Cell

Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears

Shiping Liu,[1,2,20] Eline D. Lorenzen,[3,4,20] Matteo Fumagalli,[3,20] Bo Li,[1,20] Kelley Harris,[5] Zijun Xiong,[1] Long Zhou,[1] Thorfinn Sand Korneliussen,[4] Mehmet Somel,[3,21] Courtney Babbitt,[6,7,22] Greg Wray,[6,7] Jianwen Li,[1] Weiming He,[1,2] Zhuo Wang,[1] Wenjing Fu,[1] Xueyan Xiang,[1,8] Claire C. Morgan,[9] Aoife Doherty,[10] Mary J. O'Connell,[9] James O. McInerney,[10] Erik W. Born,[11] Love Dalén,[12] Rune Dietz,[13] Ludovic Orlando,[4] Christian Sonne,[13] Guojie Zhang,[1,14] Rasmus Nielsen,[1,3,15,16,*] Eske Willerslev,[4,*] and Jun Wang[1,16,17,18,19,*]

[1]BGI-Shenzhen, Shenzhen 518083, China
[2]School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China
[3]Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA
[4]Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark
[5]Department of Mathematics, 970 Evans Hall, University of California, Berkeley, CA 94720, USA
[6]Department of Biology, 124 Science Drive, Duke Box # 90338, Duke University, Durham, NC 27708, USA
[7]Institute for Genome Sciences & Policy, 101 Science Drive, DUMC Box 3382, Duke University, Durham, NC 27708, USA
[8]College of Life Sciences, Sichuan University, Chengdu 610064, China
[9]Bioinformatics and Molecular Evolution Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland
[10]Bioinformatics and Molecular Evolution Unit, Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland
[11]Greenland Institute of Natural Resources, c/o Government of Greenland Representation in Denmark, Strandgade 91, 3. Floor, PO Box 2151, 1016 Copenhagen K, Denmark
[12]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, PO Box 50007, 10405, Stockholm, Sweden
[13]Department of Bioscience, Arctic Research Centre, Aarhus University, Frederiksborgvej 399, PO Box 358, 4000 Roskilde, Denmark
[14]Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
[15]Department of Statistics, 367 Evans Hall, University of California, Berkeley, CA 94720, USA
[16]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen Ø, Denmark
[17]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[18]Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau 999078, China
[19]Department of Medicine, University of Hong Kong, Sassoon Road, Pokfulam, Hong Kong
[20]Co-first authors
[21]Present address: Middle East Technical University, Department of Biological Sciences, 06800, Ankara, Turkey
[22]Present address: Department of Biology, 611 North Pleasant St, University of Massachusetts Amherst, Amherst, MA, 01003, USA
*Correspondence: rasmus_nielsen@berkeley.edu (R.N.), ewillerslev@snm.ku.dk (E.W.), wangj@genomics.org.cn (J.W.)
http://dx.doi.org/10.1016/j.cell.2014.03.054

SUMMARY

Polar bears are uniquely adapted to life in the High Arctic and have undergone drastic physiological changes in response to Arctic climates and a hyper-lipid diet of primarily marine mammal prey. We analyzed 89 complete genomes of polar bear and brown bear using population genomic modeling and show that the species diverged only 479–343 thousand years BP. We find that genes on the polar bear lineage have been under stronger positive selection than in brown bears; nine of the top 16 genes under strong positive selection are associated with cardiomyopathy and vascular disease, implying important reorganization of the cardiovascular system. One of the genes showing the strongest evidence of selection, *APOB*, encodes the primary lipoprotein component of low-density lipoprotein (LDL); functional mutations in *APOB* may explain how polar bears are able to cope with life-long elevated LDL levels that are associated with high risk of heart disease in humans.

INTRODUCTION

The polar bear (*Ursus maritimus*) is uniquely adapted to the extreme conditions of life in the High Arctic and spends most of its life out on the sea ice. In cold Arctic climates, energy is in high demand. Lipids are the predominant energy source and the polar bear has a lipid-rich diet throughout life. Young nurse on milk containing ~27% fat (Hedberg et al., 2011) and adults feed on a marine mammal diet, primarily consisting of seals and their blubber (Thiemann et al., 2008). Polar bears have substantial adipose deposits under the skin and around organs, which can comprise up to 50% of the body weight of an individual, depending on its nutritional state (Atkinson and Ramsay, 1995; Atkinson et al., 1996).

The polar bear is most closely related to the brown bear (*Ursus arctos*), a widely distributed omnivore found in a variety of

Figure 1. Sampling Localities
Polar and brown bear distributions are shown in blue and brown shading, respectively. See also Table S2.

habitats across the Holarctic (Figure 1). The two species differ fundamentally in their ecology, behavior, and morphology, reflecting adaptations to different ecological niches. Despite ample data there is still no consensus regarding when the two species diverged. Inferences based on the fossil record suggest polar bears diverged from brown bears some 800–150 thousand years ago (kya) (Kurtén, 1964). Estimates based on genomic data span an order of magnitude from 5–4 million years ago (Mya) (Miller et al., 2012) to ca. 600 kya (Hailer et al., 2012), depending on assumptions about effective population size and migration in the period when the two populations were drifting apart. Establishing a reliable time frame for when the polar bear emerged as a species is essential for our understanding of what evolutionary processes drove speciation, and how fast novel adaptations to extreme environments can arise in a large mammal.

Here, we apply a population genomic framework to analyze complete nuclear genomes of polar bear and brown bear populations (Figure 1, Tables S1 and S2 available online) to (1) estimate when polar bears and brown bears diverged; (2) infer the joint demographic history of the two species to elucidate what happened after they diverged; and (3) detect genes under positive selection in polar bears to gain insight into polar bear evolution and the genetic background of its unique adaptations to life in the High Arctic.

To address these issues, we deep-sequenced and de novo assembled a polar bear reference genome at a depth of 101X (See "Polar Bear Reference Genome and de novo Assembly" in Extended Experimental Procedures) and resequenced at 3.5X to 22X coverage 79 Greenlandic polar bears and ten brown bears from Fennoscandia; mainland US; and the Admiralty, Baranof, and Chichagof (ABC) Islands off the coast of Alaska (Figure 1) (see "Samples" in Extended Experimental Procedures, Tables S1 and S2, and Figure S1).

RESULTS AND DISCUSSION

Joint Demographic History of Polar Bears and Brown Bears

To infer the joint demographic history of polar bears and brown bears, we used a novel method based on identity by state (IBS) tracts of DNA shared within and between populations (Harris and Nielsen, 2013) and ∂a∂i (diffusion approximation for demographic inference [Gutenkunst et al., 2009]), which infers demographic parameters based on a diffusion approximation to the site frequency spectrum. The two models differed in their individual parameter estimates (Table S3), in part reflecting the fact that the IBS tract method uses both recombination rate and mutation rate, and ∂a∂i uses only the latter. However, despite the inherent uncertainty in the genome-wide mutation rate estimate, which we calibrated using deep fossil divergence dates (Figure S2A), the estimates from the two models are in fact quite similar with regards to divergence time, relative effective population sizes, and direction of gene flow.

We find evidence of smaller long-term effective population sizes in polar bears than in brown bears (Figure 2A). Genetic diversity is a function of effective population size, and the number of private SNPs in polar bears (2.6 million, Figure S1B) is about one third of that in brown bears (7.7 million, Figure S1C). Similarly, patterns of linkage disequilibrium (LD) can be informative about demographic history (Reich et al., 2001) and we find a slower rate of LD decay in polar bears (Figure S3A).

Prior to divergence, we find a 10-fold decline in the global joint ancestral population (Table S3). Polar bears declined in population size after the split from brown bears, although we were unable to confidently estimate the timing of the bottleneck. However, both the IBS tract method and ∂a∂i indicate that the population size decrease in polar bears was either of a greater magnitude or of a longer duration than in brown bears, in agreement with our other indicators of relative population sizes.

The Age of the Polar Bear as a Species

To reliably estimate when polar bears and brown bears diverged, we used the IBS tract method (Harris and Nielsen, 2013) and ∂a∂i (Gutenkunst et al., 2009), which both take past population size changes into account. The approaches indicated that the two species diverged only ca. 479–343 kya (Figure 2A, Table S3). Because genotyping errors appear as singletons and given that in both methods singletons lead to increasing divergence time estimates, we can conclude that the polar bear likely emerged closer to the lower bound of our estimate. Our date greatly decreases the age of polar bear origin and agrees with fossil evidence (Kurtén, 1964; Miller et al., 2012).

We assessed the effect of using our more complex demographic models versus simpler models by analyzing our data using a simple isolation-with-migration model similar to that used by Hailer et al. (2012) and Miller et al. (2012). The procedure yielded an older divergence date in the range of 1.6–0.8 Mya (Table S3). However, we found that our complex model with a more recent divergence time estimate was a better fit

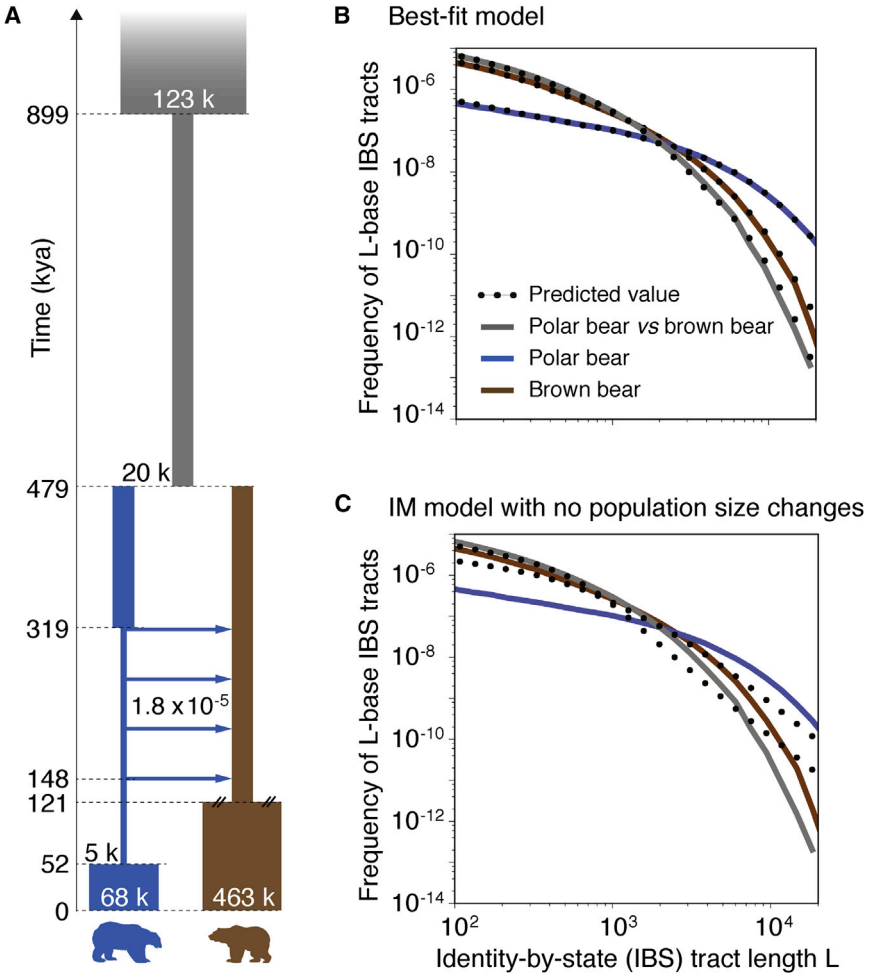

Figure 2. Demographic Inference

(A–C) Joint demographic model for polar bear and North American brown bear populations inferred using the IBS tract method (A). Joint past population is in gray, polar bear in blue and brown bear in brown. Estimated effective population sizes are indicated and the migration rate is in genetic replacements per generation. The recent brown bear population size has been downscaled by a factor of 20, the recent polar bear population size is to scale. (B), (C) Distribution of IBS tract length from our observed data (solid line) and from model prediction (dotted line) inferring gene flow from polar bear into brown bear (B) or using a simple isolation-with-migration (IM) model (C), which does not account for past population size changes. There are only two black dotted curves in (C) because the IM model constrains the within-polar bear and within-brown bear tract lengths to be the same. See also Figure S4B and Table S3.

to our empirical data (Figures 2B and 2C). Discrepancies between our divergence date and previous genomic estimates highlight the impact of accounting for past population size changes on divergence time estimates, suggesting that models that do not account for past population size changes have the potential to overestimate divergence times (e.g., Hailer et al., 2012; Miller et al., 2012).

The timing of polar bear origin coincides with Marine Isotope Stage (MIS) 11. MIS 11 was a warm period, which spanned ca. 424–374 kya. It was the longest interglacial in half a million years (Dickson et al., 2009) and lasted almost 50 kyr (de Vernal and Hillaire-Marcel, 2008). The period was associated with a substantial decrease in Greenland ice-sheet volume; DNA from the basal part of the Dye 3 ice core from southern central Greenland (Willerslev et al., 2007) and abundant spruce pollen from the shore off southwest Greenland (de Vernal and Hillaire-Marcel, 2008) both suggest that boreal coniferous forest developed at least over southern Greenland. Such a prolonged interglacial could have enabled an ancestral brown bear population to colonize northern latitudes that were previously uninhabitable for the species, setting the stage for future allopatric speciation, as subsequent climatic and environmental change caused population isolation (Stewart et al., 2010).

Gene Flow between Polar Bears and Brown Bears after Divergence

Based on morphology and a phylogenetic analysis of their nuclear genomes, polar bears and brown bears are monophyletic sister species (Figure S2B; Pagès et al., 2008). Nevertheless, the mitochondrial genomes of brown bear are paraphyletic and extant polar bear sequences are recovered as a monophyletic sister clade to the brown bear population from Alaska's ABC Islands, within the diversity of brown bear (Figure S2C). The consensus has been that this pattern reflects female-mediated gene flow from brown bears into polar bears ca. 150 kya, with subsequent fixation of the brown bear mitochondrial lineage in polar bears (Lindqvist et al., 2010). However, this was not supported by a recent genomic study, which presented evidence that gene flow historically took place from polar bears into ABC brown bears and not the other way around (Cahill et al., 2013).

Based on the IBS tract method, we find strong evidence of continuous gene flow from polar bears into North American brown bears after the species diverged (Figure 2A, Table S3). We used the IBS tract method to compare likelihoods of two scenarios with parsimonious one-way gene flow, finding that gene flow from polar bears to North American brown bears explained the data better than the reverse scenario (Figures 2B, S4B, Table S3). In the former scenario, we estimate a migration rate of 0.0018% genetic replacement per generation. As a complementary approach, we used ∂a∂i to infer the parameters of a model with asymmetric two-way gene flow between polar bears and North American brown bears. With this approach, we observe nonzero migration in both directions but infer a substantially higher migration rate in the polar-to-brown bear direction (Table S3). These results suggest that the major direction of introgression has historically been from polar bears into North American brown bears, in agreement with Cahill et al. (2013).

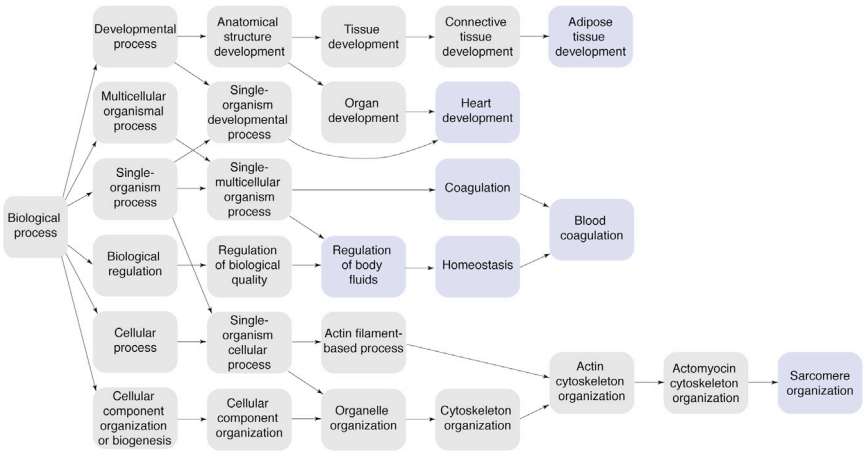

Figure 3. Enrichment Analysis

Gene Ontology enrichment analysis for putative genes under positive selection in the polar bear lineage. We ranked genes based on their homogeneity test score by first considering genes where the ratio between polymorphisms and divergence was lower in the polar bear than in the brown bear samples. We used the web application GOrilla (http://cbl-gorilla.cs.technion.ac.il) to detect biological process terms enriched with top genes in the ranked list. Blue shading indicates biological categories significantly enriched with genes under positive selection in the polar bear lineage, after correction for multiple tests.

We were unable to confidently infer when admixture took place. With the IBS tract method, we estimate gene flow from the timing of the polar bear bottleneck 319 kya to 148 kya (Figure 2A), but the method has limited power to detect migration that occurred very close to the initial divergence time. With ∂a∂i, we infer continuous gene flow until the present. We see no admixture using classical structure analyses (Tang et al., 2005) (Figure S4A), suggesting admixture is not a recent or current phenomenon. However, we note that the number of analyzed brown bear samples is limited, and none of our brown bear samples originate from regions where polar and brown bears are currently sympatric (i.e., where recently admixed individuals are most likely to be found).

To further investigate the question of admixture, we split the genomic data into 100 kbp regions (Table S4) and calculated the length distribution of regions that were introgressed between species. We find that the longest blocks were a maximum of 1.1 Mbp length. If admixture between species had taken place within the last hundreds of generations, we would expect longer tracts of shared DNA (Gravel, 2012; Pool and Nielsen, 2009). Hence the limited length of admixture blocks supports the hypothesis that admixture was an old event, and that enough time has passed for recombination to break up the long stretches of introgressed DNA.

In order to determine whether gene flow happened before or after the divergence of brown bear populations from different parts of the Holarctic, we used the *D* statistic (Durand et al., 2011; Green et al., 2010). We find evidence of gene flow between polar bears and all brown bear populations, suggesting that some gene flow took place prior to the divergence of the brown bear populations (Table S5). The strongest evidence is found with brown bears from the ABC Islands and the weakest with brown bear populations from North America and Fennoscandia, suggesting gene flow continued between polar bears and ABC brown bears also after the brown bear populations diverged. In addition, we find evidence of recent migration between brown bear populations. Our data included six brown bear samples from the ABC Islands (Figure 1, Table S2). One of these individuals (ABC06) was from Admiralty, the island located closest to the US mainland. The mitochondrial genome of ABC06 clustered with the other five ABC individuals from Baranof and Chichagof Islands, as a sister group to the polar bear (Figure S2C). We observe substantial levels of gene flow between polar bears and the Baranof and Chichagof individuals using the *D* statistic, as expected (Table S5). However, we find no signal of polar bear admixture in ABC06, which clustered with the Glacier National Park individual from Montana in the principal component analysis (Figure S3D). We do not find evidence of polar bear admixture in the Glacier NP individual either, the mitochondrial genome of which clustered with European brown bears (Figure S2C). The patterns in ABC06 reflect migration between the Admiralty Island and mainland US, in agreement with previous inferences based on nuclear microsatellites (Paetkau et al., 1999).

Genes under Positive Selection in Polar Bears

Despite being closely related species, the polar bear differs from the brown bear in ecology, behavior, and morphology and is a prime example of what happens when a species evolves through selection and adaptation to a novel environment/lifestyle. Our remarkably recent divergence time estimate of only ca. 479–343 kya, coupled with stable isotope analysis of an ancient jawbone from Svalbard that indicates that polar bears were adapted to a marine diet and life in the High Arctic by at least 110 kya (Lindqvist et al., 2010), provides us with an unprecedented timeframe for rapid evolution. Assuming an average generation time of 11.35 years (Cronin et al., 2009; De Barba et al., 2010), the distinct adaptations of polar bears may have evolved in less than 20,500 generations; this is truly exceptional for a large mammal. In this limited amount of time, polar bears became uniquely adapted to the extremities of life out on the Arctic sea ice, enabling them to inhabit some of the world's harshest climates and most inhospitable conditions.

The observation of rapid evolutionary changes in the polar bear genome raises the question of what signatures of selection are to be found in the extant genomes. We find that the enrichment categories of the top candidate genes under positive selection (i.e., the genes that showed greater values for all of our three test statistics—homogeneity test, Hudson-Aguade-Kreitman test, and F_{st} estimation—in the polar bear, Table S6) are associated with sarcomere organization, blood coagulation, heart development, and adipose tissue development (Figure 3, "Positive Selection" in Extended Experimental Procedures). In brown

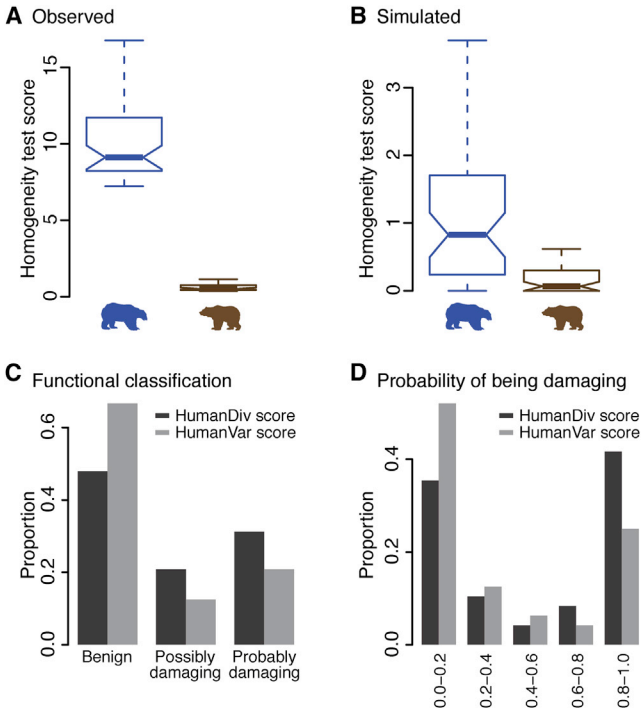

Figure 4. Positive Selection Analysis

(A and B) Distribution of the homogeneity test scores for the top-50 genes in polar bear and brown bear. We compared the observed distribution versus the expected distribution under neutrality, using the demographic model presented in Table S3; full range of values is represented, excluding outliers.

(C and D) Predicted functional impact of polar bear-specific protein substitutions. We reported the functional classification and probability of being damaging for polar bear-specific missense mutations located in the top 20 genes under positive selection, according to the two metrics HumanDiv and HumanVar computed by PolyPhen-2. See also Table S7.

bears, we do not find significant enrichment categories for the top 20 candidate genes under selection.

In general, we also find evidence of more positive selection acting on the polar bear lineage than on the brown bear lineage. Polar bears had markedly higher values in the distribution of our primary test score, the homogeneity test, compared to brown bears (Figure 4A). These patterns may only in part be explained by variation in sample size and effective population size (Figure 4B). Overall, our data support a scenario of polar bears evolving rapidly and being under strong positive selection following the divergence from brown bears.

Genes Associated with Adipose Tissue Development and Fatty Acid Metabolism

The enrichment of genes associated with adipose tissue development (Figure 3) reflects the crucial role lipids play in the ecology and life history of polar bears; the species is adapted to cope with a diet rich in fatty acids (e.g., Smith, 1980; Stirling and Archibald, 1977) and has substantial adipose deposits (Atkinson and Ramsay, 1995; Atkinson et al., 1996). Cholesterol levels in blood plasma of polar bears are extreme (e.g., Ormbostad, 2012); in humans, elevated cholesterol levels are a major risk factor for the development of cardiovascular disease (Cannon et al., 2010). It remains an enigma how polar bears are able to deal with such lifelong elevated levels of cholesterol.

The enriched categories may highlight the genes that have been important in polar bear adaptation to a lipid-rich diet. A top gene in our selection scan was *APOB* (Table 1), which produces apolipoprotein B (apoB), the primary lipid-binding protein of chylomicrons and low-density lipoproteins (LDL) (Whitfield et al., 2004). LDL cholesterol is a major risk factor for heart disease and is also known as "bad cholesterol." ApoB enables the transport of fat molecules in blood plasma and lymph and acts as a ligand for LDL receptors, facilitating the movement of molecules such as cholesterol into cells (Benn, 2009). The extreme signal of *APOB* selection implies an important role for this protein in the physiological adaptations of the polar bear. The gene is ranked second using our homogeneity test score, has an F_{st} ranking in the top 3% of the empirical distribution, and the ratio of fixed-to-polymorphic mutations in the polar bear lineage is 1:2, compared with 1:162 in the brown bear lineage—an 80-fold reduction.

Due to a lack of appropriate functional studies of polar bears, we were unable to directly identify causal variants. Nevertheless, we assessed the impact of polar bear–specific substitutions on human proteins for top-20 genes under positive selection by computational predictions: a large proportion (ca. 50%) of mutations were predicted to be functionally damaging (Figures 4C and 4D, Table S7). Substantial work has been done on the functional significance of *APOB* mutations in other mammals. In humans and mice, genetic *APOB* variants associated with increased levels of apoB are also associated with unusually high plasma concentrations of cholesterol and LDL, which in turn contribute to hypercholesterolemia and heart disease in humans (Benn, 2009; Hegele, 2009). In contrast with brown bear, which has no fixed *APOB* mutations compared to the giant panda genome, we find nine fixed missense mutations in the polar bear (Figure 5A). Five of the nine cluster within the N-terminal βα1 domain of the *APOB* gene, although the region comprises only 22% of the protein (binomial test p value = 0.029). This domain encodes the surface region and contains the majority of functional domains for lipid transport. We suggest that the shift to a diet consisting predominantly of fatty acids in polar bears induced adaptive changes in *APOB*, which enabled the species to cope with high fatty acid intake by contributing to the effective clearance of cholesterol from the blood.

Genes Associated with Cardiovascular Function

We find that nine out of the top 16 genes showing the strongest evidence of positive selection in polar bears are directly related to heart function in humans (Table 1). Mutations in all nine genes, including *APOB*, are associated with either atherosclerosis or cardiomyopathy in humans and other mammalian model organisms. *TTN* encodes Titin, an abundant protein of striated muscle, which includes cardiac muscle tissue; mutations in *TTN* are associated with familial dilated cardiomyopathy (Herman et al., 2012). *XIRP1*, also known as Cardiomyopathy-associated protein 1, is associated with the development of cardiac muscle cells (van der Ven et al., 2006). *ALPK3* encodes a kinase and

Table 1. Top-20 Genes under Positive Selection in Polar Bears

Gene	Length (bp)	Homogeneity Test Score	HKA Test p Value	F_{st}
TTN	99,416	16.76	2.47E-03	0.93
APOB	13,264	13.16	1.54E-05	0.89
OR5D13	871	8.08	4.93E-10	0.82
FCGBP	5,216	6.36	8.20E-04	0.85
XIRP1	3,848	6.05	1.50E-05	0.88
COL5A3	4,402	5.89	1.38E-02	0.81
LYST	11,172	5.58	1.08E-03	0.89
ALPK3	3,007	5.34	1.32E-05	0.91
VCL	3,106	4.87	7.51E-03	0.82
SH3PXD2B	2,458	4.34	2.81E-05	0.88
EHD3	1,230	4.28	1.62E-03	0.90
IPO4	1,260	4.18	1.81E-04	0.86
ARID5B	3,109	4.14	1.31E-02	0.84
ABCC6	3,346	4.02	9.26E-03	0.85
LAMC3	1,885	3.93	9.25E-03	0.85
CUL7	2,701	3.86	4.43E-03	0.83
C15orf55	3,001	3.86	1.71E-02	0.89
POLR1A	4,499	3.85	1.89E-02	0.82
AIM1	4,344	3.8	2.03E-02	0.92
OR8B8	965	3.71	7.37E-06	0.87

We used several statistics to analyze the coding regions of 19,822 genes annotated across the polar and brown bear population samples, using the giant panda (*Ailuropoda melanoleuca*) genome sequence as an outgroup. Genes were ordered based on their homogeneity test score; we only considered genes with (i) a significant nominal p value for the HKA test for selection in the polar bear lineage only, and (ii) a ranked F_{st} over the 90th percentile. See also Tables S6 and S7.

plays a role in cardiomyocyte differentiation; knockout genes in mice show both hypertrophic and dilated forms of cardiomyopathy (Van Sligtenhorst et al., 2012). *VCL* encodes vinculin, a cytoskeletal protein associated with cell-cell and cell-matrix junctions, which is also the major talin-binding protein in platelets. Defects in *VCL* are associated with dilated cardiomyopathy in humans (Olson et al., 2002). *EHD3* encodes a class of cardiac trafficking proteins and plays a role in endocytic transport (Galperin et al., 2002). Regulation of *EHD3* plays a role in a molecular pathway related to heart failure (Gudmundsson et al., 2012). *ARID5B* is involved in pathogenesis of atherosclerosis and adipogenesis (Wang et al., 2012). *ABCC6* is associated with transport of molecules across membranes and is associated with premature atherosclerosis (Trip et al., 2002), and *CUL7* plays a role in vascular morphogenesis (Arai et al., 2003).

Based on this evidence, we argue that potentially important reorganization of the cardiovascular system has taken place in polar bears since their divergence from brown bears, which may be related to polar bear ecology. Chronically elevated serum cholesterol, particularly LDL, contribute to the degenerative accumulation of plaques in the arteries, which can lead to progressive narrowing or blocking of blood vessels (Klop et al., 2013). Alternatively, smaller plaques may rupture and cause a clot to form and obstruct blood flow, leading to reduced blood supply of the heart muscle and eventually heart attack. Changes in behavior, including long distance swimming (Pagano et al., 2012), may also have imposed selection on other aspects of the cardiovascular system, including cardiac morphology.

Genes Associated with White Fur

A white phenotype is usually selected against in natural environments, but is common in the Arctic (e.g., beluga whale, arctic hare, and arctic fox), where it likely confers a selective advantage. A key question in the evolution of polar bears is which gene(s) cause the white coat color phenotype. The white fur is one of the most distinctive features of the species and is caused by a lack of pigment in the hair. We find evidence of strong positive selection in two candidate genes associated with pigmentation, *LYST* and *AIM1* (Table 1). *LYST* encodes the lysosomal trafficking regulator Lyst. Melanosomes, where melanin production occurs, are lysosome-related organelles and have been implicated in the progression of disease associated with Lyst mutation in mice (Trantow et al., 2010). The types and positions of mutations identified in *LYST* vary widely, but Lyst mutant phenotypes in cattle, mice, rats, and mink are characterized by hypopigmentation, a melanosome defect characterized by light coat color (Kunieda et al., 1999; Runkel et al., 2006; Gutiérrez-Gil et al., 2007). *LYST* contains seven polar bear-specific missense substitutions, in contrast to only one in brown bear. One of these, a glutamine to histidine change within a conserved WD40-repeat containing domain, is predicted to significantly affect protein function (Figure 5B, Table S7). Three polar bear changes in *LYST* are located in proximity to the N-terminal structural domain and map close to human mutations associated with Chediak-Higashi syndrome, a hair and eyes depigmentation disease (Figure 5C). We predict that all these protein-coding changes, possibly aided by regulatory mutations or interactions with other genes, dramatically suppress melanin production and transport, causing the lack of pigment in polar bear fur. Variation in expression of the other color-associated gene, *AIM1* (absent in melanoma 1), has been associated with tumor suppression in human melanoma (Trent et al., 1990), a malignant tumor of melanocytes that affects melanin pigment production.

CONCLUSIONS

Our study reveals the strength of using a population genomic approach to resolve the evolutionary history of a nonmodel organism in terms of divergence time, demographic history, selection, and adaptation. We find it remarkable that a majority of the top genes under positive selection in polar bears have functions related to the cardiovascular system and most of them to cardiomyopathy, in particular when considering their divergence from brown bears no more than ca. 479–343 kya. Such a drastic genetic response to chronically elevated levels of fat and cholesterol in the diet has not previously been reported. It certainly encourages a move beyond the standard model organisms in our search for the underlying genetic causes of human cardiovascular diseases.

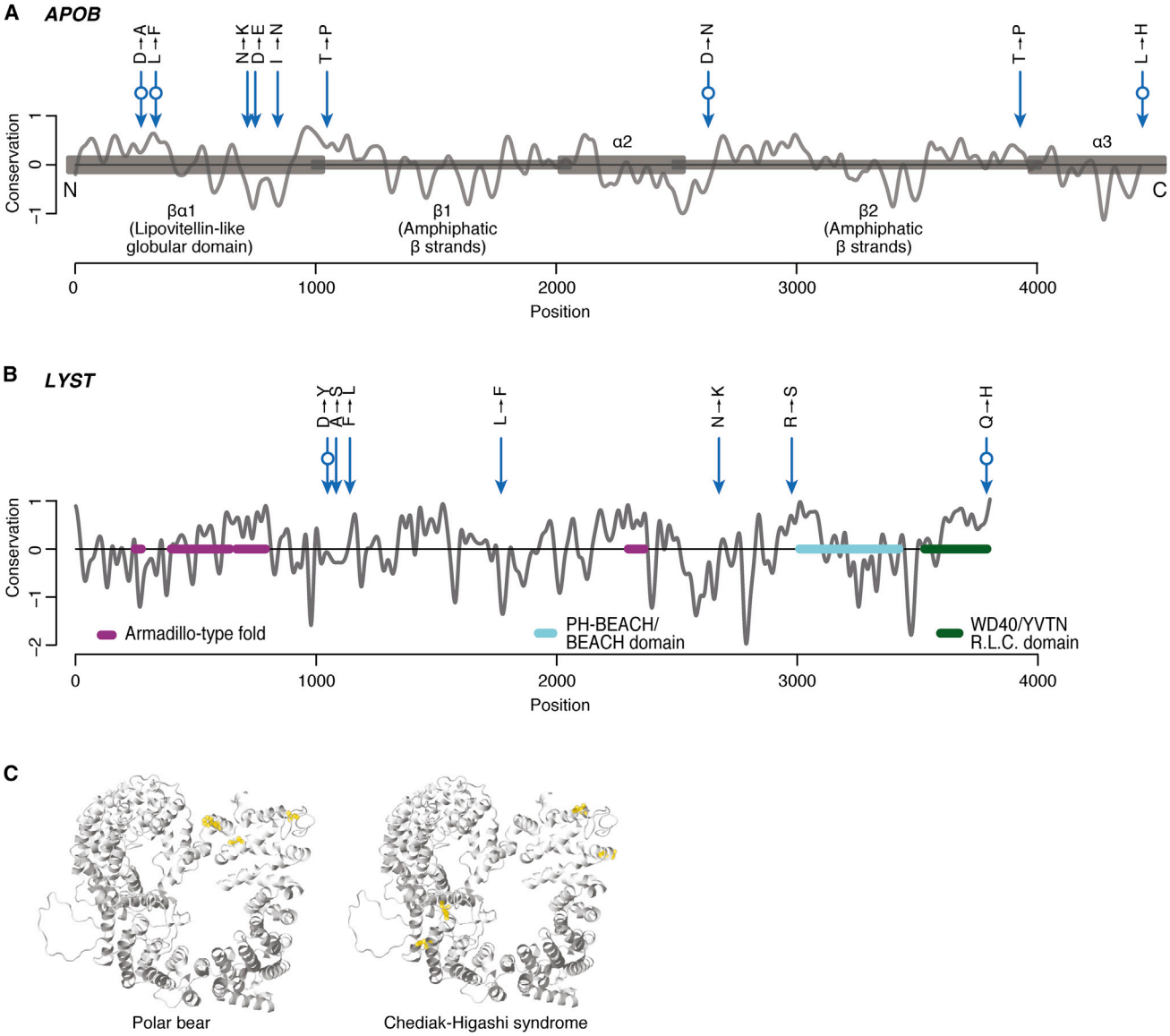

Figure 5. The apoB and LYST Protein Sequences

The distribution of fixed nonsynonymous polar bear mutations (blue arrows) compared to the brown bear, using the giant panda sequence as an outgroup.

(A) Mutations predicted to affect protein structure based on apoB alignments across 20 vertebrate species, using the SIFT algorithm (Sim et al., 2012), are indicated with hollow circles on arrows. The gray curve shows the cubic smoothing spline of the amino acid conservation scores; higher scores indicate higher conservation across 20 vertebrate species. The x axis shows the amino acid position from the N-terminal, the five domains are based on the human apoB sequence (Prassl and Laggner, 2009).

(B) The same representation as in (A), but for the LYST protein sequence. The domains are based on http://www.ebi.ac.uk/interpro/protein/LYST_HUMAN.

(C) Mapping of polar bear-specific substitutions and Chediak-Higashi syndrome causing variants on the protein structure of LYST N-terminal domain.

EXPERIMENTAL PROCEDURES

Detailed extended experimental procedures can be found in Extended Experimental Procedures in the Supplementary Information.

Samples and Data

We deep-sequenced and de novo assembled a polar bear reference genome at a depth of 101X using the Illumina HiSeq 2000 sequencing platform (Table S1; see "Polar Bear Reference Genome and De Novo Assembly" in Extended Experimental Procedures). The scaffold N50 size of the genome was ca. 16 Mb (http://dx.doi.org/10.5524/100008). In addition, we generated complete genomes of multiple polar bears from three management areas around Greenland (Kane Basin and Baffin Bay in Northwest Greenland, and Scoresbysound/Ittoqqortoormiit in Central East Greenland) and several brown bears from Fennoscandia, mainland US and the Admiralty, Baranof, and Chichagof (ABC) Islands off the coast of Alaska (Figure 1, Table S2; see "Samples" in Extended Experimental Procedures). We resequenced 18 polar bear and 10 brown bear genomes at high coverage (an average sequencing depth of ~22X), and an additional 61 polar bear genomes at low coverage (an average sequencing depth of 3.5X) ("Data Generation and QC Measures" in Extended Experimental Procedures). We filtered data with a dedicated pipeline and removed low quality reads as well as sites showing unusual coverage

compared to the empirical distribution, base quality score bias (p < 1e-5), strand bias (p < 1e-5), and deviation from Hardy-Weinberg Equilibrium (p < 1e-3) ("Data Generation and QC Measures" in Extended Experimental Procedures). We analyzed the data within a population genomic framework (Extended Experimental Procedures).

Divergence Time and Joint Demographic History of Polar Bears and Brown Bears

We applied two approaches to estimate reliably when polar bears and brown bears diverged. Importantly, both methods incorporate past population size changes. We used a novel method based on IBS tracts of DNA shared within and between populations (Harris and Nielsen, 2013) and ∂a∂i (diffusion approximation for demographic inference [Gutenkunst et al., 2009]), which infers demographic parameters based on a diffusion approximation to the site frequency spectrum ("Demographic History" in Extended Experimental Procedures).

Gene Flow between Polar Bears and Brown Bears after Divergence

To fully elucidate patterns of gene flow between polar and brown bear populations since their divergence, we used several methods: (1) the IBS tract method; (2) ∂a∂i; and (3) *D* statistics, also known as the ABBA-BABA test (Durand et al., 2011; Green et al., 2010) ("Gene Flow and Introgression" in Extended Experimental Procedures).

Genes under Positive Selection in Polar Bears

We used several complementary approaches to investigate evolutionary changes in protein sequences and analyzed the coding regions of 19,822 genes annotated across the polar bear and brown bear samples, using the giant panda genome sequence (Li et al., 2010) as an outgroup ("Positive Selection" Extended Experimental Procedures). We (1) computed a homogeneity test statistic to identify genes with a low polymorphism-to-divergence ratio in polar bears relative to brown bears; (2) used the Hudson-Aguade-Kreitman (HKA) test to verify that selection had acted specifically on the polar bear lineage and not on the brown bear lineage; (3) estimated F_{st} to identify genes that were highly differentiated between polar bears and brown bears; and (4) used a novel approach to estimate nucleotide diversity within species and divergence between species from low- to medium-quality sequencing data by taking genotype call uncertainty into account (Nielsen et al., 2012; Fumagalli et al., 2014). Because we were interested primarily in identifying completed sweeps unique to the polar bear, we did not apply haplotype-based tests aimed at identifying ongoing selective sweeps. We did not assign simulation-based p values based on specific demographic models to the test statistics. Rather, we used the computed statistics to generate ranked lists of candidate genes, and then further subjected them to statistical enrichment analyses, an approach often referred to as outlier analyses (e.g., Voight et al., 2006).

ACCESSION NUMBERS

The Short Read Archive accession number for the short raw reads reported in this paper is SRA092289.

The DDBJ/EMBL/GenBank the accession number for the Whole-Genome Shotgun project reported in this paper is AVOR00000000. The version described in this paper is version AVOR01000000.

The GigaDB DOI for the remaining data, including the gene set and SNPs, is 10.5524/100008 (http://gigadb.org/dataset/100008).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.03.054.

AUTHOR CONTRIBUTIONS

E.W., J.W., G.Z., and R.N. conceived and supervised the project. E.W.B., R.D., and C.S. provided the polar bear samples. L.D. and L.O. obtained the brown bear samples. E.D.L. extracted the samples. S.L., B.L., W.H., X.X. performed SNP calling, population and gene function analyses. T.S.K. provided support during the bioinformatic analyses of the sequencing data. B.L., Z.X., L.Z., J.L., Z.W., W.F., A.D., C.C.M., M.J.O.C., J.O.M. performed analyses in genome sequencing, assembly, annotation, evolution, and alignment. S.L., M.F., and R.N. designed and performed the population genomic analyses. M.F. and K.H. performed the demographic inference analyses. E.D.L. performed the mitogenome analyses with input form M.F. and R.N. S.L. performed the introgression and selection analyses with input from M.F., E.D.L., and R.N. M.F., M.S., C.B., and G.W. performed the functional analyses of genes under selection. E.D.L. produced the figures with input from K.H. and M.S. E.D.L. wrote the manuscript, with critical input from R.N., M.F., and the remaining authors.

ACKNOWLEDGMENTS

We thank the subsistence Greenland hunters for their valuable participation in obtaining the polar bear samples used in this study, which were collected through a number of projects funded under the DANCEA (Danish Cooperation for Environment in the Arctic) programme. We thank the University of Alaska Museum of the North for providing the ABC brown bear samples and Ilpo Kojola, Katherine Kendall, and John Waller for providing additional samples. We thank Zhaolei Zhang, Yaping Zhang, Fang Li, and Weilin Qiu for help and Xiaoning Wang (South China University of Technology) and the National Gene Bank Project of China and Shenzhen Municipal Government of China (NO.JC201005260191A, CXB201108250096A). E.D.L. was supported by grants from the Danish Council for Independent Research | Natural Sciences (09-069307) and a Marie Curie International Outgoing Fellowship within the European Community 7th Framework Programme (PIOF-GA-2009-253376). M.F. and M.S. were supported by EMBO Long-term Postdoctoral Fellowships (ALTF-229-2011 and ALTF-1475-2010) and grant PJ008116 from Next-Generation BioGreen 21 Program, Rural Development Administration, Republic of Korea. K.H. was supported by a National Science Foundation Graduate Research Fellowship and a UC Berkeley Chancellors Fellowship. R.N. was supported by NIH (2R14003229-07). G.A.W. was supported by NIH (5P-50-GM-081883 to the Duke Center for Systems Biology). C.C.M. and M.J.O'C. thank the Irish Research Council for Science, Engineering and Technology (Embark Initiative to CCM (RS2000172) and Science Foundation Ireland Research Frontiers Program (EOB2673). L.D. was supported by grants from the Swedish Research Council and the BiodivERsA ERA-NET project Climigrate.

Received: September 6, 2013
Revised: December 20, 2013
Accepted: March 4, 2014
Published: May 8, 2014

REFERENCES

Arai, T., Kasper, J.S., Skaar, J.R., Ali, S.H., Takahashi, C., and DeCaprio, J.A. (2003). Targeted disruption of p185/Cul7 gene results in abnormal vascular morphogenesis. Proc. Natl. Acad. Sci. USA *100*, 9855–9860.

Atkinson, S.N., and Ramsay, M.A. (1995). The effects of prolonged fasting of the body composition and reproductive success of female polar bears (Ursus maritimus). Funct. Ecol. *9*, 559–567.

Atkinson, S.N., Nelson, R.A., and Ramsay, M.A. (1996). Changes in the body composition of fasting polar bears (Ursus maritimus): the effect of relative fatness on protein conservation. Physiol. Zool. *69*, 304–316.

Benn, M. (2009). Apolipoprotein B levels, APOB alleles, and risk of ischemic cardiovascular disease in the general population, a review. Atherosclerosis *206*, 17–30.

Cahill, J.A., Green, R.E., Fulton, T.L., Stiller, M., Jay, F., Ovsyanikov, N., Salamzade, R., St John, J., Stirling, I., Slatkin, M., and Shapiro, B. (2013). Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. PLoS Genet. *9*, e1003345.

Cannon, C.P., Shah, S., Dansky, H.M., Davidson, M., Brinton, E.A., Gotto, A.M., Jr., Stepanavage, M., Liu, S.X., Gibbons, P., Ashraf, T.B., et al.; Determining the Efficacy and Tolerability Investigators (2010). Safety of anacetrapib in patients with or at high risk for coronary heart disease. N. Engl. J. Med. *363*, 2406–2415.

Cronin, M.A., Amstrup, S.C., Talbot, S.L., Sage, G.K., and Amstrup, K.S. (2009). Genetic variation, relatedness, and effective population size of polar bears (Ursus maritimus) in the southern Beaufort Sea, Alaska. J. Hered. *100*, 681–690.

De Barba, M., Waits, L.P., Garton, E.O., Genovesi, P., Randi, E., Mustoni, A., and Groff, C. (2010). The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. Mol. Ecol. *19*, 3938–3951.

de Vernal, A., and Hillaire-Marcel, C. (2008). Natural variability of Greenland climate, vegetation, and ice volume during the past million years. Science *320*, 1622–1625.

Dickson, A.J., Beer, C.J., Dempsey, C., Maslin, M.A., Bendle, J.A., McClymont, E.L., and Pancost, R.D. (2009). Oceanic forcing of the Marine IsotopeStage 11 interglacial. Nat. Geosci. *2*, 428–433.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252.

Fumagalli, M., Vieira, F.G., Linderoth, T., and Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics. http://dx.doi.org/10.1093/bioinformatics/btu041.

Galperin, E., Benjamin, S., Rapaport, D., Rotem-Yehudar, R., Tolchinsky, S., and Horowitz, M. (2002). EHD3: a protein that resides in recycling tubular and vesicular membrane structures and interacts with EHD1. Traffic *3*, 575–589.

Gravel, S. (2012). Population genetics models of local ancestry. Genetics *191*, 607–619.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

Gudmundsson, H., Curran, J., Kashef, F., Snyder, J.S., Smith, S.A., Vargas-Pinto, P., Bonilla, I.M., Weiss, R.M., Anderson, M.E., Binkley, P., et al. (2012). Differential regulation of EHD3 in human and mammalian heart failure. J. Mol. Cell. Cardiol. *52*, 1183–1190.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. *5*, e1000695.

Gutiérrez-Gil, B., Wiener, P., and Williams, J.L. (2007). Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais x Holstein population. BMC Genet. *8*, 56.

Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J.A., Arnason, U., and Janke, A. (2012). Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science *336*, 344–347.

Harris, K., and Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet. *9*, e1003521.

Hedberg, G.E., Derocher, A.E., Andersen, M., Rogers, Q.R., DePeters, E.J., Lönnerdal, B., Mazzaro, L., Chesney, R.W., and Hollis, B. (2011). Milk composition in free-ranging polar bears (Ursus maritimus) as a model for captive rearing milk formula. Zoo Biol. *30*, 550–565.

Hegele, R.A. (2009). Plasma lipoproteins: genetic influences and clinical implications. Nat. Rev. Genet. *10*, 109–121.

Herman, D.S., Lam, L., Taylor, M.R., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S.R., McDonough, B., Sparks, E., et al. (2012). Truncations of titin causing dilated cardiomyopathy. N. Engl. J. Med. *366*, 619–628.

Klop, B., Elte, J.W.F., and Cabezas, M.C. (2013). Dyslipidemia in obesity: mechanisms and potential targets. Nutrients *5*, 1218–1240.

Kunieda, T., Nakagiri, M., Takami, M., Ide, H., and Ogawa, H. (1999). Cloning of bovine LYST gene and identification of a missense mutation associated with Chediak-Higashi syndrome of cattle. Mamm. Genome *10*, 1146–1149.

Kurtén, B. (1964). The Evolution of the Polar Bear. Ursus maritimus (Phipps). Acta Zool. Fenn. *108*, 1–26.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., et al. (2010). The sequence and de novo assembly of the giant panda genome. Nature *463*, 311–317.

Lindqvist, C., Schuster, S.C., Sun, Y., Talbot, S.L., Qi, J., Ratan, A., Tomsho, L.P., Kasson, L., Zeyl, E., Aars, J., et al. (2010). Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. Proc. Natl. Acad. Sci. USA *107*, 5053–5057.

Miller, W., Schuster, S.C., Welch, A.J., Ratan, A., Bedoya-Reina, O.C., Zhao, F., Kim, H.L., Burhans, R.C., Drautz, D.I., Wittekindt, N.E., et al. (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc. Natl. Acad. Sci. USA *109*, E2382–E2390.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., and Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. PLoS ONE *7*, e37558.

Olson, T.M., Illenberger, S., Kishimoto, N.Y., Huttelmaier, S., Keating, M.T., and Jockusch, B.M. (2002). Metavinculin mutations alter actin interaction in dilated cardiomyopathy. Circulation *105*, 431–437.

Ormbostad, I. (2012). Relationships Between Persistent Organic Pollutants (POPs) and Plasma Clinical-Chemical Parameters in Polar Bears (*Ursus maritimus*) from Svalbard, Norway. Student thesis (Trondheim, Norway: Norwegian University of Science and Technology).

Paetkau, D., Amstrup, S.C., Born, E.W., Calvert, W., Derocher, A.E., Garner, G.W., Messier, F., Stirling, I., Taylor, M.K., Wiig, O., and Strobeck, C. (1999). Genetic structure of the world's polar bear populations. Mol. Ecol. *8*, 1571–1584.

Pagano, A.M., Durner, G.M., Amstrup, S.C., Simac, K.S., and York, G.S. (2012). Long-distance swimming by polar bears (Ursus maritimus) of the southern Beaufort Sea during years of extensive open water. Can. J. Zool. *90*, 663–676.

Pagès, M., Calvignac, S., Klein, C., Paris, M., Hughes, S., and Hänni, C. (2008). Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. Mol. Phylogenet. Evol. *47*, 73–83.

Pool, J.E., and Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics *181*, 711–719.

Prassl, R., and Laggner, P. (2009). Molecular structure of low density lipoprotein: current status and future challenges. Eur. Biophys. J. *38*, 145–158.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., and Lander, E.S. (2001). Linkage disequilibrium in the human genome. Nature *411*, 199–204.

Runkel, F., Büssow, H., Seburn, K.L., Cox, G.A., Ward, D.M., Kaplan, J., and Franz, T. (2006). Grey, a novel mutation in the murine Lyst gene, causes the beige phenotype by skipping of exon 25. Mamm. Genome *17*, 203–210.

Smith, T.G. (1980). Polar bear predation of ringed and bearded seals in the land-fast sea ice habitat. Can. J. Zool. *58*, 2201–2209.

Stewart, J.R., Lister, A.M., Barnes, I., and Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. Proc. Biol. Sci. *277*, 661–671.

Stirling, I., and Archibald, W.R. (1977). Aspects of Predation of Seals by Polar Bears. J. Fish. Res. Bd. Can. *34*, 1126–1129.

Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005). Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. *28*, 289–301.

Thiemann, G.W., Iverson, S.J., and Stirling, I. (2008). Polar bear diets and arctic marine food webs: insights from fatty acid analysis. Ecol. Monogr. *78*, 591–613.

Trantow, C.M., Hedberg-Buenz, A., Iwashita, S., Moore, S.A., and Anderson, M.G. (2010). Elevated oxidative membrane damage associated with genetic modifiers of Lyst-mutant phenotypes. PLoS Genet. *6*, e1001008.

Trent, J.M., Stanbridge, E.J., McBride, H.L., Meese, E.U., Casey, G., Araujo, D.E., Witkowski, C.M., and Nagle, R.B. (1990). Tumorigenicity in human

melanoma cell lines controlled by introduction of human chromosome 6. Science *247*, 568–571.

Trip, M.D., Smulders, Y.M., Wegman, J.J., Hu, X., Boer, J.M., ten Brink, J.B., Zwinderman, A.H., Kastelein, J.J., Feskens, E.J., and Bergen, A.A. (2002). Frequent mutation in the ABCC6 gene (R1141X) is associated with a strong increase in the prevalence of coronary artery disease. Circulation *106*, 773–775.

van der Ven, P.F.M., Ehler, E., Vakeel, P., Eulitz, S., Schenk, J.A., Milting, H., Micheel, B., and Fürst, D.O. (2006). Unusual splicing events result in distinct Xin isoforms that associate differentially with filamin c and Mena/VASP. Exp. Cell Res. *312*, 2154–2167.

Van Sligtenhorst, I., Ding, Z.M., Shi, Z.Z., Read, R.W., Hansen, G., and Vogel, P. (2012). Cardiomyopathy in α-Kinase 3 (ALPK3)-Deficient Mice. Vet. Pathol. *49*, 131–141.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.

Wang, G., Watanabe, M., Imai, Y., Hara, K., Manabe, I., Maemura, K., Horikoshi, M., Ozeki, A., Itoh, C., Sugiyama, T., et al. (2012). Associations of variations in the MRF2/ARID5B gene with susceptibility to type 2 diabetes in the Japanese population. J. Hum. Genet. *57*, 727–733.

Whitfield, A.J., Barrett, P.H., van Bockxmeer, F.M., and Burnett, J.R. (2004). Lipid disorders and mutations in the APOB gene. Clin. Chem. *50*, 1725–1732.

Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M.B., Brand, T.B., Hofreiter, M., Bunce, M., Poinar, H.N., Dahl-Jensen, D., et al. (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science *317*, 111–114.