Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change

Webb Miller[a,1], Stephan C. Schuster[a,b,1], Andreanna J. Welch[c], Aakrosh Ratan[a], Oscar C. Bedoya-Reina[a], Fangqing Zhao[b,d], Hie Lim Kim[a], Richard C. Burhans[a], Daniela I. Drautz[b], Nicola E. Wittekindt[b], Lynn P. Tomsho[a], Enrique Ibarra-Laclette[e], Luis Herrera-Estrella[e,2], Elizabeth Peacock[f], Sean Farley[g], George K. Sage[f], Karyn Rode[h], Martyn Obbard[i], Rafael Montiel[e], Lutz Bachmann[j], Ólafur Ingólfsson[k,l], Jon Aars[m], Thomas Mailund[n], Øystein Wiig[j], Sandra L. Talbot[f], and Charlotte Lindqvist[c,1,2]

[a]Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802; [b]Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551; [c]Department of Biological Sciences, University at Buffalo, Buffalo, NY 14260; [d]Computational Genomics Laboratory, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China; [e]Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, 36821, Mexico; [f]US Geological Survey, Alaska Science Center, Anchorage, AK 99508; [g]Alaska Department of Fish and Game, Anchorage, AK 99518; [h]US Fish and Wildlife Service, Anchorage, AK 99503; [i]Ontario Ministry of Natural Resources, Peterborough, ON, Canada K9J 8M5; [j]National Centre for Biosystematics, Natural History Museum, University of Oslo, 0318 Oslo, Norway; [k]Department of Earth Sciences, University of Iceland, 101 Reykjavik, Iceland; [l]University Centre in Svalbard, 9171 Longyearbyen, Norway; [m]Fram Centre, Norwegian Polar Institute, 9296 Tromsø, Norway; and [n]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark

Polar bears (PBs) are superbly adapted to the extreme Arctic environment and have become emblematic of the threat to biodiversity from global climate change. Their divergence from the lower-latitude brown bear provides a textbook example of rapid evolution of distinct phenotypes. However, limited mitochondrial and nuclear DNA evidence conflicts in the timing of PB origin as well as placement of the species within versus sister to the brown bear lineage. We gathered extensive genomic sequence data from contemporary polar, brown, and American black bear samples, in addition to a 130,000- to 110,000-y old PB, to examine this problem from a genome-wide perspective. Nuclear DNA markers reflect a species tree consistent with expectation, showing polar and brown bears to be sister species. However, for the enigmatic brown bears native to Alaska's Alexander Archipelago, we estimate that not only their mitochondrial genome, but also 5–10% of their nuclear genome, is most closely related to PBs, indicating ancient admixture between the two species. Explicit admixture analyses are consistent with ancient splits among PBs, brown bears and black bears that were later followed by occasional admixture. We also provide paleodemographic estimates that suggest bear evolution has tracked key climate events, and that PB in particular experienced a prolonged and dramatic decline in its effective population size during the last ca. 500,000 years. We demonstrate that brown bears and PBs have had sufficiently independent evolutionary histories over the last 4–5 million years to leave imprints in the PB nuclear genome that likely are associated with ecological adaptation to the Arctic environment.

demographic history | hybridization | mammalian genomics | phylogenetics

Genome-scale studies of speciation and admixture have become essential tools in evolutionary analyses of recently diverged lineages. For example, paradigm-shifting genomic research on archaic and anatomically modern humans has identified critical gene flow events during hominin history (1, 2). However, aside from several analyses of domesticated species and their wild relatives (e.g., ref. 3), studies that use whole-genome sequencing to investigate admixture in wildlife populations are only now beginning to emerge.

The bear family (Ursidae, Mammalia) represents an excellent, largely untapped model for investigating complex speciation and rapid evolution of distinct phenotypes. Although polar bears (PBs; *Ursus maritimus*) and brown bears (*Ursus arctos*) are considered separate species, analyses of fossil evidence and mitochondrial sequence data have indicated a recent divergence of PBs from

within brown bears (surveyed in ref. 4). For example, phylogenetic analyses of complete mitochondrial genomes, including from a unique 130,000- to 110,000-y-old PB jawbone from Svalbard, Norway, confirmed a particularly close relationship between PB and a genetically isolated population of brown bears from the Admiralty, Baranof, and Chichagof islands in Alaska's Alexander Archipelago (hereafter referred to as ABC brown bears) and suggested a split of their maternal lineages ~150 kya (4). This recent divergence and paraphyletic relationship raises questions whether there has been sufficient time for full reproductive isolation to develop (5). Despite being fully distinct species throughout most of their ranges (6), interbreeding between polar and brown bears has occurred in captivity (7), and although extremely rare, natural hybrids have recently been documented. Indeed, limited evidence from short stretches of mitochondrial DNA suggests that hybridization may have occurred between polar and brown bears shortly after they diverged from one another (8). However, further evidence from biparentally inherited nuclear DNA is required to critically evaluate this possibility, and in particular to determine what fraction of the extant bear genome has been sculpted by gene flow between brown bears and PBs. Although nuclear DNA sequence data recently indicated that the PB may have become genetically distinct from brown bears approximately 600 kya (9), a gene-by-gene sequencing approach of single nuclear markers clearly lacks sufficient power to detect potential ancient admixture.

Although field and molecular studies have produced critical information about the ecology and evolutionary relationships of the PB, advances in next-generation sequencing technology have only recently made full-genome studies of such wildlife species possible (10, 11). A whole-genome analysis of the PB may provide a more complete picture of its evolutionary history, possible signatures of admixture, and clues about the genetic basis of adaptive phenotypic features facilitating life in the Arctic, all of which are imperative for gaining a better understanding of possible responses to current and future climatic changes.

We gathered extensive genome sequence data from modern polar, brown, and American black bear samples, plus a ~120,000-y-old PB, to address the following questions. (*i*) What is the more precise association between the PB and its sister species, the brown bear; and do we find any signatures of past genetic interchange between the two species? (*ii*) Did the PB indeed evolve recently, as suggested by mitochondrial DNA and fossil evidence, or did it have an older origin, as demonstrated by nuclear DNA loci? (*iii*) Can we deduce any past responses in ancient bear population histories that may be connected with climatic changes? Our comparative genomic analyses have facilitated investigation of the timing of divergence and hybridization in brown and PB lineages through the last 4 to 5 million years of climatic variation, and provide a window into genetic underpinnings of adaptive features, making it possible to begin to investigate the unique characteristics of the PB and how these may reflect its excellent adaptation to the extreme Arctic environment.

Results and Discussion

Genome-Scale Sequencing of Three Bear Species. We performed deep genome sequencing (*SI Appendix*) for a PB (referred to as "PB7"), two ABC brown bears ("ABC1" and "ABC2"), a non-ABC brown bear ("GRZ"), and an American black bear (*Ursus americanus*, hereafter referred to as black bear). We generated from 25-fold to 100-fold coverage for each of these individuals, with deepest coverage for the PB (*SI Appendix*, Table S1). We assembled the 3.15 billion reads for PB7 using SOAPdenovo (12). Mate-pair information produced 1,229,681 scaffolds and contigs, with an N50 length of 61 kb and span of 2.53 billion bases (Gb), which is close to the estimated genome size of the giant panda (2.4 Gb) (13). Low-coverage genome sequence data, between threefold and 10-fold each, were produced for an additional 22 PBs (*SI Appendix*, Table S1). As a means of validating our PB assembly and making nuclear genomic sequences more useful, we also sequenced transcriptomes of individual polar and brown bears (*SI Appendix*, Table S2).

We aligned the reads from all 27 individuals to our PB assembly and searched for genomic positions where two different nucleotides could be confidently identified (*SI Appendix*). We term such positions SNPs, even though the different nucleotides can be from different species, e.g., a nucleotide fixed in PBs but differing from the orthologous black bear nucleotide. This procedure yielded 13,038,705 SNPs, of which 2,540,010 are polymorphic in PB. We produced a second set of SNPs by aligning the reads from all samples to the dog genome (*Canis familiaris* assembly version 2.0) and applying the same SNP-calling procedure. That approach produced 12,023,192 SNPs, of which 1,914,757 are polymorphic among PBs. This latter set contains a slightly smaller number of SNP calls (compared with using the PB assembly), which could be a result of genomic regions deleted in the dog lineage since the split between dog and bear. SNPs based on the dog genome, however, benefit from excellent gene annotations, which make it simple to identify synonymous and nonsynonymous coding SNPs with good reliability, and allow SNPs to be mapped to the corresponding locations on dog chromosomes. Having two sets of SNPs produced by different but complementary approaches helped us to evaluate the robustness of the observations made below. Among our discovered SNPs, 26,001 result in amino acid replacements

[single amino acid polymorphisms (SAPs)] as traced to coding regions of putatively orthologous genes known from the dog genome (canFam2). Of these SAPs, 7,014 were found to be possibly or probably "damaging" by computational predictions, i.e., ostensibly causing protein functional changes (*SI Appendix*). The number of alleles in all polymorphic SNP loci and number of high-quality polymorphic SAP alleles shared among the three bear species (Fig. 1) demonstrate a significant number of fixed alleles within each species, as well as a higher number of shared alleles between polar and brown bear, and brown and black bear, than between polar and black bear. The relatively lower overall number of alleles in the black bear may, however, derive from the lower sequence coverage compared with the other two deeply sequenced species (*SI Appendix*, Table S1).

We also produced 164 Gb of nonamplified shotgun sequence data for a 130,000- to 110,000-y-old PB specimen from Svalbard, Norway (*SI Appendix*, Table S1). The sequence reads from this ancient bear were aligned to our PB genome assembly, and, for 3,293,332 SNPs, we could confidently determine at least one allele in the ancient sample. We observed 16,338 alleles in the ancient PB sample that were similar to alleles in the three brown bears but not to modern PB alleles. With the SNPs based on the dog assembly, we determined one or both alleles for 2,837,892 SNPs in the ancient sample, and found 15,527 ancient PB alleles that were identical to alleles in brown bears but not modern PBs.

Discordance of Mitochondrial and Nuclear Evolutionary Histories. A fundamental question in the evolution of the PB concerns the exact nature of its close relationship with the brown bear. For all the bear individuals sequenced, even those with a low average coverage of the nuclear genome, we had ample data to assemble their mitochondrial genomes. Phylogenetic analyses (*SI Appendix*), which also included previously assembled and publicly available mitochondrial genomes (*SI Appendix*, Table S3), confirm that ABC brown bears are, with these data, more closely related to PBs than to other brown bears (Fig. 2*A*). As such, maternal relationships among these bear species do not follow taxonomic boundaries. To further address the enigmatic phylogenetic position of the ABC brown bears, we reconstructed a phylogenetic tree based on genome-wide SNPs from nuclear autosomal loci. These markers exhibit a different pattern from the mitochondrial genome: the two ABC brown bear individuals, which group together, are as their morphology suggests, more closely related to the non-ABC brown bear (GRZ) than they are to PBs (Fig. 2*B*). Moreover, these analyses confirm the ancient PB groups as sister to all modern PBs sampled, including modern bears from Svalbard, in the Barents Sea (4). Although this demonstrates that the ancient PB is indeed most closely related to modern PBs, it is important that extant Svalbard PBs are more closely related to extant PBs from Alaska than they are to this extinct Svalbard lineage. Although data from more extant and extinct PBs is needed for confirmation, these relationships support a hypothesis of extinction and replacement of earlier progenitor lineages in the Barents Sea PB population, and perhaps the entire extant PB genetic stock.

The clear discordance between mitochondrial and nuclear genomes in the phylogenetic placement of the ABC brown bears (Figs. 2 and 3) mirrors that found in the evolutionary histories of archaic and anatomically modern human lineages (2). Similarly, we hypothesize that two main evolutionary scenarios may explain these patterns. First, the ABC brown bears and PBs may share a maternal history as a result of admixture between ancestors of these two lineages, followed by capture of one bear mitochondrial genome by the other bear lineage. The discordance may also result from random assortment of ancestral genetic lineages due to genetic drift (incomplete lineage sorting), which may have allowed divergent mitochondrial lineages to survive, perhaps in small, isolated bear populations, while becoming lost in others. It has also been suggested that phylogenies inferred from sequence data on
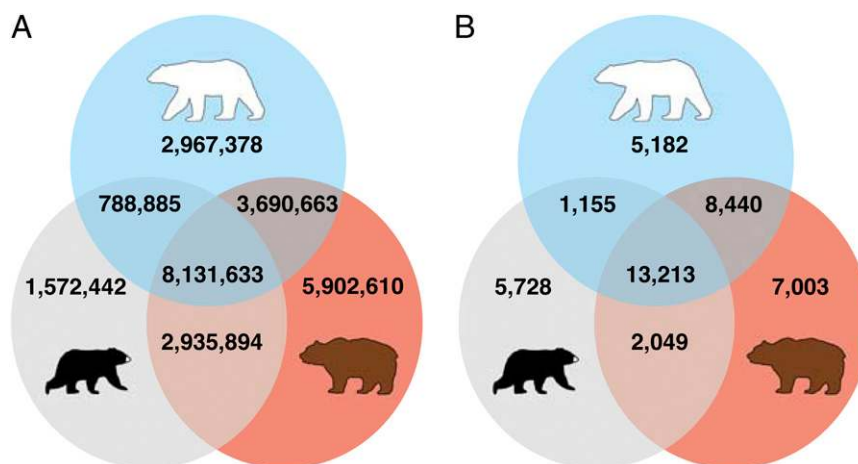
Fig. 1. The number of alleles unique to and shared by three bear species—polar, brown, and American black bear—in all SNP loci (*A*), and high-quality polymorphic variants resulting in an SAP (*B*).

nonrecombining chromosomes, such as mitochondrial DNA, may be distorted by selection (14). Therefore, we tested for evidence of positive selection in the 13 protein coding genes of the mitochondrial genome by using the dataset of 36 brown bears and PBs (*SI Appendix*). Results of our analysis demonstrated no evidence for positive selection on the branches leading to ABC brown bear plus PB, nor to PB itself, excluding a hypothesis that selective sweeps on mitochondrial DNA could be a major force in driving maternal evolutionary relationships between brown bears and PBs.

PB Divergence and Signals of Admixture. To make inferences about split times and gene flow, we applied a coalescence hidden Markov model to four of our deep-coverage bear genomes: the PB, one ABC brown bear, the non-ABC brown bear, and the black bear.

We first used a simple isolation model (15), comparing pairs of genomes under the assumption of allopatric speciation. However, for all comparisons, we obtained estimates of very recent split times and very large ancestral effective population sizes (N_e), consistent with mis-specification of the demographic model and suggesting that the true demographics involved are not simple splits but instead initial splits followed by prolonged periods with structured populations and gene flow. We therefore applied an extended model estimating an initial split time followed by a period of gene flow before a complete split (*SI Appendix*). We estimated an initial split between black bears and their sister lineage 4 to 5 Mya followed by gene flow until 100 to 200 kya (Fig. 3). This split time estimate is within the range of previously reported estimates based on molecular data (*SI Appendix*, Table S4). Within the
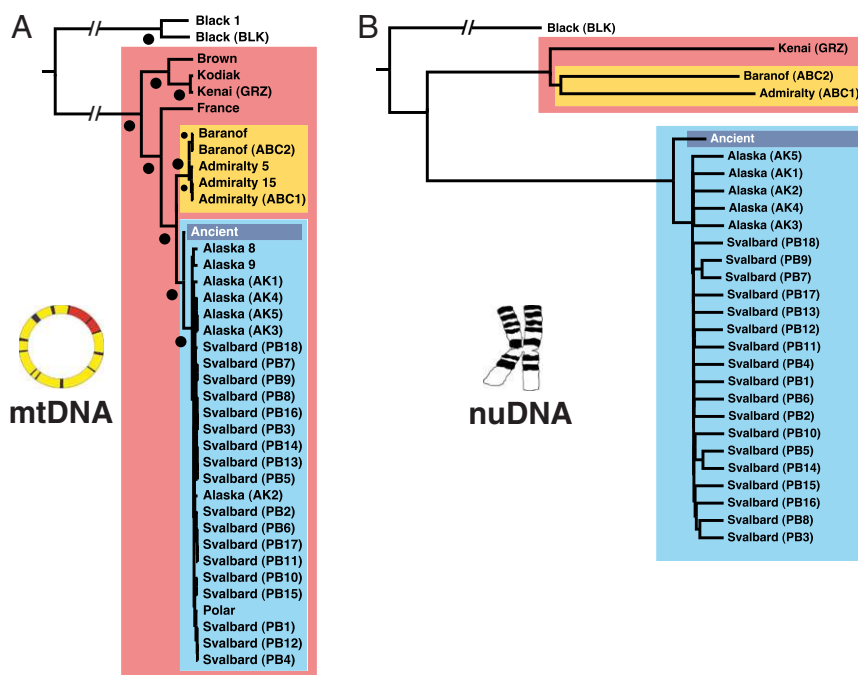

Fig. 2. Phylogenetic reconstruction of relationships among bears. (*A*) Bayesian maximum clade credibility tree reconstructed from mitochondrial genomes (mtDNA), with filled black circles at individual nodes indicating posterior probabilities greater than 0.99 and bootstrap support greater than 99% (diagonal lines indicate truncated branches, and, for the purpose of display, the cave bear genomes have been excluded; *SI Appendix*, Fig. S5, shows a complete tree). (*B*) Neighbor joining tree of genetic distances calculated from ~12 million nuclear-genome SNP markers (nuDNA).
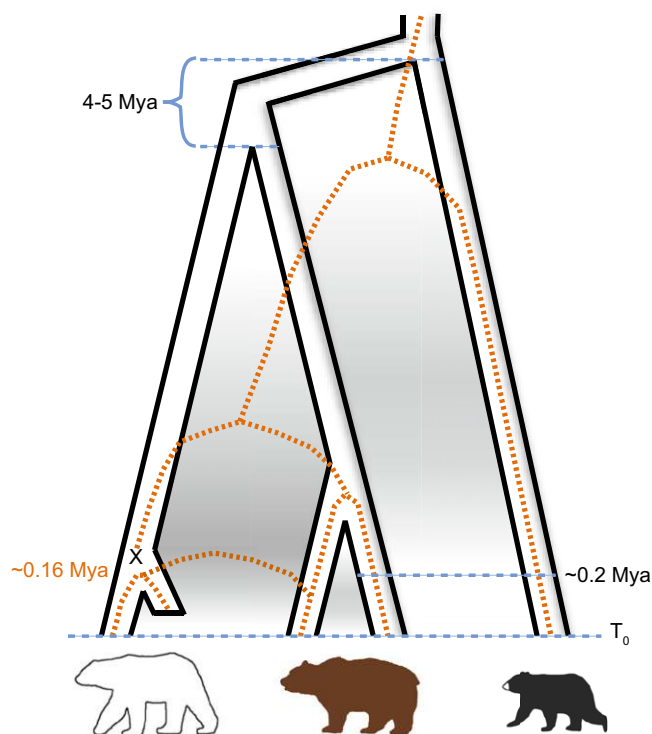
Fig. 3. Phylogenetic discordance and divergence among bears. A schematic species tree (black outline) highlights the discordance between mtDNA and nuclear histories, with the dashed orange line representing the mtDNA gene tree. The figure illustrates introgression and replacement (marked with an X) of PB mitochondrial DNA with an ABC brown bear mitochondrial genome, although the opposite scenario, i.e., capture of PB mitochondrial genome in modern ABC brown bear, cannot be excluded with these data. Two hypotheses of lineage splitting and admixture based on a coalescence hidden Markov model are indicated: (*i*) an ancient split (4–5 Mya) between the black bear and the brown bear/PB lineage, followed by intermittent gene flow (gray shading) ending by 200 to 100 kya, and (*ii*) a similarly ancient split between the PBs and brown bears followed by extensive gene flow (gray shading) between ABC brown bears and PBs until very recently. The ancient PB's lineage is indicated as extinct.

brown bear/PB lineage, the estimate for their initial split had a bimodal time distribution, indicating a very recent split or one close to the split with black bear (*SI Appendix*, Fig. S11). The estimate for when gene flow ceased between polar and ABC brown bears was not significantly different from 0 y. Although the extended model may not completely reflect the underlying demographics and further analyses are necessary, this result is most consistent with a hypothesis of an ancient initial split between brown bears and PBs at approximately the same time as brown vs. black bear, approximately 4 to 5 Mya, followed by a period of complete lineage separation (or one with very little gene flow) and later by recent admixture. It is noteworthy that these ancient split estimates of approximately 5 Mya coincide with the Miocene–Pliocene boundary, a period of environmental change that may have launched a radiation of bear species (16), and that perennial sea ice, possibly providing suitable polar habitat, has existed in the central Arctic Ocean since the Mid-Miocene (17).

To further investigate possible signals of admixture in brown bears and PBs, we searched for intervals in the ABC brown bear genomes that are more similar to those of PBs than to non-ABC brown bears (i.e., those that are "PB-like" as opposed to "GRZ-like"). We first estimated the fractions of the two sequenced ABC brown bear genomes that are PB-like, using a principal component analysis (PCA) of the SNP data (18, 19) (*SI Appendix*). Although this analysis shows considerable diversity between the two ABC brown bears (Fig. 4*A*), it estimates that 5.5% of the ABC1 genome and 9.4% of the ABC2 genome are PB-like. If we apply the same approach to the non-ABC brown bear (i.e., GRZ), only ~1.5% of its genome is PB-like.

We then partitioned the autosomes in the two ABC brown bear genomes into intervals that are PB-like in neither, one, or both chromosomes (*SI Appendix*). We mapped these intervals (Fig. 4*B* and *SI Appendix*, Fig. S12), which estimated that 7.5% of the ABC1 genome and 10.6% of the ABC2 genome are PB-like, in reasonable agreement with the PCA-based analysis (simulations indicate that the PCA tool systematically underestimates the admixture fraction; *SI Appendix*). Importantly, many of the PB-like intervals are likely too long and too divergent between the two ABC brown bears sequenced (*SI Appendix*, Fig. S12) to be easily accounted for by ancestral lineage sorting, as recombination should have fragmented these regions over time (20). In some cases, intervals are approximately four million bases long and mostly PB-like in one ABC brown bear and GRZ-like in the other (*SI Appendix*, Fig. S12). As such, geologically recent admixture gains credence for explaining the pattern observed in the nuclear and mitochondrial data.

Admixture maps (Fig. 4*B* and *SI Appendix*, Fig. S12) can also be used to identify genomic regions introduced into a bear lineage by admixture and perhaps preferentially retained by selective forces (21) For example, it has recently been suggested that modern humans may have inherited alleles critical to immune function from an advantageous ancient admixture event with archaic humans (22). One such interval with selective potential in bears contains the homologue of *ALDH7A1*. This gene has been shown to protect humans against hyperosmotic stress (23), and it is conceivable that PB-like variants may have provided a selective advantage for PBs and coastal brown bears, such as the ABC brown bears, in a shift to a marine environment (Fig. 4*C* and *SI Appendix*).

Differentiation Among Populations of PBs and Brown Bears. To assess whether the patterns of admixture identified from analysis of the genome-wide SNP dataset hold over a broader geographical range of brown bears and PBs, we extended our sampling to include individuals from three different continents (Fig. 5*A*). We selected a subset of 100 SNPs (identified from the genome sequences of the ancient PB and 23 modern PBs) and amplified and genotyped these SNPs in 118 bears, including 58 PB individuals and nine ABC and 51 non-ABC brown bears (*SI Appendix*, Table S5). Analyses of this population-level dataset indicate that although the ABC brown bears are closer to other brown bears than they are to PBs (Fig. 5*B*), some individuals share as much as 11% of their genetic background with PBs (Fig. 5*C*), which is consistent with the estimates reported above. Additionally, two of the modern PBs share as much as 20% of their genetic background with all brown bears, whereas the ancient PB shares as much as 25% (Fig. 5*C*).

Further, to investigate if the SNPs we discovered can be used to detect population genetic structure within PBs, we analyzed a subset of data representing only the modern PB individuals. From our genome sequencing, we identified 2,540,010 polymorphic SNPs among PBs, which is low compared with the number found with comparable data from humans, a relatively monomorphic species (*SI Appendix*). The estimated average fixation index (F_{ST}) between the five Alaskan PBs (as one population) and six of the highest coverage Svalbard (Barents Sea) PBs (PB1–4, 6, and 8, as a second population) was only 0.0785. By comparison, applying the same procedure to a closely analogous dataset of five individuals from each of two human groups, European and Asian, we found an average F_{ST} of 0.1928, indicating considerably less differentiation between the two PB populations. However, our PB F_{ST} may be underestimated for the species as a whole, as the two
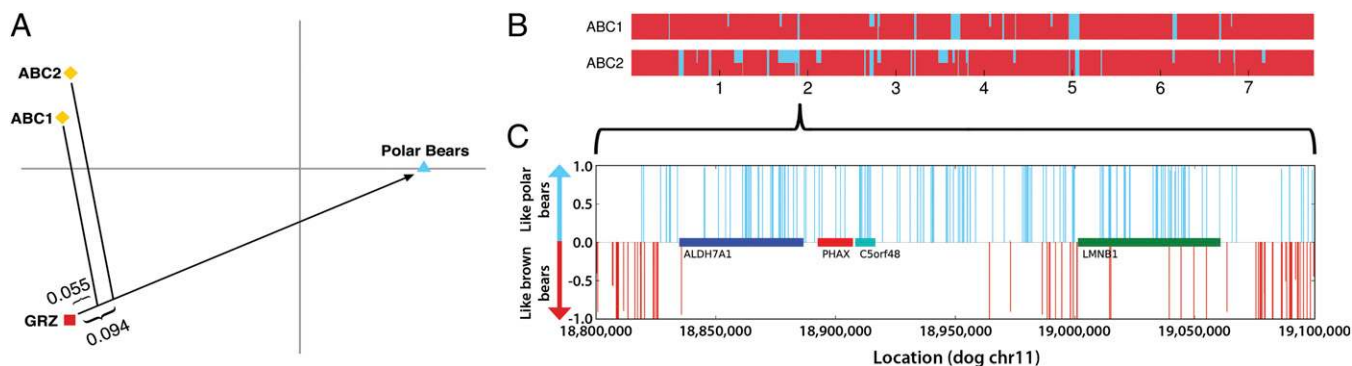
Fig. 4. Potential admixture between polar and ABC brown bears. (*A*) PCA plot showing the two "projections" that give the estimations of admixed fractions. The reason why the projections do not appear to form right angles is discussed in *SI Appendix*. (*B*) An admixture map corresponding to dog chromosome 11, for two ABC brown bears. The scale is in units of 10 million bases. Red areas denote chromosomal regions shared with non-ABC brown bears, and blue indicates where one or both chromosomes are shared with PBs (*SI Appendix*, Fig. S12, shows a complete map). (*C*) A magnification of *B* that includes a 250-kb interval in which both chromosomes in ABC brown bears are very similar to those of our sequenced PBs. The region contains four genes, including the orthologue of *ALDH7A1*, which may be related to salt tolerance (*SI Appendix*). The vertical axis is *P–Q*, where *P* is the probability of generating the genotype observed in ABC2 given the genotypes observed in PBs and *Q* is the probability assuming alleles in the non-ABC brown bear (GRZ). Thus, an SNP with a positive value provides evidence that ABC2 is PB-like in this region, whereas a negative value suggests that ABC2 is genetically like the non-ABC brown bear.

populations sampled represent only part of the global genetic diversity suggested by microsatellite data (24).

Despite the relatively low genetic diversity of PBs, our SNP dataset representing broader geographical sampling indicates the presence of genetic structuring among several extant PB populations (*SI Appendix*). F_{ST} ranged between 0.004 and 0.105 ($P < 0.005$) among all pairs of populations, except between PBs from the Southern Beaufort and Chukchi Seas ($F_{ST} = 0.004$, $P = 0.2$; *SI Appendix*, Table S7), which also group together in PCA and Bayesian clustering analyses (Fig. 5 *B* and *D*) comparably to patterns identified by using 16 microsatellite loci (24). However, PBs from the Barents Sea (Svalbard) and from southern Hudson Bay, Canada, are both significantly different from the Alaskan group (Fig. 5*D*). These results demonstrate that our nuclear SNP set, unlike sparse nuclear intron data that did not show similar structuring (9), could be useful for global population genetic analyses of PBs.

Bear Demographic History and Paleoclimatic Evidence. Deeply sequenced genomes such as those we report here allow for relatively reliable identification of heterozygous positions and the use of recently developed genomic coalescent analyses designed to estimate past population structure. We applied a method based on a pairwise sequentially Markovian coalescent (PSMC) model (25) to estimate historical fluctuations in N_e based on the distribution of the time since the most recent common ancestor for alleles within a diploid whole-genome sequence. We applied this method to our five deep-coverage bear genomes: the PB, two ABC brown bears, the non-ABC brown bear, and the black bear, in addition to a second, lower-coverage PB (*SI Appendix*, Figs. S14 and S15).

The brown bears exhibit roughly similar trends in their estimated N_e history over the past ~5 million years (Fig. 6*A*), reflecting the Pleistocene and Pliocene geologic epochs. In particular, it is noteworthy that an extended decline in N_e begins by the end of
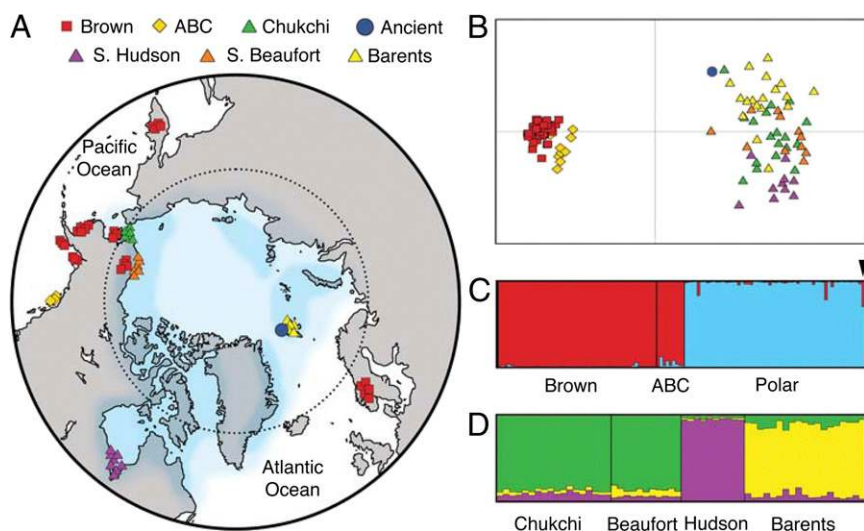

Fig. 5. Population-level analyses of 118 brown bears and PBs. (*A*) Approximate geographical locations of samples. Blue shading represents the current known range of PBs, with the darker shading indicating higher density. (*B*) PCA plot based on genotypes from ~100 SNPs. Colors and symbols for samples are as indicated in the legend for *A*, with triangles representing modern PBs. (*C*) Structure analysis of brown bears, ABC brown bears, and PBs, with the number of genetic populations set to 2, indicating low levels of admixture present in the genomes of some ABC bears and PBs (black arrowhead indicates position of the ancient PB). (*D*) Structure analysis with the number of genetic populations set to 3 for 58 PBs from across their range.

the Early Pleistocene (∼1 Mya), when the dominant wavelength of climate cycles switched from 41 ky cycles to 100 ky cycles and global glaciations and climate became much more severe (26). Conversely, black bear N_e, which began declining earlier than that of the brown bear, stabilizes and recovers during this time, suggesting less impact for Middle Pleistocene cooling and drying on this species, perhaps because of a lower-latitude paleogeographic range. Brown bear N_e recovers later, peaking at approximately 140 kya (Fig. 6B), which roughly correlates with the increased temperatures of the last interglacial (Eemian; 130–114 kya). A sharp decline in N_e follows as Earth experienced overall cooling during the last glaciation.

Historical trends in PB N_e show a very different pattern during the Pleistocene. First, our results indicate that the Barents Sea PB population coalesced at least 1 Mya (Fig. 6A). Although this timing contrasts with the much more recent estimate of PB origin based on mitochondrial genomes (160–150 kya; *SI Appendix*, Fig. S5), it does not contradict the results derived from the extended hidden Markov model described earlier. With PSMC, the marked increase in N_e between 800 and 600 kya (Fig. 6B), possibly facilitated by Middle Pleistocene cooling, is approximately bounded by Marine Isotope Stage 11 (420–360 kya), the longest and possibly warmest interglacial interval of the past 500,000 y and a potential analogue for the current and future climate (27). Although PB N_e remains low thereafter, a small recovery roughly coincident with the ABC bear–PB maternal split (*SI Appendix*, Fig. S5) could be associated with post-Eemian cooling, although this could also indicate an increase in population structure (25). The very recent, slight increase in N_e during the Holocene might reflect cooling during the Last Glacial Maximum, although genomic signatures of such recent events are known to have less power (25). Overall, this analysis strongly suggests that although PB N_e might have been considerably larger in the past, it appears to have experienced a prolonged and drastic decline for the past 500,000 y, being significantly smaller than brown bear N_e, and perhaps explaining the observed lower genetic diversity in PBs compared with brown bears. Taken together, our results strongly indicate that key climatic events have played a significant formative role in bear effective population size.

Adaptation of PBs to Arctic Environment: Potential Signals of Positive Selection. Deeply sequenced genomes also allow us to begin to investigate the unique characteristics of the PB and how this may reflect its excellent adaptation to the extreme Arctic environment. The PB exhibits extensive adaptations for life in the Arctic. These include mechanisms that allow them to maintain homeostasis under low temperatures, and extend from evident morphological features to more subtle physiological traits (28), such as a thick layer of body fat, an almost complete coverage by pigment-free fur that provides camouflage for hunting on the ice, and black skin that can absorb heat from the sun (29). Furthermore, although their characteristic pattern of energy expenditure and fat accumulation is shared with brown and black bears, physiological states in PBs are quite dynamic, and only pregnant females go through an extended winter denning (30).

We used multiple approaches to look for signatures of genetic underpinnings related to phenotypic differences between brown bears and PBs. First, we identified regions of the genome that may potentially harbor genes and/or other functional elements under positive selection in one or both of the lineages (31). We used the F_{ST} value at each SNP to measure the difference in allele frequencies between PBs and brown bears, and investigated genomic intervals (relative to the dog assembly) at which the genomes of the two species are more different than could be explained by chance ($P \leq 0.01$), as indicated by using a randomization approach (*SI Appendix*). We identified 1,374 such intervals of an average length of 20,943 bp. Genes (with annotations in the dog genome) in the highest scoring intervals (*SI Appendix*, Table S8), include *DAG1* (dystroglycan), which is a central component of the skeletal muscle dystrophin-associated glycoprotein complex, and is a candidate gene that in mutant form is involved with muscular dystrophy (32). Hibernating black bears lose significant muscle strength, although they retain considerably more muscle strength than humans and rodents do during the same period of inactivity (33). It is possible that *DAG1* provides PBs with a mechanism to reduce muscle atrophy, perhaps in combination with urea recycling and protein conservation (34).

Another high-scoring interval contains the *BTN1A1* gene (Fig. 7), which is highly expressed in the secretory epithelium of the mammary gland during lactation, and is essential for the regu-
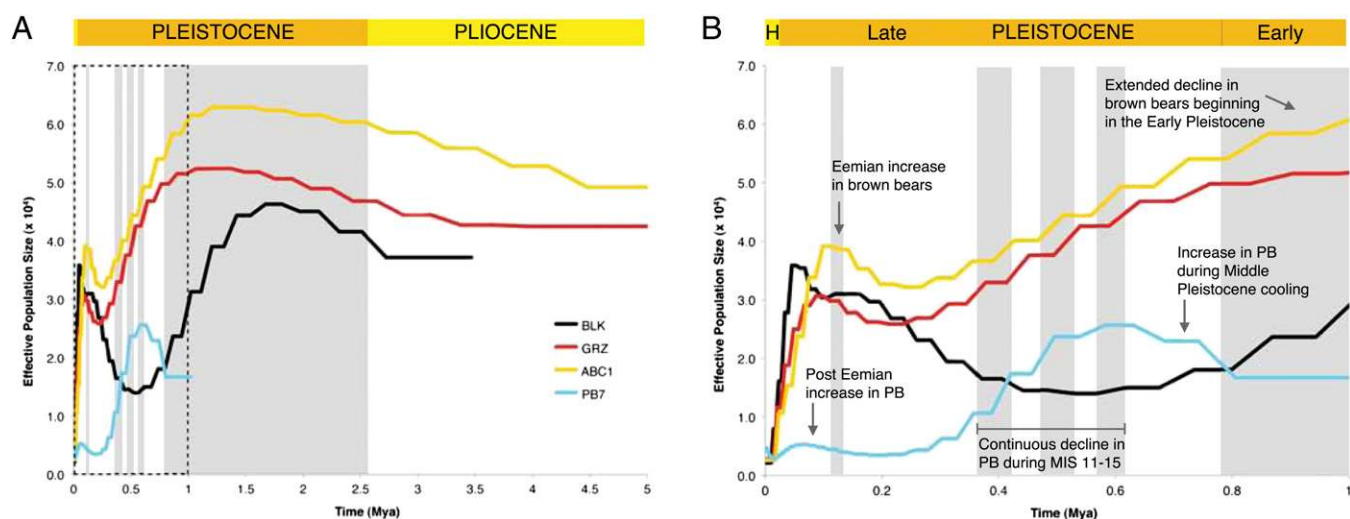

Fig. 6. Bear demographic history. (A) PSMC (25) estimates of bear N_e history inferred from four bear genomes, shown in a time span of 5 million years (only one of the ABC brown bears is shown here; *SI Appendix*, Fig. S14, shows results from all genomes analyzed as well as bootstrap resampling results). The dashed box indicates the time span shown in more detail in B. The large gray-shaded box illustrates the Early Pleistocene, whereas the smaller gray areas refer to key geological events shown in more detail in B and discussed in the text. (B) The larger gray-shaded area to (Right) refers to the Early Pleistocene, and the other gray areas (from right to left) refer to the interglacial Marine Isotope Stages (MIS) 15, 13, and 11, and the Eemian, respectively. The arrows point to major events in bear population history discussed in the text. H, Holocene epoch.

lated secretion of milk lipid droplets (35). Previous experiments have shown that certain polymorphisms in *BTN1A1* significantly increase the fat content of milk in cattle and goats, whereas its deletion has been found to greatly impact the survival and weaning weight of mice (35, 36). We hypothesize that mutations (possibly noncoding) leading to greater expression of this gene or activity of its protein product might result in a higher fat content in PB milk. This could be an adaptation in PBs to increase uptake of lipids by cubs in comparison with brown bears.

Amino Acid Polymorphisms. We also conducted a search for amino acid variants in PB genes potentially associated with phenotypic adaptations (*SI Appendix*, Table S9 and Fig. S16) and classified as "damaging" by computational predictions (*SI Appendix*, Fig. S17). Although the exact function of these genes in PBs clearly needs to be investigated further, we identified substitutions in putative orthologues possibly associated with fatty acid metabolism, hibernation, and pigmentation (further detail is provided in *SI Appendix*). For example, we found a variant of an *FTO* orthologue that appears to be fixed in PBs. Mutations in this gene have previously been correlated with severe obesity in humans, and postnatal growth retardation, lean body mass, and hyperphagia in mice (37). We also found an apparently fixed and damaging variant of a *PLTP* orthologue, the protein product of which is involved in the synthesis and catabolism of high-density lipo-protein (38) and may affect the availability of triglycerides and weight gain during hyperphagia in PBs. In addition, we found fixed variants in two putative orthologues involved in long-chain fatty acid synthesis and catabolism: *CPT1B* and *ACSL6* (39, 40). The fixed variants in these genes may be related to the fatty acid profile of PBs.

Furthermore, we looked for amino acid polymorphisms in genes previously associated with skin and hair pigmentation (*SI Appendix*, Table S13). PB fur is made up of two layers: a dense underfur and long guard hairs that are pigment-free with a hollow core. *EDNRB* was the only gene associated with pigmentation that presented a fixed variant in PBs. This gene is required for the migration of melanoblasts in dermis and enteric neuroblasts, and its mutation results in different color patterns (41). Variants that delimited the black bear from the other ursids studied were found in the genes *MC1R*, *OCA2*, *TYRP1*, and *DCT*, the interaction of which determines coat color patterns in mice, horses, and dogs (42). An SNP at the *MC1R* locus has been found to cause the recessive white-phased (i.e., Kermode) black bear found on the northwest coast of British Columbia (43). Among the various bears studied, we also found an interesting set of nine amino acid polymorphisms in a homologue of *TRPM1*

(*MLSN1*). This gene is responsible for the different coat color patterns found in Appaloosa horses (44). The high number of variants found is unusual, and might be reflective of retained duplicate genes (particularly in brown bears, in which the polymorphisms were prevalent and could be associated with their polytypic pelage color). We hypothesize that the combined effects of *EDNRB* and *TRMP1* could be involved in PB skin pigmentation, with the lack of hair pigmentation possibly caused by disrupted migration of melanoblasts to hair shafts because of their hollow anatomy.

Conclusions

Hybridization is a widespread evolutionary phenomenon that can play a role in speciation, in particular among closely related species. Today we know that evolutionary progression can continue even while species undergo "genomic invasions" from other species (45). In plants, this phenomenon is well known and often associated with whole-genome duplication (i.e., allopolyploid speciation). It is quite possible that interbreeding and introgression have been important factors during the evolution of many mammal species as well, likely associated with range contractions and expansions during times of climatic fluctuation. For example, human evolution may have been closely associated with climate change, during which survival in refugia and differential ecological adaptation may have shaped the divergence of early human species (46). Likewise, interstadial migration and expansion may have brought anatomically modern humans out of Africa and in contact with contracted-range Neanderthals and led to interbreeding between the two hominin lineages (46).

Because of an altered Arctic environment, which has experienced the strongest effects of global climate change in recent decades (47), the PB and its uncertain long-term status has become a focal point for discussions concerning the impact of global climate change on biodiversity (48, 49). Although our results clearly demonstrate that American black bears, brown bears, and PBs have had largely independent evolutionary histories over millions of years, leaving imprints in the PB nuclear genome likely associated with ecological adaptation to the Arctic environment, we also show that gene flow between bear species, possibly as a result of shifts in distribution of formerly isolated populations, has occurred during their evolutionary history. As this gene flow has apparently left inconsistent imprints in different sections of the bear genomes, as discerned by coalescence hidden Markov analysis and admixture maps, divergence time might vary substantially across the genome (50), which may in part explain the different divergence time estimates reported for PB speciation events (e.g., ref. 9). As gene flow has been sufficiently prominent, but unexpectedly so, between the American black bear and brown
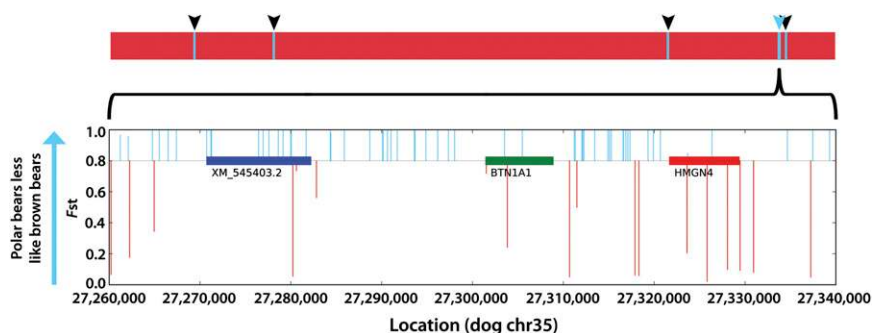

Fig. 7. Potential region of selective sweep in PBs. *Top*: Position of five short intervals of high F_{ST} between PBs and brown bears that map onto dog chromosome 35. Below that is a magnified view of the highest-scoring of the five intervals (and 59th highest scoring genome-wide). This region contains several genes, including a homologue of *BTN1A1*, which may be related to lipid uptake by cubs (as detailed in the text). For 76 SNPs, we plot the F_{ST}, with distances greater than 0.8 shown in blue (i.e., less like brown bears) and less than 0.8 in red (i.e., more like brown bears). Intervals were scored by subtracting 0.8 from F_{ST} and summing over all SNPs in the interval. Genome-wide, only 17.1% of SNPs have positive score, and the high density of SNPs with large F_{ST} in this region is statistically significant (*SI Appendix*).

bear so as to leave a clear signature in their genomes today, ancient admixture in response to climate change may prove to have been a more general phenomenon in the ursid lineage as a whole. However, the current trend toward an increasingly warmer climate in the Arctic has caused sea ice to retreat dramatically and possibly at a pace never experienced before (51). If this trend continues, it is possible that future PBs throughout most of their range may be forced to spend increasingly more time on land, perhaps even during the breeding season, and therefore come into contact with brown bears more frequently. Recently, wild hybrids and even second-generation offspring have been documented in the Northern Beaufort Sea of Arctic Canada, where the ranges of brown bears and PBs appear to overlap, perhaps as a recent response to climatic changes. Similar interbreeding among other Arctic species could have large impacts on polar biodiversity in general (52). Perhaps even more dramatically, we show that a prolonged decline in PB N_e has occurred during the last approximately 500,000 y, possibly followed by recent expansions from small founder populations that may have survived in refugia during interglacials. One such "surviving population" may be represented by our 130,000- to 110,000-y-old PB specimen, suggesting that Svalbard may have constituted an important refugium during past interglacials, acting as a source population for range expansions during subsequent cooler periods. If modern PB populations result from Holocene range expansions from a few small, contracted populations in Middle-Late Pleistocene refugia, this may explain the observed low genetic diversity in PBs today, and possibly leave modern PB populations even more vulnerable to future climatic and other environmental disturbances.

Many questions concerning evolution of the PB and its close relatives still remain, and future paleogenomic and expanded population genomic approaches will undoubtedly shed further light on some of these questions. Until the present, only limited genetic resources have been available to investigate questions of divergence and adaptation of bears, but here we have developed and analyzed extensive genomic and transcriptomic sequence data from three different species. Thus, the resources and results presented here are important for gaining a better understanding of past responses during bear evolutionary history, which may inform predictions of current and future responses. In addition to making our assembly available, we have placed our SNP calls on the Galaxy web server (usegalaxy.org) and provided a suite of tools to analyze them, which can be run through a straightforward interface that requires only an internet browser. For example, the PCA, admixture, and selective-sweep analyses were performed with Galaxy commands, and we provide Galaxy workflows such that all command parameters and options used in our analysis are made explicit.

Materials and Methods

Details of sampling and experimental procedures are provided in *SI Appendix*.

Genome Sequencing. For genomic DNA sequencing, blood and tissue samples were collected from one American black bear (*U. americanus*), one brown bear (*U. arctos*) from the Kenai Peninsula, two brown bears from Alaska's Alexander Archipelago (i.e., ABC brown bears), five PBs (*U. maritimus*) from Alaska (Chukchi Sea and Southern Beaufort Sea populations), and 18 PBs from Svalbard (Barents Sea population). Except for one ABC brown bear blood sample (ABC2), which was collected from a rescued cub in a rehabilitation center in Sitka, AK, all samples were collected from wild populations by using standard procedures and with proper permits. All samples were sequenced to generate paired-end reads of average length 101 bp by using the Illumina HiSeq 2000 sequencing platform. The sample PB7 was sequenced to a greater depth of coverage by using multiple paired-end libraries with span sizes of 160 bp, 180 bp, and 300 bp, in addition to a mate-pair library of 3 kb (*SI Appendix*, Table S1).

Genome Assemblies and Identification of SNPs. Illumina short reads and the paired-end reads from the short-insert libraries were assembled into contigs

for PB7 using SOAPdenovo (12). All available sequence data were then aligned to these contigs by using the mate-pair information in the order of estimated insert size (160 bp to 3 kbp) to generate the scaffolds. Scaffolding was followed by local reassembly of sequences in the intrascaffold gaps. To evaluate the accuracy of the draft sequence, we aligned all the paired-end reads from PB7 back to the assembly by using Burrows–Wheeler alignment (BWA) (53). Mitochondrial reads were mapped to a reference sequence (GenBank accession no. NC003428) by using BWA with default parameters or, in the case of the black bear, using LASTZ (54). These alignments were input to a reference-based assembler to assemble the mtDNA of the black bear sample, and BWA to assemble the mtDNA for the remaining samples. The identification of nuclear SNPs, as aligned to both our draft assembly and the dog genome (*C. familiaris* assembly version 2.0), and the identification of mitochondrial SNPs are described in *SI Appendix*.

Transcriptome Sequencing. Transcriptomes of two bear individuals, a PB and a brown bear, were sequenced on a Personal Genome Machine sequencer, generating a total of 9 million reads, with an estimated average length of 190 bases. Using Genome Sequencer Reference Mapper software (Newbler version 2.6), masked reads from each bear were mapped against our PB assembly as well as 22,291 dog (*C. familiaris*) cDNA sequences (mRNA; *SI Appendix*, Table S2 and Fig. S3).

SNP Genotyping. From the SNPs discovered by genome sequencing of 23 modern PBs, a set of 100 high-quality, autosomal SNPs were identified based on the criteria that loci would be variable among the 23 PBs and that both alleles in each locus were unambiguous in the ancient PB (primers are listed in *SI Appendix*, Table S6). From a total of 118 modern polar and brown bears (Fig. 5 and *SI Appendix*, Table S5), we conducted multiplex PCR amplification of the SNP loci. Barcoded Illumina TruSeq libraries were prepared for each multiplex PCR product. Twenty-four libraries were pooled in equimolar amounts and sequenced in a single run on the Illumina MiSeq platform.

Analytical Procedures. Phylogenetic analyses of nuclear genome SNP data, calculations of estimates of admixture, as well as search for genomic intervals possibly affected by selective sweeps, were conducted by using tools available in Galaxy (usegalaxy.org). Bootstrap analysis of the nuclear SNP data was not successful as a result of computational limits imposed by the required number of massive pseudoreplicate data sets. Mitochondrial genome phylogenetic analyses were conducted by using maximum likelihood and Bayesian inference, and the estimates for dates of molecular divergence were calculated by using the software BEAST (55). Positive selection analyses were carried out using the program TreeSAAP (56). Analyses of genetic diversity and differentiation among samples amplified for selected SNP loci were performed in GenALEx 6.41 (57). The Bayesian clustering program Structure version 2.3.3 (58) was used to investigate the presence of admixture between brown bears and PBs, and of population structuring inside PBs. To analyze bear demographic history, we applied a coalescence hidden Markov model to four of our deep-coverage bear genomes: the PB (PB7), one ABC brown bear (ABC1), the non-ABC brown bear (i.e., GRZ), and the black bear. To obtain a detailed history of ancestral population sizes, we implemented the PSMC (25) to analyze the aforementioned four genomes in addition to the second ABC brown bear (ABC2) and a second, lower-coverage PB (PB3).

Amino Acid Polymorphisms. Among our discovered SNPs, we searched for SNPs that could be traced to putative orthologue genes known from the dog genome (canFam2) and that resulted in SAPs. The functional effect of each SAP was predicted with PolyPhen 2 (59), which provides a classification of "possibly damaging," "probably damaging," or "benign." By using those SAPs classified as damaging (i.e., possibly damaging and probably damaging), we ranked a set of KEGG pathways based on three different metrics: percentage of genes affected, change in the number of paths, and change in the length of paths. We defined high-quality SAPs as those being called with at least two sequences, and as many as 60 sequences, having a quality value higher than 48. To determine the possible role of these mutations in the evolution of PBs, we traced them into metabolic pathways and analyzed their possible role in specific adaptations of interest.

1. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722.
2. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.
3. vonHoldt BM, et al. (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res* 21:1294–1305.
4. Lindqvist C, et al. (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proc Natl Acad Sci USA* 107:5053–5057.
5. Stirling I (2011) *Polar Bears: A Natural History of a Threatened Species* (Fitzhenry and Whiteside, Brighton, MA).
6. 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.
7. Gray AP (1972) *Mammalian Hybrids. A Check-List With Bibliography* (Commonwealth Agricultural Bureaux, Slough, UK), 2nd Ed.
8. Edwards CJ, et al. (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol* 21:1251–1258.
9. Hailer F, et al. (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* 336:344–347.
10. Miller W, et al. (2011) Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proc Natl Acad Sci USA* 108:12348–12353.
11. Perry GH, et al. (2012) A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol Evol* 4:126–135.
12. Li R, et al. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20:265–272.
13. Li R, et al. (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463:311–317.
14. Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13:729–744.
15. Mailund T, Dutheil JY, Hobolth A, Lunter G, Schierup MH (2011) Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS Genet* 7:e1001319.
16. Krause J, et al. (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220.
17. Polyak L, et al. (2010) History of sea ice in the Arctic. *Quat Sci Rev* 29:1757–1778.
18. Paschou P, et al. (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 3:1672–1686.
19. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* 107:786–791.
20. Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M (2011) Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol* 12:R19.
21. Tang H, et al. (2007) Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am J Hum Genet* 81:626–633.
22. Abi-Rached L, et al. (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89–94.
23. Brocker C, et al. (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J Biol Chem* 285:18452–18463.
24. Paetkau D, et al. (1999) Genetic structure of the world's polar bear populations. *Mol Ecol* 8:1571–1584.
25. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
26. Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* 20:PA103.
27. Howard WR (1997) Palaeoclimatology: A warm future in the past. *Nature* 388:418–419.
28. Øritsland NA (1970) Temperature regulation of the polar bear (*Thalarctos maritimus*). *Comp Biochem Physiol* 37:225.
29. Harington CR (2008) The evolution of Arctic marine mammals. *Ecol Appl* 18(2, Suppl):S23–S40.
30. Bruce DS, et al. (1990) Is the polar bear (Ursus maritimus) a hibernator? Continued studies on opioids and hibernation. *Pharmacol Biochem Behav* 35:705–711.
31. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 12:1805–1814.
32. Ibraghimov-Beskrovnaya O, et al. (1992) Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* 355:696–702.
33. Harlow HJ, Lohuis T, Beck TDI, Iaizzo PA (2001) Muscle strength in overwintering bears. *Nature* 409:997.
34. Barboza PS, Farley SD, Robbins CT (1997) Whole-body urea cycling and protein turnover during hyperphagia and dormancy in growing bears (*Ursus americanus* and *U. arctos*). *Can J Zool* 75:2129–2136.
35. Ogg SL, Weldon AK, Dobbie L, Smith AJH, Mather IH (2004) Expression of butyrophilin (Btn1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proc Natl Acad Sci USA* 101:10084–10089.
36. Qu Y, et al. (2011) Novel SNPs of butyrophilin (BTN1A1) and milk fat globule epidermal growth factor (EGF) 8 (MFG-E8) are associated with milk traits in dairy goat. *Mol Biol Rep* 38:371–377.
37. Fischer J, et al. (2009) Inactivation of the Fto gene protects from obesity. *Nature* 458:894–898.
38. Kaser S, et al. (2004) Effects of weight loss on PLTP activity and HDL particle size. *Int J Obes Relat Metab Disord* 28:1280–1282.
39. Miljkovic I, et al. (2009) Association of the CPT1B gene with skeletal muscle fat infiltration in Afro-Caribbean men. *Obesity (Silver Spring)* 17:1396–1401.
40. Soupene E, Dinh NP, Siliakus M, Kuypers FA (2010) Activity of the acyl-CoA synthetase ACSL6 isoforms: Role of the fatty acid Gate-domains. *BMC Biochem* 11:18.
41. Shin MK, Levorse JM, Ingram RS, Tilghman SM (1999) The temporal requirement for endothelin receptor-B signalling during neural crest development. *Nature* 402:496–501.
42. Cieslak M, Reissmann M, Hofreiter M, Ludwig A (2011) Colours of domestication. *Biol Rev Camb Philos Soc* 86:885–899.
43. Hedrick PW, Ritland K (2012) Population genetics of the white-phased "Spirit" black bear of British Columbia. *Evolution* 66:305–313.
44. Bellone RR, et al. (2008) Differential gene expression of TRPM1, the potential cause of congenital stationary night blindness and coat spotting patterns (LP) in the Appaloosa horse (*Equus caballus*). *Genetics* 179:1861–1870.
45. Mallet J (2005) Hybridization as an invasion of the genome. *Trends Ecol Evol* 20:229–237.
46. Stewart JR, Stringer CB (2012) Human evolution out of Africa: The role of refugia and climate change. *Science* 335:1317–1321.
47. Hansen J, et al. (2006) Global temperature change. *Proc Natl Acad Sci USA* 103:14288–14293.
48. Hunter CM, et al. (2010) Climate change threatens polar bear populations: A stochastic demographic analysis. *Ecology* 91:2883–2897.
49. Wiig Ø, Aars J, Born EW (2008) Effects of climate change on polar bears. *Sci Prog* 91:151–173.
50. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–1108.
51. MacDonald GM (2010) Global warming and the Arctic: A new world beyond the reach of the Grinnellian niche? *J Exp Biol* 213:855–861.
52. Kelly BP, Whiteley A, Tallmon D (2010) The Arctic melting pot. *Nature* 468:891.
53. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
54. Harris RS (2007) Improved pairwise alignment of genomic DNA. PhD dissertation (Pennsylvania State Univ, University Park, PA).
55. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
56. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671–672.
57. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.
58. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
59. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249.