant-short faced bear. The spectacled bear and the giant short-faced bear both share a long history of small population size. Yet the dietary specialization and large mass of the giant short-faced bear likely made it vulnerable to extinction, while the hypocarnivorous spectacled bear survived the end Pleistocene extinctions.

2.2 Background

2.2.1 South American tremarctine bears

The sole remaining member of the Tremarctinae subfamily, *Tremarctos ornatus*, is commonly known as the spectacled bear. Today the only extant bear species to live in South America (Figure 2.1) [181], fossils for the ancestor of the spectacled bear are restricted to North America [278]. The dispersal of species between North and South America, termed the Great American Biotic Interchange (GABI), spurred migration into new ranges [57]. The first tremarctine bears presumably arrived in South America during the height of the migration period when the Panamanian isthmus land bridge first connected the two Americas roughly 2.5 million years ago (Mya) [174].

The first South American tremarctine bears were of the genus *Arctotherium*. *Arc-*

totherium species faced a scarcity of competing terrestrial carnivores in South America right after their continental arrival [258]. Early South American *Arctotherium* species were large in size. Fossils indicate *Arctotherium angustidens*, the oldest of the genus, to represent the largest bear ever recorded, making it one of the dominant carnivores on the continent [258]. However, over time, *Arctotherium* species became progressively smaller as the size of carnivore guild on the South American continent increased [83].

While a number of other tremarctine bears were present in South America during the Pleistocene based on the fossil record [258, 278], remains for the presumed ancestor of the spectacled bear, *Tremarctos floridanus*, have only been found in North America [278]. At present, the oldest specimen uncovered for the spectacled bear dates to six thousand years old [266]. The absence of the spectacled bear in the fossil record of either American continent during the Pleistocene suggests that this species diverged from its ancestor *T. floridanus* recently [259], though the timing of its migration from North to South American remains unclear.

The timing of the spectacled bear's arrival to South American had major implications for its diet. The spectacled bear arrived in South America when a plethora of rival large carnivores had already claimed dominance in the environment [258]. The abundance of competitors likely drove the spectacled bear into the overwhelmingly plant-based diet they maintain today.

Figure 2.1: The current range of the spectacled bear. The spectacled bear is the only bear to inhabit the South American continent. Spectacled bears occupy a variety of ecosystems in South America along the Tropical Andes, but are mostly found in Tropical moist forests and Tropical high-altitude grasslands [89]. Spectacled bears require access to forested areas for survival [206]. Current range data is from the IUCN Red List of Threatened Species [120].

As the closest living outgroup species to the Ursidae family, the spectacled bear represents a sister taxon to all ursines. Yet there is surprisingly little genomic data for the spectacled bear, given the wealth of research on many bear species. To date, two mid coverage genomic datasets (10X) of captive spectacled bears have been generated [144], but the lack of a reference genome meant analyses were performed using a reference species that diverged 11 Mya.

2.2.2 A North American tremarctine species: the giant short-faced bear

The giant short-faced bear, *Arctodus simus*, is estimated to have lived in North America as early as the Pleistocene, roughly 2.58 Mya [243]. Scientists have learned much about the giant short-faced bear due to its extensive North American fossil record (Figure 2.2) [181]. Yet by the end of the Pleistocene epoch ∼11,700 years ago, the bear went extinct during a large megafaunal global extinction event [138].

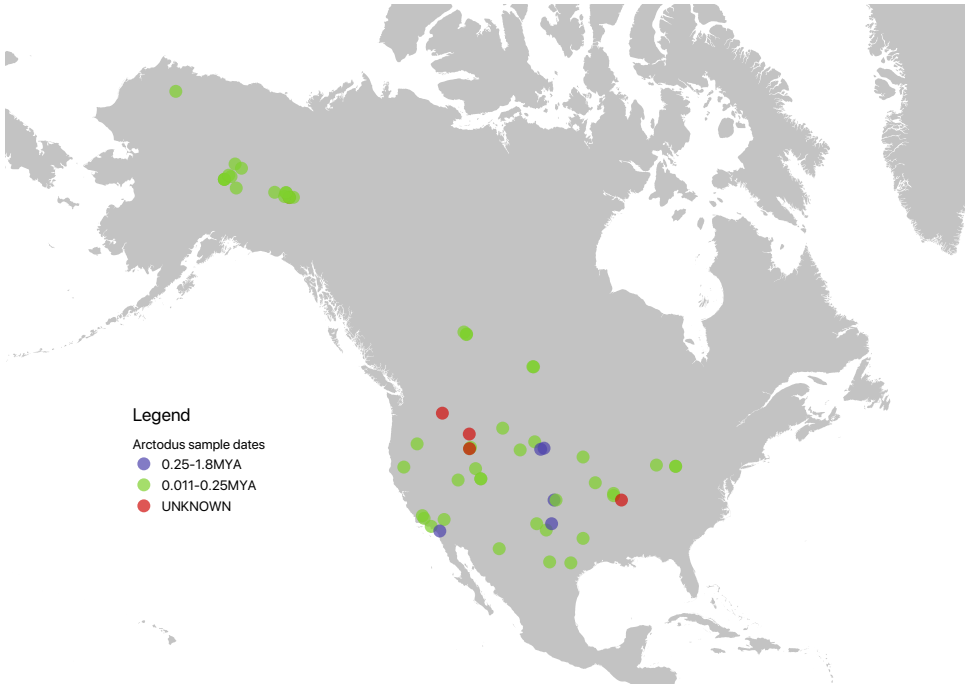

Figure 2.2: The distribution and age of giant-short faced bear fossil ages. Coordinate data downloaded from FAUNMAP [96].

The cause of these global extinctions, bar Africa, is complex [138]. It's believed that the overwhelming extinction of large mammals can be explained by the coinciding expansion of modern humans and rapid climate change. Possibly, the lethal combination of hunting with the

alteration of landscapes and range due to human expansion combined with a changing climate were what made the role of humans so key to extinction [138].

The giant short-faced bear would have been an intimidatingly large carnivore for early humans to come across in North America. The species was one of the largest bear species to have existed, standing at almost ten feet on its back legs [44]. During the Pleistocene epoch, three large bear species, representing some of the largest terrestrial mammalian carnivores, thrived: 1) *Arctotherium angustidens* in South America, 2) the cave bear, (*Ursus spelaeus*, in Europe, and 3) the giant-short faced bear in North America [258].

Mitchell *et al* investigated the similarities between two of the largest tremarctine bears: the giant short-faced bear and a South American short-faced bear, an *Arctotherium* species from Chile [190]. Phylogenetic analysis of a mitochondrial sequence from an *Arctotherium* species grouped *Arctotherium* with the spectacled bear, not the North American giant short-faced bear (Figure 2.3) [190]. Although morphologically the giant short-faced bear and *Arctotherium* are more similar, *Arctotherium* is more closely related to the only remaining tremarctine bear, the spectacled bear [190]. Presumably, the environment drove the morphological convergence of *Arctotherium* and the giant short-faced bear. Given that both *Arctotherium* and the giant short-faced bear lived in settings dominated by herbivores, both species adapted to fit a scavenger niche to survive.

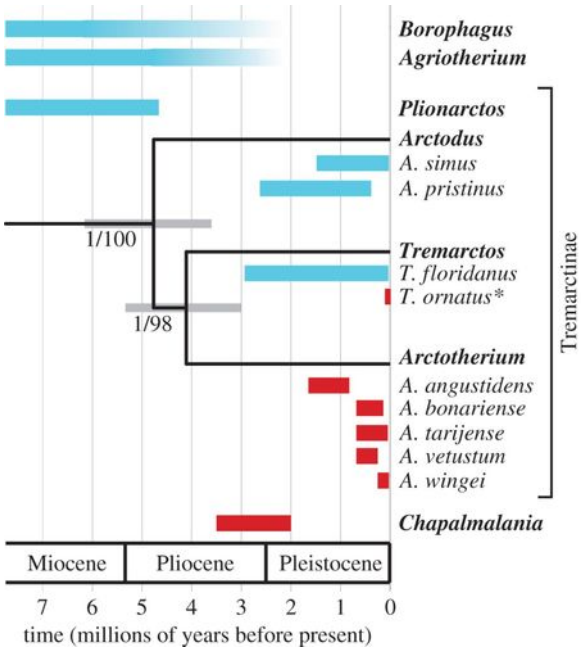

Figure 2.3: The relationships and temporal ranges of the tremarctine bears estimated from mitochondrial genomes for *Arctodus*, *Arctotherium*, and the spectacled bear (*Tremarctos ornatus*). Red denotes South American fossil distribution, and blue denotes North American fossil distribution. The extant species is denoted with an asterisk. *Borophagus*, *Agriotherium*, and *Chapalmalania* were carnivores in North and South America that went extinct by the end of the Pliocene. *Plionarctos* is the oldest known tremarctine genus [273]. Taken from Mitchell *et al*., 2016 [190].

Given the important role large carnivores play in their ecosystems, there has been much scientific research on determining the dietary behavior of the giant short-faced bear [44, 83, 260] (Figure 2.4). Prior belief held that the giant short-faced bear lived on a diet composed of largely meat, making it a hypercarnivore [145]. However, some claimed that the giant short-faced bear was mainly herbivorous, based on its morphological similarity to its closest living relative, the spectacled bear [75].

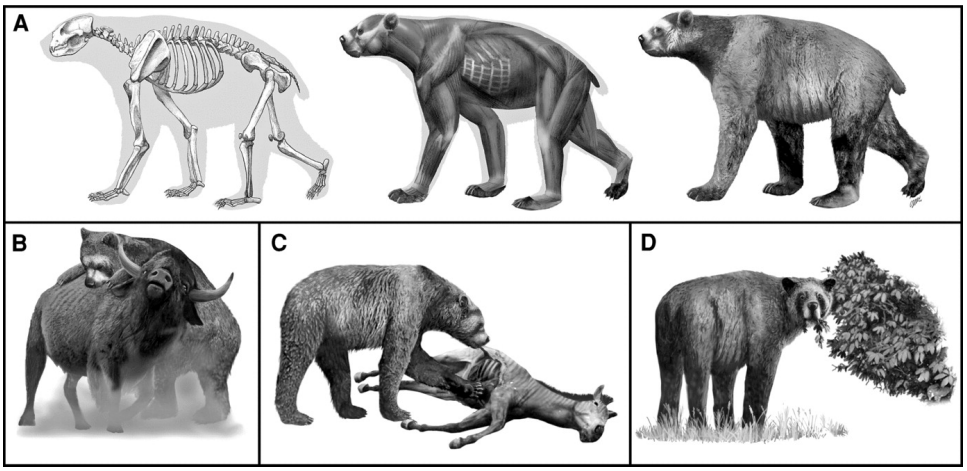

Figure 2.4: The competing theories on the dietary habits of the giant short-faced bear. Reconstruction of the giant short-faced bear based on a skull specimen. A) The skeletal reconstruction of the species. B) The giant short-faced bear as a predator as stated by Kurtèn [146]. C) The giant short-faced bear as a scavenger [177]. D) The giant short-faced bear as a herbivore as claimed by Emslie and Czaplewski [75]. Drawings by Oscar San-Isidro. Image taken from [83].

During the late Pleistocene, the giant short-faced bear and brown co-existed in North America. Both species were top predators in the region, and comparing the diets of the bears could show how the predators co-existed across their range during the late Pleistocene. Though isotopic measures do have some limitations, the method can be used to determine relative food sources for extinct and extant animals [63]. To address the question of diet, Matheus used isotopic analysis of ^{13}C and ^{15}N to predict the diet of fossil giant short-faced bears. Levels of ^{15}N for all giant-short faced bears sampled in Alaska and Yukon showed that they were highly carnivorous [177]. Meanwhile, isotopic nitrogen measures for contemporaneous brown bears show a greater range of diets, with a terrestrial vegetation component and a marine carnivore component.

^{13}C is useful for indicating carbon fixation pathways in the ecosystem's primary producers, and thus can help determine the types of prey being eaten by secondary consumers [177]. ^{13}C levels for giant short-faced bears revealed a diet composed of terrestrial herbivores [177]. Contemporaneous brown bears show significant variation in the amount of salmon and vegetation in an individual's diet [177]. Comparing the diets of the contemporaneous giant short-faced bear and the brown bear suggest little direct competition for food, meaning that the brown bear's invasion of North America was not the cause of the extinction of the giant short-faced bear. Later work has shown that brown bears likely filled the niche voided by the extinct giant short-faced bear in Beringia from 21kya to 10kya, as indicated by enriched ^{15}N values representative of carnivorous animals [16].

Combining the isotopic evidence with the morphology of the giant short-faced bear, Matheus hypothesized that the species was a scavenger of widely dispersed large mammal carcasses, requiring a large body mass for both highly efficient locomotion and intimidating other large carnivores [177]. As a scavenger in North America, the giant short-faced bear's large size would have scared off competing large felids, such as the saber-toothed cat [282], and other predators from their kills. The giant short-faced bear coincided with an explosion in ice age megafauna, including ground sloths, camels and mammoths [282]. The mass extinctions at the end of the Pleistocene would have depleted much of the large prey, thus the niche of the giant short-faced bear would have also disappeared, claims Matheus [177].

The range of the giant short-faced bear and the brown bear in North America overlapped during the Pleistocene epoch [177, 153, 16]. While Matheus showed that competition for food between the two species did not result in the demise of the giant short-faced bear

[177], the two species may have interbred. Reproductive isolation among bear species is not complete, even though some species diverged from their common ancestor several million years ago [144, 147]. A nuclear genome for the extinct giant short-faced bear enables us to determine if admixture within the ursid lineages extends further than within the ursine bears.

While population expansion in a species can be explained by adaptations or surrounding ecology, understanding the decline of a previously successful species thousands of years later is more arduous. Since data quality of ancient DNA limits our ability to produce a *de novo* assembly for extinct species, using a genome sequence for the closest living species allows us to learn about the population history and selective forces acting on giant-short faced bears to gain insight into possible causes behind its extinction. It is helpful to use a closely related species as a reference genome to obtain greater successful mappings and higher quality scores for an ancient dataset. The spectacled bear and the giant short-faced bear are believed to have diverged around 5.66 Mya [142], and thus as the sole remaining tremarctine species, the spectacled bear serves as the reference genome for this subfamily.

Here, we assemble, annotate, and analyze nuclear genomes from two bear species, the spectacled bear, *Tremarctos ornatus*, and the extinct giant short-faced bear, *Arctodus simus*. We use these data to infer the timing of divergence among these and other bears, reconstruct the demographic histories of these two species, and test for archaic admixture between the giant short-faced bear and brown bears in the late Pleistocene in Beringia.

2.3 Methods

2.3.1 DNA extraction and genome resequencing

2.3.1.1 Spectacled bear data generation

2.3.1.2 Short-read library

We obtained a sample from the Frozen Zoo of a frozen fibroblast cell line of a captive female spectacled bear, Mischief. We performed one DNA extraction using the Qiagen DNeasy Blood & Tissue kit following the manufacturer's protocol for non-nucleated erythrocytes. We made one indexed Illumina library, following the Meyer Kircher protocol [186], targeting an insert size of 330bp. The DNA was not concentrated enough to precipitate, so instead we centrifuged the sample, washed it with cold 70% EtOH, and further centrifuged it. We removed the supernatant and air dried the pellet. We resuspended the pellet in 50 µL of TE buffer at 55°C for 2 hours. We sent out the paired-end shotgun library for sequencing at UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory on an Illumina HiSeq 2500 (2x100bp) [23].

2.3.1.3 CHiCago library

We extracted additional DNA from the spectacled bear cell line using the Qiagen Blood and Cell Midi Kit (catalog no. 13343). We quantified DNA using a Qubit fluorometer. We assembled 1µg of in vitro chromatin using the Active Motif In Vitro Chromatin Assembly kit, according to the manufacturer's instructions. We crosslinked the chromatin in 1% formaldehyde for 15 minutes at room temperature, and quenched with 2.5M Glycine. We immobilized chromatin on SPRI beads at a SPRI-lysate ratio of 2:1 [62], and washed with 10mM Tris and

50mM NaCl to remove non-histone-associated DNA. We resuspended the bead and chromatin mixture in 49.5µL 1X NEBuffer 2, and digested with 5 units of DpnII enzyme for one hour at 37°C in a thermal-mixer. After digesting, we concentrated the beads and washed twice with the wash buffer. We resuspended the sample in a 50µL reaction containing dATP, dTTP, dGTP, biotinylated dCTP, and Klenow. We performed end-labeling at 25°C for 30 minutes, after which we washed the sample twice with the wash buffer. The sample was ligated overnight in a 250µL reaction containing 1X NEB T4 ligase buffer, 0.1mg/ml BSA, 0.25% Triton X-100, and 50 units T4 DNA ligase. After ligation, we added 2.5µL 10mM dNTPs and 5.5 units T4 DNA polymerase to remove unligated biotin-dCTP. After concentrating the sample and removing the ligation buffer, the crosslinks were reversed and the sample was deproteinated in 50 µL cross-link reversal buffer (50mM Tris pH=8.0, 1% SDS, 0.25mM CaCl2, and 0.5 mg per mL Proteinase K). We incubated the sample at 55°C to digest the histones, then increased the temperature to 68°C to reverse crosslinks. We then separated the sample from the beads and purified on fresh SPRI beads at a ratio of 2:1 [220]. DNA recovery was quantified with a Qubit fluorometer. We prepared 400ng of sample for sequencing using the NEB Ultra library preparation kit according to the manufacturer's instructions, with one exception: prior to the indexing PCR, the sample was enriched by pulldown on 30µL Invitrogen C1 Streptavidin beads, then washed to remove non-biotinylated DNA fragments. The washes were as follows: 1 wash with lithium wash buffer (1M LiCl, 0.5mM EDTA, 0.05% Tween-20), followed by 2 washes with sodium wash buffer (1M NaCl, 0.5mM EDTA, 0.05% Tween-20), and 2 washes with TE-Tween. We sent the library for sequencing at the UC San Diego Institute for Genomic Medicine Genomics Center on an Illumina HiSeq 2500 (2x125bp reads) [23].

2.3.1.4 Oxford Nanopore library

We extracted genomic DNA from the spectacled bear cell line sample with the Qiagen Blood and Cell Culture Mini Kit, using 750µL of starting material. We followed the protocol exactly, up until the spooling step. The DNA was not concentrated enough to precipitate, so we instead centrifuged the sample at 5000g for 15 minutes, washed it with cold 70% EtOH, and centrifuged it again at 5000g for 10 minutes. We removed the supernatant and air-dried the pellet. We resuspended the pellet in 50µL of TE buffer on a shaker at 22°C overnight. We quantified DNA with a Qubit 2.0 dsDNA HS kit at 71.5 ng per µL (3.58 µg total). We verified the DNA size distribution with a pulse-field gel using a 0.75% agarose TAE gel, run at 75V for 16 hours, with the preset 5-150 kb program on the Pippin Pulse power Supply (version 1.3.2), and estimated the DNA fragments to average roughly 23 kb in length.

We performed a Rapid Sequencing run (Oxford Nanopore, SQK-RAD004). For the Rapid library preparation, we used an input of 200ng of DNA, as recommended by the manufacturer. We first treated the high molecular weight DNA with 2.5µL of fragment repair mix and incubated for 1 minute at 30°C, followed by 1 minute at 80°C. We ligated the repaired DNA with Rapid Adapters using Blunt/TA Ligase (NEB). We quantified the Rapid libraries using a Qubit prior to sequencing. The Rapid sequencing library was run on an R9.5 flow cell using the NC_48hr_Sequencing_FLO-MIN107_SQK_RAD004 protocol. We base called Fast5 raw reads generated from the sequencing run using the latest Albacore software (Oxford Nanopore proprietary).

2.3.1.5 RNA-Seq Illumina library

We followed the NEBNext Poly(A) mRNA magnetic isolation protocol to make a cDNA library from fibroblast cells from Mischief. We assessed the RNA-Seq library size distribution using an Agilent 2100 Bioanalyzer. We sent the RNA-Seq library for sequencing at the UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory on an Illumina HiSeq 2000 (2x75bp).

2.3.1.6 Oxford Nanopore RNA library

We generated a long-read RNA library for the spectacled bear, extracting genomic DNA from the spectacled bear cell line sample with the Qiagen Blood and Cell Culture Mini Kit. For one library, we followed the protocol for the 1D2 genomic DNA sequence for the ONT using SQK-LSK308. The Rapid sequencing library was run on an R9.5 flow cell using the NC_48hr_Sequencing_FLO-MIN107_SQK_LSK308 protocol. We base called Fast5 raw reads generated from the sequencing run using the latest Albacore software (Oxford Nanopore proprietary).

2.3.1.7 Giant short-faced bear data generation and radiocarbon dating

We obtained three bone samples from the Yukon Government: YG 76.4, YG 24.1, and YG 546.562, all identified as *Arctodus simus*. YG 76.4 was a radius bone obtained from Hester Creek in Yukon, Canada. YG24.1 was a petrous bone found in Ophir Creek near Dawson Creek in Yukon, Canada. YG 546.562 was a right femur found in Canyon Creek, Yukon Canada. We subsampled the two of the three fossil bone specimens with a handheld rotating cutting

tool and submitted these to the Keck Carbon Cycle AMS facility at the University of California, Irvine (UCIAMS) and Lawrence Livermore Laboratory for radiocarbon dating. At each facility, collagen was extracted using a modified version of the Longin [31] method with ultrafiltration collagen pretreatment [19]. We calibrated radiocarbon dates to calendar years using OxCal (version 4.3) [224] using the atmospheric radiocarbon curve [226] (Table 2.1).

Sample name	Museum ID	Sample location	Radiocarbon date	Calendar age	Radiocarbon accession
Hester	YG 76.4	Hester Creek, Yukon	26,520 $\pm$ 110	29,070 to 28,631calBC	CAMS-166313
Ophir	YG 24.1	Ophir Creek, Yukon	20,250 $\pm$ 110	22,693 to 22,037 calBC	Beta-79852
Canyon	YG 546.562	Canyon Creek, Yukon	27,850 $\pm$ 220	30,403 to 29,242 calBC	UCIAMS 186671

Table 2.1: Sample details for giant short-faced bear samples used in this study. Sample names, locations, and radiocarbon dates for the samples we present in this study.

All work was performed in a dedicated ancient DNA facility at UC Santa Cruz, following standard clean room criteria [46]. For Ophir, we received powdered petrosal bone which we extracted following the Dabney protocol [51]. For Hester, we sampled the bone, and reduced to powder using a Mixer Mill MM 400 (Retsch). We collected 100 mg for DNA extraction for both samples. Bone powder was incubated for 24 hours at 37°C in an extraction buffer optimized from bone (0.45M EDTA, 0.25mg/mL Proteinase K). DNA was isolated using the silica column based Dabney method [51] with a final elution of 50µL buffer EBT (10mM Tris, 0.05% Tween-20).

We made four DNA libraries for Ophir and one library for Hester following the Meyer Kircher protocol [186], except for reducing the input volume of template to 20 µL and following the Pennsylvania State University library protocol [286]. All libraries underwent indexing PCR after which their concentrations were estimated using quantitative PCR. We pooled and sequenced all libraries for these two samples at the UC San Francisco Center for Advanced

Technology on an Illumina HiSeq 2500 (2x50bp).

The Canyon sample was crushed and reduced to powder using a Mixer Mill MM 400 (Retsch). We collected 50-115 mg for DNA extraction. Prior to extraction, a bleach pre-treatment was performed as outlined in prior work [235]. Bone powder was incubated for 24 hours at 37°C in an extraction buffer optimized from bone (0.45M EDTA, 0.25mg/mL Proteinase K). DNA was isolated using the silica column based method outlined in [51] with a final elution of 50µL buffer EBT (10mM Tris, 0.05% Tween-20).

We prepared single-stranded libraries using an ancient DNA optimized version [132] of the method outlined in [279] for DNA extracts. Single-stranded libraries were indexed and amplified in 100µL reactions containing 48µL pre-amplified library, 50µL AmpliTaq Gold 360 Master Mix, 1µM i7 indexing primer, and 1µM i5 indexing primer. Libraries were amplified in a Bio-Rad T100 thermocycler using the following conditions: 95°C for 10m, followed by 15 cycles of 95°C for 30s, 60°C for 30s, and 72°C for 60s, followed by 72°C for 7m. Post-amplified libraries were purified using a 1.2X SPRI clean. All post-amplified and cleaned libraries were quantified using a Qubit 4 (Invitrogen) and the Qubit 1X dsDNA HS assay kit (Invitrogen). Finally, all post-amplified libraries were visualized on a TapeStation 2200 (Agilent) using a D1000 High Sensitivity Assay (Agilent). Single-stranded libraries were pooled and sequenced at UC San Francisco Center for Advanced Technology on a NovaSeq S4 (2x100bp).

2.3.2 Genome assemblies and annotation

2.3.2.1 *de novo* nuclear genome assembly of the spectacled bear

We removed adapters from the shotgun Illumina library for Mischief with SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15) and reducing the minimum length of trimmed reads to 25 bp (-L 25) [262]. We used Trimmomatic [27] to 1) remove additional Meyer Kircher IS3 adapters [186] using a seed mismatch of 2 and a simple clip threshold reduced to 5 for the shorter adapter sequence, 2) quality end trim using minimum qualities of 2 or 5 for leading and trailing ends of reads, respectively, 3) window quality trim, using a window size of 4 and a minimum quality of 15, and 4) remove reads shorter than 50 bp. We used this processed shotgun data to assemble a de novo genome using the Meraculous-2D Genome Assembler [39], with diploid mode set to 1 and a kmer size of 51. The contig N50 of the shotgun assembly was 20.1 kb and the scaffold N50 was 26.2 kb. We had approximately 64X coverage of the genome from the one Illumina shotgun library.

We scaffolded the Meraculous assembled genome using HiRise [220], run in Chicago mode using the default parameters with the Chicago library as input. The resulting scaffold N50 was 20.96 Mb. We performed gap filling on the HiRise scaffolded assembly using Oxford Nanopore Technologies (ONT) long reads with the tool PBJelly, part of the PBSuite [77]. We set PBJelly to correct only intrascaffold gaps, and used the default minimum of 1 read spanning a gap. PBJelly also resolved the sizes of the gaps introduced by HiRise. We used Porechop [234] to adapter trim the ONT reads. The resulting reads provided 1.6X coverage of the genome. We reduced the number of gaps (represented in the assembly as a series of Ns) in the genome from

171,229 to 162,727 and reduced the total number of Ns in the genome from 131,322,142 to 125,562,584.

We used the genome improvement tool Pilon [289] to correct sequence errors in the gaps that were filled with the high-error ONT data. We first aligned the Illumina shotgun data back to the PBJelly genome using bwa mem [159] and marked duplicates with Picard toolkit MarkDuplicates [4]. We then ran Pilon with the alignment file as the "– frags" input, the PBJelly genome as the genome file, specified the genome as diploid, and used the default setting to fix all types of changes. We ran two iterations of alignment and consensus sequence calling.

We used the genome assessment tool BUSCO [252] to evaluate genome completeness based on a set of conserved single-copy orthologous genes (human gene set; n=4104). In the TreOrn1.0 genome, 93.5% of these genes are complete and present only in single copy. The final genome assembly was 2,232,736,973 bp in length and had an N50 of 21.11 Mb, with 151,338 gaps and 123,688,823 Ns. Scaffold larger than 1Mb constitute roughly 85% of the spectacled bear genome.

2.3.2.2 Evidence-driven spectacled bear genome annotation

The TreOrn1.0 genome was run through RepeatScout [216] to identify repeat family sequences based on high-frequency repetitive kmers. The result of RepeatScout was used an input into the genome masking tool Repeat Masker [256]. The resulting repeat-masked fasta file would be used as the reference to map RNA sequencing data, as well as input for the annotation tool Augustus in gene feature file (gff) [264].

Short-read Illumina RNA data was adapter trimmed using SeqPrep2 using the default

parameters except for increasing the quality score cutoff to 15 (-q 15) and reducing the minimum length of trimmed reads to 30 bp (-L 30) [262]. Adapter-trimmed reads were aligned to the the repeat-masked TreOrn1.0 genome using TopHat2 [137]. The alignment file was filtered using bam2hints from the Augustus toolkit [264] to only include intronic regions when converted to gff to be used as a hint for annotation [264]. Adapter-trimmed ONT RNA reads were mapped to the repeat-masked TreOrn1.0 genome using gmap [301], keeping only alignments with maximal scores. Alignments were converted into hints for input into Augustus using blat2hints.psl [264]. Publicly available proteins from the polar bear [169] were download from gigadb and mapped to the masked TreOrn1.0 genome using Exonerate [255] and converted into hints using the 'exonerate2hints' perl script from Augustus [264]. Hints from repeat masking, long read data, short-read data, and polar bear proteins were incorporated into one file for input as hints into Augustus.

Evidence-guided annotation of the spectacled bear genome was performed using the annotation tool Augustus [264], using the human as the species, allowing alternatives from evidence, hinted splice sites, untranslated regions, no in frame stop codons, and a complete gene model. Augustus predicted a total of 19,289 genes in the spectacled bear. The output gff file was converted into amino acid sequences using the getAnnoFasta.pl script from Augustus [264]. Predicted proteins were run through blastp [12] using the Swiss-Prot database downloaded on May 2019 [7]. Hits were filtered such that only the three best hits were returned, requiring all hits to have at least an e-value cutoff of 0.00001. Hits were filtered such that the hit with the highest bit score was kept for each predicted protein sequence. We used the UniProt website [7] to convert the Swiss-Prot IDs of the hits into gene names and GO terms for each hit. The

number of annotated predicted proteins was 15,504. We subclassified the annotated genes into the three GO domains, biological process, molecular function and cellular component, to see the distribution of GO names per category visualized using Blast2GO [95] (Figure 2.5).

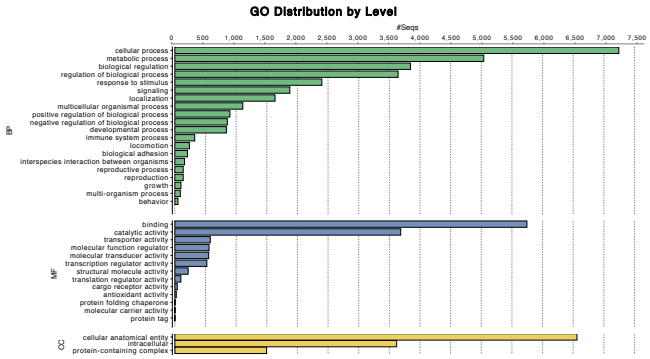

Figure 2.5: Twenty most common GO term names for each GO category based on the annotated sequences from the spectacled bear visualized using Blast2GO [95].

2.3.2.3 Modern samples nuclear data processing

We downloaded two publicly available shotgun datasets of roughly 10X nuclear coverage for two captive spectacled bears, Chaparri and Nobody [144] (Tables 2.5). We also downloaded shotgun datasets for six brown bears, two polar bears, and one giant panda (Tables 2.5, 2.6). We removed adapters using SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15) [262] and mapped reads to the TreOrn1.0 genome using the bwa mem algorithm [159], filtering for map quality of 30, removing alignments that are not primary, or have an unmapped mate. We removed PCR duplicates using samtools rmdup

[158]. We performed realignment around insertions and deletions was performed using GATK Realigner Target Creator and Indel Realignment [180].

We adapter trimmed reads using SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15) [262] for all three spectacled bears, Chaparri, Nobody and Mischief. We mapped reads to the TreOrn1.0 genome using the bwa mem algorithm [159], filtering for map quality of 30, removing alignments that are not primary, or have an unmapped mate. We removed PCR duplicates using samtools rmdup [158]. The average nuclear coverage for Mischief was 45.97X, for Chaparri was 11.12X, and 10.81X for Nobody. Realignment around insertions and deletions was performed using GATK Realigner Target Creator and Indel Realignment [180].

We assess the level of genetic variation in three bear species by creating pseudohaploid fastas for each of the six samples (PB12, PB42, OFS01, SJS01, Mischief, Nobody). We used the realigned alignment file as input, and randomly selected a single high quality base using ANGSD makeFasta 1 to generate pseudohaploid fasta files [141]. We measured the number of pairwise differences in non-overlapping 100kb windows between each pair of bear species. We plotted the number of pairwise differences per 100,000 sites which existed between each pair of bears (Figure 2.14).

Sample name	Species	Location	Coverage	SRA/ERR	Study
Nobody	*T. ornatus*	Zoo Basel	10.81	ERR946789	[144]
Chaparri	*T. ornatus*	Zoo Basel	11.12	ERR946788	[144]
Mischief	*T. ornatus*	San Diego Zoo	45.967	tbc	this study
OFS01	*U. arctos*	Östanvik, Furudal, Sweden	26.93	SRR935590, SRR935593-4, SRR935626	[169]
SJS01	*U. arctos*	Slakka, Jokkmokk, Sweden	16.34	SRR935591, SRR935625, SRR935627	[169]
PB12	*U. maritimus*	Iskanten Hvalsund, West Greenland	27.98	SRR942301, SRR942293	[169]
PB42	*U. maritimus*	Melville Bugt, West Greenland	22.80	SRR942281, SRR942282	[169]

Table 2.2: Bears used for pairwise diversity analysis. Pairs of brown bears and polar bears to assess the number of pairwise differences in bear species compared to the spectacled bear. Coverage is calculated using the spectacled bear genome as the reference.

2.3.2.4 Ancient samples nuclear data processing

We adapter trimmed and merged reads using SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15) and the overlap for merging (-o 20 for Ophir and Hester, -o 15 for Canyon) [262]. For Ophir and Hester, we used Trimmomatic [27] to 1) remove additional Meyer Kircher IS3 adapters from the merged reads [186] using a seed mismatch of 2 and a simple clip threshold reduced to 4 for the shorter adapter sequence, 2) quality end trim using minimum qualities of 2 or 5 for leading and trailing ends of reads, respectively, 3) window quality trim, using a window size of 4 and a minimum quality of 15, and 4) remove reads shorter than 25 bp (Table 2.3.

To prevent any non-authentic ancient DNA from being included in further analyses, we aligned only the merged reads to the TreOrn1.0 genome sequence using bwa aln [162] using ancient DNA parameters (-l 1024 -n 0.01 -o 2). We filtered such that all alignments had map quality of at least 30 and all alignments were primary. We removed PCR duplicates using samtools rmdup [158]. We sought to prevent any spurious mismappings that may occur due to the short read length of aDNA by removing reads shorter than 35bp. We chose 35bp as our threshold due to prior work which has shown that this fragment length considerably decreased the number of microbial alignments to the genome [185, 59]. We also lost only 1X genome coverage for our high coverage giant short-faced bear by filtering reads of less than 35bp in length compared to 25bp in length, and thus used the more conservative threshold for further analyses (Figure 2.3).

Sample name	25bp Nuclear genome coverage	25bp read count	35bp Nuclear genome coverage	35bp read count
Canyon	27.20 X	839,492,728	26.01 X	758,541,872
Hester	2.07 X	72,239,835	1.82 X	50,492,295
Ophir	1.89 X	63,330,908	1.66 X	43,107,072

Table 2.3: Effects of read length filtering in coverage for aDNA samples. Read counts were generated using 'samtools view -c'.

We realigned around insertions and deletions using GATK Realigner Target Creator and Indel Realignment [180]. We used mapDamage2 to assess the 5' and 3' deamination characteristic of ancient DNA to determine the validity of the data (Figure 2.6) [130]. All giant short-faced bear samples showed damage characteristic of ancient DNA.

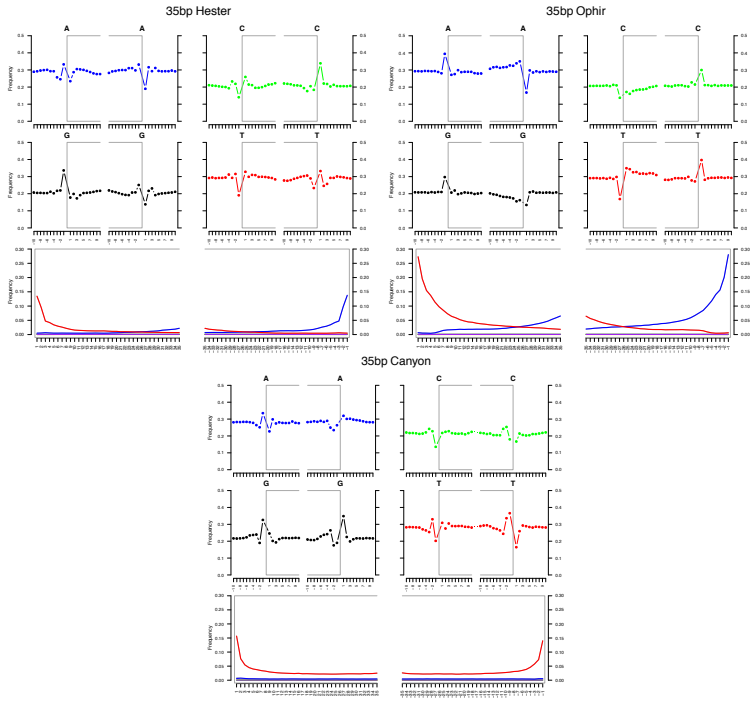

Figure 2.6: DNA fragmentation profiles show the base composition of the first/last 10 read positions and within the 10 genomic positions preceding and following read ends. Nucleotide mis-incorporation profiles provide C to T (red), G to A (blue) and other (grey) occurrences of mutations relative to the occurrence of the reference nucleotide along the first/last 10 read positions. We observe elevated levels of nucleotide mis-incorporations and a damage profile that is characteristic of ancient DNA for Hester, Ophir, and Canyon samples. We observe only damage associated with aDNA on the original template strand for the Canyon sample as a result of the single-stranded Santa Cruz method [132].

2.3.2.5 Mitochondrial genome assemblies of ursids

We used Unicycler [298] in hybrid assembly mode to assemble a mitochondrial sequence for Mischief the spectacled bear. To decrease our initial input dataset, we first identified mitochondrial mapping reads. To do this, we mapped adapter-trimmed Illumina and ONT reads from Mischief to a publicly available spectacled bear reference mitochondrial sequence (FM177764.1) [142] using bwa mem [159] and Minimap2 [160], respectively. We took the reads in the Illumina and ONT alignment files that mapped to the reference mitochondrial sequence and converted them into fastq format using bedtools bamToFastq [222]. We used both the Illumina and ONT mitochondrial-mapped readsets as input into Unicycler as long reads using the flag "l" and unpaired reads using the flag"-", respectively. The assembly created with Unicycler was 16,722 bp in length and circular. To validate this assembly, we used the Unicycler output as the reference sequence in an assembly using mapping iterative assembler (mia) [98] using roughly 24 million randomly selected adapter-trimmed Illumina shotgun reads (mia flags: -i -c -C -U -F -k 13). We noticed a slight dip in coverage in Hyper Variable Region (HVR) 2, a region of the mitochondria which is known to have a highly variable number of repeats. We assessed this region by eye to look for inaccuracies in the alignment of reads over this segment of the mitochondria. To fix the dip in coverage in this repeat region, we manually removed one of the 17 bp repeats using Jalview [293]. We checked this edited assembly by using the sequence as the reference for mia [98], again using roughly 24 million adapter-trimmed Illumina shotgun reads. We observed no changes in sequence in the outputted assembly, and no longer observed a dip in coverage across HVR2. Thus we used this sequence as the final mitochondrial

genome for Mischief.

We aligned 10 million randomly selected adapter-trimmed reads for Nobody and Chaparri to the Mischief spectacled bear mitochondrial assembly using mia (mia flags: -i -c -C -U -F -k 13) [98]. The coverage of the Nobody assembly was 31.55X and the Chaparri mitochondrial assembly coverage was 29.43X.

Shotgun reads for Hester, Ophir and Canyon were adapter trimmed using SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15), merging reads with an overlap of 20bp (-o 20), and reducing the minimum length of trimmed reads to 25 bp (-L 25) [262]. We used the merged reads as input into the mitochondrial assembler mia [98], with the publicly available *Arctodus simus* mitochondrial genome as the reference (NC011116.1) [142], and an ancient DNA substitution matrix. The assembly coverage for Hester was 19.22X, Ophir was 66.12X, and the assembly coverage for Canyon was 44.88X.

We downloaded a publicly available dataset of roughly 1 million filtered, de-multiplexed, and collapsed Illumina MiSeq reads from an mitochondrial enriched *Arctotherium* genomic library [190]. We assembled a mitochondrial sequence with the Arctotherium dataset using the publicly available spectacled bear (FM177764.1) [142] mitochondrial sequence as the reference for a mia assembly [98], using an ancient DNA substitution matrix to account for deamination. The final assembly for *Arctotherium* had 126X mitochondrial coverage.

We filtered the seven assemblies using a consensus threshold of 66% and 3X coverage per site. Any site which did not meet these requirements was changed to an 'N'.

We dowloaded publicly available mitochondrial sequences for a sun bear, sloth bear, cave bear, Asiatic black bear, giant panda, brown bear, polar bear, spectacled bear, ABC is-

land brown bear, and American black bear for the construction of a mitochondrial phylogeny (FM177765.1, FM177763.1, FM177760.1, FM177759.1, EF212882.1, EU497665.1, GU573490.1, FM177764.1, JX196368.1, AF303109.1, respectively) [142] [211] [28] [167] [189] [61].

2.3.3 Phylogenetics of ursids

2.3.3.1 Mitochondrial phylogeny of ursids

We used clustal omega [251] to align the bear mitochondrial genomes. Using Geneious [135], we partitioned the alignment into six datasets: free sites, control region, rRNA, tRNA, 1st and 2nd coding positions, 3rd coding positions based on annotations from the Asiatic black bear (FM177759.1), polar bear (GU573490.1), and spectacled bear (FM177764.1). We ran PartitionFinder [150] to determine the partitions and best substitution model, with branch lengths unlinked using the Bayesian Information Criterion. PartitionFinder separated the data into three mitochondrial partitions; 1) control region (Hasegawa-Kishino-Yano + gamma), 2) free sites, rRNA, tRNA ,1st and 2nd coding positions (general time reversible + gamma + invariable sites), and 3) 3rd codon positions (general time reversible + gamma + invariable sites).

We performed Maximum likelihood (ML) and Bayesian phylogenetic analyses using the partitioning specified above. We used RAxML [263] to produce a ML phylogeny, running one hundred bootstrap replicates. We created a phylogeny using a Bayesian approach using MrBayes [232] with the same partitioning as above. We ran four chains (one hot, three cold) for 10 million generations, with trees and model parameters sampled every 1,000 generations, with the first 25% discarded as burn-in.

2.3.3.2 Nuclear divergence times of ursids

We used ProteinOrtho [151] to generate a set of single copy orthologs 1:1 using the codon sequences of two bear species, giant panda and polar bear. We removed orthologs shorter than 300bp in length to generate a set of 14,332 orthologs. We mapped data for eight extant bear species and the giant short-faced bear to the identified orthologs using bwa mem [159] (Table 2.4). For the giant short-face bear, we used ancient DNA parameters (bwa aln -l 1024 -n 0.01 -o 2) [162]. We filtered to keep all primary alignments with a map quality of at least 30. We removed PCR duplicates using samtools rmdup [158]. We realigned around insertions and deletions using GATK Realigner Target Creator and Indel Realignment [180]. We combined all realigned bam files for the nine bears and called variants using freebayes requiring a minimum mapping quality of 25, a minimum allele fraction of 0.3, minimum alternate count of 4 and minimum coverage of 4 [90]. We removed indels and created fasta files from the vcf for each species using vcf-consensus [55]. We filtered to keep sequences that maintained codon integrity, keeping 13,713 sequences. We extracted only four fold degenerate sites, for a total of 3,415,480 bp of sequence. We aligned the fasta sequences for the nine bears using mafft [134] and used this as input for MCMCTree [302]. We ran MCMCTree with one fossil constraint and one tip date for the giant short-faced bear using the HKY model, running for 10 million generations with the first 5 million generations discarded as burn in. We checked for convergence of the posterior estimates using Tracer [223]. We repeated the analysis again to validate results.

Bear species	SRA/ERR	Study
American black bear	SRR7813600-1	[261]
Asiatic black bear	ERR946787	[144]
brown bear	SRR935590, SRR935593-4, SRR935626	[169]
giant panda	SRR019741-50, SRR019778-84, SRR019835-9, SRR019843, SRR019845-8, SRR019851-5, SRR019857-62, S RR019871-3, SRR019875-8	[164]
giant short-faced bear	tbc	this study
polar bear	SRR942293	[169]
sloth bear	ERR946786	[144]
spectacled bear	tbc	this study
sun bear	ERR946784	[144]

Table 2.4: Bears used for nuclear divergence time estimation. Species, read data, and study details for the eight extant bears and one extinct bear species used to construct a nuclear phylogeny with MCMCTree [302].

2.3.4 Demographics of the spectacled bear and the giant short-faced bear

2.3.4.1 Modeling the population history of tremarctine bears

We used the pairwise sequentially Markovian coalescent (PSMC) model [163] to estimate the historical effective population size of the spectacled bear. The input was a realigned alignment file, where scaffolds less than 1Mb in length were removed. For each bear we used sites between one third and twice the average coverage for that bear. We used a generation time of 6 years and a mutation rate of 0.6e-8 per bp per generation, based on previous estimates of the ursid mutation rate [144]. Due to the lower coverage of the two captive spectacled bears, Chaparri and Nobody, compared to Mischief we downsampled the alignment file for Mischief to observe the effects of coverage on the PSMC model. We downsampled to approximately the same coverage as Chaparri and Nobody 'samtools view -s 0.23' [158]. We observe that lower

coverage (10-11X) alters the PSMC model, especially at the oldest time points, and thus is not an accurate representation of the demographic history of the spectacled bear. We also performed one hundred replicate bootstraps for each individual per the software instructions.

With our high coverage dataset for Canyon, the giant short-faced bear, we were able to use the PSMC model [163] to estimate its historical effective population size. The input was a realigned alignment file mapped to the spectacled bear genome, where scaffolds less than 1Mb in length were removed. We used sites between one third and twice the average coverage. We used a generation time of 6 years and a mutation rate of 0.6e-8 per bp per generation, based on previous estimates of the ursid mutation rate [144]. To account for the age of our ancient sample, we rescaled the x axis, by adding the calibrated age (29,242 calBC) to the scaled time points, thus pushing back the start of the PSMC model to 29,242ya. Given that the most recent time points in the PSMC model (typically younger than 10ky) are less accurate [163], we ignore demographic estimates younger than 40kya in the plot. We also performed ten bootstrap replicates, scaling each by the sample age per the software instructions [163].

2.3.4.2 Pairwise diversity of bears

We assess the level of genetic variation in three bear species (polar bear, brown bear, and spectacled bear) by creating pseudohaploid fastas for each of the six samples (PB12, PB42, OFS01, SJS01, Mischief, and Nobody). We used the realigned alignment file as input, and randomly selected a single high quality base using ANGSD 'makeFasta 1' to generate pseudohaploid fasta files [141]. We measured the number of pairwise differences in non-overlapping 100kb windows between each pair of bear species. We plotted the number of pairwise differ-

ences per 100,000 sites which existed between each pair of bears.

Sample name	Species	Location	Coverage	SRA/ERR	Study
Nobody	*T. ornatus*	Zoo Basel	10.81	ERR946789	[144]
Chaparri	*T. ornatus*	Zoo Basel	11.12	ERR946788	[144]
Mischief	*T. ornatus*	San Diego Zoo	45.967	tbc	this study
OFS01	*U. arctos*	Östanvik, Furudal, Sweden	26.93	SRR935590, SRR935593-4, SRR935626	[169]
SJS01	*U. arctos*	Slakka, Jokkmokk, Sweden	16.34	SRR935591, SRR935625, SRR935627	[169]
PB12	*U. maritimus*	Iskanten Hvalsund, West Greenland	27.98	SRR942301, SRR942293	[169]
PB42	*U. maritimus*	Melville Bugt, West Greenland	22.80	SRR942281, SRR942282	[169]

Table 2.5: Bears used for pairwise diversity analysis. Pairs of brown bears and polar bears to assess the number of pairwise differences in bear species compared to the spectacled bear. Coverage is calculated using the spectacled bear genome as the reference.

2.3.4.3 Mappability masks for spectacled bear genome

We generated mappability masks for each scaffold in the spectacled bear genome to identify regions on which short reads can uniquely map. We used seqbility on the adapter trimmed data for Mischief using kmers of 35bp and a stringency of 0.90 to generate the mask files as described in prior work [217].

2.3.4.4 Archaic admixture between brown bears and giant short-faced bears

We used the D statistic to look for evidence of archaic admixture between the giant short-faced bear and the brown bear due to their range overlap in the late Pleistocene in Beringia. In addition to the alignment file we had created previously for the our high coverage giant short-faced bear Canyon, we included alignment files for six brown bears and a giant panda (Table 2.6). From the alignment files, we generated pseudohaploid fasta files for all eight bears using ANGSD '-dofasta 1 -minMapQ 30 -minQ 30' to randomly select a base for each scaffold larger than 1Mb [141]. We compared every pair of brown bears, using the giant short-face bear as the introgressor and the giant panda as the outgroup, with quad-aln-report from ChromCompare [6] to generate counts of ABBA and BABA sites in non overlapping 100kb windows, masking CpG sites, and looking only at transversions. We calculated the D-statistic using quad_dstat_wt_jk from ChromCompare [6], as well as the weighted block jackknife to estimate standard error and significance. No brown bear showed signs of archiac admixture with the giant short-faced bear based on this analysis.

Sample name	Species	Brown bear clade	Location	Coverage	SRA	Study
OFS01	*U. arctos*	I	Östanvik, Furudal, Sweden	26.93	SRR935590, SRR935593-4, SRR935626	[169]
Adm2	*U. arctos*	II	Admiralty Island, Alaska, USA	60.39	SRR518710-1	[189]
GRZ100	*U. arctos*	IIIA	Kenai Peninsula, Alaska, USA	30.93	SRR518712-3	[189]
SJS01	*U. arctos*	IIIA	Slakka, Jokkmokk, Sweden	16.34	SRR935591, SRR935625, SRR935627	[169]
GT1	*U. arctos*	IIIB	Alaska, USA	16.11	SRR7758718	[272]
GP01	*U. arctos*	IV	Montana, USA	20.49	SRR935609, SRR935616-7, SRR941811, SRR941814	[169]
Panda	*A. melanoleuca*	NA	Chengdu, China	15.78	SRR019741-50, SRR019778-84, SRR019835-9, SRR019843, SRR019845-8, SRR019851-5, SRR019857-62, SRR019871-3, SRR019875-8	[164]

Table 2.6: Bears used for D-statistic analysis. Brown bears and giant panda samples used to look for signs of admixture between brown bears and the giant short-faced bears.

2.4 Results

2.4.1 Genome assemblies and annotation

2.4.1.1 *de novo* genome assembly of spectacled bear

We assembled a nuclear genome for the spectacled bear using a combination of short-read Illumina data, proximity ligation Hi-C data, and long read ONT data (Table 2.7). The scaffold N50 of the TreOrn1.0 is 21.11 Mb and a BUSCO score of 93.5% [252] (Table 2.8). This genome represents the only nuclear assembly for a tremarctine bear.

Assembly version	Meraculous	HiRise	PB Jelly	Pilon iteration 1	Pilon iteration 2
Assembly step	shotgun assembly	scaffolding	gap filling	error correcting	error correcting
Input data	Illumina shotgun	Chicago	ONT	Illumina shotgun	Illumina shotgun
Genome length (bp)	2,086,202,895	2,215,996,330	2,232,824,675	2,232,974,266	2,232,736,973
Contig N50 (kb)	210.1	20.3	21.8	23.2	23.5
Contig L50	31,094	30,649	28,724	27,028	26,657
Scaffold N50 (Mb)	0.0261	20.92	21.12	21.12	21.12
Scaffold L50	23,560	29	29	29	29
# of gaps	41,345	171,229	162,727	153,671	151,388
# of Ns	1,438,142	131,322,142	125,625,584	124,345,965	123,688,823
% of Ns in genome	0.00198	5.93	5.62	5.57	5.54

Table 2.7: Genome metrics of spectacled bear assembly stages. The metrics of the spectacled bear genome at different stages of the assembly process. The final column represents the final assembly, TreOrn1.0. We saw a marked improvement in N50 as a result of scaffolding with HiRise. Gap filling with PB Jelly notably decreased the numbers of Ns and strings of Ns. As a result of this conversion of Ns to useful sequence, the number of Illumina reads that mapped to the assembly increased.

Complete BUSCOs	3839 (93.5%)
Complete and single-copy BUSCOs	3826 (93.2%)
Complete and duplicated BUSCOs	13 (0.3%)
Fragmented BUSCOs	122 (3.0%)
Missing BUSCOs	143 (3.5%

Table 2.8: Benchmarking Universal Single-Copy Orthologs (BUSCO) gene completeness score. The results of running BUSCO [252] on the TreOrn1.0 genome using the human gene set (n=4,104).

2.4.1.2 Evidence driven annotation of spectacled bear genome

We also generated a two types of RNA sequencing data; long transcript cDNA reads using long-read technology with ONT and traditional cDNA sequencing to generate short-reads. We performed an *ab initio* genome annotation using Augustus [264], predicting a set of 19,289 genes. From the predicted genes, we were able to annotate 15,504.

2.4.1.3 Modern samples nuclear data processing

We had three nuclear datasets for spectacled bears, two publicly downloaded (Chaparri and Nobody) [144] and our high coverage dataset (Mischief). We also downloaded shotgun datasets for six brown bears, two polar bears, one giant panda, one American black bear, one Asiatic black bear, one sloth bear, and one sun bear (Tables 2.4, 2.5, 2.6). We removed adapters [262] and mapped reads to the TreOrn1.0 genome using the bwa mem algorithm [159], filtering for map quality of 30, removing alignments that are not primary, or have an unmapped mate. We removed PCR duplicates using samtools rmdup [158] and performed realignment

around insertions and deletions was performed using GATK Realigner Target Creator and Indel Realignment [180]. The average nuclear coverage for Mischief was 45.97X, for Chaparri was 11.12X, and 10.81X for Nobody.

2.4.1.4 Ancient samples nuclear data processing

We obtained three bone samples from the Yukon government identified as giant short-faced bears (Table 2.1) from the late Pleistocene (27-20 kya). The sample locations are shown relative to other identified giant short-faced bear remains (Figure 2.7). We aligned adapter trimmed and merged shotgun data from our three giant short-faced bear samples to the Tre-Orn1.0 genome using bwa aln, using parameters optimized for ancient DNA [162]. We filtered all alignments for quality, unique mapping, duplicates, and reads shorter than 35bp. The average nuclear coverage for Canyon was 26.01X, for Hester was 1.82X, and 1.66X for Ophir.

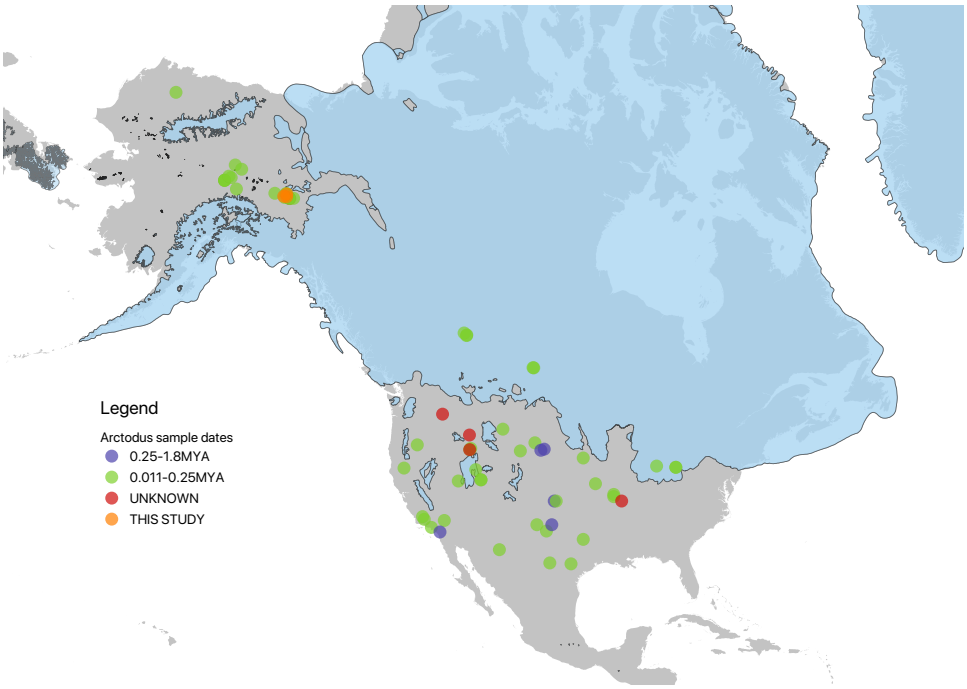

Figure 2.7: The distribution and age of giant-short faced bear fossil ages. Samples used in this study shown orange. Coordinate data downloaded from FAUNMAP [96]. Blue depicts the peak extent of the ice sheets during the Last Glacial Maximum 21,000 years ago downloaded from PaleoMaps [72].

2.4.1.5 Mitochondrial genome assemblies for ursids

We assembled a mitochondrial genome for Mischief the spectacled bear with Unicycler [298] in hybrid mode, incorporating long read and short read data. We edited the sequence in the Hyper Variable Region 2 (HVR2) based on coverage across the region, and confirmed our results by running a reference based mitochondrial assembler (mia) [98]. We observed no changes in sequence in the outputted assembly, and no longer observed a dip in coverage across HVR2. Thus we used this sequence as the final mitochondrial genome for Mischief. We generated mitochondrial assemblies for Nobody and Chaparri using our Mischief sequence as the

reference with mia [98]. We assembled mitochondrial sequences for Hester, Ophir and Canyon using shotgun data, with mia [98], and ancient DNA substitution matrix, and the *Arctodus simus* mitochondrial genome as the reference (NC011116.1) [142].

We downloaded a publicly available dataset of mitochondrial enriched *Arctotherium* reads [190], which we assembled with mia [98] using the publicly available spectacled bear (FM177764.1) [142] mitochondrial sequence, using an ancient DNA substitution matrix to account for deamination. We also dowloaded publicly available mitochondrial sequences for a sun bear, sloth bear, cave bear, Asiatic black bear, giant panda, brown bear, polar bear, spectacled bear, ABC island brown bear, and American black bear to represent the diversity of ursid species.

2.4.2 Phylogenetics of ursids

2.4.3 Mitochondrial phylogeny of ursids

We constructed a maximum likelihood and a Bayesian mitochondrial phylogeny for all eight extant bear species, in addition to three extinct bear species: the cave bear, *Arctotherium*, and the giant short-faced bear (Figure 2.8). We observe that *Arctotherium* clusters with to the spectacled bears, not the giant short-faced bears, confirmation of results of prior research [190]. This clustering indicates that the large size of the *Arctotherium* and giant short-faced bear lineages was the result of convergent evolution, not shared evolution.

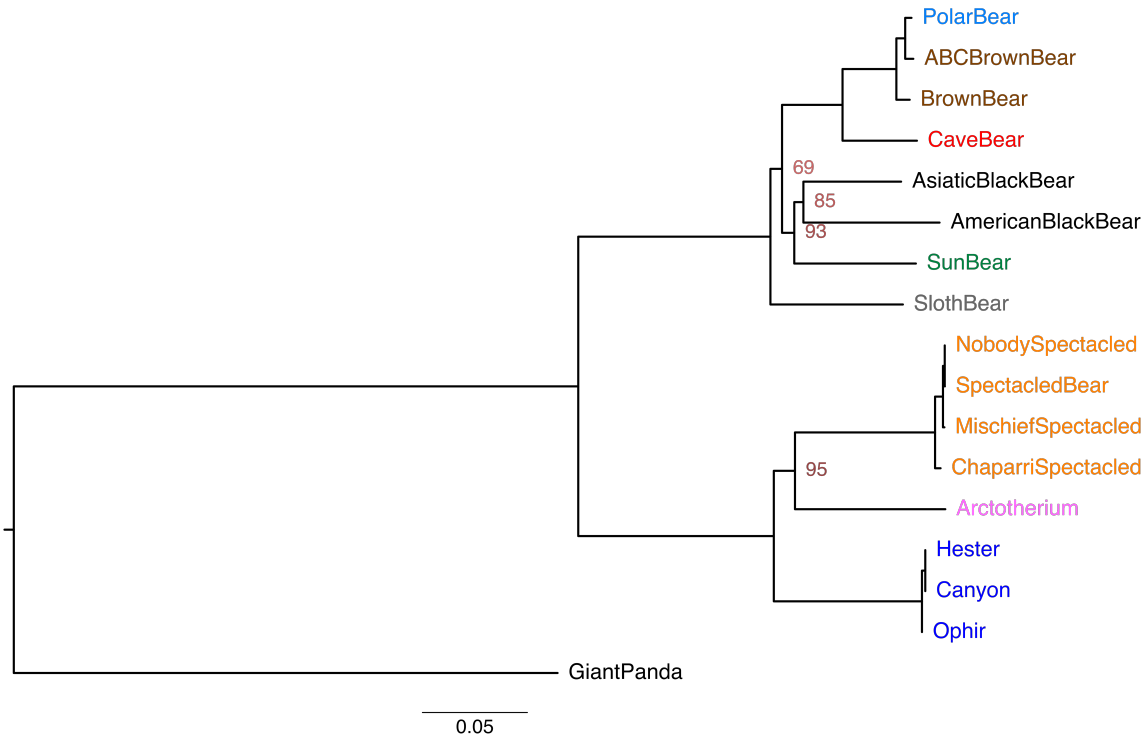

Figure 2.8: Mitochondrial phylogeny of eight extant bear species and three extinct bear species. Phylogeny constructed using relaxed filters (3X, 66% consensus) on assembled mitochondria using both Maximum Likelihood (ML) and Bayesian methods [263]. Bootstrap support values for ML are based on running one hundred bootstrap replicates, where only values below 96 are reported in red. All posterior support values are greater than 0.99, and not reported. The *Arctotherium* mitochondrial sequences cluster together and fall sister to the spectacled bear clade, and not the giant short-faced bear, as presumed based on morphology.

2.4.4 Nuclear divergence time of ursids

Using our high coverage giant short-faced bear and spectacled bear, we sought to date the divergence time of the spectacled bear and the giant-short faced bear. We identified a set of single copy orthologous coding sequences based on the annotations of the giant panda and polar bear. We identified 13,713 genes in this manner, and generated fasta sequences for

each of the eight extant bear species and the giant short-faced bear (Table 2.4) using only the four fold degenerate codon positions, for at total of 3,415,480 bases. We estimated divergence times among the bear species using MCMCTree [302] under an independent clock model with one fossil calibration and one tip date for the giant short-faced bear (Figure 2.9). We found that the giant short-faced bear and spectacled bear diverged roughly 5.5 million years ago. The tremactine bears diverged from the ursine bears roughly 13.4 million years ago.

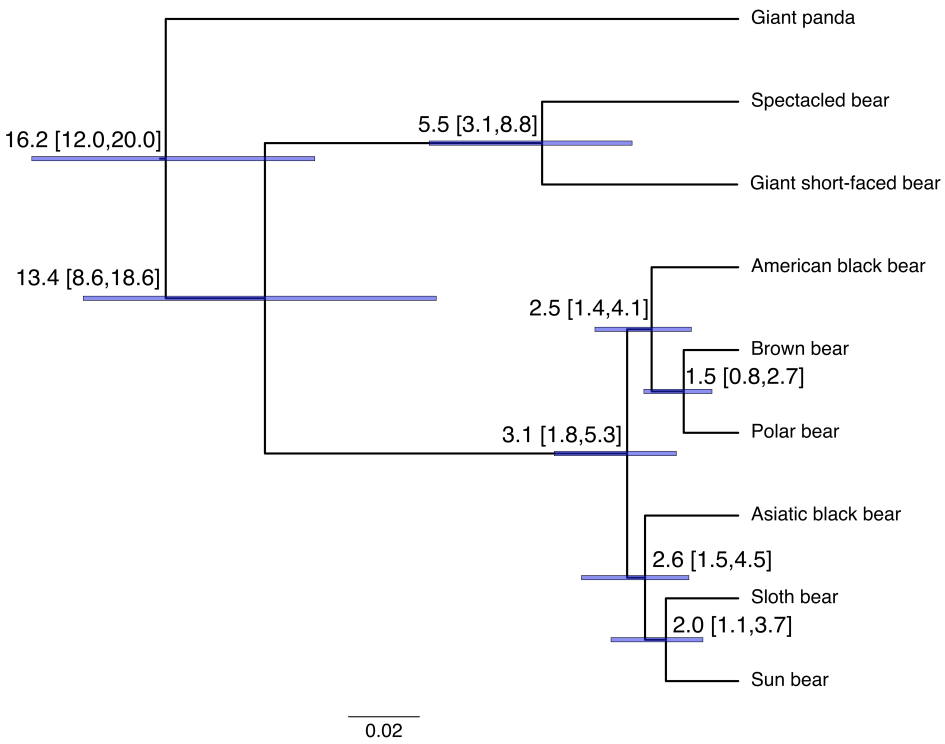

Figure 2.9: Phylogenetic tree and divergence times of eight extant bear species and the extinct giant short-faced bear. The phylogeny and divergence time estimation was based on four fold degenerate sites across 13,713 single copy orthologs. Time was calibrated at the root and based on the sample age of the giant short-faced bear. The giant panda served as the outgroup. Branch lengths represent time before present (million years ago (Mya)). The mean age of each node is shown, with 95% credibility intervals in parentheses and depicted as purple bars around each node.

2.4.5 Demographics of the spectacled bear and giant short-faced bear

2.4.5.1 Modeling the population history of tremarctine bears

Using the TreOrn1.0 genome, we sought to model the population history of the spectacled bear. We ran the pairwise sequentially Markovian coalescent (PSMC) model [163] for

three spectacled bears, as well as a downsampled version of our high coverage spectacled bear to roughly the same coverage as Nobody and Chaparri ($\sim$10X) (Figure 2.10). We also performed one hundred replicate bootstraps for each individual per the software instructions (Figure 2.11). We observe that the three mid coverage samples (Downsampled Mischief, Nobody, and Chaparri) underestimate the effective population size for the oldest timepoints in the PSMC model compared to the non-downsampled dataset for Mischief. PSMC is recommended to be run on samples with at least 20X coverage. The lower coverage samples do not recover all of the heterozygous sites across the genome due to their lack of sampling depth. For the regions that are used to infer the oldest coalescent events in the PSMC model, we observe smaller estimated effective population sizes in the lower coverage samples due to this lack of depth.

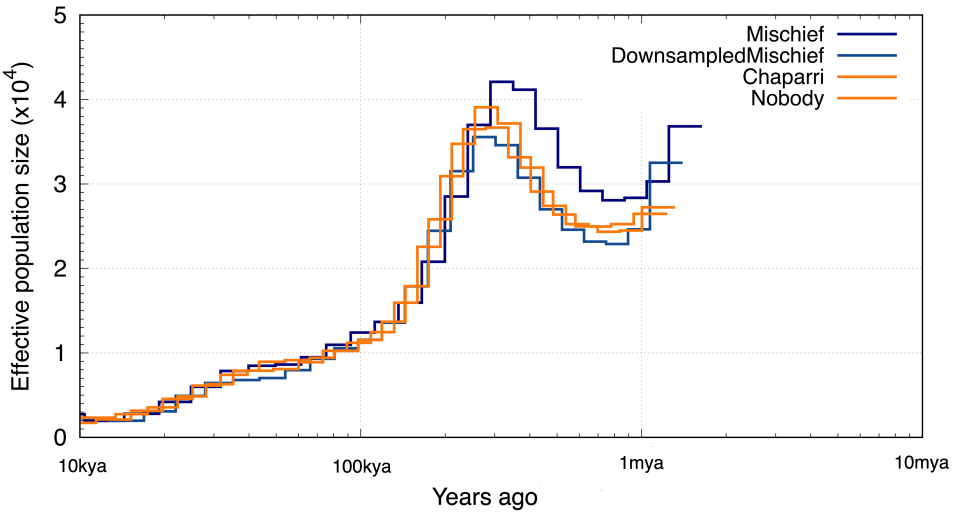

Figure 2.10: The demographic history of the spectacled bear. The inferred changes in effective population size (N_e) over time using the PSMC model [163] for three spectacled bears, Mischief (the reference genome sample, 46X), Nobody, and Chaparri. We also modeled the effective population size for a subsampled Mischief (10X), such that it had roughly the same coverage as Nobody (10X) and Chaparri (11X). We assume a generation time of 6 years and a per generation mutation rate of 0.6e-8 [144] per bp per generation. We note that the lower coverage samples show a different history than the high coverage bear. PSMC is recommended to to be run on samples with coverage greater than 20X.

The history of the spectacled bear is one of a decreasing population size. The demographic history of the spectacled bear as inferred by the PSMC model, indicates that the species was most successful roughly 450 kya, during the end of a glacial period, where the species had its peak effective population size. After which, the spectacled bear undergoes a steady decline in effective population size in response to the continued drastic climatic fluctuations during the middle Pleistocene. By the peak warm temperatures of the last interglacial maximum $\sim$130 kya, the spectacled bear population had significantly shrunk in size, as their preferred forest habitat shrunk in size [111]. The final decrease in effective population size at the end of the

Pleistocene epoch, 30-10 kya coincides with the late Pleistocene extinctions.

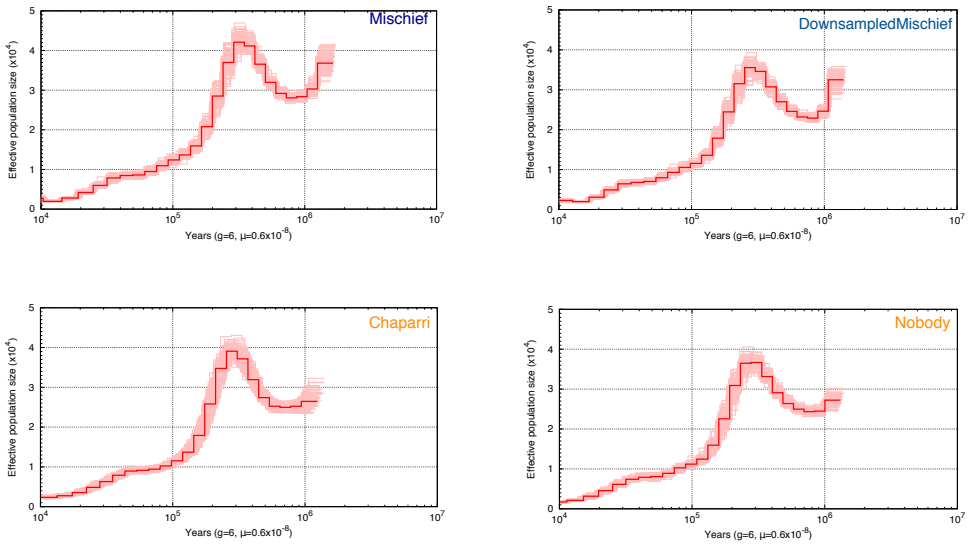

Figure 2.11: Bootstrap replicate PSMC plots for three spectacled bears, Mischief (the reference genome sample, 46X), Nobody, and Chaparri. We ran one hundred bootstrap replicates for each spectacled bear. We also modeled the effective population size for a subsampled Mischief dataset, such that it had roughly the same coverage as Nobody (10X) and Chaparri (11X). We assume a generation time of 6 years and a per generation mutation rate of 0.6e-8 [144] per bp per generation.

We used the high coverage genome of the giant short-faced bear to model its past effective population size. Using the spectacled bear genome as the reference, we ran the (PSMC) model [163] for the our high coverage giant short-faced bear, Canyon (Figure 2.12), rescaling to account for the age of the sample. We also ran ten bootstrap replicates where we rescaled to account for sample age (Figure 2.13).

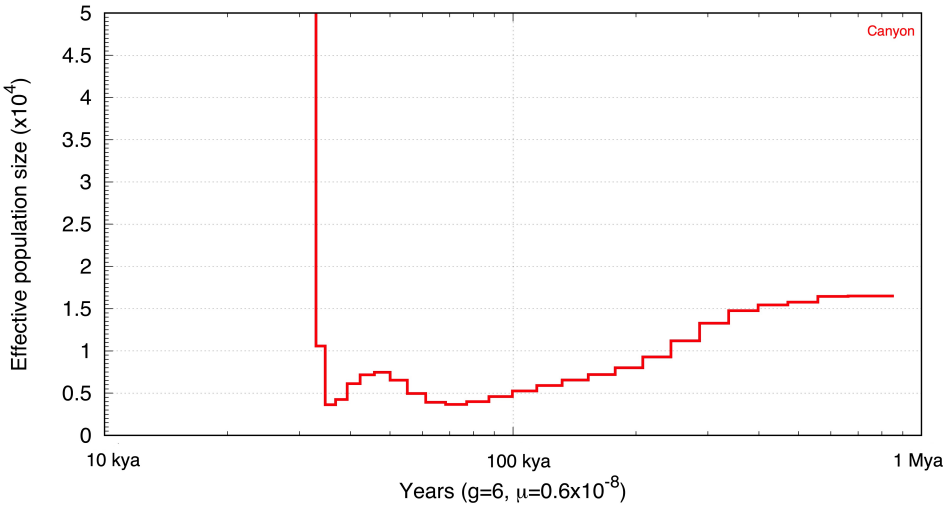

Figure 2.12: The demographic history of the giant short-faced bear. We inferred changes in effective population size (N_e) over time using the PSMC model [163]. We assume a generation time of 6 years and a per generation mutation rate of 0.6e-8 [144] per bp per generation. To account for the age of our sample, we rescaled the time axis. The most recent time points in the PSMC model, which are known to be less accurate, are thus the same timepoints which show a significant spike in N_e estimates, and can be ignored.

The demographic history of the giant short-faced bear based on the PSMC model shows a species with a long history at a small effective population size. Steady declines in effective population size throughout most of its history were followed by a slight rise in numbers roughly 60kya. By 48kya, the start of main isotopic stage 3 interstadial (MIS 3), changes in the climate in Yukon resulted in the expansion of forests, a habitat that would have been poorly suited for the scavenging giant short-faced bear. Continued regional warming from 39-34kya meant an increase in forest cover [13]. By which time, many of the herbivorous megafauna prey of the giant short-faced bear were experiencing population declines, which would have been a main contribution behind the local extinction of the giant short-faced bear in Beringia by 22kya.

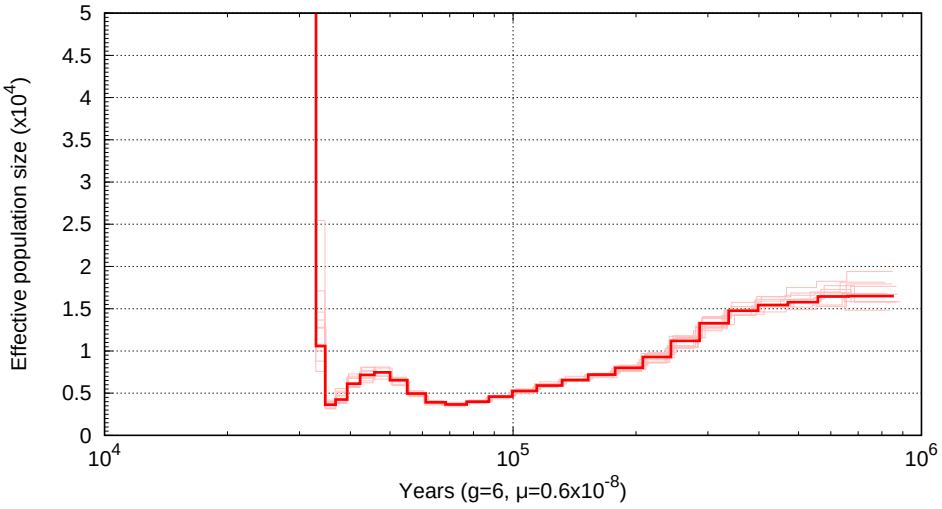

Figure 2.13: Bootstrap replicates of the demographic history of the giant short-faced bear. We ran ten bootstrap replicates of the inferred changes in effective population size (N_e) over time using the PSMC model [163], rescaling to account for the age of the sample (29,242 calBP). We assume a generation time of 6 years and a per generation mutation rate of 0.6e-8 [144] per bp per generation. We observe very few differences in the PSMC model between bootstrap replicates.

2.4.5.2 Pairwise diversity of bears

We used datasets for two captive spectacled bears, Chaparri and Mischief, as well as two brown bears and two polar bears to compare levels of pairwise diversity among bear species. We adapter trimmed reads for the two spectacled bears, two brown bears, and two polar bears using SeqPrep2 [262], and mapped filtered reads to the TreOrn1.0 genome using bwa [159]. We performed realignments around insertions and deletions using GATK Realigner Target Creator and Indel Realignment [180]. We restricted further analyses to scaffolds larger than 1Mbp in size.

We generated pseudohaploid fasta sequences by randomly selecting a base at each

position for each of the three pairs of bear species. We then measured the number of pairwise differences in non-overlapping 100kb windows for each bear species (Figure 2.14). We observe that spectacled bears have very low levels of genetic variation, on the same order as polar bears, albeit with a slightly wider distribution. In comparison, brown bears have a broader distribution of pairwise differences, and thus greater genetic variation within the species.

To assess individual levels of diversity we estimated the genome wide heterozygosity of our high coverage spectacled bear, Mischief. The spectacled bear carries roughly 3 heterozygous sites per 10,000 base pairs, on the same order of magnitude as the inbred Florida panthers [237] (Table 2.9). The heterozygosity levels of the spectacled bear are as low as those of a well-known inbred and small population of pumas, indicating the long history of population size decline has reduced the diversity of the spectacled bears to an alarmingly low degree.

Sample name	Species	Nuclear genome coverage	Heterozygosity
Mischief	*T. ornatus*	45X	0.000262
Chaparri	*T. ornatus*	11X	0.000273
Nobody	*T. ornatus*	10X	0.000289
Canyon	*A. simus*	26X	0.000150

Table 2.9: Genome-wide heterozygosity for Tremarctine bears in our study. The nuclear heterozygosity values for the three spectacled bears and one high coverage giant short-faced bear used in this study calculated using the pileup method described previously [237].

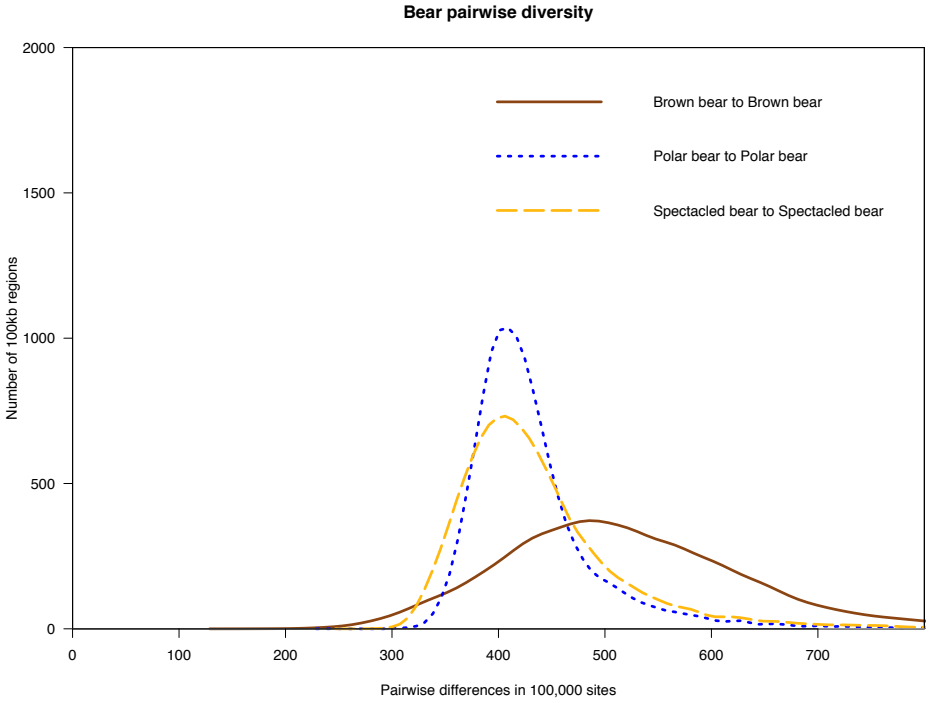

Figure 2.14: Number of pairwise differences per 100,000 bases measured in 100kb windows between three pairs of bear species. We see that the level of genetic diversity is very low for spectacled bears, similar to that observed for polar bears.

Our PSMC results for the recent history of the spectacled bear show a similar trend to the genome wide heterozygosity measures for our three spectacled bears. All three spectacled bears have low levels of genome wide heterozygosity, on the order of 3 heterozygous sites per 10,000 sites. The dramatic decrease in effective population size of the spectacled bear has thus resulted in a species with very little genetic diversity.

2.4.5.3 Archaic admixture between brown bears and giant short-faced bears

Based on the known overlap in time and range between giant short-faced bears and brown bears, we looked for signals of archaic admixture between the species (Table 2.6). We generated pseudohaploid fasta sequences for brown bears, giant short-faced bears, and the giant panda, using the spectacled bear genome as the reference. In our species tree, bear 1 and bear 2 were brown bears, bear 3 was the introgressing giant short-faced bear, and bear 4 was the giant panda, which served as the outgroup. We used the D-statistic to look for any incongruencies in the species tree in shared derived alleles between the giant short-faced bear and one of the brown bears. The D-statistic only considers sites where one of the brown bears and the giant short-faced bear carry the derived allele while the panda and the other brown bear carry the ancestral allele. We looked only at transversions to exclude any effects of ancient DNA damage.

We observe no signs of an excess of shared derived alleles between the giant short-faced bear and any brown bear (Figures 2.15, 2.16). The estimated divergence time for the giant short-faced bear and the ursine bears determined from MCMCTree of roughly 13 million years is likely to exclude the ability for these two species to produce fertile offspring. Thus the genome of the giant short-faced bear does not live on in extant brown bears.

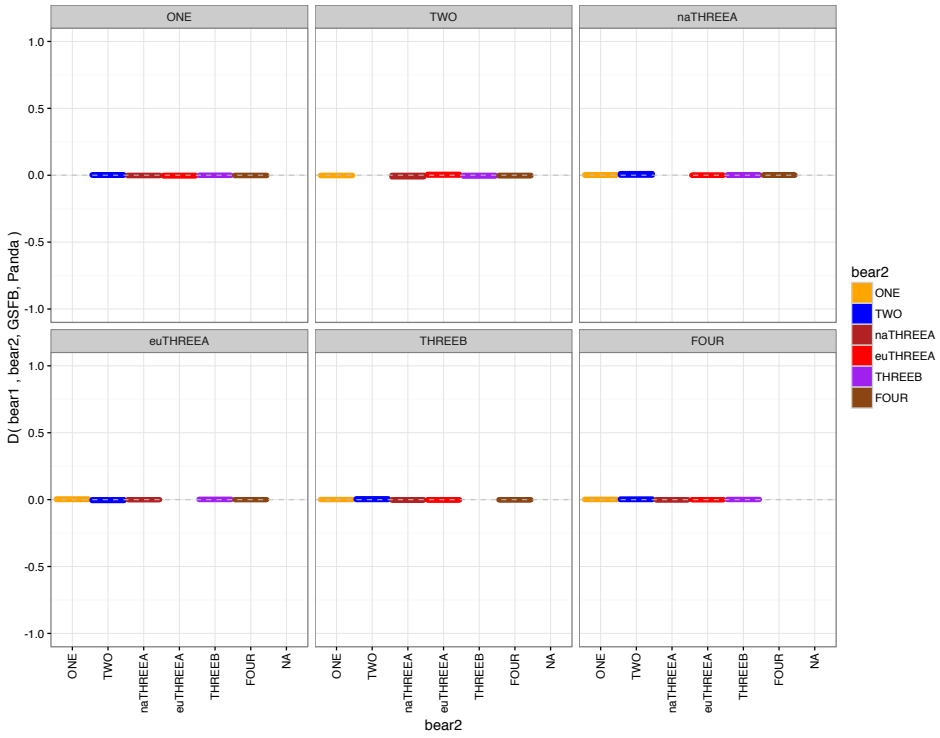

Figure 2.15: D-statistic analysis to look for signs of shared ancestry between the giant short-faced bear and the brown bear. Bear 1 is identified at the top of each panel, and comparison with a different bear 2 is shown along the x-axis colored by bear 2. Positive D-statistic scores indicate bear 1 has greater allele sharing with the polar bear than bear 2. Negative D-statistic scores indicate bear 2 has greater allele sharing with polar bears than bear 1. Significance was determined based on weighted block jackknife calculations for non overlapping 100kb windows (Figure 2.16). Bear 2 names depict the clade of each brown bear, where naTHREEA refers to a North American Clade 3A bear and euTHREEA refers to a European Clade 3A bear. GSFB refers to the giant short-faced bear. Details on the bears used for this analysis are provided in Table 2.6.

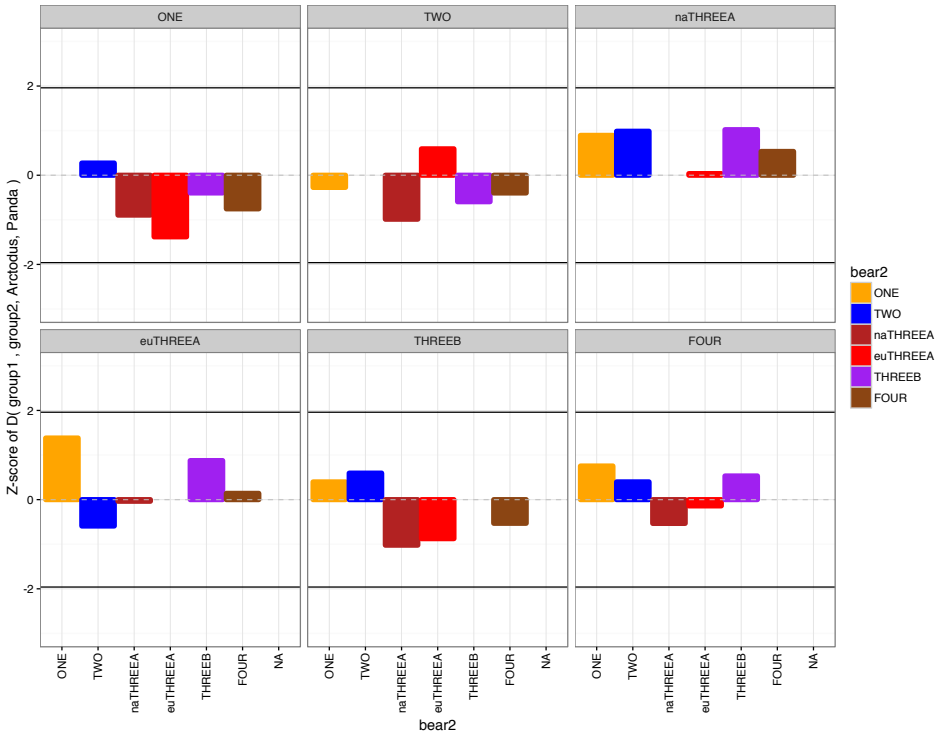

Figure 2.16: Z-scores for the D-statistic analysis looking for signs of shared ancestry between the giant short-faced bear and the brown bear. Z-scores were calculated from the standard error estimation determined by the weighted block jackknife of non-overlapping 100kb windows. Bear 1 is identified at the top of each panel, and comparison with a different bear 2 is shown along the x-axis colored by bear 2. Positive D-statistic scores indicate bear 1 has greater allele sharing with the polar bear than bear 2. Negative D-statistic scores indicate bear 2 has greater allele sharing with polar bears than bear 1. Scores of above 1.96 or below -1.96 indicate significant associated D-statistic values (Figure 2.15).

2.5 Discussion and next steps

We present a genome assembly for the spectacled bear, the sole surviving bear of the tremactine bears. We also generate three genomes for an extinct tremarctine bear, the giant short-faced bear, using the spectacled bear as the reference. The tremarctine genome means a

better reference genome for analysis for other ancient tremarctine bear samples, as opposed to the polar bear genome which had been used previously [144]. A new reference genome for a tremarctine bear species provides a closer related species to which ancient DNA from the family can be mapped, which given the rapid advancements in ancient DNA recovery, is particularly important to recover more information from sequencing [92, 88]. Continued improvement of DNA retrieval from ancient samples means having a reference genome for the last extant member of the tremarctines will allow for greater data retrieval than with a distant reference genome. Reference bias is a known problem in mapping and downstream analyses, especially in ancient samples [100], and thus having a genome assembly for the tremarctine lineage will minimize these effects for other tremarctine samples.

A reference genome and high-coverage shotgun dataset for the only extant subfamily previously without a nuclear genome will also allow for more in-depth studies of intra-familial diversity. Most ursid studies use the giant panda genome as the outgroup to ursine bears. However, the giant panda is more than 16.2 million years diverged from these species, while the spectacled bear is 5.5 million years diverged [147]. A more closely related outgroup will allow for greater assessment of ursine specialization, and possibly uncover more gene flow events within the ursine lineage.

Using high coverage genomes, we modeled demographic history of the spectacled bear and the giant short-faced bear, two species characterized by steadily declining populations. Today, spectacled bears are seen at a range of altitudes, though they prefer the higher mountainous cloud forests. Thus, the disappearance of forests during the middle Pleistocene considerably impacted their range and population size during the middle and late Pleistocene.

In more recent times, the preferred forest ecosystem of the spectacled bear would have likely afforded it protection from early human settlers, allowing it to survive to present day, though their habitat is currently at risk due to mining, farming, illegal killing, and climate change.

The tremarctine bears, bar the spectacled bear, were unable to survive into the Holocene. Early bear migrants into South America during GABI were able to take advantage of the initially slim competition and plentiful prey. The tremarctine bears of South America, *Arctotherium* spp., sustained five species on the continent throughout the Pleistocene, including the largest bear species recorded. While all five *Arctotherium* species were extirpated by the end of the Pleistocene, the lineage evolved to become increasingly smaller and more omnivorous, which allowed it to survive in South America with ever increasing competition for prey from other large carnivores [258].

The carnivorous niche of the giant short-faced bear in North America, made it highly susceptible to the end Pleistocene loss of megafauna. However, in comparison to *Arctotherium* sp., the giant short-faced bear went extinct only at the very end of the Pleistocene, implying a different set of pressures for large bears in South America compared to North America. While the giant short-faced bear is thought to capable of omnivory and diet shift based on cranial structure [83], it was likely not adept to such a significant niche shift, especially given the energy costs of such a large animal.

The giant short-faced bear highly specialized for its carnivorous niche, its large size and craniodental anatomy as evidence, but this specialization also made it susceptible to changes in its environment [282]. The mass extinctions during the end Pleistocene would have dramatically reduced the availability of large prey for the giant short-faced bear. Competition

from brown bears was not the primary reason for the demise of the giant short-faced bear, though brown bears did inhabit the giant short-faced bear's dietary niche after their extinction in Beringia [177]. In addition, to its inability to hibernate in comparison to the co-occurring brown bears [177], the giant short-faced bear was better suited to the open woodlands of the earlier Pleistocene, where it could use its size to forage or scavenge; the closed forests of the late Pleistocene would have hampered its feeding behavior. By the peak of the last glacial maximum roughly 21,000 years ago, giant short-faced bears had disappeared from the Yukon area. The youngest sample from the region being our Ophir sample, dated to 22,693 calBC 2.1.

The giant short-faced bears showed declines in population size similar to other megafauna in Beringia at the time [16, 52, 248, 53]. As a next step, we aim to investigate whether any lineage specific adaptations better suited the spectacled bear to surviving into the Holocene than the giant short-faced bear. Or was sheer size to blame for the demise of the giant short-faced bear?

While our giant short-faced bear samples represent the Beringian population, south of the Cordilleran and Laurentide Ice sheets, giant short-faced bears survived until the late Pleistocene, the youngest specimen uncovered being roughly 11 kya [242]. Data from a high quality sample from the lower contiguous 48 states could uncover distinguishing traits of giant short-faced bear populations, such as diet north and south of the ice. Future samples may shed some light on how the bears in lower North America survived for another ten thousand years.

Chapter 3

Sequencing a symbol: the genome of the California grizzly bear

[Manuscript in preparation]

Nedda F. Saremi[1], Peter D. Heintzman[2], Joshua D. Kapp[2], Nicholas Maurer[1], Katherine L. Moon[3], Alexis M. Mychajliw[4], Peter S. Alagona[5], Joanna L. Kelley[6], Richard E. Green[1], and Beth Shapiro[1,3]

Affiliations

[1] Department of Biomolecular Engineering and Bioinformatics, UC Santa Cruz

[2] Department of Ecology and Evolutionary Biology, UC Santa Cruz

[3] Howard Hughes Medical Institute

[4] La Brea Tar Pits and Museum

[5] Department of Environmental Studies, UC Santa Barbara

[6] School of Biological Sciences, Washington State University

3.1 Abstract

North American brown bears, also known as grizzly bears, are endangered today due mainly to depredation by humans. Their range once extended from Alaska to Mexico and from the Great Plains to California, where they were known as the California golden bearthe iconic symbol of the state. Very little is known about the California grizzly bear, including its evolutionary relationships to other grizzly bears. Today, the human relationship with large predators has changed, as we increasingly recognize the role of predators in maintaining biodiverse and stable ecosystems. Working with collaborators at the Museum of Vertebrate Biology (MVZ) at UC Berkeley, the California Academy of Sciences (CAS), and the University of California Santa Barbara we generated low coverage (5X) nuclear genomic data from two of the last California grizzly bears, one of whom served as the model for the bear on the California state flag. By performing population genomics analyses we can answer questions about the history and uniqueness of the extinct California golden bear and its relationship to modern brown bears.

3.2 Background

3.2.1 Brown bears

The brown bear is one of the most widely distributed terrestrial mammals, encompassing the Holarctic. The range of the brown bear is the largest range of any bear species

(Figure 3.1. Its wide distribution has meant that it has been well studied, especially in regards to brown bear phylogeography [106, 58]. Due to variation in color and morphology, brown bears have been divided into as many as 90 subspecies in North American alone, a case of taxonomic over-splitting [184].

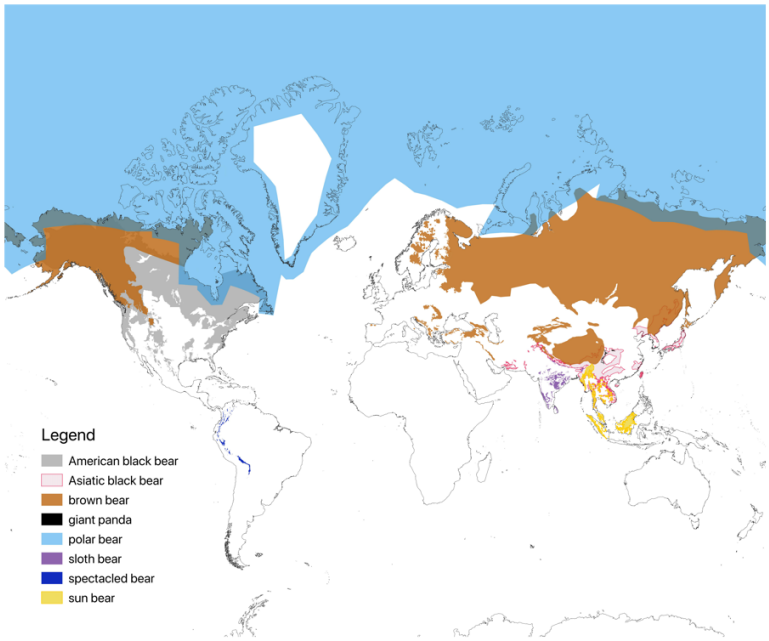

Figure 3.1: The geographical ranges of all eight extant bears. Extant range data is from the IUCN Red List of Threatened Species [119]. Base map from Natural Earth.

Today, complete mitochondrial DNA analysis of brown bears divides the bears into six major clades [108] (Figure 3.2). The major brown bear matrilineal clades show strong phylogeography. Two clades subdivide Europe and Asia, clade 1 in Western Europe and clade 3 in Eastern Europe and the Middle East [271]. These two clades show a geography that can be explained by the European paleoclimate, where clades existed in southern local ice age refugia, and expanded north after glacial periods [107].

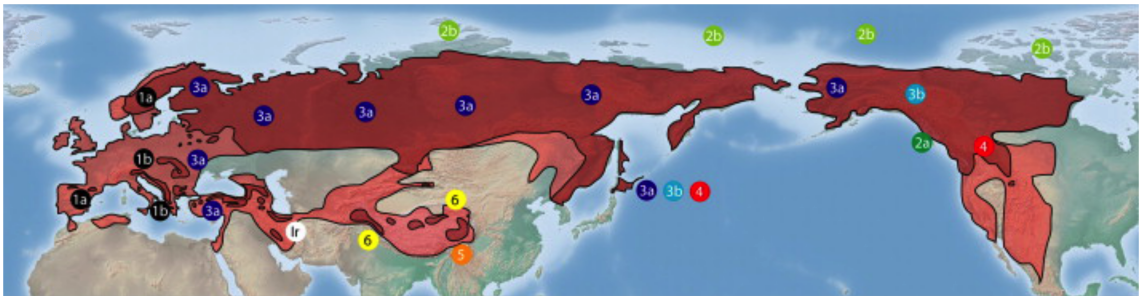

Figure 3.2: The current (dark red) and past (light red) global ranges of brown bears. Geographical distributions of brown bear clades. Brown bears have been divided into six major clades based on mitochondrial genomes. The distribution of these clades shows the strong geographic pattern of variation. Figure taken from [58].

Two less well-studied clades, 5 and 6, are endemic to Southern Asia. Complete mitogenomes only exist from brown bears from Tibetan Plateau and the Himalayas to represent these two clades [108, 148]. Clade 5 was discovered in the Tibetan Plateau and sits basal to other Eastern brown bear lineages [108]. Meanwhile, a complete mitochondrial clade 6 sequences falls basal to all other brown bear clades [148]. These basal lineages indicate a species origin in Asia, however greater sampling in the region is needed to confirm such.

In North America, strong phylogeography defines the distribution of brown bears on the continent [288]. Three clades constitute the diversity of North American brown bears based on mitochondria, clades 2, 3 and 4. While all three of these clades were present in Beringia North American at least 35,000 years ago (35 kya), some subclades have been replaced and others have migrated [16]. The North American clade which has made its way south from Beringia is clade 4. Clade 4 currently inhabits the contiguous United States and southern Canada. However, clade 4 bears were present in Beringia prior, with the oldest sample in North America thus

far for this clade dated to 36,000 years old from Eastern Beringia [16]. Clade 4 brown bears did make it south of the Laurentide and Cordilleran ice sheets prior to their coalescence into one during the Last Glacial Maximum (LGM). A fossil specimen belonging to the clade from Edmonton, Canada dated to 26 kya confirmed such [176].

North of the ice sheets in North America, local extinctions and postglacial migrations of brown bears have resulted in a largely allopatric distribution of bears. Samples belonging to subclade 3c brown bears have been dated to older than 50,000 years in the interior region of Alaska [16]. No brown bear fossils dated to between 35-21 kya have thus far been found in Eastern Beringia [16]. Presumably, local brown bear extinctions gave way to new colonization, such that by the peak of the LGM 21 kya, a new subclade, clade 3b, entered the continent by way of the Bering land bridge [16]. By 13 kya, retreat of the ice sheets enabled clade 3b bears to expand further east into Alaska and Canada [16]. Meanwhile, clade 3a brown bears managed to cross the Bering land bridge from Eurasia before its disappearance 11 kya [74]. Today, clade 3a inhabits western Alaska [16].

3.2.2 Polar bears

The history of clade 2 bears pertains to two bear species, the brown bear and the polar bear. Mitochondrial analysis has shown that brown bears are paraphyletic to polar bears [49]; polar bear mitochondria fall within the diversity of brown bear mitochondria. Based on their mitochondria, brown bears from subclade 2a are more closely related to polar bears, subclade 2b, than other brown bears. While these two species are estimated to have diverged from one another roughly half million years ago [103, 169], subsequent admixture events commonly

occurred when paleoclimatic conditions resulted in range overlaps [33, 34].

The brown bears to which the mitochondria of modern polar bears is most closely related are those found in the Admiralty, Baranof and Chicagof (ABC) islands of southeastern Alaska [249]. Nuclear genomes showed that while all North American brown bears carry 3 to 8 percent polar bear ancestry, ABC island bears have the highest levels of polar bear ancestry of all modern brown bears [33, 36, 169]. The model to explain polar bear ancestry in present day ABC island bears proposes that during a glacial period, the ice sheets extended southward. Polar bears would have populated the glaciated coast for easy food access. Upon retreat of the ice sheets, polar bears were stranded on the islands [33]. The area became suitable habitat for brown bears, who are known to migrate between the mainland and nearby islands [205]. The predominantly male migrations from the mainland meant heightened levels of polar bear ancestry were retained on the X chromosome compared to the autosomes, and explains the shared matrilieanage between ABC island brown bears and polar bears [33, 36]. It seems that island population conversion was not an isolated event between brown and polar bears, as ancient brown bears from Ireland and Kunashir have also been found to have polar bear ancestry [34]. Thus glacial refugia and retreat have played an important role in the history of brown bears.

3.2.3 California grizzly bears

Globally, brown bears have undergone a significant reduction in range since the 19th century [247] (Figure 3.3). Brown bears in North America, known as grizzly bears, have undergone a population reduction in the lower 48 states in the US and Mexico to less than 1 percent of their original size since the early 19th century [247]. At present, grizzly bears are

found in Alaska, western Canada and in the states of Wyoming, Montana, Idaho and Washington [247].

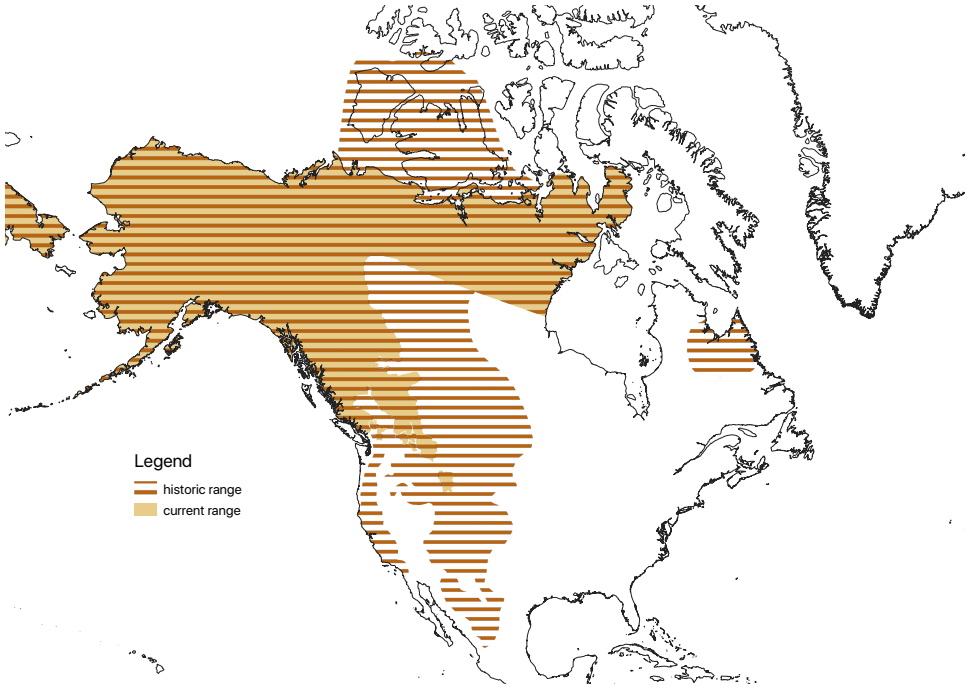

Figure 3.3: Historical versus current range of grizzly bears. Current range in light brown, historic range in striped brown. Data obtained from IUCN Red List of Threatened Species [119]. Base map from Natural Earth.

The California grizzly bear is an extinct population of brown bears that inhabited California. The last individuals of this population are believed to have died around 100 years ago, largely due to over hunting by humans. Monarch, one the last California grizzly bears, was captured in Ventura County, Southern California in 1911 as part of a publicity stunt. Monarch lived the rest of its life in captivity in zoos, with its remains preserved at both California Academy of Sciences and the Museum of Vertebrate Zoology at UC Berkeley after its death. In addition to Monarch, we sampled the remains of another California grizzly bear. Our second California grizzly came from Monterey, California from 1871 from the collection at La Brea Tar Pits and

Museum.

Clade 4 grizzly bears would have been an isolated population for some time during the Last Glacial Maximum. After making their way south by at least 26 kya, based on fossil identified in Edmonton, Canada, clade 4 brown bears would have been isolated from northern brown bear populations. Other grizzly bears in North America were restricted to north of the ice sheets until at least 13 kya when the Cordilleran and Laurentide ice sheets melted and the ice-free corridor opened [105]. However, recent research confirms a viable pathway into the lower 48 through southeastern Alaska after 17 kya [154]. The possibility of coastal migrations along Western North America may have allowed brown bears from Alaska entry into California.

Fossil remains have played an important role in brown bear research. Ancient brown bear analysis have uncovered patterns of migrations [16, 176], as well as archaic admixture events with both polar bears and cave bears [34, 14]. Here, we generate and analyze the first nuclear genomic scale data from California grizzly bears. We use fossil remains from two California grizzly bears to study their relation to modern brown bears and highlight the importance of museum specimens in revealing population histories.

3.3 Methods

3.3.1 Sampling and data generation of bear DNA

3.3.1.1 Ancient California grizzly bears

We generated resequencing data from two ancient California grizzly bear samples, Monarch and Monterey. For Monarch, we obtained bone and foot pad samples from Monarchs

remains, which are preserved at the Museum of Vertebrate Biology (MVZ) at UC Berkeley and the California Academy of Sciences (CAS), respectively. We initially gathered seven different sample types to perform an initial assessment of fragment length, endogenous content, and the the proportion of unique reads. All work was performed in a dedicated ancient DNA facility at UC Santa Cruz, following standard clean room criteria [46].

DNA was extracted and isolated from three bone fragments, two toepad samples, hair, and skin using methods optimized for the recovery of degraded DNA. For bone fragments, 100-120mg of each bone fragment was powdered using a Mixer Mill MM 400 (Retsch). Following powdering, a bleach pre-treatment was performed as outlined in [235] on one powder sample (JK378). Next, the bone powder was incubated for 24 hours at 37 °C in an extraction buffer optimized from bone (0.45M EDTA, 0.25mg/mL Proteinase K). Non-bone samples were incubated overnight at 55 °C in a hair/tissue optimized extraction buffer (1mL of 2% SDS, 10mM Tris, 2.5mM EDTA, 10mM NaCl, 5mM CaCl2, 40mM DTT, and 0.2mg/mL Proteinase K). Finally, DNA was isolated using the silica column based method outlined in [51] with a final elution of 50 µL buffer EBT (10mM Tris, 0.05% Tween-20).

Double-stranded DNA library preparation was performed using the method outlined in [186] for all extracts. Double-stranded libraries were indexed and amplified in 50 µL reactions containing 10 µL pre-amplified library, 1x AmpliTaq Gold Buffer, 2.5mM MgCl2, 0.25mM dNTPs, 0.4 µM i7 indexing primer, 0.4 µM i5 indexing primer, and 0.04 U/ µL AmpliTaq Gold Hot Start Polymerase. Libraries were amplified in a Bio-Rad T100 thermocycler using the following conditions: 94 °C for 10m, followed by 25 cycles of 94 °C for 30s, 60 °C for 30s, and 72 °C for 40s, followed by 72 °C for 7m. Post-amplified libraries were puri-

fied using a 1.5X SPRI clean. All post-amplified and cleaned libraries were quantified using a Qubit 4 (Invitrogen) and the Qubit 1X dsDNA HS assay kit (Invitrogen). Finally, libraries were visualized on a TapeStation 2200 (Agilent) using a D1000 High Sensitivity Assay (Agilent). Libraries were pooled and sequenced on a MiSeq at UC Santa Cruz (2x75bp). Foot pad and hair samples returned low endogenous DNA levels, and thus we focused on long bone samples for following work (Table 3.1).

Lab ID	Accession ID	Sample type	Mean Fragment Length (bp)	Endogenous content (%)
JK269	CAS001	Toe pad 1	38.36	30.85
JK329	CAS001	hair	56.31	0.69
JK330	CAS001	Toe pad 2	45.01	6.41
JK331	CAS001	skin	49.30	0.18
JK321	CAS001	Femur	55.66	23.38
JK322	MVZ 24537	Occipital	52.98	4.02
JK323	MVZ 24537	Tarsal	55.81	14.78
JK378 bleach	MVZ 24537	Femur	54.57	64.45

Table 3.1: Quality assessment of initial samples for Monarch. Sample ID, sample type, mean fragment length, and endogenous content for seven initial samples, and one replicate treated with bleach. Long bone samples, particularly in conjunction with bleach pre-treatment generated the highest quality libraries. Reads were mapped to the polar near genome (GCA_000687225.1)

We sampled two additional long bone samples from Monarch's remains at the Museum of Vertebrate Zoology to generate three extracts. DNA was extracted and isolated from three long bone fragments using methods optimized for the recovery of degraded DNA. First, 100-120mg of each bone fragment was powdered using a Mixer Mill MM 400 (Retsch). Following powdering, a bleach pre-treatment was performed as outlined in [235] on all powder samples. Next, the bone powder was incubated for 24 hours at 37 °C in an extraction buffer optimized from bone (0.45M EDTA, 0.25mg/mL Proteinase K). DNA was isolated using the silica column based method outlined in [51] with a final elution of 50µL buffer EBT (10mM

Tris, 0.05% Tween-20).

Double-stranded DNA library preparation was performed using the method outlined in [186] for all extracts. Double-stranded libraries were indexed and amplified in 50 µL reactions containing 10 µL pre-amplified library, 1x AmpliTaq Gold Buffer, 2.5mM MgCl2, 0.25mM dNTPs, 0.4 µM i7 indexing primer, 0.4 µM i5 indexing primer, and 0.04 U/ µL AmpliTaq Gold Hot Start Polymerase. Libraries were amplified in a Bio-Rad T100 thermocycler using the following conditions: 94C for 10m, followed by 25 cycles of 94 °C for 30s, 60 °C for 30s, and 72 °C for 40s, followed by 72 °C for 7m. Post-amplified libraries were purified using a 1.5X SPRI clean.

Single-stranded library preparation was performed using an ancient DNA optimized version [132] of the method outlined in [279] for DNA extracts from two long bones. Single-stranded libraries were indexed and amplified in 100 µL reactions containing 48 µL pre-amplified library, 50 µL AmpliTaq Gold 360 Master Mix, 1 µM i7 indexing primer, and 1 µM i5 indexing primer. Libraries were amplified in a Bio-Rad T100 thermocycler using the following conditions: 95C for 10m, followed by 15 cycles of 9 °5C for 30s, 60 °C for 30s, and 72 °C for 60s, followed by 72 °C for 7m. Post-amplified libraries were purified using a 1.2X SPRI clean.

All post-amplified and cleaned libraries were quantified using a Qubit 4 (Invitrogen) and the Qubit 1X dsDNA HS assay kit (Invitrogen). Finally, all post-amplified libraries were visualized on a TapeStation 2200 (Agilent) using a D1000 High Sensitivity Assay (Agilent). Double-stranded libraries were pooled and sequenced at UC San Francisco Center for Advanced Technology on a NovaSeq S4 (2x100bp). Single-stranded libraries were pooled and sequenced at Fulgent Genetics on an Illumina HiSeq 4000 (2x150bp).

3.3.1.2 Mitochondrial enrichment of a California grizzly bear

A hybridization enrichment was performed on the library JK321 using probes for the complete brown bear mitochondria (NC_003427.1) and the MyBaits v3 protocol (Arbor Biosciences). The library molecules were allowed to hybridize for 36 hours at 65C. Following the MyBaits v3 protocol, the enriched library was amplified in a 50 µL reaction containing 15 µL enriched library, 1X KAPA HiFi HotStart ReadyMix, 0.5 µM forward primer, and 0.5 µM reverse primer. The enriched library was amplified in a Bio-Rad T100 thermocycler using the following conditions: 98 °C for 2m, followed by 30 cycles of 98 °C for 20s, 60 °C for 30s, and 72 °C for 60s, followed by 72 °C for 5m. The post-amplified enriched library was purified using a 1.5X SPRI clean. The library was sequenced on a MiSeq at UC Santa Cruz (2x75bp).

We obtained five samples identified as California grizzly bears from the La Brea Tar Pits and Museum (Table 3.2). We sampled each, and reduced them to powder using a Mikrodismembrator. We collected 50-115 mg for DNA extraction. Prior to extraction, we incubated the bone powder at room temperature in a 0.5% bleach solution, and washed each sample three times in 1 mL of molecular grade water, to remove contaminants and increase endogenous yield [140]. We performed DNA extraction following the Dabney protocol [51], with in-house modifications by J. Kapp. We converted the resulting DNA extracts into single-stranded genomic sequencing libraries following the Santa Cruz protocol [132]. We estimated the number of cycles to amplify each library via quantatative PCR. We cleaned each amplified library with a 1.2X SPRI solution, and quantified DNA concentration using a high sensitivity Qubit Fluorometer 4 and an Agilent 5200 Fragment Analyzer. We pooled the resulting libraries and performed

preliminary sequencing on a MiSeq at UC Santa Cruz (2x75bp) to assess library quality.

Of the five samples, only one had an endogenous content high enough (LaBrea YPM MAM 7043) to allow for generation of high coverage nuclear dataset (Table 3.2). Using the already made extract, we generated eleven more libraries following the Santa Cruz protocol [132]. We again estimated the number of cycles to amplify each library via quantatative PCR, and cleaned each amplified library with a 1.2X SPRI solution. We quantified DNA concentration using a high sensitivity Qubit Fluorometer 4 and an Agilent 5200 Fragment Analyzer. We pooled the twelve libraries and sequenced them on an Illumina HiSeq 4000 (2x150bp) at Fulgent Genetics.

Lab ID	Accession ID	Locality	Sample type	Sample age	Endogenous content
SC19KLM001	LaBrea YPM MAM 7036	Alameda county	skull	1871 AD	0.14%
SC19KLM002	LaBrea YPM MAM 7043	Monterey county	skull	1871 AD	42.20%
SC19KLM003	LaBrea YPM MAM 7044	Monterey county	skull	1871 AD	0.09%
SC19KLM004	LaBrea LACM 31148	Sacramento county	pelvis and sacrum	1177 ybp	0.05%
SC19KLM005	LaBrea CALPOLY M 595	San Luis Obispo county	skull	276 ybp	0.03%

Table 3.2: Quality assessment of additional ancient brown bear samples. Sample IDs, locality, type, age, and endogenous content for five brown bear samples. We generated additional libraries for only sample LaBrea YPM MAM 7043 based on initial sequencing results. Note: 'ybp' denotes years before present.

3.3.1.3 Modern bears

Whole blood was collected from an adolescent North American male brown bear (Dodge) *Ursus arctos* in captivity at Washington State University with approval by the Washington State University Institutional Animal Care and Use Commission (IUCAC) (Protocol #03875 and #04922). Following storage at -80C in an EDTA vacutainer for several days, genomic DNA was extracted from five milliliters of whole blood using a Blood and Cell Culture

DNA Midi Kit (Qiagen catalog no. 13343) following the manufacturers recommended protocol. The resulting high molecular weight DNA was serially diluted based on Qubit dsDNA HS assay readings (Invitrogen Q32851) before input to the Chromium Genome Library Kit and Gel Bead Kit v2 (10X Genomics catalog no. PN-120258). The library produced following 10X Genomics protocol (Manual CG00043 Rev A) was assessed via Qubit assay and TapeStation D1000 screen tape (Agilent catalog no. 5067-5582) before sequencing at Fulgent Genetics on an Illumina HiSeq X Ten (2x150bp).

We also downloaded publicly available data for thirteen brown bears, 2 polar bears and one American black bear (Table 3.3).

Sample name	Species	location	sex	brown bear clade	mitochondrial coverage	nuclear coverage	ID	Study
Monarch	*U. arctos*	Ventura, California, USA	male	4	424	5.44	tbc	this study
Monterey	*U. arctos*	Monterey, California, USA	male	4	64	5.18	tbc	this study
Dodge	*U. arctos*	unknown, lower 48 USA	male	4	34	23.75	tbc	this study
Montana	*U. arctos*	Glacier Park, Montana,USA	male	4	58	20.25	SRR935609, SRR935616-7, SRR941811, SRR941814	[169]
KenaiAK	*U. arctos*	Kenai Peninsula, AK, USA	female	3a	81	31.21	SRR518712-3	[189]
DenaliAK	*U. arctos*	Denali Natl Park, AK, USA	female	3a	470	16.46	SRR830337	[33]
EAK	*U. arctos*	unknown, Eastern AK, USA	male	3b	30	16.00	SRR7758718	[272]
Finland	*U. arctos*	Ruokolahti, Finland	female	3a	363	24.71	SRR935592, SRR935595, SRR935624, SRR935628	[169]
NSweden	*U. arctos*	Slaka, Jokkmokk, Sweden	female	3a	274	16.65	SRR935591, SRR935625, SRR935627	[169]
SSweden	*U. arctos*	Furudal, Sweden	female	1	274	27.21	SRR935590, SRR935593-4, SRR935626	[169]
Adm2	*U. arctos*	Admiralty Island, AK, USA	female	2b	35	49.48	SRR518710-1	[189]
ABC06	*U. arctos*	Admiralty Island, AK, USA	female	2b	43	23.44	SRR935601, SRR935608, SRR935615, SRR935619, SRR941810, SRR941813	[169]
Bar	*U. arctos*	Baranof Island, AK, USA	male	2b	36	29.57	SRR518717	[189]
ABC01	*U. arctos*	Baranof Island, AK, USA	male	2b	98	23.77	SRR935596, SRR935603, SRR935610, SRR935618, SRR941809, SRR941812	[169]
ABC04	*U. arctos*	Chichagof Island, AK, USA	male	2b	74	11.85	SRR935599, SRR935606, SRR935613, SRR935622, SRR935630	[169]
ABC05	*U. arctos*	Chichagof Island, AK, USA	female	2b	172	26.31	SRR935600, SRR935607, SRR935614, SRR935623, SRR941808	[169]
PB12	*U. maritimus*	Iskanten, West Greenland	female	2a	156	28.45	SRR942293	[169]
PB42	*U. maritimus*	Melville Bugt, Greenland	male	2a	185	22.65	SRR942281-2	[169]
Uam	*U. americanus*	Maine, USA	male	NA	NA	37.13	SRR7813600-1	[261]

Table 3.3: Sample details for bears used in this study. Sample names, species, locations, and coverage details for the samples used for nuclear analyses.

3.3.2 Updated Polar bear reference genome

3.3.2.1 Genome polishing and scaffolding

We generated a high quality reference genome for sample D24101 (female polar bear from Southeast Greenland) using whole-genome shotgun, Hi-C, and Oxford Nanopore Technologies (ONT) libraries. Methods are detailed in the forthcoming paper [268]. Briefly, we updated the previously published polar bear reference genome assembly (GCA_000687225.1) [169] to more closely resemble our Southeast Greenland individual by first performing Pilon polishing. We mapped shotgun data from our female polar bear to the original polar bear reference using bwa mem [159], with default parameters except marking shorter split hits as secondary (-M flag), marking duplicates using the MarkDuplicates module of Picard [4]. We input these alignments Pilon using the –frags flag and run in diploid mode [289]. We scaffolded the preliminary assembly using a Hi-C library from the blood sample for D24101 using the "hic_scaffold" module in HiRise [220]. Gaps generated from scaffolding were filled using ONT data and PBJelly [77], running with flags for no between-scaffold gap-filling and only one read required to fill a gap. We error corrected the gap filled assembly using three successive rounds of Pilon polishing [289], and removed scaffolds short than 1 kbp. The final assembly is 2,330,490,886 bases in length and has a scaffold N50 of 72,237,886.

3.3.2.2 Identification X and Y scaffolds

We sought to identify the scaffolds that represented the sex chromosomes in our updated polar bear genome assembly. The data used to generate the original polar bear assembly

was from a male, however our shotgun, Hi-C, and long-read datasets were constructed from a female polar bear. Therefore, we expected that little to no scaffolding of the Y chromosome associated scaffolds would have occurred due to our processing.

We identified candidate X and Y scaffolds by alignment to the domestic dog and original polar bear genome assemblies (GCA_000002285.2, GCA_000687225.1) [109, 169] detailed in [268].

We aligned the polar bear assembly to the dog X chromosome from CanFam3.1 (NC_006621.3) using SatsumaSynteny2 from Satsuma2 with a minimum alignment length of 1000 [8]. We identified twenty scaffolds in this manner. We visualized syntenic hits between the dog X chromosome and the polar bear genome assembly using Circos [143].

To filter the twenty scaffolds identified based on synteny with the dog X chromosome, we gathered information on the scaffolds from the old polar bear assembly (GCF_000687225.1) which had been assigned as X chromosome. Prior research had constructed a list of twelve scaffolds in the previous polar bear assembly which were identified as the X chromosome through mapping of dog X chromosome genes and validation by coverage [33]. These identified scaffolds totaled to roughly 75Mb of X chromosome sequence [33]. We extracted the twelve previously identified X chromosome scaffolds and used SatsumaSynteny2 from Satsuma2 using a minimum alignment length of 500 to align them with the new polar bear genome to identify the results of scaffolding the genome [8]. We identified three scaffolds from the new polar bear assembly based on synteny to the old polar bear assembly X chromosome scaffolds. Only one of these three scaffolds overlapped with the scaffolds that were syntenic with the dog X chromosome.

We identified the regions of the new X chromosome assembly that were identified as X chromosome in the prior polar bear genome (Figure 3.4). Roughly 52 Mbp of sequence was newly identified as X chromosome in the new polar bear genome assembly.

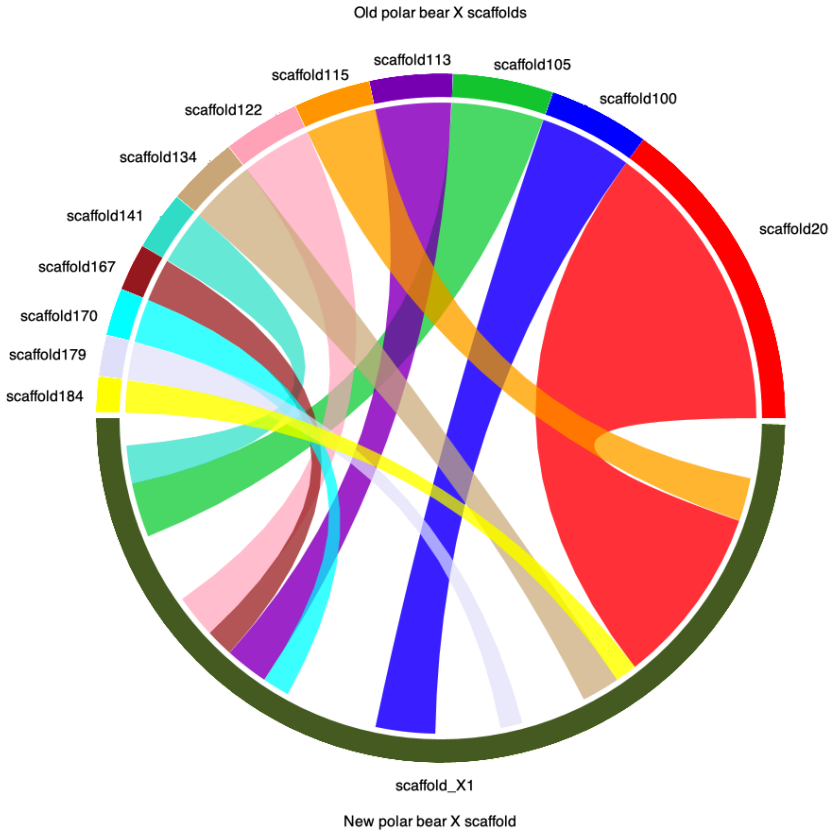

Figure 3.4: Old polar bear genome X chromosome identified scaffolds mapped to new polar bear genome largest X scaffold (X1). Synteny was performed with SynMap2 [257] and visualized using circos [143]. We assembled approximately 52 Mbp of newly identified X chromosome in the new polar bear genome.

We sought to validate the X chromosome scaffolds and distinguish them from Y chromosome scaffolds. First, based on identification of the Y chromosome scaffolds in our new polar bear assembly. Due to the Hi-C library being from a female polar bear, we expected vir-

tually no scaffolding of Y chromosome scaffolds. Using SatsumaSynteny2 [8] with a minimum alignment length of 500, we aligned twenty Y chromosome scaffolds from the old polar bear genome identified in prior work [25] to the new polar genome. We identified twenty scaffolds in our new polar bear genome that had one to one synteny with the old polar bear Y scaffolds, two of which were also identified when we mapped the old polar bear X scaffolds to the new polar bear genome.

For each identified X and Y candidate scaffold, we examined the female to male ratio of average scaffold coverage based on three previously published brown bears SRR5878341 (Female), SRR5878353 (Male), SRR5878360 (Male). We removed adapters from the brown bears using SeqPrep2 using the default parameters except for increasing the quality score cutoff to 15 (-q 15) [262]. We mapped reads to the new polar bear genome using bwa mem [159], filtering for map quality of 30. We confirmed X scaffolds as those with a ratio > 1.7, and Y scaffolds as those with a ratio < 0.3.

3.3.3 Sample processing

3.3.3.1 Modern bears processing

We adapter trimmed reads for the modern bears using SeqPrep2 using the default parameters except for increasing the quality score cutoff to 15 (-q 15) [262]. Two bears were sequenced using 10X technology (Dodge, EAK) which required the removal of barcodes and primers using proc10xGenomics.py using the default settings [9]. We mapped reads to the new polar bear genome using bwa mem [159], filtering for map quality of 30, removing alignments that are not primary, or have an unmapped mate. We removed PCR duplicates using samtools

rmdup [158]. Realignment around insertions and deletions was performed using GATK Realigner Target Creator and Indel Realignment [180].

3.3.3.2 Ancient bears processing

For libraries constructed using the single stranded Santa Cruz protocol, we first mapped reads to the PhiX genome using bwa mem [159] using default parameters. We removed any reads that mapped to PhiX. Next, we adapter trimmed and merged reads from our two ancient brown bear datasets using SeqPrep2, using the default parameters except for increasing the quality score cutoff to 15 (-q 15), decreasing the minimum read length to 25bp, and the overlap for merging (-o 15) [262]. To prevent any non-authentic ancient DNA from being included in further analyses, we aligned only the merged reads to the polar bear genome sequence using bwa aln [162] using ancient DNA parameters (-l 1024 -n 0.01 -o 2). We filtered such that all alignments had map quality of at least 30 and all alignments were primary. We removed PCR duplicates using samtools rmdup [158]. We sought to prevent any spurious mismappings that may occur due to the short read length of aDNA by removing reads shorter than 35bp. We chose 35bp as our threshold due to prior work which has shown that this fragment length considerably decreased the number of microbial alignments to the genome [185, 59].

We realigned around insertions and deletions using GATK Realigner Target Creator and Indel Realignment [180]. We used mapDamage2 [130] to assess the 5' and 3' deamination characteristic of ancient DNA to determine the authenticity of the data (Figures 3.5, 3.6). Both California grizzly bear samples showed damage characteristic of ancient DNA.

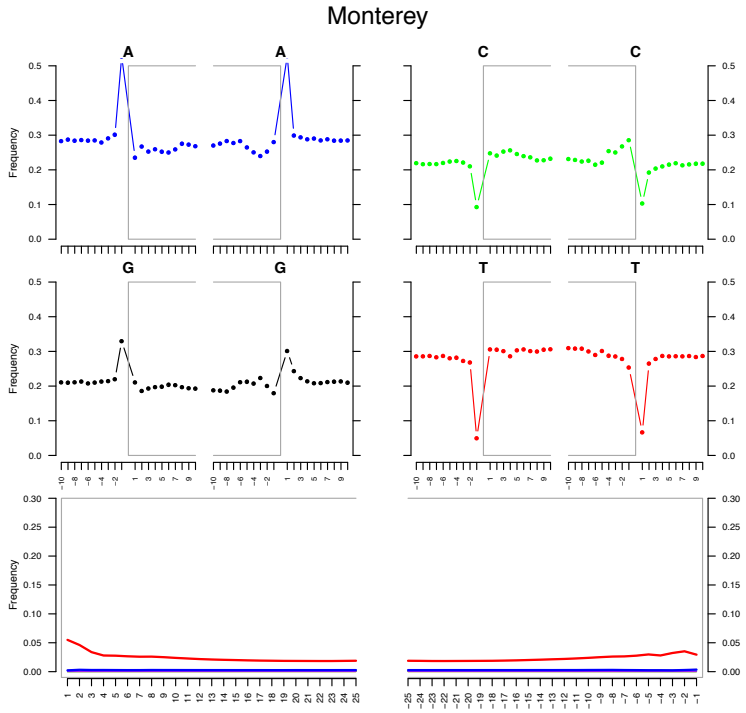

Figure 3.5: DNA fragmentation and nucleotide mis-incorporation profiles for Monterey sample as described by mapDamage2 [130]. DNA fragmentation profiles show the base composition of the first and last 10 read positions and within the 10 genomic positions preceding and following read ends. Nucleotide mis-incorporation profiles provide C to T (red), G to A (blue) and other (grey) occurrences of mutations relative to the occurrence of the reference nucleotide along the first/last 10 read positions. We observe elevated levels of nucleotide mis-incorporations and a damage profile that is characteristic of ancient DNA for Monterey one only the original template strand as a result of the single-stranded Santa Cruz method [132].

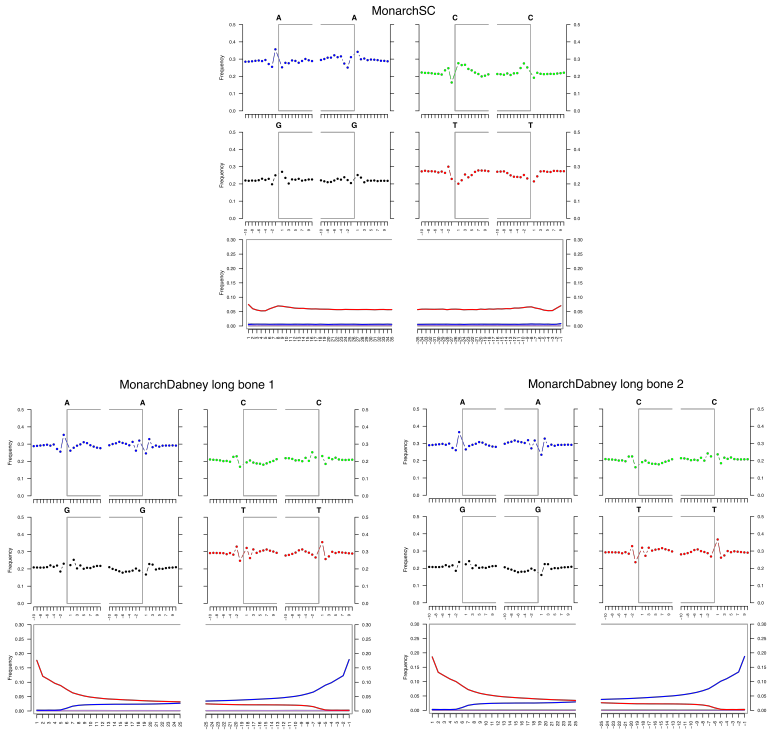

Figure 3.6: DNA fragmentation and nucleotide mis-incorporation profiles for three Monarch libraries as described by mapDamage2 [130]. DNA fragmentation profiles show the base composition of the first and last 10 read positions and within the 10 genomic positions preceding and following read ends. Nucleotide mis-incorporation profiles provide C to T (red), G to A (blue) and other (grey) occurrences of mutations relative to the occurrence of the reference nucleotide along the first/last 10 read positions. We observe elevated levels of nucleotide mis-incorporations and a damage profile that is characteristic of ancient DNA for Monarch. The Monarch library generated using the single-stranded Santa Cruz method [132] shows damage only on the original template read, where as libraries generated following the Dabney protocol show damage on both reads.

3.3.4 Population structure in brown bears

3.3.4.1 Matrilineal phylogeography in brown bears

We randomly selected at least 10 million adapter-trimmed reads for each of our samples (Table 3.3), and constructed a mitochondrial sequence using mia (mia flags: -i -c -C -U -F -k 13) [98] with EU4976651 as the reference. For Monarch, we had previously generated a mitochondrial enrichment dataset, and therefore used these reads for assembly of the mitochondria. For ancient samples, we used an ancient DNA substitution matrix to account for deamination in mia (-s). Coverage for the mitochondrial assemblies are listed in Table (3.3. We filtered assemblies using a consensus threshold of 66% and 3 X coverage per site for our relaxed filter, and a consensus threshold of 90% and 10 X coverage per site for our strict filter. Any site which did not meet these requirements was changed to an 'N'.

In addition to our assembled sequences, we downloaded mitochondrial sequences for one polar bear (NC_003428), one American black bear (AF303109), and ten brown bears (MG066705, AF303110, GU573491, AP012578, AP012585, AP012592, AP102581, AP012559, AP012563, AP012572). We included the mitochondrial sequence we used as a reference (EU4976651) in our mitochondrial dataset as well. We generated a strict and relaxed filtered alignment by aligning the respective sequences using clustal omega [251]. Using Geneious [135], we partitioned the alignment into six datasets: free sites, control region, rRNA, tRNA, 1st and 2nd coding positions, and 3rd coding positions based on mitochondrial annotations from the brown bear (AF303110, AP012592, MG593966), American black bear (JX196366), Asiatic black bear (FM177759), polar bear (GU573490), and spectacled bear (FM177764). We

ran PartitionFinder [150] to determine the partitions and best substitution model, with branch lengths unlinked using the Bayesian Information Criterion. PartitionFinder separated the relax filtered data into three mitochondrial partitions: 1) control region (Kimura two parameter + gamma), 2) 3rd codon positions (Hasegawa-Kishino-Yano + gamma), 3) free sites, rRNA, tRNA,1st and 2nd coding positions (Hasegawa-Kishino-Yano + gamma + invariable sites). The strict filtered dataset was partitioned into two mitochondrial partitions: 1) control region and 3rd codon positions (Hasegawa-Kishino-Yano + gamma), 2) free sites, rRNA, tRNA ,1st and 2nd coding positions (Hasegawa-Kishino-Yano + gamma + invariable sites). We performed Maximum likelihood (ML) and Bayesian phylogenetic analyses for both datasets using the partitioning specified above. We used RAxML-NG [263] to produce ML phylogenies, running one hundred bootstrap replicates. We also created phylogenies using a Bayesian approach using MrBayes [232] using the same partitioning as above. We ran four chains (one hot, three cold) for 10 million generations, with trees and model parameters sampled every 1,000 generations, with the first 25% discarded as burn-in. Both datasets and methods showed identical tree topologies.

3.3.4.2 Population structure in the nuclear genome of brown bears

To identify regions on which short reads can uniquely map in the new polar bear genome, we generated a mappability mask for the genome. We used seqbility (version 20091110) on the adapter trimmed data for our polar bear, using kmers of 35 bp, and a stringency of 0.90 to generate the mask file as detailed in prior work [217].

We generated a variant file using ANGSD with the command "angsd -snp_pval 1e-6 -minMapQ 25 -minQ 25 -skipTriallelic 1 -uniqueOnly 1 -minMaf 0.01 -dobcf 1 -gl 1 -dopost

1 -domajorminor 1 -domaf 1 -dogeno 1 -docounts 1 -rmTrans 1" [141] for all nineteen bears on 37 autosomal scaffolds larger than 1 Mb. We used on transversions to avoid the effects of ancient DNA damage. We kept only sites that were identified as uniquely mappable using the mappability mask we generated for the polar bear genome.

We used 6,777,432 transversions as input into SmartPCA from the EIGENSOFT package [208] to perform principal components analysis on nineteen bears encompassing three species. We saw strong separation between each species along both principal components. We also subset the data to include only North American brown bears and repeated the principal components analysis with EIGENSOFT SmartPCA [208], again using 6,777,432 SNPs.

We generated a maximum likelihood phylogeny using our whole genome SNP data with SNPhylo [152]. Using 6.78 million transversions as input, the program generated a maximum likelihood tree, using 100 bootstrap replicates, with the American black bear as the outgroup.

3.3.5 Polar bear ancestry in brown bears

We used the D-statistic to look for admixture between polar bears and brown bears. From the alignment files, we generated pseudohaploid fasta files for all bears using ANGSD '-dofasta 1 -minMapQ 30 -minQ 30 -setMaxDepthInd' to randomly select a base for each scaffold larger than 1Mb [141], using the 95th percentile of coverage as the maximum depth cutoff. We determined the 95th percentile separately for autosomes and scaffold_X1. We compared every pair of brown bears, using the polar bear as the introgressor and the American black bear as the outgroup, with quad-aln-report from ChromCompare [6] to generate counts of ABBA and

BABA sites in non overlapping windows of 500kb, masking CpG sites, and looking only at transversions. We restricted all D-statistic analyses to only sites identified in the mappability file to minimize mismappings due to short fragment lengths of ancient DNA. We calculated the D-statistic using quad_dstat_wt_jk from ChromCompare [6], as well as the weighted block jackknife to estimate standard error and significance. We used both PB12 and PB42 as the introgressing bear, testing all possible brown bear pairs. Results using either polar bear showed the same trends. We used only the PB42 polar bear for plotting.

To generate a X chromosome to autosome D-statistic ratio for pairs of brown bears, we also calculated the D-statistic for the largest autosomal scaffold, scaffold_1 (107Mb) to be able to make comparisons based on looking at a similar size portion of the genome. We followed the same methods as detailed above to calculate the D-statistic for pairs of brown bears, using PB42 as the introgressing species, and Uam as the outgroup on non overlapping 500kb windows on scaffold_1. We selected only one brown bear to represent each clade or island (SSweden for clade 1, ABC04 for Chichagof, Bar for Baranof, Adm2 for Admiralty, EAK for clade 3b, DenaliAK for clade 3a, Montana for modern clade 4, Monterey for ancient clade 4). We generated a X to autosome ratio by plotting the X D-statistic for a pair of brown bears and the autosomal D-statistic for the same pair of brown bears.

We next used the $\hat{f}$ statistic to quantify the relative polar bear ancestry identified in each brown bear for sex chromosomes and autosomes. We used three scripts [121] to calculate the amount of polar bear ancestry in each brown bear in our panel. We used 'F_hat_tv' to calculate the $\hat{f}$ statistic in non overlapping 500kb windows in both autosomes and scaffold_X1, using PB12 and PB42 as members of the introgressing species, SSweden as the least admixed

brown bear, and Uam as the outgroup. We used the script 'fhat_parser' to obtain $\hat{f}$ statistic values for each North American brown bear. To obtain standard error estimates, we ran the script 'weighted_block_jackknife_fhat' using a block size of 500kb.

3.3.6 Male-specific relationships among bears

We generated pseudohaploid fasta files for the 19 scaffolds identified as belonging to the Y chromosome for our ten male bears. We used Consensify to generate consensus sequences for each Y scaffold [15], using the 95th percentile of coverage as the maximum depth, and map and base quality thresholds of 30. We constructed a distance matrix using the 'ape' package in R, and generated a neighbor joining tree for the ten bears from the distance matrix.

3.4 Results

3.4.1 Matrilineal phylogeography in brown bears

A mitochondrial phylogeny for brown bears using the American black bear as the outgroup depicts the strong matrilineal phylogeography. Both California grizzly bears fall within the diversity of the clade 4 bears (Figures 3.7, 3.8). Our modern clade 4 bear, Dodge, also clusters with the lower 48 brown bears. Brown bears show strong structure based on the mitochondrial genome, whereby individuals of the same clade share closer ancestry with one another than with other groups of brown bears. Migration and isolation alone were not responsible for the level of differences seen between brown bear clade. The strong mitochondrial structure suggests of a deep history of monophyletic brown bear clades which went on to migrate and

occupy new ranges across the globe.

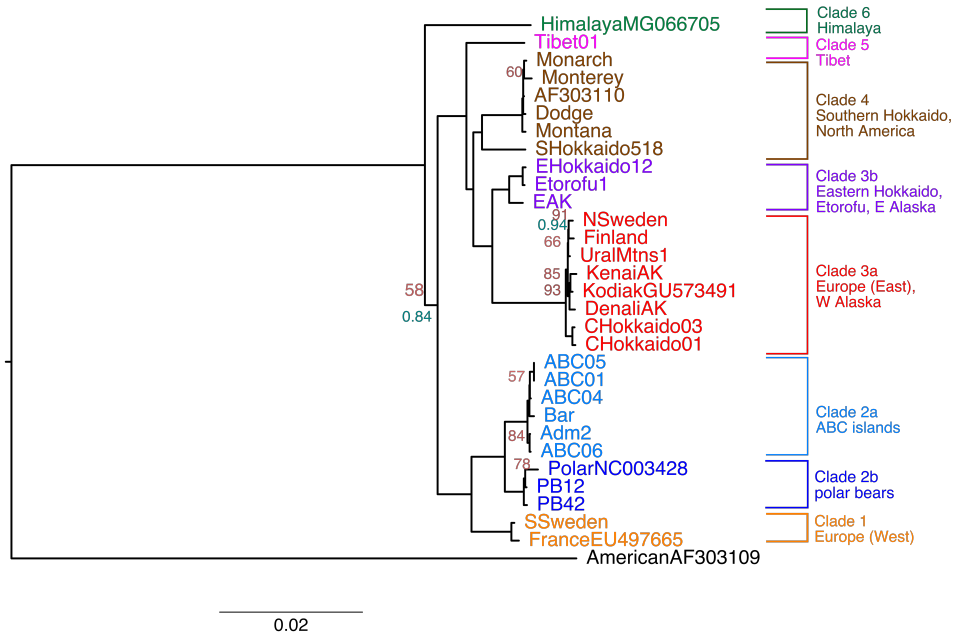

Figure 3.7: Relaxed mitochondrial phylogeny of six clades of brown bears. Detailed information on the bears used to construct this tree provided in Table 3.3. The phylogeny generated using relaxed filtering parameters (3X, 66% consensus) on the assembled mitochondrial sequences using both Maximum likelihood (ML) and Bayesian methods using three partitions [263, 232]. Bootstrap support values for ML tree of less than 96 reported in red, and posterior support values of less than 0.95 from the Bayesian tree are reported in green. Both trees show the same topology.

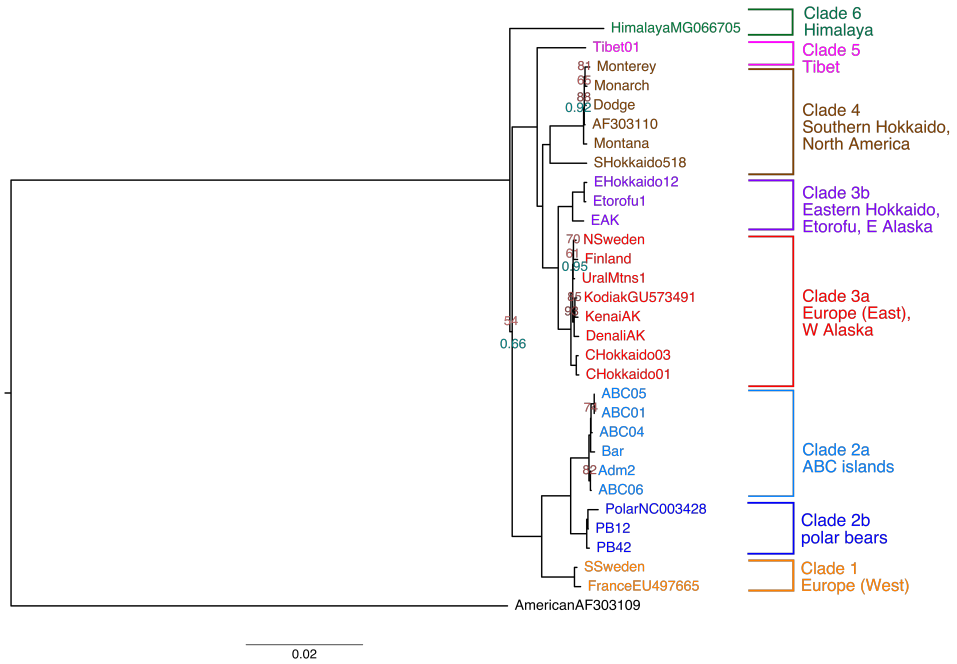

Figure 3.8: Strict mitochondrial phylogeny of six clades of brown bears. Detailed information on the bears used to construct this tree provided in Table 3.3. The phylogeny generated using strict filtering parameters (10X, 90% consensus) on the assembled mitochondrial sequences using both Maximum likelihood (ML) and Bayesian methods using two partitions [263, 232]. Bootstrap support values for ML tree of less than 96 reported in red, and posterior support values of less than 0.95 from the Bayesian tree are reported in green. Both trees show the same topology.

3.4.2 Population structure in the nuclear genome of brown bears

We used a set of 6.78 million transversions to perform a principal component analysis (PCA), both with and without the American black bear and polar bear samples. Our larger panel separates the three species (American black bear, polar bear, and brown bear) on both principal components 1 (PC1) and 2 (PC2) (Figure 3.9). PC1 explains 33.73% of the genetic variation, largely separating the American black bear from polar bears and brown bears. PC2 explains

roughly half the amount of genetic variations, but separates polar bears from brown bears and the American black bear. We see very clear separation on both PCs for each species.

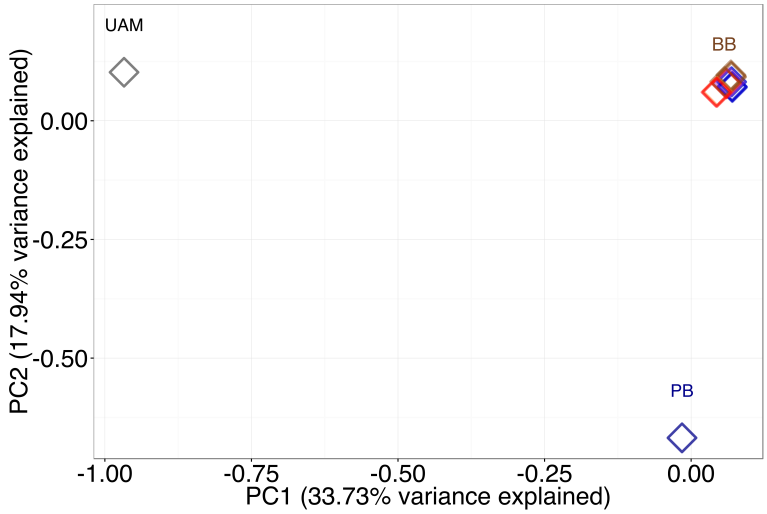

Figure 3.9: Principal component analysis for all bears in our panel using SmartPCA [208]. The first component separates the American black bear from the other bears, and the second component separates the polar bear.

Repeating the PCA using only the North American brown bears, allows for distinctions within brown bears (Figures 3.10, 3.11). The first three principal components describe 18.35%, 13.83%, and 10.51% of the total genetic variation, respectively. PC1 separates the both the clade 4 brown bears and the Baranof/Chichagof ABC bears from the mainland grizzly bears. The Baranof and Chichagof islands are the two furthest from the mainland, and based on prior work, are known to have the highest levels of polar bear ancestry of all North American grizzly bears [33, 36]. In comparison, huge recent reductions in range and numbers for clade 4 brown bears could mean that this population has experienced genetic drift, making them appear more divergent from other brown bears. PC2 and PC3 both tease apart the bears based on geog-

raphy, distinguishing between the western and eastern Alaskan brown bears, and the Admiralty and Baranof/Chichagof bears. Brown bears are highly philopatric, therefore populations retain much of their geographic structure at the nuclear level, as we see in the PCA.

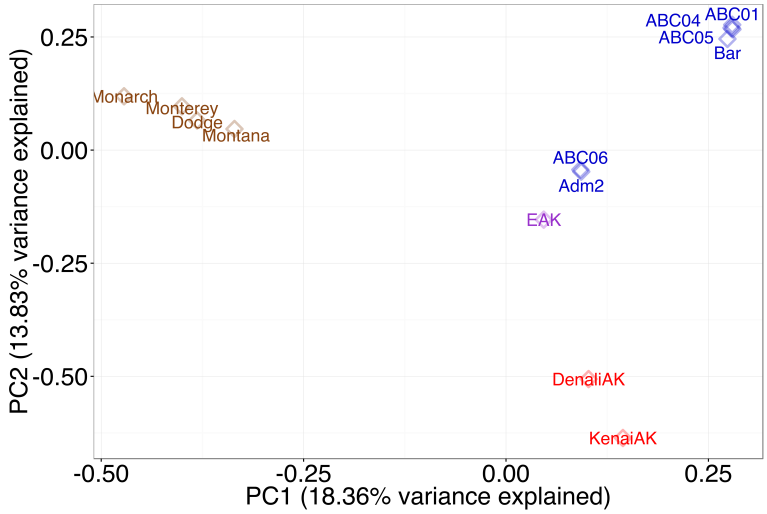

Figure 3.10: Principal component analysis (PC1 and PC2) for grizzly bears in our panel. PC1 separates clade 4 and Baranof/Chichagof bears from the other bears, and PC2 distinguishes the remaining bears based on geography.

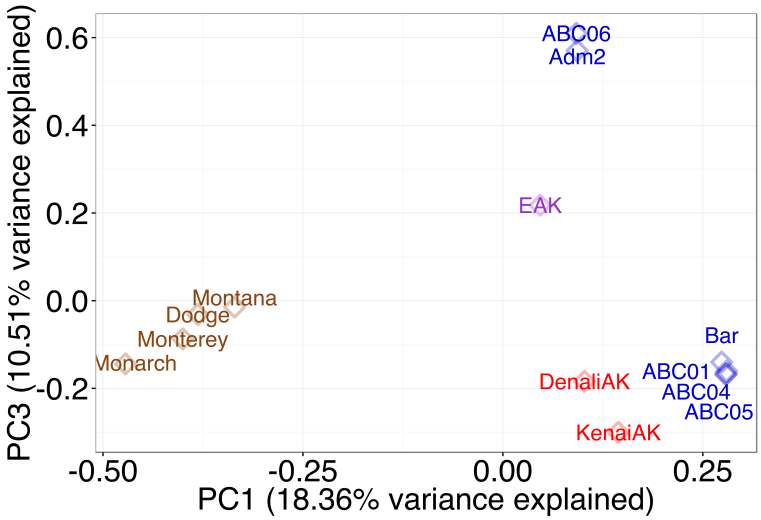

Figure 3.11: Principal component analysis (PC1 and PC3) for grizzly bears in our panel. PC1 separates clade 4 and Baranof/Chichagof bears from the other bears. PC3 distinguishes Admiralty island and clade 3b.

A maximum likelihood phylogeny constructed using 6.78 million transversions recapitulates the geographic clustering of brown bear in our panel (Figure 3.12). Bears group geographically; European brown bears, North American clade 3a bears, clade 4 bears, clade 3b, Admiralty bears, and Baranof/Chichagof bears cluster separately. Alaskan mainland bears of clade 3a appear the least similar to other North American brown bears. This clade is the most recent to have entered North America, sometime before the Bering land bridge disappeared 11kya. Clade 3a is also the most widely distributed, however in Alaska it remains restricted to the western region of the state.

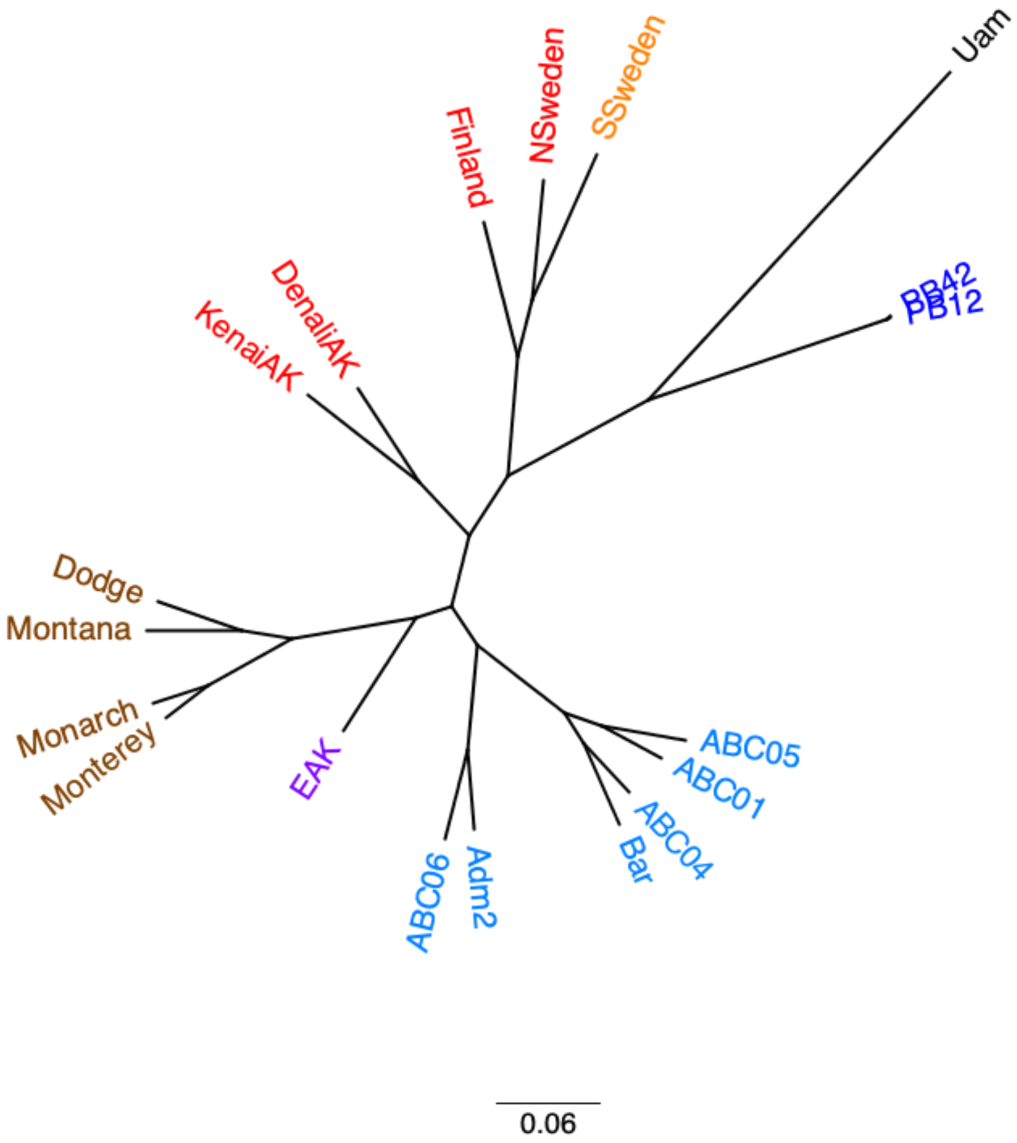

Figure 3.12: Bear SNP tree using 6.78 million transversions. [119].

3.4.3 Polar bear ancestry in brown bears

To detect admixture between polar bears and brown bears in our panel, we used the D-statistic. We used the American black bear to represent the ancestral allele (outgroup), po-

lar bears as the derived allele (introgressor), and every pair of possible brown bears from our sixteen bears to quantify relative levels of polar bear ancestry in brown bears (Figure 3.13). For the 37 scaffolds larger than 1 Mb in length, we observe that the ABC island brown bears share the greatest amount of alleles with polar bears compared to other brown bears sampled. All comparisons between an ABC island brown bear and a non-ABC island brown bear have significant Z-scores (Figure 3.14). A D-statistic value with a Z-score between 1.96 and -1.96 is not considered significant. Thus, ABC island bears share significantly more ancestry with polar bears than non-ABC island bears on the autosomes. Comparisons between the ABC islands validate a known trend; that the islands furthest from the mainland (Chichagof and Baranof) share a greater number of derived alleles with polar bears compared to bears from Admiralty island, the island nearest the mainland. The distance from the mainland made it more difficult for mainland brown bears to migrate to the Chichagof and Baranof islands. Therefore these bears retained higher levels polar bear ancestry, as seen in the higher numbers of shared alleles for polar bears and Chichagof/Baranof brown bears. Our clade 3b bear, EAK, displays the highest rate of shared derived alleles with polar bears of any non-ABC island bear. This clade is closest geographically to the ABC islands, and may even be source population of brown bears that colonized the ABC islands after the LGM. Thus migrations between the mainland clade 3b bears and the ABC islands would have been commonplace, and polar bear ancestry would have accumulated in the nearby mainland brown bears subsequent to admixture on the ABC islands.

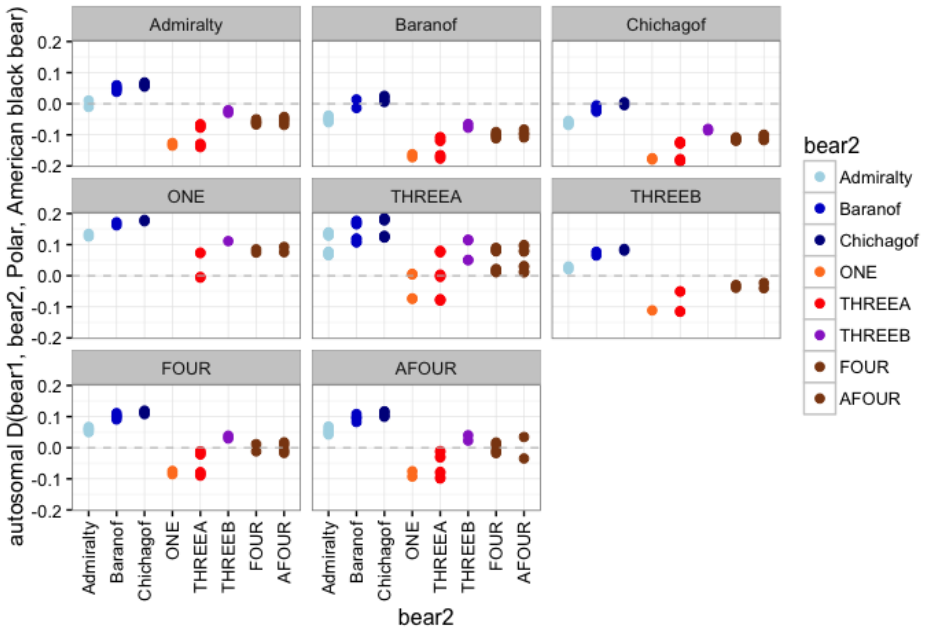

Figure 3.13: Admixture between polar bears and brown bears on the autosomes. Using the D-statistic, we looked for evidence of allele sharing between brown bears and polar bears on the 37 autosomal scaffolds larger than 1 Mb. Bear 1 is identified at the top of each panel, and comparison with a different bear 2 is shown along the x-axis colored by bear 2. Positive D-statistic scores indicate bear 1 has greater allele sharing with the polar bear than bear 2. Negative D-statistic scores indicate bear 2 has greater allele sharing with polar bears than bear 1. Significance was determined based on weighted block jackknife calculations for non overlapping 500kb windows (Figure 3.14). Bears from Chichagof Island show the highest levels of polar bear ancestry on the X scaffold relative to all other brown bears included this study. The island is the furthest from the mainland, and thus likely received fewer migrations than Admiralty and Baranof islands, which are both closer to the mainland.

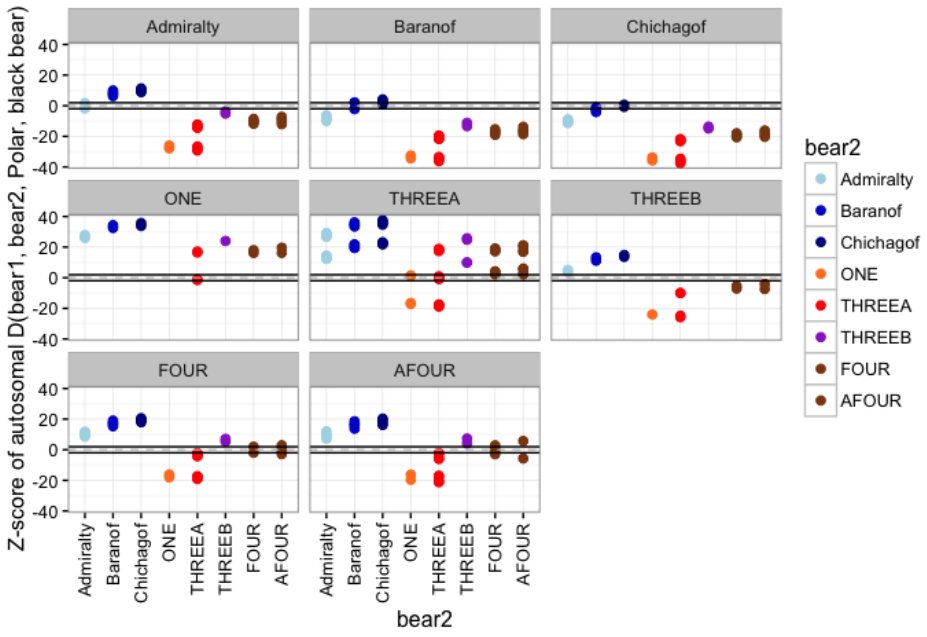

Figure 3.14: Significance scores for admixture between polar bears and brown bears on the autosomes. Using the D-statistic, we looked for evidence of allele sharing between brown bears and polar bears on the 37 autosomal scaffolds larger than 1 Mb. Bear 1 is identified at the top of each panel, and with comparison with a different bear 2 shown on the x-axis colored by bear 2. We used the weighted block jackknife on non overlapping 500kb windows to calculate Z score using the standard error estimation, shown here. Z scores greater than between 1.96 and -1.96 indicate significant amounts of allele sharing between one of the brown bears and the polar bear. Positive Z scores indicate bear 1 has greater allele sharing with the polar bear than bear 2. Negative Z scores indicate bear 2 has greater allele sharing with polar bears than bear 1.

We next quantified the amount of allele sharing between brown bears and polar bears on the X chromosome. As admixture on the ABC islands was sex biased, we expect to see greater amounts of allele sharing between ABC island bears and polar bears on the X chromosome than we do on the autosomes. Our results show that the ABC island bears have the greatest amount of allele sharing with polar bears relative to non-ABC bears, the Chichagof and

Baranof islands being the greatest (Figures 3.15, 3.16). Within ABC island comparisons only show a significant amount of allele sharing with polar bears when comparing the Chichagof and Admiralty islands, where Chichagof bears have a significant amount more shared alleles with polar bears than Admiralty bears. Comparisons between Chichagof and Baranof islands and Baranof and Admiralty islands do not show significant levels of alleles sharing with polar bears for the Chichagof and Baranof bears, respectively. All non-ABC island bears have non significant levels of allele sharing with polar bears, with the exception of EAK, whose population is closest to the ABC islands. In both the autosomal and X chromosomes we see greater levels of allele sharing between our eastern Alaskan bear and polar bears than other non-ABC island brown bears. The vicinity of this population to the polar bear source population on the ABC islands facilitated more admixture events between eastern Alaska mainland bears and the islands off the coast. The higher level of allele sharing on the X chromosome for EAK and polar bears compared to non-ABC island bears likely results from dispersal of the ABC island bears to the nearby mainland. All other non-ABC island brown bears have minimal allele sharing with polar bears compared to other non-ABC island brown bears.

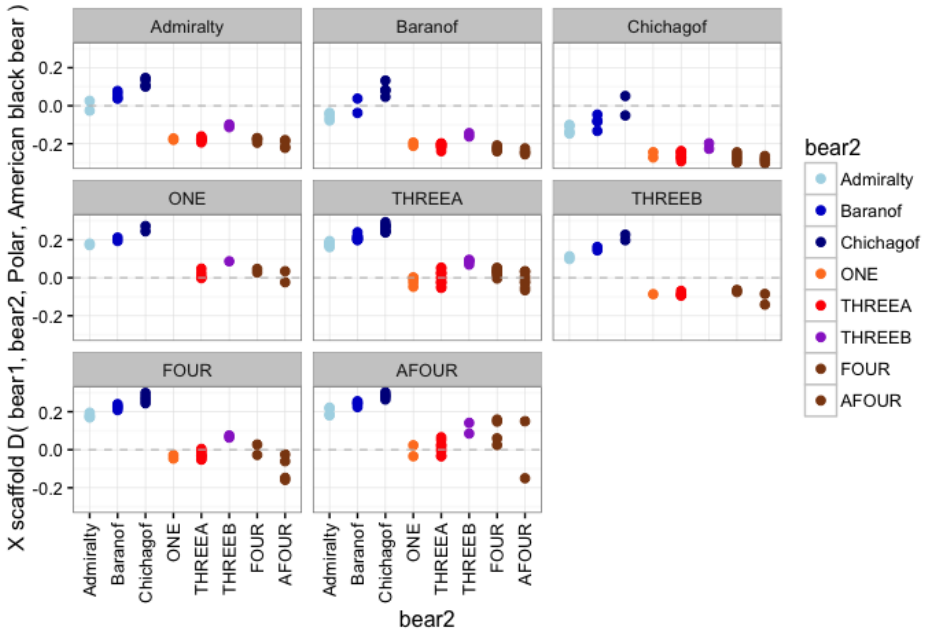

Figure 3.15: Admixture between polar bears and brown bears on the X scaffold. Using the D-statistic, we looked for evidence of allele sharing between brown bears and polar bears on the X scaffold. Bear one is identified at the top of each panel, and each bear two comparison is shown along the x-axis colored by bear two. Brown bear populations which show greater allele sharing with the polar bear on the X scaffold have positive D-statistic values. Significance was determined based on weighted block jackknife calculations for non overlapping 500kb windows (Figure 3.16). Bears from Chichagof Island show the highest levels of polar bear ancestry on the X scaffold relative to all other brown bears included this study. The island is the furthest from the mainland, and thus likely received fewer male brown bear migrations than Admiralty and Baranof islands, which are both closer to the mainland.

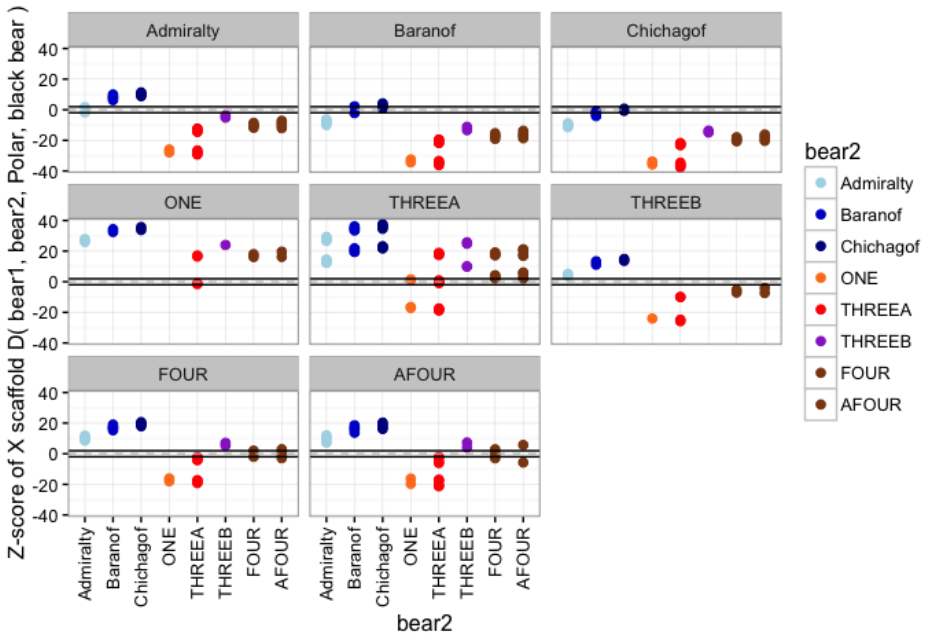

Figure 3.16: Significance scores for admixture between polar bears and brown bears on the X scaffold. Using the D-statistic, we looked for evidence of allele sharing between brown bears and polar bears on the X scaffold. Bear one is identified at the top of each panel, and each bear two comparison is shown along the x-axis colored by bear two. We used the weighted block jackknife on non overlapping 500kb windows to calculate Z score using the standard error estimation, shown here. Z scores greater than between 1.96 and -1.96 indicate significantly greater amounts of allele sharing between one of the brown bears and the polar bear.

We assess the relative levels of polar bear ancestry on the autosomes and X scaffold by looking at the ratio for the D-statistic of the X to autosomal chromosomes. We see evidence of sex-biased polar bear ancestry in ABC island bears based on the X to autosome D-statistic ratio (Figure 3.17). The sex-biased polar bear ancestry is most dramatic in Admiralty island samples. These bears exhibit high (5:1 to 4:1) ratios of ancestry on their X chromosome relative to autosomes. This trend is the result of the heavily male-biased brown bear admixture with polar bears on the ABC island nearest the mainland, Admiralty island. Conversely, mainland

brown bears exhibit bias towards ancestry on the autosome relative to the X chromosome.

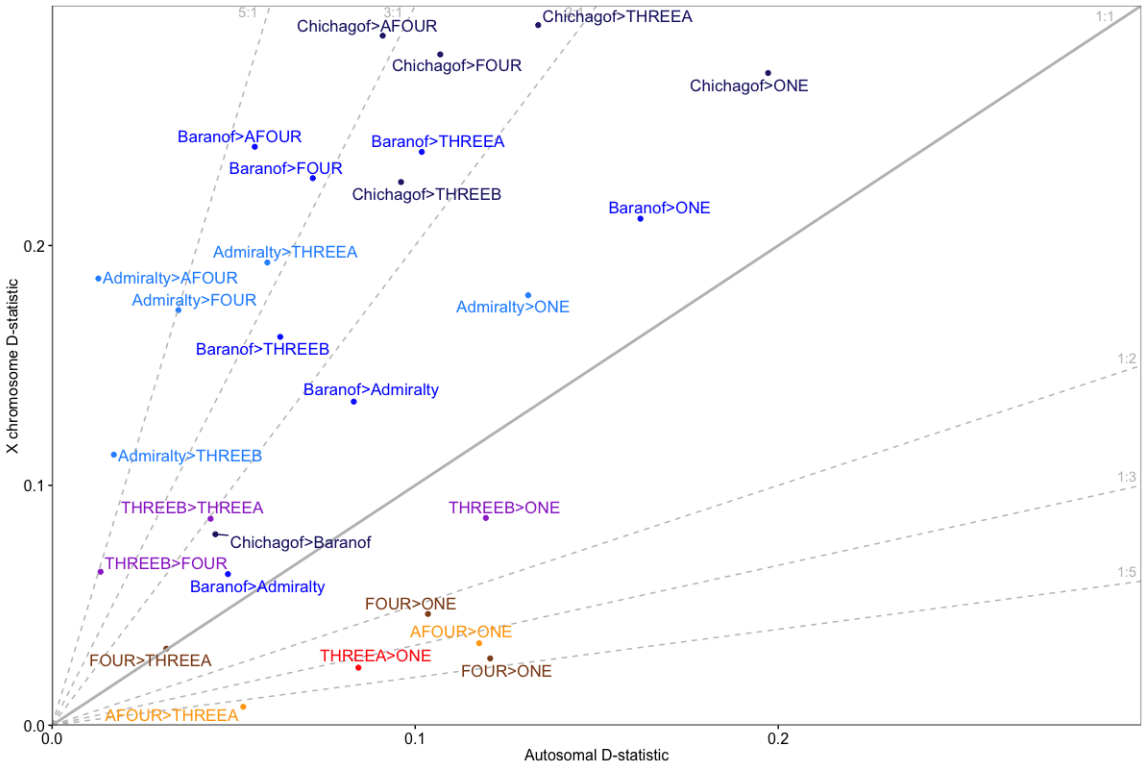

Figure 3.17: X chromosome to autosome D-statistic ratio for brown bears. We compare the X chromosome to autosome D-statistic values for each possible pair of brown bears. Values in the figure are only those where the bear listed first has greater allele sharing with polar bears than the second brown bear in the pair. We used only scaffold_1 as the autosomal D-statistic value so as to compare scaffolds of similar size (scaffold_1 is 107 Mbp, scaffold_X1 is 125 Mbp). Admiralty island brown bears have higher levels of allele sharing with polar bears on the X chromosome than the autosome. This effect is due to the greater number of migrations from mainland male brown bears onto Admiralty island compared islands further from the mainland (Baranof and Chichagof). We used one bear to represent each island or clade for this analysis (SSweden for clade 1, ABC04 for Chichagof, Bar for Baranof, Adm2 for Admiralty, EAK for clade 3b, DenaliAK for clade 3a, Montana for modern clade 4, Monterey for ancient clade 4).

We quantified the amount of polar bear ancestry in our brown bear panel using the $\hat{f}$

statistic for the autosomes and X chromosome (Table 3.4). The ABC islands brown bears have the highest proportion of polar bear ancestry on the autosomes or X chromosome of any brown bear. Specifically, the Chichagof island bears have the highest polar bear ancestry proportions as measured by the $\hat{f}$ statistic. Brown bears from Baranof island have the next highest levels of polar bear ancestry on both autosomes and the X chromosomes, followed by Admiralty islands. North American brown bears have between 3 to 10% polar bear ancestry on their autosomes. The $\hat{f}$ statistic is considered to be a conservative minimum of ancestry estimates [68], thus the true polar bear ancestry levels in the North American brown bears are higher. Polar bear ancestry estimates for brown bears on the X chomosome show a drastic difference between the ABC island and non-ABC island brown bears. ABC island bears show the same trend as on the autosomes, Chichagof island bears having the highest proportion of polar bear ancestry, followed by Baranof and Admiralty islands. Mainland brown bears migrated onto the ABC islands and mated with polar bears, slowly dissipating the polar bear ancestry in the local island bears. Islands closer to the mainland (Admiralty) had greater numbers of migrations, and therefore showed greater X bias in polar bear ancestry. Non-ABC island bears have considerably lower levels of polar bear ancestry on the X chromosome, only 1 to 3%. Male migrants from the ABC islands to the mainland would carry less polar bear ancestry each generation, and thus mainland populations would receive progressively less polar bear ancestry over time on the X chromosome.

Sample name	location	sex	brown bear clade	autosomes $\hat{f}$ (SE)	X chrom $\hat{f}$ (SE)
ABC04	Chichagof Island, AK, USA	male	2b	9.40% (0.42%)	9.08%(2.06%)
ABC05	Chichagof Island, AK, USA	female	2b	8.53% (0.44%)	7.92%(2.01%)
Bar	Baranof Island, AK, USA	male	2b	7.82% (0.41%)	5.91%(2.12%)
ABC01	Baranof Island, AK, USA	male	2b	8.30%(0.41%)	6.20%(2.04%)
ABC06	Admiralty Island, AK, USA	female	2b	6.36%(0.38%)	5.28%(1.62%)
Adm2	Admiralty Island, AK, USA	female	2b	5.67%(0.39%)	5.06%(1.90%)
EAK	unknown, Eastern AK, USA	male	3b	5.60%(0.35%)	2.80%(1.35%)
Monterey	Monterey, California, USA	male	4	4.99%(0.37%)	2.24%(1.26%)
Monarch	Ventura, California, USA	male	4	4.50%(0.38%)	1.17%(1.32%)
Montana	Glacier Park, Montana,USA	male	4	4.15%(0.38%)	1.55%(1.33%)
Dodge	unknown, lower 48, USA	male	4	3.80%(0.37%)	0.92%(1.40%)
DenaliAK	Denali Natl Park, AK, USA	female	3a	3.37%(0.31%)	0.13%(1.21%)
KenaiAK	Kenai Peninsula, AK, USA	female	3a	2.98%(0.31%)	1.15%(1.04%)

Table 3.4: Proportion of polar bear ancestry measured using the $\hat{f}$ statistic for autosomes and X chromosome for North American brown bears. Standard error in parentheses calculated on non-overlapping 500kb windows. ABC island bears have the highest proportion of polar bear ancestry on autosomes and X chromosome. Non-ABC islands bears have lower proportions of polar bear ancestry on the X chromosomes than autosomes. We used brown bear SSweden as the unadmixed bear, polar bears PB12 and PB42 were used as the other source population, and the American black bear served as the outgroup.

3.4.4 Male-specific relationships among bears

While the X scaffold shows strong geographical trends in the levels of polar bear ancestry in brown bears, the Y chromosome does not. As the Y chromosome does not undergo recombination, we could not use the D-statistic to look for evidence of allele sharing between polar bears and brown bears. To assess the relatedness of brown bears and polar bears on the Y, we generated a neighbor joining tree for the ten male bears in our panel (Figure 3.18). The Y chromosome phylogeny is monophyletic for brown bears in respect to polar bears and American black bears. We also do not see the strong phylogeography evident in mitochondrial and nuclear DNA in the Y chromosome of brown bears. The longer branch length for Monarch is presumed to be due to higher levels of ancient DNA damage. The consensus method of

pseudohaploidization removes most of the effects of DNA deamination in the final fasta file.

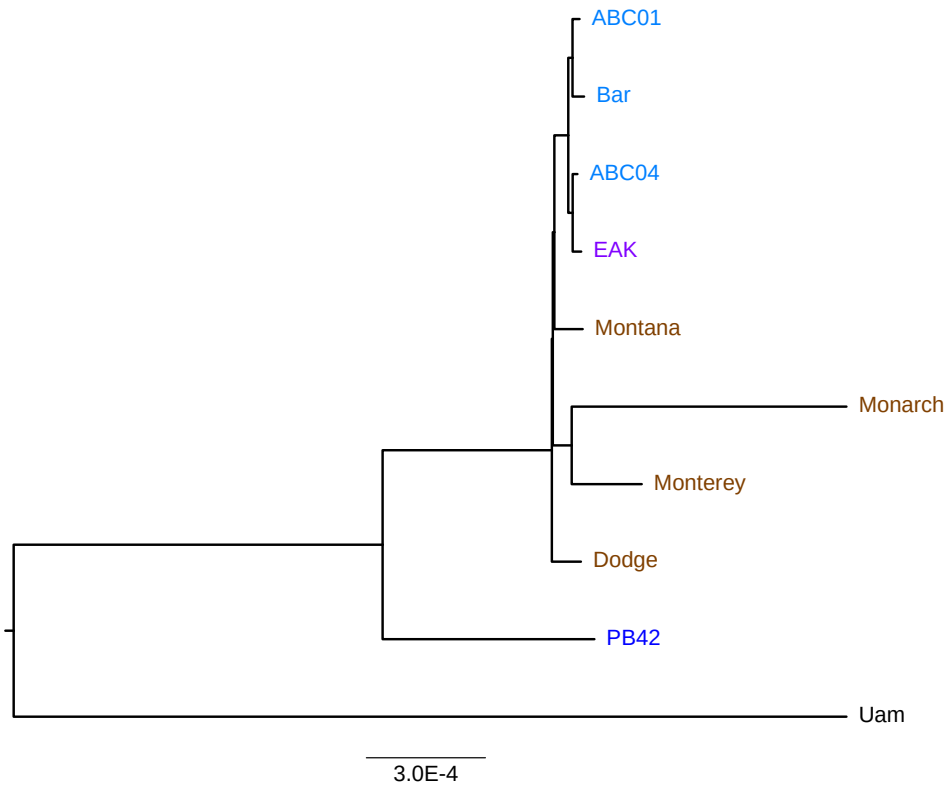

Figure 3.18: Neighbor joining tree of 19 Y scaffolds. Haploid fasta sequences for the 19 Y scaffolds were generated using Consensify to minimize the effects of aDNA [15] using the 95th percentile of coverage as the maximum coverage cutoff, and map and base quality thresholds of at least 30. The Y phylogeny distinguishes the three bear species, American black bear, polar bear and brown bear. ABC island brown bears cluster with brown bears, not polar bears.

3.5 Discussion and next steps

The brown bear has been well studied as a species that has been shaped by its environment. Throughout the Pleistocene, cycles of glaciation forced populations to find refugium. After the ice melted, brown bears colonized new ranges [107], or ranges left unoccupied due to

local population extinctions [16].

Our work recapitulates prior research which uncovered male biased brown bear gene flow from the mainland onto the ABC islands [33, 36]. Stranded off the coast after the ice sheets retreated, polar bears on the ABC islands were visited by male brown bears from the nearby mainland. Over time, the polar bear ancestry on the ABC islands decreased, and the population was converted into brown bears that retained some polar bear ancestry. In our analysis, we found the greatest proportion of X chromosome to autosome polar bear ancestry on brown bears from the Admiralty islands. The proximity of Admiralty island to the mainland meant a higher number of male brown bear migrants mating with the local polar bears. In comparison, Baranof and Chichagof islands, island more distant from the mainland, have a smaller proportion of X chromosome to autosome polar bear ancestry in comparison to Admiralty bears. Thus the legacy of male biased gene flow is seen today in the genomes of the ABC island brown bears.

In the past two hundred years, humans have shaped the history of the brown bear, by reducing both their range and numbers. North American brown bears, also known as grizzly bears, are endangered today, having survived extinction in isolated reserves and parks in Alaska, and the northwest [188]. Their range, however, once extended from Alaska to Mexico and from the Great Plains to California, where they were known as the California grizzly bearthe iconic symbol of the state. While historical records show that the California grizzly bear was once abundant, and in fact a major predator during the Gold Rush Era, it became extinct in the early 1900s, thanks mainly to a concerted depredation effort by humans. Today, the human relationship with large predators has changed, as we increasingly recognize the role of predators in maintaining biodiverse and stable ecosystems.

Brown bear ancient DNA has revealed past migrations, extinctions, and the diversity of the species [153, 16]. Here, we used two museum samples from California grizzly bears to assess the relationship of this extinct population to extant brown bears. Both mitochondrial and nuclear analyses place the extinct California brown bear with our other clade 4 brown bears. California grizzly bears carry at least 4-5% polar bear ancestry in their genome, the result of migrations off the ABC islands onto the North American mainland. Interestingly, clade 4 brown bears harbor greater levels of polar bear ancestry than the clade 3a brown bears of Western Alaska. This suggests migrations between ABC island brown bears and brown bears in the contiguous US and southwestern Canada were frequent. A coastal migration route would have been viable after 17,000 years ago, allowing brown bears a path south before the ice sheets melted [154]. Nuclear genomic data from brown bear fossils from the Pacific northwest and Southeastern Alaska could confirm coastal migrations of North American brown bears after the Last Glacial Maximum. Museum collections continue to be a unique source of genetic information, and our ability to now generate whole genome data enable us to form a fuller picture of paleoecology and past biodiversity of species.

Chapter 4

Conclusion

For my dissertation, I presented *de novo* genome assemblies for two megafauna, the puma and the spectacled bear, generated from second and third generation sequencing data. Using this genome assembly, I learned about the diversity of the species, a Central American admixture event in Florida pumas, and provided an example of the short-lived benefits to translocations. This genome allows researchers to learn more about the effects of habitat fragmentation on pumas across their range at a genome scale, and has the potential to aid in conservation management.This chapter was published in Nature Communications in October of 2019, for which I served as co-first author.

Next, I presented the genome assembly of the spectacled bear, the last of the tremarctine bears. The spectacled bear is the closest living relative to the extinct giant short-faced bears that roamed North America during the last Ice Age. I constructed a genome for the giant-short faced bear using the spectacled bear genome as the reference. I was able to uncover the demographic histories of both species, and explore their relationships to other extant bears. I am

working with collaborators on an expansion of this project to prepare the work presented in this chapter for submission.

In my third chapter, I generated genomes for two extinct California grizzly bears, the symbol of the state. Using population level genomic analyses, I determined that the California grizzly bears I performed population level analyses to assess the relationship of extant brown bears to the extinct California population. I used an updated genome assembly of the polar bear for this work, in which I helped identify sex chromosome scaffolds in the genome. This work is also currently in preparation for submission.

In addition to the work described above, I have also co-authored published research on horses and fruit flies. I assembled a mitochondrial genome for Twilight, a female Thoroughbred horse used to construct the EquCab3 reference genome [131]. My contribution to fruit fly research entailed scaffolding the *de novo* assembly for a single outbred fruit fly to construct a chromosome-scale assembly [11].

Together, I have contributed to the field with the addition of at least three genome assemblies, and new reference-guided genomes for a number of extinct species. These genome-scale data sets add to the growing field of work conserving top predators, and have the potential to inform conservation management.

Bibliography

[1] GenBank and WGS Statistics. https://www-ncbi-nlm-nih-gov.oca.ucsc.edu/genbank/statistics/.

[2] NovaSeq System Specifications — The next era of sequencing starts now. https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html.

[3] pairtools. https://github.com/mirnylab/pairtools.

[4] Picard tools - by broad institute. http://broadinstitute.github.io/picard/.

[5] Dna sequencing costs: data from the nhgri genome sequencing program. www.genome.gov/sequencingcostsdata, 2017.

[6] Paleogenomics/Chrom-Compare. UCSC Paleogenomics Lab, August 2019.

[7] UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515, January 2019.

[8] Bioinfologics/satsuma2. bioinfologics, March 2020.

[9] Ucdavis-bioinformatics/proc10xG. UC Davis Bioinformatics Core, June 2020.

[10] Mark D. Adams, Susan E. Celniker, Robert A. Holt, Cheryl A. Evans, Jeannine D. Gocayne, Peter G. Amanatides, Steven E. Scherer, Peter W. Li, Roger A. Hoskins, Richard F. Galle, Reed A. George, Suzanna E. Lewis, Stephen Richards, Michael Ashburner, Scott N. Henderson, Granger G. Sutton, Jennifer R. Wortman, Mark D. Yandell, Qing Zhang, Lin X. Chen, Rhonda C. Brandon, Yu-Hui C. Rogers, Robert G. Blazej, Mark Champe, Barret D. Pfeiffer, Kenneth H. Wan, Clare Doyle, Evan G. Baxter, Gregg Helt, Catherine R. Nelson, George L. Gabor, Miklos, Josep F. Abril, Anna Agbayani, Hui-Jin An, Cynthia Andrews-Pfannkoch, Danita Baldwin, Richard M. Ballew, Anand Basu, James Baxendale, Leyla Bayraktaroglu, Ellen M. Beasley, Karen Y. Beeson, P. V. Benos, Benjamin P. Berman, Deepali Bhandari, Slava Bolshakov, Dana Borkova, Michael R. Botchan, John Bouck, Peter Brokstein, Phillipe Brottier, Kenneth C. Burtis, Dana A. Busam, Heather Butler, Edouard Cadieu, Angela Center, Ishwar Chandra, J. Michael Cherry, Simon Cawley, Carl Dahlke, Lionel B. Davenport, Peter Davies, Beatriz de Pablos, Arthur Delcher, Zuoming Deng, Anne Deslattes Mays, Ian Dew, Suzanne M. Dietz, Kristina Dodson, Lisa E. Doup, Michael Downes, Shannon

Dugan-Rocha, Boris C. Dunkov, Patrick Dunn, Kenneth J. Durbin, Carlos C. Evangelista, Concepcion Ferraz, Steven Ferriera, Wolfgang Fleischmann, Carl Fosler, Andrei E. Gabrielian, Neha S. Garg, William M. Gelbart, Ken Glasser, Anna Glodek, Fangcheng Gong, J. Harley Gorrell, Zhiping Gu, Ping Guan, Michael Harris, Nomi L. Harris, Damon Harvey, Thomas J. Heiman, Judith R. Hernandez, Jarrett Houck, Damon Hostin, Kathryn A. Houston, Timothy J. Howland, Ming-Hui Wei, Chinyere Ibegwam, Mena Jalali, Francis Kalush, Gary H. Karpen, Zhaoxi Ke, James A. Kennison, Karen A. Ketchum, Bruce E. Kimmel, Chinnappa D. Kodira, Cheryl Kraft, Saul Kravitz, David Kulp, Zhongwu Lai, Paul Lasko, Yiding Lei, Alexander A. Levitsky, Jiayin Li, Zhenya Li, Yong Liang, Xiaoying Lin, Xiangjun Liu, Bettina Mattei, Tina C. McIntosh, Michael P. McLeod, Duncan McPherson, Gennady Merkulov, Natalia V. Milshina, Clark Mobarry, Joe Morris, Ali Moshrefi, Stephen M. Mount, Mee Moy, Brian Murphy, Lee Murphy, Donna M. Muzny, David L. Nelson, David R. Nelson, Keith A. Nelson, Katherine Nixon, Deborah R. Nusskern, Joanne M. Pacleb, Michael Palazzolo, Gjange S. Pittman, Sue Pan, John Pollard, Vinita Puri, Martin G. Reese, Knut Reinert, Karin Remington, Robert D. C. Saunders, Frederick Scheeler, Hua Shen, Bixiang Christopher Shue, Inga Sidén-Kiamos, Michael Simpson, Marian P. Skupski, Tom Smith, Eugene Spier, Allan C. Spradling, Mark Stapleton, Renee Strong, Eric Sun, Robert Svirskas, Cyndee Tector, Russell Turner, Eli Venter, Aihui H. Wang, Xin Wang, Zhen-Yuan Wang, David A. Wassarman, George M. Weinstock, Jean Weissenbach, Sherita M. Williams, Trevor Woodage, Kim C. Worley, David Wu, Song Yang, Q. Alison Yao, Jane Ye, Ru-Fang Yeh, Jayshree S. Zaveri, Ming Zhan, Guangren Zhang, Qi Zhao, Liansheng Zheng, Xiangqun H. Zheng, Fei N. Zhong, Wenyan Zhong, Xiaojun Zhou, Shiaoping Zhu, Xiaohong Zhu, Hamilton O. Smith, Richard A. Gibbs, Eugene W. Myers, Gerald M. Rubin, and J. Craig Venter. The Genome Sequence of Drosophila melanogaster. *Science*, 287(5461):2185–2195, March 2000.

[11] Matthew Adams, Jakob McBroome, Nicholas Maurer, Evan Pepper-Tunick, Nedda F. Saremi, Richard E. Green, Christopher Vollmers, and Russell B. Corbett-Detig. One fly–one genome: Chromosome-scale genome assembly of a single outbred Drosophila melanogaster. *Nucleic Acids Research*, 48(13):e75–e75, July 2020.

[12] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990.

[13] Patricia M. Anderson and Anatoly V. Lozhkin. The Stage 3 interstadial complex (Karginskii/middle Wisconsinan interval) of Beringia: Variations in paleoenvironments and implications for paleoclimatic interpretations. *Quaternary Science Reviews*, 20(1):93–125, January 2001.

[14] Axel Barlow, James A. Cahill, Stefanie Hartmann, Christoph Theunert, Georgios Xenikoudakis, Gloria G. Fortes, Johanna L. A. Paijmans, Gernot Rabeder, Christine Frischauf, Aurora Grandal-d'Anglade, Ana García-Vázquez, Marine Murtskhvaladze, Urmas Saarma, Peeter Anijalg, Tomaž Skrbinšek, Giorgio Bertorelle, Boris Gasparian, Guy Bar-Oz, Ron Pinhasi, Montgomery Slatkin, Love Dalén, Beth Shapiro, and Michael

Hofreiter. Partial genomic survival of cave bears in living brown bears. *Nature Ecology & Evolution*, 2(10):1563–1570, October 2018.

[15] Axel Barlow, Stefanie Hartmann, Javier Gonzalez, Michael Hofreiter, and Johanna L. A. Paijmans. Consensify: A Method for Generating Pseudohaploid Genome Sequences from Palaeogenomic Datasets with Reduced Error Rates. *Genes*, 11(1), January 2020.

[16] I. Barnes, P. Matheus, B. Shapiro, D. Jensen, and A. Cooper. Dynamics of Pleistocene Population Extinctions in Beringian Brown Bears. *Science*, 295(5563):2267–2270, March 2002.

[17] Ross Barnett, Ian Barnes, Matthew J. Phillips, Larry D. Martin, C. Richard Harington, Jennifer A. Leonard, and Alan Cooper. Evolution of the extinct sabretooths and the american cheetah-like cat. *Current Biology*, 15(15):R589–R590, August 2005.

[18] M A Barone, M E Roelke, J Howard, J L Brown, A E Anderson, and D E Wildt. Reproductive characteristics of male florida panthers: Comparative studies from florida, texas, colorado, latin america, and north american zoos. *J. Mammal.*, 75(1):150–162, 1994.

[19] Will Beaumont, Robert Beverly, John Southon, and R. E. Taylor. Bone preparation at the KCCAMS laboratory. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 268(7):906–909, April 2010.

[20] Paul Beier. Dispersal of juvenile cougars in fragmented habitat. *J. Wildl. Manage.*, 59(2):228, 1995.

[21] Christopher J Bell, Ernest L. Lundelius, Anthony D. Barnosky, Russell W. Graham, Everett H. Lindsay, Dennis R. Ruez, Holmes A. Semken, S. David Webb, and Richard J. Zakrzewski. *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology*, chapter 7. The Blancan, Irvingtonian, and Rancholabrean Mammal Ages, pages 232–314. Columbia University Press, 2004.

[22] John F Benson, Jeff A Sikich, and Seth P D Riley. Individual and population level resource selection patterns of mountain lions preying on mule deer along an Urban-Wildland gradient. *PLoS One*, 11(7):e0158006, July 2016.

[23] David R. Bentley, Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, Dirk J. Evers, Colin L. Barnes, Helen R. Bignell, Jonathan M. Boutell, Jason Bryant, Richard J. Carter, R. Keira Cheetham, Anthony J. Cox, Darren J. Ellis, Michael R. Flatbush, Niall A. Gormley, Sean J. Humphray, Leslie J. Irving, Mirian S. Karbelashvili, Scott M. Kirk, Heng Li, Xiaohai Liu, Klaus S. Maisinger, Lisa J. Murray, Bojan Obradovic, Tobias Ost, Michael L. Parkinson, Mark R. Pratt, Isabelle M. J. Rasolonjatovo, Mark T. Reed, Roberto Rigatti, Chiara Rodighiero, Mark T. Ross, Andrea Sabot, Subramanian V. Sankar, Aylwyn Scally, Gary P. Schroth, Mark E. Smith, Vincent P. Smith, Anastassia Spiridou, Peta E. Torrance, Svilen S. Tzonev, Eric H. Vermaas, Klaudia Walter, Xiaolin Wu, Lu Zhang, Mohammed D. Alam,

Carole Anastasi, Ify C. Aniebo, David M. D. Bailey, Iain R. Bancarz, Saibal Banerjee, Selena G. Barbour, Primo A. Baybayan, Vincent A. Benoit, Kevin F. Benson, Claire Bevis, Phillip J. Black, Asha Boodhun, Joe S. Brennan, John A. Bridgham, Rob C. Brown, Andrew A. Brown, Dale H. Buermann, Abass A. Bundu, James C. Burrows, Nigel P. Carter, Nestor Castillo, Maria Chiara E. Catenazzi, Simon Chang, R. Neil Cooley, Natasha R. Crake, Olubunmi O. Dada, Konstantinos D. Diakoumakos, Belen Dominguez-Fernandez, David J. Earnshaw, Ugonna C. Egbujor, David W. Elmore, Sergey S. Etchin, Mark R. Ewan, Milan Fedurco, Louise J. Fraser, Karin V. Fuentes Fajardo, W. Scott Furey, David George, Kimberley J. Gietzen, Colin P. Goddard, George S. Golda, Philip A. Granieri, David E. Green, David L. Gustafson, Nancy F. Hansen, Kevin Harnish, Christian D. Haudenschild, Narinder I. Heyer, Matthew M. Hims, Johnny T. Ho, Adrian M. Horgan, Katya Hoschler, Steve Hurwitz, Denis V. Ivanov, Maria Q. Johnson, Terena James, T. A. Huw Jones, Gyoung-Dong Kang, Tzvetana H. Kerelska, Alan D. Kersey, Irina Khrebtukova, Alex P. Kindwall, Zoya Kingsbury, Paula I. Kokko-Gonzales, Anil Kumar, Marc A. Laurent, Cynthia T. Lawley, Sarah E. Lee, Xavier Lee, Arnold K. Liao, Jennifer A. Loch, Mitch Lok, Shujun Luo, Radhika M. Mammen, John W. Martin, Patrick G. McCauley, Paul McNitt, Parul Mehta, Keith W. Moon, Joe W. Mullens, Taksina Newington, Zemin Ning, Bee Ling Ng, Sonia M. Novo, Michael J. O'Neill, Mark A. Osborne, Andrew Osnowski, Omead Ostadan, Lambros L. Paraschos, Lea Pickering, Andrew C. Pike, Alger C. Pike, D. Chris Pinkard, Daniel P. Pliskin, Joe Podhasky, Victor J. Quijano, Come Raczy, Vicki H. Rae, Stephen R. Rawlings, Ana Chiva Rodriguez, Phyllida M. Roe, John Rogers, Maria C. Rogert Bacigalupo, Nikolai Romanov, Anthony Romieu, Rithy K. Roth, Natalie J. Rourke, Silke T. Ruediger, Eli Rusman, Raquel M. Sanches-Kuiper, Martin R. Schenker, Josefina M. Seoane, Richard J. Shaw, Mitch K. Shiver, Steven W. Short, Ning L. Sizto, Johannes P. Sluis, Melanie A. Smith, Jean Ernest Sohna Sohna, Eric J. Spence, Kim Stevens, Neil Sutton, Lukasz Szajkowski, Carolyn L. Tregidgo, Gerardo Turcatti, Stephanie vandeVondele, Yuli Verhovsky, Selene M. Virk, Suzanne Wakelin, Gregory C. Walcott, Jingwen Wang, Graham J. Worsley, Juying Yan, Ling Yau, Mike Zuerlein, Jane Rogers, James C. Mullikin, Matthew E. Hurles, Nick J. McCooke, John S. West, Frank L. Oaks, Peter L. Lundberg, David Klenerman, Richard Durbin, and Anthony J. Smith. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–59, November 2008.

[24] Matthias Bernt, Alexander Donath, Frank Jühling, Fabian Externbrink, Catherine Florentz, Guido Fritzsch, Joern Pütz, Martin Middendorf, and Peter F Stadler. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.*, 69(2):313–319, November 2013.

[25] Tobias Bidon, Nancy Schreck, Frank Hailer, Maria A. Nilsson, and Axel Janke. Genome-Wide Search Identifies 1.9 Mb from the Polar Bear Y Chromosome for Evolutionary Analyses. *Genome Biology and Evolution*, 7(7):2010–2022, July 2015.

[26] Frederick R. Blattner, Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F.

Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, and Ying Shao. The Complete Genome Sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462, September 1997.

[27] Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, August 2014.

[28] Céline Bon, Nicolas Caudy, Maud de Dieuleveult, Philippe Fosse, Michel Philippe, Frédéric Maksud, Éliane Beraud-Colomb, Eric Bouzaid, Rym Kefi, Christelle Laugier, Bernard Rousseau, Didier Casane, Johannes van der Plicht, and Jean-Marc Elalouf. Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proceedings of the National Academy of Sciences*, 105(45):17447–17452, November 2008.

[29] Kirsten I. Bos, Verena J. Schuenemann, G. Brian Golding, Hernán A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, Sharon N. DeWitte, Matthias Meyer, Sarah Schmedes, James Wood, David J. D. Earn, D. Ann Herring, Peter Bauer, Hendrik N. Poinar, and Johannes Krause. A draft genome of Yersinia pestis from victims of the Black Death. *Nature*, 478(7370):506–510, October 2011.

[30] Adrian W. Briggs, Udo Stenzel, Philip L. F. Johnson, Richard E. Green, Janet Kelso, Kay Prüfer, Matthias Meyer, Johannes Krause, Michael T. Ronan, Michael Lachmann, and Svante Pääbo. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, 104(37):14616–14621, September 2007.

[31] T. A. Brown, D. E. Nelson, J. S. Vogel, and J. R. Southon. Improved Collagen Extraction by Modified Longin Method. *Radiocarbon*, 30(2):171–177, 1988/ed.

[32] Joshua N. Burton, Andrew Adey, Rupali P. Patwardhan, Ruolan Qiu, Jacob O. Kitzman, and Jay Shendure. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, 31(12):1119–1125, December 2013.

[33] James A. Cahill, Richard E. Green, Tara L. Fulton, Mathias Stiller, Flora Jay, Nikita Ovsyanikov, Rauf Salamzade, John St John, Ian Stirling, Montgomery Slatkin, and Beth Shapiro. Genomic Evidence for Island Population Conversion Resolves Conflicting Theories of Polar Bear Evolution. *PLOS Genet*, 9(3):e1003345, March 2013.

[34] James A. Cahill, Peter D. Heintzman, Kelley Harris, Matthew D. Teasdale, Joshua Kapp, Andre E. R. Soares, Ian Stirling, Daniel Bradley, Ceiridwen J. Edwards, Kiley Graim, Aliaksandr A. Kisleika, Alexander V. Malev, Nigel Monaghan, Richard E. Green, and Beth Shapiro. Genomic Evidence of Widespread Admixture from Polar Bears into Brown Bears during the Last Ice Age. *Molecular Biology and Evolution*, 35(5):1120–1129, May 2018.

[35] James A Cahill, André E R Soares, Richard E Green, and Beth Shapiro. Inferring species divergence times using pairwise sequential markovian coalescent modelling and low-coverage genomic data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 371(1699), July 2016.

[36] James A. Cahill, Ian Stirling, Logan Kistler, Rauf Salamzade, Erik Ersmark, Tara L. Fulton, Mathias Stiller, Richard E. Green, and Beth Shapiro. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, 24(6):1205–1217, March 2015.

[37] Francisco C Ceballos, Peter K Joshi, David W Clark, Michèle Ramsay, and James F Wilson. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.*, 19(4):220–234, April 2018.

[38] Christopher C Chang, Carson C Chow, Laurent Cam Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), 2015.

[39] Jarrod A Chapman, Isaac Ho, Sirisha Sunkara, Shujun Luo, Gary P Schroth, and Daniel S Rokhsar. Meraculous: de novo genome assembly with short paired-end reads. *PLoS One*, 6(8):e23501, August 2011.

[40] Nicolás R Chimento and Alejandro Dondas. First record of puma concolor (mammalia, felidae) in the Early-Middle pleistocene of south america. *J. Mamm. Evol.*, 2017.

[41] Lynda Chin, Jannik N. Andersen, and P. Andrew Futreal. Cancer genomics: From discovery science to personalized medicine. *Nature Medicine*, 17(3):297–303, March 2011.

[42] Asif T. Chinwalla, Lisa L. Cook, Kimberly D. Delehaunty, Ginger A. Fewell, Lucinda A. Fulton, Robert S. Fulton, Tina A. Graves, LaDeana W. Hillier, Elaine R. Mardis, John D. McPherson, Tracie L. Miner, William E. Nash, Joanne O. Nelson, Michael N. Nhan, Kymberlie H. Pepin, Craig S. Pohl, Tracy C. Ponce, Brian Schultz, Johanna Thompson, Evanne Trevaskis, Robert H. Waterston, Michael C. Wendl, Richard K. Wilson, Shiaw-Pyng Yang, Peter An, Eric Berry, Bruce Birren, Toby Bloom, Daniel G. Brown, Jonathan Butler, Mark Daly, Robert David, Justin Deri, Sheila Dodge, Karen Foley, Diane Gage, Sante Gnerre, Timothy Holzer, David B. Jaffe, Michael Kamal, Elinor K. Karlsson, Cristyn Kells, Andrew Kirby, Edward J. Kulbokas, Eric S. Lander, Tom Landers, J. P. Leger, Rosie Levine, Kerstin Lindblad-Toh, Evan Mauceli, John H. Mayer, Megan McCarthy, Jim Meldrim, Jill P. Mesirov, Robert Nicol, Chad Nusbaum, Steven Seaman, Ted Sharpe, Andrew Sheridan, Jonathan B. Singer, Ralph Santos, Brian Spencer, Nicole Stange-Thomann, Jade P. Vinson, Claire M. Wade, Jamey Wierzbowski, Dudley Wyman, Michael C. Zody, Ewan Birney, Nick Goldman, Arkadiusz Kasprzyk, Emmanuel Mongin, Alistair G. Rust, Guy Slater, Arne Stabenau, Abel Ureta-Vidal, Simon Whelan, Rachel Ainscough, John Attwood, Jonathon Bailey, Karen Barlow, Stephan Beck, John Burton, Michele Clamp, Christopher Clee, Alan Coulson, James Cuff, Val Curwen, Tim Cutts, Joy Davies, Eduardo Eyras, Darren Grafham, Simon Gregory, Tim Hubbard, Adrienne Hunt, Matthew Jones, Ann Joy, Steven Leonard, Christine Lloyd,

Lucy Matthews, Stuart McLaren, Kirsten McLay, Beverley Meredith, James C. Mullikin, Zemin Ning, Karen Oliver, Emma Overton-Larty, Robert Plumb, Simon Potter, Michael Quail, Jane Rogers, Carol Scott, Steve Searle, Ratna Shownkeen, Sarah Sims, Melanie Wall, Anthony P. West, David Willey, Sophie Williams, Josep F. Abril, Roderic Guigó, Genís Parra, Pankaj Agarwal, Richa Agarwala, Deanna M. Church, Wratko Hlavina, Donna R. Maglott, Victor Sapojnikov, Marina Alexandersson, Lior Pachter, Stylianos E. Antonarakis, Emmanouil T. Dermitzakis, Alexandre Reymond, Catherine Ucla, Robert Baertsch, Mark Diekhans, Terrence S. Furey, Angela Hinrichs, Fan Hsu, Donna Karolchik, W. James Kent, Krishna M. Roskin, Matthias S. Schwartz, Charles Sugnet, Ryan J. Weber, Peer Bork, Ivica Letunic, Mikita Suyama, David Torrents, Evgeny M. Zdobnov, Marc Botcherby, Stephen D. Brown, Robert D. Campbell, Ian Jackson, Nicolas Bray, Olivier Couronne, Inna Dubchak, Alex Poliakov, Edward M. Rubin, Michael R. Brent, Paul Flicek, Evan Keibler, Ian Korf, S. Batalov, Carol Bult, Wayne N. Frankel, Piero Carninci, Yoshihide Hayashizaki, Jun Kawai, Yasushi Okazaki, Simon Cawley, David Kulp, Raymond Wheeler, Francesca Chiaromonte, Francis S. Collins, Adam Felsenfeld, Mark Guyer, Jane Peterson, Kris Wetterstrand, Richard R. Copley, Richard Mott, Colin Dewey, Nicholas J. Dickens, Richard D. Emes, Leo Goodstadt, Chris P. Ponting, Eitan Winter, Diane M. Dunn, Andrew C. von Niederhausern, Robert B. Weiss, Sean R. Eddy, L. Steven Johnson, Thomas A. Jones, Laura Elnitski, Diana L. Kolbe, Pallavi Eswara, Webb Miller, Michael J. O'Connor, Scott Schwartz, Richard A. Gibbs, Donna M. Muzny, Gustavo Glusman, Arian Smit, Eric D. Green, Ross C. Hardison, Shan Yang, David Haussler, Axin Hua, Bruce A. Roe, Raju S. Kucherlapati, Kate T. Montgomery, Jia Li, Ming Li, Susan Lucas, Bin Ma, W. Richard McCombie, Michael Morgan, Pavel Pevzner, Glenn Tesler, Jörg Schultz, Douglas R. Smith, John Tromp, Kim C. Worley, and Eric D. Green. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562, December 2002.

[43] Yun Sung Cho, Li Hu, Haolong Hou, Hang Lee, Jiaohui Xu, Soowhan Kwon, Sukhun Oh, Hak-Min Kim, Sungwoong Jho, Sangsoo Kim, Young-Ah Shin, Byung Chul Kim, Hyunmin Kim, Chang-Uk Kim, Shu-Jin Luo, Warren E Johnson, Klaus-Peter Koepfli, Anne Schmidt-Küntzel, Jason A Turner, Laurie Marker, Cindy Harper, Susan M Miller, Wilhelm Jacobs, Laura D Bertola, Tae Hyung Kim, Sunghoon Lee, Qian Zhou, Hyun-Ju Jung, Xiao Xu, Priyvrat Gadhvi, Pengwei Xu, Yingqi Xiong, Yadan Luo, Shengkai Pan, Caiyun Gou, Xiuhui Chu, Jilin Zhang, Sanyang Liu, Jing He, Ying Chen, Linfeng Yang, Yulan Yang, Jiaju He, Sha Liu, Junyi Wang, Chul Hong Kim, Hwanjong Kwak, Jong-Soo Kim, Seungwoo Hwang, Junsu Ko, Chang-Bae Kim, Sangtae Kim, Damdin Bayarlkhagva, Woon Kee Paek, Seong-Jin Kim, Stephen J O'Brien, Jun Wang, and Jong Bhak. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.*, 4:2433, 2013.

[44] Per Christiansen. What size were Arctodus simus and Ursus spelaeus (Carnivora: Ursidae)? *Annales Zoologici Fennici*, 36(2):93–102, 1999.

[45] The C. elegans Sequencing Consortium*. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396):2012–2018, December 1998.

[46] Alan Cooper and Hendrik N. Poinar. Ancient DNA: Do It Right or Not at All. *Science*, 289(5482):1139–1139, August 2000.

[47] Russell Corbett-Detig and Rasmus Nielsen. A hidden markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLoS Genet.*, 13(1):e1006529, January 2017.

[48] Russell B Corbett-Detig. Heterozygosity_HMM. `https://github.com/russcd/Heterozygosity_HMM`.

[49] Matthew A. Cronin, Steven C. Amstrup, Gerald W. Garner, and Ernest R. Vyse. Interspecific and intraspecific mitochondrial DNA variation in North American bears (*Ursus*). *Canadian Journal of Zoology*, 69(12):2985–2992, December 1991.

[50] M. Culver, W. E. Johnson, J. Pecon-Slattery, and S. J. OBrien. Genomic ancestry of the american puma (puma concolor). *Journal of Heredity*, 91(3):186–197, May 2000.

[51] Jesse Dabney, Michael Knapp, Isabelle Glocke, Marie-Theres Gansauge, Antje Weihmann, Birgit Nickel, Cristina Valdiosera, Nuria García, Svante Pääbo, Juan-Luis Arsuaga, and Matthias Meyer. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39):15758–15763, September 2013.

[52] R. Dale Guthrie. Rapid body size decline in Alaskan Pleistocene horses before extinction. *Nature*, 426(6963):169–171, November 2003.

[53] R. Dale Guthrie. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature*, 441(7090):207–209, May 2006.

[54] Kevin G. Daly, Pierpaolo Maisano Delser, Victoria E. Mullin, Amelie Scheu, Valeria Mattiangeli, Matthew D. Teasdale, Andrew J. Hare, Joachim Burger, Marta Pereira Verdugo, Matthew J. Collins, Ron Kehati, Cevdet Merih Erek, Guy Bar-Oz, François Pompanon, Tristan Cumer, Canan Çakırlar, Azadeh Fatemeh Mohaseb, Delphine Decruyenaere, Hossein Davoudi, Özlem Çevik, Gary Rollefson, Jean-Denis Vigne, Roya Khazaeli, Homa Fathi, Sanaz Beizaee Doost, Roghayeh Rahimi Sorkhani, Ali Akbar Vahdati, Eberhard W. Sauer, Hossein Azizi Kharanaghi, Sepideh Maziar, Boris Gasparian, Ron Pinhasi, Louise Martin, David Orton, Benjamin S. Arbuckle, Norbert Benecke, Andrea Manica, Liora Kolska Horwitz, Marjan Mashkour, and Daniel G. Bradley. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, 361(6397):85–88, July 2018.

[55] Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E Handsaker, Gerton Lunter, Gabor T Marth, Stephen T Sherry, Gilean

McVean, Richard Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, August 2011.

[56] Diego Darriba, Guillermo L Taboada, Ramón Doallo, and David Posada. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*, 9(8):772, July 2012.

[57] S. David Webb. The great american biotic interchange: Patterns and processes. *Annals of the Missouri Botanical Garden*, 93(2):245–257, August 2006.

[58] John Davison, Simon Y. W. Ho, Sarah C. Bray, Marju Korsten, Egle Tammeleht, Maris Hindrikson, Kjartan Østbye, Eivind Østbye, Stein-Erik Lauritzen, Jeremy Austin, Alan Cooper, and Urmas Saarma. Late-Quaternary biogeographic scenarios for the brown bear (Ursus arctos), a wild mammal model species. *Quaternary Science Reviews*, 30(3–4):418–430, February 2011.

[59] Cesare de Filippo, Matthias Meyer, and Kay Prüfer. Quantifying and reducing spurious alignments for the analysis of ultra-short ancient DNA sequences. *BMC Biology*, 16(1):121, October 2018.

[60] A. P. Jason de Koning, Wanjun Gu, Todd A. Castoe, Mark A. Batzer, and David D. Pollock. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLOS Genetics*, 7(12):e1002384, December 2011.

[61] Isabelle Delisle and Curtis Strobeck. Conserved Primers for Rapid Sequencing of the Complete Mitochondrial Genome from Carnivores, Applied to Three Species of Bears. *Molecular Biology and Evolution*, 19(3):357–361, March 2002.

[62] Xinxian Deng, Wenxiu Ma, Vijay Ramani, Andrew Hill, Fan Yang, Ferhat Ay, Joel B. Berletch, Carl Anthony Blau, Jay Shendure, Zhijun Duan, William S. Noble, and Christine M. Disteche. Bipartite structure of the inactive mouse X chromosome. *Genome Biology*, 16(1):152, August 2015.

[63] Michael J Deniro and Samuel Epstein. Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta*, 45(3):341–351, March 1981.

[64] Clio Der Sarkissian, Morten E. Allentoft, María C. Ávila-Arcos, Ross Barnett, Paula F. Campos, Enrico Cappellini, Luca Ermini, Ruth Fernández, Rute da Fonseca, Aurélien Ginolhac, Anders J. Hansen, Hákon Jónsson, Thorfinn Korneliussen, Ashot Margaryan, Michael D. Martin, J. Víctor Moreno-Mayar, Maanasa Raghavan, Morten Rasmussen, Marcela Sandoval Velasco, Hannes Schroeder, Mikkel Schubert, Andaine Seguin-Orlando, Nathan Wales, M. Thomas P. Gilbert, Eske Willerslev, and Ludovic Orlando. Ancient genomics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660):20130387, January 2015.

[65] Pavel Dobrynin, Shiping Liu, Gaik Tamazian, Zijun Xiong, Andrey A Yurchenko, Ksenia Krasheninnikova, Sergey Kliver, Anne Schmidt-Küntzel, Klaus-Peter Koepfli, Warren Johnson, Lukas F K Kuderna, Raquel García-Pérez, Marc de Manuel, Ricardo Godinez,

Aleksey Komissarov, Alexey Makunin, Vladimir Brukhin, Weilin Qiu, Long Zhou, Fang Li, Jian Yi, Carlos Driscoll, Agostinho Antunes, Taras K Oleksyk, Eduardo Eizirik, Polina Perelman, Melody Roelke, David Wildt, Mark Diekhans, Tomas Marques-Bonet, Laurie Marker, Jong Bhak, Jun Wang, Guojie Zhang, and Stephen J O'Brien. Genomic legacy of the african cheetah, acinonyx jubatus. *Genome Biol.*, 16:277, December 2015.

[66] A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, 22(5):1185–1192, May 2005.

[67] Beth L Dumont and Bret A Payseur. Evolution of the genomic rate of recombination in mammals. *Evolution*, 62(2):276–294, February 2008.

[68] Eric Y. Durand, Nick Patterson, David Reich, and Montgomery Slatkin. Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, 28(8):2239–2252, August 2011.

[69] Richard M. Durbin, David Altshuler, Richard M. Durbin, Gonçalo R. Abecasis, David R. Bentley, Aravinda Chakravarti, Andrew G. Clark, Francis S. Collins, Francisco M. De La Vega, Peter Donnelly, Michael Egholm, Paul Flicek, Stacey B. Gabriel, Richard A. Gibbs, Bartha M. Knoppers, Eric S. Lander, Hans Lehrach, Elaine R. Mardis, Gil A. McVean, Deborah A. Nickerson, Leena Peltonen, Alan J. Schafer, Stephen T. Sherry, Jun Wang, Richard K. Wilson, Richard A. Gibbs, David Deiros, Mike Metzker, Donna Muzny, Jeff Reid, David Wheeler, Jun Wang, Jingxiang Li, Min Jian, Guoqing Li, Ruiqiang Li, Huiqing Liang, Geng Tian, Bo Wang, Jian Wang, Wei Wang, Huanming Yang, Xiuqing Zhang, Huisong Zheng, Eric S. Lander, David Altshuler, Lauren Ambrogio, Toby Bloom, Kristian Cibulskis, Tim J. Fennell, Stacey B. Gabriel, David B. Jaffe, Erica Shefler, Carrie L. Sougnez, David R. Bentley, Niall Gormley, Sean Humphray, Zoya Kingsbury, Paula Kokko-Gonzales, Jennifer Stone, Kevin J. McKernan, Gina L. Costa, Jeffry K. Ichikawa, Clarence C. Lee, Ralf Sudbrak, Hans Lehrach, Tatiana A. Borodina, Andreas Dahl, Alexey N. Davydov, Peter Marquardt, Florian Mertes, Wilfiried Nietfeld, Philip Rosenstiel, Stefan Schreiber, Aleksey V. Soldatov, Bernd Timmermann, Marius Tolzmann, Michael Egholm, Jason Affourtit, Dana Ashworth, Said Attiya, Melissa Bachorski, Eli Buglione, Adam Burke, Amanda Caprio, Christopher Celone, Shauna Clark, David Conners, Brian Desany, Lisa Gu, Lorri Guccione, Kalvin Kao, Andrew Kebbel, Jennifer Knowlton, Matthew Labrecque, Louise McDade, Craig Mealmaker, Melissa Minderman, Anne Nawrocki, Faheem Niazi, Kristen Pareja, Ravi Ramenani, David Riches, Wanmin Song, Cynthia Turcotte, Shally Wang, Elaine R. Mardis, Richard K. Wilson, David Dooling, Lucinda Fulton, Robert Fulton, George Weinstock, Richard M. Durbin, John Burton, David M. Carter, Carol Churcher, Alison Coffey, Anthony Cox, Aarno Palotie, Michael Quail, Tom Skelly, James Stalker, Harold P. Swerdlow, Daniel Turner, Anniek De Witte, Shane Giles, Richard A. Gibbs, David Wheeler, Matthew Bainbridge, Danny Challis, Aniko Sabo, Fuli Yu, Jin Yu, Jun Wang, Xiaodong Fang, Xiaosen Guo, Ruiqiang Li, Yingrui Li, Ruibang Luo, Shuaishuai Tai, Honglong Wu, Hancheng Zheng, Xiaole Zheng, Yan Zhou, Guoqing Li, Jian Wang,

Huanming Yang, Gabor T. Marth, Erik P. Garrison, Weichun Huang, Amit Indap, Deniz Kural, Wan-Ping Lee, Wen Fung Leong, Aaron R. Quinlan, Chip Stewart, Michael P. Stromberg, Alistair N. Ward, Jiantao Wu, Charles Lee, Ryan E. Mills, Xinghua Shi, Mark J. Daly, Mark A. DePristo, David Altshuler, Aaron D. Ball, Eric Banks, Toby Bloom, Brian L. Browning, Kristian Cibulskis, Tim J. Fennell, Kiran V. Garimella, Sharon R. Grossman, Robert E. Handsaker, Matt Hanna, Chris Hartl, David B. Jaffe, Andrew M. Kernytsky, Joshua M. Korn, Heng Li, Jared R. Maguire, Steven A. McCarroll, Aaron McKenna, James C. Nemesh, Anthony A. Philippakis, Ryan E. Poplin, Alkes Price, Manuel A. Rivas, Pardis C. Sabeti, Stephen F. Schaffner, Erica Shefler, Ilya A. Shlyakhter, David N. Cooper, Edward V. Ball, Matthew Mort, Andrew D. Phillips, Peter D. Stenson, Jonathan Sebat, Vladimir Makarov, Kenny Ye, Seungtai C. Yoon, Carlos D. Bustamante, Laura Clarke, Paul Flicek, Fiona Cunningham, Javier Herrero, Stephen Keenen, Eugene Kulesha, Rasko Leinonen, William M. McLaren, Rajesh Radhakrishnan, Richard E. Smith, Vadim Zalunin, Xiangqun Zheng-Bradley, Jan O. Korbel, Adrian M. Stütz, Sean Humphray, Markus Bauer, R. Keira Cheetham, Tony Cox, Michael Eberle, Terena James, Scott Kahn, Lisa Murray, Aravinda Chakravarti, Kai Ye, Francisco M. De La Vega, Yutao Fu, Fiona C. L. Hyland, Jonathan M. Manning, Stephen F. McLaughlin, Heather E. Peckham, Onur Sakarya, Yongming A. Sun, Eric F. Tsung, Mark A. Batzer, Miriam K. Konkel, Jerilyn A. Walker, Ralf Sudbrak, Marcus W. Albrecht, Vyacheslav S. Amstislavskiy, Ralf Herwig, Dimitri V. Parkhomchuk, Stephen T. Sherry, Richa Agarwala, Hoda M. Khouri, Aleksandr O. Morgulis, Justin E. Paschall, Lon D. Phan, Kirill E. Rotmistrovsky, Robert D. Sanders, Martin F. Shumway, Chunlin Xiao, Gil A. McVean, Adam Auton, Zamin Iqbal, Gerton Lunter, Jonathan L. Marchini, Loukas Moutsianas, Simon Myers, Afidalina Tumian, Brian Desany, James Knight, Roger Winer, David W. Craig, Steve M. Beckstrom-Sternberg, Alexis Christoforides, The 1000 Genomes Project Consortium, Corresponding author, Steering committee, Production group: Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, Illumina, Life Technologies, Max Planck Institute for Molecular Genetics, Roche Applied Science, Washington University in St Louis, Wellcome Trust Sanger Institute, Analysis group: Agilent Technologies, Baylor College of Medicine, Boston College, Brigham and Women's Hospital, The Human Gene Mutation Database Cardiff University, Cold Spring Harbor Laboratory, Cornell and Stanford Universities, European Bioinformatics Institute, European Molecular Biology Laboratory, Johns Hopkins University, Leiden University Medical Center, Louisiana State University, US National Institutes of Health, Oxford University, and The Translational Genomics Research Institute. A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–1073, October 2010.

[70] A S Dyke, J T Andrews, P U Clark, J H England, G H Miller, J Shaw, and J J Veillette. The laurentide and innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews*, 21(1-3):9–31, 2002.

[71] Robert C Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5):1792–1797, March 2004.

[72] Juergen Ehlers, Philip L. Gibbard, and Philip D. Hughes. QUATERNARY GLACIATIONS - EXTENT AND CHRONOLOGY A Closer Look Introduction. In J. Ehlers, P. L. Gibbard, and P. D. Hughes, editors, *Quaternary Glaciations - Extent and Chronology: A Closer Look*, volume 15, pages 1–14. Elsevier Science Bv, Amsterdam, 2011.

[73] L Mark Elbroch and Anna Kusler. Are pumas subordinate carnivores, and does it matter? *PeerJ*, 6:e4293, January 2018.

[74] Scott A. Elias, Susan K. Short, C. Hans Nelson, and Hilary H. Birks. Life and times of the Bering land bridge. *Nature*, 382(6586):60–63, July 1996.

[75] S.D. Emslie and N.J. Czaplewski. A new record of giant short-faced bear, Arctodus simus, from western North America with a reevaluation of its paleobiology. *Contributions in Science*, (371):1–12, November 1985.

[76] Adam C. English, Stephen Richards, Yi Han, Min Wang, Vanesa Vee, Jiaxin Qu, Xiang Qin, Donna M. Muzny, Jeffrey G. Reid, Kim C. Worley, and Richard A. Gibbs. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE*, 7(11):e47768, November 2012.

[77] Adam C English, Stephen Richards, Yi Han, Min Wang, Vanesa Vee, Jiaxin Qu, Xiang Qin, Donna M Muzny, Jeffrey G Reid, Kim C Worley, and Richard A Gibbs. Mind the gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. *PLoS One*, 7(11):e47768, November 2012.

[78] Holly B Ernest, T Winston Vickers, Scott A Morrison, Michael R Buchalski, and Walter M Boyce. Fractured genetic connectivity threatens a southern california puma (puma concolor) population. *PLoS One*, 9(10):e107985, October 2014.

[79] G Evanno, S Regnaut, and J Goudet. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, 14(8):2611–2620, July 2005.

[80] Antoine Fages, Kristian Hanghøj, Naveed Khan, Charleen Gaunitz, Andaine Seguin-Orlando, Michela Leonardi, Christian McCrory Constantz, Cristina Gamba, Khaled A. S. Al-Rasheid, Silvia Albizuri, Ahmed H. Alfarhan, Morten Allentoft, Saleh Alquraishi, David Anthony, Nurbol Baimukhanov, James H. Barrett, Jamsranjav Bayarsaikhan, Norbert Benecke, Eloísa Bernáldez-Sánchez, Luis Berrocal-Rangel, Fereidoun Biglari, Sanne Boessenkool, Bazartseren Boldgiv, Gottfried Brem, Dorcas Brown, Joachim Burger, Eric Crubézy, Linas Daugnora, Hossein Davoudi, Peter de Barros Damgaard, María de los Ángeles de Chorro y de Villa-Ceballos, Sabine Deschler-Erb, Cleia Detry, Nadine Dill, Maria do Mar Oom, Anna Dohr, Sturla Ellingvåg, Diimaajav Erdenebaatar, Homa Fathi, Sabine Felkel, Carlos Fernández-Rodríguez, Esteban García-Viñas, Mietje Germonpré, José D. Granado, Jón H. Hallsson, Helmut Hemmer, Michael Hofreiter, Aleksei Kasparov, Mutalib Khasanov, Roya Khazaeli, Pavel Kosintsev, Kristian Kristiansen, Tabaldiev Kubatbek, Lukas Kuderna, Pavel Kuznetsov, Haeedeh Laleh, Jennifer A. Leonard, Johanna Lhuillier, Corina Liesau von Lettow-Vorbeck, Andrey Logvin,

Lembi Lõugas, Arne Ludwig, Cristina Luis, Ana Margarida Arruda, Tomas Marques-Bonet, Raquel Matoso Silva, Victor Merz, Enkhbayar Mijiddorj, Bryan K. Miller, Oleg Mochlov, Fatemeh A. Mohaseb, Arturo Morales, Ariadna Nieto-Espinet, Heidi Nistelberger, Vedat Onar, Albína H. Pálsdóttir, Vladimir Pitulko, Konstantin Pitskhelauri, Mélanie Pruvost, Petra Rajic Sikanjic, Anita Rapan Papeša, Natalia Roslyakova, Alireza Sardari, Eberhard Sauer, Renate Schafberg, Amelie Scheu, Jörg Schibler, Angela Schlumbaum, Nathalie Serrand, Aitor Serres-Armero, Beth Shapiro, Shiva Sheikhi Seno, Irina Shevnina, Sonia Shidrang, John Southon, Bastiaan Star, Naomi Sykes, Kamal Taheri, William Taylor, Wolf-Rüdiger Teegen, Tajana Trbojević Vukičević, Simon Trixl, Dashzeveg Tumen, Sainbileg Undrakhbold, Emma Usmanova, Ali Vahdati, Silvia Valenzuela-Lamas, Catarina Viegas, Barbara Wallner, Jaco Weinstock, Victor Zaibert, Benoit Clavel, Sébastien Lepetz, Marjan Mashkour, Agnar Helgason, Kári Stefánsson, Eric Barrey, Eske Willerslev, Alan K. Outram, Pablo Librado, and Ludovic Orlando. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, 0(0), May 2019.

[81] Daniel Falush, Matthew Stephens, and Jonathan K Pritchard. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587, August 2003.

[82] George A Feldhamer, Bruce C Thompson, and Joseph A Chapman. *Wild Mammals of North America: Biology, Management, and Conservation*. JHU Press, October 2003.

[83] Borja Figueirido, Juan A. Pérez-Claros, Vanessa Torregrosa, Alberto Martín-Serra, and Paul Palmqvist. Demythologizing Arctodus simus, the 'Short-Faced' Long-Legged and Predaceous Bear that Never Was. *Journal of Vertebrate Paleontology*, 30(1):262–275, January 2010.

[84] Jean-Franois Flot, Herv Marie-Nelly, and Romain Koszul. Contact genomics: scaffolding and phasing (meta)genomes using chromosome 3d physical signatures. *FEBS Letters*, 589(20PartA):29662974, 2015.

[85] Thibaud-Nissen Françoise, Souvorov Alexander, Murphy Terence, and Dicuccio,Michael. *Eukaryotic Genome Annotation Pipeline*, November 2013.

[86] R Frankham. Conservation genetics. *Annu. Rev. Genet.*, 29:305–327, 1995.

[87] Duane Froese, Mathias Stiller, Peter D. Heintzman, Alberto V. Reyes, Grant D. Zazula, Andr E. R. Soares, Matthias Meyer, Elizabeth Hall, Britta J. L. Jensen, Lee J. Arnold, and et al. Fossil and genomic evidence constrains the timing of bison arrival in north america. *Proceedings of the National Academy of Sciences*, 114(13):3457–3462, March 2017.

[88] Marie-Theres Gansauge, Tobias Gerber, Isabelle Glocke, Petra Korlević, Laurin Lippik, Sarah Nagel, Lara Maria Riehl, Anna Schmidt, and Matthias Meyer. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*, 45(10):e79–e79, June 2017.

[89] Shaenandhoa García-Rangel. Andean bear Tremarctos ornatus natural history and conservation. *Mammal Review*, 42(2):85–119, 2012.

[90] Erik Garrison and Gabor Marth. Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907 [q-bio]*, July 2012.

[91] Aurelien Ginolhac, Morten Rasmussen, M. Thomas P. Gilbert, Eske Willerslev, and Ludovic Orlando. mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics*, 27(15):2153–2155, August 2011.

[92] Isabelle Glocke and Matthias Meyer. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Research*, 27(7):1230–1237, July 2017.

[93] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 Genes. *Science*, 274(5287):546–567, October 1996.

[94] David Gordon, John Huddleston, Mark J. P. Chaisson, Christopher M. Hill, Zev N. Kronenberg, Katherine M. Munson, Maika Malig, Archana Raja, Ian Fiddes, LaDeana W. Hillier, Christopher Dunn, Carl Baker, Joel Armstrong, Mark Diekhans, Benedict Paten, Jay Shendure, Richard K. Wilson, David Haussler, Chen-Shan Chin, and Evan E. Eichler. Long-read sequence assembly of the gorilla genome. *Science*, 352(6281):aae0344, April 2016.

[95] Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10):3420–3435, June 2008.

[96] R.W. Graham and Jr. E.L. Lundelius. Faunmap ii: New data for north america with a temporal extension for the blancan, irvingtonian and early rancholabrean. `https://ucmp.berkeley.edu/faunmap/index.html`, 2010. data retrieved from FAUNMAP II Database, version 1.0.

[97] Richard E. Green, Johannes Krause, Susan E. Ptak, Adrian W. Briggs, Michael T. Ronan, Jan F. Simons, Lei Du, Michael Egholm, Jonathan M. Rothberg, Maja Paunovic, and Svante Pääbo. Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444(7117):330–336, November 2006.

[98] Richard E Green, Anna-Sapfo Malaspinas, Johannes Krause, Adrian W Briggs, Philip L F Johnson, Caroline Uhler, Matthias Meyer, Jeffrey M Good, Tomislav Maricic, Udo Stenzel, Kay Prüfer, Michael Siebauer, Hernán A Burbano, Michael Ronan, Jonathan M Rothberg, Michael Egholm, Pavao Rudan, Dejana Brajković, Zeljko Kućan, Ivan Gusić, Mårten Wikström, Liisa Laakkonen, Janet Kelso, Montgomery Slatkin, and Svante Pääbo. A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, 134(3):416–426, August 2008.

[99] A. D. Greenwood, C. Capelli, G. Possnert, and S. Pääbo. Nuclear DNA sequences from late Pleistocene megafauna. *Molecular Biology and Evolution*, 16(11):1466–1473, November 1999.

[100] Torsten Günther and Carl Nettelblad. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*, 15(7):e1008302, July 2019.

[101] Kyle D Gustafson, T Winston Vickers, Walter M Boyce, and Holly B Ernest. A single migrant enhances the genetic diversity of an inbred puma population. *Royal Society Open Science*, 4(5):170115, 2017.

[102] Erika Hagelberg, Mark G. Thomas, Charles E. Cook, Andrei V. Sher, Gennady F. Baryshnikov, and Adrian M. Lister. DNA from ancient mammoth bones. *Nature*, 370(6488):333–334, August 1994.

[103] Frank Hailer, Verena E. Kutschera, Björn M. Hallström, Denise Klassert, Steven R. Fain, Jennifer A. Leonard, Ulfur Arnason, and Axel Janke. Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. *Science*, 336(6079):344–347, April 2012.

[104] O. Handt, M. Krings, R. H. Ward, and S. Pääbo. The retrieval of ancient human DNA sequences. *American Journal of Human Genetics*, 59(2):368–376, August 1996.

[105] Peter D. Heintzman, Duane Froese, John W. Ives, André E. R. Soares, Grant D. Zazula, Brandon Letts, Thomas D. Andrews, Jonathan C. Driver, Elizabeth Hall, P. Gregory Hare, Christopher N. Jass, Glen MacKay, John R. Southon, Mathias Stiller, Robin Woywitka, Marc A. Suchard, and Beth Shapiro. Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29):8057–8063, July 2016.

[106] Godfrey Hewitt. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789):907–913, June 2000.

[107] GODFREY M Hewitt. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68(1):87–112, September 1999.

[108] Daisuke Hirata, Tsutomu Mano, Alexei V. Abramov, Gennady F. Baryshnikov, Pavel A. Kosintsev, Alexandr A. Vorobiev, Evgeny G. Raichev, Hiroshi Tsunoda, Yayoi Kaneko, Koichi Murata, Daisuke Fukui, and Ryuichi Masuda. Molecular Phylogeography of the Brown Bear (Ursus arctos) in Northeastern Asia Based on Analyses of Complete Mitochondrial DNA Sequences. *Molecular Biology and Evolution*, page mst077, April 2013.

[109] Marc P. Hoeppner, Andrew Lundquist, Mono Pirun, Jennifer R. S. Meadows, Neda Zamani, Jeremy Johnson, Görel Sundström, April Cook, Michael G. FitzGerald, Ross Swofford, Evan Mauceli, Behrooz Torabi Moghadam, Anna Greka, Jessica Alföldi, Amr

Abouelleil, Lynne Aftuck, Daniel Bessette, Aaron Berlin, Adam Brown, Gary Gearin, Annie Lui, J. Pendexter Macdonald, Margaret Priest, Terrance Shea, Jason Turner-Maier, Andrew Zimmer, Eric S. Lander, Federica di Palma, Kerstin Lindblad-Toh, and Manfred G. Grabherr. An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts. *PLOS ONE*, 9(3):e91172, March 2014.

[110] M. Hofreiter, V. Jaenicke, D. Serre, A. von Haeseler, and S. Pääbo. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, 29(23):4793–4799, December 2001.

[111] Henry Hooghiemstra and Thomas van der Hammen. Neogene and Quaternary development of the neotropical rain forest: The forest refugia hypothesis, and a literature overview. *Earth-Science Reviews*, 44(3):147–183, September 1998.

[112] Maurice Hornocker and Sharon Negri. *Cougar: Ecology and Conservation*. University of Chicago Press, December 2009.

[113] M. Hoss, P. Jaruga, T. H. Zastawny, M. Dizdaroglu, and S. Paabo. DNA Damage and DNA Sequence Retrieval from Ancient Tissues. *Nucleic Acids Research*, 24(7):1304–1307, April 1996.

[114] Matthias Hoss, Svante Paabo, and N. K. Vereshchagin. Mammoth DNA sequences. *Nature*, 370(6488):333–333, August 1994.

[115] R R Hudson. Gene genealogies and the coalescent process. In Futuyma, D , Antonovics,, editor, *Oxford Surveys in Evolutionary Biology*, volume 7, pages 1–44. Oxford University Press, 1990.

[116] Thomas J. Hudson, Lincoln D. Stein, Sebastian S. Gerety, Junli Ma, Andrew B. Castle, James Silva, Donna K. Slonim, Rafael Baptista, Leonid Kruglyak, Shu-Hua Xu, Xintong Hu, Angela M. E. Colbert, Carl Rosenberg, Mary Pat Reeve-Daly, Steve Rozen, Lester Hui, Xiaoyun Wu, Christina Vestergaard, Kimberly M. Wilson, Jane S. Bae, Shanak Maitra, Soula Ganiatsas, Cheryl A. Evans, Margaret M. DeAngelis, Kimberly A. Ingalls, Robert W. Nahf, Lloyd T. Horton, Michele Oskin Anderson, Alville J. Collymore, Wenjuan Ye, Vardouhie Kouyoumjian, Irena S. Zemsteva, James Tam, Richard Devine, Dorothy F. Courtney, Michelle Turner Renaud, Huy Nguyen, Tara J. O'Connor, Cécile Fizames, Sabine Fauré, Gabor Gyapay, Colette Dib, Jean Morissette, James B. Orlin, Bruce W. Birren, Nathan Goodman, Jean Weissenbach, Trevor L. Hawkins, Simon Foote, David C. Page, and Eric S. Lander. An STS-Based Map of the Human Genome. *Science*, 270(5244):1945–1954, December 1995.

[117] National Human Genome Research Institute. The Cost of Sequencing a Human Genome. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.

[118] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, October 2004.

[119] IUCN. Puma concolor: Nielsen, c., thompson, d., kelly, m. & Lopez-Gonzalez, C.A, 2014.

[120] IUCN. Tremarctos ornatus: Velez-liendo, x. & Garca-Rangel, S., 2017.

[121] jacahill. Jacahill/Admixture, February 2020.

[122] K Jae-Heup, E Eizirik, S J O'Brien, and W E Johnson. Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats. *Mitochondrion*, 1(3):279–292, October 2001.

[123] Viviane Jaenicke-Després, Ed S. Buckler, Bruce D. Smith, M. Thomas P. Gilbert, Alan Cooper, John Doebley, and Svante Pääbo. Early Allelic Selection in Maize as Revealed by Ancient DNA. *Science*, 302(5648):1206–1208, November 2003.

[124] Miten Jain, Sergey Koren, Karen H. Miga, Josh Quick, Arthur C. Rand, Thomas A. Sasani, John R. Tyson, Andrew D. Beggs, Alexander T. Dilthey, Ian T. Fiddes, Sunir Malla, Hannah Marriott, Tom Nieto, Justin O'Grady, Hugh E. Olsen, Brent S. Pedersen, Arang Rhie, Hollian Richardson, Aaron R. Quinlan, Terrance P. Snutch, Louise Tee, Benedict Paten, Adam M. Phillippy, Jared T. Simpson, Nicholas J. Loman, and Matthew Loose. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4):338–345, April 2018.

[125] Miten Jain, Hugh E Olsen, Benedict Paten, and Mark Akeson. The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.*, 17(1):239, November 2016.

[126] Mattias Jakobsson and Noah A Rosenberg. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14):1801–1806, July 2007.

[127] Warren E. Johnson, Eduardo Eizirik, Jill Pecon-Slattery, William J. Murphy, Agostinho Antunes, Emma Teeling, and Stephen J. OBrien. The late miocene radiation of modern felidae: A genetic assessment. *Science*, 311(5757):73–77, January 2006.

[128] Warren E Johnson, David P Onorato, Melody E Roelke, E Darrell Land, Mark Cunningham, Robert C Belden, Roy McBride, Deborah Jansen, Mark Lotz, David Shindle, Jogayle Howard, David E Wildt, Linda M Penfold, Jeffrey A Hostetler, Madan K Oli, and Stephen J O'Brien. Genetic restoration of the florida panther. *Science*, 329(5999):1641–1645, September 2010.

[129] Hákon Jónsson, Aurélien Ginolhac, Mikkel Schubert, Philip L. F. Johnson, and Ludovic Orlando. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29(13):1682–1684, July 2013.

[130] Hákon Jónsson, Aurélien Ginolhac, Mikkel Schubert, Philip L. F. Johnson, and Ludovic Orlando. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29(13):1682–1684, July 2013.

[131] Theodore S. Kalbfleisch, Edward S. Rice, Michael S. DePriest, Brian P. Walenz, Matthew S. Hestand, Joris R. Vermeesch, Brendan L. O′Connell, Ian T. Fiddes, Alisa O. Vershinina, Nedda F. Saremi, Jessica L. Petersen, Carrie J. Finno, Rebecca R. Bellone, Molly E. McCue, Samantha A. Brooks, Ernest Bailey, Ludovic Orlando, Richard E. Green, Donald C. Miller, Douglas F. Antczak, and James N. MacLeod. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, 1(1):1–8, November 2018.

[132] Joshua Kapp. Santa cruz library protocol.

[133] Marty Kardos, Mikael Åkesson, Toby Fountain, Øystein Flagstad, Olof Liberg, Pall Olason, Håkan Sand, Petter Wabakken, Camilla Wikenros, and Hans Ellegren. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol*, 2(1):124–131, January 2018.

[134] Kazutaka Katoh, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066, July 2002.

[135] Matthew Kearse, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, Chris Duran, Tobias Thierer, Bruce Ashton, Peter Meintjes, and Alexei Drummond. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649, June 2012.

[136] Peter Kerpedjiev, Nezar Abdennur, Fritz Lekschas, Chuck McCallum, Kasper Dinkla, Hendrik Strobelt, Jacob M Luber, Scott B Ouellette, Alaleh Azhir, Nikhil Kumar, Jeewon Hwang, Soohyun Lee, Burak H Alver, Hanspeter Pfister, Leonid A Mirny, Peter J Park, and Nils Gehlenborg. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.*, 19(1):125, August 2018.

[137] Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, April 2013.

[138] Paul L. Koch and Anthony D. Barnosky. Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):215–250, 2006.

[139] Jan O. Korbel, Alexander Eckehart Urban, Jason P. Affourtit, Brian Godwin, Fabian Grubert, Jan Fredrik Simons, Philip M. Kim, Dean Palejev, Nicholas J. Carriero, Lei Du, Bruce E. Taillon, Zhoutao Chen, Andrea Tanzer, A. C. Eugenia Saunders, Jianxiang Chi, Fengtang Yang, Nigel P. Carter, Matthew E. Hurles, Sherman M. Weissman, Timothy T. Harkins, Mark B. Gerstein, Michael Egholm, and Michael Snyder. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science*, 318(5849):420–426, October 2007.

[140] Petra Korlević, Tobias Gerber, Marie-Theres Gansauge, Mateja Hajdinjak, Sarah Nagel, Ayinuer Aximu-Petri, and Matthias Meyer. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques*, 59(2):87–93, August 2015.

[141] Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15(1):356, November 2014.

[142] Johannes Krause, Tina Unger, Aline Noon, Anna-Sapfo Malaspinas, Sergios-Orestis Kolokotronis, Mathias Stiller, Leopoldo Soibelzon, Helen Spriggs, Paul H. Dear, Adrian W. Briggs, and et al. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the miocene-pliocene boundary. *BMC Evolutionary Biology*, 8(1):220, Jul 2008.

[143] Martin I. Krzywinski, Jacqueline E. Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, June 2009.

[144] Vikas Kumar, Fritjof Lammers, Tobias Bidon, Markus Pfenninger, Lydia Kolter, Maria A. Nilsson, and Axel Janke. The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*, 7:46487, April 2017.

[145] B. Kurtén. *Pleistocene Bears of North America: Genus Arctodus, Short-Faced Bears*. Number v. 2 in Acta Zoologica Fennica. Societas pro fauna et flora Fennica, 1967.

[146] Björn Kurtén. Pleistocene bears of North America 2. Genus Arctodus, Short-Faced bears. March 1967.

[147] Verena E. Kutschera, Tobias Bidon, Frank Hailer, Julia L. Rodi, Steven R. Fain, and Axel Janke. Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. *Molecular Biology and Evolution*, 31(8):2004–2017, August 2014.

[148] Tianying Lan, Stephanie Gill, Eva Bellemain, Richard Bischof, Muhammad Ali Nawaz, and Charlotte Lindqvist. Evolutionary history of enigmatic bears in the Tibetan Plateau–Himalaya region and the identity of the yeti. *Proceedings of the Royal Society B: Biological Sciences*, 284(1868):20171804, December 2017.

[149] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, Roel Funke, Diane Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie LeVine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian, Dudley Wyman, Jane Rogers, John Sulston,

Rachael Ainscough, Stephan Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Shownkeen, Sarah Sims, Robert H. Waterston, Richard K. Wilson, LaDeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx, Sandra W. Clifton, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan-Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher, Marvin Frazier, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson, George M. Weinstock, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki, Todd Taylor, Jean Weissenbach, Roland Heilig, William Saurin, Francois Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert, Patrick Wincker, André Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien, Andreas Rump, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Hong Mei Lee, JoAnn Dubois, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang, Jun Gu, Leroy Hood, Lee Rowen, Anup Madan, Shizen Qin, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola, Michael J. Proctor, Bruce A. Roe, Feng Chen, Huaqin Pan, Juliane Ramser, Hans Lehrach, Richard Reinhardt, W. Richard McCombie, Melissa de la Bastide, Neilay Dedhia, Helmut Blöcker, Klaus Hornischer, Gabriele Nordsiek, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu-Chuan Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kaspryzk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran, Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, Jörg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw-Pyng Yang, Ru-Fang Yeh, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld, Kris A. Wetterstrand, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood, David R. Cox, Maynard V. Olson, Rajinder Kaul, Christopher Raymond, Nobuyoshi Shimizu, Kazuhiko Kawasaki, Shinsei Minoshima, Glen A. Evans, Maria Athanasiou, Roger Schultz, Aristides Patrinos, and Michael J. Morgan. Initial sequencing and analy-

sis of the human genome. *Nature*, 409(6822):860–921, February 2001.

[150] Robert Lanfear, Brett Calcott, Simon Y W Ho, and Stephane Guindon. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.*, 29(6):1695–1701, June 2012.

[151] Marcus Lechner, Sven Findeiß, Lydia Steiner, Manja Marz, Peter F. Stadler, and Sonja J. Prohaska. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, 12(1):124, April 2011.

[152] Tae-Ho Lee, Hui Guo, Xiyin Wang, Changsoo Kim, and Andrew H. Paterson. SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, 15(1):162, February 2014.

[153] Jennifer A. Leonard, Robert K. Wayne, and Alan Cooper. Population genetics of Ice Age brown bears. *Proceedings of the National Academy of Sciences*, 97(4):1651–1654, February 2000.

[154] Alia J. Lesnek, Jason P. Briner, Charlotte Lindqvist, James F. Baichtal, and Timothy H. Heaton. Deglaciation of the Pacific coastal corridor directly preceded the human colonization of the Americas. *Science Advances*, 4(5):eaar5040, May 2018.

[155] Samuel Levy, Granger Sutton, Pauline C. Ng, Lars Feuk, Aaron L. Halpern, Brian P. Walenz, Nelson Axelrod, Jiaqi Huang, Ewen F. Kirkness, Gennady Denisov, Yuan Lin, Jeffrey R. MacDonald, Andy Wing Chun Pang, Mary Shago, Timothy B. Stockwell, Alexia Tsiamouri, Vineet Bafna, Vikas Bansal, Saul A. Kravitz, Dana A. Busam, Karen Y. Beeson, Tina C. McIntosh, Karin A. Remington, Josep F. Abril, John Gill, Jon Borman, Yu-Hui Rogers, Marvin E. Frazier, Stephen W. Scherer, Robert L. Strausberg, and J. Craig Venter. The Diploid Genome Sequence of an Individual Human. *PLOS Biology*, 5(10):e254, September 2007.

[156] Timothy J Ley, Elaine R Mardis, Li Ding, Bob Fulton, Michael D McLellan, Ken Chen, David Dooling, Brian H Dunford-Shore, Sean McGrath, Matthew Hickenbotham, Lisa Cook, Rachel Abbott, David E Larson, Dan C Koboldt, Craig Pohl, Scott Smith, Amy Hawkins, Scott Abbott, Devin Locke, LaDeana W Hillier, Tracie Miner, Lucinda Fulton, Vincent Magrini, Todd Wylie, Jarret Glasscock, Joshua Conyers, Nathan Sander, Xiaoqi Shi, John R Osborne, Patrick Minx, David Gordon, Asif Chinwalla, Yu Zhao, Rhonda E Ries, Jacqueline E Payton, Peter Westervelt, Michael H Tomasson, Mark Watson, Jack Baty, Jennifer Ivanovich, Sharon Heath, William D Shannon, Rakesh Nagarajan, Matthew J Walter, Daniel C Link, Timothy A Graubert, John F DiPersio, and Richard K Wilson. DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. *Nature*, 456(7218):66–72, November 2008.

[157] Gang Li, Brian W Davis, Eduardo Eizirik, and William J Murphy. Phylogenomic evidence for ancient hybridization in the genomes of living cats (felidae). *Genome Res.*, 26(1):1–11, January 2016.

[158] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, and 1000 Genome Project Data. The sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.

[159] Heng Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ARXIV*, 2013.

[160] Heng Li. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, September 2018.

[161] Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, July 2009.

[162] Heng Li and Richard Durbin. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5):589–595, March 2010.

[163] Heng Li and Richard Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496, 2011.

[164] Ruiqiang Li, Wei Fan, Geng Tian, Hongmei Zhu, Lin He, Jing Cai, Quanfei Huang, Qingle Cai, Bo Li, Yinqi Bai, Zhihe Zhang, Yaping Zhang, Wen Wang, Jun Li, Fuwen Wei, Heng Li, Min Jian, Jianwen Li, Zhaolei Zhang, Rasmus Nielsen, Dawei Li, Wanjun Gu, Zhentao Yang, Zhaoling Xuan, Oliver A. Ryder, Frederick Chi-Ching Leung, Yan Zhou, Jianjun Cao, Xiao Sun, Yonggui Fu, Xiaodong Fang, Xiaosen Guo, Bo Wang, Rong Hou, Fujun Shen, Bo Mu, Peixiang Ni, Runmao Lin, Wubin Qian, Guodong Wang, Chang Yu, Wenhui Nie, Jinhuan Wang, Zhigang Wu, Huiqing Liang, Jiumeng Min, Qi Wu, Shifeng Cheng, Jue Ruan, Mingwei Wang, Zhongbin Shi, Ming Wen, Binghang Liu, Xiaoli Ren, Huisong Zheng, Dong Dong, Kathleen Cook, Gao Shan, Hao Zhang, Carolin Kosiol, Xueying Xie, Zuhong Lu, Hancheng Zheng, Yingrui Li, Cynthia C. Steiner, Tommy Tsan-Yuk Lam, Siyuan Lin, Qinghui Zhang, Guoqing Li, Jing Tian, Timing Gong, Hongde Liu, Dejin Zhang, Lin Fang, Chen Ye, Juanbin Zhang, Wenbo Hu, Anlong Xu, Yuanyuan Ren, Guojie Zhang, Michael W. Bruford, Qibin Li, Lijia Ma, Yiran Guo, Na An, Yujie Hu, Yang Zheng, Yongyong Shi, Zhiqiang Li, Qing Liu, Yanling Chen, Jing Zhao, Ning Qu, Shancen Zhao, Feng Tian, Xiaoling Wang, Haiyin Wang, Lizhi Xu, Xiao Liu, Tomas Vinar, Yajun Wang, Tak-Wah Lam, Siu-Ming Yiu, Shiping Liu, Hemin Zhang, Desheng Li, Yan Huang, Xia Wang, Guohua Yang, Zhi Jiang, Junyi Wang, Nan Qin, Li Li, Jingxiang Li, Lars Bolund, Karsten Kristiansen, Gane Ka-Shu Wong, Maynard Olson, Xiuqing Zhang, Songgang Li, Huanming Yang, Jian Wang, and Jun Wang. The sequence and de novo assembly of the giant panda genome. *Nature*, 463(7279):311–317, January 2010.

[165] Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke,

John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950):289–293, October 2009.

[166] Tomas Lindahl. Instability and decay of the primary structure of DNA. *Nature*, 362(6422):709–715, April 1993.

[167] Charlotte Lindqvist, Stephan C. Schuster, Yazhou Sun, Sandra L. Talbot, Ji Qi, Aakrosh Ratan, Lynn P. Tomsho, Lindsay Kasson, Eve Zeyl, Jon Aars, Webb Miller, Ólafur Ingólfsson, Lutz Bachmann, and Øystein Wiig. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences*, 107(11):5053–5057, March 2010.

[168] H E L Lischer and L Excoffier. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2):298–299, January 2012.

[169] Shiping Liu, Eline D. Lorenzen, Matteo Fumagalli, Bo Li, Kelley Harris, Zijun Xiong, Long Zhou, Thorfinn Sand Korneliussen, Mehmet Somel, Courtney Babbitt, Greg Wray, Jianwen Li, Weiming He, Zhuo Wang, Wenjing Fu, Xueyan Xiang, Claire C. Morgan, Aoife Doherty, Mary J. O'Connell, James O. McInerney, Erik W. Born, Love Dalén, Rune Dietz, Ludovic Orlando, Christian Sonne, Guojie Zhang, Rasmus Nielsen, Eske Willerslev, and Jun Wang. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, 157(4):785–794, May 2014.

[170] J V Lopez, M Culver, J C Stephens, W E Johnson, and S J O'Brien. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Mol. Biol. Evol.*, 14(3):277–286, March 1997.

[171] J V Lopez, N Yuhki, R Masuda, W Modi, and S J O'Brien. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.*, 39(2):174–190, August 1994.

[172] Wai Yee Low, Rick Tearle, Derek M. Bickhart, Benjamin D. Rosen, Sarah B. Kingan, Thomas Swale, Françoise Thibaud-Nissen, Terence D. Murphy, Rachel Young, Lucas Lefevre, David A. Hume, Andrew Collins, Paolo Ajmone-Marsan, Timothy P. L. Smith, and John L. Williams. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nature Communications*, 10(1):260, January 2019.

[173] Guillaume Marçais and Carl Kingsford. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6):764–770, March 2011.

[174] Larry G. Marshall, S. David Webb, J. John Sepkoski, and David M. Raup. Mammalian Evolution and the Great American Interchange. *Science*, 215(4538):1351–1357, March 1982.

[175] Larry D. Martin, B. M. Gilbert, and Daniel B. Adams. A cheetah-like cat in the north american pleistocene. *Science*, 195(4282):981–982, 1977.

[176] Paul Matheus, James Burns, Jaco Weinstock, and Michael Hofreiter. Pleistocene Brown Bears in the Mid-Continent of North America. *Science*, 306(5699):1150–1150, November 2004.

[177] Paul E. Matheus. Diet and Co-ecology of Pleistocene Short-Faced Bears and Brown Bears in Eastern Beringia. *Quaternary Research*, 44(3):447–453, November 1995.

[178] Eunice M. Matte, Camila S. Castilho, Renata A. Miotto, Denis A. Sana, Warren E. Johnson, Stephen J. OBrien, Thales R. O. de Freitas, and Eduardo Eizirik. Molecular evidence for a recent demographic expansion in the puma (puma concolor) (mammalia, felidae). *Genetics and Molecular Biology*, 36(4):586597, December 2013.

[179] O Mazet, W Rodríguez, S Grusea, S Boitard, and L Chikhi. On the importance of being structured: instantaneous coalescence rates and human evolution–lessons for ancestral population size inference? *Heredity*, 116(4):362–371, April 2016.

[180] Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A DePristo. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20(9):1297–1303, September 2010.

[181] B.N. McLellan and D Reiner. A review of bear evolution. *Int Conf Bear Res and Manage*, (9):85–96, 1994.

[182] B H McRae, P Beier, L E Dewald, L Y Huynh, and P Keim. Habitat barriers limit gene flow and illuminate historical events in a wide-ranging carnivore, the american puma. *Mol. Ecol.*, 14(7):1965–1977, June 2005.

[183] Paloma Medina, Bryan Thornlow, Rasmus Nielsen, and Russell Corbett-Detig. Estimating the timing of multiple admixture pulses during local ancestry inference. *Genetics*, 210(3):1089–1107, November 2018.

[184] CH Merriam. Review of the grizzly and big brown bears of North America. Technical report, US Government Printing Office, Washington, D.C., 1918.

[185] Matthias Meyer, Juan-Luis Arsuaga, Cesare de Filippo, Sarah Nagel, Ayinuer Aximu-Petri, Birgit Nickel, Ignacio Martínez, Ana Gracia, José María Bermúdez de Castro, Eudald Carbonell, Bence Viola, Janet Kelso, Kay Prüfer, and Svante Pääbo. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature; London*, 531(7595):504–507K, March 2016.

[186] Matthias Meyer and Martin Kircher. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.*, 2010(6), June 2010.

[187] Matthias Meyer, Martin Kircher, Marie-Theres Gansauge, Heng Li, Fernando Racimo, Swapan Mallick, Joshua G. Schraiber, Flora Jay, Kay Prüfer, Cesare de Filippo, Peter H. Sudmant, Can Alkan, Qiaomei Fu, Ron Do, Nadin Rohland, Arti Tandon, Michael Siebauer, Richard E. Green, Katarzyna Bryc, Adrian W. Briggs, Udo Stenzel, Jesse Dabney, Jay Shendure, Jacob Kitzman, Michael F. Hammer, Michael V. Shunkov, Anatoli P. Derevianko, Nick Patterson, Aida M. Andrés, Evan E. Eichler, Montgomery Slatkin, David Reich, Janet Kelso, and Svante Pääbo. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, 338(6104):222–226, October 2012.

[188] C. R. Miller, L. P. Waits, and P. Joyce. Phylogeography and mitochondrial diversity of extirpated brown bear (Ursus arctos) populations in the contiguous United States and Mexico. *Molecular Ecology*, 15(14):4477–4485, December 2006.

[189] Webb Miller, Stephan C. Schuster, Andreanna J. Welch, Aakrosh Ratan, Oscar C. Bedoya-Reina, Fangqing Zhao, Hie Lim Kim, Richard C. Burhans, Daniela I. Drautz, Nicola E. Wittekindt, Lynn P. Tomsho, Enrique Ibarra-Laclette, Luis Herrera-Estrella, Elizabeth Peacock, Sean Farley, George K. Sage, Karyn Rode, Martyn Obbard, Rafael Montiel, Lutz Bachmann, Ólafur Ingólfsson, Jon Aars, Thomas Mailund, Øystein Wiig, Sandra L. Talbot, and Charlotte Lindqvist. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences*, 109(36):E2382–E2390, April 2012.

[190] Kieren J. Mitchell, Sarah C. Bray, Pere Bover, Leopoldo Soibelzon, Blaine W. Schubert, Francisco Prevosti, Alfredo Prieto, Fabiana Martin, Jeremy J. Austin, and Alan Cooper. Ancient mitochondrial DNA reveals convergent evolution of giant short-faced bears (Tremarctinae) in North and South America. *Biology Letters*, 12(4):20160062, April 2016.

[191] Andrew W. Murray and Jack W. Szostak. Construction of artificial chromosomes in yeast. *Nature*, 305(5931):189–193, September 1983.

[192] Gemma G. R. Murray, André E. R. Soares, Ben J. Novak, Nathan K. Schaefer, James A. Cahill, Allan J. Baker, John R. Demboski, Andrew Doll, Rute R. Da Fonseca, Tara L. Fulton, M. Thomas P. Gilbert, Peter D. Heintzman, Brandon Letts, George McIntosh, Brendan L. O'Connell, Mark Peck, Marie-Lorraine Pipes, Edward S. Rice, Kathryn M. Santos, A. Gregory Sohrweide, Samuel H. Vohr, Russell B. Corbett-Detig, Richard E. Green, and Beth Shapiro. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, 358(6365):951–954, November 2017.

[193] National Institutes of Health and Department of Energy. Understanding Our Genetic Inheritance: The US Human Genom Project: The First Five Years, FY 1991-1995. Technical Report Document DOE/ER-0452P, National Institutes of Health, U.S. Department of Health and Human Services; and U.S. Department of Energy, 1990.

[194] James P. Noonan, Graham Coop, Sridhar Kudaravalli, Doug Smith, Johannes Krause, Joe Alessi, Feng Chen, Darren Platt, Svante Pääbo, Jonathan K. Pritchard, and Ed-

ward M. Rubin. Sequencing and Analysis of Neanderthal Genomic DNA. *Science*, 314(5802):1113–1118, November 2006.

[195] James P. Noonan, Michael Hofreiter, Doug Smith, James R. Priest, Nadin Rohland, Gernot Rabeder, Johannes Krause, J. Chris Detter, Svante Pääbo, and Edward M. Rubin. Genomic sequencing of Pleistocene cave bears. *Science (New York, N.Y.)*, 309(5734):597–599, July 2005.

[196] Kristin Nowell, Peter Jackson, and IUCN/SSC Cat Specialist Group. *Wild Cats: Status Survey and Conservation Action Plan*. World Conservation Union, 1996.

[197] Stephen J O'Brien. *Tears of the Cheetah: The Genetic Secrets of Our Animal Ancestors*. St. Martin's Griffin, October 2015.

[198] Stephen J O'Brien, Melody E Roelke, Naoya Yuhki, Karen W Richards, Warren E Johnson, William I Franklin, Allen E Anderson, Oron L Bass, JR., Robert C Belden, and Janie S Martenson. Genetic introgression within the florida panther (felis concolor coryi). *National Geographic Research*, 6(4):484–494, 1990.

[199] Alexander Ochoa, David P Onorato, Robert R Fitak, Melody E Roelke-Parker, and Melanie Culver. Evolutionary and functional mitogenomics associated with the genetic restoration of the florida panther. *J. Hered.*, 108(4):449–455, June 2017.

[200] Genome 10K Community of Scientists. Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *Journal of Heredity*, 100(6):659–674, November 2009.

[201] Maynard Olson, Leroy Hood, Charles Cantor, and David Botstein. A Common Language for Physical Mapping of the Human Genome. *Science; Washington*, 245(4925):1434, September 1989.

[202] Dave Onorato, Chris Belden, Mark Cunningham, Darrell Land, Roy McBride, and Melody Roelke. *The Biology and Conservation of Wild Felids*. Oxford University Press, June 2010.

[203] Ludovic Orlando, Aurélien Ginolhac, Guojie Zhang, Duane Froese, Anders Albrechtsen, Mathias Stiller, Mikkel Schubert, Enrico Cappellini, Bent Petersen, Ida Moltke, Philip L. F. Johnson, Matteo Fumagalli, Julia T. Vilstrup, Maanasa Raghavan, Thorfinn Korneliussen, Anna-Sapfo Malaspinas, Josef Vogt, Damian Szklarczyk, Christian D. Kelstrup, Jakob Vinther, Andrei Dolocan, Jesper Stenderup, Amhed M. V. Velazquez, James Cahill, Morten Rasmussen, Xiaoli Wang, Jiumeng Min, Grant D. Zazula, Andaine Seguin-Orlando, Cecilie Mortensen, Kim Magnussen, John F. Thompson, Jacobo Weinstock, Kristian Gregersen, Knut H. Røed, Véra Eisenmann, Carl J. Rubin, Donald C. Miller, Douglas F. Antczak, Mads F. Bertelsen, Søren Brunak, Khaled A. S. Al-Rasheid, Oliver Ryder, Leif Andersson, John Mundy, Anders Krogh, M. Thomas P. Gilbert, Kurt Kjær, Thomas Sicheritz-Ponten, Lars Juhl Jensen, Jesper V. Olsen, Michael Hofreiter,

Rasmus Nielsen, Beth Shapiro, Jun Wang, and Eske Willerslev. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456):74–78, July 2013.

[204] Svante Pääbo, Hendrik Poinar, David Serre, Viviane Jaenicke-Després, Juliane Hebler, Nadin Rohland, Melanie Kuch, Johannes Krause, Linda Vigilant, and Michael Hofreiter. Genetic Analyses from Ancient DNA. *Annual Review of Genetics*, 38(1):645–679, November 2004.

[205] David Paetkau, Gerald F. Shields, and Curtis Strobeck. Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, 7(10):1283–1292, October 1998.

[206] S. Paisley and D. L. Garshelis. Activity patterns and time budgets of Andean bears (Tremarctos ornatus) in the Apolobamba Range of Bolivia. *Journal of Zoology*, 268(1):25–34, 2006.

[207] Pier Francesco Palamara, Todd Lencz, Ariel Darvasi, and Itsik Pe'er. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.*, 91(5):809–822, November 2012.

[208] Nick Patterson, Alkes L Price, and David Reich. Population structure and eigenanalysis. *PLoS Genet.*, 2(12):e190, 2006.

[209] Alison J Pearks Wilkerson, Terje Raudsepp, Tina Graves, Derek Albracht, Wesley Warren, Bhanu P Chowdhary, Loren C Skow, and William J Murphy. Gene discovery and comparative analysis of x-degenerate genes from the domestic cat Y chromosome. *Genomics*, 92(5):329–338, November 2008.

[210] Trevor J Pemberton, Devin Absher, Marcus W Feldman, Richard M Myers, Noah A Rosenberg, and Jun Z Li. Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*, 91(2):275–292, 2012.

[211] Rui Peng, Bo Zeng, Xiuxiang Meng, Bisong Yue, Zhihe Zhang, and Fangdong Zou. The complete mitochondrial genome and phylogenetic analysis of the giant panda (Ailuropoda melanoleuca). *Gene*, 397(1):76–83, August 2007.

[212] Simone Picelli, Åsa K Björklund, Omid R Faridani, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, 10(11):1096–1098, November 2013.

[213] Joseph K Pickrell and Jonathan K Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8(11):e1002967, November 2012.

[214] Hendrik N. Poinar, Carsten Schwarz, Ji Qi, Beth Shapiro, Ross D. E. MacPhee, Bernard Buigues, Alexei Tikhonov, Daniel H. Huson, Lynn P. Tomsho, Alexander Auch, Markus

Rampp, Webb Miller, and Stephan C. Schuster. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science*, 311(5759):392–394, January 2006.

[215] Mihai Pop, Adam Phillippy, Arthur L. Delcher, and Steven L. Salzberg. Comparative genome assembly. *Briefings in Bioinformatics*, 5(3):237–248, September 2004.

[216] Alkes L. Price, Neil C. Jones, and Pavel A. Pevzner. De novo identification of repeat families in large genomes. *Bioinformatics*, 21(suppl 1):i351–i358, January 2005.

[217] Kay Prüfer, Fernando Racimo, Nick Patterson, Flora Jay, Sriram Sankararaman, Susanna Sawyer, Anja Heinze, Gabriel Renaud, Peter H. Sudmant, Cesare de Filippo, Heng Li, Swapan Mallick, Michael Dannemann, Qiaomei Fu, Martin Kircher, Martin Kuhlwilm, Michael Lachmann, Matthias Meyer, Matthias Ongyerth, Michael Siebauer, Christoph Theunert, Arti Tandon, Priya Moorjani, Joseph Pickrell, James C. Mullikin, Samuel H. Vohr, Richard E. Green, Ines Hellmann, Philip L. F. Johnson, Hélène Blanche, Howard Cann, Jacob O. Kitzman, Jay Shendure, Evan E. Eichler, Ed S. Lein, Trygve E. Bakken, Liubov V. Golovanova, Vladimir B. Doronichev, Michael V. Shunkov, Anatoli P. Derevianko, Bence Viola, Montgomery Slatkin, David Reich, Janet Kelso, and Svante Pääbo. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481):43–49, January 2014.

[218] Kay Prüfer, Udo Stenzel, Michael Hofreiter, Svante Pääbo, Janet Kelso, and Richard E. Green. Computational challenges in the analysis of ancient DNA. *Genome Biology*, 11(5):R47, 2010.

[219] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3):559–575, September 2007.

[220] Nicholas H Putnam, Brendan L O'Connell, Jonathan C Stites, Brandon J Rice, Marco Blanchette, Robert Calef, Christopher J Troll, Andrew Fields, Paul D Hartley, Charles W Sugnet, David Haussler, Daniel S Rokhsar, and Richard E Green. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.*, 26(3):342–350, March 2016.

[221] Qiang Qiu, Guojie Zhang, Tao Ma, Wubin Qian, Junyi Wang, Zhiqiang Ye, Changchang Cao, Quanjun Hu, Jaebum Kim, Denis M. Larkin, Loretta Auvil, Boris Capitanu, Jian Ma, Harris A. Lewin, Xiaoju Qian, Yongshan Lang, Ran Zhou, Lizhong Wang, Kun Wang, Jinquan Xia, Shengguang Liao, Shengkai Pan, Xu Lu, Haolong Hou, Yan Wang, Xuetao Zang, Ye Yin, Hui Ma, Jian Zhang, Zhaofeng Wang, Yingmei Zhang, Dawei Zhang, Takahiro Yonezawa, Masami Hasegawa, Yang Zhong, Wenbin Liu, Yan Zhang, Zhiyong Huang, Shengxiang Zhang, Ruijun Long, Huanming Yang, Jian Wang, Johannes A. Lenstra, David N. Cooper, Yi Wu, Jun Wang, Peng Shi, Jian Wang, and

Jianquan Liu. The yak genome and adaptation to life at high altitude. *Nature Genetics*, 44(8):946–949, August 2012.

[222] Aaron R Quinlan and Ira M Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

[223] Andrew Rambaut, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, 67(5):901–904, September 2018.

[224] Christopher Bronk Ramsey. Methods for Summarizing Radiocarbon Datasets. *Radiocarbon*, 59(6):1809–1833, December 2017.

[225] Morten Rasmussen, Yingrui Li, Stinus Lindgreen, Jakob Skou Pedersen, Anders Albrechtsen, Ida Moltke, Mait Metspalu, Ene Metspalu, Toomas Kivisild, Ramneek Gupta, Marcelo Bertalan, Kasper Nielsen, M. Thomas P. Gilbert, Yong Wang, Maanasa Raghavan, Paula F. Campos, Hanne Munkholm Kamp, Andrew S. Wilson, Andrew Gledhill, Silvana Tridico, Michael Bunce, Eline D. Lorenzen, Jonas Binladen, Xiaosen Guo, Jing Zhao, Xiuqing Zhang, Hao Zhang, Zhuo Li, Minfeng Chen, Ludovic Orlando, Karsten Kristiansen, Mads Bak, Niels Tommerup, Christian Bendixen, Tracey L. Pierre, Bjarne Grønnow, Morten Meldgaard, Claus Andreasen, Sardana A. Fedorova, Ludmila P. Osipova, Thomas F. G. Higham, Christopher Bronk Ramsey, Thomas v O. Hansen, Finn C. Nielsen, Michael H. Crawford, Søren Brunak, Thomas Sicheritz-Pontén, Richard Villems, Rasmus Nielsen, Anders Krogh, Jun Wang, and Eske Willerslev. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463(7282):757–762, February 2010.

[226] Paula J. Reimer, Edouard Bard, Alex Bayliss, J. Warren Beck, Paul G. Blackwell, Christopher Bronk Ramsey, Caitlin E. Buck, Hai Cheng, R. Lawrence Edwards, Michael Friedrich, Pieter M. Grootes, Thomas P. Guilderson, Haflidi Haflidason, Irka Hajdas, Christine Hatté, Timothy J. Heaton, Dirk L. Hoffmann, Alan G. Hogg, Konrad A. Hughen, K. Felix Kaiser, Bernd Kromer, Sturt W. Manning, Mu Niu, Ron W. Reimer, David A. Richards, E. Marian Scott, John R. Southon, Richard A. Staff, Christian S. M. Turney, and Johannes van der Plicht. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, 55(4):1869–1887, 2013/ed.

[227] Seth P D Riley, John P Pollinger, Raymond M Sauvajot, Eric C York, Cassity Bromley, Todd K Fuller, and Robert K Wayne. A southern california freeway is a physical and social barrier to gene flow in carnivores. *Mol. Ecol.*, 15(7):1733–1741, June 2006.

[228] Seth P D Riley, Laurel E K Serieys, John P Pollinger, Jeffrey A Sikich, Lisa Dalbeck, Robert K Wayne, and Holly B Ernest. Individual behaviors dominate the dynamics of an urban mountain lion population isolated by roads. *Curr. Biol.*, 24(17):1989–1994, September 2014.

[229] Seth P D Riley, Trish Smith, and Winston T Vickers. Assessment of wildlife crossing sites for the interstate 15 and highway 101 freeways in southern california. Technical report, March 2018.

[230] Gene E. Robinson, Kevin J. Hackett, Mary Purcell-Miramontes, Susan J. Brown, Jay D. Evans, Marian R. Goldsmith, Daniel Lawson, Jack Okamuro, Hugh M. Robertson, and David J. Schneider. Creating a Buzz About Insect Genomes. *Science*, 331(6023):1386–1386, March 2011.

[231] Jacqueline A Robinson, Diego Ortega-Del Vecchyo, Zhenxin Fan, Bernard Y Kim, Bridgett M vonHoldt, Clare D Marsden, Kirk E Lohmueller, and Robert K Wayne. Genomic flatlining in the endangered island fox. *Curr. Biol.*, 26(9):1183–1189, May 2016.

[232] Fredrik Ronquist, Maxim Teslenko, Paul van der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3):539–542, May 2012.

[233] Jonathan M Rothberg and John H Leamon. The development and impact of 454 sequencing. *Nature Biotechnology*, 26(10):1117–1124, October 2008.

[234] Ryan R. Wick. Porechop. `https://github.com/rrwick/Porechop`.

[235] Boessenkool S, Hanghøj K, Nistelberger Hm, Der Sarkissian C, Gondek At, Orlando L, Barrett Jh, and Star B. Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*, 17(4):742–751, December 2016.

[236] F. Sanger, J. E. Donelson, A. R. Coulson, H. Kossel, and D. Fischer. Use of dna polymerase i primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 dna. *Proceedings of the National Academy of Sciences of the United States of America*, 70(4):1209–1213, April 1973.

[237] Nedda F. Saremi, Megan A. Supple, Ashley Byrne, James A. Cahill, Luiz Lehmann Coutinho, Love Dalén, Henrique V. Figueiró, Warren E. Johnson, Heather J. Milne, Stephen J. O'Brien, Brendan O'Connell, David P. Onorato, Seth P. D. Riley, Jeff A. Sikich, Daniel R. Stahler, Priscilla Marqui Schmidt Villela, Christopher Vollmers, Robert K. Wayne, Eduardo Eizirik, Russell B. Corbett-Detig, Richard E. Green, Christopher C. Wilmers, and Beth Shapiro. Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature Communications*, 10(1):4769, October 2019.

[238] Susanna Sawyer, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and Svante Pääbo. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS One; San Francisco*, 7(3):e34131, March 2012.

[239] Melissa A Wilson Sayres, Melissa A Wilson Sayres, Chris Venditti, Mark Pagel, and Kateryna D Makova. Do variations in substitution rates and male mutation bias correlate

with life-history traits? a study of 32 mammalian genomes. *Evolution*, 65(10):2800–2815, 2011.

[240] George B Schaller and Peter Gransden Crawshaw. Movement patterns of jaguar. *Biotropica*, 12(3):161, 1980.

[241] Joshua G Schraiber and Joshua M Akey. Methods and models for unravelling human evolutionary history, 2015.

[242] Blaine W. Schubert. Late Quaternary chronology and extinction of North American giant short-faced bears (Arctodus simus). *Quaternary International*, 217(1):188–194, April 2010.

[243] Blaine W. Schubert, Richard C. Hulbert, Bruce J. MacFadden, Michael Searle, and Seina Searle. Giant Short-faced Bears (Arctodus simus) in Pleistocene Florida USA, a Substantial Range Extension. *Journal of Paleontology*, 84(1):79–87, January 2010.

[244] Mikkel Schubert, Aurelien Ginolhac, Stinus Lindgreen, John F Thompson, Khaled AS AL-Rasheid, Eske Willerslev, Anders Krogh, and Ludovic Orlando. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, 13:178, May 2012.

[245] V. J. Schuenemann, K. Bos, S. DeWitte, S. Schmedes, J. Jamieson, A. Mittnik, S. Forrest, B. K. Coombes, J. W. Wood, D. J. D. Earn, W. White, J. Krause, and H. N. Poinar. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proceedings of the National Academy of Sciences*, 108(38):E746–E752, September 2011.

[246] Michael K Schwartz and Kevin S McKelvey. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, 10(2):441–452, 2009.

[247] Christopher Servheen, Stephen Herrero, and Bernard Peyton. *Bears: Status Survey and Conservation Action Plan*. IUCN, Gland, Switzerland, 6 edition, 1999.

[248] Beth Shapiro, Alexei J. Drummond, Andrew Rambaut, Michael C. Wilson, Paul E. Matheus, Andrei V. Sher, Oliver G. Pybus, M. Thomas P. Gilbert, Ian Barnes, Jonas Binladen, Eske Willerslev, Anders J. Hansen, Gennady F. Baryshnikov, James A. Burns, Sergei Davydov, Jonathan C. Driver, Duane G. Froese, C. Richard Harington, Grant Keddie, Pavel Kosintsev, Michael L. Kunz, Larry D. Martin, Robert O. Stephenson, John Storer, Richard Tedford, Sergei Zimov, and Alan Cooper. Rise and Fall of the Beringian Steppe Bison. *Science*, 306(5701):1561–1565, November 2004.

[249] Gerald F. Shields, Deborah Adams, Gerald Garner, Martine Labelle, Jacy Pietsch, Malcolm Ramsay, Charles Schwartz, Kimberly Titus, and Scott Williamson. Phylogeography of Mitochondrial DNA Variation in Brown Bears and Polar Bears. *Molecular Phylogenetics and Evolution*, 15(2):319–326, May 2000.

[250] H. Shizuya, B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proceedings of the National Academy of Sciences*, 89(18):8794–8797, September 1992.

[251] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539, October 2011.

[252] Felipe A Simão, Robert M Waterhouse, Panagiotis Ioannidis, Evgenia V Kriventseva, and Evgeny M Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, October 2015.

[253] Magda Sindičić, Tomislav Gomerčić, Ana Galov, Primož Polanc, Duro Huber, and Alen Slavica. Repetitive sequences in eurasian lynx (lynx lynx l.) mitochondrial DNA control region. *Mitochondrial DNA*, 23(3):201–207, June 2012.

[254] Guy St C Slater and Ewan Birney. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6:31, February 2005.

[255] Guy St C. Slater and Ewan Birney. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6:31, 2005.

[256] AFA Smith, R Hubley, and P Green. RepeatMasker. http://www.repeatmasker.org/, 2013.

[257] Carol Soderlund, William Nelson, Austin Shoemaker, and Andrew Paterson. SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res.*, 16(9):1159–1168, September 2006.

[258] Leopoldo H. Soibelzon and Blaine W. Schubert. The largest known bear, arctotherium angustidens, from the early pleistocene pampean region of argentina: With a discussion of size and diet trends in bears. *Journal of Paleontology*, 85(1):6975, Jan 2011.

[259] Leopoldo H. Soibelzon, Eduardo P. Tonni, and Mariano Bond. The fossil record of South American short-faced bears (Ursidae, Tremarctinae). *Journal of South American Earth Sciences*, 20(1–2):105–113, October 2005.

[260] B. Sorkin. Ecomorphology of the giant short-faced bears Agriotherium and Arctodus. *Historical Biology*, 18(1):1–20, January 2006.

[261] Anuj Srivastava, Vishal Kumar Sarsani, Ian Fiddes, Susan M. Sheehan, Rita L. Seger, Mary E. Barter, Selena Neptune-Bear, Charlotte Lindqvist, and Ron Korstanje. Genome assembly and gene expression in the American black bear provides new insights into the renal response to hibernation. *DNA Research*, 26(1):37–44, February 2019.

[262] John St. John and Jordan Eizenga. SeqPrep2.

[263] Alexandros Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, May 2014.

[264] Mario Stanke, Rasmus Steinkamp, Stephan Waack, and Burkhard Morgenstern. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, 32(suppl 2):W309–W312, January 2004.

[265] Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Miles J. Efron, Ravishankar Iyer, Michael C. Schatz, Saurabh Sinha, and Gene E. Robinson. Big Data: Astronomical or Genomical? *PLOS Biology*, 13(7):e1002195, July 2015.

[266] Marcelo Stucchi, Rodolfo Salas-Gismondi, Patrice Baby, Jean-Loup Guyot, and Bruce J. Shockey. A 6,000+ year-old specimen of a spectacled bear from an andean cave in peru. *Ursus*, 20(1):6368, Apr 2009.

[267] Mel Sunquist and Fiona Sunquist. *Wild Cats of the World*. University of Chicago Press, May 2017.

[268] Megan A. Supple. New polar bear genome.

[269] U.S. Geological Survey. Gap analysis project, 2017, cougar (puma concolor) mCOUGx_CONUS_2001v1 habitat map. `https://www.sciencebase.gov/catalog/item/58fa64d6e4b0b7ea545257e0`, July 2018.

[270] Linda L Sweanor, Kenneth A Logan, and Maurice G Hornocker. Cougar dispersal patterns, metapopulation dynamics, and conservation. *Conserv. Biol.*, 14(3):798–808, 2000.

[271] Pierre Taberlet and Jean Bouvet. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear Ursus arctos in Europe. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1344):195–200, March 1994.

[272] Gregory A. Taylor, Heather Kirk, Lauren Coombe, Shaun D. Jackman, Justin Chu, Kane Tse, Dean Cheng, Eric Chuah, Pawan Pandoh, Rebecca Carlsen, Yongjun Zhao, Andrew J. Mungall, Richard Moore, Inanc Birol, Maria Franke, Marco A. Marra, Christopher Dutton, and Steven J. M. Jones. The Genome of the North American Brown Bear or Grizzly: Ursus arctos ssp. horribilis. *Genes*, 9(12), November 2018.

[273] Richard H. Tedford and James Martin. Plionarctos, a Tremarctine Bear (Ursidae: Carnivora) from Western North America. *Journal of Vertebrate Paleontology*, 21(2):311–321, 2001.

[274] Emma C. Teeling, Sonja C. Vernes, Liliana M. Dávalos, David A. Ray, M. Thomas P. Gilbert, and Eugene Myers. Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annual Review of Animal Biosciences*, 6(1):23–46, February 2018.

[275] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, November 2012.

[276] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814):796–815, December 2000.

[277] Elizabeth A Thompson. Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2):301–326, 2013.

[278] E. Trajano and H. Ferrarezzi. A fossil bear from northeastern brazil, with a phylogenetic analysis of the south american extinct tremarctinae (ursidae). *Journal of Vertebrate Paleontology*, 14(4):552561, Feb 1995.

[279] Christopher J. Troll, Joshua Kapp, Varsha Rao, Kelly M. Harkins, Charles Cole, Colin Naughton, Jessica M. Morgan, Beth Shapiro, and Richard E. Green. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics*, 20(1):1023, December 2019.

[280] Terrence M. Tumpey, Christopher F. Basler, Patricia V. Aguilar, Hui Zeng, Alicia Solórzano, David E. Swayne, Nancy J. Cox, Jacqueline M. Katz, Jeffery K. Taubenberger, Peter Palese, and Adolfo García-Sastre. Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus. *Science*, 310(5745):77–80, October 2005.

[281] John R. Tyson, Nigel J. O'Neil, Miten Jain, Hugh E. Olsen, Philip Hieter, and Terrance P. Snutch. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. *Genome Research*, 28(2):266–274, February 2018.

[282] Blaire Van Valkenburgh. Major Patterns in the History of Carnivorous Mammals. *Annual Review of Earth and Planetary Sciences*, 27(1):463–493, 1999.

[283] Blaire Van Valkenburgh, Frederick Grady, and Björn Kurtèn. The plio-pleistocene cheetah-like cat miracinonyx inexpectatus of north america. *Journal of Vertebrate Paleontology*, 10(4):434–454, December 1990.

[284] Madelon van de Kerk, David P. Onorato, Jeffrey A. Hostetler, Benjamin M. Bolker, and Madan K. Oli. Dynamics, Persistence, and Genetic Management of the Endangered Florida Panther Population. *Wildlife Monographs*, 203(1):3–35, 2019.

[285] T Winston Vickers, T Winston Vickers, Jessica N Sanchez, Christine K Johnson, Scott A Morrison, Randy Botta, Trish Smith, Brian S Cohen, Patrick R Huber, Holly B Ernest, and Walter M Boyce. Survival and mortality of pumas (puma concolor) in a fragmented, urbanizing landscape. *PLoS One*, 10(7):e0131490, 2015.

[286] Julia T. Vilstrup, Andaine Seguin-Orlando, Mathias Stiller, Aurelien Ginolhac, Maanasa Raghavan, Sandra C. A. Nielsen, Jacobo Weinstock, Duane Froese, Sergei K. Vasiliev, Nikolai D. Ovodov, Joel Clary, Kristofer M. Helgen, Robert C. Fleischer, Alan Cooper, Beth Shapiro, and Ludovic Orlando. Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLOS ONE*, 8(2):e55950, February 2013.

[287] Gregory W Vurture, Fritz J Sedlazeck, Maria Nattestad, Charles J Underwood, Han Fang, James Gurtowski, and Michael C Schatz. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14):2202–2204, July 2017.

[288] Lisette P. Waits, Sandra L. Talbot, R. H. Ward, and G. F. Shields. Mitochondrial DNA Phylogeography of the North American Brown Bear and Implications for Conservation. *Conservation Biology*, 12(2):408–417, 1998.

[289] Bruce J Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K Young, and Ashlee M Earl. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11):e112963, November 2014.

[290] Jun Wang, Wei Wang, Ruiqiang Li, Yingrui Li, Geng Tian, Laurie Goodman, Wei Fan, Junqing Zhang, Jun Li, Juanbin Zhang, Yiran Guo, Binxiao Feng, Heng Li, Yao Lu, Xiaodong Fang, Huiqing Liang, Zhenglin Du, Dong Li, Yiqing Zhao, Yujie Hu, Zhenzhen Yang, Hancheng Zheng, Ines Hellmann, Michael Inouye, John Pool, Xin Yi, Jing Zhao, Jinjie Duan, Yan Zhou, Junjie Qin, Lijia Ma, Guoqing Li, Zhentao Yang, Guojie Zhang, Bin Yang, Chang Yu, Fang Liang, Wenjie Li, Shaochuan Li, Dawei Li, Peixiang Ni, Jue Ruan, Qibin Li, Hongmei Zhu, Dongyuan Liu, Zhike Lu, Ning Li, Guangwu Guo, Jianguo Zhang, Jia Ye, Lin Fang, Qin Hao, Quan Chen, Yu Liang, Yeyang Su, A. San, Cuo Ping, Shuang Yang, Fang Chen, Li Li, Ke Zhou, Hongkun Zheng, Yuanyuan Ren, Ling Yang, Yang Gao, Guohua Yang, Zhuo Li, Xiaoli Feng, Karsten Kristiansen, Gane Ka-Shu Wong, Rasmus Nielsen, Richard Durbin, Lars Bolund, Xiuqing Zhang, Songgang Li, Huanming Yang, and Jian Wang. The diploid genome sequence of an Asian individual. *Nature*, 456(7218):60–65, November 2008.

[291] Yiwei Wang, Maximilian L Allen, and Christopher C Wilmers. Mesopredator spatial and temporal responses to large predators and human development in the santa cruz mountains of california. *Biol. Conserv.*, 190:23–33, 2015.

[292] Ryk Ward and Chris Stringer. A molecular handle on the Neanderthals. *Nature*, 388(6639):225–226, July 1997.

[293] Andrew M. Waterhouse, James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, May 2009.

[294] Robert H. Waterson, Eric S. Lander, Richard K. Wilson, and The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055):69–87, September 2005.

[295] Robert H. Waterston, Eric S. Lander, and John E. Sulston. On the sequencing of the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6):3712–3716, March 2002.

[296] Michael V. Westbury, Stefanie Hartmann, Axel Barlow, Michaela Preick, Bogdan Ridush, Doris Nagel, Thomas Rathgeber, Reinhard Ziegler, Gennady Baryshnikov, Guilian Sheng, Arne Ludwig, Ingrid Wiesel, Love Dalen, Faysal Bibi, Lars Werdelin, Rasmus Heller, and Michael Hofreiter. Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Science Advances*, 6(11):eaay0456, March 2020.

[297] David A. Wheeler, Maithreyan Srinivasan, Michael Egholm, Yufeng Shen, Lei Chen, Amy McGuire, Wen He, Yi-Ju Chen, Vinod Makhijani, G. Thomas Roth, Xavier Gomes, Karrie Tartaro, Faheem Niazi, Cynthia L. Turcotte, Gerard P. Irzyk, James R. Lupski, Craig Chinault, Xing-zhi Song, Yue Liu, Ye Yuan, Lynne Nazareth, Xiang Qin, Donna M. Muzny, Marcel Margulies, George M. Weinstock, Richard A. Gibbs, and Jonathan M. Rothberg. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452(7189):872–876, April 2008.

[298] Ryan R Wick, Louise M Judd, Claire L Gorrie, and Kathryn E Holt. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.*, 13(6):e1005595, June 2017.

[299] Christopher C Wilmers, Yiwei Wang, Barry Nickel, Paul Houghtaling, Yasaman Shakeri, Maximilian L Allen, Joe Kermish-Wells, Veronica Yovovich, and Terrie Williams. Scale dependent behavioral responses to human development by a large predator, the puma. *PLoS One*, 8(4):e60590, April 2013.

[300] Sewall Wright. Isolation by distance. *Genetics*, 28(2):114–138, 1943.

[301] Thomas D. Wu and Colin K. Watanabe. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9):1859–1875, May 2005.

[302] Ziheng Yang. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, 24(8):1586–1591, August 2007.

[303] Guojie Zhang. Bird sequencing project takes off. *Nature*, 522(7554):34–34, June 2015.

[304] H Zischler, M Hoss, O Handt, A von Haeseler, A. van der Kuyl, and J Goudsmit. Detecting dinosaur DNA. *Science*, 268(5214):1192–1193, May 1995.