UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SPECIATION GENOMICS: INVESTIGATING THE CAUSES AND CONSEQUENCES OF ADMIXTURE BETWEEN POLAR BEARS AND BROWN BEARS USING ANCIENT AND CONTEMPORARY GENOMES

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

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

James A. Cahill

December 2016

The Dissertation of James A. Cahill is approved:

_____
Professor Beth Shapiro, Chair

_____
Professor Richard E. Green

_____
Professor Grant Pogson

_____
Professor Giacomo Bernardi

_____
Professor David Haussler

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

List of Figures

1997; Stuart *et al.* 2004). Brown bears occur in the Irish fossil record before and after the LGM but are absent from 32-14 ka(Woodman *et al.* 1997; Edwards *et al.* 2011), the most recent pre-LGM and most ancient post-LGM brown bears are included in this study.

List of Tables

To my wonderful family

Kiley, your love and companionship extend to every aspect of my life.

Mom and Dad, your love, care and support are the foundation of everything I am and everything that I hope to be.

Speciation Genomics: Investigating the Causes and Consequences of Admixture between Polar Bears and Brown Bears using Ancient and Contemporary Genomes

The concept of the species has traditionally been the principal unit for the classification of the diversity of life. Although the conceptual and legal relevance of the idea of species is profound and wide ranging it is also increasingly evident that the concept of species is a significant oversimplification of the complex processes at work in nature. The growing body of evidence from genomic data, and other sources, indicates that interspecific hybridization plays a much more substantial role in evolution than had previously been appreciated and that speciation, the formation of new species, is likely a more gradual process than had been appreciated. In my dissertation work, I have applied whole genome sequencing and ancient DNA based techniques to examine the evolutionary relationship of two closely related species, polar bears (*Ursus maritimus*) and brown bears (*Ursus arctos*), which occupy substantially different ecological niches. This work is presented in a series of three papers that have contributed to reshaping scientific understanding of the relatedness of polar bears and brown bears and suggest that climate changes can profoundly impact the frequency of hybridization between these two species. I have also contributed to method development – in a fourth paper my coauthors and I developed a novel application of existing analysis tools to estimate the end point of the speciation process, the time at which all gene flow has ceased, from genomic data. Finally, in collaboration with my co-first author Axel Barlow and colleagues, I have examined the evolutionary relationship between polar and brown bears and their closest extinct relatives, the cave bears, which has revealed previously undocumented admixture between cave bears and brown bears and provides insights into the extent of diversity within the paleo-species groups of cave bear.

Acknowledgements

I am deeply grateful to the many many mentors and collaborators that have helped me to reach the point of writing this dissertation. I have been profoundly fortunate to conduct my PhD research in a dynamic lab full of talented and generous people. This work is, I hope, humble testament to the exceptional training that I have received. To my advisor and undoubtedly the most brilliant person I have ever known Beth Shapiro, every interaction with you has been an education and a privilege. To co-advisor Ed Green, you taught me so much and were so very generous to me. To my mentors, Mathias Stiller my dear friend who taught me what it means to be a scientist, Tara Fulton who took the time to make sure that a first year who had no idea what he was doing stayed pointed in the right direction, August Woerner who taught me how to program, Rich Posner who has always been there to help me. To my labmates Sam Vohr, Pete Heintzman, Dan Chang, André Elias Rodrigues Soares, Darko Cotoras, Nathan Schaefer, Nedda Saremi, Josh Kapp, John St. John, Kelly Harkins, Logan Kistler and everyone else in the lab over the past six years it has been a tremendous pleasure. To my collaborators, Axel Barlow, Michi Hofreiter, Bob Wayne, Ian Stirling, Bridgett von Holdt, Klaus Peter Koepfli, Kelley Harris, Nikita Ovsyanikov, Daniel Bradley and so many others thank you so much for allowing me to be a part of your research. To the friends, especially, Doug Wykstra, Nick Geib and Caitlin Conn thank you for helping me to stay (mostly) sane. Finally and most importantly to the people I most love, Kiley who has been literally by my side through this, no one could ever have made me happier, and to my family Mom, Dad and Erin I love you all very very much.

General Introduction

Species Concepts, Speciation and Hybridization

The ability to distinguish between types of organisms is found in many species (Seyfarth & Cheney 1980) and formal human study of the diversity of life is one of the oldest questions in biology. In the European Scientific tradition, the classification of life into species can be attributed to Aristotle who asserted a typological and idealist conception of species. The classification of life into species was also a concern of early modern scientists, notably Linnaeus and Buffon among others, who undertook the classification of living things into a clear hierarchy, defining species and grouping more similar species together into larger organizational (taxonomic) groups. With the development of evolutionary theories by Lamarck and then Darwin, these taxonomic groups acquired a more robust biological significance as groups of greater or lesser evolutionary relatedness.

While Darwin's theory of evolution by natural selection provided a scientific basis for the classification of species, it simultaneously presented fundamental problems for the concept of species. If each species were a separate special creation, then a clear and definite distinction between species was at least a possibility, however, if similarity among species was the product of natural selection leading to differentiation of organisms with a shared ancestor then species must gradually differentiate. Gradual differentiation of organisms from a shared ancestor challenges the classifier to define some threshold for distinguishing between species.

Species Concepts

The criteria that biologists apply to determine whether two groups of related organisms are the same or different species are “species concepts”. Importantly, species concepts are postulates not hypotheses – their role is to

define criteria to test nature against, not to make a testable assertion about the natural world. Therefore, while it is reasonable to ask whether polar bears and brown bears qualify as separate species under some species concept, it is not reasonable to assert that any single species concept can provide a complete understanding of the speciation process for all organisms. Many species concepts have been proposed to aid the classification of life, these tend to capture different features of speciation, the process by which one species becomes multiple distinct species.

The most widely discussed and applied species concept is Ernst Mayr's biological species concept. The biological species concept states that "species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups" (Mayr 1942). This in effect defines species in opposition to admixture, if hybridization does not occur between naturally existing populations, or those hybrids are sufficiently evolutionarily disadvantaged to allow those hybrids to contribute to future generations, the species are considered to be truly species. It is important to recognize that the biological species concept was not intended as a general classification tool but as a descriptor of what Mayr considered to be the biologically "real" phenomenon of species (Mayr 1940).

The key strength of the biological species concept is that it captures a significant biological reality, namely species that are distinct under the biological species concept have undeniably separate evolutionary trajectories and will continue to be distinct for the remainder of their existence. Interbreeding can be a powerful homogenizing factor (Grant et al. 2004; Wright 1931), so a relatively high degree of reproductive isolation is necessary to ensure candidate species will remain independent. However, the biological species concept strikingly fails to conform to what we would expect to constitute species prior to rigorous

investigation of whether candidate species qualify as distinct species under the biological species concept. In 2005, Mallet reviewed the literature and found that at least 25% of plant species and 10% of animal species are part of a hybrid system (Mallet 2005). Although some of the "species" failing to reach species status under the biological species concept are legitimately cases of over-splitting, genomic evidence has revealed a scope of hybridization far beyond even that described by Mallet (Dasmahapatra et al. 2012; Green et al. 2010; Zinner et al. 2009; Poelstra et al. 2014; Cahill et al. 2013; Lamichhaney et al. 2015; Carbone et al. 2014).

Another widely applied species concept is the Ecological Species Concept which, defines a species as "a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in it's range and which evolves separately from all lineages outside its range" (Van Valen 1976). This is more commonly restated as, a species is a group of related organism that share an ecological niche; the "n-dimensional hypervolume" of environmental parameters within which an individual can survive and reproduce (Hutchinson 1957). This framework can produce very different species classifications than the biological species concept. For example, under the ecological species concept, most of Darwin's finch species would qualify as species as they exploit different food sources, but under the biological species concept their ability to interbreed would reject separate species status (Lamichhaney et al. 2015). As we will see in later chapters, polar bears and brown bears are clearly distinct species under the ecological species concept but are less well defined under the biological species concept (Chapter 3). In general, the ecological species concept splits species into narrower groups than the biological species concept. However, because the niche is an abstract concept the ecological species concept provides a more qualitative description of species than the

biological species concept.

It is increasingly clear that speciation should be considered as a process rather than a single event (Abbott et al. 2013). As populations diverge, they may accumulate differences that gradually increase the costs of interbreeding. It is now clear that adaptive divergence generally predates reproductive isolation (Abbott et al. 2013) and that the diversification of life exists as a continuum (Mallet 2005). Although theory places great emphasis on hybrid incompatibility (Dobzhansky 1937), the role of behavioral isolation between young species may be under-appreciated. In many fish species, for example, human-caused increases in water turbidity have led to dramatic increases in the rate of hybridization as the ability of these organisms to recognize conspecific mates is decreased (Seehausen et al. 1997). This dissertation will explore the process of speciation in polar bears and brown bears, specifically focusing on the extent of admixture (Chapters 2, 3, 5) and the forces that may promote admixture (Chapters 2, 4).

Mechanisms of Speciation

In my discussion of the definition of species, I have so far not explicitly focused on how species arise. One of the central appeals of the Biological Species Concept is that it captures a fundamental biological reality; interbreeding is a powerful homogenizing force that prevents diversification (Wright 1931). Therefore, mechanisms of speciation must first provide a means of establishing reproductive isolation, at least in the immediate term and then more permanent barriers may or may not evolve.

The most widely accepted model of speciation is therefore allopatric speciation, which proposes that speciation begins when some previously absent physical barrier separates a group of interbreeding individuals into separate mutually isolated populations. In this model, isolating mechanisms that evolve

first arise for some reason other than promoting preventing interbreeding because as the populations are isolated so preventing interbreeding is irrelevant. Following re-contact, incipient species that are without sufficient reproductive isolating mechanisms will coalesce back into a single species while those with sufficient reproductive isolation may develop further behaviors or traits to avoid interbreeding, if any interbreeding producing unfit hybrids occurs. That allopatric speciation can occur is inevitable because without the homogenizing effect of interbreeding, the constant accumulation of mutations means that isolated populations will diverge over time.

More controversial is whether sympatric speciation, speciation without geographic isolation can occur in animals. An intriguing potential case of incipient sympatric species are a group of hybrid Darwin's finches on Daphne Major, a small island in the Galapagos Archipelago. The hybrids and both parent species continue to live on the island but since the initial hybridization event the hybrids have only interbred with other hybrids and have not backcrossed with either parent species (Grant & Grant 2009). Song plays an important role in mating in finches and these finches have a unique song that is in effect a behaviorally derived isolating mechanism (Grant & Grant 2009). However, the authors note that allopatric isolation prior to the hybridization event played an important role in generating the conditions for the observed sympatric isolation (Grant & Grant 2009). It seems plausible that a similar behaviorally mediated sympatric speciation or at least maintenance of isolation between interfertile individuals may be at play in the African Great Lakes Cichlid radiation given that increased water turbidity which obscures potential mates and thereby inhibits behavioral isolation has led to an increase in hybridization (Seehausen et al. 1997).

Whether isolation between species is physical or behavioral, it likely plays

a key role in the early stages of the speciation process. Hybridization is the violation of isolation and so by studying hybridization between species we can investigate what factors regulate hybridization, how the speciation process occurs, what factors inhibit divergence, and when isolation finally occurs.

Hybridization

Hybridization is interbreeding between separate species. Hybrid individuals exhibit a range of characteristics and fitness effects that differ between species pairs. Hybrids may be completely sterile (mules), have reduced fertility potentially exhibiting infertility in some mating combinations but not others (mice, *Mus musculus, Mus domesticus* hybrids)(Good, Dean, et al. 2008; Good, Handel, et al. 2008), or be fertile (some gibbons, Darwin's finches, polar bears and brown bears)(Lamichhaney et al. 2015; Carbone et al. 2014; Preuß et al. 2009). In addition to reproductive traits, many other characteristics can be impacted by hybridization. Hybridization may give rise to new combinations of alleles that are potentially advantageous (Becker et al. 2013) or disrupt gene regulatory systems producing phenotypes dramatically outside the range of the parent species such as Lion-Tiger hybrids which are much larger than either parent species.

If some hybrids are reproductively viable, admixture – the transfer of genetic material between species – may occur. Admixed populations, those populations with ancestry from multiple species may arise. Historically, the extent and evolutionary significance of admixture was underestimated by biologists, particularly zoologists (Abbott et al. 2013). However, with the advent of genome sequencing technology, a much greater role for admixture has been revealed (Green et al. 2010; Reich et al. 2011; Lamichhaney et al. 2015; Carbone et al. 2014; Dasmahapatra et al. 2012). This transformation in our knowledge of the role of admixture in evolution from genomic data is largely a consequence of two

characteristics of these data that make them unusually suited to detecting admixture. Genome sequences provide temporal depth that is not available in field studies. Greater temporal depth allows the detection of admixture with now extinct species (Green et al. 2010) and admixture events between extant species that took place in the past but have now ceased (Schumer et al. 2016). Second, hybridization has a profound impact on genomic diversity compared to its frequency and therefore even rare hybridization events can produce substantial and detectable admixture within a population. For example, non-African modern humans have 1-4% Neanderthal ancestry (Green et al. 2010), which could be explained by a single generation where one in fifty humans were first generation hybrids. However, a more plausible explanation would be several generations with even lower rates of neanderthal parentage. As shown in figure 0.1, over 20 generations of introgression, one parent in 1,000 is sufficient to produce almost 2% introgressed ancestry, and those generations need not be consecutive.

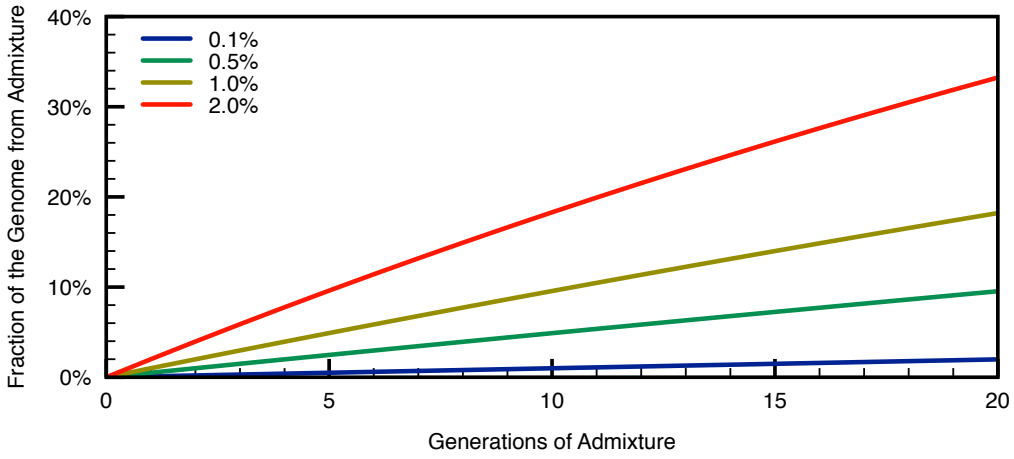


Figure 0.1. An estimation of the amount of introgressed ancestry in a population over 20 generations of admixture at different rates of introgression. For simplicity, this assumes that introgressed ancestry is selectively neutral and all introgressing individuals are themselves unadmixed.

However, as my collaborators and I demonstrate in the following studies, it is likely that admixture is not necessarily confined to peripheral drips of

introgression. Our studies (Chapters 1, 3), as well as some others in different systems (Zinner et al. 2009; vonHoldt et al. 2016; Grant & Grant 2009; Carbone et al. 2014), suggest that breakdowns in species isolation can occur in bursts when environmental conditions change in a manner that locally disrupts isolation by removing physical, ecological, or behavioral barriers to hybridization.

Genomics

The genome is the entirety of the genetic material contained within the cell nucleus. Genomes encode the entirety of the functional information that, following transcription into RNA and translation into proteins, control the development, form, and behavior of the organism. Eukaryotic genomes are also made up of functional non-genic DNA (Lander et al. 2001), which have regulatory or structural functions and apparently non-functional genomic parasites such as repetitive elements (Kellis et al. 2014). The genome also provides us with a uniquely powerful record of an organism's recent evolutionary history, which allows us to investigate many biologically important processes.

The enzymes that replicate DNA during each cell division and repair damaged DNA are imperfect and the resulting mutations occur with an approximately constant frequency for a given species (Zuckerkandl & Pauling 1962). Most mutations are weakly deleterious or selectively neutral so their continued inheritance in a population is determined by genetic drift, rather than natural selection (Ohta 1973; Kimura 1968). The random and approximately clocklike accumulation of mutations allows us to infer the time since two individuals shared a common ancestor from the number of differences between their genetic sequences (Zuckerkandl & Pauling 1962). Through recombination and inheritance the genome contains many segments that each may capture a different aspect of an organisms ancestry and therefore a different aspect of the

relationship between two or more organisms. By sequencing whole genomes we can investigate the entire available history of an organism maximizing our insight into that individual's evolutionary history.

Genome sequencing began in 1976 with the sequencing of the first RNA-virus (bacteriophage MS2) (Fiers et al. 1976) and the following year the sequencing of the first DNA virus genome (bacteriophage φX174)(Sanger et al. 1977). The first bacteria sequenced was *Haemophilus influenzea* (Fleischmann et al. 1995), the first eukaryote was *Saccharomyces cerevisiae* (Cherry et al. 1997) and the first vertebrate was the human genome (Lander et al. 2001; Venter et al. 2001). Subsequently the rate of genome sequencing has accelerated dramatically.

De novo genome assembly is an essential first step in many genomic studies, although the vast majority of genomic studies do not involve this type of genome assembly. Rather, these studies, including all of the studies in this manuscript, make use of reference-based alignments, sometimes termed re-sequencing. Once a reference genome has been generated, short DNA sequence reads (described in more detail later in this chapter), are compared to the reference genome to determine where there is a location within the reference genome sufficiently similar to the read that it is considered likely to be derived from the same position in the reference genome. An obvious limitation of this approach is that any portion of the sequenced individual's genome that is not present in the reference genome cannot be recovered.

Furthermore, the more divergent the reference genome and the sequenced individual are at homologous regions of the genome, the less likely reads are to be aligned to the reference genome (Prüfer et al. 2010). This "reference bias" can lead to the exclusion of divergent regions and more perniciously the successful mapping of only the more reference like allele at heterozygous sites. The extent of these biases depends greatly upon the specific mapping parameters used, the

evolutionary divergence separating the reference genome from the sequenced individual and the quality of the sequencing reads in terms of read length, sequencing error rate and DNA damage, such as cytosine deamination (Prüfer et al. 2010). In practice, this limitation must be addressed on a study-by-study basis depending on the reference genome and data being used so I will leave detailed methodological descriptions to their respective chapters.

Ancient DNA

DNA is a remarkably resilient molecule that can be preserved in nature for hundreds of thousands of years given suitable environmental conditions (Orlando et al. 2013; Miteva et al. 2004). Using ancient DNA, we can directly investigate how populations responded to environmental changes (Shapiro et al. 2004; Barnes et al. 2002; Stiller et al. 2010), study the diversity and evolutionary relationships of extinct species and associate fossils and ancient pathogens with extant groups (Krause et al. 2008; Heintzman et al. 2015; Raoult et al. 2000; Bos et al. 2011; Shapiro et al. 2004; Lorenzen et al. 2011). However, ancient DNA research also experiences a broad array of unique challenges: DNA breaks down into short fragments (Handt et al. 1994); Cytosine deamination damage converts Cytosine to Uracil (Hofreiter 2001); environmental (exogenous) DNA occurs at high rates within many samples (Green et al. 2006); inhibitory compounds may become associated with a sample rendering DNA impossible to extract (Handt et al. 1994); and the total amount of DNA present in any sample is very low making it highly susceptible to contamination by modern DNA (Handt et al. 1994). Despite the many difficulties associated with working with ancient DNA, high throughput sequencing, and especially Illumina short read sequencing, are very powerful tools for ancient DNA research dramatically increasing analytical power (Green et al. 2010; Orlando et al. 2013; Reich et al. 2011; Prüfer et al. 2014; Heintzman et al. 2015).

Bear Evolution

Bears (family Ursidae) are mammals in the order *Carnivora*, suborder Caniformia. Bears originated in Eurasia and began to diversify at least 11-12 million years ago (Abella et al. 2012). Their nearest extant relatives are the mustelids (weasels, ferrets, and kin; family Mustelidae) and pinnipeds (seals, sea lions; families Phocidae, Otariidae and Odobenidae). The three groups are thought to have diverged in rapid succession ~34.8-48.3 million years ago (Eizirik et al. 2010). Phylogenetic studies suggest that bears are most likely the first lineage to diverge (Eizirik et al. 2010), however, the lineages diversified quickly and which of the three clades is the outgroup to the other two is not consistent across loci (Eizirik et al. 2010; Fulton & Strobeck 2006; Delisle & Strobeck 2005).

Extant ursid species fall into three genera; *Ailuropoda*, *Tremarctos* and *Ursus*. The most basal lineage, *Ailuropoda,* is thought to have diverged from the other Ursids about 14.4-24.8 million years ago (Krause et al. 2008), with diagnostic fossils as providing a minimum divergence of more than 11-12 million years ago (Abella et al. 2012). The giant panda (*Ailuropoda melanoleuca*) is the only extant member of the genus and was the first Ursid to have it's genome sequenced in 2010 (Li et al. 2010). *Tremarctos* also contains only a single extant representative, the spectacled bear (*T. ornatus*). The spectacled bear is the only surviving representative of the subfamily Tremarctinae (short-faced bears) which diverged from *Ursus* bears about 9.8-16.6 million years ago (Krause et al. 2008) and were the first bears to enter the Americas. Extinct tremarctines include the genera, the giant short-faced bears genera *Arctodus* and *Arctotherium* and *T. floridianus* all of which persisted until the late Pleistocene (Soibelzon et al. 2005).

The remaining six extant bear species are members of the genus *Ursus.* Although *Ursus* has been subdivided into as many as 5 genera (*Ursus, Thalarctos, Melursus, Selenarctos and Helarctos*), all six species share a very recent common

ancestor within the last 4.2-6.9 million years (Krause et al. 2008). *Ursus* bears diversified very rapidly limiting phylogenetic resolution, to date the exact pattern of diversification and relationship between *Ursus* bears remains incompletely understood. The only robustly established phylogenetic relationships are the nearest relative relationship of polar bears (*U. maritimus*) and brown bears (*Ursus arctos*), and the nearest relative to those two species being the extinct cave bear (*Ursus spelaeus*).

Polar bears and Brown bears

The exact evolutionary relationship of polar bears and brown bears has been the subject of substantial debate in recent years. This is largely because post-divergence admixture between the polar and brown bears has produced an unexpectedly complicated mitochondrial phylogeny that was not consistent with the genome wide pattern. Using exclusively morphological data, Bjorn Kurten proposed that polar bears and brown bears were near relatives. He further proposed that polar bears arose from a population of brown bear like ancestors that became isolated in northern Russia and increasingly made use of, and adapted to, the sea ice habitat (Kurten 1964). Although the specific details of Kurten's hypothesis, such as the location of the initial sea ice adapted population, are difficult to substantiate the broad pattern appears to be supported by genomic data. As the specific pattern of relatedness between polar bears and brown bears is the major theme of this dissertation, I will focus here on work conducted prior to 2013, the publication date of chapter 2.

Early genetic analysis of bears, conducted using allozymes and published in 1988, identified polar bears and brown bears as one another's nearest extant relative and estimated the divergence time to be 2-3 million years (Goldman et al. 1989). More detailed and more robustly calibrated allozyme based estimates of polar bear and brown bear divergence placed the divergence time at only 90

thousand years ago (Wayne et al. 1991). Mitochondrial restriction enzyme fragment length polymorphism and cytochrome b sequencing further supported the close relationship between polar bears and brown bears (Shields & Kocher 1991).

Partial mitochondrial genome sequencing by Cronin and colleagues in 1991 revealed that polar bear's mitochondrial haplotypes fell within the diversity of North American brown bears. Specifically, Alaskan brown bears from three large islands Admiralty, Baranof and Chichagof (ABC) Islands, mitochondrial haplotypes more closely related to polar bears than to other brown bears (Cronin et al. 1991). A more diverse sampling of Alaskan brown bear mitochondrial DNA confirmed the relationship between polar bears and brown bears, including the ABC islands-polar bear nearest mitochondrial relative relationship (Talbot & Shields 1996). North American brown bear diversity was further classified in 1998 by Waits and colleagues who established four major brown bear mitochondrial clades (Waits et al. 1998).

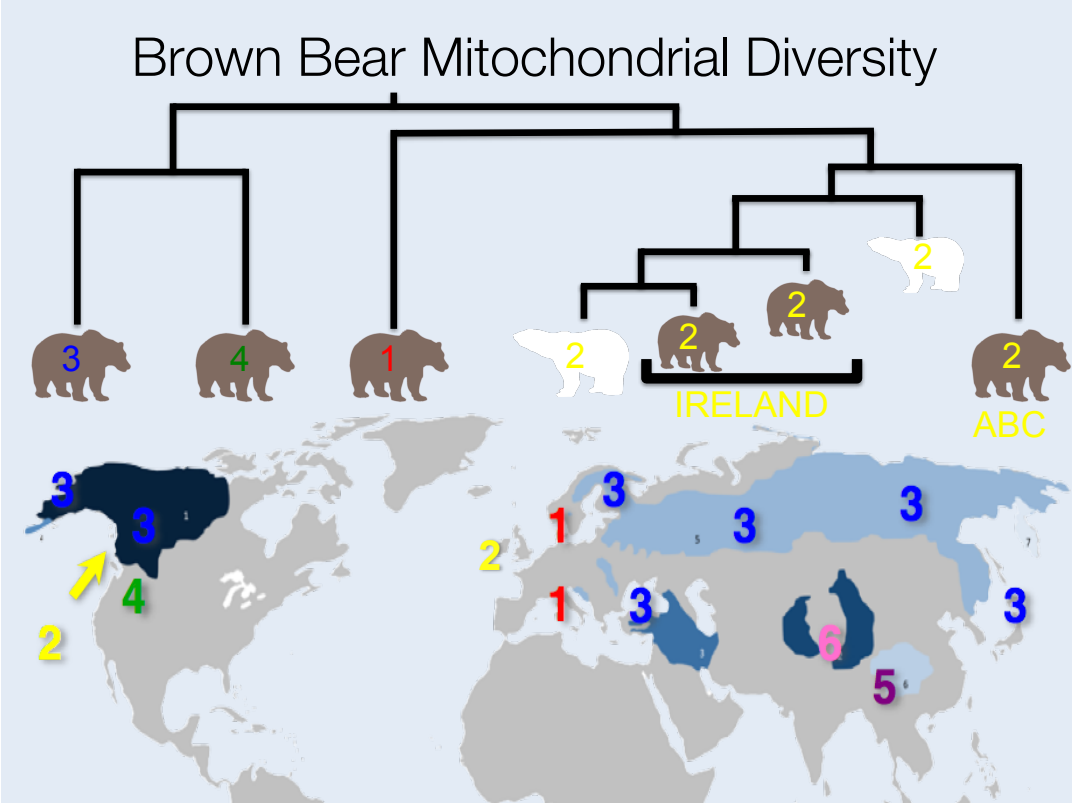


Figure 0.2. The geographic distribution of brown bear mitochondrial haplotypes (bottom) and a simplified phylogeny of extant and extinct brown bear populations based on (Edwards et al. 2011) (top). Polar bears fall within clade 2.

Following the mitochondrial analyses, early nuclear genomic analyses, primarily using microsatellites, began to assess the geographic distribution of diversity in polar bears and brown bears (Paetkau et al. 1998; Paetkau et al. 1999; TABERLET et al. 1995; Taberlet & Bouvet 1994). Microsatellite data led to the subdivision of polar bears into 16 populations (Paetkau et al. 1999). Subsequent analysis have continued to broadly support these populations (Peacock et al. 2015) although, polar bears have very little genetic diversity (Cahill et al. 2013) which raises questions as to the biological importance of these groupings.

Brown bears are much more diverse and widespread than polar bears and comparatively less well studied. As a result of these factors, brown bear studies have been more regionally focused although it should be noted that polar bear studies have historically underrepresent Russian populations. Alaskan brown

bear diversity was documented by Paetkau and colleagues who assessed a range of populations around Alaska. Paetkau et al. 1998 found island populations had reduced diversity but intriguingly they also found that ABC Islands bears group very closely with nearby mainland brown bears (Paetkau et al. 1998). These results revealed the discordance between the mitochondrial DNA which clearly distinguishes the ABC islands brown bears from mainland brown bears (Cronin et al. 1991) and the nuclear genome which do not support the same degree of isolation (Paetkau et al. 1998). Paetkau and colleagues proposed that this nuclear/mitochondrial discordance was the result of male mediated gene flow between the islands and the mainland (Paetkau et al. 1998).

Ancient DNA provided the next group of significant advances in polar and brown bear evolutionary history. In 2002 Barnes and colleagues published partial mitochondrial genome data from Beringian brown bears living from the present to the before the limit of radiocarbon dating, given in that paper as 60 thousand years (Barnes et al. 2002). Barnes and colleagues found substantial turnover in mitochondrial haplotypes in brown bears living in Beringia before and after the LGM consistent with substantial recent (<35 thousand years ago) population movement. Most importantly for this study they identified a subset of pre-LGM Beringian brown bears that possessed polar bear like mitochondrial haplotypes that form a distinct subgroup of clade II from the ABC islands brown bears (Barnes et al. 2002).

In 2010 Lindqvist and colleagues published the full mitochondrial genome of a 110-130 thousand year old polar bear mandible from Svalbard, Norway. This bear's mitochondrial haplotype was more closely related to polar bears than extant brown bears, including ABC islands brown bears which the ancient Svalbardian polar bear shared a matrilineal ancestor with 10-30 thousand years prior to it's death (Lindqvist et al. 2010). Although the Svalbard mandible had a

derived morphology very similar to modern polar bears (Ingólfsson & Wiig 2009) the mitochondrial common ancestor of the polar bears and the ABC islands brown bears was interpreted to have lived prior to the origin of polar bears. However, the next year an analysis of mitochondrial DNA from Pleistocene brown bear in Ireland would find that some Irish brown bears living between 9-32 thousand years ago (cal. BP) had mitochondrial haplotypes that were more closely related to modern polar bears than to the Svalbard fossil. This provided the first clear genetic evidence for hybridization between polar bears and brown bears, as the shared maternal ancestor of Irish bears and modern polar bears within the last 30 thousand years was not otherwise reconcilable with the 110-130 thousand year old polar bear from Svalbard (Edwards et al. 2011). At the time the mitochondrial results were interpreted as multiple introgressions of brown bear mitochondria into polar bears because polar bear mitochondria fall within the diversity of brown bear mitochondria (Edwards et al. 2011).

Nuclear genome sequence based assessment of the divergence between polar bears and brown bears began in 2012 when Hailer and colleagues published a phylogeny based on 14 nuclear loci. This phylogeny showed that polar bears and brown bears formed distinct clades as the nuclear genome level, although not all loci were clearly sorted between species (Hailer et al. 2012). Hailer et al estimated the most recent common ancestor of polar bears and brown bears to have lived about 338-934 thousand years ago (Hailer et al. 2012). Shortly thereafter Miller and colleagues published the first whole genome analysis of polar bears and brown bears. This analysis indicated gene flow between polar bears and brown bears, as well as American black bears (*Ursus americanus*) and a very ancient 4-5 million year old divergence between the three species with ongoing gene flow between all three species (Miller et al. 2012). They also estimated the ancestral population size of brown bears and polar bears using

Pairwise Sequentially Markova Coalescent (PSMC) (Li & Durbin 2011)(discussed in detail in Chapter 4), a program for estimating ancestral population size through time from the distribution of heterozygosity within a single genome. This revealed that ancestral polar bear population size declined dramatically about 300 thousand years ago and has remained very low, but relatively constant, since that time (Miller et al. 2012).

My first chapter, published in early 2013, investigated the relationship between ABC islands brown bears and polar bears. Chapter 2, published in 2015, expanded the scope of this investigation to other populations and Chapter 4, published in 2016, also contributed to the investigation of admixture in modern populations. In Chapter 3, I investigated admixture in ancient populations as well as modern populations. The implications of each of these studies are addressed in their respective chapters.

Cave Bears

In chapter 5 I present a collaborative project that I co-led with Axel Barlow investigating the genomic diversity of cave bears and their relationship to their near relatives. Cave bears (*Ursus spelaeus*) are among the most common fossils in Europe and are known to have ranged across most of Northern Eurasia during the Pleistocene (KNAPP et al. 2009). Mitochondrial DNA evidence shows that cave bears' nearest evolutionary relatives are polar bears and brown bears which share a common ancestor prior to their common ancestor with cave bears (Krause et al. 2008).

European cave bears went extinct during the last glacial maximum ~20-25 thousand years ago following a long and gradual decline in numbers as evidenced by mitochondrial haplotype diversity (Stiller et al. 2010). It is hypothesized that this gradual decline may have been caused by competition with humans for cave sites, which the cave bears used for hibernation (PACHER

& STUART 2009; Stiller et al. 2010). While this presents and intriguing possibility the story is complicated somewhat by the existence of mitochondrially divergent cave bears that lived in central Asia (KNAPP et al. 2009; Dabney et al. 2013) and may or may not have been subject to the same pressures as their European relatives at that time.

Although cave bears are relatively well studied for an ancient taxon and have been used as a model system for the development of ancient DNA technologies (Dabney et al. 2013) nuclear genomic studies are lacking from the published literature. As a result there is much to be learned about and from the genomes of these near relatives of polar bears and brown bears.

Conclusion

The studies that follow are my first author publications (Chapters 1, 2, 4) and first author and co-first author unpublished studies (Chapters 3, 5) from my graduate research conducted between November 2010 and the time of the submission of this dissertation August 2016. In that time I have been fortunate to be involved with a number of excellent collaborators and contributed to several studies in a non-first author capacity. However, these studies represent the core of my research as a graduate student and my main contribution to the corpus of scientific knowledge.

References

Abbott, R. et al., 2013. Hybridization and speciation. *Journal of evolutionary biology*, 26(2), pp.229–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23323997 [Accessed October 18, 2013].

Abella, J. et al., 2012. Kretzoiarctos gen. nov., the oldest member of the giant panda clade. *PloS one*, 7(11), p.e48985. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0048985 [Accessed February 4, 2015].

Barnes, I. et al., 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science (New York, N.Y.)*, 295(5563), pp.2267–70. Available at: http://www.sciencemag.org/content/295/5563/2267.short [Accessed October 23, 2013].

Becker, M. et al., 2013. Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change*, 3(12), pp.1039–1043. Available at: http://www.nature.com/doifinder/10.1038/nclimate2027.

Bos, K.I. et al., 2011. A draft genome of Yersinia pestis from victims of the Black Death. *Nature*, 478(7370), pp.506–510. Available at: http://www.nature.com/doifinder/10.1038/nature10549 [Accessed August 15, 2016].

Cahill, J.A. et al., 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. M. W. Nachman, ed. *PLoS genetics*, 9(3), p.e1003345. Available at: http://dx.plos.org/10.1371/journal.pgen.1003345 [Accessed October 19, 2013].

Cahill, J.A. et al., 2015. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular ecology*, 24(6), pp.1205–1217. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25490862 [Accessed December 11, 2014].

Carbone, L. et al., 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*, 513(7517), pp.195–201. Available at: http://www.nature.com/doifinder/10.1038/nature13679 [Accessed August 16, 2016].

Cherry, J.M. et al., 1997. Genetic and physical maps of Saccharomyces cerevisiae. *Nature*, 387(6632 Suppl), pp.67–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9169866 [Accessed August 16, 2016].

Cronin, M. a. et al., 1991. Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursus). *Canadian Journal of Zoology*, 69(12), pp.2985–2992. Available at: http://www.nrcresearchpress.com/doi/abs/10.1139/z91-421.

Dabney, J. et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39), pp.15758–15763. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1314445110 [Accessed August 21, 2016].

Dasmahapatra, K.K. et al., 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), pp.94–98. Available at: http://www.nature.com/doifinder/10.1038/nature11041.

Delisle, I. & Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. *Molecular Phylogenetics and Evolution*, 37(1), pp.192–201.

Dobzhansky, T., 1937. *Genetics and the Origin of Species*, New York, NY: Columbia University Press.

Edwards, C.J. et al., 2011. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology : CB*, 21(15), pp.1251–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21737280 [Accessed October 23, 2013].

Eizirik, E. et al., 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 56(1), pp.49–63.

Fiers, W. et al., 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, 260(5551), pp.500–507. Available at: http://www.nature.com/doifinder/10.1038/260500a0 [Accessed August 16, 2016].

Fleischmann, R.D. et al., 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*, 269(5223), pp.496–512. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7542800 [Accessed August 16, 2016].

Fulton, T.L. & Strobeck, C., 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. *Molecular Phylogenetics and Evolution*, 41(1), pp.165–181.

Goldman, D., Giri, P.R. & O'Brien, S.J.., 1989. Molecular Genetic-Distance Estimates Among the Ursidae as Indicated by One- and Two- Dimensional Protein Electrophoresis Published by : Society for the Study of Evolution Stable URL : *Evolution*, 43(2), pp.282–295.

Good, J.M., Dean, M.D. & Nachman, M.W., 2008. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics*, 179(4), pp.2213–28. Available at: http://www.genetics.org/content/179/4/2213.short [Accessed July 19, 2014].

Good, J.M., Handel, M.A. & Nachman, M.W., 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution; international journal of organic evolution*, 62(1), pp.50–65. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2907743&tool=pmcentrez&rendertype=abstract [Accessed July 15, 2014].

Grant, P.R. et al., 2004. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution; international journal of organic evolution*, 58(7), pp.1588–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15341160 [Accessed August 15, 2014].

Grant, P.R. & Grant, B.R., 2009. The secondary contact phase of allopatric speciation in Darwin's finches. *Proceedings of the National Academy of Sciences*, 106(48), pp.20141–20148. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0911761106 [Accessed August 16, 2016].

Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, 328(5979), pp.710–22. Available at: http://www.sciencemag.org/content/328/5979/710.abstract [Accessed October 17, 2013].

Green, R.E. et al., 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444(7117), pp.330–336. Available at: http://www.nature.com/doifinder/10.1038/nature05336 [Accessed August

15, 2016].

Hailer, F. et al., 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science (New York, N.Y.)*, 336(6079), pp.344–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22517859 [Accessed October 17, 2013].

Handt, O. et al., 1994. Ancient DNA: methodological challenges. *Experientia*, 50(524).

Heintzman, P.D. et al., 2015. Genomic Data from Extinct North American Camelops Revise Camel Evolutionary History. *Molecular Biology and Evolution*, 32(9), pp.2433–2440. Available at: http://mbe.oxfordjournals.org/lookup/doi/10.1093/molbev/msv128.

Hofreiter, M., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, 29(23), pp.4793–4799. Available at: http://nar.oxfordjournals.org/content/29/23/4793 [Accessed January 15, 2014].

Hutchinson, G.E., 1957. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0), pp.415–427. Available at: http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1957.022.01.039 [Accessed August 16, 2016].

Ingólfsson, Ó. & Wiig, Ø., 2009. Late Pleistocene fossil find in Svalbard: the oldest remains of a polar bear (Ursus maritimus Phipps, 1744) ever discovered. *Polar Research*, 28(3), pp.455–462. Available at: http://www.polarresearch.net/index.php/polar/article/view/6131 [Accessed October 24, 2013].

Kellis, M. et al., 2014. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences*, 111(17), pp.6131–6138. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1318948111 [Accessed August 17, 2016].

Kimura, M., 1968. Evolutionary Rate at the Molecular Level. *Nature*, 217(5129), pp.624–626. Available at: http://dx.doi.org/10.1038/217624a0 [Accessed January 19, 2015].

KNAPP, M. et al., 2009. First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Molecular Ecology*, 18(6), pp.1225–1238. Available at: http://doi.wiley.com/10.1111/j.1365-294X.2009.04088.x [Accessed August 21, 2016].

Krause, J. et al., 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC evolutionary biology*, 8(1), p.220. Available at: http://www.biomedcentral.com/1471-2148/8/220 [Accessed October 24, 2013].

Kurten, B., 1964. The evolution of the Polar Bear, Ursus maritimus Phipps. *Acta zoologica Fennica*, 108, p.30.

Lamichhaney, S. et al., 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), pp.371–375. Available at: http://dx.doi.org/10.1038/nature14181 [Accessed February 11, 2015].

Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome.

Nature, 409(6822), pp.860–921. Available at: http://www.nature.com/doifinder/10.1038/35057062 [Accessed August 16, 2016].

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), pp.493–6. Available at: http://dx.doi.org/10.1038/nature10231 [Accessed July 10, 2014].

Li, R. et al., 2010. The sequence and de novo assembly of the giant panda genome. *Nature*, 463(7279), pp.311–7. Available at: http://dx.doi.org/10.1038/nature08696 [Accessed October 18, 2013].

Lindqvist, C. et al., 2010. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), pp.5053–7. Available at: http://www.pnas.org/content/107/11/5053.short [Accessed October 19, 2013].

Lorenzen, E.D. et al., 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479(7373), pp.359–364. Available at: http://www.nature.com/doifinder/10.1038/nature10574 [Accessed August 15, 2016].

Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), pp.229–237.

Mayr, E., 1940. Speciation Phenomena in Birds. *The American Naturalist*, 74(752), pp.249–278.

Mayr, E., 1942. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist - Ernst Mayr - Google Books*, Cambridge, Massachusetts: Harvard University Press. Available at: https://books.google.com/books?hl=en&lr=&id=mAIjnLp6r_MC&oi=fnd&pg=PR9&dq=systematics+and+the+origin+of+species&ots=TSKrIrCfRF&sig=HTyy6o3GeUrOWy2oUI9cdKXeCnI#v=onepage&q=systematics and the origin of species&f=false.

Miller, W. et al., 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), pp.E2382–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3437856&tool=pmcentrez&rendertype=abstract [Accessed October 21, 2013].

Miteva, V.I., Sheridan, P.P. & Brenchley, J.E., 2004. Phylogenetic and Physiological Diversity of Microorganisms Isolated from a Deep Greenland Glacier Ice Core. *Applied and Environmental Microbiology*, 70(1), pp.202–213. Available at: http://aem.asm.org/cgi/doi/10.1128/AEM.70.1.202-213.2004 [Accessed August 16, 2016].

Ohta, T., 1973. Slightly Deleterious Mutant Substitutions in Evolution. *Nature*, 246(5428), pp.96–98. Available at: http://dx.doi.org/10.1038/246096a0 [Accessed March 11, 2016].

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–8. Available at: http://dx.doi.org/10.1038/nature12323 [Accessed October 18, 2013].

PACHER, M. & STUART, A.J., 2009. Extinction chronology and palaeobiology of the cave bear (*Ursus spelaeus*). *Boreas*, 38(2), pp.189–206. Available at: http://doi.wiley.com/10.1111/j.1502-3885.2008.00071.x [Accessed August 21, 2016].

Paetkau, D. et al., 1999. Genetic structure of the world's polar bear populations. *Molecular Ecology*, 8(10), pp.1571–1584. Available at: http://doi.wiley.com/10.1046/j.1365-294x.1999.00733.x [Accessed October 25, 2013].

Paetkau, D., Shields, G.F. & Strobeck, C., 1998. Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, 7(10), pp.1283–1292. Available at: http://doi.wiley.com/10.1046/j.1365-294x.1998.00440.x [Accessed October 24, 2013].

Peacock, E. et al., 2015. Implications of the circumpolar genetic structure of polar bears for their conservation in a rapidly warming arctic. *PloS one*, 10(1), p.e112021. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112021#pone-0112021-g003 [Accessed January 12, 2015].

Poelstra, J.W. et al., 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science (New York, N.Y.)*, 344(6190), pp.1410–4. Available at: http://www.sciencemag.org/content/344/6190/1410.short [Accessed October 29, 2015].

Preuß, A. et al., 2009. Bear-hybrids: behaviour and phenotype. *Der Zoologische Garten*, 78(4), pp.204–220. Available at: http://www.sciencedirect.com/science/article/pii/S0044516909000276 [Accessed October 24, 2013].

Prüfer, K. et al., 2010. Computational challenges in the analysis of ancient DNA. *Genome biology*, 11(5), p.R47. Available at: http://genomebiology.com/2010/11/5/R47 [Accessed May 13, 2015].

Prüfer, K. et al., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), pp.43–9. Available at: http://dx.doi.org/10.1038/nature12886 [Accessed July 10, 2014].

Raoult, D. et al., 2000. Molecular identification by "suicide PCR" of Yersinia pestis as the agent of Medieval Black Death. *Proceedings of the National Academy of Sciences*, 97(23), pp.12800–12803. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.220225197 [Accessed August 15, 2016].

Reich, D. et al., 2011. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics*, 89(4), pp.516–528. Available at: http://www.sciencedirect.com/science/article/pii/S0002929711003958 [Accessed October 24, 2013].

Sanger, F. et al., 1977. Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, 265(5596), pp.687–695. Available at: http://www.nature.com/doifinder/10.1038/265687a0 [Accessed August 16, 2016].

Schumer, M. et al., 2016. Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology*, 25(11), pp.2661–2679. Available at:

http://doi.wiley.com/10.1111/mec.13602 [Accessed August 17, 2016].

Seehausen, O., Alphen, J.J.M. van & Witte, F., 1997. Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. *Science*, 277(5333).

Seyfarth, R.M. & Cheney, D.L., 1980. The Ontogeny of Vervet Monkey Alarm Calling Behavior: A Preliminary Report. *Zeitschrift für Tierpsychologie*, 54(1), pp.37–56. Available at: http://doi.wiley.com/10.1111/j.1439-0310.1980.tb01062.x [Accessed August 24, 2016].

Shapiro, B. et al., 2004. Rise and fall of the Beringian steppe bison. *Science (New York, N.Y.)*, 306(5701), pp.1561–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15567864 [Accessed August 15, 2016].

Shields, G.F. & Kocher, T.D., 1991. Phylogenetic Relationships of North American Ursids Based on Analysis of Mitochondrial DNA. *Evolution*, 45(1), pp.218–221.

Soibelzon, L.H., Tonni, E.P. & Bond, M., 2005. The fossil record of South American short-faced bears (Ursidae, Tremarctinae). *Journal of South American Earth Sciences*, 20(1), pp.105–113.

Stiller, M. et al., 2010. Withering Away--25,000 Years of Genetic Decline Preceded Cave Bear Extinction. *Molecular Biology and Evolution*, 27(5), pp.975–978. Available at: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msq083 [Accessed August 16, 2016].

TABERLET, P. et al., 1995. Localization of a Contact Zone between Two Highly Divergent Mitochondrial DNA Lineages of the Brown Bear Ursus arctos in Scandinavia. *Conservation Biology*, 9(5), pp.1255–1261. Available at: http://doi.wiley.com/10.1046/j.1523-1739.1995.951255.x [Accessed August 17, 2016].

Taberlet, P. & Bouvet, J., 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear Ursus arctos in Europe. *Proceedings. Biological sciences / The Royal Society*, 255(1344), pp.195–200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8022838 [Accessed August 17, 2016].

Talbot, S.L. & Shields, G.F., 1996. Phylogeography of Brown Bears (Ursus arctos) of Alaska and Paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, 5(3), pp.477–494. Available at: http://www.sciencedirect.com/science/article/pii/S1055790396900445 [Accessed October 24, 2013].

Van Valen, L., 1976. Ecological Species, Multispecies, and Oaks. *Taxon*, 25(2/3), p.233. Available at: http://www.jstor.org/stable/1219444?origin=crossref [Accessed August 16, 2016].

Venter, J.C. et al., 2001. The Sequence of the Human Genome. *Science*, 291(5507), pp.954–956.

vonHoldt, B.M. et al., 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances*, 2(7).

Waits, L.P. et al., 1998. Mitochondrial DNA American for Brown Bear

Phylogeography of the and North. *Conservation Biology*, 12(2), pp.408–417.

Wayne, R., Van Valkenburgh, B. & O'Brien, S., 1991. Molecular distance and divergence time in carnivores and primates. *Mol. Biol. Evol.*, 8(3), pp.297–319. Available at: http://mbe.oxfordjournals.org/content/8/3/297 [Accessed July 31, 2015].

Wright, S., 1931. Evolution in mendelian populations. *Genetics*, 16, pp.97–159.

Zinner, D. et al., 2009. Mitochondrial phylogeography of baboons (Papio spp.) – Indication for introgressive hybridization? *BMC Evolutionary Biology*, 9(1), p.83. Available at: http://www.biomedcentral.com/1471-2148/9/83.

Zuckerkandl, E. & Pauling, L., 1962. Molecular Disease, Evolution, and Genic Heterogeneity. In M. Kasha & B. Pullman, eds. *Horizons in Biochemistry*. New York, NY: Academic Press, pp. 189–222.

Chapter 1: Genomic Evidence for Island Population Conversion Resolves Conflicting Theories of Polar Bear Evolution

James A. Cahill, Richard E. Green, Tara L. Fulton, Mathias Stiller, Flora Jay, Nikita Ovsyanikov, Rauf Salamzade, John St. John, Ian Stirling, Montgomery Slatkin, Beth Shapiro

Originally published in *PLoS Genetics*, 9(3), 2013.

Abstract

Despite extensive genetic analysis, the evolutionary relationship between polar bears (*Ursus maritimus*) and brown bears (*U. arctos*) remains unclear. The two most recent comprehensive reports indicate a recent divergence with little subsequent admixture or a much more ancient divergence followed by extensive admixture. At the center of this controversy are the Alaskan ABC Islands brown bears that show evidence of shared ancestry with polar bears. We present an analysis of genome-wide sequence data for seven polar bears, one ABC Islands brown bear, one mainland Alaskan brown bear, and a black bear (*U. americanus*), plus recently published datasets from other bears. Surprisingly, we find clear evidence for gene flow from polar bears into ABC Islands brown bears but no evidence of gene flow from brown bears into polar bears. Importantly, while polar bears contributed <1% of the autosomal genome of the ABC Islands brown bear, they contributed 6.5% of the X chromosome. The magnitude of sex-biased polar bear ancestry and the clear direction of gene flow suggest a model wherein the enigmatic ABC Island brown bears are the descendants of a polar bear population that was gradually converted into brown bears via male-dominated brown bear admixture. We present a model that reconciles heretofore-conflicting genetic observations. We posit that the enigmatic ABC Islands brown bears derive from a population of polar bears likely stranded by the receding ice at the end of the last glacial period. Since then, male brown bear migration onto the island has gradually converted these bears into an admixed population whose phenotype and genotype are principally brown bear, except at mtDNA and X-linked loci. This process of genome erosion and conversion may be a common outcome when climate change or other forces cause a population to become isolated and then overrun by species with which it can hybridize.

Author Summary

The evolutionary genetic relationship between polar bears (*Ursus maritimus*) and brown bears (*U. arctos*) is a subject of continuing controversy. To address this we generated genome-wide sequence data for seven polar bears, two brown bears (including one from the enigmatic ABC Islands population), and a black bear (*U. americanus*). These data reveal remarkable genetic homogeneity within polar bears and clear evidence of past hybridization with brown bears. Hybridization, however, appears to be limited to habitat islands, where isolated populations of polar bears are gradually converted into brown bears via male-mediated dispersal and sex-biased gene flow. Our simplified and comprehensive model for the origin and evolution of polar bears resolves conflicting interpretations of mitochondrial and nuclear genetic data, and highlights the potential effect of natural climate change on long-term evolutionary processes.

Introduction

Despite polar bears' clear morphological and behavioral adaptations to their arctic environment(Slater *et al.* 2010; Stirling 2011) their genetic relationship to brown bears remains unclear (Kurten 1964; Edwards *et al.* 2011; Hailer *et al.* 2012; Miller *et al.* 2012). Analysis of maternally inherited mitochondrial DNA (mtDNA) shows that polar bears fall within the range of variation of brown bears. Extant brown bears from Alaska's ABC (Admiralty, Baranof and Chichagof) Islands, some extinct brown bears from Ireland and mainland Alaska, and two ~115,000-year-old polar bears share the mtDNA haplotype of all extant polar bears (Cronin *et al.* 1991; Talbot & Shields 1996; Barnes *et al.* 2002; Lindqvist *et al.* 2010; Davison *et al.* 2011; Edwards *et al.* 2011). The time to most recent common ancestor (TMRCA) of this mtDNA haplotype has been estimated at ~160 thousand years ago (kya) (Figure 1.11) (Lindqvist *et al.* 2010; Davison *et al.* 2011; Edwards *et al.* 2011; Miller *et al.* 2012). Recent analysis of data from a panel

of brown and polar bears at 14 nuclear loci showed that polar bears are generally distinct from brown bears, with genomic TMRCA averaging ~600 kya (Hailer *et al.* 2012). Under a simple population split model without subsequent admixture, the population divergence should be more recent than average genomic divergence and thus polar bears became a distinct species more recently than 600 kya. A separate recent genome sequencing survey concluded that brown bear and polar bear lineages are much older. Miller and colleagues concluded that the lineage that would become polar bears diverged from that which would become brown bears more than 4 million years ago, followed by admixture that continues to the present (Miller *et al.* 2012). Consistent with this, the past and present geographic ranges of both species overlap at their margins (Figure 1.1) and fertile hybrids are known in both captive and wild populations (Preuß *et al.* 2009; Stirling 2011).

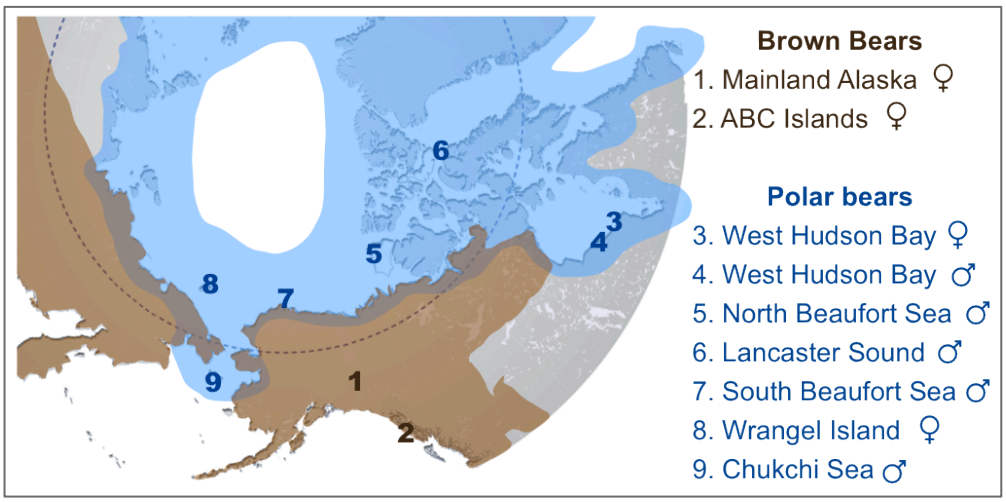


Figure 1.1. Map showing the approximate current geographic ranges of brown bears (brown) and polar bears (blue). Numbers indicate the geographic location of origin of two brown bears and seven polar bears polar analyzed here. An American black bear from central Pennsylvania was also sequenced as part of this study. Shotgun data amounting to 4-6X coverage for polar bears and 11-12X coverage for brown and black bears (Supplement) was aligned to the current distribution of the polar bear genome (Li *et al.* 2011; Liu *et al.* 2014).

The current consensus is that mtDNA and perhaps other polar bear loci are the result of past introgressions from brown bears into polar bears (Edwards

et al. 2011; Hailer *et al.* 2012; Miller *et al.* 2012). One scenario that has been proposed to reconcile the complicated discordance between the mtDNA trees and the species trees requires at least two instances of hybridization (Edwards *et al.* 2011). The first, which must have occurred before ~115kya, passed the mtDNA haplotype from polar bears into brown bears, including the ancient Irish brown bears and ancestors of the ABC Island brown bears. The second passed this mtDNA haplotype back into polar bears, after which it came to fixation in all extant polar bears. This convoluted scenario is necessary if, in fact, polar bears derive their mtDNA haplotype and other loci from brown bears. Unfortunately, this prevailing consensus has gone unquestioned. Here, we present an analysis of published and newly generated genome-wide data for brown bears and polar bears. We find extensive evidence of previous admixture, from polar bears into brown bears, especially of X-linked genes.

Results

To more fully delineate the genetic relationship between polar bears and brown bears, we sequenced random genomic shotgun libraries from seven polar bears, two brown bears and one black bear to learn the ancestral state for alleles (Figure 1.1, Supplemental Text). We mapped these reads to the assembled genome scaffolds of polar bear (Supplemental Text)(Li *et al.* 2011). Because the sequence coverage of each bear was uneven and too low to reliably call heterozygous sites, we down-sampled the sequence data from each bear to 1×. That is, we randomly picked a high-quality base from amongst all reads that mapped reliably at each position in the bear genome. In this way, we generated a composite haplotype for each bear and used these data for further analysis.

To gauge the level of diversity within and divergence between bear species, we made pairwise comparison between each bear, in 50 kb windows, across the bear genome (Figure 1.2). In agreement with previous reports (Paetkau

et al. 1999; Hailer *et al.* 2012), we find that polar bears are remarkably homogeneous: polar bear alleles differ at ~4 sites in 10,000. In contrast, brown bears have roughly four times as much genetic diversity, differing at ~17 in 10,000 sites. We note that the level of diversity among brown bears is nearly as high as the divergence between brown and polar bears. As expected, polar bears and brown bears show similar pairwise genomic divergence from the black bear. Likewise, the polar bears, brown bears, and black bears all show similar genomic divergence from the giant panda (*Ailuropoda melanoleuca*) (Li *et al.* 2010).

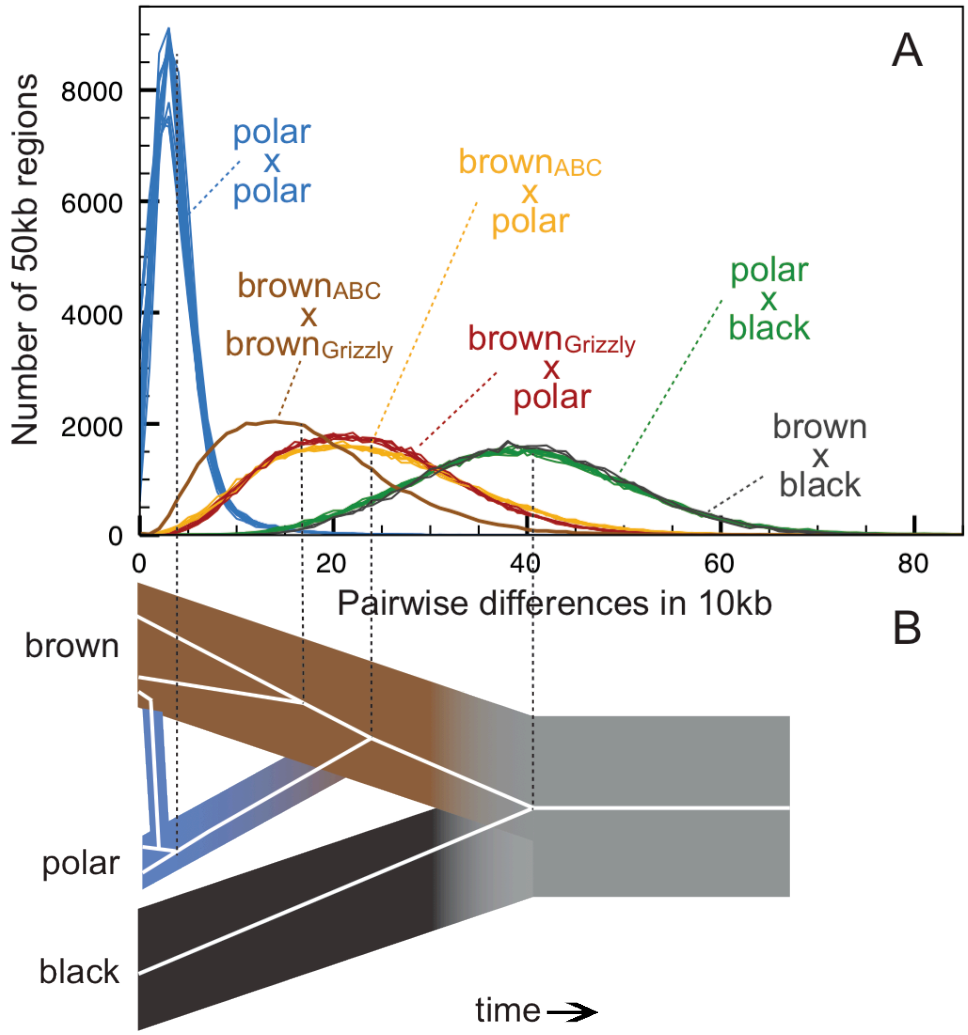


Figure 1.2. Genetic diversity within and between bear species. (A) Pairwise differences between individuals estimated as the average number of differences per 10 thousand bases (kb) in 42,000 non-overlapping 50 kb regions. After strict quality filtering, within-sample heterozygosity was resolved by selecting a single, high-quality base at random. The Lancaster Sound polar bear showed an excess

of postmortem damage, as expected for historic specimens (Green *et al.* 2008), and is shown in Figure 1.5. Polar bears are remarkably homogenous compared to brown bears, and both polar bears and brown bears are approximately equally diverged from the American black bear. Consistent with the results of the *D*-statistic test, pairwise distance between the ABC Islands brown bear and all polar bears (yellow lines) is less than that between the mainland brown bear and all polar bears (red lines). (B) Schematic diagram of a representative gene tree within brown bear, polar bear, and black bear populations, with the present day at the left of the diagram. For this locus, admixture occurring more recently than the population divergence of polar bears leads to the introgression of a polar bear haplotype into brown bears. Estimate of average genomic distance for brown, black, and polar bears and for population divergence between brown bears and polar bears given different calibration points are provided in Table 1.1.

We quantified admixture between brown and polar bears using the *D*-statistic (Green *et al.* 2010). In brief, *D* is the excess fraction of derived alleles shared between one of two conspecific individuals with a candidate admixing individual (Figure 1.3). Note that both incomplete lineage sorting (ILS) and admixture can lead to sharing of derived alleles, in this case between polar bears and brown bears. ILS, being a stochastic process, will result in equivalent numbers of shared, derived alleles between any two brown bears and a polar bear. Admixture, on the other hand, will result in more shared, derived alleles in the more admixed bear. Thus, under the null model of no admixture, $D=0$. A significant non-zero value of *D* indicates more admixture with one of the two individuals.

Comparison of any two polar bears for admixture with brown bears found little evidence for admixture. All *D*-statistics comparing two polar bears to a brown bear were statistically indistinguishable from 0 (Figure 1.3 top and middle panels).

Conversely, *D*-statistic comparisons between the ABC Islands and mainland brown bears for polar bear admixture were consistently and equivalently non-zero (Figure 1.3, bottom), regardless of the polar bear used in the comparison ($D=0.016$, which translates to roughly 0.75% of the genome; *Z*-score =1.24). Remarkably, when the analysis is restricted to the 12 scaffolds (~74

Mb of sequence) identified as X-chromosome (Supplemental Text), D=0.22, or ~6.5% of the X-chromosome (Z-score=4.52) (Figure 1.6; Tables 1.4, 1.5, 1.6). We find this same enrichment of the X chromosome, compared to the autosome, for admixture with polar bears when analyzing genome sequence data from two additional, recently published ABC Islands brown bears (Figure 1.7, Tables 1.7)(Miller *et al.* 2012). The ABC Islands bears therefore share not only their mtDNA but also a significant portion of their X-chromosomes with polar bears. A parsimonious explanation for these observations is that the same admixture event that resulted in sharing of the polar bear mtDNA haplotype with ABC Island brown bears also results in sharing of much of the X-chromosome.

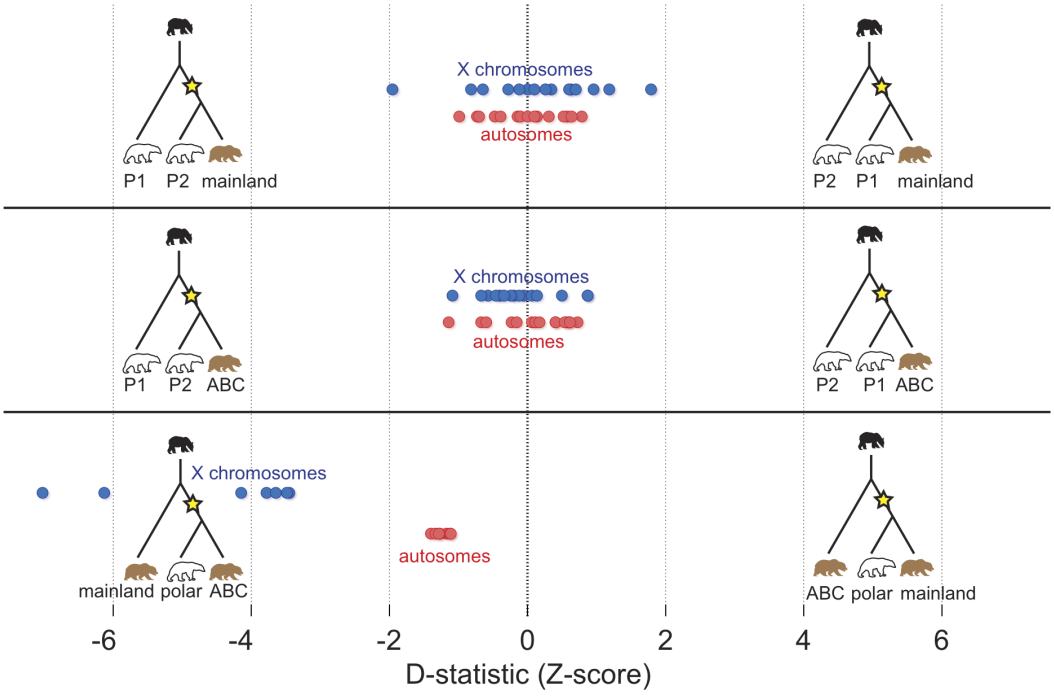


Figure 1.3. Summary of D-statistic comparisons between polar bears and brown bears. In each comparison, the black bear was used to define the ancestral allele. The Z-score of the D-statistic for each comparison is shown for autosomes (red) and X-chromosome (blue). Each dot represents the data from comparison of one pair of bears. In the top panel, all pairs of polar bears are compared for excess derived allele matching against the mainland brown bear. In the middle panel, all pairs of polar bears are compared against the ABC Island brown bear. The bottom panel shows the comparison of the two brown bears for excess allele matching to polar bears with each dot representing a different polar bear.

To test the direction of X-chromosome gene flow between polar bears and the ABC Islands bear we simulated the effect of having 6.5% ancestry (roughly the amount estimated above) in either polar bear or mainland brown bear X chromosome from the reciprocal species (Figure 1.4, Figure 1.9). The simulation was carried out by randomly selecting 6.5% of the X-chromosome of the candidate recipient species to be replaced by sequence from the candidate donor species (Figure 1.8, Supplemental Text). We then measured the distribution of pairwise divergences that would result following this simulated admixture.

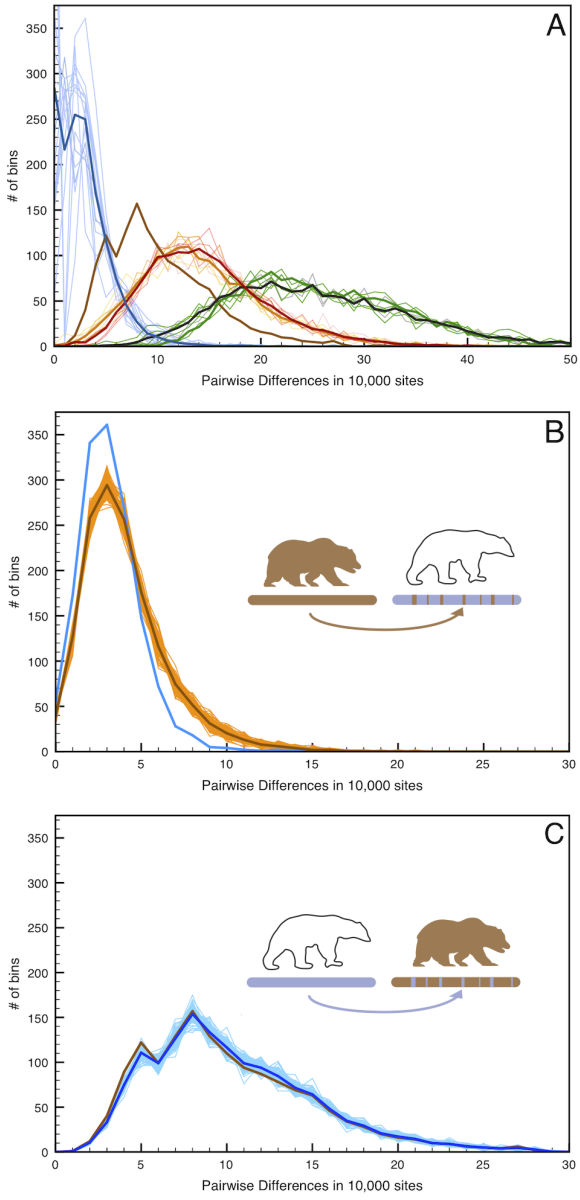


Figure 1.4. Simulated admixture reveals the direction of gene flow on the X chromosome. (A) Pairwise distance as in Figure 1.2 but limited to the 12 scaffolds identified as X-chromosome. (B) 100 replicate simulations in which 6.5% of the female West Hudson Bay polar bear X-chromosome is replaced with that of the mainland Alaska brown bear in randomly inserted 20 kb fragments, simulating admixture from the brown bear genome into polar bear ~50kya. Pairwise differences are calculated between the simulated genome (light brown lines; mean highlighted in dark brown) and the plot comparing the two female polar bears (blue line), to maximize the number of informative sites in the test. The addition of brown bear DNA to the polar bear genome markedly increases the number of high-diversity bins (>10 differences/10 kb), indicating that any introgression of brown bear DNA into polar bears should be easily detectable. (C). As in (B), but with 6.5% of the mainland Alaska brown bear X-chromosome is replaced with that of the female West Hudson Bay polar bear. In this instance, we find no difference between the simulated (blue lines) and real (brown line) data.

Given the low genetic diversity within polar bears, this amount of brown bear ancestry would be clearly identifiable as an excess of deeply diverging regions between polar bears, even in unphased data from which a random allele is chosen at each site. Conversely, simulating 6.5% polar bear ancestry in the mainland brown bear X-chromosome is more consistent with the observed level of genomic regional divergence between brown bear X-chromosomes. Thus, we deduce that the direction of gene flow was from polar bear into the ABC Islands brown bear X-chromosome.

Recently published genome sequence data from a ~115ky polar bear (Miller *et al.* 2012) allow us to further probe when and in which direction admixture might have happened. Using this ancient polar bear in the *D*-statistic test gives nearly identical results to the extant polar bears (autosome $D=0.015$; X-chromosome $D=0.212$). That is, ABC island brown bears are equally enriched for polar bear matching derived alleles, even when this ~115ky polar bear is used in the comparison. Therefore, if the admixture was from the ABC Island brown bears (or a closely related population) into polar bears, it must have occurred prior to ~115ky. Furthermore, no significant subsequent admixture could have occurred, since the modern polar bears are nearly homogeneous for the ABC

Islands brown bear D-statistic signal. Finally, if gene flow was from brown bears into polar bears, it would had to have been from a population of brown bears that lived more than ~115kya that today finds itself restricted to a group of islands that only became habitable for brown bears since the end of the last glacial maximum, about ~16kya. Given the unlikeliness of this scenario and the incompatibility of polar bear X-chromosomes genetic divergence with brown bear ancestry, we conclude that the direction of gene flow was from polar bears into the ABC Island brown bears.

Discussion

The genome-wide analysis presented here indicates that (1) polar bears are a remarkably homogeneous species and show no evidence of brown bear ancestry, (2) the ABC Islands brown bears show clear evidence of polar bear ancestry, and (3) this polar bear ancestry of ABC Islands brown bears is conspicuously enriched in the X-chromosome. ABC Islands brown bears show a simple positive correlation between how maternally biased a genetic locus is (mtDNA>X chromosome>autosomes) and how much polar bear ancestry is present (100%, 6.5%, 1%). Given this observation, and our knowledge about the natural history of these islands through the Pleistocene and Holocene, we present the following model.

During the peak of the last ice age, brown bears were likely absent from the region that now comprises the ABC Islands. Although fossil remains dating to this period are abundant on the more southerly islands of the Alexander Archipelago, brown bears are not among the species present during the period spanning 26-12kya, when glacial conditions were at their peak (Heaton *et al.* 1996; Heaton & Grady 2003; Carrara *et al.* 2007; Chiverrell & Thomas 2010). Geological and climatological data suggest that if any habitat suitable for brown bears persisted on the ABC Islands during the LGM it would have been limited to the

western part of Baranof Island, the most distant of the ABC Islands from the Alaskan mainland (Carrara *et al.* 2007). By itself, however, this potential refugium would have been too small to support viable populations of brown bears (Alaska Department of Fish and Game).

Polar bears, alternately, would likely have colonized the sea ice adjacent to the ABC Islands as the ice advanced southward. Notably, marine mammals dominate the fossil remains dating to this interval (Heaton & Grady 2003), including ringed seals, an ideal food source for polar bears (Stirling 2011). As the climate warmed and ice retreated, polar bears may have been stranded on or near the ABC Islands. As the habitat became increasingly hospitable to brown bears (Carrara *et al.* 2007), the early colonizers from the mainland would have been predominantly the more peripatetic sub-adult males (Paetkau *et al.* 1998). Admixture involving an influx of mostly or exclusively male brown bears with the stranded polar bears would have resulted in a gradual erosion of the polar bear genome within the isolated population. The sex bias of admixing brown bears would have made genomic erosion more rapid in the autosomes, confining the vestiges of polar bear ancestry in extant ABC Islands bears primarily to matrilineal-biased genetic loci (Figure 1.15).

Our simplified model - little or no brown bear ancestry in polar bears and matrilineal-biased polar bear ancestry in the ABC Islands brown bears - is consistent with several important comparative genomic observations. First, mtDNA and nuclear genome diversity within both extant and a ~115kya polar bear is extremely low. This low level of polar bear diversity is consistent with no admixture from brown bears. Brown bears, in contrast, show much higher levels of diversity including many deep genetic lineages that have not completely sorted since their population divergence from polar bears. The ABC Islands brown bears show genome-wide evidence of admixture with polar bears

concentrated on the X chromosome. Importantly, the level of admixture inferred from *D*-statistic analyses is only compatible with polar bear admixture into the ABC Islands brown bear X chromosomes and not the other way around. Conveniently, this model explains the presence of the polar bear mtDNA haplotype in all ABC Islands brown bears: the mtDNA haplotype of the male brown bear immigrants is lost, regardless of how many male brown bear immigrants arrived.

The model for historic admixture proposed here is distinct from the traditional framework for admixture, including the scenario involving early humans and Neandertals for which the *D*-statistic analysis was originally developed (Green *et al.* 2010; Reich *et al.* 2011). Usually, the goal is to find the signal of a potentially small amount of admixture from a single or few admixture episodes that took place many generations ago (Figure 1.12A). While such a model is consistent with the ABC Islands brown bear autosomal *D*-statistic results, it is insufficient to explain the large difference in the X-chromosome or the fixation of the polar bear mtDNA haplotype in the ABC Islands brown bears (Supplemental Text). In fact, reasonable parameter values for a model that assumes a single episode of admixture from polar bears into brown bears do not result in a ratio of *D* for the X and autosomes that exceeds 2.7; our observed ratio is ~14. Alternately, a long process of sex-biased immigration of brown bears into what was initially a polar bear population can result in much higher ratios of polar bear ancestry for the X and autosomes (Table 1.9; Figures 1.12B, 1.13, 1.14), consistent with the empirical observations presented here.

Spatially explicit modeling has been used to probe the dynamics of gene flow from introgression during species expansions (Currat *et al.* 2008). These simulations have yielded insight into the often non-intuitive patterns seen in various loci such as the apparent asymmetry in gene flow from the native species

into the invading species. An extension of this approach to incorporate a migration barrier to female, but not male, gene flow and a dwindling native population of polar bears, may more fully reveal the demographic details of the brown bear invasion. Of particular note, there is evidence that brown bear migration between the mainland and ABC Islands may be ongoing. Analysis of variation at 17 rapidly evolving microsatellite loci indicated that brown bears from Admiralty Island, the closest of the ABC Islands to the mainland, are more similar to mainland Alaskan brown bears than were bears from Baranof and Chichagof Islands (Paetkau *et al.* 1998). Assuming no disruption of the salient features of this migration, its final state, which has not yet been realized, would be complete conversion of the population, i.e., the fixation of brown bear alleles in all genomic loci in the ABC Island bears except the strictly maternal mtDNA.

We note that our data cannot resolve the *timing* of the origin of polar bears as a distinct lineage. Such an estimate has been hindered mainly by the paucity of preserved ancient polar bear remains (Kurten 1964; Ingólfsson & Wiig 2009), and consequent lack of fossil calibrations. However, our data do provide insight into the *relative timing* of divergence between the three bear lineages sampled here. To generate a hypothetical scenario for the timing of the origin of polar bears, we apply several previously suggested calibration strategies to our data (Table 1.1; Figure 1.2B). Regardless of the calibration strategy applied, our data support a long interval between the initial divergence between black bears and the brown bear/polar bear lineage, and the later divergence between brown bears and polar bears. This is similar to that observed by Hailer *et al* (Hailer *et al.* 2012), and in contrast to the scenario predicted by the model of Miller *et al* (Miller *et al.* 2012).

	Scaled	Method A	Method B	Method C
giant panda/ black bear	5.99	*8-16 Mya*	23.3-38.8 Mya	12.0 Mya
black bear/ brown bear	*1*	1.34-2.67 Mya	*3.90-6.48 Mya*	2.00 Mya
brown bear/ polar bear	0.6	0.80-1.60 Mya	2.43-3.89 Mya	1.20 Mya
brown bears (population)	0.43	0.57-1.15 Mya	1.68-2.79 Mya	0.86 Mya
polar bears (population)	0.1	0.13-0.27 Mya	0.39-0.65 Mya	0.19 Mya

Table 1.1. Estimates of average genomic TMRCA for black, brown and polar bear lineages, and average population TMRCA for brown bears and polar bears estimated from our data, using three calibration methods (calibrated notes are listed in italics). Estimates are scaled based on an average pairwise distance between sampled brown bears and polar bears of 1 (Figure 1.2A). Method A assumes divergence between the giant panda and polar bear lineage 12±4 My (Hailer *et al.* 2012). Method B assumes an average TMRCA between brown bears and black bears 3.9-6.48 Mya (Krause *et al.* 2008). Method C assumes a mammalian mutation rate of $1x10^{9}$ substitutions/site/year, the basis for the very old estimates presented in (Miller *et al.* 2012).

From analysis of the data presented here, we infer that polar bears most likely became a distinct lineage sometime during the Pleistocene. This timing is consistent with previous molecular (Table 1.1) and morphological (Kurten 1964) estimates. Polar bears and brown bears were clearly established as a morphologically distinct species by at least ~115kya – the age of the oldest known polar bear fossil (Ingólfsson & Wiig 2009; Lindqvist *et al.* 2010). Regardless of this timing, our data suggest that polar bears have remained a small, distinct lineage since their origin (Figure 1.10), with lineage-specific adaptations reinforced by the ecological constraints of their extreme environment (Supplemental Text, Table 1.8)(Miller *et al.* 2012). Brown bears, in contrast, have had a larger effective population size (Figure 1.10), with segregating polymorphism that often predates their split with polar bears (Figure 1.2B).

The process of genomic erosion we propose here may not be unique to the stranded ABC Islands polar bears. Past changes in the distribution of polar ice, for example, may have also stranded polar bears or hybrids on present-day Ireland, explaining the appearance of polar bear mtDNA in the remains of extinct Irish brown bears (Edwards *et al.* 2011). Long-term climate change may often strand populations on islands or island-like habitats, such as lakes or mountain plateaus. If these stranded populations then hybridize with closely related immigrants, we predict substantial variability in the apparent level of admixture indicated by *D*-statistics. Furthermore, in the case of sex-biased immigration, the ratio of *D*-statistics for the X and autosomes will be highly dependent on the rate and duration of immigration.

Materials and Methods

We extracted DNA from nine of the ten bears in a modern DNA laboratory using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's specifications. The historic Lancaster Sound polar bear (Smithsonian Natural History Museum ID 512133; Table 1.2) was extracted in a dedicated ancient DNA laboratory at Penn State University that is geographically isolated from modern molecular biology research, using a column-based extraction protocol for ancient DNA (Rohland *et al.* 2010).

We physically sheared the DNA of the modern bears using a Diagenode Bioruptor UCD-200 instrument. Fifty μl of each of the six modern polar bear extracts were transferred into 1.5 ml tubes and exposed to four rounds of sonication for 7 min, using the energy setting "HIGH" and an "ON/OFF interval" of 30 seconds. To attain a longer insert size, we slightly modified the procedure to include two 7-min rounds and one 5-min round of sonication for the brown bears, black bear, and second round of sequencing for two polar bears (West Hudson Bay X3249106A; and Chukchi Sea UP08.010; Table 1.2). We then

purified and concentrated the extracts using the Agencourt AMPure XP PCR purification kit, according to manufacturer's instructions, and eluted in 20 μl of 1×TE, with 0.05% Tween20. The historic bear sample was already fragmented due to degradation, and was not sonicated.

We prepared indexed Illumina libraries using 15 μl of each extract following the protocol described in (Meyer & Kircher 2010), with reaction volumes scaled to total volume of 40 μl. To verify final DNA concentration and the distribution of insert sizes, we ran each library on an Agilent 2100 Bioanalyzer. We then sequenced each polar bear on a separate lane of an Illumina HiSeq 2000 instrument using 100 base-pair (bp) paired-end chemistry at the UC Santa Cruz Core Genomics Facility. We sequenced one lane each of the two brown bears, the black bear, and an additional lane for two polar bears (Table 1.2) using an Illumina HiSeq 2000 instrument with 150-bp paired-end chemistry at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

From the Illumina sequence data, we removed the index and adapter sequence and merged paired reads using a script provided by M. Kircher (Kircher 2012). We then trimmed each read to remove low quality bases by trimming inward from the 3′-end of the read until detecting a base with quality score ≥13 (~95% confidence). We mapped the resulting data to the draft polar bear genome (Li *et al.* 2011; Liu *et al.* 2014) using *BWA* (Li & Durbin 2010). We removed duplicated reads created by PCR amplification using *rmdup* program from samtools (Li *et al.* 2009). We then applied GATK's (McKenna *et al.* 2010) base quality score recalibration and indel realignment, and performed SNP genotyping across all samples simultaneously using default settings in GATK (DePristo *et al.* 2011). Total coverage is shown in Table 1.2.

Acknowledgments

We thank E. Richardson for helpful comments on the manuscript and for providing the Canadian polar bear samples, and T. Evans of the U.S. Fish and Wildlife Service and E. Regehr, now of the Alaska Science Center, for providing Alaskan polar bear tissue samples. We thank K. Helgen and S. Peurach of the Smithsonian Natural History Museum for providing access to the historic polar bear from Lancaster Sound, and L. Olsen and the University of Alaska at Fairbanks Museum of Natural History for providing access to brown bear specimens. We thank B. Li, G. Zhang, E. Willersleve, J. Wang, and J. Wang for generation and public release of the reference polar bear genome.

Author Contributions

Conceived and designed the experiments: JA Cahill, RE Green, B Shapiro. Performed the experiments: TL Fulton, M Stiller, N Ovsyanikov. Analyzed the data: JA Cahill, RE Green, F Jay, R Salamzade, J St. John, I Stirling, M Slatkin, B Shapiro. Contributed reagents/materials/analysis tools: N Ovsyanikov, M Slatkin, B Shapiro. Wrote the paper: JA Cahill, RE Green, B, Shapiro. Assisted with data interpretation and preparation of the manuscript: JA Cahill, RE Green, TL Fulton, M Stiller, F Jay, N Ovsyanikov, R Salamzade, J St. John, I Stirling, M Slatkin, B Shapiro.

Supplementary Materials:

1. Data collection

1.1 Samples Collected

To assess the global genomic diversity of extant polar bears, we collected tissue specimens from seven polar bears (*Ursus maritimus*) from across their present-day range and two Alaskan brown bears (*U. arctos*) from the University of Alaska Museum of Natural History, one from the enigmatic ABC Islands population, and the other from the Alaskan mainland (Figure 1.1, Table 1.2) for random shotgun sequencing. To learn the ancestral state of polar bear and brown bear alleles, we performed identical random shotgun sequencing on a single American black bear (*U. americanus*), provided by Anthony Ross, North-Central Regional Wildlife Supervisor of the Pennsylvania Game Commission.

Species	Collection	Accession number	Geographic origin	Sex
U. maritimus	Canadian Wildlife Service; Edmonton, Alberta, Canada	X3249106A	West Hudson Bay	female
U. maritimus	Canadian Wildlife Service; Edmonton, Alberta, Canada	X3312806A	West Hudson Bay	male
U. maritimus	Canadian Wildlife Service; Edmonton, Alberta, Canada	X3292306A	North Beaufort Sea	male
U. maritimus	United States Fish and Wildlife Service; Anchorage, Alaska, USA	990083KD	North Beaufort Sea	male
U. maritimus	United States Fish and Wildlife Service; Anchorage, Alaska, USA	940090KB	Chukchi Sea	male
U. maritimus	Smithsonian Natural History Museum; Washington DC, USA	512133	Lancaster Sound	male
U. maritimus	Wrangel Island State Nature Reserve	UP08.010	Wrangel Island	female

U. arctos	University of Alaska Museum of Natural History; Fairbanks, Alaska, USA	UAM63857	Admiralty Island, Alaska	female
U. arctos	University of Alaska Museum of Natural History; Fairbanks, Alaska, USA	UAM33812	Wood River, Denali National Park, Alaska	female
U. americanus	Pennsylvania Game Commission, Jersey Shore, Pennsylvania, USA	JC012	Central Pennsylvania	female

Table 1.2: Sample Details.

1.2 Data resulting from Illumina Sequencing after quality control and filtering

Species	Sampling Location (Abbreviation)	Number of reads	Coverage	Gender
Polar bear	Chukchi Sea (CS)	1.24E+08	4.3X (15.8X)	Male
Polar bear	Wrangel Island (WI)	1.90E+08	4.7X	Female
Polar bear	West Hudson Bay (WHB_f)	1.69E+08	4.9X (17.9X)	Female
Polar bear	West Hudson Bay (WHB_m)	1.50E+08	4.5X	Male
Polar bear	North Beaufort Sea (NBS)	1.86E+08	5.7X	Male
Polar bear	South Beaufort Sea (SBS)	1.44E+08	4.7X	Male
Polar bear	Lancaster Sound (LS)	1.78E+08	4.8X	Male
Brown bear	ABC Islands (Admiralty Island) (ABC)	3.54E+08	12.1X	Female
Brown bear	Denali NP, Alaska (Grizzly)	3.61E+08	12.1X	Female
American black bear	Pennsylvania (Black)	3.28E+08	11.6X	Male

Table 1.3: Data collected for this analysis. Whole genome shotgun sequence data were collected from ten bears from the locations listed. Number of reads corresponds to the number of reads that mapped to the draft polar bear genome (Li *et al.* 2011; Liu *et al.* 2014) using BWA (Li & Durbin 2010). Coverage is estimated by averaging the number of reads that map to each site of the draft polar bear genome, after extensive filtering as described in in section 1.2. For two polar bears, we sequenced an additional Illumina lane to increase coverage. The augmented data set (coverage in parentheses) was used for the analysis described in section 2.5.

1.3 Identifying the X chromosome

Most analyses described below are of data mapped to the draft polar bear genome assembly (Li *et al.* 2011; Liu *et al.* 2014). This assembly contains 72214 scaffolds with N50 of 15.9Mb. Of these, we restrict our analysis to the 238 scaffolds >1 Mb in length, for a combined total genome size ~2.2Gb. Importantly, these scaffolds are not anchored to chromosomes. In order to contrast patterns of divergence and admixture from autosomes to X chromosomes, we took the following approach to identify scaffolds from the polar bear genome that are likely X chromosome.

First, since the X chromosome is homologous across mammals, we used the closest, well annotated genome to bears, i.e., the domestic dog (Lindblad-Toh *et al.* 2005) to find polar bear genome scaffolds that contain X-linked genes. We mapped 549 X-linked dog genes to the polar bear scaffolds using *genBLASTa* (She *et al.* 2008). From this mapping, we counted the number of dog X chromosome genes that mapped to each polar bear scaffold >1Mb in length. Second, we compared the coverage by shotgun sequencing, assuming that scaffolds of X chromosome should have about half the coverage of autosomal scaffolds in males. We identified scaffolds as putatively X if they showed typical genome-wide coverage in the female brown bears but approximately half of the genome-wide coverage in the male black bear. (We chose the male black bear for comparison as it had higher genome coverage than any of the male polar bears). Finally, we combined those tests, and selected scaffolds as deriving from the X chromosome if they met the coverage criteria *and* contained at least 10 dog X chromosome genes. This resulted in 12 scaffolds designated as X-chromosome (20, 100, 105, 113, 115, 122, 134, 141, 167, 170, 179, 184) with a combined length of 73.7 Mb. This is likely nearly half of the polar bear X chromosome, as the dog X chromosome assembly spans 123.9 Mb.

1.4 Mapping the panda genome to the polar bear genome

As an alternative for identifying the ancestral allele for polymorphic sites in brown and polar bears, we also used the reference giant panda (*Ailuropoda melanoleuca*) genome, ailMel1 (Li *et al.* 2010). We split this genome sequence into non-overlapping segments of 256 base-pairs and used *BWA* to map these data to the polar bear genome. We increased the –n option of *BWA* to 24 for increased mapping sensitivity. For each position of the polar bear genome covered by a panda read, 96.5% were covered by exactly one panda read. We selected that panda aligned base if the read's map-quality was ≥30. For sites covered by two reads (3.4%), we selected one of the two bases randomly; note that in nearly all cases the two bases are identical. Sites covered by three or more reads were excluded from analysis.

2. Data Analysis

2.1 Estimating pairwise distances between individuals

We calculated the number of pairwise differences between all individuals by selecting, for each individual, a single base call mapped to each position in the 238 scaffolds of >1Mb length from the polar bear draft genome (Li *et al.* 2011; Liu *et al.* 2014). Base calls were limited to those with phred >61. This effectively limits our analysis to only positions where the forward and reverse reads overlapped and agreed, i.e., where confidence in the base quality is maximal. We subdivided the scaffolds into non-overlapping 50kb windows, and calculated the number of pairwise differences between individuals, normalized by the total number of sites where both individuals met base quality criteria. We generated histograms of divergence in these 50kb windows by calculating the number of differences in 10,000 bases. In nearly all 50kb bins, more than 10,000 sites were observed, so binning artifacts were minimal. However, some comparisons involving male individuals on the X chromosome scaffolds did show binning

artifacts because the lower coverage decreased the total number of sites observed in pairs of individuals. In general, X chromosome scaffolds showed far fewer pairwise differences per 10,000 sites than the autosomes, presumably due to the smaller effective population size of the X chromosome.

Figures 1.2A and 1.3A of the main text show histograms of pairwise distance estimates calculated for the autosomes and X-chromosome, *excluding* the historic polar bear from Lancaster Sound. This sample was collected from Cornwallis Island in October of 1973, and is currently part of the collection at the National Museum of Natural History (Smithsonian Institution, Washington DC). As expected from historic and ancient samples, the genomic data from this specimen shows an excess (nearly twice as many) C to T and G to A transitions when compared to the polar bear reference genome than the other, modern polar bears. This pattern is likely due to cytosine deamination to uracil, which is the most common form of post-mortem DNA damage (Hofreiter 2001). While including the Lancaster Sound polar bear does not significantly influence the results presented here or in the main text, the effect of the excess of damaged sites is clearly visible in these pairwise divergence plots (Figure 1.5).

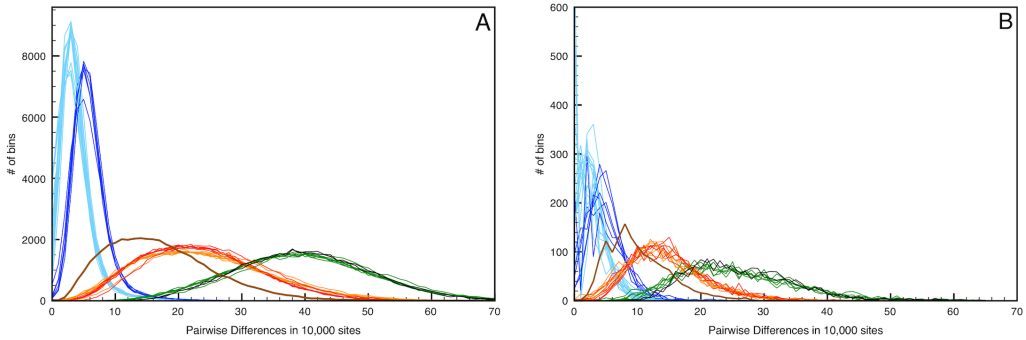


Figure 1.5. Pairwise distances between all pairs of bears including the historic bear from Lancaster Sound. Plots show histograms for (A) all autosomal data and (B) X chromosome only. The color scheme matches Figure 1.2A and Figure 1.3A from the main text. The Lancaster Sound polar bear data are highlighted in dark blue.

2.2 Calculating the *D*-statistic

We estimated admixture using the *D*-statistic (Green *et al.* 2010; Durand *et al.* 2011). The implementation here closely follows that described in (Green *et al.* 2010). Informally, the test is a comparison of sharing of derived alleles between two individuals of the same species, I1 and I2, with a candidate admixing individual, M, of a different species. Derived alleles are defined by using an out group individual, O. In this case, the American black bear or giant panda, where indicated, is used as outgroup. At each position in the genome, a random single allele is chosen from amongst reads that pass various filtering criteria. We filtered: (i) only analyzing map data from scaffolds ≥1Mb, (ii) from genomic sites where overall read coverage was between the 5th and 95th percentile, genome-wide, (iii) base-quality ≥61, (iv) read-map quality ≥10, (v) uniquely mappable 35-mers (Derrien *et al.* 2012).

For genomic sites with a suitable available allele from each individual in the comparison, we write the alleles in the following order: I1, I2, M, and O. Designating the allele from O as A (for ancestral) and an alternate allele as B, we restrict our focus to sites of either ABBA or BABA configuration. That is, we consider *only* sites where M is different from O, i.e., the candidate introgressor has a derived allele and this allele is seen in *either* I1 or I2. Counting the number of such sites, we can calculate:

$$D = (\text{ABBA-BABA})/(\text{ABBA+BABA})$$

ABBA and BABA sites may be generated by one of four phenomena: admixture, incomplete lineage sorting, multiple mutations occurring at the same site, and machine error leading to incorrect identification of alleles. In the absence of significant ancient population structure, one can expect incomplete lineage sorting, multiple mutations at the same site and machine error to be evenly distributed between ABBA and BABA sites. Any imbalance between the number of ABBA and BABA sites is thus attributed to admixture. Note that even when

admixture can be detected, the *D*-statistic does not indicate the direction of gene flow.

To establish statistical confidence for the results, a weighted block jackknifing approach was employed with blocks of 5Mb (Green *et al.* 2010). Because the polar bear genome consists of scaffolds not mapped to chromosomes, we required complete 5Mb blocks within scaffolds. This requirement will over-estimate the variance and thus make our significance test conservative. A *Z*-score is then calculated by taking the absolute value of *D*/standard error, as per (Green *et al.* 2010). Tables 1.4 and 1.5 show the *Z* score for admixture analysis for all configurations of bears using either the black bear (Table 1.4) or giant panda (Table 1.5) as outgroup.

I1 bear	**I2 bear**	**M bear**	**O bear**	***D* (auto.)**	***Z* (auto.)**	***D* (X)**	***Z* (X)**
ABC	Grizzly	WHB_f	Black	-0.014	-1.176	**-0.228**	**-4.149**
ABC	Grizzly	WHB_m	Black	-0.016	-1.272	**-0.235**	**-7.021**
ABC	Grizzly	SBS	Black	-0.014	-1.119	**-0.204**	**-3.449**
ABC	Grizzly	CS	Black	-0.017	-1.399	**-0.223**	**-3.777**
ABC	Grizzly	NBS	Black	-0.017	-1.330	**-0.207**	**-3.483**
ABC	Grizzly	WI	Black	-0.014	-1.107	**-0.227**	**-6.135**
ABC	Grizzly	LS	Black	-0.016	-1.279	**-0.200**	**-3.646**
WHB_f	WHB_m	ABC	Black	-0.022	-1.143	-0.051	-0.188
WHB_f	SBS	ABC	Black	-0.014	-0.676	0.200	0.874
WHB_f	CS	ABC	Black	-0.005	-0.228	-0.175	-0.569
WHB_f	NBS	ABC	Black	-0.013	-0.600	0.060	0.868
WHB_f	WI	ABC	Black	0.001	0.058	-0.107	-0.400
WHB_m	SBS	ABC	Black	-0.003	-0.157	0.143	0.498
WHB_m	CS	ABC	Black	0.009	0.405	-0.030	-0.053
WHB_m	NBS	ABC	Black	0.011	0.596	-0.027	-0.117
WHB_m	WI	ABC	Black	0.015	0.726	0.018	0.061
SBS	CS	ABC	Black	0.013	0.560	-0.290	-1.086
SBS	NBS	ABC	Black	0.009	0.408	-0.176	-0.667
SBS	WI	ABC	Black	0.011	0.538	-0.158	-0.457
CS	NBS	ABC	Black	0.003	0.116	-0.118	-0.245
CS	WI	ABC	Black	0.013	0.613	0.022	0.134
NBS	WI	ABC	Black	0.004	0.173	-0.099	-0.339
WHB_f	WHB_m	Grizzly	Black	-0.021	-0.994	-0.233	-0.817
WHB_f	SBS	Grizzly	Black	-0.010	-0.475	0.184	1.788
WHB_f	CS	Grizzly	Black	-0.003	-0.145	0.000	0.000

WHB_f	NBS	Grizzly	Black	-0.015	-0.730	0.051	0.342
WHB_f	WI	Grizzly	Black	-0.002	-0.103	0.117	0.630
WHB_m	SBS	Grizzly	Black	-0.015	-0.700	0.388	0.955
WHB_m	CS	Grizzly	Black	-0.010	-0.387	0.241	0.601
WHB_m	NBS	Grizzly	Black	0.000	0.000	0.219	0.698
WHB_m	WI	Grizzly	Black	0.012	0.573	0.237	1.184
SBS	CS	Grizzly	Black	0.013	0.522	-0.556	-1.956
SBS	NBS	Grizzly	Black	0.003	0.135	-0.182	-0.643
SBS	WI	Grizzly	Black	0.017	0.789	0.051	0.100
CS	NBS	Grizzly	Black	0.002	0.098	0.143	0.255
CS	WI	Grizzly	Black	0.007	0.307	-0.020	-0.124
NBS	WI	Grizzly	Black	0.013	0.641	-0.067	-0.275
Mean Values							
ABC	Grizzly	Any Polar	Black	-0.015	-1.240	**-0.218**	**-4.523**

Table 1.4. *D*-statistic and *Z* scores using American black bear as outgroup. Significant deviations from zero are highlighted in bold. Abbreviations are as in Table 1.3. The Lancaster Sound polar bear is not included in tests as I1 or I2.

I1 bear	**I2 bear**	**M bear**	**O bear**	***D* (auto.)**	***Z* (auto.)**	***D* (X)**	***Z* (X)**
ABC	Grizzly	WHB_f	Panda	-0.015	-1.353	-0.212	-4.077
ABC	Grizzly	WHB_m	Panda	-0.019	-1.639	-0.196	-6.634
ABC	Grizzly	SBS	Panda	-0.016	-1.378	-0.219	-3.974
ABC	Grizzly	CS	Panda	-0.016	-1.358	-0.210	-3.582
ABC	Grizzly	NBS	Panda	-0.018	-1.539	-0.200	-3.426
ABC	Grizzly	WI	Panda	-0.016	-1.377	-0.214	-3.960
ABC	Grizzly	LS	Panda	-0.017	-1.448	-0.185	-5.101
WHB_f	WHB_m	ABC	Panda	-0.011	-0.735	-0.133	-0.633
WHB_f	SBS	ABC	Panda	-0.003	-0.200	-0.006	-0.041
WHB_f	CS	ABC	Panda	-0.004	-0.209	-0.156	-1.232
WHB_f	NBS	ABC	Panda	-0.003	-0.207	0.014	0.111
WHB_f	WI	ABC	Panda	-0.002	-0.117	0.020	0.225
WHB_m	SBS	ABC	Panda	-0.001	-0.041	-0.042	-0.443
WHB_m	CS	ABC	Panda	0.006	0.313	-0.018	-0.071
WHB_m	NBS	ABC	Panda	0.010	0.646	-0.138	-0.492
WHB_m	WI	ABC	Panda	-0.002	-0.144	0.053	0.295
SBS	CS	ABC	Panda	0.003	0.172	0.019	0.069
SBS	NBS	ABC	Panda	0.004	0.240	-0.056	-0.249
SBS	WI	ABC	Panda	-0.001	-0.031	-0.046	-0.455
CS	NBS	ABC	Panda	0.006	0.328	0.103	0.572
CS	WI	ABC	Panda	-0.001	-0.028	0.050	0.205
NBS	WI	ABC	Panda	-0.007	-0.432	-0.053	-0.488
WHB_f	WHB_m	Grizzly	Panda	-0.011	-0.736	-0.172	-0.942
WHB_f	SBS	Grizzly	Panda	0.000	0.017	-0.069	-0.564
WHB_f	CS	Grizzly	Panda	-0.004	-0.255	-0.161	-1.265

WHB_f	NBS	Grizzly	Panda	-0.008	-0.522	-0.019	-0.107
WHB_f	WI	Grizzly	Panda	-0.003	-0.183	0.053	0.454
WHB_m	SBS	Grizzly	Panda	-0.003	-0.186	0.035	0.247
WHB_m	CS	Grizzly	Panda	-0.004	-0.201	0.018	0.072
WHB_m	NBS	Grizzly	Panda	0.001	0.044	-0.009	-0.035
WHB_m	WI	Grizzly	Panda	0.000	-0.024	0.143	1.541
SBS	CS	Grizzly	Panda	-0.002	-0.123	-0.106	-0.373
SBS	NBS	Grizzly	Panda	-0.008	-0.488	-0.052	-0.214
SBS	WI	Grizzly	Panda	0.000	-0.003	0.006	0.024
CS	NBS	Grizzly	Panda	-0.002	-0.105	0.222	1.307
CS	WI	Grizzly	Panda	0.000	-0.021	0.000	0.000
NBS	WI	Grizzly	Panda	0.005	0.309	-0.024	-0.094
Mean Values							
ABC	Grizzly	Any Polar	Panda	-0.017	-1.442	-0.205	-4.394

Table 1.5. *D*-statistic and *Z* scores using giant panda as outgroup. Significant deviations from zero are highlighted in bold. Abbreviations are as in Table 1.3. The Lancaster Sound polar bear is not included in tests as I1 or I2.

We note that when I1 or I2 contains an excess of postmortem damaged sites, these can influence estimates of both *D* and *Z*. In these instances, damage affects a false positive match between the outgroup (O) and the damaged individual, and consequent identification of the non-damaged individual as potentially admixed. This effect increases as the evolutionary distance between I1/I2 and the outgroup increases. We therefore exclude the Lancaster Sound bear from analyses in which it would be either I1 or I2.

Table 1.6 shows the results of the *D*-statistic test for admixture between the ABC Island and mainland Alaska brown bears and the American black bear, as recently proposed by Miller and colleagues (Miller *et al.* 2012). We find no support for admixture between brown bears and the American black bear using our approach.

I1 bear	**I2 bear**	**M bear**	**O bear**	***D* (auto.)**	***Z* (auto.)**	***D* (X)**	***Z* (X)**
ABC	Grizzly	Black	Panda	-0.002	-0.354	0.014	0.263
WHB_f	ABC	Black	Panda	0.002	0.364	-0.010	-0.314
WHB_f	WI	Black	Panda	0.004	0.243	-0.018	-0.086

Table 1.6. *D*-statistic and *Z* score for admixture test between brown bears, polar bears and the American black bear. The highest coverage polar bears were selected for this analysis. Abbreviations are as in Table 1.3.

Table 1.7 shows the results of the *D*-statistic test for our ABC Island brown bear (ABC (Adm)) and the two ABC Island brown bears recently sequenced by Miller and colleagues (Admiralty and Baranof) (Miller *et al.* 2012). We generated reference-based alignments for each bear as described above, and selected a random base-call from each site in the reference genome again as described above.

I1 bear	I2 bear	M bear	O bear	*D* (auto.)	*Z* (auto.)	*D* (X)	*Z* (X)
Admiralty	Baranof	WHB_f	Black	**0.053**	**3.722**	0.080	0.769
Admiralty	Baranof	WHB_m	Black	**0.053**	**3.689**	0.087	0.746
Admiralty	Baranof	SBS	Black	**0.051**	**3.563**	0.058	0.683
Admiralty	Baranof	CS	Black	**0.049**	**3.296**	0.024	0.197
Admiralty	Baranof	NBS	Black	**0.051**	**3.552**	0.050	0.352
Admiralty	Baranof	WI	Black	**0.054**	**3.739**	0.063	0.556
Admiralty	Baranof	LS	Black	**0.052**	**3.543**	0.078	0.667
Admiralty	Baranof	SVB	Black	**0.050**	**3.505**	0.035	0.323
Admiralty	ABC (Adm)	WHB_f	Black	-0.021	-1.670	0.000	-0.002
Admiralty	ABC (Adm)	WHB_m	Black	-0.019	-1.490	0.040	0.252
Admiralty	ABC (Adm)	SBS	Black	-0.022	-1.673	-0.029	-0.183
Admiralty	ABC (Adm)	CS	Black	-0.019	-1.440	-0.060	-0.380
Admiralty	ABC (Adm)	NBS	Black	-0.022	-1.631	-0.031	-0.187
Admiralty	ABC (Adm)	WI	Black	-0.021	-1.643	-0.038	-0.215
Admiralty	ABC (Adm)	LS	Black	-0.019	-1.417	0.002	0.014
Admiralty	ABC (Adm)	SVB	Black	-0.020	-1.528	0.012	0.071
Admiralty	Grizzly	WHB_f	Black	**-0.035**	**-2.908**	**-0.250**	**-3.357**
Admiralty	Grizzly	WHB_m	Black	**-0.036**	**-2.972**	**-0.245**	**-3.483**
Admiralty	Grizzly	SBS	Black	**-0.037**	**-3.013**	**-0.244**	**-3.208**
Admiralty	Grizzly	CS	Black	**-0.037**	**-2.932**	**-0.270**	**-2.645**
Admiralty	Grizzly	NBS	Black	**-0.038**	**-3.122**	**-0.252**	**-4.326**
Admiralty	Grizzly	WI	Black	**-0.035**	**-2.947**	**-0.275**	**-3.567**
Admiralty	Grizzly	LS	Black	**-0.035**	**-2.889**	**-0.232**	**-2.736**
Admiralty	Grizzly	SVB	Black	**-0.038**	**-3.137**	**-0.232**	**-2.370**
Baranof	ABC (Adm)	WHB_f	Black	**-0.070**	**-4.868**	-0.099	-0.995
Baranof	ABC (Adm)	WHB_m	Black	**-0.070**	**-4.690**	-0.086	-0.996
Baranof	ABC (Adm)	SBS	Black	**-0.071**	**-4.853**	-0.090	-1.009
Baranof	ABC (Adm)	CS	Black	**-0.066**	**-4.324**	-0.066	-0.569
Baranof	ABC (Adm)	NBS	Black	**-0.070**	**-4.691**	-0.086	-0.809

Baranof	ABC (Adm)	WI	Black	**-0.072**	**-4.869**	-0.110	-0.984
Baranof	ABC (Adm)	LS	Black	**-0.068**	**-4.612**	-0.107	-1.285
Baranof	ABC (Adm)	SVB	Black	**-0.067**	**-4.490**	-0.018	-0.215
Baranof	Grizzly	WHB_f	Black	**-0.082**	**-6.260**	**-0.298**	**-3.170**
Baranof	Grizzly	WHB_m	Black	**-0.083**	**-6.291**	**-0.312**	**-3.511**
Baranof	Grizzly	SBS	Black	**-0.083**	**-6.376**	**-0.267**	**-2.511**
Baranof	Grizzly	CS	Black	**-0.080**	**-6.040**	**-0.274**	**-2.645**
Baranof	Grizzly	NBS	Black	**-0.084**	**-6.303**	**-0.280**	**-2.447**
Baranof	Grizzly	WI	Black	**-0.083**	**-6.332**	**-0.296**	**-3.162**
Baranof	Grizzly	LS	Black	**-0.081**	**-6.156**	**-0.285**	**-3.366**
Baranof	Grizzly	SVB	Black	**-0.081**	**-6.269**	**-0.255**	**-2.325**
WHB_f	WHB_m	Admiralty	Black	-0.013	-0.595	-0.115	-0.434
WHB_f	SBS	Admiralty	Black	-0.002	-0.100	0.196	2.149
WHB_f	CS	Admiralty	Black	-0.006	-0.268	-0.088	-0.284
WHB_f	NBS	Admiralty	Black	-0.001	-0.048	-0.017	-0.100
WHB_f	WI	Admiralty	Black	0.008	0.409	0.026	0.140
WHB_m	SBS	Admiralty	Black	-0.002	-0.085	0.222	0.988
WHB_m	CS	Admiralty	Black	0.002	0.085	0.091	0.364
WHB_m	NBS	Admiralty	Black	0.023	1.177	-0.026	-0.104
WHB_m	WI	Admiralty	Black	0.024	1.166	0.098	0.303
SBS	CS	Admiralty	Black	0.000	-0.015	**-0.394**	**-2.077**
SBS	NBS	Admiralty	Black	0.016	0.776	-0.254	-1.245
SBS	WI	Admiralty	Black	0.013	0.638	-0.083	-0.292
CS	NBS	Admiralty	Black	0.015	0.715	-0.053	-0.101
CS	WI	Admiralty	Black	0.003	0.149	0.051	0.242
NBS	WI	Admiralty	Black	0.002	0.112	0.024	0.129
WHB_f	WHB_m	Baranof	Black	-0.014	-0.686	-0.059	-0.149
WHB_f	SBS	Baranof	Black	-0.011	-0.544	0.235	1.857
WHB_f	CS	Baranof	Black	-0.016	-0.643	-0.161	-0.491
WHB_f	NBS	Baranof	Black	-0.019	-0.932	0.167	1.877
WHB_f	WI	Baranof	Black	-0.002	-0.119	0.023	0.089
WHB_m	SBS	Baranof	Black	-0.020	-0.966	0.279	0.926
WHB_m	CS	Baranof	Black	-0.015	-0.613	0.273	1.247
WHB_m	NBS	Baranof	Black	-0.006	-0.292	0.063	0.318
WHB_m	WI	Baranof	Black	0.007	0.333	0.113	0.397
SBS	CS	Baranof	Black	0.008	0.350	-0.556	-1.666
SBS	NBS	Baranof	Black	0.007	0.343	-0.036	-0.119
SBS	WI	Baranof	Black	0.013	0.654	-0.091	-0.414
CS	NBS	Baranof	Black	0.003	0.138	0.154	0.345
CS	WI	Baranof	Black	0.015	0.702	-0.050	-0.205
NBS	WI	Baranof	Black	0.014	0.703	-0.171	-0.959
Mean Values							
Admiralty	Baranof	Any Polar	Black	**0.052**	**3.576**	0.060	0.537
Admiralty	ABC (Adm)	Any Polar	Black	-0.020	-1.561	-0.013	-0.079

Admiralty	Grizzly	Any Polar	Black	**-0.036**	**-2.990**	**-0.250**	**-3.212**
Baranof	ABC (Adm)	Any Polar	Black	**-0.069**	**-4.675**	-0.083	-0.858
Baranof	Grizzly	Any Polar	Black	**-0.082**	**-6.253**	**-0.283**	**-2.892**

Table 1.7. *D*-statistic and *Z* score for admixture test between three ABC Islands brown bear and polar bears, using the American black bear as outgroup. Brown bears *Admiralty* and *Baranof* are the two ABC Islands brown bears recently published by Miller and colleagues, and are labeled according to island of origin. Our ABC Island brown bear is also from Admiralty Island, and is labeled ABC (Adm). Other abbreviations are as in Table 1.3. Significant deviations from $D=0$ are highlighted in bold. The Lancaster Sound polar bear is not included as either I1 or I2.

Although the amount of admixture detected on both autosomal and X-chromosome scaffolds is greater for the two additional ABC Islands brown bears than for our ABC Island brown bear, these two bears also show surprisingly large differences between their autosomal and X-chromosome *D*-statistics. Notably, the brown bear from Baranof Island, the most distant island from the Alaskan mainland, appears to have the most polar bear admixture. The other two bears are from Admiralty Island, the closest island to the Alaskan mainland. The observed differences in estimated *D*-statistics may therefore reflect a gradient of admixture on the ABC Islands. Additional sampling of ABC Island bears will be necessary to fully understand the process of admixture and to determine the role geography and stochastic genetic or demographic processes in determining the distribution of residual polar bear DNA in the ABC Islands brown bear population.

Finally, we performed a *D*-statistic test using the genomic data set from a ~115,000 year old polar bear fossil from Poolepynten, Svalbard, Norway (Miller *et al.* 2012). As above, we generated a reference-based alignment of the ancient polar bear and selected a random single base call from each site in the reference genome. Because the ancient polar bear has very low coverage, we decreased the minimum base call confidence cutoff to phred=20.

If the admixture event(s) detectable in the genomes of the ABC Islands brown bears was, in fact, confined to the ABC Islands, we should also see no sign of admixture in the ancient polar bear relative to other polar bears. To test for a different signature of admixture in the ancient polar bear we performed the *D*-statistic test as:

D(ABC, Grizzly, Ancient Polar Bear, Am. Black Bear)

We find that the ancient polar bear results in the same pattern of admixture as the modern polar bears (autosome D=-0.015; X-chromosome D=-0.213). These results support the hypothesis that the observed admixture was confined to the ABC Islands.

To investigate the local decay of the admixture signal from polar bears into brown bears, we performed the following analysis. We chose ABBA and BABA sites in the autosomal and X-chromosome scaffolds and then measured the *D*-statistic downstream of this focal site in 5kb windows extending out to 50kilobases. The ABBA and BABA focal sites were chosen such that windows would not overlap. The results of this analysis, shown in Figure 1.6, are consistent with two important expectations for local *D*-statistics locally around such sites. First, the presence of an ABBA or BABA site indicates specific tree topology for at least one of the haplotypes at that site. This topology should extend until recombination changes it. Thus, a single ABBA or BABA sites implies more of the same, locally, and a strong skew of the *D*-statistic in the direction of the focal observation. This result is observed for both autosomes and the X-chromosome. Second, recombination away from the ABBA or BABA implied topology should cause the *D*-statistic to regress toward the genome-wide mean at increasing distance from the focal ABBA or BABA site. Again, the results are consistent with our expectation.

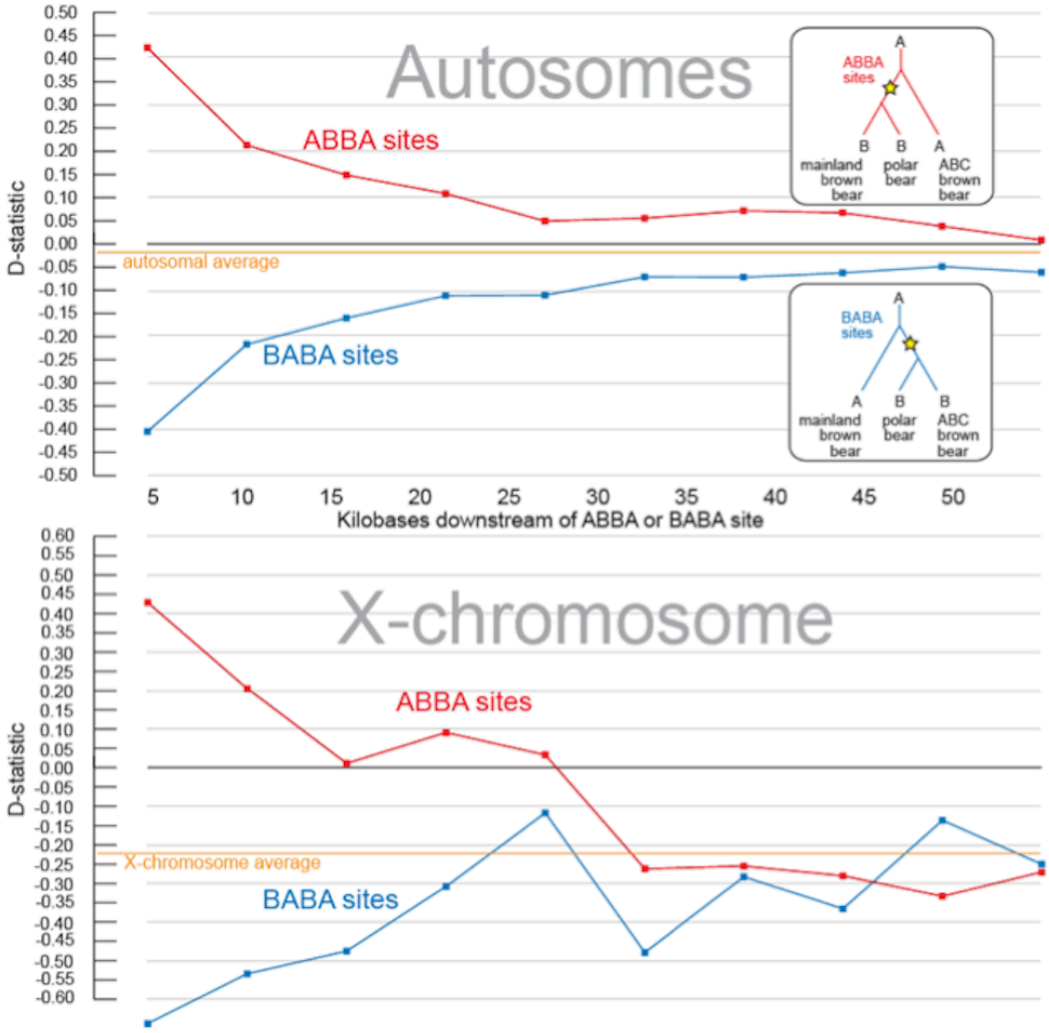


Figure 1.6. Decay of *D*-statistic downstream of ABBA and BABA sites. ABBA and BABA sites for (mainland brown bear, ABC island bear, polar bear, black bear) imply a specific topology (insets) at that site for the sampled haplotypes. *D*-statistics in the downstream vicinity of this focal SNP are heavily biased in the direction of the original observation, as expected.

2.3 Determining the Percentage of the Genome Resulting from Admixture:

We used the *D*-statistic to calculate the proportion of the various genomes that might be derived from admixture using the $\hat{f}$ estimator (Green *et al.* 2010; Durand *et al.* 2011). Informally, this estimates the proportion of an admixed genome deriving from admixture by comparing the rate of derived allele sharing as a proportion of that which would result from a completely admixed individual, M2.

S(I1,I2,M1,O) is the numerator of the *D*-statistic: ABBA-BABA

$$\hat{f} = (S(I1,I2,M2,O)/S(I1,M1,M2,O))$$

We calculated $\hat{f}$ for all combinations of two brown bears and two polar bears in our study. When M1 and M2 are members of the same species as is the case here $\hat{f}$ will underestimate the amount of the genome resulting from admixture (f) by the following amount (Durand *et al.* 2011):

$$f = \hat{f} \text{ (time\_of\_speciation} - \text{time\_of\_admixture) / time\_of\_speciation}$$

As the timing of the admixture event predicted here is much more recent than the timing of initial species divergence, we assume that $\hat{f}$ is a reasonable approximation of *f*.

Although some comparisons of differential admixture between polar bears produced non-zero *D*-statistic values, none were significant via the block jackknife test. We note also that the *number* of ABBA and BABA sites is extremely small when comparing two polar bears. This is a straightforward consequence of the small amount of genetic variation within polar bears as ABBA and BABA sites require that the two polar bears *differ* from one another at that site.

In contrast, the brown bears show a clear and consistent pattern of differential admixture regardless of which polar bear is used as the candidate for admixture. The average $\hat{f}$ values are 0.75% (stdev = 0.094%) admixture on the autosomes and 6.5% (stdev = 0.483%) admixture on the X chromosome, for an average percent ratio of 8.8 (Figure 1.7).

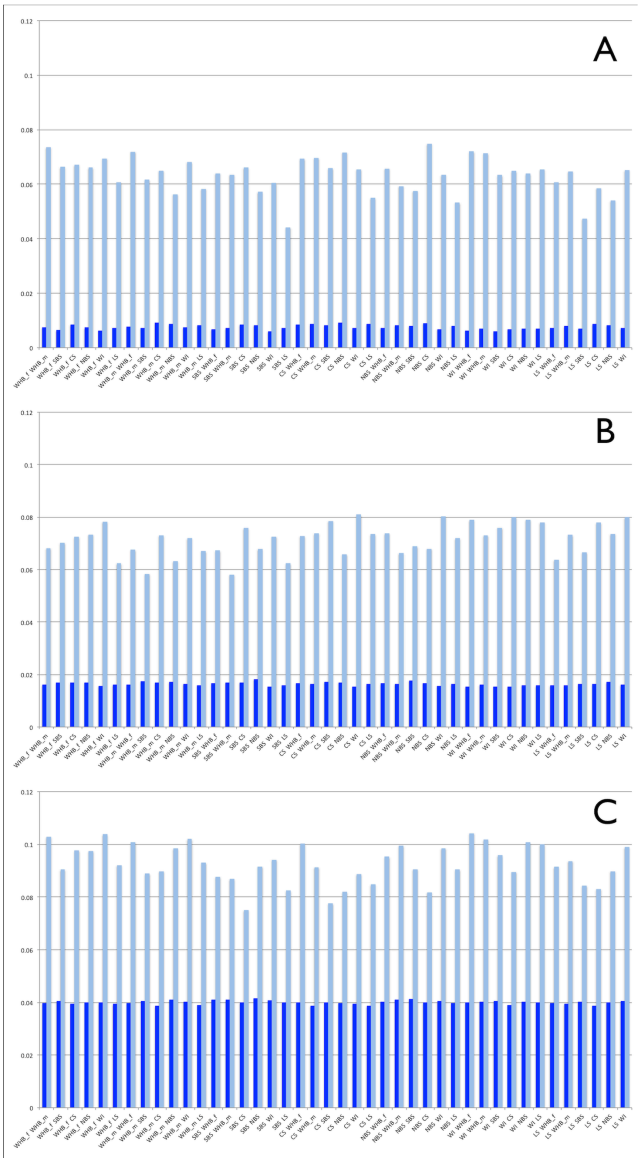


Figure 1.7. Proportion of polar bear ancestry of the ABC Islands brown bears calculated using f. The proportion of polar bear ancestry inferred for the autosomes (dark blue) and X chromosome (light blue) is shown for each ABC Islands brown bear; (A) the Admiralty Island brown bear sequenced in this study, (B) the Admiralty Island brown bear of Miller et al, (C) the Baranof Island brown bear of Miller et al (Miller *et al.* 2012). The bears from Admiralty Island show similar amounts of polar bear ancestry but the amount inferred for the Baranof Island bear is much greater. This may be due to the greater distance from the mainland of Baranof Island limiting brown bear immigration to a greater degree than on the more accessible Admiralty Island. The inverse correlation of X chromosome: autosome ratio and total amount of polar bear ancestry is also consistent with our model of population and genome conversion form polar bears to brown bears via sex biased brown bear introgression (Figure 1.13).

2.4 Characterizing the direction of gene flow

To test the direction of gene flow between polar bears and brown bears, we substituted 6.5% (a close approximation of the amount of admixture estimated for the X-chromosome, see section 2.3) of the female West Hudson Bay polar bear X chromosome with corresponding X chromosome data from mainland Alaska brown bear (Figure 1.3B) and vice versa (Figure 1.3C). For each simulation we randomly selected multiple regions of the X chromosome to be replaced by introgression from the opposite species. To capture the approximate size of haplotypes that would be introduced if the admixture occurred 50kya (assuming 1cM/Mb; generation time of 10 years), admixed regions were 20kb in length. To simulate more recent admixture, we used a smaller number of longer regions, and to simulate more ancient admixture, we used a larger number of shorter regions. To maximize resolution, we conducted this test on female bears only.

Because our pairwise difference method (described in Supplementary section 2.1) uses only a single high quality base call to represent each site, we must distinguish in our simulations between heterozygous introgessed regions (the vast majority) and homozygous introgessed regions. Heterozygous introgressed regions are those where randomly selected admixed regions do not overlap; in these regions we select the base call from the original individual and the introgressing individual randomly with equal probability to represent sampling from one introgressed and one nonintrogressed chromosome. For cases where two introgressed regions overlap, we consider the introgression to be homozygous. In such cases we select exclusively from the introgressing individual in creating the simulated introgressed chromosome. Note that because introgressed regions are simulated as occurring independently on each chromosome the sum of the length of all introgressed haplotypes is equal to two times the amount of admixture predicted in section 2.3, in this case 13% of the length of the X chromosome recovered as described in section 1.3.

Figure 1.8. Simulated introgression. To simulate introgression of the amount predicted from our data, we randomly replace sections of the original sequence, shown in blue, with sequence from the introgressor species, shown in red. When only a single introgressed region covers a site in the reference genome it is considered heterozygous, shown in purple, and is represented by either the introgressed or original sequence with equal probability. If two introgressed regions overlap then it is considered to be homozygously introgressed, as is the case on the right side of this figure and in the red region only introgressor sites are selected to represent the individual for the pairwise difference calculation.

We then count the number of pairwise differences in non-overlapping 50kb bins between the simulated introgressed chromosome and the actual data from an individual of the same species (as for Figure 1.2A) and compare the results to the true data.

Substituting 6.5% of the polar bear X chromosome with brown bear X chromosome results in an excess of highly diverging bins (bins in which the number of pairwise differences are greater than ~8 in 10,000) compared to the real data for all 100 simulations (Figure 1.3B). In contrast, substituting 6.5% of the mainland brown bear X chromosome with polar bear X chromosome results in no observable difference compared to the real data (Figure 1.3C).

To test whether the size of the admixture blocks (and associated estimated time of introgression) influences the results, we performed several additional simulations in which block sizes were selected to correspond to admixture occurring in 10,000 year increments spanning the period 10 kya - 200 kya. In all cases, substituting 6.5% of the brown bear X chromosome into polar bears yields an excess of deeply diverging regions beyond what is seen in real polar bear X chromosome sequence data (Figure 1.9). This suggests that while Figure 1.2 depicts results with an assumed time of admixture of 50 kya, similar results are expected for any time of admixture in the range 10-200 kya.

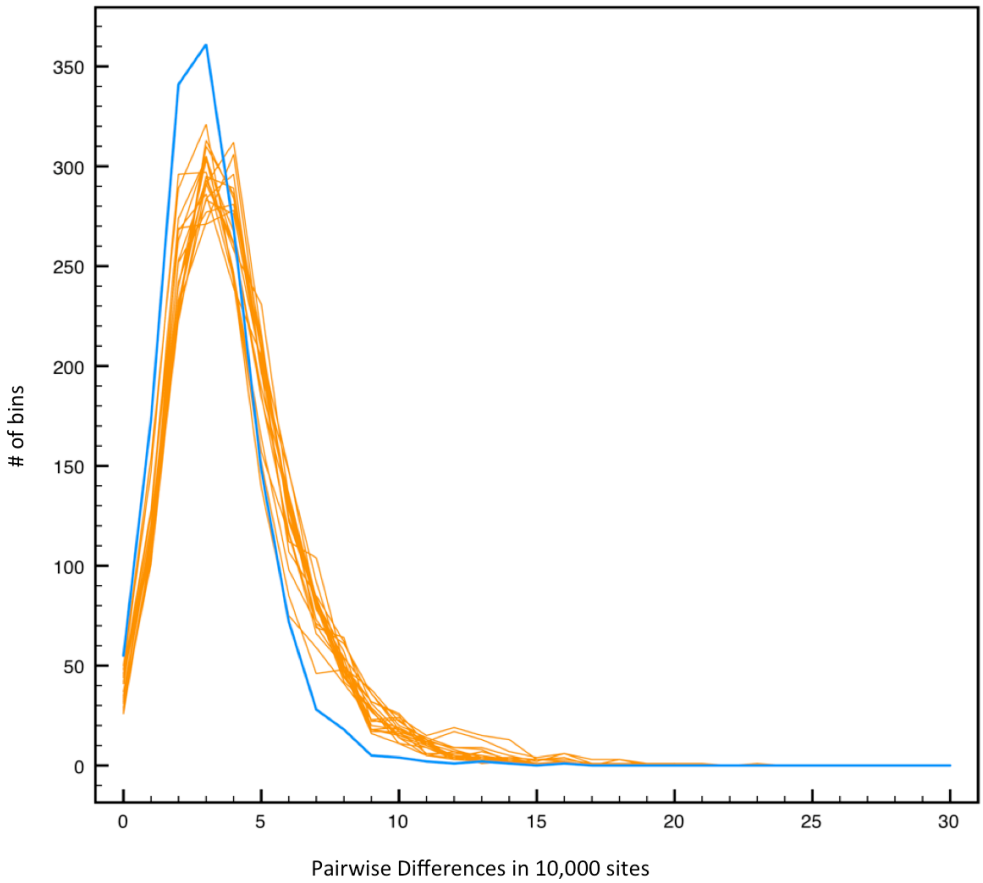


Figure 1.9. Simulations of brown bear into polar bear admixture of various block lengths. In orange are simulations of 6.5% admixture into polar bears in 10,000-year time intervals from 10Kya to 100Kya. The observed pairwise difference between the two female polar bears in the study is shown in blue. There is no systematic effect from different hypothetical times of admixture and all show the same pattern of increased numbers of highly divergent regions of the X chromosome.

Since the higher within-species diversity of brown bears overlaps with the distribution of divergence between polar bears and brown bears (see Figures. 1.2B, 1.5), this test is less powerful to detect the presence of admixture of polar bears into brown bears than *vice versa*. Thus, at present we cannot rule out the possibility of polar bear admixture into *all* brown bears. However, these results do argue against polar bears being the *recipient* of admixture, especially the admixture seen in ABC Island brown bear versus mainland brown bear X chromosome comparison. Thus, we conclude that the ABC Island brown bear

and the population it represents are the *recipient* and not the *donor* of X-chromosome genetic material.

2.5 Inferring population size through time using PSMC

We calculated the effective population size through time of each of the three species in our study using Li and Durban's PSMC (Figure 1.10) (Li & Durbin 2011). PSMC utilizes the density of heterozygous sites within a single individual to infer population size through time. Because identification of heterozygous sites is sensitive to the depth of sequencing, we restricted our analysis to just those individuals with at least 10X coverage. These were the two female brown bears, the male American black bear and two polar bears following additional sequencing (see section 1.1, Table 1.3): the female West Hudson Bay bear and the male Chukchi Sea bear. PSMC was limited to scaffolds mapping to the autosomes.

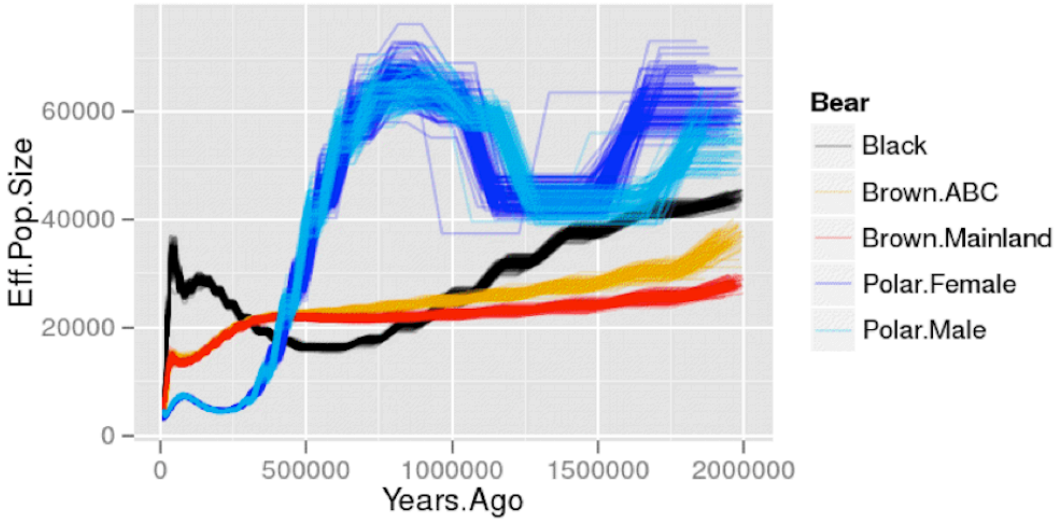


Figure 1.10. Autosomal population sizes through time as estimated with PSMC. 100 bootstrap replicates are shown for the 5 bears listed. We assume a generation time of 10 years and a mutation rate of 1×10^{-9} substitutions/site/year. Note that individuals of the same species show similar profiles. However, polar bears and brown bear profiles do not converge over the time period shown.

We note that although the shapes of the curves are similar to those of Miller *et al.* [9], we did not observe as large an inferred effective population size for brown bears. The lack of converging population size histories between polar

bears and brown bears suggests either that these species were separate populations through the time period in which PSMC has resolution (~1,000,000 years using the generation time and mutation rate described above), or that population structure is complicating the inference of population size through time (Li & Durbin 2011). Similarly, the very large population size of polar bears in the past may reflect either a much larger effective population size or a high degree of past population structure.

2.6 Identifying genomic regions that may be under adaptive evolution in the polar bear

We performed a screen for regions of the polar bear genome that are potentially under very strong or recent positive selection. We identified regions of the genome that satisfy all of the following criteria: (i) polar bears have low within-species variation (ii) polar bears are distantly diverged from brown bears, and (iii) brown bears have average amounts of within-species variation. These criteria were designed to identify candidate regions of polar bear adaptation that are now evolving under strong purifying selection in polar bears.

We generated a vcf file sing GATK with all of the individuals in our sample. Then we generated a sliding window of 50kb width across each scaffold in 1kb increments. In each region we calculated the number of variable sites within polar bears, the number of variable sites in brown bears and the number of fixed differences between polar bears and brown bears. Then for regions where there were at least 20 variable sites in brown bears we define a PBAR (polar bear accelerated region) score as (polar bear variable sites / fixed differences). At that point we set an arbitrary score cutoff of 0.0218. We merged all overlapping regions with a score less than the cutoff and assigned the lowest score in the merged region. After merging we selected the 100 lowest-scoring regions for analysis.

Because the polar bear genome is still lacking gene annotation, we used *genBLASTa* (She *et al.* 2008) to map dog genes from ENSEMBL to the polar bear genome. We then identified genes that intersected with the PBARs. Although many PBARs do not contain known genes or contain genes of unknown function we did identify several genes that are of potential interest in relation to polar bear evolution. These are shown in Table 1.8. We note that KCNT2 was also identified as a candidate for selection in polar bears by Miller and colleagues (Miller *et al.* 2012).

polar bear scaffold	start : end	PBAR score	Gene Name
scaffold120	1778000 : 1906000	0.0000	
scaffold33	4186000 : 4254000	0.0000	
scaffold65	3241000 : 3308000	0.0000	
scaffold23	2212000 : 2278000	0.0000	
scaffold15	23999000 : 24056000	0.0000	
scaffold15	14041000 : 14094000	0.0000	
scaffold62	12365000 : 12415000	0.0000	NFKB1
scaffold33	879000 : 929000	0.0000	
scaffold101	368000 : 418000	0.0000	
scaffold6	27302000 : 27378000	0.0047	
scaffold66	12141000 : 12200000	0.0061	
scaffold17	19851000 : 19905000	0.0068	
scaffold16	7384000 : 7437000	0.0072	
scaffold17	20451000 : 20504000	0.0072	
scaffold6	27302000 : 27498000	0.0077	
scaffold11	26036000 : 26086000	0.0084	
scaffold11	26040000 : 26090000	0.0089	TTC3
scaffold29	1399000 : 1461000	0.0090	
scaffold78	5958000 : 6009000	0.0094	
scaffold2	15636000 : 15709000	0.0095	
scaffold83	5896000 : 5956000	0.0097	
scaffold19	6882000 : 6935000	0.0098	SAG S-antigen/ ATG16L1
scaffold13	6092000 : 6154000	0.0099	TMEM39B
scaffold19	6684000 : 6738000	0.0099	USP40
scaffold13	6106000 : 6156000	0.0100	KHDRBS1
scaffold33	13477000 : 13535000	0.0102	LCA5
scaffold139	4232000 : 4293000	0.0102	KIAA0232

scaffold30	5340000 : 5396000	0.0102	
scaffold78	7593000 : 7643000	0.0106	
scaffold27	5780000 : 5831000	0.0108	ANAPC10
scaffold9	6803000 : 6853000	0.0112	PIK3R1
scaffold55	1305000 : 1367000	0.0114	
scaffold30	16056000 : 16111000	0.0114	
scaffold11	19774000 : 19829000	0.0115	
scaffold82	3569000 : 3636000	0.0115	
scaffold10	9791000 : 9844000	0.0116	
scaffold66	522000 : 574000	0.0119	
scaffold15	23999000 : 24049000	0.0120	KIAA0196
scaffold220	268000 : 325000	0.0122	PDSS1
scaffold6	35754000 : 35818000	0.0123	
scaffold95	3142000 : 3192000	0.0128	
scaffold1	20665000 : 20717000	0.0131	EHBP1
scaffold48	218000 : 268000	0.0131	
scaffold155	1921000 : 1986000	0.0132	SERPIND1
scaffold55	2900000 : 2950000	0.0132	
scaffold15	11408000 : 11458000	0.0133	EBAG9/ SYBU/ SNPH
scaffold60	11749000 : 11801000	0.0137	
scaffold140	3455000 : 3511000	0.0144	
scaffold21	15269000 : 15319000	0.0144	CLVS2
scaffold250	188000 : 243000	0.0145	
scaffold75	7819000 : 7870000	0.0145	
scaffold14	2533000 : 2583000	0.0145	
scaffold5	28373000 : 28427000	0.0149	GPHN
scaffold84	7504000 : 7558000	0.0149	COX10
scaffold1	44825000 : 44876000	0.0149	
scaffold43	13061000 : 13114000	0.0154	VMP1
scaffold55	4615000 : 4665000	0.0154	
scaffold28	4113000 : 4164000	0.0155	
scaffold48	4000 : 57000	0.0156	
scaffold17	8070000 : 8123000	0.0159	
scaffold6	35121000 : 35177000	0.0160	ELAC1
scaffold13	8710000 : 8767000	0.0160	GMEB1
scaffold48	9700000 : 9752000	0.0160	
scaffold181	2392000 : 2444000	0.0169	
scaffold45	15653000 : 15707000	0.0172	SCMH1
scaffold57	3713000 : 3763000	0.0174	PDE10A
scaffold55	2536000 : 2591000	0.0175	LOC612664
scaffold11	3255000 : 3327000	0.0175	
scaffold40	8193000 : 8248000	0.0180	
scaffold45	14544000 : 14594000	0.0180	

scaffold41	8699000 : 8750000	0.0182	
scaffold55	4311000 : 4362000	0.0182	
scaffold55	2312000 : 2363000	0.0182	PREP
scaffold8	14356000 : 14406000	0.0182	
scaffold18	17382000 : 17439000	0.0189	ASTN1
scaffold75	7261000 : 7314000	0.0189	
scaffold155	1950000 : 2004000	0.0190	SNAP29
scaffold7	512000 : 562000	0.0190	
scaffold13	20852000 : 20904000	0.0192	
scaffold82	3503000 : 3557000	0.0194	
scaffold81	524000 : 574000	0.0194	
scaffold6	27439000 : 27491000	0.0195	
scaffold38	7100000 : 7150000	0.0196	
scaffold255	70000 : 123000	0.0197	KCNT1/ KCNT2
scaffold8	25229000 : 25281000	0.0203	C28H10orf22
scaffold5	35857000 : 35907000	0.0203	
scaffold133	2333000 : 2385000	0.0203	
scaffold12	3411000 : 3467000	0.0204	
scaffold67	6066000 : 6117000	0.0204	
scaffold14	9003000 : 9053000	0.0205	
scaffold1	9082000 : 9137000	0.0206	FSHR
scaffold57	3691000 : 3741000	0.0206	
scaffold5	27529000 : 27579000	0.0208	
scaffold45	7733000 : 7784000	0.0210	
scaffold71	5482000 : 5533000	0.0211	
scaffold8	3258000 : 3310000	0.0212	
scaffold131	2224000 : 2277000	0.0213	RNGTT
scaffold41	3223000 : 3273000	0.0213	
scaffold40	8204000 : 8254000	0.0214	SLC46A3
scaffold9	19506000 : 19556000	0.0216	SH3RF2

Table 1.8. Candidate genetic regions for polar bear adaptation. The genomic coordinates of each of the 100 lowest Polar Bear Accelerated Regions (PBAR) scoring regions are shown along with the dog genes, if any, map to these regions.

2.7 Inferring the timing of the origin of the polar bear lineage

The evolutionary relationship between brown bears and polar bears has been a contentious issue for over half of a century. The fossil record for polar bears is markedly lacking (Kurten 1964; Harington 2008; Ingólfsson & Wiig 2009), due to their preference for arctic shelf ice and continental edges, where remains are unlikely to be preserved over geological time. The two oldest known polar

bear fossils date to around the same age: one from Poolepynten, Svalbard, is estimated to have lived around 110-130 kya (Ingólfsson & Wiig 2009), and another from Kiøpsvik, Norway, dates to around 115 kya (Davison *et al.* 2011). Both of these are clearly identifiable as polar bears rather than brown bears, placing a strict lower bound on when polar bears first appeared as a distinct lineage.

In 1964, Björn Kurtén concluded that polar bears evolved from a brown bear–like species no earlier than the Mindel glacial stage of the early Middle Pleistocene (ca. 750-675 kya) (Kurten 1964). He based this estimate on a comparative analysis of allometric growth patterns in the modern polar bear skull with fossil remains from brown bears and cave bears dating to the early Middle Pleistocene. He noted, however, that some morphological characteristics of modern polar bears, specifically tooth patterns, may have continued to evolve until as recently as the last 10-20,000 years, as polar bear have become increasingly specialized in their extreme environment. Kurtén therefore suggested that the divergence between brown and polar bears may also have occurred much more recently. In making this conclusion, he noted that a more recent divergence could explain why the two species continue to produce fertile hybrid offspring both in zoos (Kowalska 1964; Preuß *et al.* 2009) and in the wild (Stirling 2011).

The first decade or so of genetic analyses of brown and polar bears focused mainly on maternally-inherited mitochondrial DNA, and provided additional support for the recent-divergence hypothesis. Polar bears and brown bears share a common mitochondrial ancestor within the within the last 110-160 kya (Lindqvist *et al.* 2010; Davison *et al.* 2011; Edwards *et al.* 2011), a period that spans the interval from which the two oldest polar bear fossils are known. In addition, the mitochondrial lineage to which all living polar bears belong falls

within the diversity of brown bears, with their closest *living* relatives a population of brown bears from the ABC Islands of southeastern Alaska (Shields & Kocher 1991; Cronin *et al.* 1991; Talbot & Shields 1996; Waits *et al.* 1998). This lineage (Figure 1.11), which also includes ancient brown bears from Ireland (Edwards *et al.* 2011) and Beringia (Barnes *et al.* 2002) and mitochondrial lineages from the two polar bear fossils (Lindqvist *et al.* 2010; Davison *et al.* 2011; Miller *et al.* 2012) is by convention referred to as *mitochondrial clade II* (Leonard *et al.* 2013).

Several hypotheses have been proposed to explain the branching order within mitochondrial clade II. Initially, when genetic data were available only from living individuals (ABC Island brown bears and polar bears), it was hypothesized that the ABC Islands brown bears represented a very old brown bear lineage, and that their close relationship to polar bears reflected the divergence between the two species (Heaton *et al.* 1996). However, this hypothesis is inconsistent with four lines of evidence.

First, geological data suggest that the ABC islands were mostly overridden by ice during the peak of the last ice age (Last Glacial Maximum, or LGM; ca 26-14 kya) and therefore not habitable by brown bears (Carrara *et al.* 2007). This makes it unlikely that a very old lineage would survive, isolated, in this region. Fossil evidence supports the geological data; brown bear fossils are known from Prince of Wales Island (just south of the ABC Islands) both before (two bones dating to 26 and 31 kya) and after (several bones dating to less than 12 kya) the LGM (Heaton & Grady 2003). However, no brown bear bones are found on either Prince of Wales or the ABC Islands during the LGM. Ample fossil remains of ringed seals (*Phoca hispida*), as well as bones from Arctic foxes (*Alopex lagopus*) and Steller sea lion (*Eumetopias jubatus*) recovered from Prince of Wales Island (Heaton & Grady 2003) suggest that taphonomic conditions were favorable for preservation during this period.

Second, nuclear microsatellite data show no evidence of restricted gene flow between brown bear populations living on the ABC Islands today and brown bear populations in mainland Alaska (Paetkau *et al.* 1998). However, this same analysis does identify restricted gene flow between the Kodiak Islands and mainland Alaska, suggesting that if such a restriction did exist it would be observable from the microsatellite data (Paetkau *et al.* 1998).

Third, the most recent matrilineal ancestor of the present-day brown bear population on the ABC Islands lived 37-10kya (Edwards *et al.* 2011), well after morphologically distinct polar bears are known (Ingólfsson & Wiig 2009; Davison *et al.* 2011).

Finally, when the data from the two fossil polar bears are included in the mitochondrial tree (Figure 1.11), both polar bears and brown bears become paraphyletic with respect to each other (Edwards *et al.* 2011). These data suggest a different evolutionary scenario is required to explain the mitochondrial phylogeny.

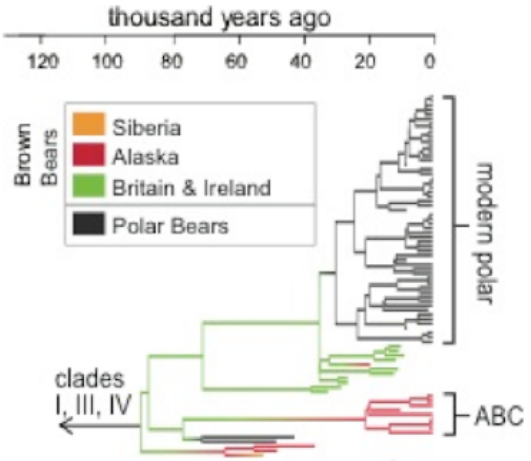


Figure 1.11. Mitochondrial phylogeny for polar bears, ABC Island brown bears and extinct Irish brown bears. Adapted from Edwards *et al* (Edwards *et al.* 2011).

Precisely when polar bears and brown bears initially split remains controversial. Estimates from nuclear DNA range from ~600 kya (Hailer *et al.* 2012), in agreement with Kurtén's allometry-based estimate, to 4-5 Mya (Miller *et*

al. 2012). The large amount of uncertainty in these estimates stems mainly from the poor fossil record within the ursid radiation, which has resulted in a lack of an appropriate calibration. The 600kya estimate of Hailer and colleagues (Hailer *et al.* 2012) was generated by calibrating the rate of molecular evolution using seven fossils, the most proximate of which was the divergence between the giant panda and the polar bear at around 12Mya. The oldest estimate, that of 4-5 Mya from Miller and colleagues (Miller *et al.* 2012), was based on a serial coalescent model that used an evolutionary rate of 10^{-9} substitutions per site per year. We note that our genomic data do not address the lack of a universally adopted fossil calibration. However, we attempt to compile and contrast various approaches as a necessary precursor to integrating our observations.

Miller and colleagues recently presented full-genome shotgun sequencing data to address the issues of the polar bear origin and subsequent hybridization more thoroughly than had been attempted previously (Miller *et al.* 2012). They identified sites in the nuclear genome that were polymorphic among and between polar and brown bears from high-coverage genome sequences of 27 bears, including 3 brown bears, a black bear, and 23 polar bear. They concluded that polar bears and brown bears diverged 4-5 Mya, at approximately the same time as the polar/brown bear lineage diverged from black bears (Miller *et al.* 2012). Interestingly, the model they propose includes a long period of admixture between brown, polar, and black bear lineages, lasting up until around 100-200 kya for the black bear and the brown/polar lineage, and until the present day between brown bears and polar bears (Miller *et al.* 2012). The finding that polar bears and brown bears formed sister lineages (in contrast to the paraphyletic relationship recovered from analysis of mitochondrial DNA) agreed with previous analyses of a small number of nuclear loci (Hailer *et al.* 2012). However the very ancient early divergence was much older than previously suggested,

either from morphological (Kurten 1964) or other nuclear data (Yu *et al.* 2004; Edwards *et al.* 2011; Hailer *et al.* 2012).

The American black bear divergence date given by Miller *et al.* is similar to the time to most recent common ancestor (TMRCA) of American black bears, brown bears and polar bears of 3.9-6.48 Mya estimated by Krause and colleagues (Krause *et al.* 2008). This estimate was based only on a mitochondrial phylogeny, but included both extinct and extant bears. However, Krause *et al.* estimate the TMRCA of polar bears and a non-ABC islands brown bear to be 0.66-1.17 Mya, again suggesting a recent divergence of brown bears and polar bears, and a long evolutionary distance between the divergence of these two lineages from black bears, and the subsequent divergence of these two lineages from each other. While our genomic data cannot confirm a specific time in which these two divergences occurred, they strongly support a long evolutionary distance between these two events, rather than a rapid radiation of all three lineages around the same time.

2.8 Modeling admixture between ABC Islands brown bears and polar bears

The *D*-statistics calculated for the four populations: mainland brown bear, ABC Islands brown bear, polar bear and either the panda or black bear as an outgroup, show that the ABC Islands brown bears are more similar than the mainland brown bear to the polar bear, suggesting admixture between ABC Islands brown bears and polar bears. A surprising result was the very large value of *D* for the X chromosome compared to the value for the autosome (hereafter denoted by D_x and D_{auto} respectively). The ratio of D_x to D_{auto} is 13.98 when the black bear is the outgroup.

Here, we investigate what underlying model of admixture can explain this discrepancy. In particular, we show that a single episode of admixture from polar bears into brown bears is highly improbable, whereas a continuous sex-

biased immigration of brown bears to a polar bear population is more likely to account for these observations.

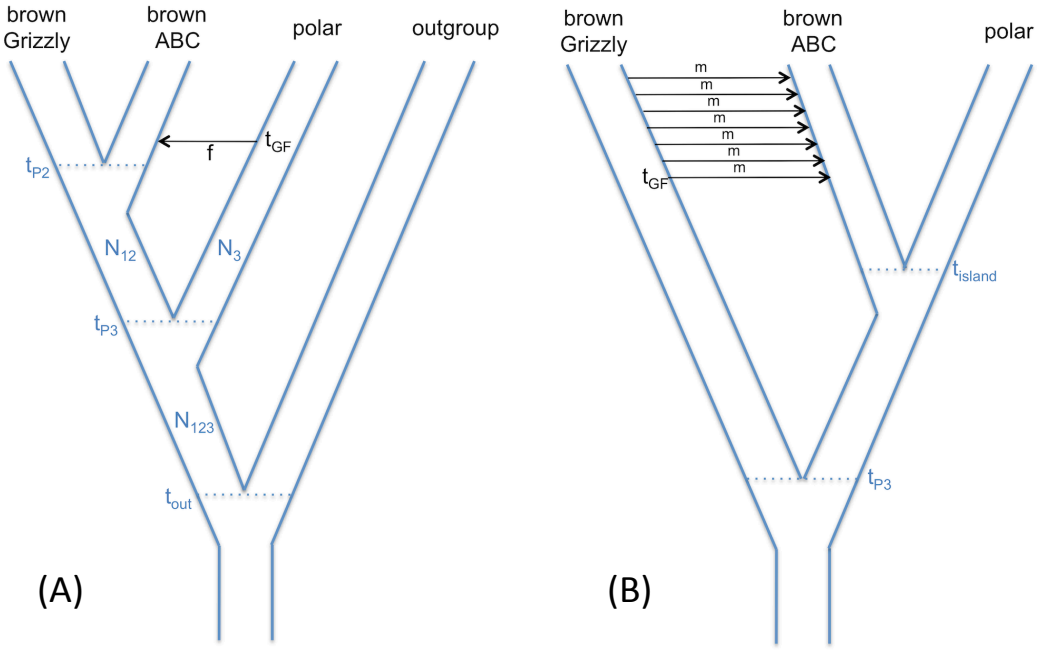


Figure 1.12. The two scenarios for admixture tested in the following section. (A) Shows a classic, single episode gene flow from polar bears into ABC bears with a magnitude of *f*, where *f* is the total amount of polar bear ancestry in ABC bears. (B) Shows the genome erosion model whereby the same amount of polar bear ancestry is achieved by continuous introgression of brown bears into an initially polar bear population.

I. Scenario 1: single episode of admixture from polar bears into the brown bear population.

Single episode of gene flow

Under the simple model of gene flow described in Figure 1.12, the analytical expectation of the *D*-statistic, *D(mainland, ABC, polar, outgroup)*, for autosomal sites can be derived (Durand *et al.* 2011). The expectation of *D* for X-linked sites ($E[D_x]$) can easily be obtained using equation (1) with effective population sizes scaled by ¾.

Single episode and sex-biased gene flow

If we assume a sex-biased process in which female polar bears mate with male brown bears but not the reverse, the expectation of D for autosomal sites is unchanged. However, if the admixing individual were female, D for X-linked sites will be larger than for autosomal sites because females bring relatively more X chromosomes into the brown bear population, X ($f_x > f$). The parameter f is the probability that an autosomal allele is inherited from a migrant polar bear at the time of gene flow:

$$f = P(allele\ inherited\ from\ a\ polar\ bear\ at\ time\ T_{GF})$$

$$f = P(allele\ from\ mother) * P(mother\ is\ polar\ bear) + P(allele\ from\ father) * P(father\ is\ a\ polar\ bear$$

In the case of completely sex-biased gene flow

$$\Pr(\text{the father is a polar bear}) = 0$$

Therefore,

$$P(mother\ is\ a\ polar\ bear) = \frac{f}{P(allele\ from\ mother)} = 2f$$

Similarly, f_x for the X chromosome is given by

$$f_X = P(allele\ from\ mother) * P(mother\ is\ a\ polar\ bear) + P(allele\ from\ father) * P(father\ is\ a\ polar\ bear)$$

$$f_X = \frac{2}{3}2f + \frac{1}{3}0$$

$$F_X = \frac{4}{3}f$$

The expectation of D_X in the case of sex-biased migration can thus be obtained from equation (1) by scaling the effective population sizes N_3, N_{12}, and N_{123} by ¾ and f by 4/3.

Based on the above results, the ratio $E[D_x]/E[D_{auto}]$ is larger if there is sex-bias than if there is no sex bias. Consequently, if the sex-biased model fails to predict large ratios, the non sex-biased model will fail as well. For this reason, we

focus on the sex-biased model and show that under this model the ratio cannot be as large as 14 for reasonable parameter values.

We set f to 0.007 (see section 2.2) and calculated the ratio D_X/D_{auto} for different set of parameters using the analytical expression. We explored the 6-dimension parameter space defined by $(t_{GF}, t_{P2}, t_{P3}, N_3, N_{12}, N_{123})$ as described in Table 1.9, resulting in 8,697,500 calculated ratios. All values were in the range [1.13, 2.67], and thus much smaller than the observed D_X/D_{auto} ratio of 14. Using different values of f does not change the range substantially.

Conclusion: A simple episode of polar bear gene flow into brown bears cannot explain the discrepancy observed between D calculated for autosome and D calculated for X, even in case of an extremely sex-biased gene flow.

II. Scenario 2: continuous migration of mainland male brown bears to the ABC islands.

Deterministic approach

In the second scenario, the ABC islands were initially populated with polar bears exclusively (Figure 1.12B). At some time t_{GF} in the past, male brown bears started to immigrate from the Alaskan mainland and replace a fraction m of the male bears in the ABC Islands at each generation. Assuming random mating on the island, we can obtain the recurrence equations that describe the change in frequency of a "polar bear allele" (PB allele) at an autosomal locus and at an X-linked locus. When studying alleles that are initially population-specific, and under the assumption of independence between sites, the frequencies of the PB allele can be interpreted as the expected percentage of polar bear ancestry on the island for the autosome and the X chromosome respectively. The aim is thus to compare the expected frequencies to the estimates of polar bear ancestry (~0.75% for the autosome, ~6.5% for the X, section 2.2) and to investigate if Scenario 2 can

explain a ratio of ancestry percentage as high as that observed in the real data (~8.8).

Starting with a PB allele frequency of 100% in ABC islands bears and 0% in mainland bears, after *t+1* generations of ongoing migration it can be shown that

- the frequency p_{t+1} of an autosomal PB allele in the island is given by

$$p_{t+1} = (1 - \frac{m}{2})\, p_t \qquad (2)$$

- the frequency g_{t+1} of an X-linked PB allele in the island female bears is given by

$$g_{t+1} = \frac{(1 - m)\, g_{t-1} + g_t}{2}$$

- the frequency h_{t+1} of an X-linked PB allele in the island males is given by

$$h_{t+1} = (1 - m)\, g_t$$

From equation (2) we obtain that the frequency of the PB autosomal allele is given by

$$p_{t+1} = (1 - \frac{m}{2})^{t-1} \qquad (3)$$

Assuming that the migration started 12kya (1200 generations; approximately when the ABC Islands would have again been habitable by brown bears) and is ongoing, and that the present frequency of the PB allele is 0.7%, we obtain from equation (3) that the migration rate *m* is approximately 0.0083. Figure 1.13 shows the frequencies of the PB autosomal allele and the PB X-linked allele as a function of the time period of undergoing immigration, for *m* set to 0.0083. The X-linked frequency decreases more slowly than the autosomal frequency, and after 1200 generations the ratio between the two frequencies is 5.22. Note that considering different starting times of immigration, and different migration rates (*m* in [0, 0.1]), did not substantially changed the ratio of frequencies (which was always in the range [5.1, 5.3]).

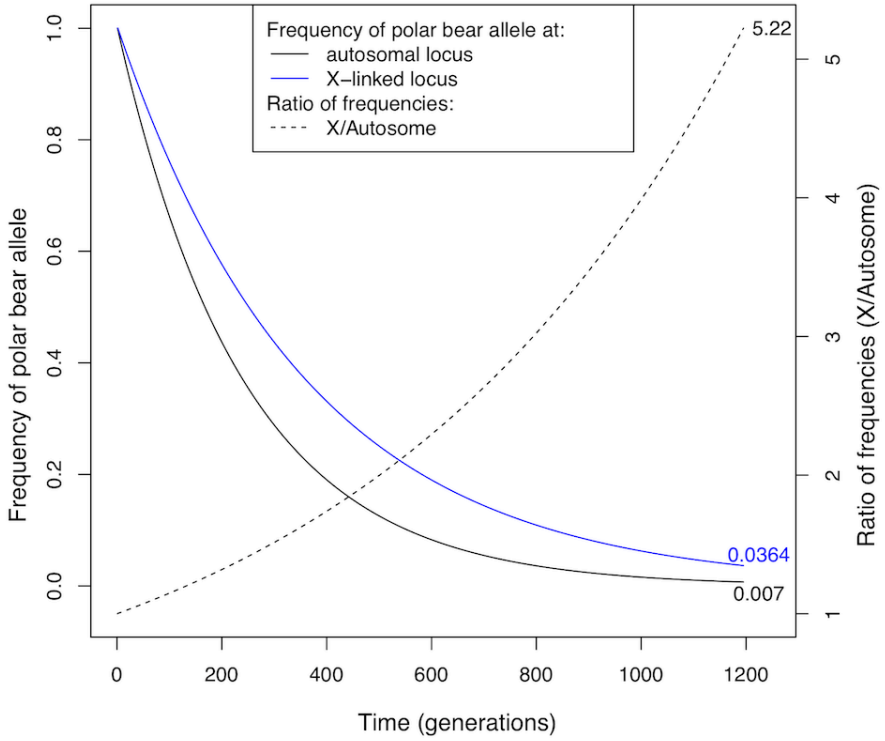


Figure 1.13. Changes in allele frequency through time with immigration. Left scale: Frequency of a polar bear allele for an autosomal locus (black line) and an X-linked locus (blue line) as a function of the time period of ongoing mainland brown bear immigration. Right scale: Ratio of the frequency for X and for the autosome. For this graph the migration rate *m* was set to 0.0083.

Although the expected ratio is smaller than the observed ratio (~8.8), Scenario 2 (continuous immigration) provides a much better fit to the data than does Scenario 1. Because stochastic variation in the migration and recombination processes, as well as sampling errors, are expected to generate some discrepancy between the expected and observed ratios (especially since the amount of data for X is limited, we further investigate the scenario using simulations.

Stochastic approach

Using the software ms (Hudson 2002) we simulated sequences for a Grizzly Brown bear, an ABC Brown bear and a Polar bear under scenarios of continuous immigration of Brown bear to the ABC islands (Figure 1.12B). We run simulations under non sex-biased, partially sex-biased, and sex-biased gene flow.

We set the demographic parameters to reasonable values *(N_1=N_{123}=30k, N_2=N_3=N_{23}=4k, t_{P3}=6000kya, t_{island}=100kya, t_{GF}=12kya)* and then calibrated the migration rate so that the *D* statistics calculated from the simulated autosomal data fit the observed autosomal *D(ABC, Grizzly, Polar, Panda)* (*-0.017*). Assuming non sex-biased gene flow, a migration rate of 0.525% (% of island bears coming from mainland at each generation) was found to be appropriate; in case of complete sex-biased gene flow this corresponds to a rate of 0.525x2=1.05% for males and 0% for females. We then simulated 12 independent X-linked loci each of length 6Mb (mimicking the 12 X-linked scaffolds identified as X chromosome), with the demographic parameters scaled for X, and recombination occurring within each locus at rate 10^{-8} per site. The mutation rate per site was set to $5x10^{-9}$, so that the number of ABBA patterns simulated roughly equals the one in real X-linked data. Dependence between sites is expected to increase the variance of the simulated *D* values.

Figure 1.14 shows the distribution of the simulated D statistics for the X chromosome under 5 scenarios of continuous immigration with different strengths of sex-bias (200 independent simulations for each). The blue density line corresponds to a scenario with no sex-bias (the ratio of female migration rate by male migration rate is 1), whereas the red line corresponds to an extreme sex-bias (only males migrate). The observed *D* for X (vertical dotted line) is in the range of the *D* values for simulations under extreme sex-biased gene flow (red line), but not in the range of values for simulations under non sex-biased gene flow or partially sex-biased gene flow (blue, cyan, purple, and orange lines).

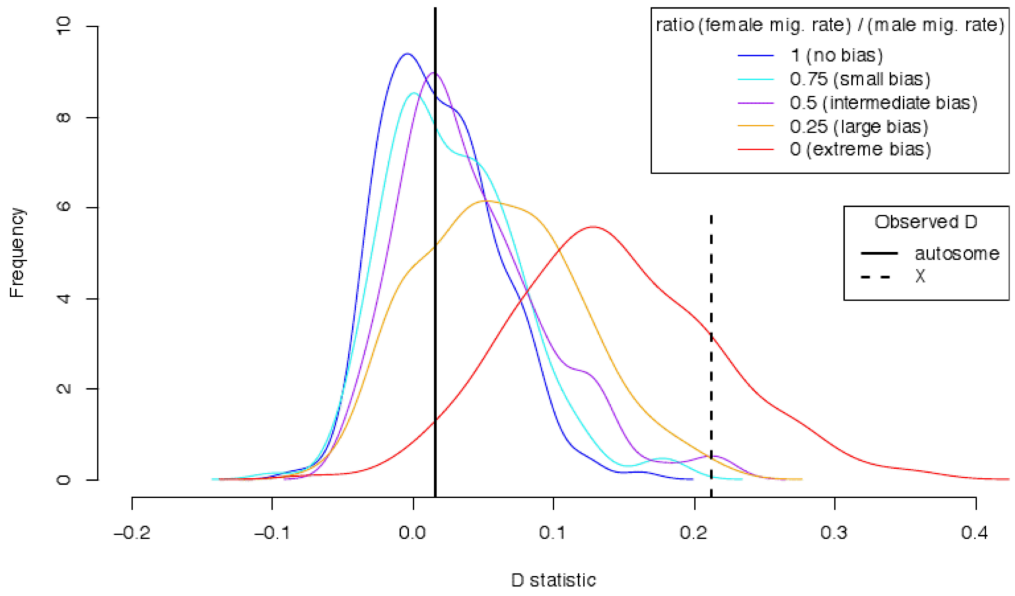


Figure 1.14. Effect of sex biased gene flow on X vs Autosome ratio of D statistics. Distribution of D(ABC, Grizzly, Polar, Panda) calculated from data simulated at 12 independent X-linked scaffolds of length 6 Mb with recombination occurring within each locus at rate of 1×10^{-8} per site. Data were simulated using the same parameters as before, but the strength of the sex-bias varies. The ratio of female migration rate by male migration rate ranges from R=1 (no sex-bias, blue line) to 0 (extreme sex-bias, red line).

Conclusion: The scenario of continuous immigration with an extremely large sex-bias from mainland brown bear to ABC islands is a likely scenario of admixture. Figure 1.14 shows that even an intermediate bias (purple and orange lines) is unlikely to explain the observed D statistic for X.

Parameter	Description	Range	#sampled
t_{GF}	Time of gene flow	[5kya, 50kya]	10
t_{P2}	Divergence time between mainland and ABC brown bears	[60kya, 300kya]	25
t_{P3}	Divergence time between brown and polar bears	[300kya, 1000kya]	71
N_3	Effective population size of polar bears	[1k, 10k]	10
N_{12}	Effective population size of the ancestors of Alaskan mainland and ABC brown bears	[10k, 70k]	7
N_{123}	Effective population size of the ancestors of brown and polar bears	[10k, 70k]	7

Table 1.9. Parameter space for simulations

2.9 An ecological assessment of our model for the ABC Islands and other island habitats.

Our model suggests that the genomic composition of today's ABC Islands bears is the result of a unique combination of ecological, behavioral and climatological circumstances. In contrast to previous models that suggest polar bears originated on the ABC Islands (see section 2.7), our model, proposes that today's ABC Islands brown bears share a common ancestor very recently, dating at the earliest to within the last glacial period. We note that this model is in agreement with other recently published data (Lindqvist *et al.* 2010; Edwards *et al.* 2011; Hailer *et al.* 2012; Miller *et al.* 2012), indicating a recent common ancestor for these bears.

We propose that the ABC Islands population began as a population of polar bears, and, through time, was gradually converted into a population of brown bears via a process of genomic erosion. Continuous input of brown bear DNA, mostly or exclusively from male brown bears, eventually replaced most of the original polar bear genome (Figure 1.15). The ecology and behavior of the two bear species, in particular behavioral differences between males and females, both enabled this process and made it possible to recognize it from the genomic data.

It is notable that the polar bear/brown bear hybrids that are observed in the Canadian Arctic today have exclusively, as far as is known, been born from crosses of polar bear females (or hybrids) and brown bear males (Stirling 2011). Adult male brown bears are known to emerge from their winter dens prior to the spring snowmelt and move onto sea ice to scavenge seals killed by polar bears. Although the peak breeding season for polar bears lasts from April into early May and for brown bears from the end of May through June, male

spermatogenesis in brown bears covers a much wider temporal range, and female polar bears are induced ovulators (Stirling 2011). Therefore, although their breeding periods do not overlap entirely, the production of hybrids is thus possible. Hybrid cubs will stay with their mothers, likely following the adult females onto the ice until weaning. However, they may be more predisposed to make use of terrestrial habitat, and therefore more vulnerable to "stranding" or at least spending more time on terrestrial habitats as they reach adulthood.

While we cannot know the precise process leading to the stranding or isolation of the colonizing population of polar/hybrid bears on the ABC Islands, we propose the following scenario, based on both ecological and genomic data (Figure 1.15): As the ice retreated around the ABC Islands toward the end of the last glacial period, polar bears living on the edges of this ice may have come into contact with male brown bears under a scenario similar to that seen today in the Canadian Arctic. Somehow, a population of polar bear or polar/brown hybrids settled on these islands. Over time, sub-adult male brown bears, the main class of dispersing brown bears individuals (Stirling 2011), dispersed from the mainland to become part of the resident adult population on the ABC Islands (Figure 1.12B). As this happened, the genomes of the bears isolated on the ABC Islands would became more and more brown bear-like, in a process we term *genomic erosion*. This process essentially converts the population that was once polar bears into the population of brown bears observed today. With sufficient time, we expect this process to lead to the complete erosion of the polar bear genome within this population, with the exception of the mitochondrial genome, as no mitochondrial input is received from the dispersing brown bear males.

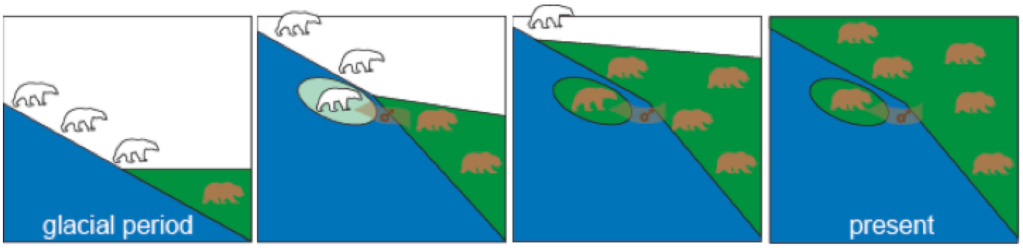


Figure 1.15. Population conversion/genomic erosion model. The salient features of this model are shown schematically. Starting during the last glacial period (left panel), the region is inhabited by polar bears. As the ice retreats and the oceans rise, islands form, cutting off a polar bear or hybrid population from the mainland. Over time, continuous male-dominated or male-exclusive gene flow converts the island population to be of predominantly brown bear ancestry. The remnants of polar bear ancestry are most prevalent in female-associated loci: the mtDNA and X-chromosome.

While we focus here on inferring what may have happened in the ABC Islands, similar processes of hypothetical "stranding" followed by introgression may have occurred in other island or island-like habitats, in particular during periods of climate change as species' ranges change and temporarily overlap. The resulting genetic imprint will depend on several demographic parameters, including the starting population size, whether the immigrant admixers are restricted to one sex, and how long it has been since the process began. Further genomic assessment, for example of the extinct Irish brown bear population, will help to refine these theories and better understand how natural climate change can determine the distribution and genetic diversity of species.

References

Alaska Department of Fish and Game, Brown Bear Species Profile.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science (New York, N.Y.)*, **295**, 2267–70.

Carrara PE, Ager TA, Baichtal JF (2007) Possible refugia in the Alexander Archipelago of southeastern Alaska during the late Wisconsin glaciation. *Canadian Journal of Earth Sciences*, **44**, 229–244.

Chiverrell RC, Thomas GSP (2010) Extent and timing of the Last Glacial Maximum (LGM) in Britain and Ireland: a review. *Journal of Quaternary Science*, **25**, 535–549.

Cronin M a., Amstrup SC, Garner GW, Vyse ER (1991) Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursus). *Canadian Journal of Zoology*, **69**, 2985–2992.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution; international journal of organic evolution*, **62**, 1908–20.

Davison J, Ho SYW, Bray SC *et al.* (2011) Late-Quaternary biogeographic scenarios for the brown bear (Ursus arctos), a wild mammal model species. *Quaternary Science Reviews*, **30**, 418–430.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491–8.

Derrien T, Estellé J, Marco Sola S *et al.* (2012) Fast computation and applications of genome mappability. (CA Ouzounis, Ed,). *PloS one*, **7**, e30377.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Edwards CJ, Suchard MA, Lemey P *et al.* (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology : CB*, **21**, 1251–8.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, **328**, 710–22.

Green RE, Malaspinas A-S, Krause J *et al.* (2008) A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell*, **134**, 416–426.

Hailer F, Kutschera VE, Hallström BM *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science (New York, N.Y.)*, **336**, 344–7.

Harington CR (2008) The evolution of Arctic marine mammals. *Ecological applications : a publication of the Ecological Society of America*, **18**, S23–S40.

Heaton TH, Grady F (2003) The Late Wisconsin Vertebrate History of Prince of Wales Island, Southeast Alaska. In: *Ice Age Cave Faunas of North America* (eds Schubert, B. W., Mead JI, Graham RW), pp. 17–53. Indiana University Press.

Heaton TH, Talbot SL, Shields GF (1996) An Ice Age Refugium for Large Mammals in the Alexander Archipelago, Southeastern Alaska. *Quaternary Research*, **46**, 186–192.

Hofreiter M (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of

genetic variation. *Bioinformatics*, **18**, 337–338.

Ingólfsson Ó, Wiig Ø (2009) Late Pleistocene fossil find in Svalbard: the oldest remains of a polar bear (Ursus maritimus Phipps, 1744) ever discovered. *Polar Research*, **28**, 455–462.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods in molecular biology (Clifton, N.J.)*, **840**, 197–228.

Kowalska Z (1964) Cross-breeding between a female European brown bear (Ursus arctos) and a male polar bear (U. maritimus) in the Logzkim Zoo. *Przeglad Zoologiczny*, **9**.

Krause J, Unger T, Noçon A *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC evolutionary biology*, **8**, 220.

Kurten B (1964) The evolution of the Polar Bear, Ursus maritimus Phipps. *Acta zoologica Fennica*, **108**, 30.

Leonard SA, Risley CL, Turvey ST (2013) Could brown bears (Ursus arctos) have survived in Ireland during the Last Glacial Maximum? *Biology letters*, **9**, 20130281.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **26**, 589–95.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493–6.

Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–7.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–9.

Li B, Zhang G, Willersleve E, Wang J (2011) GigaDB Dataset - DOI 10.5524/100008 - Genomic data from the polar bear (Ursus maritimus).

Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–19.

Lindqvist C, Schuster SC, Sun Y *et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5053–7.

Liu S, Lorenzen ED, Fumagalli M *et al.* (2014) Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, **157**, 785–794.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297–303.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, **2010**, pdb.prot5448.

Miller W, Schuster SC, Welch AJ *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2382–90.

Paetkau D, Amstrup SC, Born EW *et al.* (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, **8**, 1571–1584.

Paetkau D, Shields GF, Strobeck C (1998) Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, **7**, 1283–1292.

Preuß A, Ganslosser U, Purschke G, Magiera U (2009) Bear-hybrids: behaviour and phenotype. *Der Zoologische Garten*, **78**, 204–220.

Reich D, Patterson N, Kircher M *et al.* (2011) Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics*, **89**, 516–528.

Rohland N, Siedel H, Hofreiter M (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular ecology resources*, **10**, 677–83.

She R, Chu JS-C, Wang K, Pei J, Chen N (2008) genBlastA: Enabling BLAST to identify homologous gene sequences. *Genome Research*, **19**, 143–149.

Shields GF, Kocher TD (1991) Phylogenetic Relationships of North American Ursids Based on Analysis of Mitochondrial DNA. *Evolution*, **45**, 218–221.

Slater GJ, Figueirido B, Louis L, Yang P, Van Valkenburgh B (2010) Biomechanical consequences of rapid evolution in the polar bear lineage. *PloS one*, **5**, e13870.

Stirling I (2011) *Polar Bears: The Natural History of a Threatened Species*. Fitzhenry and Whiteside, Brighton, MA.

Talbot SL, Shields GF (1996) Phylogeography of Brown Bears (Ursus arctos) of Alaska and Paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, **5**, 477–494.

Waits LP, Talbot SL, Ward RH, Shields GF (1998) Mitochondrial DNA American for Brown Bear Phylogeography of the and North. *Conservation Biology*, **12**, 408–417.

Yu L, Li Q, Ryder O., Zhang Y (2004) Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Molecular Phylogenetics and Evolution*, **32**, 480–494.

Chapter 2: Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears

James A Cahill, Ian Stirling, Logan Kistler, Rauf Salamzade, Erik Ersmark, Tara L Fulton, Mathias Stiller, Richard E Green, and Beth Shapiro

Originally published in *Molecular Ecology*, 24(6);1205–1217, 2015.

Abstract

Polar bears are an arctic, marine adapted species that is closely related to brown bears. Genome analyses have shown that polar bears are distinct and genetically homogeneous in comparison to brown bears. However, these analyses have also revealed a remarkable episode of polar bear gene flow into the population of brown bears that colonized the Admiralty, Baranof, and Chichagof Islands (ABC Islands) of Alaska. Here, we present an analysis of data from a large panel of polar bear and brown bear genomes that includes brown bears from the ABC Islands, the Alaskan mainland and Europe. Our results provide clear evidence that gene flow between the two species had a geographically wide impact, with polar bear DNA found within the genomes of brown bears living both on the ABC Islands and in the Alaskan mainland. Intriguingly, while brown bear genomes contain up to 8.8% polar bear ancestry, polar bear genomes appear to be devoid of brown bear ancestry, suggesting the presence of a barrier to gene flow in that direction.

Introduction

Polar bears (*Ursus maritimus*) have evolved numerous morphological, behavioral, and physiological specializations for their arctic habitat, including white coat color, a reduced hibernation regime, and a strictly carnivorous diet with corresponding changes in tooth morphology and cranial structure (Sacco & Van Valkenburgh 2004; Slater *et al.* 2010). These adaptations distinguish polar bears from their closely related sister taxon, brown bears (*U. arctos*), who have a far more diverse morphology, ecology and geographic range than do polar bears. Brown bears vary widely in size, coloration, and diet regimes that range from primarily herbivorous to populations that are largely dependent on salmon. The historical range of brown bears includes most of Northern Eurasia and Western North America, while polar bears are found in the continental shelf sea ice

regions of the north Arctic (Figure 2.1).

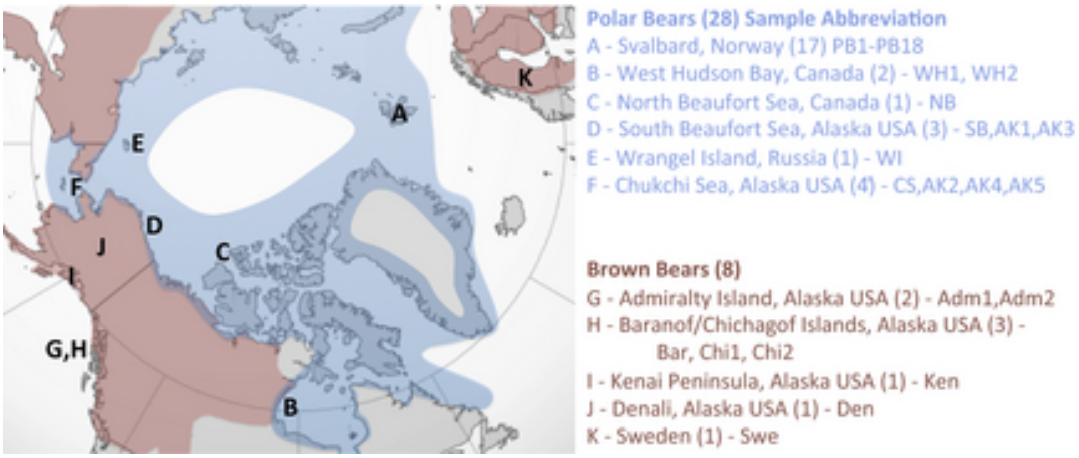


Figure 2.1 – Sample Map. Map of the present-day geographic range of brown bears (red) and polar bears (blue). Letters indicate location from which bears were sampled.

Despite the substantial morphological, behavioral and ecological differences between brown bears and polar bears, the two species share a very close evolutionary relationship. Precisely how close, however, remains a subject of substantial ongoing debate. At many loci, lineages have not yet sorted between the two species (Hailer *et al.* 2012, 2013; Cahill *et al.* 2013), which complicates estimates of when the two species diverged. In addition, polar bears and brown bears produce viable and fertile hybrids both in the wild and in captivity (Preuß *et al.* 2009; Stirling 2011), suggesting a recent divergence between the two lineages.

Published estimates of the time of divergence between brown bears and polar bears using genetic data range from 340 thousand to 4-5 million years ago (Hailer *et al.* 2012; Miller *et al.* 2012; Cahill *et al.* 2013; Cronin *et al.* 2014; Liu *et al.* 2014). Not all of these estimates are directly comparable, however. Divergence estimates based on the molecular clock assume limited or no gene flow and range from ~600 thousand to 3 million years ago, depending largely on how the molecular clock is calibrated (Hailer *et al.* 2012; Cahill *et al.* 2013; Cronin *et al.* 2014). Importantly, these estimates are genomic divergence times and not

population divergence times, and will therefore predate the origin of polar bears as a distinct lineage.

Two studies Miller et al 2012 and Liu et al 2014, have also attempted to estimate the time of population divergence from whole genome data using population-modeling frameworks. Liu et al estimated population divergence to have occurred 343-479 thousand years ago with post-divergence gene flow from polar bears to brown bears (Liu *et al.* 2014). Miller et al estimated a much earlier 4-5 million years ago divergence followed by bidirectional post-divergence gene flow (Miller *et al.* 2012). Given the greater concordance of the 343-479 thousand years ago population divergence with the genetic divergence estimates it appears to be a more plausible framework.

Analyses of mitochondrial DNA further complicated interpretation of polar bear and brown bear history. Most brown bear mitochondrial haplotypes show strikingly strong geographic structure, likely resulting from female philopatry (Korsten *et al.* 2009; Davison *et al.* 2011; Edwards *et al.* 2011). However, the first genetic studies of bear mitochondrial DNA identified a strange exception to this rule: polar bear mitochondrial haplotypes fall within the range of diversity of brown bear mitochondrial haplotypes, rather than outside of brown bear mitochondrial diversity, as would be expected of separate species (Cronin *et al.* 1991; Talbot & Shields 1996; Waits *et al.* 1998; Lindqvist *et al.* 2010). The brown bear mitochondrial lineage that is most closely related to those of polar bears is found today only on Alaska's ABC Islands (Figure 2.1, Locations G and H). In fact, these brown bear mitochondria are more similar to the mitochondria of polar bears than they are to other brown bear mitochondrial lineages. This finding led to early speculation that the ABC Islands population was a very ancient population of brown bears and therefore the most closely related population to polar bears (Talbot et al 1996).

Later surveys of bear mitochondrial DNA included geographically diverse samples of both living brown bears and extinct populations (Leonard *et al.* 2000; Barnes *et al.* 2002; Valdiosera *et al.* 2007; Edwards *et al.* 2011). These analyses further complicated the scenario by revealing three geographically and temporally distinct brown bear populations that had mitochondrial haplotypes that were very similar to those of polar bears (hereafter the brown/polar mtDNA clade). In addition to the ABC Islands population, this includes an extinct population that lived on present-day Ireland until around 9,000 years ago (Edwards *et al.* 2011) and another that lived in Pleistocene Beringia—the once-contiguous landmass that connected Alaska to northeastern Siberia—more than 50,000 years ago (Barnes *et al.* 2002). These findings led to further speculation that the geographic distribution of mitochondrial haplotypes in the brown/polar mitochondrial clade was due not to long-term evolution, but instead to multiple instances of hybridization, during which the mitochondrial lineage was passed between the two species, eventually leading to the geographic pattern of the present day (Edwards *et al*. 2011).

More recently, analyses of nuclear genomic and Y-chromosome data have provided additional insights into the evolutionary relationship between brown and polar bears. Consensus nuclear DNA phylogenies (Hailer *et al.* 2012; Cahill *et al.* 2013; Cronin *et al.* 2014; Bidon *et al.* 2014) and Y-chromosome phylogenies (Bidon *et al.* 2014) indicate that, on average, polar bears and brown bears form two distinct lineages. However, many nuclear loci and the mitochondria deviate from this pattern (Hailer *et al.* 2012, 2013; Cahill *et al.* 2013). This deviation from the average species tree topology is most likely due to the effects of incomplete lineage sorting and, possibly, admixture.

Nuclear genomic data also reveal a remarkable difference in the amount of diversity within the two bear lineages. Mirroring their respective levels of

ecological and biological diversity, polar bears are much more genetically homogenous than are brown bears (Miller *et al.* 2012; Cahill *et al.* 2013). Two polar bear individuals differ at only about 0.03% of sites, while two brown bears from Alaska differed at 0.16% of sites (Cahill *et al.* 2013). On average, a polar bear differs from a brown bear at roughly 0.24% of sites in the genome (Cahill *et al.* 2013).

We recently proposed that the ABC islands brown bear population descends from an admixture event with polar bears (Cahill *et al.* 2013). We observed that ABC Islands brown bears show evidence of increased polar bear ancestry relative to other brown bear populations throughout their autosomal genomes. However, polar bear ancestry is further elevated on ABC Island brown bear X chromosomes compared to their autosomes (Cahill *et al.* 2013), and they contain mitochondria that are more similar to those of polar bears than to other brown bears.

These genetic observations and the natural history of the ABC Islands are consistent with a model wherein ABC Islands brown bears are the descendants of an original population of polar bears. Immigration of primarily or exclusively male brown bears gradually converted the phenotype and genotype of the ABC islands bears into those of brown bears. This male-biased gene flow did not convert the strictly maternal mitochondrial DNA and was less pronounced in converting X chromosome loci, but completely converted the paternal Y chromosome. This result overturns the previous hypothesis that polar bears received their mitochondrial haplotype via introgression from a population of brown bears closely related to the ABC Island brown bears or ancient Irish brown bears (Edwards *et al*. 2011). Rather, the mitochondrial haplotype found in polar bears is of polar bear origin. Under this scenario, brown bears in the brown/polar mtDNA clade are the recipients of introgression from polar bears.

Previously, we measured the impact of admixture on the nuclear genomes of ABC bears using the *D*-statistic test for admixture (Green *et al.* 2010; Durand *et al.* 2011). We showed that a brown bear from the ABC Islands carried more polar-bear matching alleles than a brown bear from Denali National Park. While the result was statistically well supported, this framework—using a single, non-ABC Islands brown bear for comparison—lacks the power to assess whether the mainland brown bear is devoid of polar bear ancestry or simply has less polar bear ancestry than does the ABC Islands bear. We were therefore unable to explore the geographic extent to which this admixture event affected bear populations or to explore hypotheses about frequency of admixture between the two bear lineages. Recently a study calculated *D*-statistics suggesting that other ABC islands brown bears and a brown bear from Glacier National Park in Montana, USA, possessed polar bear ancestry (Liu *et al.* 2014).

Here, we test the extent of polar bear ancestry within and beyond the ABC islands population by analyzing genome-wide data from a more diverse panel of brown bears (Figure 2.1). Comparisons between bears in this larger panel reveal a more widespread pattern of polar bear admixture into brown bears. We find evidence of polar bear admixture in ABC Islands brown bears and in brown bears from the Alaskan mainland. Within the ABC Islands, we identify a geographic cline of admixture, with more retained polar bear ancestry in bears further from the mainland. Finally, we find no evidence of gene flow from brown bears into polar bears, in stark contrast to the widespread signal of gene flow in the other direction.

Methods

Assembling a large panel of brown bear and polar bear genome sequences

Previous analysis indicated that the genomes of ABC Islands brown bears contain more polar bear ancestry than do those of mainland Alaskan brown

bears. However, this analysis was performed using only a single genome representing each population. To more fully explore the spatial distribution of the signal of polar bear ancestry within the ABC Islands, we collected a larger panel of brown and polar bear genomes. We sequenced 3 previously unpublished brown bears one from Sweden and two from Chichagof island. We analyzed these samples along with samples from two previously published data sets; two brown bears, seven polar bears and an American black bear published in Cahill et al 2013 (Cahill *et al.* 2013) and three brown bears and twenty-three polar bears published in Miller et al 2012 (Miller *et al.* 2012) (Figure 2.1, Table 2.3).

For many bear genomes, the depth of coverage was insufficient to call heterozygote sites reliably. We therefore used the strategy described previously (Green *et al.* 2010; Cahill *et al.* 2013) to sample one high quality base at each genomic position from each bear, thereby creating a pseudo-haploid sequence. Two of the polar bear samples were unsuitable for this approach and we excluded these from further analysis (Supplemental Materials, Figure 2.5, Figure 2.6). As described previously, we partitioned the polar bear genome assembly (Li *et al.* 2011) to which all sequence data were mapped into scaffolds that are likely to be on autosomes and those likely to be X-chromosomes (Cahill *et al.* 2013). No Y chromosome scaffolds were found to meet our minimum scaffold length filtering criteria (see below). To infer the history of the Y chromosome, we therefore assessed separately only the largest Y chromosome scaffold (Supporting Information).

DNA extraction, library preparation, and sequencing

We extracted DNA from the Chichagof Island brown bears and the Swedish brown bear using the DNeasy Blood & Tissue Kit (Qiagen) and the QIAmp Micro Kit (Qiagen) according to the manufacturer's specifications. We physically sheared the DNA of the three new brown bear samples using a

Diagenode Bioruptor NGS instrument. Extracts were transferred into 1.5 ml tubes and exposed to seven rounds of sonication, using the energy setting "HIGH" and an "ON/OFF interval" of 30/30 seconds. We then purified and concentrated the extracts using the Agencourt AMPure XP PCR purification kit, according to manufacturer's instructions, and eluted in 20μl of 1xTE, with 0.05% Tween20.

We prepared indexed Illumina libraries using 15μl of each extract following the protocol described by Meyer & Kircher (Meyer & Kircher 2010), with reaction volumes scaled to total volume of 40μl. To verify final DNA concentration and the distribution of insert sizes, we ran each library on an Agilent 2100 Bioanalyzer. We then sequenced each bear on half of a lane of an Illumina HiSeq 2000 instrument using 150 base-pair (bp) paired-end chemistry at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Quality control, mapping, and pseudo-haploidization

From the Illumina sequence data, we removed the index and adapter sequence and merged paired reads using a script provided by M. Kircher (Kircher 2012). We then trimmed each read to remove low quality bases by trimming inward from the 3'-end of the read until detecting a base with quality score ≥13 (~95% confidence) (Lohse *et al.* 2012). We mapped the resulting data to the draft polar bear genome (Li *et al.* 2011) using BWA (Li & Durbin 2010). We removed duplicated reads created by PCR amplification using the rmdup program from samtools (Li *et al.* 2009).

For all individuals at each position in the genome, we chose a random, single allele from amongst reads that passed the following filtering criteria: (i) read map quality Phred-30 or greater; (ii) Illumina base quality Phred-30 or greater; (iii) genomic position is within a uniquely mappable 35-mers identified using GEM (Derrien *et al.* 2012); (iv) read coverage at that position is between the 5th and 95th percentiles genome wide identified using BEDTools coverageBed

(Quinlan & Hall 2010); and (v) discarding scaffolds <1MB in length. This approach is designed to have uniform power to detect rare alleles at all mappable positions in all individuals. In contrast, using a genotyping inference program to call heterozygous sites would have greater power to detect rare, non-reference alleles in higher coverage individuals. This would, however, confound downstream analysis. The result of our approach is that a single base call, randomly selected from among the mapped reads, represents the individual at every site in the reference genome. Because bears are diploid, this single base call necessarily only represents one of the two alleles at sites where the bear in heterozygous.

Detecting admixture

We used the *D*-statistic framework (Green *et al.* 2010; Durand *et al.* 2011) to measure the relative amounts of polar bear ancestry between pairs of brown bears. The *D*-statistic is a comparison between four individuals: two conspecific individuals, P1 and P2, a candidate introgressor, P3, and an outgroup, O. At each site in the genome, we test whether the relationship between these four individuals is inconsistent with the species tree topology. Sites that are considered inconsistent with the species tree are those at which P2 shares a derived allele with P3 but not P1 (ABBA sites) or sites where P1 shares a derived allele with P3 but not P2 (BABA sites). In the absence of ancestral population structure, processes other than admixture that produce loci inconsistent with the species tree, such as incomplete lineage sorting and error, are expected to produce an equal number of ABBA and BABA sites (Green *et al.* 2010; Durand *et al.* 2011). An excess of either ABBA or BABA sites is evidence of admixture. In this framework, admixture between P1 and P3, for example, is expected to produce an excess of BABA sites compared to ABBA sites. As described previously, we used the black bear genome to determine the ancestral state at

each polymorphic genomic position (Cahill *et al.* 2013).

We performed analyses for all combinations of pairs of conspecific individuals and candidate admixers. This amounted to 720 comparisons for 8 brown bears with 28 candidate introgressor polar bears (Table 2.1; Figure 2.2; Table 2.4), and 3360 comparisons for 28 polar bears with 8 candidate introgressor brown bears (Table 2.5). For separate analysis of the X chromosome, we used the X chromosome polar bear genome scaffolds identified in our previous study (Cahill *et al.* 2013). To be classified as an X chromosome scaffold a scaffold must meet two criteria: differences in male versus female shotgun sequence coverage and the presence of orthologs to X-linked genes from the dog genome (Lindblad-Toh *et al.* 2005).

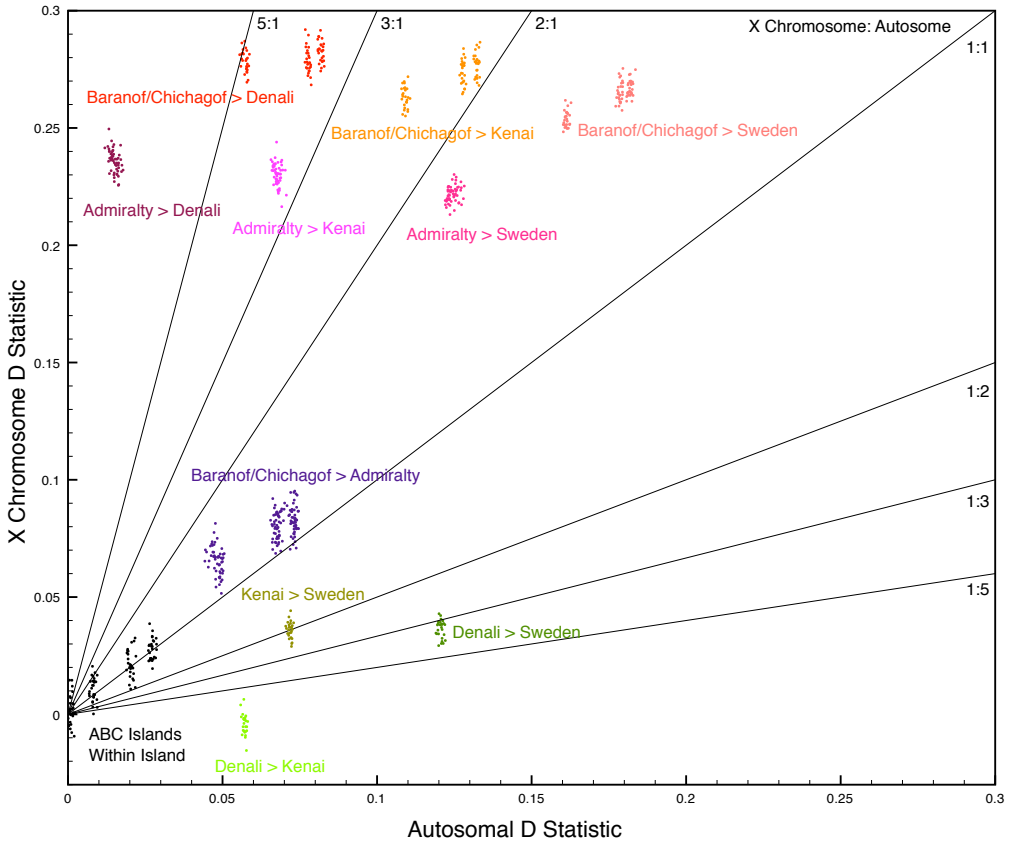


Figure 2.2 - *D*-statistic measure of admixture in brown bears. Distribution of *D*-statistic tests between two brown bears and a polar bear candidate introgressor with an American black bear outgroup. Each dot represents an independent test with a different polar bear as the candidate introgressor. ABC Islands bears,

particularly those from Baranof and Chichagof Islands, show the highest amount of polar bear introgression. Admiralty Island brown bears show the greatest bias toward polar bear ancestry on the X chromosome versus the autosomes. The Denali brown bear shows the greatest bias toward polar bear ancestry on the autosomes relative to the X chromosome.

P1	P2			% Polar Bear
	Sweden	Kenai	Denali	
Adm1	0.1258 (12.8)	0.0685 (5.9)	0.0160 (1.3)	5.99 (12.2)
Adm2	0.1231 (12.2)	0.0669 (6.1)	0.0139 (1.1)	5.88 (11.7)
Bar	0.1613 (14.7)	0.1091 (8.9)	0.0573 (4.3)	7.82 (13.9)
Chi1	0.1786 (17.7)	0.1278 (11.3)	0.0777 (6.4)	8.68 (16.0)
Chi2	0.1819 (18.3)	0.1323 (12.1)	0.0819 (6.7)	8.83 (16.5)
Den	0.1267 (14.3)	0.0571 (5.6)	N/A	5.38 (12.7)
Ken	0.0719 (9.6)	N/A	-0.0571 (5.6)	3.17 (9.2)

Table 2.1 - Polar bear ancestry in brown bear autosomes. Average autosomal *D*-statistic values reflecting the amount of polar bear ancestry in each brown bear (P1) that results from tests in which the Swedish, Kenai, or Denali brown bears (P2) are used as the polar bear-free baseline. For each *D*-statistic reported, the corresponding *Z* score (Green *et al.* 2010; Durand *et al.* 2011), estimated using a weighted block jackknife approach with 5MB blocks (Green *et al.* 2010; Cahill *et al.* 2013, Materials and Methods), is indicated in parentheses. The final column shows the average proportion of polar bear ancestry in each brown bear autosomal genome ($\hat{f}$ estimator) and corresponding *Z* score. A summary of all *D*-statistic comparisons performed in this study is provided in Table S2.

The related $\hat{f}$ estimator (Green *et al.* 2010; Durand *et al.* 2011) was used to estimate the proportion of the genome derived from admixture. Ideally, this test requires two individuals of the candidate introgressor species that are not themselves admixed. This was not possible in all cases, however, because all of the brown bears except the Swedish brown bear were found to be admixed with polar bears. To minimize bias, we used the two least admixed brown bears—the Swedish and Kenai individuals—to estimate the fraction of brown bear genomes that had introgressed from polar bears (Supporting Table S5).

For both the *D* and $\hat{f}$ -statistics, we measure *D*-statistical significance using a weighted block jackknife using 5Mb blocks (Green *et al.* 2010; Durand *et al.* 2011). The weighted block jackknife tests whether admixture signals are uniform across the genome and therefore reflect the same population history.

The weighted block jackknife produces a standard error value. The number of standard errors by which the observed value of *D* or $\hat{f}$ differs from the null expectation of zero is the *Z* score. In keeping with Green et al 2010 (Green *et al.* 2010), we define significant *D* and $\hat{f}$ -statistic results as having *Z* scores greater than three. Note that we do not perform multiple-test correction, as it is not clear how to correctly account for multiple tests in this exploratory framework in which most tests are not independent.

Results

Admixture analysis

D-statistic comparisons between pairs of brown bears from different populations revealed statistically significant differences in all comparisons (Figure 2.2; Tables 2.1, 2.2 and 2.4). Consistent with previous observations using only a single mainland brown bear, comparison between any ABC Islands bear and any non-ABC Islands brown bear showed an excess of polar bear ancestry in the ABC Islands bear. Also as before, the excess of polar bear ancestry in ABC Islands bears was greater on the X chromosomes than on the autosomes (Figure 2.2, Table 2.1). The lower statistical significance of X chromosome results compared to autosomal results (Tables 2.1, 2.4, 2.5, 2.6, 2.7) is due to the much smaller size of the X chromosome and corresponding increased influence of removing 5MB blocks.

P1	**P2**			**% Polar Bear**
	Sweden	**Kenai**	**Denali**	
Adm1	0.2226 (1.9)	0.2285 (2.7)	0.2330 (4.5)	7.63 (1.8)
Adm2	0.221 (1.8)	0.2323 (2.9)	0.2388 (3.2)	7.53 (1.7)
Bar	0.2538 (1.8)	0.2632 (2.6)	0.2785 (2.7)	9.35 (1.5)
Chi1	0.2654 (2.2)	0.2736 (3.3)	0.2787 (4.0)	9.59 (1.9)
Chi2	0.2669 (2.6)	0.2769 (4.1)	0.2826 (3.9)	9.71 (2.3)
Den	0.0364 (0.4)	-0.0041 (0.1)	N/A	1.04 (0.3)
Ken	0.0360 (0.7)	N/A	0.0041 (0.1)	1.04 (0.7)

Table 2.2 - Polar bear ancestry in brown bear X chromosomes. Average X-chromosome *D*-statistic values reflecting the amount of polar bear ancestry in each brown bear (P1) that results from tests in which the Swedish, Kenai, or Denali brown bears (P2) are used as the polar bear-free baseline. For each *D*-statistic reported, the corresponding *Z* score is reported as in Table 2.1. The final column shows the average proportion of polar bear ancestry in each brown bear X-chromosome ($\hat{f}$ -estimator) and corresponding *Z* score.

The brown bear from Sweden had the lowest rate of matching polar bear alleles in all pairwise comparisons with other brown bears. The Kenai brown bear had the next lowest rate, with $\hat{f}$ -estimated polar bear ancestry of 3.17% on the autosomes and 1.04% on the X chromosome. The Denali brown bear, which was the only mainland brown bear sample from our previous report, had the highest rate of polar bear allele matching amongst all non-ABC Islands brown bears in this larger sample of brown bears (Figure 2.2). Using f, we estimate polar bear ancestry in the Denali bear to be at least 5.38% on the autosomes and 1.04% on the X chromosome.

D-statistic measurements of polar bear ancestry on the X-chromosome versus autosomes reveal a striking difference between ABC Islands bears and non-ABC Islands bears. The pattern of *enriched* polar bear ancestry on the X chromosome of ABC Islands brown bears is *reversed* in non-ABC Islands brown bears (Figure 2.2, Table 2.1). That is, non-ABC Island brown bears have less polar bear ancestry on the X chromosome relative to their autosomes.

Within the ABC islands the brown bears from Baranof and Chichagof islands have more polar bear ancestry than the brown bears from Admiralty island (Figure 2.2, Tables 2.1, 2.2). $\hat{f}$ -estimation of polar bear ancestry in ABC islands brown bears ranges from 5.9% to 8.8% for the autosomes and 7.5 to 9.7% of the X chromosome. As noted above, all of the ABC islands brown bears' X chromosomes are enriched for polar bear ancestry compared to their autosomes. Within the ABC islands, X chromosome bias of polar bear ancestry increases as

total polar bear ancestry decreases. The brown bears of Admiralty island, the island closest to the mainland, have the least polar bear ancestry and polar bear ancestry is the most X Chromosome biased (Tables 2.1, 2.2).

We also tested for a signal of admixture within polar bear genomes. *D*-statistic tests between pairs of polar bear genomes for unequal rates of matching derived brown bear alleles resulted in no *D*-statistics that differed statistically from zero (weighted block jackknife $Z>3$) (Figure 2.3, 2.5; Table 2.5). $\hat{f}$ -estimators also indicated an absence of detectable gene flow from brown bears into polar bears, with no comparisons deviating statistically from zero ($\hat{f}$ -statistics with weighted block jackknife $Z>3$).

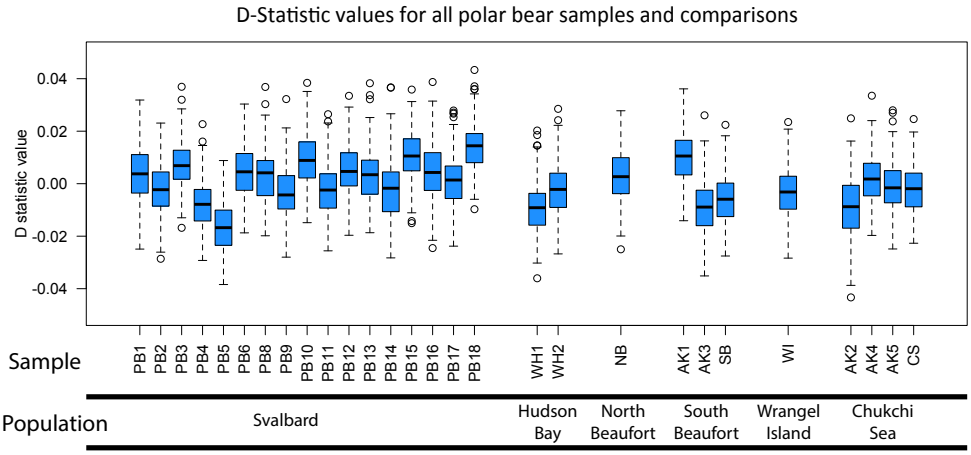


Figure 2.3 - *D*-statistic measure of admixture in polar bears. Box-and-whisker plots showing the range of *D*-statistic values for a single polar bear (Sample), arranged along the x-axis by geographic location (Population), compared to every other polar bear with every brown bear as a candidate introgressor. For each box-and-whisker-plot, boxes range from the 25th-75th percentiles, whiskers are 1.5 times the distance from the 25th to 75th percentile, or the most extreme result if it is less than 1.5 times the distance from the 25th to 75th percentile. Circles indicate data that fall outside of 25th to 75th percentile (outliners). Statistically significant *D*-statistic values indicate that the subject polar bear shares an excess of derived alleles with brown bears. None of the comparisons, including the outliners, resulted in *D*-statistic values that differed significantly from zero ($Z > 3$).

Discussion

Uneven amounts of polar bear ancestry among brown bears

Our previous observations about polar bear ancestry within ABC Islands bears relied on comparison to a single mainland brown bear from Denali National Park, Alaska. While the comparisons were consistently and strongly in the direction of excess polar bear ancestry in ABC Islands bears, use of a single comparison genome is limiting in an important way. Quantifying the absolute amount of polar bear ancestry requires making an assumption about the amount of polar bear ancestry in the comparison brown bear. For our previous work, we made the assumption that the Denali bear was free of polar bear ancestry. Following this assumption, the excess polar bear ancestry in ABC Islands brown bears was interpreted as a measure of the absolute amount of polar bear ancestry in the ABC Islands brown bears.

Our new results indicate that this assumption was incorrect – the Denali brown bear is *not* free of polar bear ancestry. In fact, among the non-ABC Islands brown bears analyzed here, the Denali bear has the *greatest* polar bear ancestry: at least 5.38% of the autosomes and 1.04% of the X chromosome derives from polar bear. The Swedish brown bear has the least polar bear ancestry in all pairwise comparisons and thus establishes a new baseline for admixture-free brown bear.

A further unexpected result is that in contrast to the excess of polar bear ancestry on the X chromosomes relative to autosomes among ABC Islands bears, we see the opposite pattern within non-ABC Island brown bears. That is, these bears have lower levels of polar bear ancestry within their X chromosomes versus their autosomes (Figure 2.2).

These results have several important implications for understanding the admixture event on the ABC Islands. Most importantly, we now estimate a much greater amount of polar bear ancestry in each ABC Islands brown bear than

previously reported. Re-calculating the absolute amount of polar bear ancestry for the five ABC Islands bears using the Swedish bear as the non-admixed standard results in higher estimated proportions of polar bear ancestry than when using the Denali bear (Table 2.1). Whereas we previously reported an absolute amount of 6.5% polar bear ancestry on the X-chromosome and 0.5% polar bear ancestry on the autosomes of the bear from Admiralty Island, estimates based on the Swedish bear indicate that 7.6% of the X-chromosomes of this bear are derived from polar bear ancestry, as is 6.0% of the autosomes. These results may explain the very high X: autosome ratio of polar bear ancestry that was estimated while using the Denali brown bear as standard, which fell outside of the distribution of ratios predicted by demographic simulation (Cahill *et al.* 2013). Because the Denali bear has a significant amount of polar bear ancestry on the autosomes, this led to an underestimate of the amount of polar bear ancestry on the autosomes of the ABC Islands brown bear.

We observe a geographic signal wherein bears from the islands less accessible from the mainland—Baranof and Chichagof Islands—have more polar bear ancestry than the bears of Admiralty Island (Figure 2.2). A previous analysis of microsatellite data from a large sample of Alaskan brown bears similarly reported more gene flow between Admiralty Island and the Alaskan mainland than between the more distant Baranof and Chichagof Islands and the Alaskan mainland, and very little gene flow between Baranof and Chichagof Islands and Admiralty Island (Waits *et al.* 1998; Paetkau *et al.* 1999). These data support a model of brown bear dispersal from the Alaskan mainland that is limited mainly by long-distance water crossings (Figure 2.2).

Notably, Baranof and Chichagof Islands brown bears' polar bear ancestry is less X chromosome biased than Admiralty Island bears' polar bear ancestry (Figure 2.2). This result is consistent with and extends the model we proposed

previously wherein male-dominated gene flow from the mainland onto these islands carries brown bear genetic material into a population that was initially polar bears. More distal bears are further from the source of this gene flow and thus less impacted by migration from the mainland. In this way, the brown bears of Baranof and Chichagof Islands retain more of their polar bear genetic ancestry, but exhibit comparatively less bias toward the X chromosome. Lending further support to this hypothesis of female biased polar bear ancestry and male biased brown bear ancestry, a recent study by Bidon and colleagues found Y chromosome ancestry on the ABC islands to be exclusively brown bear (Bidon *et al.* 2014). Similarly, we find no evidence of polar bear ancestry in any brown bears when analyzing the largest Y-chromosome scaffold in the polar bear assembly (Supporting Information).

Polar bear ancestry in non-ABC islands brown bears

For each of the non-ABC Islands Alaskan brown bears we analyzed, excess polar bear ancestry is observed on the autosomes over the X-chromosome when compared to the Swedish brown bear. The opposite is observed in the ABC islands brown bears. This X chromosome *depletion* is more pronounced than the X chromosome enrichment on the ABC islands. There are multiple plausible hypotheses that could explain this result, both demographic and selective.

Demographically, brown bear dispersal is primarily male-mediated (McLellan & Reiner 1994; Støen *et al.* 2006). Populations that are located further from the site of hybridization would be expected, therefore, to have less polar bear ancestry on the female-biased X chromosome than on the autosomes. This behavioral process is supported by genetic evidence: while brown bear mitochondrial haplotypes, which are exclusively maternally-inherited, show strong geographic structuring (Korsten *et al.* 2009; Davison *et al.* 2011; Edwards *et al.* 2011), Y chromosome haplotypes, which are exclusively paternally-inherited,

show no such geographic structure (Bidon *et al.* 2014). Thus, one hypothesis that is consistent with our data is that brown bears from the ABC Islands, and perhaps other regions of polar bear admixture, will have their polar bear ancestry dispersed primarily by male brown bears. As males carry fewer X chromosomes than autosomes, polar bear ancestry will become increasingly less visible on the X chromosome than on autosomes as one samples brown bears farther from the site of admixture.

From a selective standpoint, it has been suggested that loci involved in hybrid incompatibility are overrepresented on the X chromosome (Masly & Presgraves 2007). This is because, in the heterogametic sex, the presence of only one copy of any incompatible allele prevents a homologous compatible allele from masking the incompatibility. In theory, this should lead to a reduction in introgressed ancestry on the X chromosome relative to the rest of the genome. Such an effect was recently observed in the case of Neandertal introgression into non-African humans, where Neandertal ancestry is almost absent on the X chromosome (Sankararaman *et al.* 2014). In that case, the authors' simulations appeared to reject a demographic explanation.

These processes are not mutually exclusive, and both biased dispersal and selection against polar bear ancestry on the X chromosome may play a role in explaining the lower polar bear ancestry on the X chromosome of non-ABC islands brown bears. At this stage, it seems likely that there is insufficient understanding of the demography of bears throughout Alaska and Northern Canada to make a definitive assessment of the role of each.

Our results showing excess polar bear ancestry in every brown bear we sampled compared with the Swedish brown bear, whose own polar bear ancestry we cannot reliably estimate for the reasons noted above, suggest a higher rate of introgression of polar bear DNA into brown bear genomes than previously

calculated. It may be that the epicenter of this introgression is in the ABC Islands and the surrounding area. Following the initial introgression event, predominantly male migration could have then carried polar bear ancestry away from the Islands. This model is simple and not obviously contradictory to the data. However, further study will be required to refine the number, timing, and geographic locations of polar bear admixture into Alaskan brown bears. One particularly intriguing possibility concerns the other brown bears that were found with a mitochondrial haplotype similar to the polar bear haplotype: some Beringian brown bears that lived more than 50 thousand years ago (Barnes *et al.* 2002) and a population of now extinct Irish brown bears (Edwards *et al.* 2011). It will be interesting to know if these results are due to a similar process of male-mediated gene flow from brown bears into other polar bear populations in the past and, if so, if there is any remaining polar bear ancestry from those introgression events in brown bear populations today.

Absence of detectable brown bear ancestry in polar bears

Admixture is more easily detected by the *D*-statistic when the population receiving gene flow is small (Durand *et al.* 2011). Given that the effective population size of polar bears is and likely has been small for many thousands of years (Miller *et al.* 2012), our observation that none of the 28 polar bear genomes used in this analysis have detectable brown bear ancestry is even more striking, and has important implications for understanding both the relationship between the two species and for predicting the long-term consequences of hybridization.

Despite the widespread impact of admixture on brown bear genomes, the genetic data indicate no corresponding effect on polar bears. The absence of detectable gene flow into polar bears may therefore reflect an ecological barrier to admixed individuals surviving as polar bears, where any introduction of brown bear DNA into polar bears may be strongly deleterious (Schluter 2009). Within

the polar bear lineage, there is evidence of powerful episodes of positive selection (Liu *et al.* 2014). One possibility is that the extremely specialized adaptations of polar bears may quickly place phenotypically intermediate hybrids at more of a disadvantage in the polar bear environment than the brown bear environment.

One simple example of this could be coat color - a trait that would likely play a more severe role in decreasing fitness of F1 hybrids in the polar bear habitat than the brown bear habitat. Like other arctic predators hunting on snow or ice, such as arctic wolves (*Canis lupus arctos*) or arctic foxes (*Vulpes lagopus*) in winter, polar bears are uniformly white except for their eyes and nose. In contrast, hybrids have darker patches and sometimes overall coloration (Gray 1971; Stirling 2011). When a polar bear stalks a seal on the ice, it holds the head down low and walks in a straight line toward the intended prey (Stirling 1974), presumably because that minimizes contrasting dark spots that a seal may notice. An important time for polar bear feeding is late spring when the new crop of young ringed seals, with little experience with predators, is weaned (Stirling & Øristland 1995) and later in the spring, as the snow melts that covers breathing holes and birth lairs when a high proportion of seals are out on the ice basking and molting (Kelly & Quakenbush 1990). Much of the hunting of these seals is done by stalking (Stirling 1974). Hybrid bears with patches or darker pelage, or darker shades would be more visible to the seals and therefore less successful hunters. In contrast, variation in coloration may be less of a constraint for a hybrid bear feeding on brown bear food sources; vegetation, salmon, or carrion. While this model - extreme reduction in F1 hybrid fitness in the polar bear ecological environment - is speculative and simplistic, it would suffice to explain the striking absence of brown bear genetic introgression into polar bear populations.

Another possible explanation for the observed imbalance in admixture proportions may be that brown bear DNA did introgress into polar bears, but that all polar bears have equivalent levels of brown bear ancestry. The *D*-statistic, which is a pairwise comparison method, has no power to detect admixture in this unique scenario. Such a scenario could manifest in two ways. First, all of the polar bears sampled here may have received an exactly equal amount of brown bear ancestry via introgression more recently than the time of the polar bear populations shared common ancestor. This scenario seems unlikely due to the size and geographic diversity of the panel of polar bears analyzed here. Widespread brown bear into polar bear introgression might be expected to result in at least some variation in the amount of introgressed brown bear ancestry in one of these bears. Second, brown bear DNA may have introgressed into the polar bear population that was ancestral to all extant populations of polar bears. In this second scenario, all polar bears would have exactly the same brown bear ancestry, thereby masking the signal of admixture. Under this scenario, no living polar bear would have detectable excess brown bear ancestry.

One way to investigate this second scenario is to examine the number of *D*-statistic informative sites. *D*-statistic informative sites are incongruous with the species tree and can arise from a variety of processes including incomplete lineage sorting, sequencing errors, multiple mutations at a site, and admixture. If two conspecific individuals, P1 and P2, received the same amount of introgression from P3, then the *D*-statistic for *D*(P1,P2,P3,O) would be zero. However, the number of species tree incongruous sites would be greater than if no introgression had occurred. In addition to *D*-statistics of zero, we observe very few sites that are incongruous with the species tree (Figure 2.4) when comparing two polar bears for brown bear matching alleles. This suggests that the second

scenario—brown bear admixture into the ancestral population of polar bears—is also unlikely.

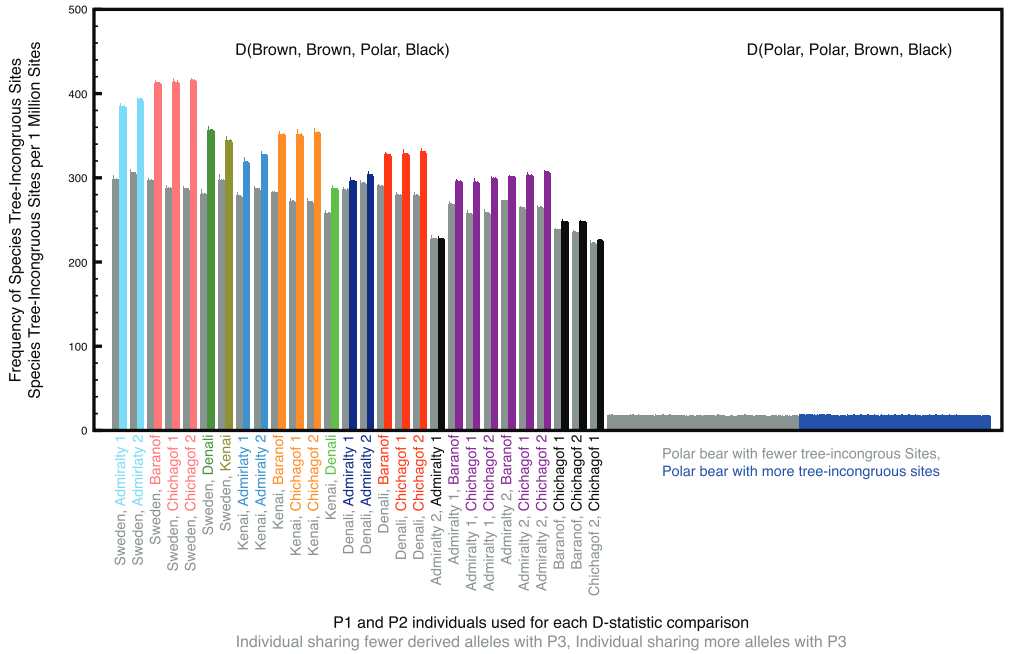


Figure 2.4 – Frequency of sites informative to the *D*-statistic. The frequency of ABBA sites (grey bars) and BABA sites (colored bars) for each *D*-statistic comparison. Both ABBA and BABA sites are considered species tree incongruent sites. Processes other than admixture, such as incomplete lineage sorting and sequencing error, are expected to produce an equal number of ABBA and BABA sites. Any difference between the number of ABBA and BABA sites—here, the difference between colored and grey bars—is interpreted as evidence of admixture. Comparisons involving pairs of polar bears show very few tree-incongruent sites and no evidence of admixture from brown bears.

If brown bear introgression were very ancient and took place prior to the most recent common ancestor of polar bears, then it would be impossible to detect this admixture event as all polar bears would carry these introgressed and fixed alleles. However, estimates of the timing of both the genetic time to most recent common ancestor between polar bears (130-650 thousand years ago; Cahill *et al.* 2013) and speciation between polar bears and brown bears (343-479 thousand years ago; Liu *et al.* 2014) indicate that these two events were nearly simultaneous, making this situation unlikely. While it is not possible to exclude the possibility of minimal, evenly distributed brown bear introgression into polar bears, any introgression that did occur must have been limited and makes no

impact on the extant genetic diversity of polar bears.

Wider implications of asymmetric gene flow

Gene flow asymmetry is not an obvious or expected outcome of admixture. Nevertheless, evidence for asymmetric gene flow has been presented in other instances, notably between modern humans and Neandertals (Green *et al.* 2010), between subspecies of house mouse (*Mus musculus*) (Good *et al.* 2008; Teeter *et al.* 2008) and among some Darwin's finches (Grant & Grant 2010). The asymmetric hybridization we observe between polar bears and brown bears differs in important ways from these examples. While the impact of human and Neandertal admixture was geographically widespread, it is currently not possible to know the extent of gene flow into Neandertal populations since population data from this extinct species is scarce. In any case, it is possible that the asymmetry observed thus far is due to demographic phenomenon – a growing, expanding human population entering and replacing a dwindling Neandertal population (Mellars & French 2011; Prüfer *et al.* 2014).

Hybridization between house mouse subspecies *M. m. musculus* and *M. m. domesticus* occurs in a large hybrid zone across central Europe, in which genes flow more readily from *M. m. domesticus* introgressing into *M. m. musculus* than they do in the reverse direction (Teeter *et al.* 2008; Wang *et al.* 2011; Staubach *et al.* 2012). However, unlike the strictly asymmetric gene flow from polar bears into brown bears, genes are also known to flow from *M. m. musculus* into *M. m. domesticus* (Teeter *et al.* 2008; Wang *et al.* 2011; Staubach *et al.* 2012). Thus the asymmetry is a quantitative and not a qualitative phenomenon. Among Darwin's finches, F1 hybrids exhibit biased backcrossing into the paternal species, whichever that may be. This bias is believed to be mediated by imprinting of paternal song (Grant & Grant 1997) and in any case is not a strict barrier to gene flow in either direction (Grant *et al.* 2004).

Our observation of extreme asymmetry in gene flow between polar bears and brown bears suggests that the impacts of admixture may differ considerably among these closely related species. More generally, these results may present a new challenge to the concept of species. The model of past hybridization and gene flow that we present is consistent with a biological species definition of brown bears that includes polar bears, but *inconsistent* with a biological species definition of polar bears that includes brown bears. To our knowledge, there is no current species concept that fully accommodates an *asymmetric* definition of species.

Understanding *why* brown bear alleles do not introgress into polar bear populations may provide important insights into how polar bears survive in their extreme, arctic habitat. The consequences of this observation for conservation of polar bears is clear: polar bears have very little genetic diversity, and this pattern has persisted despite geographically widespread signals of admixture within brown bears. It seems unlikely, therefore, that hybridization or the paucity of genetic diversity among polar bears represents the principle threat to the long-term survival of polar bears. Rather, the rapid rate of recent climate change and consequent disappearance of their habitat (Stirling & Derocher 2012) remain the most proximate and serious threats to polar bears.

Conclusion

Hybridization between polar bears and brown bears has exerted a surprisingly large and asymmetrical influence on the genomes of polar bears and brown bears carrying polar bear genes into brown bears inhabiting a wide geographic area. Interestingly, while brown bears possess polar bear ancestry across significant portions of their genomes, brown bear ancestry appears absent from polar bears. This suggests that an as yet unidentified barrier to gene flow exists that prevents hybrid individuals from successfully backcrossing with the

polar bear population. This one-way barrier to gene flow provides an interesting new framework for the study of the interactions between climate, ecology and speciation.

Acknowledgements

JC and BS were supported by the Packard Foundation and by NSF ARC-09090456 and ARC-1203990. REG was supported by the Searle Scholars Program. BS and REG were additionally supported by the Gordon and Betty Moore Foundation. We thank the University of Alaska at Fairbanks Museum of Natural History and the Natural History Museum in Stockholm for providing brown bear specimens.

Author Contributions

J.C., R.E.G., and B.S. designed the experiment and wrote the manuscript. J.C. performed experiments. L.K. and M.S. generated sequencing libraries. E.E. extracted DNA from the Swedish bear. B.S. extracted DNA from the Chichagof bears. J.C. and R.S. prepared sequence data for analysis.

Data Accessibility

Whole genome shotgun sequencing data produced for this study is available in the NCBI Short Read Archive as SRX795188, SRX796430 and SRX796442. Data from previously published studies are available at the NCBI Short Read Archive SRX155945-51, SRX155953-62, SRX156012-08, SRX156136, SRX156156-63, SRX265152, SRX265434-36, SRX265452-54, SRX265456, SRX265457, SRX265459.

Additional Files

To accommodate space limitations, large supplementary files, Tables 2.3-2.7 are available online along with the original publication of this chapter (Cahill *et al.* 2015).

Appendix

Exclusion of PB7 and LS

Figure 2.5 describes the results of *D*-statistic tests (Green *et al.* 2010; Durand *et al.* 2011) for brown bear admixture into pairs of individual polar bears. Two outliers are observed: LS, which has the most positive distribution of *D*-statistics (indicating that the LS polar bear contains less admixture from brown bears than any other polar bear), and PB7, the only polar bear for which the distribution of *D*-statistics does not include zero (suggesting that PB7 may have *bona fide* admixture from brown bears). We suspect that both of these results are artefacts resulting from errors in the sequencing libraries rather than admixture with brown bears.

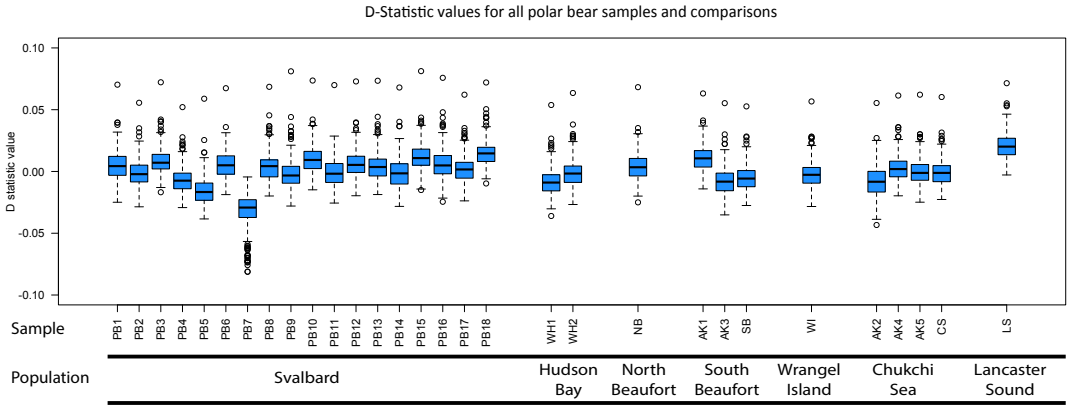


Figure 2.5 - *D*-statistic tests for brown bear admixture into individual polar bears. The boxplot shows the autosomal *D*-statistic for each polar bear with all possible combinations of polar bear and brown bear introgressors. Negative values indicate that the individual listed on the x-axis has more polar bear ancestry.

As reported previously (Cahill *et al.* 2013), the LS polar bear data contain DNA damaged sites. This ancient DNA associated error biases the *D*-statistic away from zero. The LS data were generated from a 40-year-old bone that is currently stored at the National Museum of National History in Washington, DC (Cahill *et al.* 2013). These data show nearly twice the number of C to T and G to A transitions when compared to the polar bear reference genome than the other polar bears. This is most likely due to miscoding lesions resulting from cytosine deamination to uracil, which is the most common form of ancient DNA damage (Hofreiter 2001). Although damage does not affect the *D*-statistic when the damaged individual is in the position of the potential introgressor, it will bias the *D*-statistic when the damaged sample is included in either the P1 or P2 positions (potential recipient of introgression). In this case, the effect is a false positive match between the damaged individual and the outgroup, and consequent identification of any undamaged comparison individual as admixed, as seen for LS in Figure 2.5.

D-statistic results show that the PB7 polar bear has a stronger signal for

brown bear ancestry than any other polar bear (mean D= -0.033). However, *D*-statistic tests involving PB7 and the Kenai brown bear (Ken) as potential introgressor are particularly extreme, with an average *D* of -0.066 for D(PB7, other polar bear, Ken, black bear), compared to -0.028 for *D*(PB7, other polar bear, other brown bear, black bear). Given that the PB7 and Ken data were generated as part of the same project (Miller *et al.* 2012), we explored the possibility that contamination between these data may exist.

We mapped the published genomic data from PB7 and Ken to a reference polar bear mitochondrial genome (Delisle & Strobeck 2002) bwa (Li & Durbin 2010) and samtools (Li *et al.* 2009) under the same parameters as the analyses described in the main text (Methods). We then identified all positions in the mitochondrial genome where the consensus mitochondrial genome sequences assembled for PB7 and Ken differed from each other. At each of these positions, we then calculated the number of times the consensus allele for Ken was observed in the PB7 data set, and vice versa. For most sites, we observed zero reads in the Ken data set that matched the consensus assembly for PB7 (Figure 2.6A). However, in the PB7 data set, 0.5-1.0% of PB7 reads match the Ken allele (Figure 2.6A). To determine whether this result could be due to differences in sequencing depth, we down sampled the Ken and PB7 data sets to an equal number of reads, and counted the number of sites in the down sampled data sets in which the consensus allele for the opposite species was observed (Figure 2.6B). We observed that as the number of reads sampled increased so did the number of Ken alleles observed in the PB7 data set. These results are most simply explained by small amounts of contamination of the PB7 data set with data from Ken. We therefore chose to exclude PB7 from further analysis.

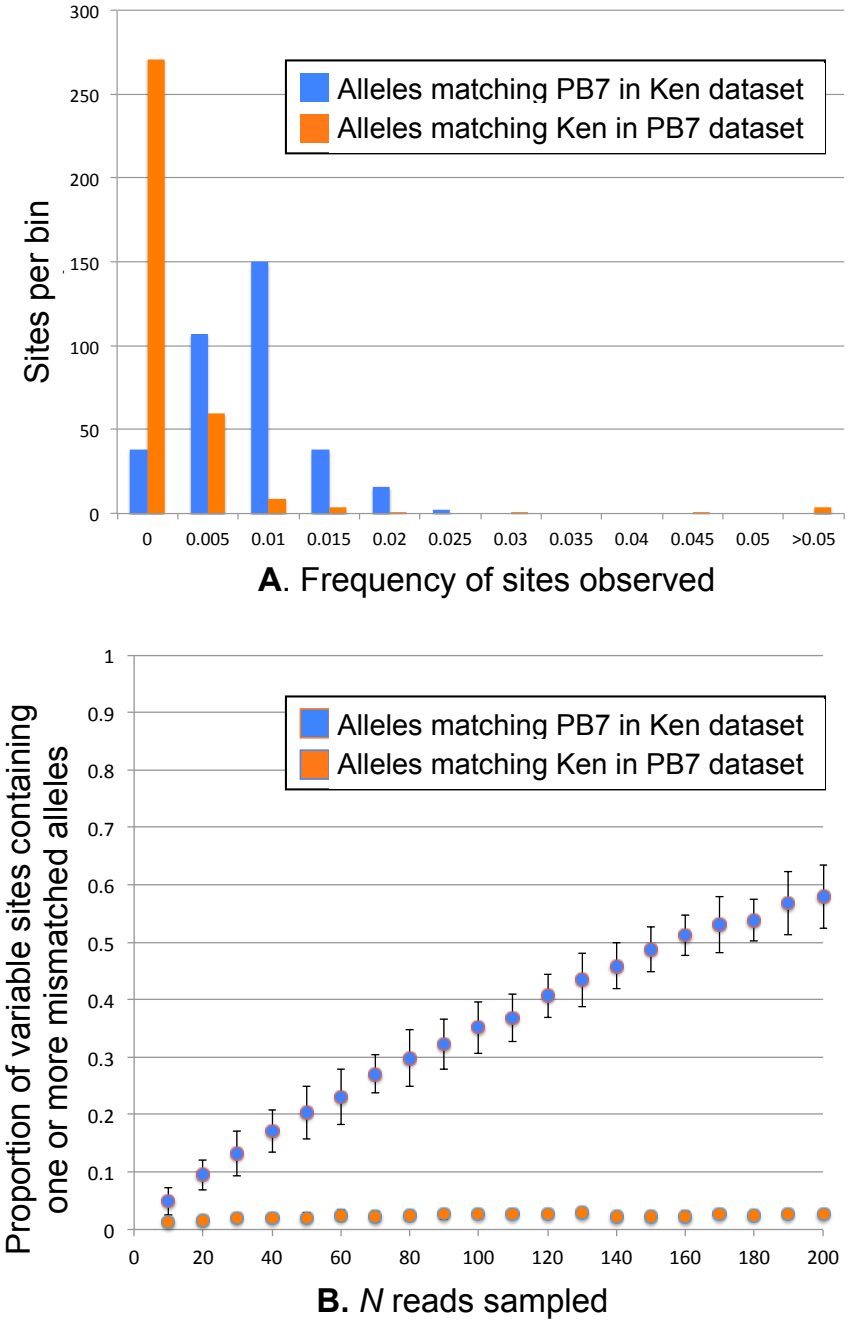


Figure 2.6 – Tests for contamination of PB7 by Ken
The frequency of reads in the PB7 and Ken data sets that may be derived from contamination by the other data set (A). To control for the detection of potential contaminant reads based on differences in coverage, we sampled fixed numbers of reads from each individual at variable sites. (B) The mean frequency of sites containing one or more potentially contaminant reads of 20 random draws of *N* reads. Error bars equal two standard deviations.

Y-chromosome Analysis

Recent studies have shown that the Y-chromosome haplotypes found in polar bears and form reciprocally monophyletic clades (Bidon *et al.* 2014). These findings include the ABC islands brown bears which at the Y-chromosome fall within the brown bear clade (Bidon *et al.* 2014). We assessed the same ~390 KB Y-chromosome scaffold of the polar bear reference genome (Li *et al.* 2011) as Bidon et al (scaffold 297) and calculated the sequence divergence between all male bears in our panel.

Using the pseudohaploidization method used elsewhere in this study (Methods) we generated representative sequences for each individual at scaffold 297 and compared pairwise differences. Within polar pairwise differences ranged from 1-4 differences per 10,000 sites. Polar bear to Baranof island brown bear differences range from 11-12.5 differences per 10,000 sties. Polar bear to American black bear differences range from 29-31 differences per 10,000 sites and the Baranof brown bear has 29 differences to the American black bear per 10,000 sites.

Our analytical power in this analysis is limited by only having a single male brown bear. However, insofar as our results are interpretable they show that the Baranof island brown bear (a male ABC islands bear) possesses a Y-chromosome haplotype that falls well outside the diversity of polar bear Y-chromosomes. Bidon et al observed brown bear to polar bear divergence to be ~35% of brown/polar bear divergence to black bear (Bidon *et al.* 2014). Our observation of ~39% (Figure 2.7) is qualitatively consistent with previous results and probably indicative of minor filtering differences.

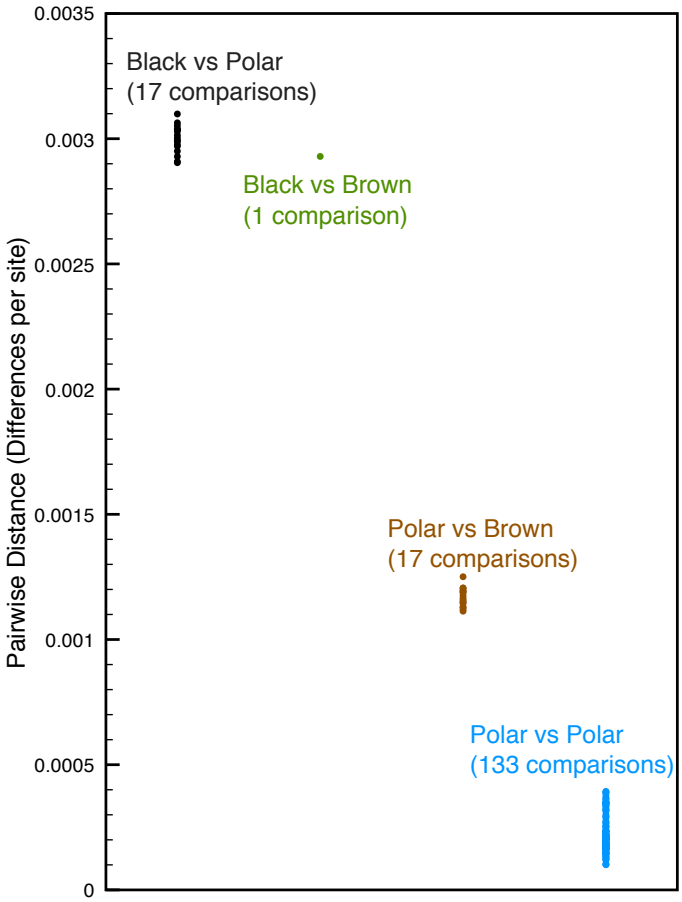


Figure 2.7 – Y-chromosome pairwise difference
The number of pairwise differences per site between male individuals in our panel of bears at a ~390KB Y-chromosome scaffold. We find that the Baranof sample (the only male brown bear in this study) falls outside the range of divergences observed between polar bears. The level of divergence is consistent with previous studies analyzing the same scaffold with different individuals suggesting that brown bears and polar bears form reciprocally monophyletic clades at the Y-chromosome (Bidon *et al.* 2014).

References

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science (New York, N.Y.)*, **295**, 2267–70.

Bidon T, Janke A, Fain SR *et al.* (2014) Title: Brown and polar bear Y chromosomes

reveal extensive male-biased gene flow within brother lineages. *Molecular biology and evolution*, msu109–.

Cahill JA, Green RE, Fulton TL *et al.* (2013) Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. (MW Nachman, Ed,). *PLoS genetics*, **9**, e1003345.

Cahill JA, Stirling I, Kistler L *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular ecology*, **24**, 1205–1217.

Cronin M a., Amstrup SC, Garner GW, Vyse ER (1991) Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursus). *Canadian Journal of Zoology*, **69**, 2985–2992.

Cronin MA, Rincon G, Meredith RW *et al.* (2014) Molecular Phylogeny and SNP Variation of Polar Bears (Ursus maritimus), Brown Bears (U. arctos), and Black Bears (U. americanus) Derived from Genome Sequences. *The Journal of heredity*, **105**, 312–23.

Davison J, Ho SYW, Bray SC *et al.* (2011) Late-Quaternary biogeographic scenarios for the brown bear (Ursus arctos), a wild mammal model species. *Quaternary Science Reviews*, **30**, 418–430.

Delisle I, Strobeck C (2002) Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Molecular biology and evolution*, **19**, 357–61.

Derrien T, Estellé J, Marco Sola S *et al.* (2012) Fast computation and applications of genome mappability. (CA Ouzounis, Ed,). *PloS one*, **7**, e30377.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Edwards CJ, Suchard MA, Lemey P *et al.* (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology : CB*, **21**, 1251–8.

Good JM, Handel MA, Nachman MW (2008) Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution; international journal of organic evolution*, **62**, 50–65.

Grant PR, Grant BR (1997) Mating patterns of Darwin's Finch hybrids determined by song and morphology. *Biological Journal of the Linnean Society*, **60**, 317–343.

Grant PR, Grant BR (2010) Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 1065–76.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution;*

international journal of organic evolution, **58**, 1588–99.

Gray AP (1971) *Mammalian Hybrids a Check-list with Bibliography*. Commonwealth Agricultural Bereaux.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, **328**, 710–22.

Hailer F, Kutschera VE, Hallström BM *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science (New York, N.Y.)*, **336**, 344–7.

Hailer F, Kutschera VE, Hallström BM *et al.* (2013) Response to comment on "Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage". *Science (New York, N.Y.)*, **339**, 1522.

Hofreiter M (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.

Kelly BP, Quakenbush LT (1990) Spatiotemporal use of lairs by ringed seals (Phoca hispida). *Canadian Journal of Zoology*, **68**, 2503–2512.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods in molecular biology (Clifton, N.J.)*, **840**, 197–228.

Korsten M, Ho SYW, Davison J *et al.* (2009) Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Molecular Ecology*, **18**, 1963–1979.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of ice age brown bears. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1651–4.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **26**, 589–95.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–9.

Li B, Zhang G, Willersleve E, Wang J (2011) GigaDB Dataset - DOI 10.5524/100008 - Genomic data from the polar bear (Ursus maritimus).

Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–19.

Lindqvist C, Schuster SC, Sun Y *et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5053–7.

Liu S, Lorenzen ED, Fumagalli M *et al.* (2014) Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, **157**, 785–794.

Lohse M, Bolger AM, Nagel A *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic acids research*, **40**, W622–7.

Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in Drosophila. (NH Barton, Ed,). *PLoS biology*, **5**, e243.

McLellan B, Reiner DC (1994) A Review of Bear Evolution. *Ursus*, **9**, 85–96.

Mellars P, French JC (2011) Tenfold population increase in Western Europe at the Neandertal-to-modern human transition. *Science (New York, N.Y.)*, **333**, 623–7.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, **2010**, pdb.prot5448.

Miller W, Schuster SC, Welch AJ *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2382–90.

Paetkau D, Amstrup SC, Born EW *et al.* (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, **8**, 1571–1584.

Preuß A, Gansloßer U, Purschke G, Magiera U (2009) Bear-hybrids: behaviour and phenotype. *Der Zoologische Garten*, **78**, 204–220.

Prüfer K, Racimo F, Patterson N *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–9.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, **26**, 841–2.

Sacco T, Van Valkenburgh B (2004) Ecomorphological indicators of feeding behaviour in the bears (Carnivora: Ursidae). *Journal of Zoology*, **263**, 41–54.

Sankararaman S, Mallick S, Dannemann M *et al.* (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507**, 354–357.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science (New York, N.Y.)*, **323**, 737–41.

Slater GJ, Figueirido B, Louis L, Yang P, Van Valkenburgh B (2010) Biomechanical consequences of rapid evolution in the polar bear lineage. *PloS one*, **5**, e13870.

Staubach F, Lorenc A, Messer PW *et al.* (2012) Genome patterns of selection and

introgression of haplotypes in natural populations of the house mouse (Mus musculus). (MH Kohn, Ed,). *PLoS genetics*, **8**, e1002891.

Stirling I (1974) Midsummer observations on behavior of wild polar bears (Ursus maritimus). *Canadian Journal of Zoology*, **52**, 1191–1198.

Stirling I (2011) *Polar Bears: The Natural History of a Threatened Species*. Fitzhenry and Whiteside, Brighton, MA.

Stirling I, Øristland NA (1995) Relationships between estimates of ringed seal (Phoca hispida) and polar bear (Ursus marifimus) populations in the Canadian Arctic. *Canadian Journal of FIsh and Aquatic Sciences*, **52**, 2594–2612.

Støen O-G, Zedrosser A, Saebø S, Swenson JE (2006) Inversely density-dependent natal dispersal in brown bears Ursus arctos. *Oecologia*, **148**, 356–64.

Talbot SL, Shields GF (1996) Phylogeography of Brown Bears (Ursus arctos) of Alaska and Paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, **5**, 477–494.

Teeter KC, Payseur BA, Harris LW *et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome research*, **18**, 67–76.

Valdiosera CE, García N, Anderung C *et al.* (2007) Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Molecular ecology*, **16**, 5140–8.

Waits LP, Talbot SL, Ward RH, Shields GF (1998) Mitochondrial DNA American for Brown Bear Phylogeography of the and North. *Conservation Biology*, **12**, 408–417.

Wang L, Luzynski K, Pool JE *et al.* (2011) Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular ecology*, **20**, 2985–3000.

Chapter 3: Genomic evidence of globally widespread admixture from polar bears into brown bears during the last ice age

James A. Cahill, Peter D. Heintzman, Kelley Harris, Matthew Teasdale, Joshua Kapp, André E Rodrigues Soares, Ian Stirling, Nigel Monaghan, Ceiridwen J. Edwards, Alexander V. Malev, Aliaksandr A. Kisleika, Daniel Bradley, Richard E. Green and Beth Shapiro

Unpublished, author list subject to change prior to peer review publication

Abstract

Periods of rapid climate change can create opportunities for evolutionary adaptation, including by facilitating admixture between closely related species. Recent genomic analyses have provided substantial evidence for gene flow from polar bears (*Ursus maritimus*) into Alaskan brown bears (*Ursus arctos*) (Cahill *et al.* 2013, 2015; Liu *et al.* 2014) and mitochondrial DNA supports polar bear introgression into Irish brown bears (Edwards *et al.* 2011). However, the timing, frequency, and evolutionary significance of admixture between bear species remains debated. Here, we explore the link between ice age climate change and admixture directly by analyzing genomic DNA from ten ancient brown bears that lived in Ireland during the last 40,000 years, a period that spans the peak of the last ice age and subsequent rapid warming. We find the brown bears that lived closer in time to the maximum local ice extent (Clark *et al.* 2012) had the most polar bear ancestry, up to 20%, but that polar bear ancestry declined to undetectable levels by the time the 4,000 years ago. In addition to these Irish bears, we find traces of polar bear ancestry in two previously unstudied brown bear populations in Canada, and Russia, indicating that admixture with polar bears was a globally widespread phenomenon. Our results show that introgression of polar bear alleles into brown bears has occurred multiple times, and support a model in which climate change promotes episodes of admixture followed by a gradual decline in polar bear ancestry. This model may be informative for many admixing species pairs impacted by climate change.

Main Text

Post-divergence gene flow between species is increasingly understood to have been common in evolutionary history (Green *et al.* 2010; Dasmahapatra *et al.* 2012; Poelstra *et al.* 2014; Lamichhaney *et al.* 2015). Also known as admixture, this

process most commonly occurs when two species that have been geographically isolated suddenly overlap in range and are reproductively compatible. Genomic analyses across hybrid zones have revealed considerable variation among species pairs in both the spatial patterns and evolutionary consequences of admixture (Good *et al.* 2008; Dasmahapatra *et al.* 2012; Poelstra *et al.* 2014). In some cases, genomic incompatibilities lead to hybrid phenotypes that are less fit than either parent species (Good *et al.* 2008). In other cases, new combinations of alleles may provide local adaptive advantages (Garroway *et al.* 2010). Hybridization may therefore be an important source of evolutionary novelty, for example during periods of rapid climate change, when shifting habitats may form communities comprising previously isolated populations and species (Graham *et al.* 1996; Parmesan & Yohe 2003; Hoffmann & Sgrò 2011).

Here, we use a paleogenomic approach to directly explore the role of climate change in facilitating admixture between brown bears (*Ursus arctos*) and polar bears (*U. maritimus*) (Figure 3.1). Focusing on a now-extinct population of brown bears from present-day Ireland, we isolate genomic DNA from ten cave-preserved bones that were morphologically and isotopically identified as brown bears (Edwards *et al.* 2011) and that range in age from 37.5-3.9 thousand calibrated years before present (cal. ka BP). This interval spans the local peak of the last ice age *ca.* 24.7 cal. ka BP (Peters *et al.* 2015), when polar bear populations would have been most proximate to present-day Ireland. To explore the geographic extent of potential admixture, we also extracted and analyzed DNA from ancient and modern brown bear remains from northeastern Canada and northeastern Asia.

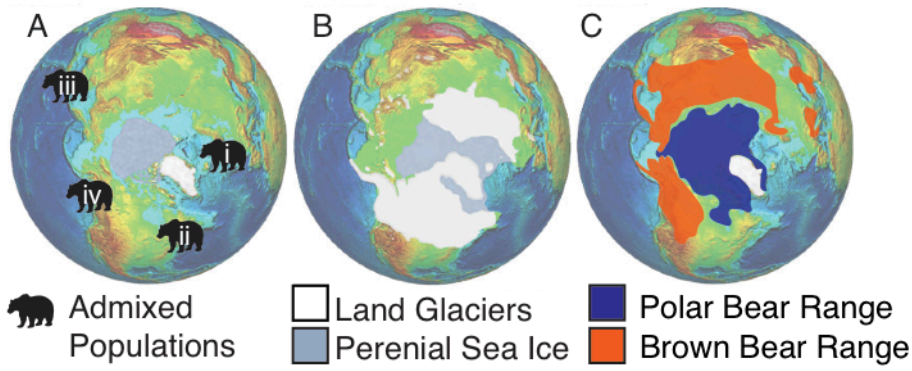


Figure 3.1. (A) Geographic locations of brown bear populations identified here and in previous analyses(Cahill *et al.* 2013) as having some component of polar bear ancestry: (i) present-day Ireland; (ii) Chaplain Sea, Quebec, Canada; (iii) Kunashir Island, Russia; (iv) ABC Islands, Alaska, USA. Panel A shows the present day distribution of glaciers and sea ice. Details of samples used here are provided in Extended Data Table 3.1. Each of these admixed populations is located near the extent of sea and/or glacial ice at the last glacial maximum, *ca*, 24ky BP(Peters *et al.* 2015), which is depicted in panel B, but far from the present-day range of polar bears(Schliebe *et al.* 2008), as shown in panel C. Base image from (http://earthobservatory.nasa.gov/Features/BorealMigration/boreal_migration2.php).

To estimate the amount of polar bear ancestry within each Irish bear genome, we used the D and $\hat{f}$ statistics (Green *et al.* 2010; Durand *et al.* 2011) (Figure 3.2, 3.3, 3.4; Tables 3.4, 3.5), which infer admixture based on an excess of shared derived ancestry between polar bears and Irish brown bears compared to that between polar bears and an unadmixed brown bear. Strikingly, the Irish brown bears with the largest proportion of polar bear ancestry lived temporally closest to the peak of the last ice age, with the most admixed bear, 20% polar bear ancestry (Z=7.2), dating to *ca.* 14 cal. ka BP. We found that Irish brown bears that lived after this time had less polar bear ancestry (Figure 3.2).

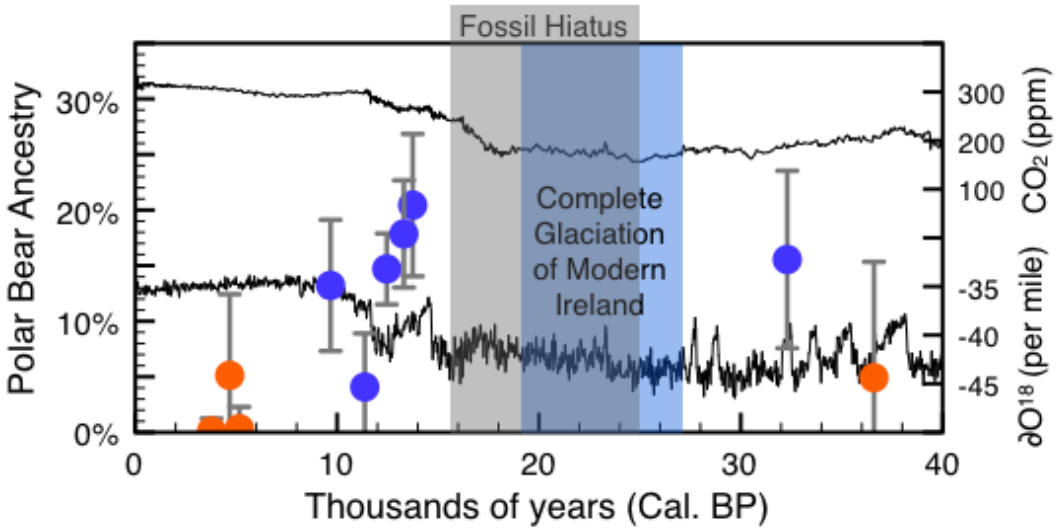


Figure 3.2. We estimate the percentage of each Irish brown bear's genome derived from polar bear ancestry using $\hat{f}$ and plotted it against it's calibrated age (Methods). Error bars show 95% confidence intervals estimated by weighted block jackknife (1.96 standard errors). Mitochondrial haplotype(Edwards *et al.* 2011) is indicated by color; polar bear like, clade 2 (blue) and brown bear like, clade 1 (orange). To show the correspondence between polar bear ancestry and climates we show two climate proxies: dO18 from NGRIP and CO2 from Vostok, in both cases values closer to the top of the figure are indicative of warmer temperatures. Glacial reconstructions indicate that all of modern Ireland was glaciated during the LGM from 27-19 ka(Clark *et al.* 2012) although carbon dates indicate that some areas in the far south-east may have been ice free as late as 25 ka(Woodman *et al.* 1997). During the LGM there is a general hiatus in the vertebrate fossil record until 15 ka(Woodman *et al.* 1997; Stuart *et al.* 2004). Brown bears occur in the Irish fossil record before and after the LGM but are absent from 32-14 ka(Woodman *et al.* 1997; Edwards *et al.* 2011), the most recent pre-LGM and most ancient post-LGM brown bears are included in this study.

In addition to having the most polar bear ancestry in their nuclear genomes (Figure 3.2), the Irish brown bears that lived around the time of the LGM have a mitochondrial haplotype that is very closely related to that of present day polar bears (Edwards *et al.* 2011) (Figure 3.2). Polar bear-like mitochondrial haplotypes are also found in an admixed population of brown bears that live on the ABC (Admiralty, Baranof, Chichagof) Islands of southeast Alaska (Cronin *et al.* 1991) (Figure 3.1). Because the ABC Island bears also have an excess of polar bear ancestry in the X chromosomes compared the autosomes, this pattern been attributed to sex-

biased admixture (Cahill *et al.* 2013, 2015). Because brown bears are maternally philopatric (Zedrosser *et al.* 2007), the earliest dispersers to uncolonized habitats are likely be male. While we were unable to recover sufficient comparative information from the X chromosomes of the ancient Irish bears, this hypothesis is consistent with the early predominance of polar bear-like mitochondrial haplotypes in the Irish brown bears (Edwards *et al.* 2011).

We hypothesize that a combination of local changes in habitat availability, lower sea levels, and species-specific natural histories facilitated the observed admixture between brown and polar bears. The paleoecological and fossil records of Ireland suggest that all or most of Ireland was glaciated throughout much of the last ice age, leaving little to no habitat for brown bears (Clark *et al.* 2012; Ó Cofaigh *et al.* 2012; Edwards *et al.* 2014). At the same time, major tidewater glaciers on the western shelf and down the Irish Sea basin and offshore iceberg scouring of the sea floor suggest the possibility of productive sea ice habitat for polar bears along the Irish coast (Edwards *et al.* 2011; Clark *et al.* 2012; Ó Cofaigh *et al.* 2012; Peters *et al.* 2015). As resident brown bear populations declined during the approach of the last glacial maximum (LGM), this proximity in range probably led to admixture, as it can in the present day populations of brown bears and polar bears whose ranges overlap (Stirling 2011). After the LGM, brown bears probably re-colonized Ireland via a land-bridge that connected the Island to mainland Europe (Clark *et al.* 2012). These colonizing bears would have encountered and hybridized with resident polar bears or hybrid bears. Such encounters probably decreased in frequency as the ice receded. Continuing dispersal from the mainland of unadmixed brown bears would gradually reduce the proportion of polar bear ancestry in the Irish brown bear

population, leading to the pattern observed in the Irish brown bear genomes (Figure 3.2).

In addition to neutral demographic processes, it is possible that selection against polar bear alleles may have played a role in the rapid post-glacial decline of polar bear ancestry in the admixed Irish population. Not only do polar bears and brown bears differ morphologically, physiologically, and behaviorally (Stirling 2011), but the long term, very small effective population size of polar bears (Miller *et al.* 2012) may lead to an overall higher genetic load compared to brown bears. To test whether selection against maladaptive polar bear alleles may have contributed to the post-glacial decline in polar bear ancestry, we performed simulations using a recently proposed population history model (Liu *et al.* 2014) (see Methods). First, we estimated the expected relative abundance of deleterious alleles in polar bears and brown bears prior to admixture. Then, we simulated introgression of this predicted genetic load into the Irish brown bear genomes at the proportions inferred in this study (Figure 3.2). Despite their comparatively small effective population size, polar bears are expected to have only a slightly higher genetic load than brown bears, resulting in a median decrease in fitness of 4% (Figure 3.8). Simulations showed that polar bears' greater genetic load reduced polar bear ancestry from a starting simulated amount of 25% to 22.4% before stabilizing (Figure 3.9). The post-LGM reduction in polar bear ancestry is therefore not explainable by genetic load, and the observed decline in polar bear ancestry is probably due to backcrossing with less admixed brown bears.

To explore the global influence of habitat changes associated with the last ice age on bear genomic diversity, we estimated the proportion of polar bear ancestry from two brown bears from Kunashir Island, Russia (Figure 3.1), which is located

between the Northwest Pacific Ocean and the Okhotsk Sea, The two Kunashir Island bears have 2.2% (Z=3.3) and 12.4% (Z=6.8) polar bear ancestry (Table 3.5), strongly supporting past admixture with polar bears. Sediment cores taken from the southern Okhotsk sea show substantial changes in primary productivity between the LGM and the present-day, consistent with perennial sea ice cover 30-15 ka BP(Seki *et al.* 2004). Kunashir Island would have been at the southernmost extent of this perennial ice, and therefore close to polar bear habitat. Interestingly, white coat color occurs in some Kunashir Island brown bears (Sato *et al.* 2011) and, while the source of this color variation is unknown, intermediate coat color is typical of polar/brown bear hybrids (Preuß *et al.* 2009).

We also recovered and analyzed DNA from a brown bear bone recovered from Champlain Sea deposits in Quebec, Canada (Figure 3.1), that dates to 11.3 cal. ka BP (Harington *et al.* 2014). We found this bear to have at least 4.2% (Z=1.9) polar bear ancestry (Table 3.5). Although, this result is not significant tests that include transition sites produce a similar value for polar bear ancestry but because of the larger number of sites included those comparisons not restricted to transversions do provide significant evidence of polar bear ancestry (Table 3.5; Figure 3.3). Given that the inclusion of transition sites biases D to underestimate the amount of admixture it is likely that further sequencing, now underway will provide sufficient data to more robustly describe polar bear ancestry in this sample.

The Quebec bone was discovered near the south-eastern limit of the Laurentide ice sheet along with remains of arctic marine mammals including Ringed seals (*Pusa hispida*), Bearded seals (*Erignathus barbatus*) and Atlantic walrus (*Odobenus rosmarus rosmarus*) (Harington *et al.* 2014), suggesting that this location may also have supported polar bears after the LGM. Importantly, while both the Kunashir Island

and the Champlain Sea would have been near the maximum extent of the polar bear range at the peak of the LGM, major geographic barriers isolate both of these localities from the ABC Islands and Ireland, indicating at least four distinct admixture events associated with the last ice age.

Admixture between brown bears and polar bears has also been observed in the present-day Canadian Arctic (Doupé *et al.* 2007), and has been attributed to climate-induced overlap between the two species (Kelly *et al.* 2010). Together, these data reveal the ongoing and dynamic nature of gene flow between brown bears and polar bears, and the important role that climate change and consequent habitat redistribution plays in facilitating admixture. Intriguingly, the evolutionary consequences of this admixture appear to be mediated by ecological and behavioral differences between the two species, which maintain polar bears as a genetically distinct lineage that is devoid of brown bear DNA (Peacock *et al.* 2015; Cahill *et al.* 2015). These results highlight the complicated nature of speciation, and suggest that bears may be a useful lineage in which to explore the formation of incompatibilities between diverging lineages.

Correlation between recent climate change and admixture has been observed recently for several related species pairs, including trout (Muhlfeld *et al.* 2014), flying squirrels (Garroway *et al.* 2010), *Pachycladon* grasses (Becker *et al.* 2013) and damselflies (Sánchez-Guillén *et al.* 2014). While the long-term evolutionary consequences to these species pairs are not yet known, preliminary evidence suggests a wide range of possible outcomes, from extinction via genetic replacement (Muhlfeld *et al.* 2014) to the creation of hybrid phenotypes with higher fitness in the new habitat relative to the parental lineages (Becker *et al.* 2013). While it is tempting to consider these as localized examples, and therefore unlikely to have widespread

evolutionary consequences, introgressed DNA will in many instances spread to non-admixed populations over time. Introgressed polar bear DNA has been observed in mainland Alaskan brown bears, for example, probably due to post-glacial dispersal from the ABC Islands (Liu *et al.* 2014; Cahill *et al.* 2015). Admixture resulting from climate-related habitat redistribution is likely, therefore, to have long-term and widespread evolutionary consequences, and may be an important mechanism for generating and maintaining diversity.

Methods

DNA extraction, library preparation and sequencing

We tested bones and teeth from cave sites across Ireland (Edwards *et al.* 2011), one bone from Champlain Sea deposits in Quebec (Harington *et al.* 2014), Canada, and two skin specimens from Kunashir Island, Russia for DNA preservation (Figure 3.1; Tables 3.1, 3.2, 3.6). All 18 samples were morphologically identified as brown bears and previous isotopic analysis of the Irish bones indicated that the Irish bears consumed a terrestrial rather than a marine diet , consistent with the bone coming from a brown bear. We calibrated radiocarbon dates for each ancient sample (Woodman *et al.* 1997; Edwards *et al.* 2011; Harington *et al.* 2014) and the non-bear samples used to estimate the timing of the LGM fossil hiatus in Ireland (Woodman *et al.* 1997) using OxCal 4.2 (Bronk Ramsey 2009), with the IntCal13 terrestrial calibration curve (Reimer *et al.* 2013). Unless otherwise noted, all sample ages are reported as median calibrated dates with 95% confidence intervals.

All pre-amplification laboratory work on the ancient bone specimens was conducted in a clean lab facility at UC Santa Cruz that is dedicated to ancient DNA research, following standard procedures for working with degraded DNA (Fulton 2012). We extracted DNA from 100-120 mg of powdered bone or tooth root

following Dabney *et al*. (Dabney *et al.* 2013). For the skin samples, we homogenized ~120mg, which was extracted using a modified version of the Dabney *et al*. (Dabney *et al.* 2013) protocol. This involved substituting the lysis buffer and overnight incubation steps with those described in Gilbert *et al*. 2007(Gilbert *et al.* 2007). Following extraction, we converted the DNA extracts into double-stranded, indexed sequencing libraries following Meyer and Kircher (Meyer & Kircher 2010), as modified by Heintzman *et al*. (Heintzman *et al.* 2015). We pooled and sequenced libraries using v3 2× 75 base pair (bp), paired-end chemistry on the Illumina MiSeq platform. Following sequencing, we merged overlapping reads using SeqPrep (https://github.com/jstjohn/SeqPrep)(St John 2011) using the default parameters, resulting in the removal of reads shorter than 30 bp and requiring an overlap of 10 bp between the paired-end reads for read merging. We mapped the reads to the polar bear reference genome (Li *et al.* 2011) using bwa aln v0.7.7 (Li & Durbin 2010). We eliminated reads with map quality scores less than 25 (MapQ<25), and removed PCR duplicates with samtools rmdup v0.1.19 (Li *et al.* 2009). We estimated endogenous DNA content by dividing the number of reads mapped by the total number of reads (Tables 3.2, 3.6). For the Irish brown bear samples, we selected ten for additional sequencing, based on endogenous DNA content and sample age (Table 3.2). These were re-pooled and sequenced on two lanes of the Illumina HiSeq-2500 using 2× 50 bp paired-end sequencing in rapid run mode.

Mapping and reference bias correction

DNA decay results in deamination (C to T) damage(Stiller *et al.* 2006) and short fragment lengths(Poinar *et al.* 2006) which makes reads less mappable to a reference than modern DNA reads. To identify optimal mapping parameters, we mapped reads from an Irish brown bear, Leitrim-4, to a polar bear mitochondrial

genome (NC_003428.1)(Delisle & Strobeck 2002) using mia (https://github.com/mpieva/mapping-iterative-assembler)(Briggs *et al.* 2009). Mia is an alignment program that is designed to compensate for ancient DNA damage and that conducts a full Smith-Waterman alignment for each read. However, while mia provides high quality mitochondrial mappings, the computational demands do not scale to whole nuclear genomes. To identify an appropriate whole genome mapper and set of mapping parameters, we mapped reads from Leitrim-4 using bwa v0.7.7 (Li & Durbin 2010) and bowtie2 v2.1.0 (Langmead & Salzberg 2012) under various parameters, ultimately selecting a set of parameters resulted in <2% false positive mappings, which are defined as mappings in which the whole genome mapper aligned a read that was not aligned by mia, and < 2% false negative mappings, which are defined as mappings in which whole genome aligner failed to map a read that was mapped by mia (Table 3.7).

Based on the analysis described above, we mapped ancient nuclear genomic reads using bowtie2 v2.1.0 (Langmead & Salzberg 2012) under the local alignment setting. We set the –N flag to 1 to allow a mismatch in the alignment seed and the –mp flag to 4 to reduce the maximum penalty for mismatches from a default value of 6 to 4. We eliminated reads with map quality scores less than 30 (MapQ <30). We removed PCR duplicates with samtools rmdup v0.1.19 (Li *et al.* 2009). We estimated endogenous DNA content by dividing the number of reads mapped by the total number of reads (Table 3.4).

To counteract potential reference genome bias effects in our nuclear genome analyses, we mapped our reads to both the polar bear reference genome (Li *et al.* 2011) and a brown bear "pseudo-reference," which we created by calling the consensus allele of an unadmixed Swedish brown bear (SAMN02256313) (Liu *et al.*

2014) at each site in the polar bear reference genome (Table 3.7). We used the union of high confidence read mappings to both reference genomes for downstream analyses. If a read mapped to different locations in the two genomes, we retained the mapping with the higher map quality score; in the case of a tie a randomly chosen mapping was retained. To test whether the samples exhibited patterns of cytosine deamination (C to T) damage consistent with ancient DNA and quantify the samples fragment lengths we ran mapDamage v2.0.5 (Jónsson *et al.* 2013) for each Map Quality and PRC duplicate filtered bam file (Figures 3.6, 3.7).

For comparison between the <1× coverage ancient Irish brown bears and 4-127× coverage genomes from present-day bears, we reduced all samples to a single representative haploid sequence. For each individual at each position in the reference genome, we randomly selected a single base with Base-Quality ≥ 30 that was on a read with Map-Quality ≥ 30. To avoid regions of poor assembly, we excluded scaffolds in the polar bear reference genome (Li *et al.* 2011) less than 1 Mb in length. We also excluded any sites where an individual's coverage fell outside the 5^{th}-95^{th} percentiles of its coverage distribution (for present-day samples) or was greater than 2× (for ancient samples). Finally, we excluded any sites in the polar bear reference sequence that were not identified as part of a uniquely mappable 35-mer by GEM (Derrien *et al.* 2012).

Estimating the proportion of polar bear ancestry

We tested for admixture using the *D*-statistic (also known as the ABBA/BABA test)(Green *et al.* 2010; Durand *et al.* 2011). This test detects differences in derived allele sharing between two individuals of the same species and a candidate introgressing species. We tested for an excess of derived allele sharing between polar bears and Irish/Quebec/Kunashir brown bears using *D*(our brown

bear, Fennoscandian brown bear, polar bear, American black bear). Previous studies have shown Fennoscandian brown bears share fewer derived alleles with polar bears than any other brown bear population for which genome sequence data is currently available (Liu *et al.* 2014; Cahill *et al.* 2015). We calculated the *D*-statistic with all combinations of the 4 published Fennoscandian brown bears (Liu *et al.* 2014; Cahill *et al.* 2015), the 28 published polar bears (Miller *et al.* 2012; Cahill *et al.* 2013) (Table 3.3) and each of our newly sequenced brown bears. We calculated the *D*-statistic separately for the X-chromosome scaffolds identified in (Cahill *et al.* 2013) and for other, putatively autosomal, scaffolds greater than 1Mb in length. We tested for statistically significant divergence from the unadmixed expectation, $D=0$, using a weighted block jackknife (Kunsch 1989) with 5Mb blocks. Low coverage in the Irish brown bears provided an insufficient number of informative sites on scaffolds mapping to the X-chromosome for meaningful interpretation (Table 3.4). To quantify the potential impact of reference bias, we calculated *D*-statistics separately for the candidate hybrid brown bears aligned to each reference genome and for reads mapping to either genome. To examine the potential impact of cytosine deamination damage on *D*-statistic estimates, we conducted tests both using all sites and restricted to transversions only, the latter excluding cytosine deamination but reducing the number of informative sites (Table 3.4). In the main text, we conservatively report the transversion only analysis of the union of reads mapped to either reference genome. Results with other mappings are provided in Tables 3.4, 3.5 and Figure 3.3.

To quantify the fraction of each brown bear genome that was derived from polar bear ancestry, we used the $\hat{f}$ statistic (Green *et al.* 2010; Durand *et al.* 2011). $\hat{f}$ is used to estimate the fraction of the genome derived from introgression by dividing

the value of ABBA-BABA observed by the expected value of ABBA-BABA if the hybrid were 100% polar bear. $\hat{f}$ is biased toward underestimation of the fraction of the admixed individual's genome derived from polar bear, and should be considered a lower bound estimate of the proportion of the genome introgressed from polar bear. This bias increases in proportion to the antiquity of the admixture event, as described by Durand et al (Durand *et al.* 2011). As with the *D*-statistic, we calculated $\hat{f}$ with and without transitions and using the polar bear reference, brown bear pseudo-reference and the union of reads mapped to either reference sequence. We tested for statistical significance using the weighted block jackknife with a 5 Mb block, significant results are more than three standard errors from zero ($Z \geq 3$).

Recalibrating Admixture Estimates for Low Coverage

To test the impact of using low <1X coverage samples for D-statistic inference we applied two tests, first, we randomly sampled mapped reads from two brown bears, one with 9% detectable polar bear ancestry (Chi1) and one without detectable polar bear ancestry (Swe) (Cahill *et al.* 2015). We generated 20 random resamplings for each individual at 0.1 million, 1 million, 10 million and 20 million reads, which correspond roughly to 0.01X, 0.1X, 1X and 2X coverage. Second, we randomly resampled unlinked sites from the haploidized files generated from the full coverage data sets of Chi1 and Swe. We sampled the average number of covered sites from the 0.1 million, 1 million, 10 million and 20 million read samples (14,231,218; 137,044,256; 998,411,021 and 1,491,931,181 sites respectively). Our expectation given the lack of any previously reported coverage induced bias in the literature was that with fewer reads variance of the inferred value would increase, as captured by the weighted block jackknife, but the mean would be unaffected.

In contrast to this expectation we found that down-sampling the reads up to and including 2X coverage consistently led to inference of greater amounts of admixture than the entire Chi1 and Swe datasets Figure 3.10A. This effect was more pronounced at lower coverage, but at a given coverage level had an equal magnitude for both Chi1 and Swe Figure 3.10A. This has two implications, first all *D* and $\hat{f}$ statistic estimates of 2X coverage or less may be overestimates, and second this effect is independent of the amount of introgression into the individual. The independence of the effect from the absolute amount of introgressed ancestry allows us to propose a single recalibration approach that is independent of the amount of admixture detected for all brown bear *D* and $\hat{f}$ statistics. It is not clear whether the same calibration applies to all species our is impacted by a species demographic history in ways other than the total amount of admixture detected.

Intriguingly down sampling the number of sites included from the haploidized sequence generated from the entire data set did not produce a bias Figure 3.10B. This demonstrates that any recalibration should be conducted on the basis of the number of reads in the sample rather than the coverage of the sample. That bias results from resampling of reads but not sites is also consistent with the source of the bias being the linkage of data by reads. If so, the samples in this study, which have, on average, shorter reads than the Chi1 and Swe samples (Cahill *et al.* 2015) may be less biased.

To correct the *D* and $\hat{f}$ estimates we reduce the *D* or $\hat{f}$ estimate by, the difference between the mean of the Chi1 simulations, with the number of reads nearest to the low coverage sample, from the Chi1 result from the whole Chi1 data set (Eq. 1).

$$D_{corrected} = D - \left(\overline{D}_{Chi1\ nearest\ read\ resampling} - D_{Chi1\ All\ data}\right) \quad (Eq.1)$$

We report the number of mapped reads per sample, and the correction used for each sample's D and $\hat{f}$ estimates in Table 3.8. Although the discovery of a previously unidentified bias in the D-statistic method is concerning the scale of the bias, up to ~2% introgressed ancestry does not significantly impact the conclusions of this study. Where applied, coverage recalibrated estimates are explicitly noted in the text.

Direction of gene flow

Neither the D nor the $\hat{f}$ statistic explicitly test for the direction of gene flow(Green *et al.* 2010; Durand *et al.* 2011). A significant D-statistic result indicates only an elevation in the frequency of shared derived alleles in conflict with the species tree expectation, but does not identify which individual was the recipient of introgression. To test whether our candidate hybrid populations were the recipients of gene flow, we tested for genomic regions in which the candidate hybrids exhibited low polar bear divergence and high brown bear divergence, as in(Green *et al.* 2010). We subdivided the reference genome into 1 Mb non-overlapping bins. For each bin we calculated the frequency of transversion differences between a candidate hybrid brown bear and unadmixed representatives of each parent species: polar bear, SAMN01057676 (Miller *et al.* 2012), and Swedish brown bear, SAMN03252407 (Cahill *et al.* 2015). We excluded bins with fewer than 10,000 informative sites to minimize stochastic noise in the result. We compared the results for the candidate hybrids to an analysis testing for polar bear introgression into a Finnish brown bear, SAMN02256315 (Liu *et al.* 2014) without detectable polar bear ancestry. (Figure 3.5).

Influence of selection against polar bear alleles in hybrids

To test whether the accumulation of weakly deleterious alleles in polar bears could be responsible for the decline in polar bear ancestry observed in the Irish brown bear population (Figure 3.2) we used a forward-simulation approach of

Harris and Nielsen (Harris & Nielsen 2016), and the simulator SLiM (Messer 2013). We assumed the same mutation rate, 1.3e-8 per site per generation, as in Liu *et al* 2014 (Liu *et al.* 2014) and that 70% of mutations in exons are nonsynonymous. We assumed nonsynonymous mutation in bears have the same gamma-shaped distribution of fitness effects that has been inferred for humans (Eyre-Walker *et al.* 2006). For genome structure information we used the polar bear reference genome and associated exon map (Liu *et al.* 2014). We assume recombination between scaffolds and a recombination rate of 1.35e-8 per site per generation within scaffolds, as in Liu *et al* 2014 (Liu *et al.* 2014). We neglect the fitness effects of mutations that occur outside exons, as in pervious applications of this method (Harris & Nielsen 2016).

To conduct our forward in time simulations we used a model of polar bear and brown bear population history we used a model based on the population model of Liu *et al* 2014 (Liu *et al.* 2014). We began with 17,797 generations of mutation accusation in an ancestral population N_e=10,000. Then polar and brown bear populations diverge into separate populations each N_e=10,000. At generation 31,894 the size of the polar bear population is reduced to N_e=5,000. At generation 58,678, we model the at least 20% introgression of polar bear ancestry into Irish brown bears (Figure 3.2, Table 3.5) as a single pulse of polar bear into brown bear introgression with 25% population replacement. Forward in time simulation continued until generation 60,000. We measured both the genetic load accumulated in polar bears prior to the admixture event (Figure 3.8) and the amount of polar bear ancestry persisting in the Irish brown bears after admixture (generations 58,678-60,000)(Figure 3.9).

Mitochondrial genome haplotyping of the Kunashir Island bears

To determine the mitochondrial haplotypes of the Kunashir Island brown bears, which were the only samples used herein that have not been typed previously, we used mia (Green *et al.* 2008) to map SeqPrep merged reads to a polar bear mitochondrial genome reference (NC_003428.1)(Delisle & Strobeck 2002). We called the consensus sequence for sites with at least 3× coverage and masked all other sites. We preformed a global alignment of each Kunashir mitochondrial haplotype to 39 previously published polar bear and brown bear mitochondrial haplotypes (Hirata *et al.* 2013), including at least one individual from each major mitochondrial clade of polar bears and brown bears, with MAFFT default parameters online version 7.245 (Katoh & Standley 2013). We constructed a neighbor joining tree (Saitou & Nei 1987) using the included MAFFT program from all conserved sites (16,288 bp for Kunashir 1 and 11,533 bp for Kunashir 2) where all sequences were represented under a Jukes-Cantor mutation model (Jukes & Cantor 1969). The Kunashir samples both fall into clade 3b, with previously published Kunashir brown bear mitochondrial sequences (Hirata *et al.* 2013).

Additional Figures and Tables

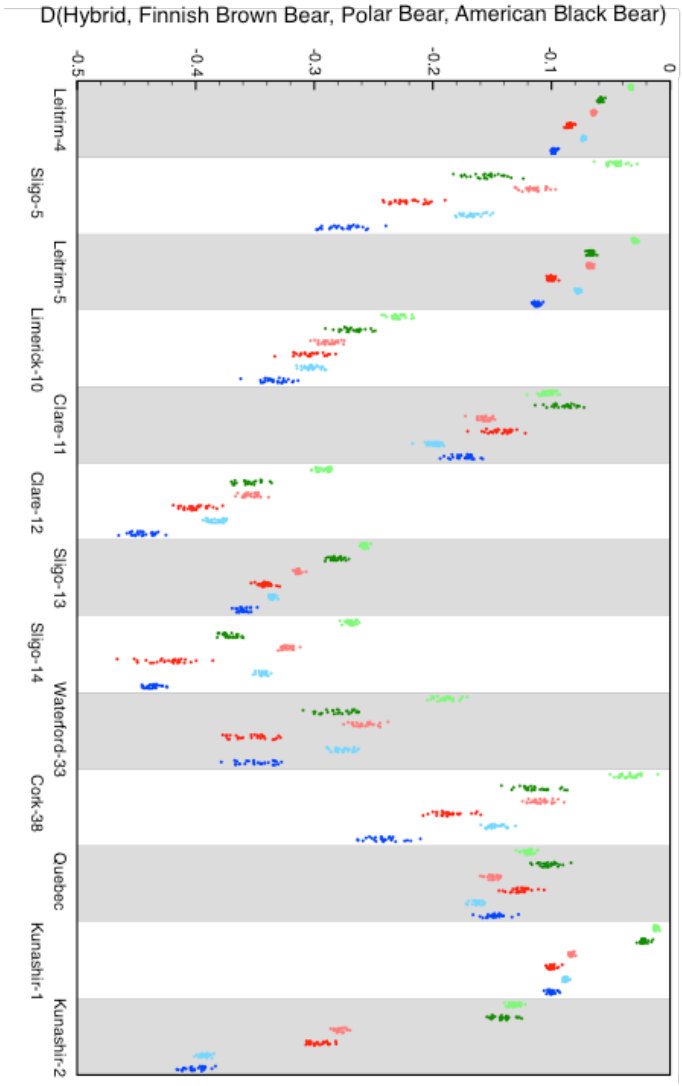


Figure 3.3 | Uncalibrated *D*-statistics calculated under different mapping and filtering criteria. *D*-statistics calculated for each of the Irish brown bear (P1) compared to a Finnish brown bear (P2)(SAMN02256315)(Liu *et al.* 2014) and each Polar bear (P3)(Miller *et al.* 2012; Cahill *et al.* 2013)(Table 3.3). An American black bear(Cahill *et al.* 2013) was used as an outgroup for all comparisons. Comparisons using all sites (light colored dots) are less different from zero than corresponding comparisons using only transversions (dark colored dots). Comparisons of Irish brown bears mapped to the polar bear reference(Li *et al.* 2011)(blue dots) have increased allele sharing with polar bear; samples mapped to the brown bear, SAMN02256314(Liu *et al.* 2014), (green dots) have decreased allele sharing with polar bear; and the union of reads mapped to either reference (red dots) is intermediate.

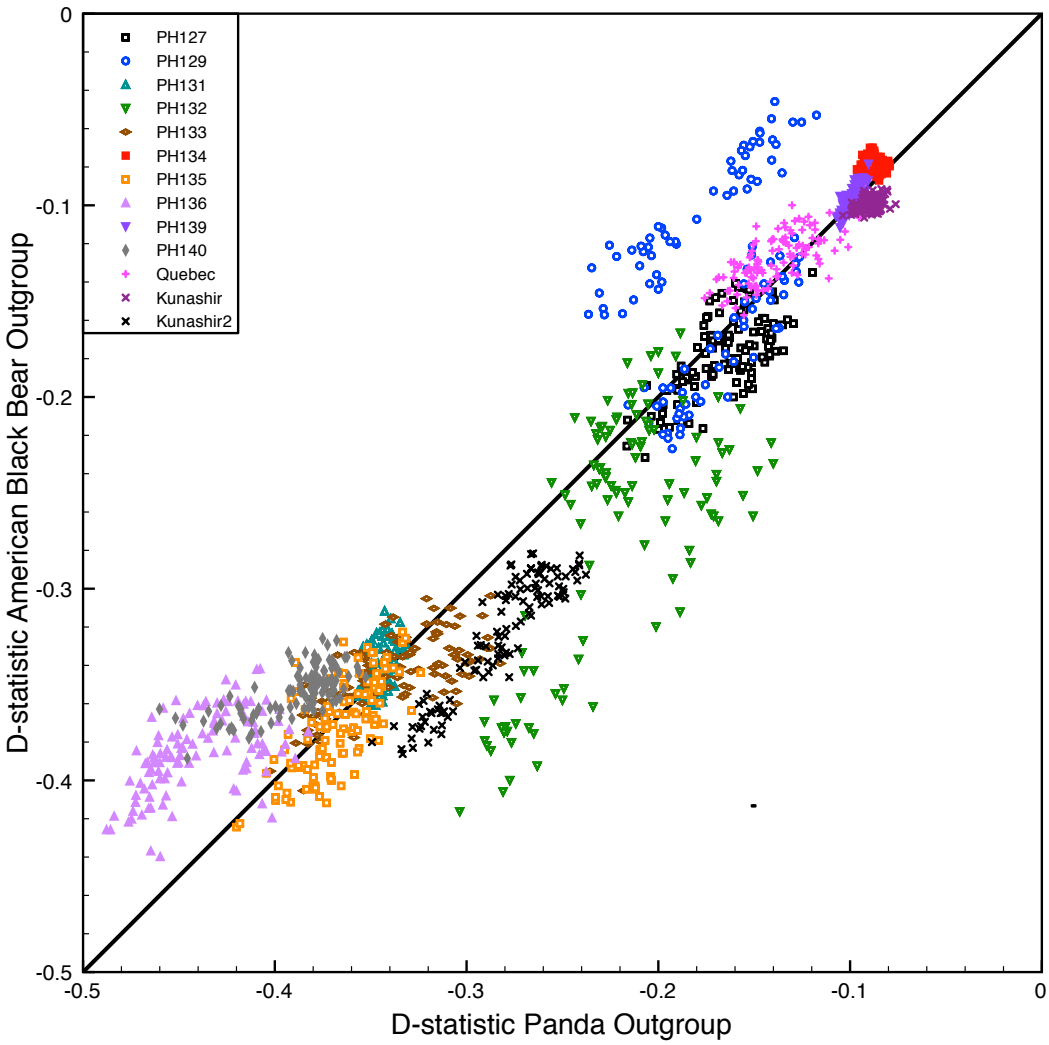


Figure 3.4 | The Influence of outgroup selection on uncalibrated *D*-statistic results. To test the impact of using different outgroups, we show the relationship between *D*-statistics calculated with a Giant Panda outgroup(Li *et al.* 2010) (x-axis) and an American black bear outgroup (y-axis). All comparisons are calculated from the union of reads mapped to the polar bear reference and the unadmixed brown bear pseudo-reference, and exclude transition sites. Each Irish bear (P1) is compared against 4 Fennoscandian brown bears (P2)(Liu *et al.* 2014; Cahill *et al.* 2015) and 28 Polar bears (P3)(Miller *et al.* 2012; Cahill *et al.* 2013) for a total of 112 comparisons per Irish bear using each outgroup.

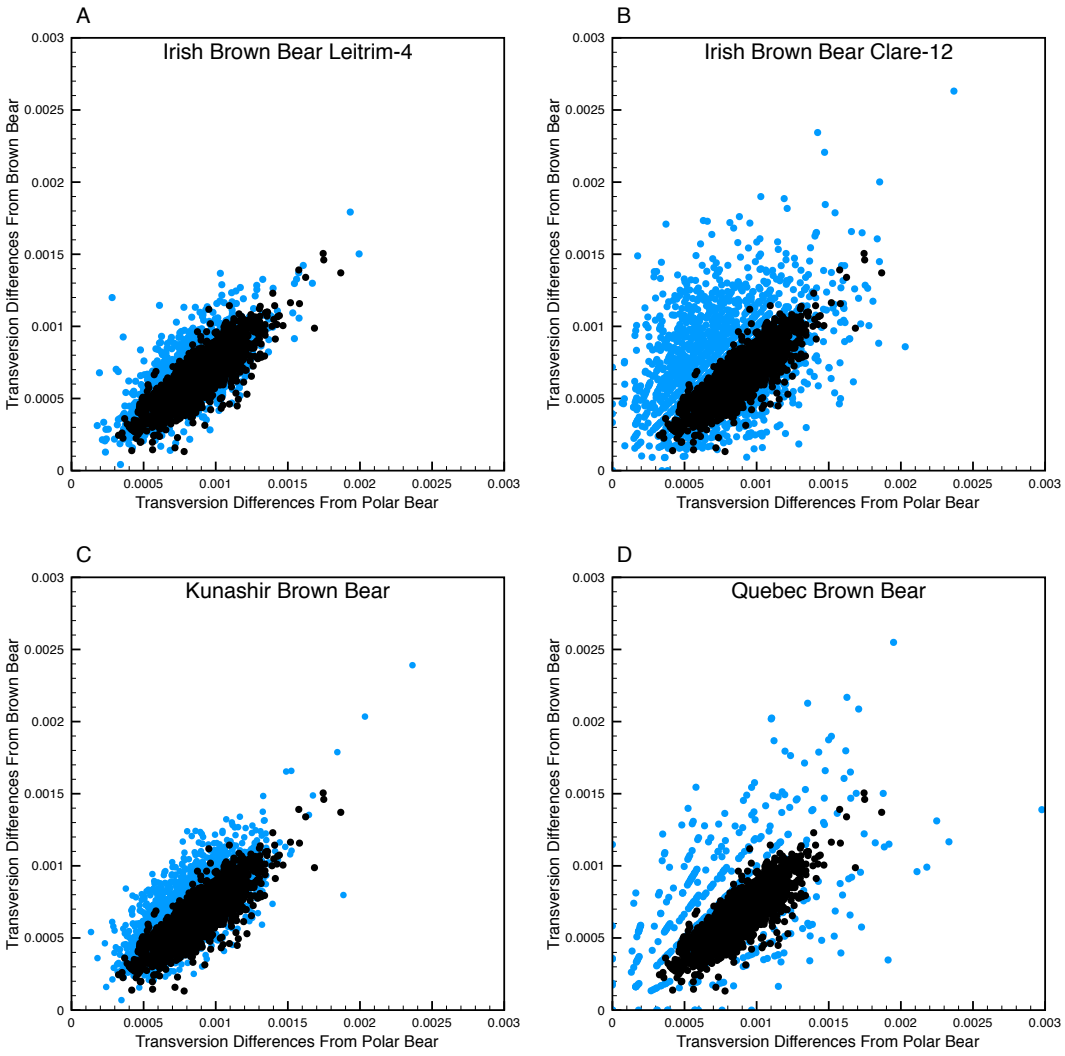


Figure 3.5 | Direction of gene flow. If the candidate hybrid brown bears are the recipients of introgression from polar bears we would expect them to contain genomic regions of low polar bear divergence and higher brown bear divergence. Here we show the distribution of divergence from polar bear and brown bear in 1Mb bins for the two highest coverage Irish bears (A and B), the higher coverage Kunashir bear (C) and the Quebec bear (D)(blue dots). We compared these to the result from a Finnish brown bear with no detectable polar bear ancestry (black dots). We find that all three candidate hybrid populations have an excess of regions of lower polar bear divergence than expected from the Finnish bear result, and the signal is much more pronounced in Clare-12 sample with the highest polar bear ancestry among the four samples tested. These result support brown bears as the recipients of polar bear introgression.

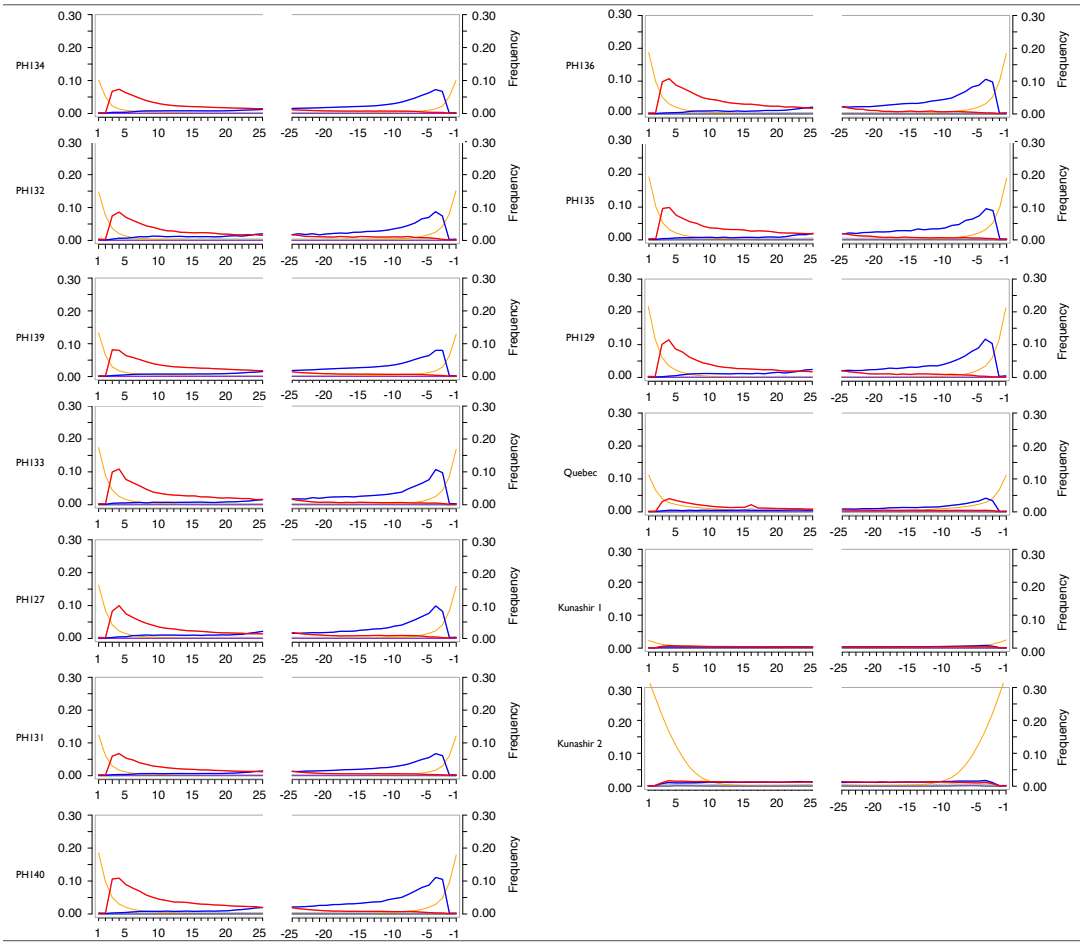


Figure 3.6 | Ancient DNA damage patterns inferred with mapDamage. Cytosine deamination damage (C to T) is diagnostic of ancient DNA. Cytosine deamination is characterized by an excess of thymine observations at the 5′ end of the reads (red line) and an excess of the reverse compliment, adenine, observed at the 3′ end of reads (blue line). We used mapDamage v2.0.5(Jónsson *et al.* 2013) to visualize the damage pattern of mapped reads. The ancient samples reflect patterns typical of ancient DNA beginning at the 3rd position in the read. This appears to be the product of damaged bases in the 1st or 2nd position in the read being soft clipped by bowtie2(Langmead & Salzberg 2012) as indicated by the corresponding increase in soft clipping (orange line).

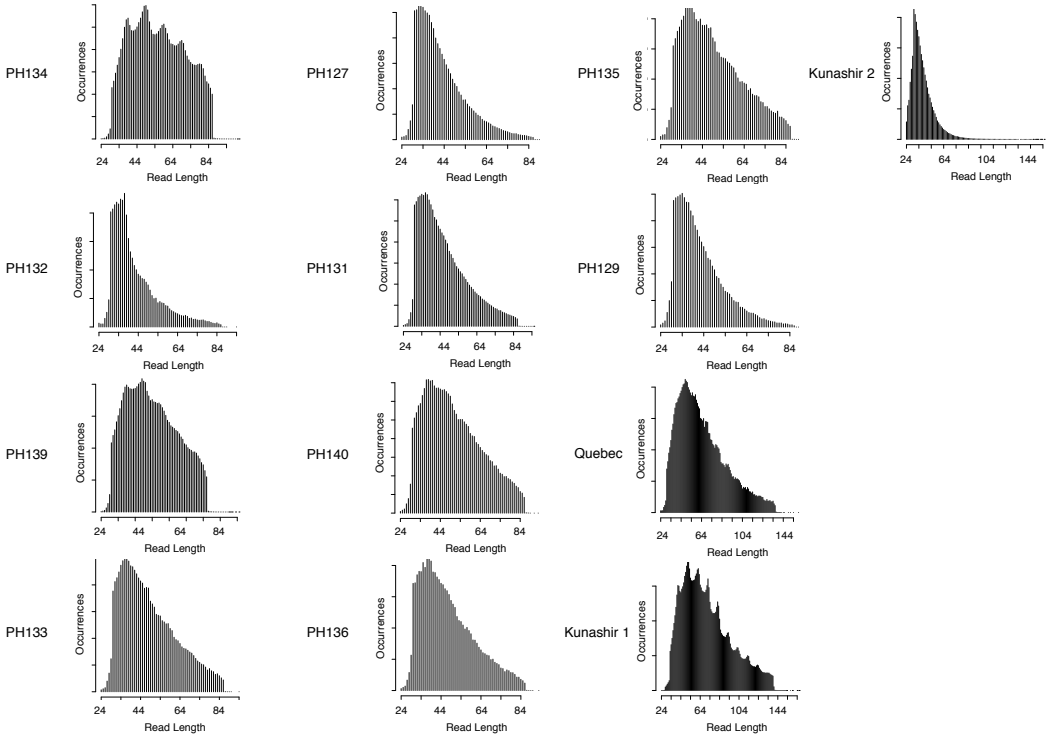


Figure 3.7 | Sequenced DNA fragment length distributions. Length distributions for SeqPrep (St John 2011) merged reads from the 10 HiSeq-sequenced Irish brown bears and the MiSeq-sequenced Kunashir and Quebec brown bears. These distributions were visualized with mapDamgae v2.0.5(Jónsson *et al.* 2013) and are reported as mapped read length after soft clipping by bowtie2 during mapping(Langmead & Salzberg 2012). The difference between the Kunashir/Quebec and Irish brown bear maximum insert lengths are due to the use of 2× 75 and 2× 50 paired-end sequencing, respectively. Read lengths shorter than 30 bp are the result of bowtie2 soft clipping.

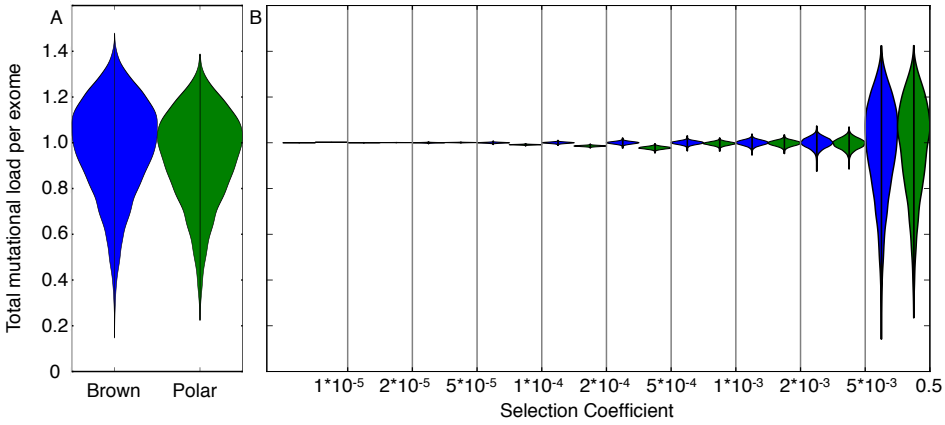


Figure 3.8 | Simulated selection coefficient distributions. These violin plot shows the distribution of fitness within the respective brown bear and polar bear populations one generation before admixture, normalized such that the median brown bear (blue) has fitness 1. At the whole individual fitness level (A) there is a subtle shift in the polar bear population (green) toward lower fitness values, with a median fitness of 96% indicating that overall polar bears have a slightly greater genetic load. We also tested the relative selective impact of alleles of varying selective coefficients (B).

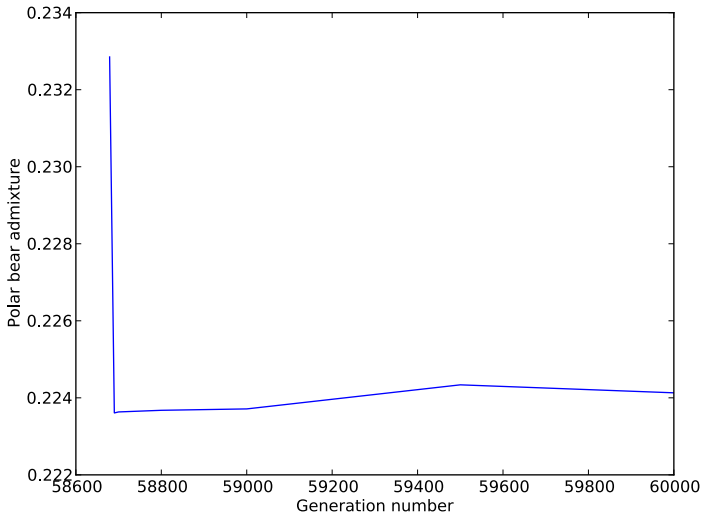
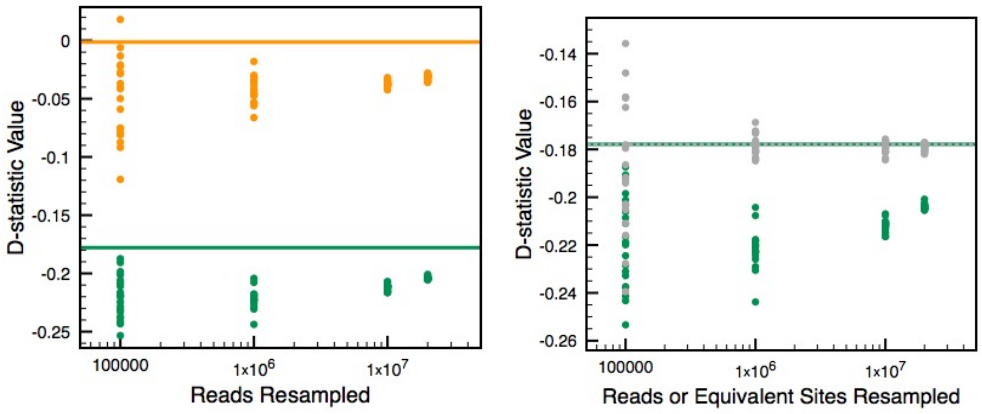


Figure 3.9 | Simulated estimate of polar bear ancestry lost from Irish brown bears as a result of selection against introgressed deleterious alleles. Forward in time population simulations of polar bears and brown bears suggest that selection against polar bear genetic load is unlikely to be the principal mechanism for reducing polar bear ancestry in Irish brown bears. Here we show the amount of polar bear ancestry retained in our population simulations (see Methods) assuming that the admixed Irish population did not receive any outside gene flow after polar bear admixture. After the simulated admixture pulse replaces 25% of the brown bear population with polar bears, selection against weakly deleterious polar bear alleles reduced this

ancestry fraction to 22.4%. However, the polar bear ancestry fraction stabilized at this level and did not decline further, indicating that selection against polar bears' greater genetic load is not sufficient to explain the declines in polar bear ancestry observed in the Irish brown bears (Figure 3.2).



(A) (B)

Figure 3.10 | D-statistic values from random downsampling. Here we show the impact of decreasing the amount of available data on D-statistic analysis. In (A) we show the impact of random downsampling of sequencing reads on two multi-fold coverage brown bears, a 9% polar bear ABC island bear (Chi1, green) (SRX795188)(Cahill *et al.* 2015) and an unadmixed Swedish brown bear (Swe, orange) (SRX796442)(Cahill *et al.* 2015). All D-statistic comparisons are made against another unadmixed Swedish brown bear [SAMN02256314] (Liu *et al.* 2014)and an Alaskan brown bear (SRX156102)(Miller *et al.* 2012). In (B) we show Chi1 results for read resampling (green) and site resampling (grey). In both subfigures horizontal lines indicate the whole data-set D-statistic value. These results show that read downsampling leads to increased gene flow detected between the downsampled individual and the polar bear (more negative D statistic), but site downsampling does not introduce a bias.

Table 3.1 | Basic Sample Information

Sample ID	Museum ID	Tissue	Locality	mtDNA clade	Sample age (Cal. yr BP)
PH134: Leitrim-4	F21458	Tooth: molar	Poll na mBéar Cave, Glenade, Co. Leitrim, Ireland	1(Edwards et al. 2011)	3,791 ± 50
PH132: Sligo-5	F21439/85	Bone: metatarsal	Polldownin Cave, Co. Sligo, Ireland	1(Edwards et al. 2011)	4,687 ± 94
PH139: Leitrim-5	F21456/8	Bone: mandible	Poll na mBéar Cave, Glenade, Co. Leitrim, Ireland	1(Edwards et al. 2011)	5,180 ± 93
PH133: Limerick-10	F21749	Bone: calcaneum	Red Cellar Cave, Co. Limerick, Ireland	2(Edwards et al. 2011)	9,702 ± 96
PH127: Clare-11	F21752	Bone: vertebrae	Newhall Cave, Edenvale, Co. Clare, Ireland	2(Edwards et al. 2011)	11,391 ± 125
PH131: Clare-12	F21750	Bone: calcaneum	Newhall Cave, Edenvale, Co. Clare, Ireland	2(Edwards et al. 2011)	12,393 ± 199
PH140: Sligo-13	F21748	Bone: femur	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland	2(Edwards et al. 2011)	13,219 ± 105
PH136: Sligo-14	F21119	Bone: humerus	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland.	2(Edwards et al. 2011)	13,683 ± 182
PH135: Waterford-33	F21753	Bone: astragalus	Shandon Cave, Dungarvan, Co. Waterford, Ireland	2(Edwards et al. 2011)	33,067 ± 625
PH129: Cork-38	F21751	Bone: humerus	Mammoth Cave, Castlepook, Co. Cork, Ireland	1(Edwards et al. 2011)	37,970 ± 613
TF106: Quebec	MPEP 82.1	Bone: metatarsal	Saint-Nicolas, Quebec, Canada	4(Harington et al. 2014)	11,279 ± 30
JK022: Kunashir1	N/A	Skin	Kunashir Island, Sakhalin, Russian Federation	3b	Historic (no date)
JK328: Kunashir2	N/A	Skin	Kunashir Island, Sakhalin, Russian Federation	3b	Historic (no date)

Sample information including: sample origin, age and mitochondrial clade.

Table 3.2 | Mapping Statistics

Sample ID	Genomic coverage (×)	% reads mapped to either reference genome	% reads mapped to the polar bear reference genome	% reads mapped to the brown bear pseudo-reference genome	% reads mapped exclusively to the polar bear reference	% reads mapped exclusively to the brown bear pseudo-reference
PH134: Leitrim-4	0.1233	15.86%	15.44%	15.36%	3.16%	2.65%
PH132: Sligo-5	0.0016	0.46%	0.44%	0.39%	15.16%	4.69%
PH139: Leitrim-5	0.0832	10.36%	10.08%	10.06%	2.89%	2.69%
PH133: Limerick-10	0.0034	0.75%	0.73%	0.72%	4.04%	3.31%
PH127: Clare-11	0.0043	0.70%	0.67%	0.67%	4.29%	4.19%
PH131: Clare-12	0.0172	3.10%	3.02%	3.00%	3.12%	2.69%
PH140: Sligo-13	0.0058	0.75%	0.73%	0.72%	4.44%	3.44%
PH136: Sligo-14	0.0026	0.53%	0.51%	0.50%	4.59%	3.89%
PH135: Waterford-33	0.0016	0.33%	0.31%	0.31%	7.36%	4.52%
PH129: Cork-38	0.002	0.48%	0.46%	0.46%	4.32%	3.21%
TF106: Quebec	0.0061	0.64%	0.62%	0.61%	4.49%	3.35%
JK022: Kunashir1	0.0588	76.86%	75.01%	74.83%	2.64%	2.40%
JK328: Kunashir2	0.0093	12.05%	11.24%	11.41%	7.21%	5.61%

Here we show the read mapping rates for the 13 samples used in this study, and the resulting coverage from the union of mappings to both the polar bear and brown bear references.

Table 3.3 – Sample List

Sample	Species	Location	Short Read Accession	Original Publication	Museum Accession
PH134: Leitrim-4	*Ursus arctos*	Poll na mBéar Cave, Glenade, Co. Leitrim, Ireland	SAMN04884477	This Study	F21458
PH132: Sligo-5	*Ursus arctos*	Polldownin Cave, Co. Sligo, Ireland	SAMN04884474	This Study	F21439/85
PH139: Leitrim-5	*Ursus arctos*	Poll na mBéar Cave, Glenade, Co. Leitrim, Ireland	SAMN04884476	This Study	F21456/8
PH133: Limerick-10	*Ursus arctos*	Red Cellar Cave, Co. Limerick, Ireland	SAMN04884479	This Study	F21749
PH127: Clare-11	*Ursus arctos*	Newhall Cave, Edenvale, Co. Clare, Ireland	SAMN04884482	This Study	F21752
PH131: Clare-12	*Ursus arctos*	Newhall Cave, Edenvale, Co. Clare, Ireland	SAMN04884480	This Study	F21750
PH140: Sligo-13	*Ursus arctos*	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland	SAMN04884478	This Study	F21748
PH136: Sligo-14	*Ursus arctos*	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland.	SAMN04884471	This Study	F21119
PH135: Waterford-33	*Ursus arctos*	Shandon Cave, Dungarvan, Co. Waterford,	SAMN04884483	This Study	F21753

		Ireland			
PH129: Cork-38	*Ursus arctos*	Mammoth Cave, Castlepook, Co. Cork, Ireland	SAMN04884481	This Study	F21751
TF106: Quebec	*Ursus arctos*	Saint-Nicolas, Quebec, Canada	SAMN04884484	This Study	MPEP 82.1
JK022: Kunashir1	*Ursus arctos*	Kunashir Island, Sakhalin, Russian Federation	SAMN04884485	This Study	N/A
JK328: Kunashir2	*Ursus arctos*	Kunashir Island, Sakhalin, Russian Federation	SAMN04884486	This Study	N/A
WH1	*Ursus maritimus*	Manitoba, Canada	SRX265154	(Cahill *et al.* 2013)	X3249106A
WH2	*Ursus maritimus*	Manitoba, Canada	SRX265434	(Cahill *et al.* 2013)	X3312806A
SB	*Ursus maritimus*	Barrow, AK, USA	SRX265435	(Cahill *et al.* 2013)	990083KD
CS	*Ursus maritimus*	Gambell, AK, USA	SRX265436	(Cahill *et al.* 2013)	940090KB
NB	*Ursus maritimus*	NWT, Canada	SRX265452	(Cahill *et al.* 2013)	X3292306A
WI	*Ursus maritimus*	Wrangel Island, Russia	SRX265453	(Cahill *et al.* 2013)	Unknown
LS	*Ursus maritimus*	Cornwallis Island, Canada	SRX265454	(Cahill *et al.* 2013)	512133
PB1	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155945	(Miller *et al.* 2012)	N23531
PB2	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155946	(Miller *et al.* 2012)	N23604
PB3	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155947	(Miller *et al.* 2012)	N23719
PB4	*Ursus maritimus*	Spitsbergen, Svalbard,	SRX155948	(Miller *et al.* 2012)	N23917

		Norway			
PB5	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155949	(Miller *et al.* 2012)	N23949
PB6	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155950	(Miller *et al.* 2012)	N26028
PB7	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX156156-63	(Miller *et al.* 2012)	N7773
PB8	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155951	(Miller *et al.* 2012)	N7986
PB9	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155953	(Miller *et al.* 2012)	N23985
PB10	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155954	(Miller *et al.* 2012)	N23997
PB11	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155955	(Miller *et al.* 2012)	N26029
PB12	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155956	(Miller *et al.* 2012)	N26030
PB13	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155957	(Miller *et al.* 2012)	N23355
PB14	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155958	(Miller *et al.* 2012)	N23379
PB15	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155959	(Miller *et al.* 2012)	N23694
PB16	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155960	(Miller *et al.* 2012)	N23797
PB17	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155961	(Miller *et al.* 2012)	N26024
PB18	*Ursus maritimus*	Spitsbergen, Svalbard, Norway	SRX155962	(Miller *et al.* 2012)	N26025
AK1	*Ursus maritimus*	Barrow, AK, USA	SRX156102	(Miller *et al.* 2012)	542
AK2	*Ursus maritimus*	Diomede, AK, USA	SRX156103	(Miller *et al.* 2012)	562

AK3	*Ursus maritimus*	Barrow, AK, USA	SRX156104	(Miller *et al.* 2012)	574
AK4	*Ursus maritimus*	Diomede, AK, USA	SRX156105	(Miller *et al.* 2012)	2368
AK5	*Ursus maritimus*	Savoonga, AK, USA	SRX156106	(Miller *et al.* 2012)	651
Swe	*Ursus arctos*	Dalarna, Sweden	SAMN03252406	(Cahill *et al.* 2015)	
SJS01	*Ursus arctos*	Slakka, Jokkmokk, Sweden	SAMN02256314	(Liu *et al.* 2014)	20105373
OFS01	*Ursus arctos*	Östanvik, Furudal, Sweden	SAMN02256313	(Liu *et al.* 2014)	20105434
RF01	*Ursus arctos*	Ruokolahti, Finland	SAMN02256315	(Liu *et al.* 2014)	7429
Uam	*Ursus americanus*	Pennsylvania, USA	SRX265459	(Cahill *et al.* 2013)	n/a

Table 3.4 | D-statistics Results.

Sample	Reference and Analysis Type	*D* calibrated	weighted block jackknife	Z score
Kunashir	2R-TV	-0.0515	0.0146	**3.5150**
Kunashir2	2R-TV	-0.2892	0.0387	**7.4739**
Clare-11	2R-TV	-0.1203	0.0459	2.6193
Cork-38	2R-TV	-0.1638	0.0632	2.5900
Clare-12	2R-TV	-0.2980	0.0239	**12.4425**
Sligo-5	2R-TV	-0.1095	0.0777	1.4094
Limerick-10	2R-TV	-0.3045	0.0484	**6.2928**
Leitrim-4	2R-TV	-0.0399	0.0111	**3.5804**
Waterford-33	2R-TV	-0.3145	0.0691	**4.5499**
Sligo-14	2R-TV	-0.3879	0.0462	**8.3952**
Leitrim-5	2R-TV	-0.0477	0.0120	**3.9820**
Sligo-13	2R-TV	-0.3253	0.0692	**4.7009**
Quebec	2R-TV	-0.0753	0.0425	1.7720
Kunashir	2R-all	-0.0367	0.0107	**3.4284**
Kunashir2	2R-all	-0.2567	0.0213	**12.0760**
Clare-11	2R-all	-0.1095	0.0261	**4.2031**
Cork-38	2R-all	-0.0715	0.0340	2.1001
Clare-12	2R-all	-0.2652	0.0173	**15.3367**
Sligo-5	2R-all	-0.0860	0.0393	2.1868
Limerick-10	2R-all	-0.2648	0.0272	**9.7282**
Leitrim-4	2R-all	-0.0215	0.0083	2.5975
Waterford-33	2R-all	-0.2423	0.0375	**6.4575**
Sligo-14	2R-all	-0.3188	0.0297	**10.7347**
Leitrim-5	2R-all	-0.0186	0.0085	2.1774
Sligo-13	2R-all	-0.2820	0.0235	**11.9821**
Quebec	2R-all	-0.1000	0.0243	**4.1122**
Kunashir	PB-all	-0.0458	0.0101	**4.5551**
Kunashir2	PB-all	-0.3540	0.0211	**16.7503**
Clare-11	PB-all	-0.1522	0.0257	**5.9293**
Cork-38	PB-all	-0.1094	0.0334	**3.2713**
Clare-12	PB-all	-0.2853	0.0167	**17.0864**
Sligo-5	PB-all	-0.1257	0.0389	**3.2306**
Limerick-10	PB-all	-0.2879	0.0259	**11.1048**
Leitrim-4	PB-all	-0.0288	0.0077	**3.7527**
Waterford-33	PB-all	-0.2464	0.0346	**7.1272**
Sligo-14	PB-all	-0.3402	0.0283	**12.0088**
Leitrim-5	PB-all	-0.0287	0.0081	**3.5534**

Sligo-13	PB-all	-0.3056	0.0228	**13.4327**
Quebec	PB-all	-0.1136	0.0235	**4.8307**
Kunashir	PB-TV	-0.0600	0.0139	**4.3056**
Kunashir2	PB-TV	-0.3846	0.0377	**10.2056**
Clare-11	PB-TV	-0.1584	0.0460	**3.4428**
Cork-38	PB-TV	-0.2055	0.0618	**3.3252**
Clare-12	PB-TV	-0.3177	0.0232	**13.7097**
Sligo-5	PB-TV	-0.1516	0.0769	**1.9707**
Limerick-10	PB-TV	-0.3290	0.0456	**7.2161**
Leitrim-4	PB-TV	-0.0490	0.0095	**5.1780**
Waterford-33	PB-TV	-0.3154	0.0621	**5.0775**
Sligo-14	PB-TV	-0.4049	0.0439	**9.2320**
Leitrim-5	PB-TV	-0.0611	0.0108	**5.6550**
Sligo-13	PB-TV	-0.3730	0.0343	**10.8899**
Quebec	PB-TV	-0.0914	0.0403	2.2689
Kunashir	BB-all	0.0317	0.0101	**3.1421**
Kunashir2	BB-all	-0.0848	0.0252	**3.3722**
Clare-11	BB-all	-0.0453	0.0268	1.6870
Cork-38	BB-all	0.0020	0.0349	0.0583
Clare-12	BB-all	-0.2038	0.0174	**11.7308**
Sligo-5	BB-all	-0.0124	0.0396	0.3125
Limerick-10	BB-all	-0.2148	0.0276	**7.7749**
Leitrim-4	BB-all	0.0134	0.0077	1.7362
Waterford-33	BB-all	-0.1634	0.0365	**4.4780**
Sligo-14	BB-all	-0.2522	0.0304	**8.2851**
Leitrim-5	BB-all	0.0213	0.0081	2.6348
Sligo-13	BB-all	-0.2332	0.0236	**9.8970**
Quebec	BB-all	-0.0665	0.0243	2.7430
Kunashir	BB-TV	0.0205	0.0140	1.4652
Kunashir2	BB-TV	-0.1109	0.0465	2.3863
Clare-11	BB-TV	-0.0603	0.0484	1.2473
Cork-38	BB-TV	-0.1015	0.0656	1.5476
Clare-12	BB-TV	-0.2316	0.0240	**9.6539**
Sligo-5	BB-TV	-0.0731	0.0789	0.9256
Limerick-10	BB-TV	-0.2728	0.0470	**5.8004**
Leitrim-4	BB-TV	-0.0074	0.0096	0.7662
Waterford-33	BB-TV	-0.2481	0.0663	**3.7409**
Sligo-14	BB-TV	-0.3258	0.0469	**6.9521**
Leitrim-5	BB-TV	-0.0140	0.0109	1.2842

Sligo-13	BB-TV	-0.3110	0.0360	**8.6286**
Quebec	BB-TV	-0.0336	0.0423	0.7946

D-statistics are calculated *D*(Sample, Swe, AK1, Uam)(see Table 3.3). Z-scores are calculated from low coverage calibrated D-statistics (column 3). Other polar bears and brown bears not shown due to space constraints.

Table 3.5 $\hat{f}$ results

Sample	Reference and Analysis Type	$\hat{f}$ calibrated	weighted block jackknife	Z-score
Clare-11	2R-TV	4.58%	0.0242	2.9094
Cork-38	2R-TV	2.91%	0.0338	1.5837
Clare-12	2R-TV	14.75%	0.0160	**10.6967**
Sligo-5	2R-TV	5.23%	0.0384	1.9987
Limerick-10	2R-TV	12.85%	0.0300	**5.0995**
Leitrim-4	2R-TV	0.35%	0.0058	**4.7668**
Waterford-33	2R-TV	16.65%	0.0404	**4.7229**
Sligo-14	2R-TV	21.59%	0.0327	**7.3574**
Leitrim-5	2R-TV	0.11%	0.0067	**3.7408**
Sligo-13	2R-TV	17.58%	0.0256	**7.8377**
Kunashir	2R-TV	2.10%	0.0068	**6.6573**
Kunashir2	2R-TV	13.55%	0.0178	**8.9385**
Quebec	2R-TV	4.07%	0.0233	2.7935
Clare-11	2R-all	4.49%	0.0152	**4.5645**
Cork-38	2R-all	2.14%	0.0212	2.1625
Clare-12	2R-all	14.29%	0.0127	**13.1181**
Sligo-5	2R-all	5.38%	0.0239	**3.2724**
Limerick-10	2R-all	12.23%	0.0181	**8.0945**
Leitrim-4	2R-all	-0.49%	0.0044	**4.3138**
Waterford-33	2R-all	11.17%	0.0243	**5.6039**
Sligo-14	2R-all	18.39%	0.0216	**9.6465**
Leitrim-5	2R-all	-1.11%	0.0046	2.8077
Sligo-13	2R-all	15.59%	0.0185	**9.7346**
Kunashir	2R-all	1.43%	0.0049	**7.7334**
Kunashir2	2R-all	13.21%	0.0097	**16.0915**
Quebec	2R-all	4.69%	0.0133	**5.3618**
Clare-11	BB-all	1.45%	0.0151	2.5887
Cork-38	BB-all	-1.85%	0.0211	0.2831
Clare-12	BB-all	10.40%	0.0120	**10.7027**
Sligo-5	BB-all	1.41%	0.0230	1.6762
Limerick-10	BB-all	8.60%	0.0175	**6.3293**
Leitrim-4	BB-all	-0.96%	0.0040	**3.6281**

Waterford-33	BB-all	7.25%	0.0225	**4.3043**
Sligo-14	BB-all	14.26%	0.0210	**7.9435**
Leitrim-5	BB-all	-1.41%	0.0045	2.2164
Sligo-13	BB-all	11.99%	0.0177	**8.1647**
Kunashir	BB-all	-1.87%	0.0044	1.1876
Kunashir2	BB-all	2.15%	0.0095	**4.7724**
Quebec	BB-all	2.84%	0.0130	**4.0774**
Clare-11	BB-TV	2.44%	0.0233	2.0957
Cork-38	BB-TV	-0.01%	0.0336	0.7253
Clare-12	BB-TV	10.79%	0.0147	**8.9746**
Sligo-5	BB-TV	3.37%	0.0366	1.5906
Limerick-10	BB-TV	11.14%	0.0278	**4.8850**
Leitrim-4	BB-TV	-0.05%	0.0049	**4.7711**
Waterford-33	BB-TV	13.51%	0.0367	**4.3469**
Sligo-14	BB-TV	16.93%	0.0312	**6.2125**
Leitrim-5	BB-TV	-0.10%	0.0058	**3.9594**
Sligo-13	BB-TV	13.97%	0.0244	**6.7351**
Kunashir	BB-TV	-1.36%	0.0061	1.7128
Kunashir2	BB-TV	2.17%	0.0179	2.5593
Quebec	BB-TV	1.27%	0.0229	1.6220
Clare-11	PB-all	6.96%	0.0159	**5.9157**
Cork-38	PB-all	4.55%	0.0220	**3.1778**
Clare-12	PB-all	15.88%	0.0129	**14.1883**
Sligo-5	PB-all	7.80%	0.0248	**4.1318**
Limerick-10	PB-all	13.77%	0.0180	**8.9895**
Leitrim-4	PB-all	1.01%	0.0042	**8.1930**
Waterford-33	PB-all	12.29%	0.0237	**6.2133**
Sligo-14	PB-all	20.46%	0.0228	**10.0622**
Leitrim-5	PB-all	0.95%	0.0047	**7.1358**
Sligo-13	PB-all	17.35%	0.0188	**10.5042**
Kunashir	PB-all	5.29%	0.0130	**5.9091**
Kunashir2	PB-all	1.86%	0.0048	**8.9491**
Quebec	PB-all	20.18%	0.0128	**17.6271**
Clare-11	PB-TV	7.19%	0.0252	**3.8256**
Cork-38	PB-TV	5.91%	0.0353	2.3704
Clare-12	PB-TV	16.23%	0.0162	**11.4673**
Sligo-5	PB-TV	7.31%	0.0399	2.4427
Limerick-10	PB-TV	15.03%	0.0289	**6.0383**
Leitrim-4	PB-TV	1.81%	0.0051	**8.1743**

Waterford-33	PB-TV	17.62%	0.0385	**5.2057**
Sligo-14	PB-TV	23.36%	0.0317	**8.1366**
Leitrim-5	PB-TV	2.07%	0.0060	**7.4509**
Sligo-13	PB-TV	19.17%	0.0248	**8.7282**
Kunashir	PB-TV	2.15%	0.0067	**6.8339**
Kunashir2	PB-TV	21.34%	0.0220	**10.8031**
Quebec	PB-TV	5.11%	0.0223	**3.3906**

$\hat{f}$ statistics are calculated $\hat{f}$(OFS01, Sample, AK1, WH1, Uam)(see Table 3.3). Z-scores are calculated from low coverage calibrated $\hat{f}$ statistics (column 3). Significant Z-scores (Z>3) are bold. Other polar bears and brown bears not shown due to space constraints.

Table 3.6 | Preliminary Assessment of Irish brown bear bones to identify candidates for deeper sequencing.

Sample ID	Selected for Further Sequencing	Sample Age uncalibrated and accession	Percentage of Reads Mapped to polar bear reference genome with bwa aln	Percentage of Non-redundant Reads retained after samtools rmdup	Sample Location
F16010	No	8880+/- 90 bp (OxA-3714)(Woodman *et al.* 1997)	0.05%	99.03%	Derrykeel Bog, Co. Offaly, Ireland
F21104	No	10650 +/- 100 bp (OxA-3704)	0.18%	99.53%	Red Cellar Cave, Co. Limerick, Ireland
F21119	Yes	11920 +/- 85 bp (OxA-3706)	0.42%	99.79%	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland
F21160	No	32430 +/- 670 bp (OxA-4245)	0.09%	98.16%	Shandon Cave, Co. Waterford, Ireland
F21168	No	35570 +/- 1100 bp (OxA-4252)(Woodman *et al.* 1997)	0.09%	89.03%	Ballynamintra Cave, Co. Waterford, Ireland.
F21439/85	Yes	4136 +/- 37 bp (UB-6704)	0.41%	98.76%	Polldownin Cave, Co. Sligo, Ireland
F21455/43	No	2956 +/- 33 bp (UB-6705)	0.22%	97.52%	Poll na mBéar Cave, Co. Leitrim, Ireland
F21456/8	Yes	4520 +/- 37 bp (UB-6697)	10.73%	99.84%	Poll na mBéar Cave, Co. Leitrim,

					Ireland
F21458	Yes	3517 +/- 31 bp (OxA-15479)	17.04%	99.92%	Poll na mBéar Cave, Co. Leitrim, Ireland
F21748	Yes	11460 +/- 57 bp (UB-6698)	0.63%	99.25%	Plunkett Cave, Kesh Corran, Co. Sligo, Ireland
F21749	Yes	8719 +/- 48 bp (UB-6699)	0.71%	99.95%	Red Cellar Cave, Co. Limerick, Ireland
F21750	Yes	10495 +/- 51 bp (UB-6700)	2.97%	99.96%	Edenvale Cave, Co. Clare, Ireland
F21751	Yes	32648 +/- 245 bp (UB-6701)	0.40%	99.96%	Castlepook Cave, Co. Cork, Ireland
F21752	Yes	9946 +/- 53 bp (UB-6702)	0.63%	99.94%	Newhall Cave, Edenvale, Co. Clare, Ireland
F21753	Yes	28390 +/- 177 bp (UB-6703)	0.26%	99.80%	Shandon Cave, Co. Waterford, Ireland

Preliminary assessment of Irish brown bear DNA preservation derived from barcoded pooled sequencing of Irish brown bears using Illumina MiSeq version 3 chemistry. We selected ten samples for further sequencing based on endogenous DNA content, library complexity and sample age. Endogenous DNA content is estimated based on the proportion of reads mapped to the polar bear reference genome(Li *et al.* 2011) using bwa aln(Li & Durbin 2010). Library complexity is estimated via the fraction of unique mapped reads retained after samtools rmdup(Li *et al.* 2009). All samples are from the collection of NMING (National Museum of Ireland)

Table 3.7 | Sequencing read mapping parameter testing.

Method Used	false-negative	false-positive
bowtie2 --very-fast-local	15.5%	1.1%
bowtie2 --fast-local	11.0%	1.9%
bowtie2 --sensitive-local	4.7%	3.1%
bowtie2 --very-sensitive-local	4.7%	3.1%
bowtie2 --very-fast-global	17.3%	0.3%
bowtie2 --fast-global	17.3%	0.3%
bowtie2 --sensitive-global	12.6%	0.3%
bowtie2 --very-sensitive-global	10.2%	0.3%
bwa aln	17.4%	0.2%
bwa aln -n 0.01	13.2%	0.2%
bwa aln -n 0.001	10.0%	0.3%

bwa aln -n 0.0001	9.9%	0.3%
bwa mem -T1	3.2%	3.9%
bwa mem	6.3%	2.9%
bowtie2 --sensitive-global, map quality at least 30	45.9%	0.2%
bwa mem, map quality at least 30	6.5%	2.7%
bowtie2 --sensitive-local, map quality at least 25	5.3%	1.6%
bowtie2 --sensitive-local, map quality at least 30	6.1%	0.8%
bowtie2 --sensitive-local -N1, map quality at least 25	2.1%	1.8%
bowtie2 --sensitive-local -N 1, map quality at least 30	3.9%	1.1%
bowtie2 --sensitive-local -N 1 --mp 5, map quality at least 30	2.6%	1.3%
bowtie2 --sensitive-local -N 1 --mp 4, map quality at least 30	1.6%	1.3%
bowtie2 --sensitive-local -N 1 --mp 3, map quality at least 30	1.5%	1.6%
bowtie2 --sensitive-local --mp 4, map quality at least 30	5.2%	1.0%

To identify appropriate parameters for whole genome alignment we mapped reads from an Irish brown bear Leitrim-5 to a polar bear reference mitochondrial sequence(Delisle & Strobeck 2002)(NC_003428.1). We compared these whole genome approaches results to the mitochondrial alignment program mia seeking a parameter set that minimized both the number of reads mapped by mia but not the whole genome alignment program (false negatives) and the number of reads mapped by a whole genome alignment program not mapped by mia (false positives). The parameter set selected for this study is highlighted.

Table 3.8 | D and f statistic recalibration

Sample ID	Mapped Reads	Downsampling Used for Correction	D-statistic Correction Factor (Arbitrary Units)	f estimator Correction Factor (% Polar Bear)
PH134: Leitrim-4	4,816,004	1 million reads	-0.044865683	-2.40%
PH132: Sligo-5	80,762	100 thousand reads	-0.043369463	-2.45%
PH139: Leitrim-5	3,348,544	1 million reads	-0.044865683	-2.40%
PH133: Limerick-10	150,831	100 thousand reads	-0.043369463	-2.45%
PH127: Clare-11	218,659	100 thousand reads	-0.043369463	-2.45%
PH131: Clare-12	814,831	1 million reads	-0.044865683	-2.40%
PH140: Sligo-13	247,519	100 thousand reads	-0.043369463	-2.45%

PH136: Sligo-14	120,669	100 thousand reads	-0.043369463	-2.45%
PH135: Waterford-33	70,551	100 thousand reads	-0.043369463	-2.45%
PH129: Cork-38	103,725	100 thousand reads	-0.043369463	-2.45%
TF106: Quebec	207,130	100 thousand reads	-0.043369463	-2.45%
JK022: Kunashir1	1,980,045	1 million reads	-0.044865683	-2.40%
JK328: Kunashir2	587,804	1 million reads	-0.044865683	-2.40%

Here we show the corrections applied to each sample's D and f statistic results based on the number of reads present in the data set. We note that the reads in this study are shorter than the multi-fold coverage control individuals Chi1 and Swe (Cahill *et al.* 2015), so if within read linkage is the source of this bias see Figure 3.10 these corrections may lead to underestimation of polar bear ancestry.

References

Becker M, Gruenheit N, Steel M *et al.* (2013) Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change*, **3**, 1039–1043.

Briggs AW, Good JM, Green RE *et al.* (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science (New York, N.Y.)*, **325**, 318–21.

Bronk Ramsey C (2009) Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, **51**, 337–360.

Cahill JA, Green RE, Fulton TL *et al.* (2013) Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. (MW Nachman, Ed,). *PLoS genetics*, **9**, e1003345.

Cahill JA, Stirling I, Kistler L *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular ecology*, **24**, 1205–1217.

Clark CD, Hughes ALC, Greenwood SL, Jordan C, Sejrup HP (2012) Pattern and timing of retreat of the last British-Irish Ice Sheet. *Quaternary Science Reviews*, **44**, 112–146.

Cronin M a., Amstrup SC, Garner GW, Vyse ER (1991) Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursus). *Canadian Journal of Zoology*, **69**, 2985–2992.

Dabney J, Knapp M, Glocke I *et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15758–63.

Dasmahapatra KK, Walters JR, Briscoe AD *et al.* (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Delisle I, Strobeck C (2002) Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Molecular biology and evolution*, **19**, 357–61.

Derrien T, Estellé J, Marco Sola S *et al.* (2012) Fast computation and applications of genome mappability. (CA Ouzounis, Ed,). *PloS one*, **7**, e30377.

Doupé JP, England JH, Furze M, Paetkau D (2007) Most Northerly Observation of a Grizzly Bear (Ursus arctos) in Canada: Photographic and DNA Evidence from Melville Island, Northwest Territories. *Arctic*, **60**, 271–276.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Edwards CJ, Ho SYW, Barnett R *et al.* (2014) Continuity of brown bear maternal lineages in northern England through the Last-glacial period. *Quaternary Science Reviews*, **96**, 131–139.

Edwards CJ, Suchard MA, Lemey P *et al.* (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology : CB*, **21**, 1251–8.

Eyre-Walker A, Woolfit M, Phelps T (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, **173**, 891–900.

Fulton TL (2012) Setting up an ancient DNA laboratory. In: *Ancient DNA, methods and protocols* (eds Shapiro B, Hofreiter M), pp. 1–11. Springer protocols.

Garroway CJ, Bowman J, Cascaden TJ *et al.* (2010) Climate change induced hybridization in flying squirrels. *Global Change Biology*, **16**, 113–121.

Gilbert MTP, Tomsho LP, Rendulic S *et al.* (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science (New York, N.Y.)*, **317**, 1927–30.

Good JM, Dean MD, Nachman MW (2008) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics*, **179**, 2213–28.

Graham RW, Lundelius EL, Graham MA *et al.* (1996) Spatial Response of Mammals to Late Quaternary Environmental Fluctuations. *SCIENCE*, **272**, 1601–1606.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, **328**, 710–22.

Green RE, Malaspinas A-S, Krause J *et al.* (2008) A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell*, **134**, 416–426.

Harington CR, Cournoyer M, Chartier M, Fulton TL, Shapiro B (2014) Brown bear (Ursus arctos) (9880 ± 35bp) from late-glacial Champlain Sea deposits at Saint-Nicolas, Quebec, Canada, and the dispersal history of brown bears. *Canadian Journal of Earth Sciences*, **535**, 527–535.

Harris K, Nielsen R (2016) The Genetic Cost of Neanderthal Introgression. *Genetics*, genetics.116.186890–.

Heintzman PD, Zazula GD, Cahill JA *et al.* (2015) Genomic Data from Extinct North American Camelops Revise Camel Evolutionary History. *Molecular Biology and Evolution*, **32**, 2433–2440.

Hirata D, Mano T, Abramov A V *et al.* (2013) Molecular phylogeography of the brown bear (Ursus arctos) in Northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Molecular biology and evolution*, **30**, 1644–52.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–85.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics (Oxford, England)*, **29**, 1682–4.

Jukes TH, Cantor CR (1969) Evolution of Protein Molecules. In: *Mammalian Protein Metabolism* (ed Munro RE), pp. 21–123. Academic Press, New York, NY.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**, 772–80.

Kelly BP, Whiteley A, Tallmon D (2010) The Arctic melting pot. *Nature*, **468**, 891.

Kunsch HR (1989) The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics*, **17**, 1217–1241.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**, 357–9.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **26**, 589–95.

Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–7.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–9.

Li B, Zhang G, Willersleve E, Wang J (2011) GigaDB Dataset - DOI 10.5524/100008 - Genomic data from the polar bear (Ursus maritimus).

Liu S, Lorenzen ED, Fumagalli M *et al.* (2014) Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, **157**, 785–794.

Materials and Methods

Messer PW (2013) SLiM: Simulating evolution with selection and linkage. *Genetics*, **194**, 1037–1039.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, **2010**, pdb.prot5448.

Miller W, Schuster SC, Welch AJ *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2382–90.

Muhlfeld CC, Kovach RP, Jones LA *et al.* (2014) Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change*, **4**, 620–624.

Ó Cofaigh C, Telfer MW, Bailey RM, Evans DJA (2012) Late Pleistocene chronostratigraphy and ice sheet limits, southern Ireland. *Quaternary Science Reviews*, **44**, 160–179.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Peacock E, Sonsthagen SA, Obbard ME *et al.* (2015) Implications of the circumpolar genetic structure of polar bears for their conservation in a rapidly warming arctic. *PloS one*, **10**, e112021.

Peters JL, Benetti S, Dunlop P, Ó Cofaigh C (2015) Maximum extent and dynamic behaviour of the last British–Irish Ice Sheet west of Ireland. *Quaternary Science Reviews*, **128**, 48–68.

Poelstra JW, Vijay N, Bossu CM *et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science (New York, N.Y.)*, **344**, 1410–4.

Poinar HN, Schwarz C, Qi J *et al.* (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science (New York, N.Y.)*, **311**, 392–4.

Preuß A, Ganslo̊ßer U, Purschke G, Magiera U (2009) Bear-hybrids: behaviour and phenotype. *Der Zoologische Garten*, **78**, 204–220.

Reimer PJ, Bard E, Bayliss A *et al.* (2013) IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, **55**, 1869–1887.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

Sánchez-Guillén RA, Muñoz J, Hafernik J *et al.* (2014) Hybridization rate and climate change: are endangered species at risk? *Journal of Insect Conservation*, **18**, 295–305.

Sato Y, Nakamura H, Ishifune Y, Ohtaishi N (2011) The white-colored brown bears of the Southern Kurils. *Ursus*, **22**, 84–90.

Schliebe S, Wiig Ø, Derocher A, Lunn N, (IUCN SSC Polar Bear Specialist Group) (2008) Ursus maritimus (Polar Bear). *The IUCN Red List of Threatened Species. Version 2015.2.*

Seki O, Ikehara M, Kawamura K *et al.* (2004) Reconstruction of paleoproductivity in the Sea of Okhotsk over the last 30 kyr. *Paleoceanography*, **19**, n/a–n/a.

St John J (2011) SeqPrep: Tool for stripping adaptors and/or merging paired reads with overlap into single reads.

Stiller M, Green RE, Ronan M *et al.* (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 13578–84.

Stirling I (2011) *Polar Bears: The Natural History of a Threatened Species*. Fitzhenry and Whiteside, Brighton, MA.

Stuart AJ, Kosintsev PA, Higham TFG, Lister AM (2004) Pleistocene to Holocene extinction dynamics in giant deer and woolly mammoth. *Nature*, **431**, 684–9.

Woodman P, McCarthy M, Monaghan N (1997) The Irish Quaternary Fauna Project. *Quaternary Science Reviews*, **16**, 129–159.

Zedrosser A, Støen O-G, Sæbø S, Swenson JE (2007) Should I stay or should I go? Natal dispersal in the brown bear. *Animal Behaviour*, **74**, 369–376.

Chapter 4: Inferring species divergence times using Pairwise Sequential Markovian Coalescent (PSMC) modeling and low coverage genomic data

James A. Cahill, André E. R. Soares, Richard E. Green, and Beth Shapiro

Originally published in *Philosophical Transactions of the Royal Society B*, 371(1699); 624-626, 2016.

Abstract

Understanding when species diverged aids in identifying the drivers of speciation, but the end of gene flow between populations can be difficult to ascertain from genetic data. We explore the use of Pairwise Sequential Markovian Coalescent (PSMC) modeling to infer the timing of divergence between species and populations. PSMC plots generated using artificial hybrid genomes show rapid increases in effective population size at the time when the two parent lineages diverge, and this approach has been used previously to infer divergence between human lineages. We show that, even without high coverage or phased input data, PSMC can detect the end of significant gene flow between populations by comparing the PSMC output from artificial hybrids to the output of simulations with known demographic histories. We then apply PSMC to detect divergence times among lineages within two real data sets: great apes and bears within the genus *Ursus*. Our results confirm most previously proposed divergence times for these lineages, and suggest that gene flow between recently diverged lineages may have been common among bears and great apes, including up to one million years of continued gene flow between chimpanzees and bonobos after the formation of the Congo River.

Introduction

The assumption that lineages accumulate sequence-level changes at an approximately constant rate through time, also known as the molecular clock hypothesis, makes it possible to place estimates of evolutionary divergence on calendar and geological timescales using DNA sequence data. The molecular clock hypothesis was first proposed in the 1960's (Kimura, 1968; Zuckerkandl & Pauling, 1962) and has been widely used across evolutionary biology. Inference of the time to most recent common ancestor (TMRCA) of two or more lineages has been used, for

example, to provide insights into environmental events that may have driven evolutionary radiations (Bromham & Woolfit, 2004), episodes of cross-species transmission in viruses (Worobey, Han, & Rambaut, 2014), and the colonization of new habitats by a dispersing species (Fleischer-Dogley et al., 2011).

Prior to recent advances in genome-scale sequencing, most studies incorporating a molecular clock approach estimated the TMCRA of two lineages using phylogenies inferred from one or a few loci, or "gene trees". However, except for instances of post-divergence admixture, the TMRCA of a particular locus within the genome will be more ancient than the population-level divergence of the two lineages. Genome-scale data present an opportunity to unravel the divergence histories of two lineages with increased accuracy, as each of the many "gene trees" within the genome describes a different aspect of the "species tree" (Richard R. Hudson, 1991; Pamilo & Nei, 1988).

PSMC (Pairwise Sequentially Markovian Coalescent) (H. Li & Durbin, 2011) and MSMC (Multiple Sequentially Markovian Coalescent) (Schiffels & Durbin, 2014) are two new computational approaches that are capable of estimating the demographic history of a lineage from genome-scale data. Both PSMC and MSMC infer fluctuations in ancestral effective population size (N_e) from either a single genome (PSMC) or from multiple genomes sampled from the same population (MSMC). PSMC and MSMC estimate ancestral population size by measuring the rate of heterozygosity across regions of the genome. Because heterozygous sites are nucleotide positions where the two parental chromosomes differ, genomic estimates of heterozygosity can be paired with the molecular clock to infer when an individual's parents shared a common ancestor. The distribution across the genome of these times to parental common ancestry, or coalescence events, can then be used

to infer changes in the ancestral population size over time, as the probability of observing a coalescence event at some time in the past is inversely proportional to the ancestral population size.

In addition to inferring changes in effective population size over time, PSMC and MSMC have been used to estimate divergence times between species (Freedman et al., 2014; Lamichhaney et al., 2015; H. Li & Durbin, 2011; Prado-Martinez et al., 2013; Prüfer et al., 2014; Schiffels & Durbin, 2014). The most common approach has been to first infer PSMC plots for each species separately and then to overlay these plots. When reading the plots from the present into the past, between-species divergence is inferred to have occurred at the point in time when the trajectories of two overlain plots become identical, which presumably reflects the transition to shared population histories (e.g., the time prior to divergence). This approach has been used, for example, to estimate the timing of interspecific divergences within the great apes (Prado-Martinez et al., 2013), between modern humans and Neandertals and Denisovans (Prüfer et al., 2014), and between dogs and wolves (Freedman et al., 2014). This approach does not, however, account for the possibility of demographic processes other than complete divergence, such as population structure or that the two lineages might have the same effective size due to chance (H. Li & Durbin, 2011). The second approach, which we investigate further here, uses phased data, which has either been the X chromosomes of male individuals (H. Li & Durbin, 2011), or from multiple whole genomes drawn from high quality human data sets (Schiffels & Durbin, 2014). These phased data are used to create artificial F1 hybrid genomes comprising one chromosome from each of the two lineages whose divergence time is to be inferred, and plots are generated from these artificial hybrid genomes. Sites along an F1 hybrid genome cannot coalesce more recently than the

speciation of the two parent species. These plots therefore show a transition from an infinite population size during the time of lineage divergence to population sizes that reflect the shared ancestry period prior to divergence. Where this transition falls is interpreted as the time of divergence (H. Li & Durbin, 2011).

Although the artificial F1 hybrid approach is a potentially powerful method to learn when two lineages diverged, it is unknown to what extent this approach is suitable for organisms for which high quality, phased genomic data are not yet available. Here, we use a combination of simulated and real data to investigate the utility of the F1-hybrid PSMC approach, hereafter hPSMC, under a variety of demographic scenarios and with low coverage and unphased data. First, we use simulated phased and unphased data to explore the influence of (1) amount of time since divergence, (2) post-divergence gene flow, and (3) effective population size prior to divergence on the accurate recovery of divergence time using hPSMC. We then apply hPSMC to two well studied, real data sets: great apes and bears from the genus *Ursus*. We compare hPSMC estimates of divergence timing within these lineages to estimates inferred using other approaches.

Methods

1. Simulated Data

First, we explored the influence of a several demographic factors on the accuracy of divergence-time inference using hPSMC. Using ms (R. R. Hudson, 2002), we simulated chromosome sequences to generate four data sets: (1) phased haplotypes with no post-divergence gene flow between populations; (2) unphased "haploidized" sequences generated by randomly selecting from one of two haplotypes with no post divergence gene flow between populations to mimic data that are typically available from short-read shotgun sequencing; (3) unphased

sequences generated from populations with a range of pre-divergence effective sizes; and (4) phased haplotypes with varying amounts of post divergence gene flow between populations.

We simulated populations and DNA sequences using ms coalescent simulation (R. R. Hudson, 2002). For each simulation in this study we simulated 200 Mb genomes divided into chromosomes of equal length of either 5 or 10 MB (Tables 4.2, 4.3). We assume a mutation rate of 1*10^-9 mutations per site per year, a recombination rate of 1 CM/Mb per site per generation, and a generation time of 25 years. To create simulated phased data sets, we simulated one haploid sequence per chromosome each from two populations, both with an effective size of 10,000 individuals. The populations were simulated to diverge from an ancestral population of 10,000 individuals at time *t*. We then used these haploid sequences to create an artificial F1 hybrid. We converted the ms output to psmc input files (.psmcfa format) by parsing the sequences and calling a heterozygous site where the parents differ (see https://github.com/jacahill/hPSMC). Although the default settings for PSMC is to bin the genome into 100 base-pair regions (H. Li & Durbin, 2011), we reduced this binning to 10 base-pairs so as to compensate for the higher expected heterozygosity in our simulated hybrid genomes. This change also allows for greater resolution at older time periods and avoids mutation saturation. To assess the influence of the evolutionary distance between populations, we created seven phased data sets where *t*= 100,000, 500,000, 1 million, 2 million, 3 million, 4 million or 5 million years ago. To assess the influence of post-divergence migration between populations, we created five additional data sets where post-divergence migration rates were 0.1, 1, 10, 100 or 1000 migrants per generation. For each of these five data sets, *t*=1 million years ago.

To create unphased data sets, we used a process called "haploidization" (Green et al., 2010), which involves selecting a single high-confidence base call at each site where reads are mapped to the reference genome. Haploidization is useful for genomic analyses of low-coverage data, as it requires only a single high quality base call mapped to each site compared to >20X coverage needed for calling and phasing genotypes (Nielsen, Paul, Albrechtsen, & Song, 2011). To generate a haploidized data set, we simulated data using ms (R. R. Hudson, 2002) as above, but generated two haplotypes per population. Then, for each population, we randomly selected a single allele at each site where the two simulated haplotypes differed. As with the phased data above, we created eight data sets reflecting a range of divergence times between populations from 0 to 5 million years ago.

2. Real Data

We next applied hPSMC to two well-studied biological test cases where whole genome sequence data are available and for which divergence between lineages has been estimated previously: bears from the genus *Ursus* and great apes. We downloaded reads from the NCBI SRA for five bears: two polar bears (*Ursus maritimus*), one from Svalbard (SAMN01057660)(Miller et al., 2012) and another from Alaska (SAMN01057676)(Miller et al., 2012); two brown bears (*Ursus arctos* SAMN03252407, SAMN02045559)(Cahill et al., 2013, 2015), one from North America (SRA) and one from Europe; and an American black bear (*Ursus americanus* SAMN02045561)(Cahill et al., 2013), for one individual from each of the five extant great ape species: human (*Homo sapiens*, ERP001960) (Illumina, 2014), chimpanzee (*Pan troglodytes*, ERS027400) (Prüfer et al., 2012), bonobo (*Pan paniscus*, ERX012399) (Prüfer et al., 2012), gorilla (*Gorilla gorilla*, SRX339460)(Cortez et al., 2014) and orangutan (*Pongo pygmaeus ERS225256*) (Cortez et al., 2014). We aligned the bears to

the polar bear reference genome (B. Li, Zhang, Willersleve, & Wang, 2011) and the great apes to the human reference genome (hg19) (Consortium., 2001))using bwa 0.7.10 (H. Li, 2013) with the BWA-MEM algorithm and default settings. We processed the files, filtered reads with map quality scores less than 30, and removed duplicate reads using Samtools version 0.1.19 (H. Li et al., 2009). After mapping, we generated haploidized sequences as described above for each individual from base calls with minimum base quality and read mapping qualities of 30. We ran PSMC on simulated hybrids from all pairwise combinations of bears and all pairwise combinations of great apes, and estimated the divergence times between each species and between populations as described above. For all bears and great apes we used the default PSMC settings described in the original publication of the method (H. Li & Durbin, 2011). For bears, we assumed the polar bear generation time of 15 years (Schliebe, Wiig, Derocher, Lunn, & (IUCN SSC Polar Bear Specialist Group), 2008) for all comparisons, and a mutation rate of 1*10^-9 mutations per site per year (Nachman & Crowell, 2000). For great apes, we used a generation time of 25 years as per the human analyses in in the original publication of PSMC (H. Li & Durbin, 2011), and also estimated chimp generation time (Langergraber et al., 2012), with a mutation rate of 1*10^-9 mutations per site per generation.

Results

1. Simulated Data

We first explored the influences of three demographic parameters—time since divergence, pre-divergence population size, and post-divergence gene flow—on the ability of PSMC to infer the time of divergence between lineages. Using simulation, we created artificial F1 hybrid chromosomes to mimic both high coverage (phased) genomic data sets and lower coverage (unphased) genomic data

sets, and estimated PSMC plots from these artificial hybrid chromosomes. In each resulting plot, the timing of the transition between an infinite inferred population size to a population size that reflects the shared ancestry period of the two lineages is interpreted as the time of divergence between those lineages.

First, we created artificial hybrid chromosomes from simulated populations in which the time of divergence between parent populations ranged from 100 thousand years ago (ka) to five million years ago (Ma). Using the same divergence times, we created data sets that reflected both phased and unphased data. As expected, plots generated using phased, artificial F1 hybrid genomes (those generated from a single parental chromosome of each species or lineage) show a transition (a rapid change in inferred ancestral population size) at the simulated divergence time (Figure 4.1A). The plots are qualitatively similar whether they are generated from phased or unphased data, however the precise timing of transition is somewhat offset, with the unphased data often showing a transition that occurs more recently than the phased data.

Using unphased data, we then generated eight additional data sets in which the divergence occurred 1 Ma but the pre-transition effective population size ranged from 1,000 to 50,000 individuals. Pre-divergence population size influences the transition time, with larger populations resulting in more ancient transitions from infinite N_e to N_e that is reflective of shared ancestry (Figure 4.1B). This effect is also observed when the divergence occurred 100,000 years ago (data not shown), suggesting that this approach may be more likely to produce accurate estimate of divergence time when populations are small.

We next simulated data sets in which gene flow continues between the populations post-divergence. Assuming a 1 Ma divergence between lineages and

pre- and post-divergence N_s of 10,000, we varied the number of migrants per generation from 0 (complete isolation) to 1000. Figure 4.1C shows that gene flow between populations quickly erodes the precision of hPSMC to detect divergence. At low rates of migration, a transition is observed, but it is not the typical rapid change in N_e and may be challenging to interpret in real data (Figure 4.1C). A rate of one or more migrants per generation results in a plot of N_e that is the post-divergence sum of the populations exchanging migrants (here, 20,000), which is what would be expected in the absence of population divergence. Like phased data, post-divergence gene flow produces a much slower rate of hPSMC-inferred population size increase in haploidized data sets (Figure S18-21)(Cahill et al., 2016). At large population sizes, haploidized data appear to be less impacted by gene flow than phased data (Figure 4.1C) (Figures S18-21)(Cahill et al., 2016).

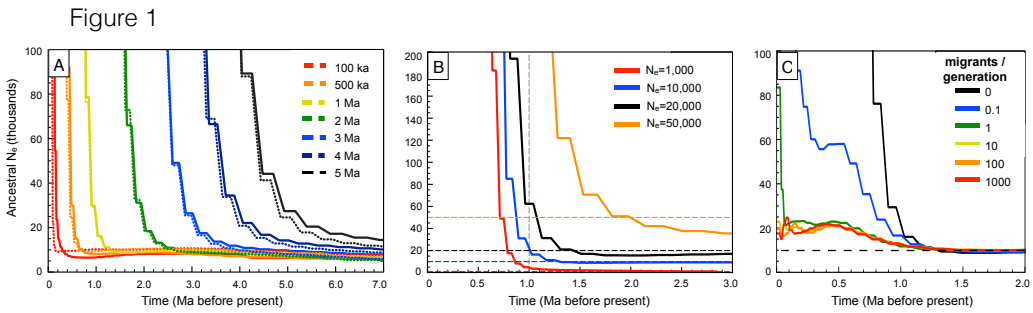


Figure 4.1: Results of simulation experiments designed to test the accuracy of hPSMC in inferring divergence time under three varying demographic scenarios: (A) The influence of using phased (dashed lines) versus unphased (solid lines) data to infer divergence times at seven different depths of divergence; (B) the influence of pre-divergence effective population size on the ability of hPSMC to detect divergence between unphased data; (C) the influence of post-divergence migration between populations. In (B) and (C), divergence between populations occurs 1Ma and the dashed vertical lines indicate the pre-divergence effective population size.

The results described in Figure 4.1 show that population divergence can be inferred as a transition between an infinite population size to population sizes that reflect the shared ancestry of the two parent lineages. However, pinpointing the

exact timing of this transition can be challenging, in particular given the demographic complexities of real data. We therefore implemented a simulation-based procedure that estimates the most likely transition time by comparing hPSMC plots estimated from analyses of simulated data generated under a range of transition times to plots estimated from the real data. This procedure assumes that, if the transition occurs more recently in a simulated data set than it does in the real data, the time of divergence assumed in the simulation was probably more recent than the truth. Likewise, if the transition is older than that observed in the real data, the time of divergence used in that simulation was probably older than the truth. We therefore consider the time range during which divergence is most likely to have occurred to be the narrowest range of simulated divergence times that include the real data without intersecting it (Figure 4.2). So as to capture the portion of the hPSMC plot that is most influenced by the divergence event, we consider only the portion of the inferred hPSMC plots where the ancestral N_e is between arbitrary thresholds of 1.5 and 10 times the pre-divergence N_e. The lower bound is to avoid conflating pre-divergence increases in population size with the signal of population divergence. The upper bound is to avoid exploring parameter space in which little information is present; in instances where inferred N_e increases exponentially after population divergence, the values reported by PSMC are informed by increasingly little data and so will eventually become a greater source of error than information for very large inferred values of N_e. The cut off values of 1.5 to 10 time pre-divergence Ne are intended as a reasonable starting point for interpretation, but may not be appropriate for all data sets.

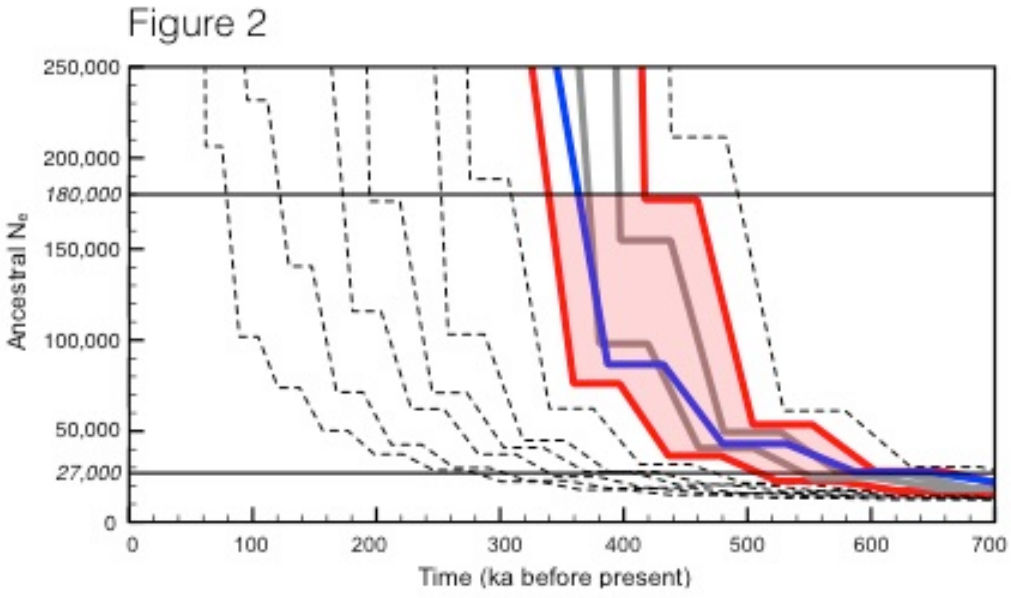


Figure 4.2: An approach to pinpoint the transition (divergence) time using simulation. Here, the hPSMC plot generated for the artificially created chimpanzee/bonobo hybrid genome (blue line) is compared to eleven simulated data sets with divergence times ranging from 0 to 500 ka. Divergence is inferred to have occurred between the simulated divergence times of 300-400 ka (red shaded region), as these are the closest simulations with transition times that do not intersect the transition time of real data. All simulations assume a pre-divergence effective population size of 18,000, which was estimated from the plot of the real data. The vertical lines delineate the range of ancestral effective population size estimates that correspond to 1.5 to 10 times the pre-divergence Ne (27,000-180,000). Plots resulting from all other comparisons are provided as Figures S1-S17)(Cahill et al., 2016)

2. Real Data

We next used hPSMC, as described above, to infer the timing of divergence between lineages of great apes and bears in the genus *Ursus*. For each comparison, we first generated hPSMC plots from artificial F1 hybrid genomes generated from two parent lineages (Figure 4.3). To infer the most likely transition intervals for each pair of lineages, we then simulated populations of each pre-divergence N_e as above, where simulated populations diverged at a range of times spanning those suggested by the hPSMC plot (Figures 4.2, S1-S17)(Cahill et al., 2016), and used these results to infer the most likely range of divergence times for each pair of lineages (Table 4.1).

For the great apes, hPSMC infers the end of gene flow between chimpanzees and bonobos to be 300-450 ka. Our results indicate that humans diverged from the common ancestor of chimpanzees and bonobos about 1.75-3.75 Ma, and that the Hominini (*Homo* and *Pan*) diverged from the lineage leading to gorillas 3.75-6.25 Ma. We find that the lineage leading to orangutans diverged from the other great apes approximately 7.5-13.0 Ma. For the bears, we infer that brown bears and polar bears diverged within the last 200 ka, and their common ancestor diverged from American black bears 500 ka - 1 Ma (Figure 4.3B). In addition to between species divergence, we also estimate the divergence between geographically disparate polar bear populations from Svalbard and Alaska and brown bear populations from Sweden and Alaska. These intraspecific divergences are also within 200ka, 50-150ka for the polar bears and less than 100ka for the brown bears.

Figure 3

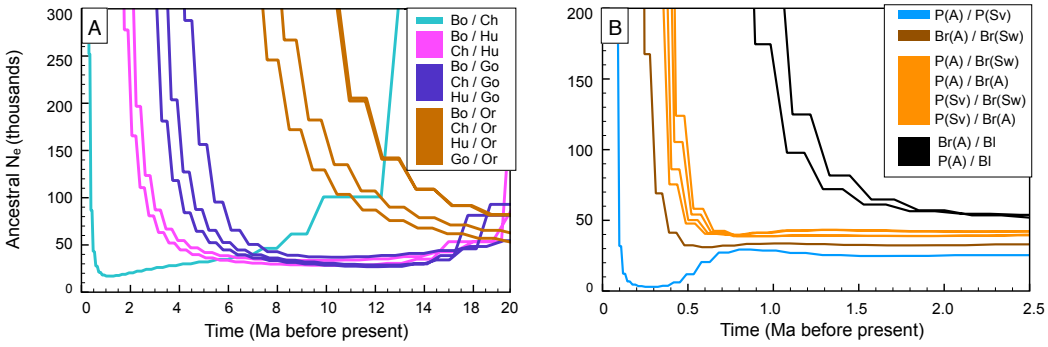


Figure 4.3: Results of hPSMC analyses of (A) five species of great apes and (B) three species of bears in the genus *Ursus*. Within the great apes (A), we observe the expected pattern of divergence in which orangutans diverge most anciently followed by gorillas and then humans, and chimpanzees and bonobos diverge most recently. Within bears (B), we also find the expected order of divergence, where the American black bear is the most ancient divergence, followed by brown bear/ polar bear divergence (light brown) and brown bear/brown bear divergence (dark brown) The polar bear/polar bear divergence (blue) is inferred to have occurred very recently and may be an artifact of the small effective population size of polar bears (see Figure 4.1B).

hPSMC "Parent" 1 Lineage	hPSMC "Parent" 2 Lineage	Inferred Recent Bound for Divergence	Inferred Ancient Bound for Divergence	Pre-Divergence N_e for Simulations
Bonobo	Chimpanzee	300,000	450,000	18,000
Bonobo	Human	2,000,000	3,750,000	40,000
Bonobo	Gorilla	4,000,000	5,250,000	40,000
Bonobo	Orangutan	8,000,000	10,500,000	80,000
Chimpanzee	Human	1,750,000	3,250,000	40,000
Chimpanzee	Gorilla	3,500,000	4,750,000	40,000
Chimpanzee	Orangutan	7,500,000	9,500,000	80,000
Human	Gorilla	5,000,000	6,250,000	40,000
Human	Orangutan	10,000,000	13,000,000	80,000
Gorilla	Orangutan	10,000,000	13,000,000	80,000
Polar Bear (Alaska)	Polar Bear (Scandinavia)	50,000	150,000	4,000
Brown Bear (Alaska)	Brown Bear (Scandinavia)	<100,000	100,000	45,000
Polar Bear (Alaska)	Brown Bear (Alaska)	<100,000	200,000	45,000
Polar Bear (Alaska)	Brown Bear (Scandinavia)	<100,000	100,000	45,000
Polar Bear (Scandinavia)	Brown Bear (Alaska)	<100,000	200,000	45,000
Polar Bear (Scandinavia)	Brown Bear (Scandinavia)	<100,000	200,000	45,000
Polar Bear (Alaska)	American Black Bear	500,000	900,000	50,000
Brown Bear (Alaska)	American Black Bear	600,000	1,000,000	50,000

Table 4.1. Corrected estimates of the inferred divergence time between lineages using hPSMC. Estimates were corrected using the procedure described in Figure 4.2.

Discussion

Our results demonstrate that PSMC can be used to infer the timing of divergence between lineages under a wide range of demographic scenarios, although the accuracy with which divergence time is detected is influenced both by demography and by the quality of the data available for analysis. The extension of the simulated F1 hybrid PSMC (hPSMC) framework to unphased haploidized data, which mirrors the type of data that are available for many published genomes,

produces results that are comparable to those from phased data (Figure 4.1A,B). However, pre-divergence population size will affect the inferred transition time, with population sizes >10,000 appearing to diverge earlier than the truth (Figure 4.1B).

A known, potentially confounding feature of PSMC is that rapid changes in ancestral effective population size are recovered as more gradual transitions (H. Li & Durbin, 2011). In the context of hPSMC this means that we cannot apply a purely qualitative approach to estimating population divergence time by increases in inferred ancestral population size. Even using simulated phased without any post divergence gene flow (Figure 4.1A) inferred ancestral population size begins to increase gradually before the divergence event. Therefore we believe that a framework of comparing simulated population divergence events hPSMC results to those for real data (see above), is essential for reliable divergence time estimation with hPSMC.

Our simulation results also show that hPSMC is highly sensitive to post-divergence gene flow (Figure 4.1C). This sensitivity suggests that the approach may be a useful tool to infer the timing of the end of gene flow between two diverging populations, rather than the time of the initial divergence. As expected from population genetic theory, no divergence is detected when one or more migrants move between populations per generation post-divergence (Slatkin, 1987; Wright, 1931). With smaller amounts of gene flow (here, 0.1 migrants per generation), divergence is detected but the precision of the time estimate of that divergence is less than with no post-divergence migration (Figure 4.1C).

The capacity of hPSMC to capture the end of gene-flow between populations differentiates this approach from the more commonly used approach to detecting

divergence using PSMC, which is to overlay PSMC plots generated separately for the two parent lineages and infer divergence by detecting shared ancestry (Lamichhaney et al., 2015; H. Li & Durbin, 2011; Prado-Martinez et al., 2013; Prüfer et al., 2014; Schiffels & Durbin, 2014). Conveniently, these two approaches appear to capture different aspects of the divergence process. Overlaying PSMC plots detects the end of panmixia, whereas hPSMC detects the end of gene flow. Very strong allopatric barriers to gene flow might result in similar estimates of divergence time from these two approaches, as the end of gene flow will occur simultaneously with divergence. However, incomplete isolation or post-divergence gene flow may cause hPSMC to produce divergence estimates that are substantially more recent than those from overlain PSMC. Using hPSMC in conjunction with overlain PSMC plots may therefore provide additional insights into the divergence process, including an indication of the likelihood of post-divergence gene flow.

Like other forms of PSMC-based analysis, hPSMC requires assumptions about the rate of mutation and generation time that can profoundly impact inference. Because the mutation rate is assumed to be constant across the genome, any local variation, as may be due to the effects of purifying or balancing selection, is ignored. In addition, if the assumed mutation rate differs from the true genome-wide average, estimates of divergence times will be skewed proportionally from the true divergence time. In addition, longer generation time estimates produce larger PSMC inferred ancestral population sizes (H. Li & Durbin, 2011), therefore incorrect inference of the ancestral population size will affect the inferred range of most likely divergence times. Mutation rates and generation times can be difficult to estimate reliably (Conrad et al., 2011; Helgason, Hrafnkelsson, Gulcher, Ward, & Stefánsson, 2003; Hobolth, Christensen, Mailund, & Schierup, 2007; Langergraber et al., 2012;

Tremblay & Vézina, 2000; Zhu, Siegal, Hall, & Petrov, 2014) and the results of any PSMC-based analysis should be interpreted within the context of these limitations.

While the simulations presented above are based on relatively simple demographic models that assume a lack of post divergence gene flow, the same simulation framework can be applied to other, more complex demographic scenarios in order to observe the influence of specific demographic parameters on the shape of the hPSMC plot. For example, if we assume models where populations are large at the time of divergence and maintain prolonged post divergence gene flow, haploidized data produce hPSMC plots that are consistent with a more ancient divergence than the truth (Figure S18-21)(Cahill et al., 2016). Finally given the influential role of pre-divergence N_e on haploidized hPSMC plots, it is important that any simulated data used for comparison to a real data set exhibit the same pre-divergence ancestral population size as the real data.

Great Apes

Our hPSMC-based estimates of divergence times within the lineages of great apes (Table 4.1) are mostly similar to estimates produced using different molecular approaches, although variation between the fossil record and molecular estimates of divergence times within the great apes has long made these estimates contentious (Green & Shapiro, 2013; Langergraber et al., 2012). One surprising result of our analysis is the extremely recent inferred divergence time between chimpanzees and bonobo (Table 4.1). Today, chimpanzees and bonobos are separated by the Congo River, which is thought to have formed ~1.5-2 Ma (Beadle, 1981) and to be a strong barrier to gene flow (Myers Thompson, 2003). Previous PSMC-based estimates of chimpanzee and bonobo divergence, which were estimated by overlaying PSMC plots, suggested that the two lineages diverged 1.5-3 Ma (Prado-Martinez et al.,

2013), supporting the hypothesis that the Congo River has always been a strong barrier to gene flow. Our hPSMC results indicate divergence between chimpanzees and bonobos occurred only 350-400 ka, which we hypothesize may reflect the different aspects of the divergence process that are captured by the two PSMC-based methods. Overlaying PSMC plots to estimate divergence assumes that any difference in ancestral population size indicates population divergence. However, population structure and admixture can also produce different ancestral population size estimates without necessarily indicating that gene flow between populations has ended (H. Li & Durbin, 2011). In contrast, hPSMC is more sensitive to gene flow (Figure 4.1C) and will therefore describe the time when significant gene flow ended. In this context, it can be interpreted that our results suggest that chimpanzees and bonobos may have experienced a long period of population structure with gene flow from 1.5 Ma to 300-450 ka. Interestingly, a comparison of excess allele sharing between a bonobo genome and the genomes of central, eastern and western chimpanzees found no evidence for gene flow between bonobos and any one particular population of chimpanzees (Prüfer et al., 2012). However, the different chimpanzee populations are estimated to have diverged within the last 500 ka (Prado-Martinez et al., 2013), and some evidence of post-divergence gene flow between chimpanzee populations has been inferred (Prado-Martinez et al., 2013; Won & Hey, 2005). If gene flow did occur between chimpanzees and bonobos it therefore must have occurred prior to the isolation between the three populations of chimpanzees.

Our estimates of the time of divergence between the African great apes (human, chimpanzee, bonobo and gorilla) all have large confidence intervals (Table 4.1). This is in part because the transition from infinite ancestral population sizes to a

period of shared ancestry is not as abrupt in these real data sets as it tends to be in simulated data sets (Figure 4.3A). Figure 4.1C shows that this same phenomenon is observed with low levels of post-divergence gene flow. The wide confidence intervals estimated for divergence among great ape lineages may be partly the result of low levels of post-divergence gene flow near the time of speciation. It is also possible that other violations of the assumptions of PSMC, including purifying selection and variation of mutation rates within or among lineages may influence these results. Another possibility is that hPSMC could be detecting the effect of a genetic mosaic of divergent and non-divergent genomic regions, as may occur when two species speciate via strong divergent selection (Via, 2012), or speciation under "genome hitchhiking" (Feder, Egan, & Nosil, 2012; Nosil, Harmon, & Seehausen, 2009). Future analysis of the patterns of genetic variation and rates of mutation along these lineages will be necessary to fully understand why these observed transitions occur more slowly than expected from simulation.

***Bears in the genus* Ursus**

The timing of divergence between bear lineages in the genus *Ursus* has been a matter of much recent debate (Cahill et al., 2013; M. a. Cronin, Amstrup, Garner, & Vyse, 1991; M. A. Cronin, McDonough, Huynh, & Baker, 2013; Edwards et al., 2011; Liu et al., 2014; Miller et al., 2012). This is due in part to the paucity of fossils representing the early divergence of this lineage (Wayne, Van Valkenburgh, & O'Brien, 1991) and to post-divergence hybridization, which may be common among bears (Cahill et al., 2013, 2015; Edwards et al., 2011; Liu et al., 2014). Molecular estimates of the timing of divergence between polar bears and brown bears range from 300 ka to 5 Ma (Liu et al., 2014; Miller et al., 2012). A recent population genetic analysis of 89 polar and brown bear genomes concluded that these two lineages

probably diverged 350-500 ka (Liu et al., 2014). In contrast, PSMC-based estimates of the divergence between brown bears and polar bears has failed to identify a period during which the two lineages converged to the same population size (Miller et al., 2012), which has been interpreted to suggest a very old divergence between the lineages (Miller et al., 2012).

In our hPSMC analyses, *all* of the divergences estimated between polar bears and brown bears—both within and between lineages—occur within the last 200,000 years. We hypothesize that these remarkably recent divergences, which disagree with evidence from the fossil record (Wayne et al., 1991), are probably the result of recent admixture among these lineages (Cahill et al., 2013, 2015; Liu et al., 2014). For example, we infer that brown bears in Alaska and Sweden diverged less than 100,000 years ago. Today, Alaskan and European brown bears are isolated by a variety of geographic and physical barriers, including the Bering Strait. During the last ice age, however, brown bears occupied a more or less continuous range from western Europe to Canada's Yukon Territory (Barnes, Matheus, Shapiro, Jensen, & Cooper, 2002; Edwards et al., 2011). Data from mitochondrial DNA indicates a major expansion of brown bears out of Beringia beginning around 30-35 ka (Edwards et al., 2011), which may explain the recent gene flow between Swedish and Alaskan brown bears (Figure 4.3B).

Similarly, hPSMC suggests that brown bears and polar bears diverged less than 200 ka (Table 4.1), which is considerably more recent than the timing of divergence inferred using a population genetics approach (Liu et al., 2014) and more recent than the age of the oldest known polar bears fossil, which dates to ~110 ka (Ingólfsson & Wiig, 2009). The hPSMC-based estimate of divergence is also probably influenced by post-divergence gene flow. Alaskan brown bears, including the

individual used in this study, are known to have a small but variable component of polar bear ancestry as the result of post-divergence hybridization with polar bears within the last 20 ka (Cahill et al., 2013, 2015). However, between-species divergences estimated using the Swedish brown bear, which has not been shown to have any polar bear ancestry, also suggest a recent divergence (Table 4.1). This is less well explained, because neither polar bears nor Swedish brown bears have been shown to have detectable introgressed ancestry from the other species (Cahill et al., 2013; Liu et al., 2014). However, as the tests used to detect introgression used the Swedish bear as the purportedly un-admixed individual, future work in comparison to other, potentially less admixed, brown bears might reveal some polar bear ancestry in this Swedish brown bear.

Conclusions

We have shown that hPSMC, or PSMC analysis of simulated F1 hybrid individuals, can be used to estimate population divergence times with low coverage unphased data. hPSMC provides a distinct perspective with regard to divergence than other methods, including overlaying PSMC plots. While overlaying PSMC plots detects the end of panmixia, hPSMC detects the end of gene flow. Very strong allopatric barriers to gene flow might make these two estimates the same. However, incomplete isolation or post-divergence gene flow may cause hPSMC to produce divergence estimates that are substantially more recent than those from overlain PSMC. In our case studies using real data from great apes and bears, we inferred divergence times that were largely consistent with estimates from other methods. However, our assessments of recently diverging lineages—chimpanzees and bonobos and polar bears and brown bears—are suggestive of a divergence process that includes post-divergence gene flow rather than an abrupt transition

from panmixia to isolation. While other methods are available to infer the timing of divergence between lineages, even under a variety of complex, demographic scenarios MSMC (Multiple Sequential Markovian Coalescent) (Schiffels & Durbin, 2014), δaδi (Diffusion Approximations for Demographic Inference)(Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) and IBS (Identity By State) tract length (Harris & Nielsen, 2013), most of these require high coverage and phased genomic data, often from multiple individuals. hPSMC is particularly useful for estimating divergence time from low coverage data sets, such as those found in ancient DNA studies, because it does not require the ability to call heterozygous sites or phase haplotypes. We suggest that hPSMC may be a valuable tool for estimating divergence in these common scenarios, and that, in combination with other approaches, can provide important new insights into the process of population subdivision and speciation.

Additional Information

Authors' Contributions

JC, REG and BS designed the experiment. JC and AERS performed the analyses. All authors contributed to the interpretation of results and writing of the manuscript.

Funding

This work was supported by grants from the Packard Foundation and the Gordon and Betty Moore Foundation.

Additional Information: Additional supplementary figures (S1-21) are available along with the original publication online (Cahill et al., 2016).

Table 4.2: ms (R. R. Hudson, 2002) simulation parameters for simulated data model testing shown in Figure 4.1.

Test	Ne	Divergence (years)	Migrants per Generation	phased or haploidized	ms command
Figure 4.1A	10000	100000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 -ej 0.1 2 1
Figure 4.1A	10000	500000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 0.5 2 1
Figure 4.1A	10000	1000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 1.0 2 1
Figure 4.1A	10000	2000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 2.0 2 1
Figure 4.1A	10000	3000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 3.0 2 1
Figure 4.1A	10000	4000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 4.0 2 1
Figure 4.1A	10000	5000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 5.0 2 1
Figure 4.1A	10000	100000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 0.1 2 1
Figure 4.1A	10000	500000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 0.5 2 1
Figure 4.1A	10000	1000000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 1.0 2 1
Figure 4.1A	10000	2000000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 2.0 2 1
Figure 4.1A	10000	3000000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 3.0 2 1
Figure 4.1A	10000	4000000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 4.0 2 1
Figure 4.1A	10000	5000000	0	haploidized	ms 4 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 5.0 2 1

Figure 4.1B	1000	1000000	0	haploidized	ms 4 40 -t 500.0 -r 200.0 5000000 -I 2 2 2 -ej 10.0 2 1
Figure 4.1B	10000	1000000	0	haploidized	ms 4 40 -t 5000.0 -r 2000.0 5000000 -I 2 2 2 -ej 1.0 2 1
Figure 4.1B	20000	1000000	0	haploidized	ms 4 40 -t 10000.0 -r 4000.0 5000000 -I 2 2 2 -ej 0.5 2 1
Figure 4.1B	50000	1000000	0	haploidized	ms 4 40 -t 25000.0 -r 10000.0 5000000 -I 2 2 2 -ej 0.2 2 1
Figure 4.1C	10000	1000000	0	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 2 2 -ej 1.0 2 1
Figure 4.1C	10000	1000000	0.1	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 0.4 -ej 1.0 2 1
Figure 4.1C	10000	1000000	1	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 4.0 -ej 1.0 2 1
Figure 4.1C	10000	1000000	10	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 40.0 -ej 1.0 2 1
Figure 4.1C	10000	1000000	100	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 400.0 -ej 1.0 2 1
Figure 4.1C	10000	1000000	1000	phased	ms 2 20 -t 10000.0 -r 4000.0 10000000.0 -I 2 1 1 4000.0 -ej 1.0 2 1

Table 4.3. To correct for the effect of population size on inferring divergence time using hPSMC, we simulated simple population divergence events with the same pre-divergence effective population size as that which hPSMC inferred for our data. Then, we compared the hPSMC result for real data to the simulated data to estimate divergence time. Here, we show the ms (R. R. Hudson, 2002) commands used to create these simulated population divergence events. (see online publication).

References

Barnes, I., Matheus, P., Shapiro, B., Jensen, D., & Cooper, A. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science (New York, N.Y.)*, *295*(5563), 2267–70. http://doi.org/10.1126/science.1067814

Beadle, L. C. (1981). *Inland waters of tropical Africa*. New York, NY: Longman.

Bromham, L., & Woolfit, M. (2004). Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. *Systematic Biology*, *53*(5), 758–66. http://doi.org/10.1080/10635150490522278

Cahill, J. A., Green, R. E., Fulton, T. L., Stiller, M., Jay, F., Ovsyanikov, N., … Shapiro, B. (2013). Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genetics*, *9*(3), e1003345. http://doi.org/10.1371/journal.pgen.1003345

Cahill, J. A., Soares, A. E. R., Green, R. E., Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1699), 624–626. http://doi.org/10.1098/rstb.2015.0138

Cahill, J. A., Stirling, I., Kistler, L., Salamzade, R., Ersmark, E., Fulton, T. L., … Shapiro, B. (2015). Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, *24*(6), 1205–1217. http://doi.org/10.1111/mec.13038

Conrad, D. F., Keebler, J. E. M., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., … Awadalla, P. (2011). Variation in genome-wide mutation rates within and between human families. *Nature Genetics*, *43*(7), 712–714. http://doi.org/10.1038/ng.862

Consortium., I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, *412*(August), 860–921. http://doi.org/10.1038/35057062

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., … Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, *508*(7497), 488–493. http://doi.org/10.1038/nature13151

Cronin, M. a., Amstrup, S. C., Garner, G. W., & Vyse, E. R. (1991). Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursus). *Canadian Journal of Zoology*, *69*(12), 2985–2992. http://doi.org/10.1139/z91-421

Cronin, M. A., McDonough, M. M., Huynh, H. M., & Baker, R. J. (2013). Genetic relationships of North American bears (Ursus) inferred from amplified fragment length polymorphisms and mitochondrial DNA sequences. *Canadian Journal of Zoology*, *91*(9), 626–634. http://doi.org/10.1139/cjz-2013-0078

Edwards, C. J., Suchard, M. A., Lemey, P., Welch, J. J., Barnes, I., Fulton, T. L., … Shapiro, B. (2011). Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current Biology : CB*, *21*(15), 1251–8. http://doi.org/10.1016/j.cub.2011.05.058

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics : TIG*, *28*(7), 342–350. http://doi.org/10.1016/j.tig.2012.03.009

Fleischer-Dogley, F., Kettle, C. J., Edwards, P. J., Ghazoul, J., Määttänen, K., & Kaiser-Bunbury, C. N. (2011). Morphological and genetic differentiation in populations of the dispersal-limited coco de mer (Lodoicea maldivica): implications for management and conservation. *Diversity and Distributions*, *17*(2), 235–243. http://doi.org/10.1111/j.1472-4642.2010.00732.x

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., … Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics*, *10*(1), e1004016. http://doi.org/10.1371/journal.pgen.1004016

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, *328*(5979), 710–22. http://doi.org/10.1126/science.1188021

Green, R. E., & Shapiro, B. (2013). Human evolution: turning back the clock. *Current Biology : CB*, *23*(7), R286–8. http://doi.org/10.1016/j.cub.2013.02.050

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695. http://doi.org/10.1371/journal.pgen.1000695

Harris, K., & Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics*, *9*(6), e1003521. http://doi.org/10.1371/journal.pgen.1003521

Helgason, A., Hrafnkelsson, B., Gulcher, J. R., Ward, R., & Stefánsson, K. (2003). A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *American Journal of Human Genetics*, *72*(6), 1370–1388. http://doi.org/10.1086/375453

Hobolth, A., Christensen, O. F., Mailund, T., & Schierup, M. H. (2007). Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genetics*, *3*(2), e7. http://doi.org/10.1371/journal.pgen.0030007

Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, *7*, 1–44.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338. http://doi.org/10.1093/bioinformatics/18.2.337

Illumina. (2014). Platinum Genomes. Retrieved from http://www.illumina.com/platinumgenomes/

Ingólfsson, Ó., & Wiig, Ø. (2009). Late Pleistocene fossil find in Svalbard: the oldest remains of a polar bear (Ursus maritimus Phipps, 1744) ever discovered. *Polar*

Research, *28*(3), 455–462. http://doi.org/10.1111/j.1751-8369.2008.00087.x

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, *217*(5129), 624–626. http://doi.org/10.1038/217624a0

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., … Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371–375. http://doi.org/10.1038/nature14181

Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., … Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(39), 15716–15721. http://doi.org/10.1073/pnas.1211740109

Li, B., Zhang, G., Willersleve, E., & Wang, J. (2011). GigaDB Dataset - DOI 10.5524/100008 - Genomic data from the polar bear (Ursus maritimus). Retrieved from http://gigadb.org/dataset/100008

Li, H. (2013). Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM, *00*(00), 1–3.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–6. http://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–9. http://doi.org/10.1093/bioinformatics/btp352

Liu, S., Lorenzen, E. D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., … Wang, J. (2014). Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, *157*(4), 785–794. http://doi.org/10.1016/j.cell.2014.03.054

Miller, W., Schuster, S. C., Welch, A. J., Ratan, A., Bedoya-Reina, O. C., Zhao, F., … Lindqvist, C. (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(36), E2382–90. http://doi.org/10.1073/pnas.1210506109

Myers Thompson, J. a. (2003). A model of the biogeographical journey from Proto-pan to Pan paniscus. *Primates; Journal of Primatology*, *44*(2), 191–7. http://doi.org/10.1007/s10329-002-0029-1

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics*, *156*(1), 297–304. Retrieved from http://www.genetics.org/content/156/1/297.full

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews. Genetics*, *12*(6), 443–51. http://doi.org/10.1038/nrg2986

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*(3), 145–156. http://doi.org/10.1016/j.tree.2008.10.011

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.*, *5*(5), 568–583. Retrieved from http://mbe.oxfordjournals.org/content/5/5/568

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., … Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, *499*(7459), 471–5. http://doi.org/10.1038/nature12228

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., … Pääbo, S. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature*, *486*(7404), 527–31. http://doi.org/10.1038/nature11128

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., … Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–9. http://doi.org/10.1038/nature12886

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–25. http://doi.org/10.1038/ng.3015

Schliebe, S., Wiig, Ø., Derocher, A., Lunn, N., & (IUCN SSC Polar Bear Specialist Group). (2008). Ursus maritimus (Polar Bear). Retrieved July 31, 2015, from http://www.iucnredlist.org/details/22823/0

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, *236*(4803), 787–792. http://doi.org/10.1126/science.3576198

Tremblay, M., & Vézina, H. (2000). New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics*, *66*(2), 651–658. http://doi.org/10.1086/302770

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1587), 451–460. http://doi.org/10.1098/rstb.2011.0260

Wayne, R., Van Valkenburgh, B., & O'Brien, S. (1991). Molecular distance and divergence time in carnivores and primates. *Mol. Biol. Evol.*, *8*(3), 297–319. Retrieved from http://mbe.oxfordjournals.org/content/8/3/297

Won, Y.-J., & Hey, J. (2005). Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, *22*(2), 297–307.

http://doi.org/10.1093/molbev/msi017

Worobey, M., Han, G.-Z., & Rambaut, A. (2014). A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*, *508*(7495), 254–7. http://doi.org/10.1038/nature13016

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*, 97–159.

Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), E2310–8. http://doi.org/10.1073/pnas.1323011111

Zuckerkandl, E., & Pauling, L. (1962). Molecular Disease, Evolution, and Genic Heterogeneity. In M. Kasha & B. Pullman (Eds.), *Horizons in Biochemistry* (pp. 189–222). New York, NY: Academic Press.

Chapter 5: Genome sequencing of cave bears (*Ursus spelaeus*) reveals introgression into living brown bears

Expected Co-First Authors of eventual publication: James Cahill, Axel Barlow

Senior Authors: Beth Shapiro, Michael Hofreiter

Expected Coauthors: Stefanie Hartmann, Gernot Rabeder, Christine Frischauf, Aurora Grandal-d'Anglade, Ana García-Vazquez, Mari Murtskhvaladze, Urmas Saarma, Peeter Anijalg, Gloria Fortes, Georgios Xenikoudakis, Tomaz Skrbinsek, Giorgio Bertorelle, Beth Shapiro, Ron Pinhasi, Love Dalén

Unpublished, author list subject to change prior to peer review publication
This chapter was written for expressly for this dissertation and has not been reviewed by the expected co-authors of the final publication.

Abstract

The emerging field of ancient genomics presents the opportunity to investigate the past diversity of life in great detail. Cave bears (*Ursus spelaeus*) with their extensive fossil record are an important model species for ancient DNA research (Dabney *et al.* 2013), and as the nearest relative of polar bears and brown bears provide a particularly useful lens for the study of their near relatives. Here we present and analysis of four multi-fold coverage cave bear genomes. We find that cave bears exhibit a deep divergence and substantial isolation between European and Caucasian populations arguably consistent with separate species designation. We also find evidence for gene flow from cave bears into brown bears contributing at least 1.0-3.5% of brown bear's genomes in all sampled brown bear populations.

Introduction

Cave bears (*Ursus spelaeus*) were widespread throughout Eurasia in the Pleistocene (KNAPP *et al.* 2009), but went extinct after a gradual decline from 50-24 thousand years ago (Stiller *et al.* 2010, 2014). Cave bears are morphologically similar to brown bears but are thought to have had a more herbivorous diet compared to the omnivorous brown bear (PACHER & STUART 2009), although cave bears were not exclusively herbivorous (Richards *et al.* 2008). Mitochondrial DNA evidence has shown that cave bears are the nearest matrilineal relative of the polar/brown bear clade (*Ursus maritimus and Ursus arctos*)(Krause *et al.* 2008). Until recently cave bears were believed have been restricted to Europe however, recent findings have shown that their range extends across Asia to eastern Siberia (KNAPP *et al.* 2009). Intriguingly, cave bears from Asia's mitochondrial genomes form a clade that is highly divergent from European cave bears (KNAPP *et al.* 2009).

Cave bears high frequency in the fossil record makes them an ideal model

system for ancient genomics. Cave bears have served as a key testing system for ancient DNA technological development including the protocol for shotgun sequencing libraries with fragment sizes less than 50 bp (Dabney *et al.* 2013). Further their nearest outgroup relationship to the polar/brown bear clade which has been the subject of substantial recent interest (Li *et al.* 2011; Hailer *et al.* 2012; Miller *et al.* 2012; Cahill *et al.* 2013, 2015; Bidon *et al.* 2014; Liu *et al.* 2014) and societal concern makes the sequencing of cave bears an important step in understanding the diversity and evolution of these extant species of interest.

To investigate the diversity within cave bears nuclear genomes and their relationship with their near evolutionary relatives we generated 2.7-4.4X coverage whole genome shotgun sequencing data for four cave bears each from a different recognized subspecies: *Ursus spelaeus spelaeus* and *U. s. ingressus* large bodied European cave bears, *U. s. eremus* a smaller bodied alpine European cave bear and *U. s. kudarensis* a Caucasian cave bear (Table 5.1). As well as several brown bears including brown bears from the same region as each of the cave bears (Table 5.1) and a late Pleistocene Austrian brown bear that was spatially and temporally sympatric with cave bears. These data not only provide the first whole genome insights into cave bears, they reveal the substantial subdivision between European cave bears (subspecies *spelaeus, eremus* and *ingressus*) and Caucasian cave bears (*kudarensis*). Strikingly these data also reveal hitherto unexpected gene flow from cave bears into brown bears meaning that some cave bear traits continue to survive within brown bears.

Methods

Laboratory methods – ancient samples

All laboratory work preceding library amplification was carried out in

dedicated ancient DNA facilities at the University of York or the University of Potsdam, following established guidelines (Fulton 2012). Bone samples were taken from the densest, interior portion of *Ursus* petrous bones, which has be shown to provide high percentages of endogenous DNA (Gamba *et al.* 2014; Pinhasi *et al.* 2015). Fifty milligrams of bone was ground to a fine powder using ceramic mortar and pestles and subjected to DNA extraction, following the protocol of (Dabney *et al.* 2013). DNA extracts were converted into Illumina sequencing libraries using a published protocol based on single-stranded DNA (Gansauge & Meyer 2013). A unique eight base-pair index sequence was incorporated within the P7 adapter sequence of each library during amplification to facilitate data demultiplexing. The optimal number of PCR cycles applied to each library during amplification was determined in advance using a qPCR methodology (Gansauge & Meyer 2013). Amplified libraries were purified using commercial silica spin-columns (Qiagen MinElute) and quantified using a Qubit 2.0 fluorometer (ThermoFisher Scientific) and 2200 TapeStation Instrument (Agilent), prior to sequencing. All samples were sequenced on an Illumina HiSeq with 75 base pair, paired end reads.

Laboratory methods – modern samples

Modern *Ursus* tissue samples were subjected to DNA extraction using a commercial kit (Qiagen Dneasy). Extract concentrations were measured using Qubit and DNA quality assessed using TapeStation genomic DNA assay. Five-hundred nanograms of DNA in a volume of 130ul was then sheared by sonication to an average fragment length of 500bp using a Covaris S220 focused ultrasonicator. Sheared DNA was then converted into Illumina sequencing libraries using a published protocol based on double-stranded DNA (Meyer & Kircher 2010) with modifications described in (Fortes & Paijmans 2015). Library amplification and

indexing was carried out as described previously for ancient samples. Library molecules corresponding to insert sized < 300bp and > 1000bp were removed prior to sequencing using a PippinPrep instrument (Sage Science). All samples were sequenced on an Illumina HiSeq with 151 base pair, paired end reads.

Sample	Species	Location	Coverage	Ancient
GS136	*Ursus spelaeus*	Spain	2.72X	Yes
UD1838	*Ursus spelaeus*	Austria	3.29X	Yes
WK01	*Ursus spelaeus*	Austria	4.39X	Yes
HV74	*Ursus spelaeus*	Russian Federation	2.71X	Yes
191Y	*Ursus arctos*	Russian Federation	4.28X	No
Uap	*Ursus arctos*	Austria	1.93X	Yes
LS039	*Ursus arctos*	Spain	3.62X	No
235	*Ursus arctos*	Russian Federation	8.03X	No
Cau	*Ursus arctos*	Russian Federation	9.35X	No
Slov-arctos	*Ursus arctos*	Slovenia	8.48X	No
Swe	*Ursus arctos*	Dalarna, Sweden	6.30X	No
Adm1	*Ursus arctos*	Admiralty Island, Alaska, USA	8.96X	No
Den	*Ursus arctos*	Denali, Alaska, USA	2.09X	No
WH1	*Ursus maritimus*	West Hudson Bay, Canada	3.43X	No
NB	*Ursus maritimus*	North Beaufort Sea, Canada	4.34X	No
SB	*Ursus maritimus*	South Beaufort Sea, Alaska, USA	3.45X	No
Uam	*Ursus americanus*	Pennsylvania, USA	7.53X	No

Table 5.1 Sample Information.

Mapping and Data preparation

We merged overlapping paired end reads with SeqPrep (St John 2011). We mapped merged and unmerged reads to the giant panda reference genome (Li *et al.* 2010) with bwa aln v0.7.7 (Li & Durbin 2010). To account for the substantial evolutionary divergence between the reference genome and the samples we increased the allowed mismatch rate by setting the –n flag to 0.01 rather than the default of 0.04. We excluded reads with a MapQuality score less than 30 and

removed duplicate reads with samtools v0.1.19 (Li *et al.* 2009).

We selected the giant panda reference genome (Li *et al.* 2010) rather than the less evolutionarily distant and more contagious polar bear reference genome (Li *et al.* 2011; Liu *et al.* 2014) because as an ingroup to the group of samples being studied the polar bear reference genome might introduce bias into our mappings that would disproportionately impact admixture inference. Cave bear reads from regions of the cave bear genome introgressed from polar bear or brown bear would have a greater probability of mapping to the polar bear genome than reads from other parts of the genome because of their lower divergence(Prüfer *et al.* 2010). This biased assembly would produce the artifact of an inflated frequency of shared derived alleles between polar bears and cave bears in the cave bear assemblies. By contrast mapping to the outgroup giant panda reference would have no bias for or against mapping introgressed regions making it a more suitable reference genome for this study.

For analysis we generated haploidized sequences for each individual by randomly selecting a single high quality base call (BaseQuality >=30, read MapQuality >=30) at each site in the panda reference genome. This method better represents nonreference alleles for low coverage samples than genotype calling which tends to be biased toward the reference allele potentially confounding downstream analysis (Green *et al.* 2010). To avoid including duplicated elements we masked sites where an individual's coverage was about the 95th percentile genome wide.

Divergence between individuals

To assess the genome wide average divergence between the cave bear lineages examined in this study we calculated the pairwise sequence divergence

between the haploidized sequences at all sites where all four cave bears and a polar bear (NB) outgroup were represented. We restricted our analysis to transversion sites only to avoid bias resulting from Cytosine deamination damage. From the pairwise distance matrix we constructed a neighbor joining tree (Saitou & Nei 1987) using TREX (Boc *et al.* 2012).

Admixture

We tested for admixture between cave bears and their nearest extant relatives, polar bears and brown bears with the D-statistic (ABBA, BABA test) (Green *et al.* 2010; Durand *et al.* 2011). To identify differential gene flow between cave bears and brown/polar bears we calculated D-statistics for all possible combinations of polar bears and brown bears. In order to avoid bias resulting from ancient DNA cytosine deamination (C->T error) damage (Hofreiter 2001) we restricted our analysis to transversion sites. To test for significance we applied the weighted block jackknife (Kunsch 1989; Green *et al.* 2010). Because of the low contiguity of the giant panda genome we used 1Mb non-overlapping blocks. However, because any cave bear introgression into brown bears would have occurred prior to cave bear's extinction 1Mb should be larger than the largest non-recombined introgressed block which is the minimum appropriate size for weighted block jackknife bins (Green *et al.* 2010; Patterson *et al.* 2012), we consider results more than three standard errors different from zero (Z>3) to be significant evidence of admixture and report confidence intervals for estimates of 1.96 (95% confidence) standard errors (Green *et al.* 2010).

To quantify the amount of admixture we used the $\hat{f}$ statistic (Green *et al.* 2010; Durand *et al.* 2011) which is the ratio of the excess of shared derived alleles between the admixed individual and candidate introgressor and the excess of shared

derived alleles expected in a 100% admixed individual. We calculated all $\hat{f}$ statisitcs that are consistent with the D-statistic results. $\hat{f}$'s expected value is best calculated by using individuals that we hypothesize best approximate the diversity within the introgressing populations. We selected consider the European cave bears to best represent the diversity within a potentially introgressing cave bear lineage. For brown bear introgressors we selected Eurasian brown bears as best representing diversity in a potential cave bear introgressor. As with the *D*-statistic we measure and report significance based on weighted block jackknife with 1 Mb blocks (Kunsch 1989; Green *et al.* 2010).

To determine whether gene flow was principally from cave bears into brown bears or brown bears into cave bears my coauthors developed a novel test. They subdivided the genome into 25kb non-overlapping bins and in each bin calculated the pairwise transversion differences between two cave bears, two brown/polar bears and an outgroup, American black bear, at all sites where all five individuals had a called base. From this distance matrix they calculated the rooted tree topology for each 25kb non-overlapping bin in the genome. Then they compared the frequency of species tree incongruent topologies. If gene flow was principally from cave bears into brown bears, we would expect the brown bears that share more alleles with cave bears to also more frequently group with cave bears in 25kb bin tree reconstructions. However, if gene flow was principally from brown bears into cave bears we would expect greater variation in the rate with which cave bears group with brown bears in bin phylogeny reconstruction.

Results

Diversity among Cave bears

We sequenced four cave bears, one each from four recognized subgroups

classified as either species or subspecies in the literature; *spelaeus, eremus, ingressus and kudarensis* (KNAPP *et al.* 2009; Dabney *et al.* 2013). To assess the relationship between these lineages we calculated the genome wide pairwise transversion rate between them and between each cave bear and a polar bear outgroup. To ensure the analysis was not biased by differences in coverage we restricted our analysis to sites where all four cave bears and the polar bear outgroup all had a called site in their haploidized sequence. We generated a neighbor joining tree (Saitou & Nei 1987) which revealed that the European cave bears *spelaeus, eremus, ingressus* group together to the exclusion of *kudarensis,* consistent with the mitochondrial phylogeny (KNAPP *et al.* 2009). However, consistent with morphology but not with the mitochondrial phylogeny (Baca *et al.* 2012) we found *spelaeus* and *ingressus* to be most closely related to one another (Table 5.2).

	GS136	UD1838	*eremus*	*kudarensis*	*maritimus*
GS136	---				
UD1838	0.00055	---			
eremus	0.00069	0.00085	---		
kudarensis	0.00094	0.00110	0.00119	---	
maritimus	0.00124	0.00140	0.00149	0.00144	---

Table 5.2. **Nuclear Genome Divergence Matrix.** Pairwise sequence difference between cave bears (listed by subspecies name, see Table 5.1) and a polar bear outgroup (*maritimus*), calculated from transversion sites only.

To assess the extent of genetic isolation between cave bear lineages we calculated the D-statistic (Green *et al.* 2010; Durand *et al.* 2011)for all possible combinations of cave bears Table 5.4. We then use D-statistic tests consistent with the primary tree topology D(ingroup cave bear, ingroup cave bear, outgroup cave bear, outgroup black bear) to test for asymmetric gene flow between outgroup and ingroup lineages, but no results were significantly different from zero (Z>3)

indicating the absence of asymmetric post-divergence gene flow.

To test the degree of genetic isolation between ingroup and outgroup lineages we applied most common tree inconsistent D-statistic tests within cave bears D(outgroup cave, ingroup cave, ingroup cave, outgroup black bear) (Figure 5.1, orange). This test is expected to range from 0 to 1 with one indicating that the ingroup always shares alleles with the other ingroup to the exclusion of the outgroup and zero indicating a perfect trifurcation where the "outgroup" is no less likely to share an allele with the "ingroup" than the "ingroups" are to share alleles with one another. In the absence of asymmetric admixture results less than D=0.33 indicate that the most common tree topology is not the majority tree topology. Within European cave bears D(WK01,GS136,UD1838,Black Bear)=0.25 consistent with low differentiation between lineage and substantial incomplete lineage sorting or admixture. But tests for the connectivity between Caucasian and European cave bears demonstrate a much greater degree of isolation indicate D(HV74,GS136,WK01,Black Bear)=0.73 (see Figure 5.1).

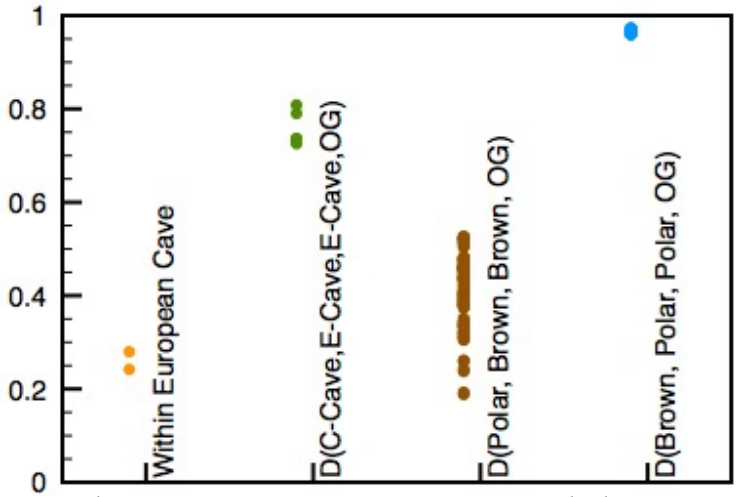


Figure 5.1. Here we show D-statistic tests inconsistent with the general topology. This tests the degree to which the P2 and P3 individuals are distinct form the P1 individual and range from 0, trifurcation to 1 total absence of post divergence gene flow or incomplete lineage sorting. We find that within European cave bears genetic isolation (orange) is low but within the range of the degree of isolation of brown

bears from polar bears (brown). However, European cave bears are quite genetically isolated from Caucasian cave bears D(Caucasus, Europe, Europe, Outgroup)=0.72 to 0.80, although still less than the genetic isolation of polar bears from brown bears D(Brown, Polar, Polar, Outgroup)=0.96 to 0.97.

Interspecies Gene Flow

We tested for interspecific gene flow between cave bears and brown/polar bears with the D-statistic (Green *et al.* 2010; Durand *et al.* 2011). We calculated all possible combinations of introgressors from the polar/brown bear clade into the cave bear clade and vis versa. We found evidence for gene flow between brown bears and cave bears but not between polar bears and cave bears. Within brown bears evidence for gene flow with cave bears was greatest in the ancient Austrian brown bear, contemporary with cave bears. Cave bear gene flow evidence was also greater in Eurasian brown bears than North American brown bears. Within cave bears we found evidence for more extensive gene flow between European cave bears and brown bears than between Caucuses (*U. kudarensis*) cave bears and brown bears D(European Cave, Caucasian Cave, Brown, Outgroup)=0.020 to 0.038 (Table 5.4).

Direction of Gene Flow

To determine the direction of gene flow my collaborators tested whether phylogenetic reconstruction from 25kb non-overlapping bins was consistent with gene flow from brown bears into cave bears or cave bears into brown bears. Briefly, if brown bears were the recipients of cave bear introgression we would expect the topology (((Cave, Admixed Brown), Cave) Polar) to be more common than the topology (((Cave, Polar), Cave), Admixed brown). By contrast if the direction of gene flow was brown bears into cave bears we would expect sites influenced by admixture to have the topology (((Admixed Cave, Brown), Polar), Cave), but we would not expect the brown bear to fall within the diversity of cave bears more frequently than the polar bear falls within the diversity of cave bears.

The 25kb-bin tree topologies show that brown bears that share the most alleles with cave bears fall within the diversity of cave bears more frequently than brown bears with sharing fewer alleles with cave bears or polar bears. This is consistent with the principal direction of gene flow being cave bear introgression into brown bears. There is also a small increase in bins where European cave bears (cave +) fall within the diversity of brown/polar bears relative to Caucasian cave bears (cave -) indicating that gene flow may have been bi-directional.

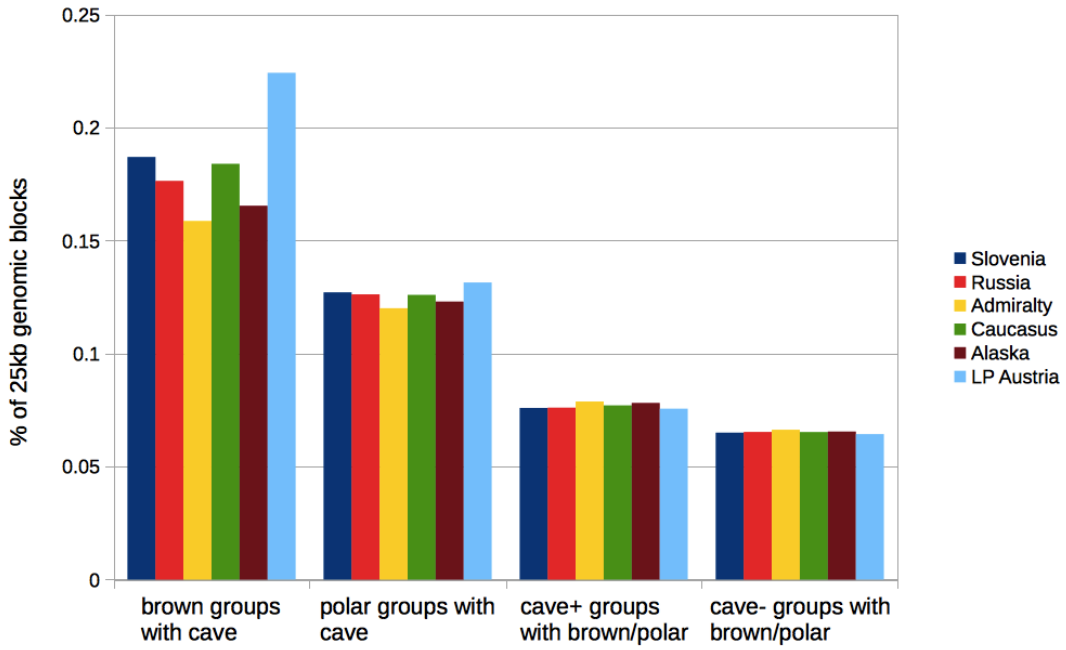


Figure 5.2. Here we show the frequency with which one individual falls within the diversity of the other clade. Columns are colored by the brown bear used in the test, cave+ is a European cave bear, cave– is a Caucasian cave bear. Variation is greatest in the rate with which brown bears group with cave bears and is ordered by the rate of admixture with cave bears as inferred by D and $\hat{f}$ statistics. LP Austria (light blue) has the most cave bear introgression, followed by Eurasian brown bears (dark blue, red, green), then North American brown bears (yellow, brown). Cave+ falls within the diversity of brown/polar bears more frequently than cave- suggesting the possibility of a smaller additional component of gene flow from brown bears into European cave bears.

Quantifying Gene Flow

Finally we quantified the amount of gene flow from cave bears into brown bears using the $\hat{f}$ statistic. We calculated $\hat{f}$ as $\hat{f}$(polar, X, European cave, European cave, black), we excluded *U. kudarensis* because it is so divergent from other cave bears that it shares substantially fewer alleles with other cave bears than would be expected, which in turn inflates our estimates of cave bear ancestry (Durand *et al.* 2011). We found extant brown bears possess 1.5-3% cave bear ancestry with greater cave bear ancestry in Eurasian brown bears and less cave bear ancestry in North American brown bears (Figure 5.3). The greatest observed cave bear ancestry was in the ancient Austrian Brown bears, which we calculated to have at least 3.5-4% cave bear ancestry.

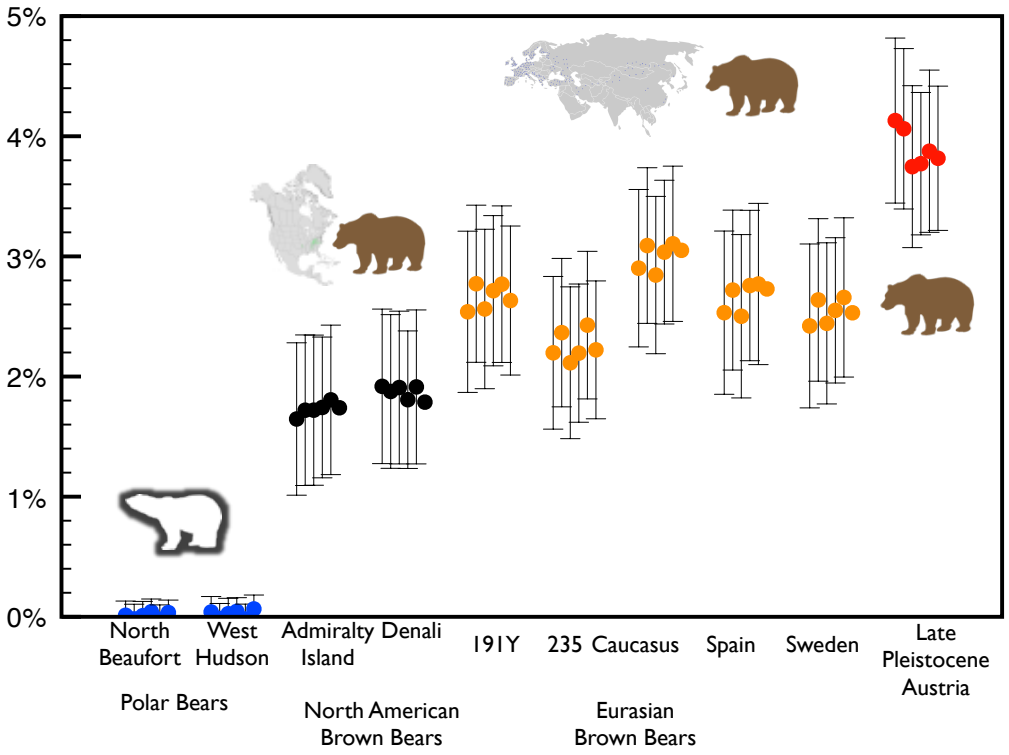


Figure 5.3. Here we show the amount of cave bear introgression inferred into polar (blue), and brown bears (black, orange and red). The Late Pleistocene Austrian brown bear (red) has the greatest cave bear ancestry, at least 3.5-4% of its genome, the Eurasian brown bears have the most cave bear ancestry among modern brown bears but all brown bears have significant cave bear ancestry. Six measurements per sample represent all possible combinations of European Cave Bear introgressors

$\hat{f}$(polar, X, European cave, European cave, black). Error bars are given as 95% confidence intervals, 1.95 weighted block jackknife standard errors.

Although, our directionality test did not detect evidence of gene flow from brown bears into cave bears we do find D-statistics consistent with greater gene flow between European cave bears and brown bears than between *U. kudarensis* and brown bears $\hat{f}$(*kudarensis, spelaeus, arctos, arctos,* Outgroup)=0.8% greater brown bear ancestry in European Cave bears. These data suggest that the major component of gene flow was from cave bears into ancient European brown bears but it is not possible to definitively determine from these data whether there was also some brown bear introgression into European cave bears or if these results are the product of the cave bear introgressors into brown bears being more closely related to European than Caucasian brown bears.

Discussion

How many species of cave bear

The relatively ancient divergence of the *kudarensis* cave bear from European cave bears and the lack of shared variation between *kudarensis* and other cave bears D(*kudarensis,* European cave bear, European cave bear, Black Bear) = 0.73 to 0.80, suggests substantial subdivision among cave bears. However, whether this warrants a separate species designation is difficult to determine. Polar bears and brown bears likely have much greater ecological divergence in terms of the niches occupied by the two species than European and Caucasian cave bears and the data here are insufficient to provide insights into the extent of reproductive isolation within the broad group cave bears. Therefore although the deep divergence and low levels of allele sharing between Caucasian and European cave bears suggest significant divergence I do not feel that either considering them multiple species or a single

species can be rejected from this study.

Admixture between Cave bears and Brown bears

Admixture between polar bears and brown bears is well documented from numerous studies (Cahill *et al.* 2013, 2015; Liu *et al.* 2014), and some studies have suggested an even greater scope for admixture within the genus *Ursus* (Kutschera *et al.* 2014). This study is the first to clearly show admixture between cave bears and brown bears and intriguingly reveals that at least some cave bear traits persist in modern brown bears.

None the less the extent of the contribution of cave bears to brown bear diversity seems limited, cave bears and brown bears were sympatric within Eurasia for many thousands of years (Stiller *et al.* 2010) but only a relatively small amount of gene flow is detected. This suggests that cave bears and brown bears likely exhibited limited reproductive isolation. Because cave bears are extinct we are limited in understanding the role of behavior and other prezygotic barriers in regulating gene flow, however, as more brown bear genomes are sequenced it may become possible to robustly test for cave bear ancestry deserts which may shed light on the limits of compatibility between these species.

By comparing the rate of cave bear ancestry in different brown bear populations we may be able constrain the timing of gene flow between the species. All brown bears exhibit greater cave bear ancestry than polar bears indicating that the gene flow detected here must be more recent than the divergence of brown bears and polar bears most recently estimated as 343-479 thousand years ago (Liu *et al.* 2014). North American brown bears possess some cave bear ancestry but less than the Eurasian brown bears. Because cave bears did not inhabit North America we can conclude that the cave bear ancestry in North American brown bears is the result of

cave bear introgression prior to their migration into North America within the last 100 thousand years (Kurtén & Anderson 1980; McLellan & Reiner 1994).

Conclusion

Whole genome sequencing of four subspecies of cave bear reveal substantial differentiation between cave bear groups (Figure 5.1, Table 5.2). It further reveals that cave bears admixed with brown bears contributing 1.5-3% of the genomes of modern brown bears (Figure 5.3). As a result a component of the genetic diversity of cave bears persists within living brown bears even after the cave bears extinction.

Additional Tables

P1	P2	D-statistic	weighted block jackknife	Z-score
191Y	235	-0.0103	0.0101	1.0275
191Y	Adm1	-0.0320	0.0107	2.9761
191Y	Cau	0.0186	0.0102	1.8191
191Y	Den	-0.0256	0.0102	2.5108
191Y	LS039	0.0053	0.0115	0.4638
191Y	NB	-0.0887	0.0101	8.7705
191Y	SB	-0.0890	0.0103	8.6472
191Y	Swe	0.0000	0.0115	0.0028
191Y	Uap	0.0344	0.0110	3.1378
191Y	WH2	-0.0847	0.0106	7.9785
235	Adm1	-0.0276	0.0102	2.6986
235	Cau	0.0268	0.0094	2.8458
235	Den	-0.0222	0.0099	2.2491
235	LS039	0.0173	0.0108	1.6052
235	NB	-0.0798	0.0097	8.1921
235	SB	-0.0800	0.0099	8.1167
235	Swe	0.0123	0.0105	1.1745
235	Uap	0.0528	0.0107	4.9117
235	WH2	-0.0766	0.0100	7.6927
Adm1	Cau	0.0489	0.0100	4.8733
Adm1	Den	0.0040	0.0108	0.3704
Adm1	LS039	0.0377	0.0110	3.4269

Adm1	NB	-0.0665	0.0103	6.4250
Adm1	SB	-0.0652	0.0105	6.2061
Adm1	Swe	0.0365	0.0107	3.4243
Adm1	Uap	0.0740	0.0112	6.6147
Adm1	WH2	-0.0594	0.0104	5.7002
Cau	Den	-0.0470	0.0096	4.9166
Cau	LS039	-0.0178	0.0104	1.7094
Cau	NB	-0.0984	0.0098	10.0876
Cau	SB	-0.0996	0.0100	9.9901
Cau	Swe	-0.0156	0.0103	1.5087
Cau	Uap	0.0268	0.0103	2.6096
Cau	WH2	-0.0951	0.0101	9.3676
Den	LS039	0.0359	0.0106	3.3905
Den	NB	-0.0676	0.0101	6.7008
Den	SB	-0.0679	0.0103	6.5943
Den	Swe	0.0337	0.0102	3.3119
Den	Uap	0.0733	0.0109	6.7496
Den	WH2	-0.0642	0.0104	6.1658
LS039	NB	-0.0859	0.0103	8.3621
LS039	SB	-0.0879	0.0105	8.3354
LS039	Swe	0.0002	0.0115	0.0143
LS039	Uap	0.0328	0.0115	2.8619
LS039	WH2	-0.0838	0.0105	7.9876
NB	SB	-0.0177	0.0355	0.4988
NB	Swe	0.0865	0.0102	8.4787
NB	Uap	0.1198	0.0109	11.0133
NB	WH2	0.0095	0.0337	0.2832
SB	Swe	0.0881	0.0103	8.5509
SB	Uap	0.1192	0.0109	10.9583
SB	WH2	0.0185	0.0377	0.4896
Swe	Uap	0.0383	0.0115	3.3302
Swe	WH2	-0.0805	0.0104	7.7776
Uap	WH2	-0.1135	0.0112	10.1740

Table 5.3 *D*-statistic tests for differential cave bear admixture with members of the Polar/Brown bear clade. *D*-statistic values for cave bear admixture with polar and brown bears. Brown and polar bears P1 and P2 are listed, all comparisons used the American black bear (Uam) as an outgroup. *D*-statistics were calculated for all four cave bears but results were similar so due to limited space we show only tests with GS136 as P3.

P1	P2	P3	D-statistic	Weighted block jackknife	Z-score
WK01	GS136	191Y	0.0027	0.0174	0.1559
WK01	GS136	235	0.0171	0.0168	1.0168
WK01	GS136	Adm1	0.0211	0.0169	1.2461
WK01	GS136	Cau	0.0143	0.0170	0.8393
WK01	GS136	Den	0.0161	0.0169	0.9560
WK01	GS136	LS039	0.0113	0.0172	0.6586
WK01	GS136	NB	0.0062	0.0175	0.3553
WK01	GS136	SB	0.0136	0.0177	0.7666
WK01	GS136	Swe	0.0092	0.0179	0.5128
WK01	GS136	Uap	-0.0284	0.0178	1.5932
WK01	GS136	WH2	0.0083	0.0176	0.4693
WK01	HV74	191Y	-0.0312	0.0099	3.1539
WK01	HV74	235	-0.0304	0.0094	3.2514
WK01	HV74	Adm1	-0.0348	0.0096	3.6165
WK01	HV74	Cau	-0.0295	0.0096	3.0869
WK01	HV74	Den	-0.0300	0.0097	3.0934
WK01	HV74	LS039	-0.0335	0.0102	3.2856
WK01	HV74	NB	-0.0380	0.0100	3.8056
WK01	HV74	SB	-0.0357	0.0099	3.5941
WK01	HV74	Swe	-0.0335	0.0099	3.3727
WK01	HV74	Uap	-0.0309	0.0102	3.0269
WK01	HV74	WH2	-0.0377	0.0103	3.6702
WK01	UD1838	191Y	-0.0024	0.0143	0.1674
WK01	UD1838	235	0.0137	0.0137	1.0006
WK01	UD1838	Adm1	0.0032	0.0142	0.2215
WK01	UD1838	Cau	0.0140	0.0138	1.0096
WK01	UD1838	Den	0.0048	0.0141	0.3403
WK01	UD1838	LS039	-0.0003	0.0148	0.0180
WK01	UD1838	NB	0.0120	0.0152	0.7872
WK01	UD1838	SB	0.0070	0.0159	0.4403
WK01	UD1838	Swe	0.0079	0.0146	0.5455
WK01	UD1838	Uap	0.0103	0.0146	0.7060
WK01	UD1838	WH2	0.0089	0.0157	0.5653
GS136	HV74	191Y	-0.0195	0.0122	1.6025
GS136	HV74	235	-0.0333	0.0117	2.8546
GS136	HV74	Adm1	-0.0398	0.0115	3.4579
GS136	HV74	Cau	-0.0197	0.0115	1.7054

GS136	HV74	Den	-0.0339	0.0118	2.8721
GS136	HV74	LS039	-0.0335	0.0122	2.7493
GS136	HV74	NB	-0.0329	0.0124	2.6437
GS136	HV74	SB	-0.0312	0.0125	2.5007
GS136	HV74	Swe	-0.0299	0.0123	2.4261
GS136	HV74	Uap	-0.0148	0.0123	1.2014
GS136	HV74	WH2	-0.0314	0.0128	2.4563
GS136	UD1838	191Y	0.0068	0.0210	0.3238
GS136	UD1838	235	0.0163	0.0193	0.8445
GS136	UD1838	Adm1	-0.0087	0.0200	0.4342
GS136	UD1838	Cau	0.0151	0.0199	0.7592
GS136	UD1838	Den	0.0108	0.0198	0.5449
GS136	UD1838	LS039	0.0083	0.0210	0.3954
GS136	UD1838	NB	0.0136	0.0216	0.6321
GS136	UD1838	SB	0.0077	0.0216	0.3562
GS136	UD1838	Swe	0.0120	0.0206	0.5815
GS136	UD1838	Uap	0.0852	0.0210	4.0503
GS136	UD1838	WH2	0.0125	0.0217	0.5734
HV74	UD1838	191Y	0.0241	0.0100	2.4176
HV74	UD1838	235	0.0302	0.0097	3.1156
HV74	UD1838	Adm1	0.0336	0.0097	3.4658
HV74	UD1838	Cau	0.0263	0.0097	2.7260
HV74	UD1838	Den	0.0305	0.0099	3.0694
HV74	UD1838	LS039	0.0278	0.0103	2.7065
HV74	UD1838	NB	0.0395	0.0103	3.8152
HV74	UD1838	SB	0.0346	0.0106	3.2559
HV74	UD1838	Swe	0.0291	0.0105	2.7722
HV74	UD1838	Uap	0.0336	0.0102	3.2968
HV74	UD1838	WH2	0.0382	0.0108	3.5380

Table 5.4 *D*-statistic tests for differential brown/polar bear admixture with members of the cave bear clade. *D*-statistic values for differential admixture between cave bears and brown or polar bears. All comparisons used the American black bear (Uam) as an outgroup.

References

Baca M, Stankovic A, Stefaniak K *et al.* (2012) Genetic analysis of cave bear specimens from Niedźwiedzia Cave, Sudetes, Poland. *Palaeontologia Electronica*, **15**.

Bidon T, Janke A, Fain SR *et al.* (2014) Title: Brown and polar bear Y chromosomes reveal extensive male-biased gene flow within brother lineages. *Molecular*

biology and evolution, msu109–.

Boc A, Diallo AB, Makarenkov V (2012) T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research*, **40**, W573–W579.

Cahill JA, Green RE, Fulton TL *et al.* (2013) Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. (MW Nachman, Ed,). *PLoS genetics*, **9**, e1003345.

Cahill JA, Stirling I, Kistler L *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular ecology*, **24**, 1205–1217.

Dabney J, Knapp M, Glocke I *et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15758–63.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Fortes GG, Paijmans JLA (2015) Analysis of Whole Mitogenomes from Ancient Samples. In:, pp. 179–195.

Fulton TL (2012) Setting up an ancient DNA laboratory. In: *Ancient DNA, methods and protocols* (eds Shapiro B, Hofreiter M), pp. 1–11. Springer protocols.

Gamba C, Jones ER, Teasdale MD *et al.* (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, **5**, 5257.

Gansauge M-T, Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, **8**, 737–748.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, **328**, 710–22.

Hailer F, Kutschera VE, Hallström BM *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science (New York, N.Y.)*, **336**, 344–7.

Hofreiter M (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.

KNAPP M, ROHLAND N, WEINSTOCK J *et al.* (2009) First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Molecular Ecology*, **18**, 1225–1238.

Krause J, Unger T, Noçon A *et al.* (2008) Mitochondrial genomes reveal an explosive

radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC evolutionary biology*, **8**, 220.

Kunsch HR (1989) The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics*, **17**, 1217–1241.

Kurtén B, Anderson E (1980) *Pleistocene mammals of North America*. Columbia University Press.

Kutschera VE, Bidon T, Hailer F *et al.* (2014) Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. *Molecular Biology and Evolution*, **31**, 2004–2017.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **26**, 589–95.

Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–7.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–9.

Li B, Zhang G, Willersleve E, Wang J (2011) GigaDB Dataset - DOI 10.5524/100008 - Genomic data from the polar bear (Ursus maritimus).

Liu S, Lorenzen ED, Fumagalli M *et al.* (2014) Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, **157**, 785–794.

McLellan B, Reiner DC (1994) A Review of Bear Evolution. *Ursus*, **9**, 85–96.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, **2010**, pdb.prot5448.

Miller W, Schuster SC, Welch AJ *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2382–90.

PACHER M, STUART AJ (2009) Extinction chronology and palaeobiology of the cave bear (*Ursus spelaeus*). *Boreas*, **38**, 189–206.

Patterson N, Moorjani P, Luo Y *et al.* (2012) Ancient admixture in human history. *Genetics*, **192**, 1065–93.

Pinhasi R, Fernandes D, Sirak K *et al.* (2015) Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone (MD Petraglia, Ed,). *PLOS ONE*, **10**, e0129102.

Prüfer K, Stenzel U, Hofreiter M *et al.* (2010) Computational challenges in the analysis of ancient DNA. *Genome biology*, **11**, R47.

Richards MP, Pacher M, Stiller M *et al.* (2008) Isotopic evidence for omnivory among European cave bears: Late Pleistocene Ursus spelaeus from the Pestera cu Oase, Romania. *Proceedings of the National Academy of Sciences*, **105**, 600–604.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

St John J (2011) SeqPrep: Tool for stripping adaptors and/or merging paired reads with overlap into single reads.

Stiller M, Baryshnikov G, Bocherens H *et al.* (2010) Withering Away--25,000 Years of Genetic Decline Preceded Cave Bear Extinction. *Molecular Biology and Evolution*, **27**, 975–978.

Stiller M, Molak M, Prost S *et al.* (2014) Mitochondrial DNA diversity and evolution of the Pleistocene cave bear complex. *Quaternary International*, **339**, 224–231.

Synthesis

The application of genome technology to the field of evolutionary biology has had profound and diverse impacts. Scientific understanding of polar bears and brown bears evolutionary history provides a striking microcosm of wider trends within the field. I feel uniquely fortunate to have had the opportunity to be involved in and contribute to some of the first whole genome based studies of these species.

At the outset of my dissertation research there were clear leading hypotheses regarding the evolutionary history of polar bears and brown bears some of which have been supported by genomic research while others have not. Polar bears and brown bears were very closely related, and Admiralty, Baranof and Chichagof (ABC) Islands brown bears were either the source population of polar bears (McLellan & Reiner 1994; Heaton et al. 1996) or polar bears had received multiple introgressions of brown bear mitochondria and presumably nuclear genes (Edwards et al. 2011).

Our first paper (Chapter 1) allowed us to question or overturn some of these assumptions. While we did find gene flow between polar bears and ABC islands brown bears (Chapter 1, Figure 1.3) we were able to show, for the first time that the direction of gene flow was from polar bears into brown bears. Polar bears' profound lack of genetic diversity (Chapter 1, Figure 1.2, 1.4, 2.4, 5.2) is inconsistent with any introgression into from other species polar bears since the common ancestors of the majority of the polar bears' genomes 130-650 thousand years ago (Chapter 1, Table 1.1).

The initial investigation of the ABC islands admixture history also revealed a fundamental surprise regarding the nature of admixture between polar bears and brown bears. Unlike the human/neaderthal admixture that is hypothesized to have been relatively peripheral even in the populations effected (Green et al. 2010) the

elevated frequency of polar bear ancestry at maternally inherited loci in ABC bears led us to propose a polar bear origin for the ABC islands brown bear population (Chapter 1).

In chapters 1 and 2 we developed a hypothesis that the pre-admixture population inhabiting the ABC islands were polar bears, but that male biased immigration of brown bears gradually decreased the amount of polar bear ancestry to the current levels 6-9% polar bear ancestry (Chapter 2). The key predictions of this hypothesis is that admixture was characterized by the formation of intermediate populations that gradually homogenized toward the continental norm rather than peripheral admixture where the frequency of introgressed alleles in the population was always low. The population conversion model of admixture is important because it predicts a much more powerful and disruptive role for admixture in the evolution of affected populations. In this model admixture is a broadly transformative process not simply an interesting side effect of post-divergence contact that may introduce a few beneficial traits (Huerta-Sánchez et al. 2014) but does not fundamentally alter the make up of the population.

In addition to refining the population conversion hypothesis, the study described in chapter 2 led us to a greater appreciation on the scale of the impact of admixture upon brown bears. In chapter 1 our limited sample set led us to detect only a small amount of polar bear introgression into ABC islands brown bears (Chapter 1) and so we predicted that admixture might be a process that impacted odd island populations such as the ABC islands (Chapter 1) and Ireland (Edwards et al. 2011) but probably did not significantly impact continental populations. Sequencing additional brown bears in mainland Alaska and Scandinavia overturned that hypothesis and revealed that all North American brown bears possessed polar

bear ancestry (Chapter 2), requiring a reassessment of the scope of the impact of admixture.

This work culminated with the study described in chapter 3. By sequencing nuclear genomic material from 10 Irish brown bears spanning the last glacial maximum we were able to directly investigate the impact of admixture on a potentially admixed population over the course of significant climatic fluctuations. The greater than 20% polar bear ancestry observed in the Irish brown bear living nearest in time to the end of the LGM and during the retreat of glaciers from Ireland provides direct evidence for the formation of intermediate polar/brown bear hybrid populations during periods of rapid climate warming. The decline in polar bear ancestry observed within Irish brown bears after the LGM further supports the population conversion hypothesis's expectation that initially high levels of admixed ancestry may be eroded if migration from unadmixed population is allowed to continue (Figure 3.2).

In chapter 3 we also sought to examine the geographic scope of admixture more fully by examining at least one population from the east and west coasts of the Atlantic and Pacific oceans that would have been near the frontier of polar and brown bears' ranges at the end of the last glacial maximum. In all four cases (ABC islands, Ireland, Kunashir Island and Quebec) we found evidence of polar bear introgression into brown bears. This indicates that the admixture observed in the ABC islands and Ireland was likely to have been a general phenomenon, while the local conditions likely play a role in the scope of admixture the potential for substantial admixture between polar bears and brown bears does not appear to be limited to any single geographic region.

My other projects, hPSMC (chapter 4) and the cave bear study (chapter 5) are

not as directly related to the core progression of polar/brown bear admixture papers (chapters 1-3) but they do share important connections. The prevalence of post divergence gene flow from polar bears into brown bears poses the question of exactly how to describe species divergence. As I discussed briefly in the introduction, speciation is best considered as a process rather than an event. An important but often difficult to diagnose stage in that process is the time at which post divergence gene flow ceases and complete genetic isolation begins.

In various projects throughout my PhD beginning in Chapter 1, I have used Heng Li and Richard Durbin's Pairwise Sequentially Markovian Coalescent (PSMC) a program which estimates ancestral effective population size through time by measuring the genetic divergence between an individual's parents as evidenced by heterozygosity. But this raises the question, what would happen if the individual were a first generation (F1) hybrid with each parent being a member of a different species. We hypothesized and then demonstrated that this would produce a characteristic pattern in the result of PSMC that could be used to determine when populations became fully isolated from one another. Applying many of the approaches that we employed to measure admixture from low coverage data (Green et al. 2010) we were able to configure hPSMC to similarly provide divergence time estimates from low coverage and even ancient DNA damaged samples.

Finally the cave bears project (chapter 5) provides an interesting expansion upon the polar and brown bear admixture studies (chapters 1-3). Using similar methods we were able to further expand the now quite extensive admixture history of brown bears. At the onset of my dissertation research there was no scientific consensus of any gene flow from other species into brown bears. With these studies completed we have not found a single brown bear that can be said to be free of

introgression from another species. Many brown bear populations have substantial detectable polar bear ancestry and all brown bears have detectable cave bear ancestry. It is unclear what, if any evolutionary consequences this capacity for receiving gene flow has for brown bears. Brown bears are by far the most wide ranging and adaptable of the three, or perhaps four if we could *kudarensis* cave bears as distinct, species involved in the admixture system and they appear to the the principal recipient of introgression. One could argue that this reception of gene flow is somehow invigorating and increases adaptive potential, and that may prove to be true. However, I personally suspect that the opposite is more likely that it is the width of the brown bears niche allows hybrids to backcross more readily.

As genomic investigations of admixture continue to expand it will be interesting to see to what degree the results in polar, brown and cave bears coincide with other species. Does admixture often take the form of the formation of genetically intermediate populations? Does climate change promote hybridization? Does ecological specialization play an important role in regulating gene flow, including a phase of divergence between generalists and specialists where gene flow can only transpire from the specialist into the generalist?

References

Edwards, C.J. et al., 2011. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology : CB*, 21(15), pp.1251–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21737280 [Accessed October 23, 2013].

Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, 328(5979), pp.710–22. Available at: http://www.sciencemag.org/content/328/5979/710.abstract [Accessed October 17, 2013].

Heaton, T.H., Talbot, S.L. & Shields, G.F., 1996. An Ice Age Refugium for Large Mammals in the Alexander Archipelago, Southeastern Alaska. *Quaternary Research*, 46(2), pp.186–192. Available at: http://www.sciencedirect.com/science/article/pii/S0033589496900587

[Accessed October 24, 2013].

Huerta-Sánchez, E. et al., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512(7513), pp.194–197. Available at: http://www.nature.com/doifinder/10.1038/nature13408 [Accessed August 24, 2016].

McLellan, B. & Reiner, D.C., 1994. A Review of Bear Evolution. *Bears: Their Biology and Management*, 9, p.85. Available at: http://www.jstor.org/stable/10.2307/3872687?origin=crossref [Accessed August 24, 2016].