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

Webb Millera,1, Stephan C. Schusterb,1, Andreanna J. Welchc, Aakrosh Ratand, Oscar C. Bedoya-Reinae, Fangqing Zhaoa,b,d, Hie Lim Kimh, Richard C. Burhansg, Daniela I. Drautzf, Nicola E. Wittekindt, Lynn P. Tomshoa, Enrique Ibarra-Lacletteg, Luis Herrera-Estrellah,2, Elizabeth Peacock,1, Sean Farleyg, George K. Sagei, Karyn Rodeh, Martyn Obbard, Rafael Montiel, Lutz Bachmann, Ólafur Ingólfsson, Jon Aars, Thomas Mailund, Øystein Wiig, Sandra L. Talbot, and Charlotte Lindqvists,1,2

aCenter for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802; bSingapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551; cDepartment of Biological Sciences, University at Buffalo, Buffalo, NY 14260; dComputational Genomics Laboratory, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China; eLaboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, 36821, Mexico; fDepartment of Earth Sciences, University of Iceland, 101 Reykjavik, Iceland; gDepartment of Biological Sciences, University at Buffalo, Buffalo, NY 14260; and hDepartment of Biological Sciences, University at Buffalo, Buffalo, NY 14260

Contributed by Luis Herrera-Estrella, June 25, 2012 (sent for review May 29, 2012)

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, enigmatic model system. For example, paradigm-shifting genomic research on arctic brown bears has demonstrated that they are genetically distinct from brown bears approximately 150,000 years ago. This result has significant implications for understanding the evolutionary history of bears and for conservation efforts.

In this study, we sought to understand the evolutionary relationships between polar and brown bears using genomic data. We analyzed whole-genome sequence data from contemporary polar, brown, and American black bear samples, in addition to a 130,000- to 110,000-y old PB specimen. Our analyses support a close relationship between PBs and brown bears, with evidence for ancient admixture between the two species. This finding has significant implications for understanding the evolutionary history of bears and for conservation efforts.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option. Data deposition. The sequence reported in this paper has been deposited in the GenBank database (accession nos. JX196366–JX196392 and SRA054912).


This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1210506109/-/DCSupplemental.
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 (“ABCI” 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 coverage, 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 theorthologous 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. 24). 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. 2F). 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
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. 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).
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 demographies 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. 4A), 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. 4B 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. 4B 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. 4C 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. 5A). 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. 5B), some individuals share as much as 11% of their genetic background with PBs (Fig. 5C), 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. 5C).

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
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, Table S7), which also group together in PCA and Bayesian clustering analyses (Fig. 5B 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. 5D). 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. 6A), reflecting the Pleistocene and Pliocene geologic epochs. In particular, it is noteworthy that an extended decline in $N_e$ begins by the end of...
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 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-
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 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 *TRPM1* 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., alloployploid 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, interstitial 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 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*<sub>ST</sub> 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*<sub>ST</sub>, 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*<sub>ST</sub> 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*<sub>ST</sub> in this region is statistically significant (*SI Appendix*).](https://www.pnas.org/content/ptl/early/2022/10/31/2204861119/F7.large.jpg)
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 refuge 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 climate-driven future changes.

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 each contig and a minimum insert size (160 bp) to generate an optical genome map. This 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 (PGM) 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 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 amount of computer time. LaVern Beier, Rod Flynn, and Neil Barten (Alaska Department of Fish and Game) for support during different stages of the project, and the Centre for Scientific Computing Aarhus (Aarhus University) for free computer time. LaVern Beier, Rod Flynn, and Neil Barten (Alaska Department of Fish and Game) and Bill Leacock (US Fish and Wildlife Service)
facilitated field sampling. This study received financial support from Penn State University; the College of Arts and Sciences, University at Buffalo; US Geological Survey’s Changing Arctic Ecosystem Initiative; US Fish and Wildlife Service; Ontario Ministry of Natural Resources, Canada; and Alaska Department of Fish and Game. Funding of the Galaxy tools to analyze medium-coverage sequence data from multiple individuals is supported by National Institutes of Health Grant UL1 RR033184-01 to the Penn State Clinical and Translational Science Institute. This work was supported by National Science Foundation Award DEB 0733029 (to W.M.), Howard Hughes Medical Institute Grant 55005946 (to L.H.-E.), and a grant from the Pennsylvania Department of Health using Tobacco Commonwealth Universal Research Enhancement Funds (W.M.).