Supporting Information

Polar and Brown Bear Genomes Reveal Ancient Admixture and Demographic Footprints of Past Climate Change

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

[1]These authors contributed equally to this work.
[2]To whom correspondence should be addressed. E-mail: cl243@buffalo.edu, lherrera@langebio.cinvestav.mx.

SI Appendix includes:

SI MATERIALS AND METHODS

Bear samples used for genome sequencing

For Illumina HiSeq 2000 genomic DNA sequencing, blood and tissue samples were collected from one American black bear (*Ursus americanus*; hereafter referred to as black bear), one brown bear (*U. arctos*) from the Kenai Peninsula, two brown bears from Alaska's Alexander Archipelago (hereafter referred to as "ABC brown bears"), five polar bears (*U. maritimus*) from Alaska (Chukchi Sea and Southern Beaufort Sea populations), and 18 polar bears from Svalbard (Barents Sea population) (Table S1). Except for one ABC brown bear blood sample (ABC2), which was collected from a rescued cub in a rehabilitation center in Sitka, Alaska, all samples were collected from wild populations.

Sampling of bears in Alaska

Blood and tissue samples were collected from polar and brown bears following standard procedures (1, 2). Briefly, blood samples were stored in EDTA, whole blood Vacutainer® tubes, or a blood preservation buffer (3). All tissue and skin biopsies were placed in a preservation buffer (4M Urea, 0.2M NaCl, 100mM Tris HCL pH 8.0, 0.5% n-lauroyl-sarcosine, 10mM EDTA).

Sampling of polar bear blood in Svalbard

Polar bear tagging and sampling was carried out by researchers from the Norwegian Polar Institute, who conduct long-term research and monitoring of polar bears from the western part of the Barents Sea population, in the Svalbard archipelago area. Data and samples were collected under proper permits using standard procedures. The search and capture for polar bears was carried out by help of a helicopter. Capture efforts focused on high-density areas, and polar bears were immobilized using remote injection of darts (Cap-Chur Equipment) containing Zoletil (Virbac) (4). Bears were individually marked using numbered ear tags, a tattoo in the upper lips, and a microchip. Blood was extracted from a femoral vein into individual anticoagulation-treated vacutainer tubes and kept frozen. Thereafter, the animals were monitored from a remote location to ensure their recovery from anesthesia. No animals were killed or captured as a result of these events. The blood samples were imported into the U.S. under proper import permit (U.S. Fish & Wildlife permit no. 14287A) and with CITES export permits.

Among the Svalbard polar bear samples, one individual female was selected for more in-depth genomic analysis (PB7), since considerable additional data and samples have been collected from her over the past 20 years. This female was captured for the first time on the east coast of Spitsbergen, Svalbard, in April 1990 as a solitary sub-adult. The number N7773 was tattooed in both of her upper lips and a plastic tag imprinted with the same number was put in each of her ears. Furthermore, she was equipped with a satellite transmitter. In addition, various tissue samples of skin, blubber, and blood were collected. Since then she was caught again in 1991 and 1993, and via satellite transmitters her movements were followed from spring 1990 through autumn 1995. In this period, she stayed within a very restricted area in eastern Svalbard (Fig. S1). N7773 was not found again until 2008; a one-year-old male cub accompanied her. In spring 2010 she was caught again (Fig. S1), still in the same area, and blood samples used in this study were collected.

Table S1. Summary of data generated for the 28 bear samples using the HiSeq 2000 sequencing platform (PE = paired-end, MP = mate-pair).

Species (common name)	Internal ID	Sample ID (Sex)	Tissue	Locality	Read type	Number of reads	Number of bases (Gb)
U. maritimus **(polar bear)**	PB1	N23531 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	179,043,011	36.17
	PB2	N23604 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	182,696,799	36.90
	PB3	N23719 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	219,528,250	44.34
	PB4	N23917 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	183,807,167	37.13
	PB5	N23949 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	93,132,954	18.81
	PB6	N26028 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	179,693,644	36.30
	PB7	N7773 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	1,538,311,984	310.74
					82 bp PE	37,254,159	6.11
					101 bp MP	455,407,073	89.97
	PB8	N7968 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	192,131,246	38.81
	PB9	N23985 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	175,731,996	35.50
	PB10	N23997 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	179,962,903	36.35
	PB11	N26029 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	100,081,588	20.22
	PB12	N26030 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	56,865,920	11.49
	PB13	N23355 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	75,944,172	15.34
	PB14	N23379 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	116,186,414	23.47
	PB15	N23694 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	73,854,578	14.92
	PB16	N23797 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	69,562,160	14.05
	PB17	N26024 (M)	Blood	Spitsbergen, Svalbard	101 bp PE	69,451,600	14.03
	PB18	N26025 (F)	Blood	Spitsbergen, Svalbard	101 bp PE	70,799,958	14.30
	AK1	542 (M)	Homogenate	Barrow, AK	101 bp PE	81,552,812	16.47
	AK2	562 (M)	Kidney	Diomede, AK	101 bp PE	80,201,740	16.20
	AK3	574 (M)	Homogenate	Barrow, AK	101 bp PE	81,942,277	16.55
	AK4	2368 (M)	Liver	Diomede, AK	101 bp PE	86,472,564	17.47
	AK5	651 (F)	Tissue	Savoonga, AK	101 bp PE	78,674,747	15.89
	Ancient	Poolepynten	Tooth	Poolepynten, Svalbard	101 bp PE	812,006,446	164.03
U. arctos **(brown bear)**	ABC1	051711 (F)	Blood	Admiralty Island, AK	101 bp PE	873,936,298	176.54
	ABC2	"Lucky" (M)	Blood	Baranof Island, AK	101 bp PE	560,618,941	113.25
	GRZ	100 (F)	Blood	Kenai Peninsula, AK	101 bp PE	955,513,125	193.01
U. americanus **(American black bear)**	BLK	0902155 (M)	Blood	Anchorage, AK	101 bp PE	445,916,120	90.08

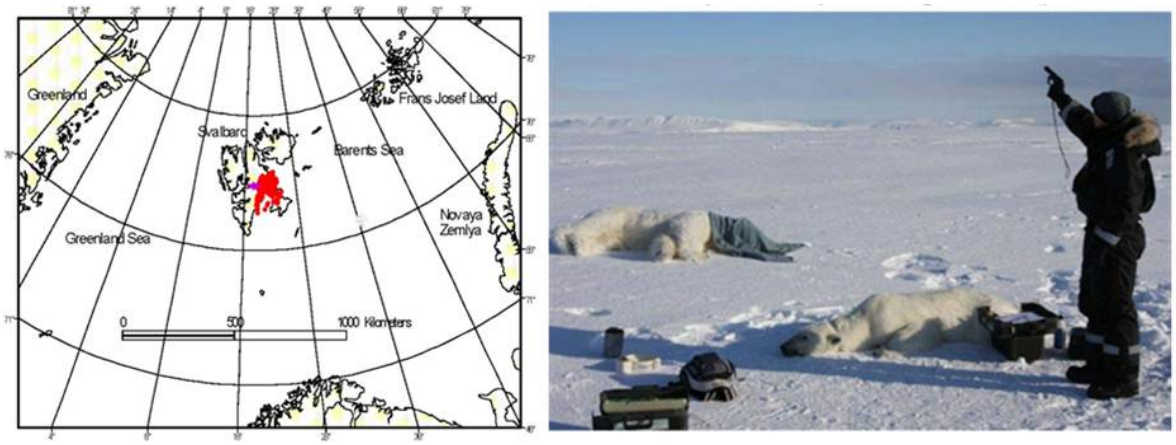


Figure S1. **Svalbard polar bear N7773 (PB7)**. Left: Positions (red dots) received from bear N7773 from spring 1990 to autumn 1995. Right: Tagging of N7773 (in foreground) with an adult male in spring 2010. Another adult male is approaching and must be scared away with a flare gun. Photo: Ø. Wiig.

The ancient Poolepynten polar bear specimen

The polar bear jawbone was excavated in-situ at the Poolepynten coastal cliffs, Svalbard, Norway (Fig. S2) (5). The stratigraphy, depositional history, and environmental development of these coastal cliffs have been studied extensively for a number of years by a team of geologists (5-7). The lowermost unit, where the polar bear jawbone was discovered, is characterized as a marine unit containing an abundant and diversified foraminiferal fauna, dominated by Arctic species similar to modern fauna in shallow sites near Svalbard. This reflects an Arctic, open-marine environment, influenced by glacier input and advection of warm North Atlantic water. The polar bear mandible was discovered at ca. 320 m in the section transect, about 3.5 m above sea level, where it was embedded in a heavily glaciotectonically deformed section of the unit. Based on ^{14}C dating of kelp and infrared-stimulated luminescence (IRSL) of a sediment sample, the unit has been suggested to be of last interglacial (Eemian) to Early Weichselian age. Recent work has revised the lithostratigraphy of the Poolepynten coastal cliffs and re-dated the sequence using the optically stimulated luminescence (OSL) dating technique to constrain the Poolepynten chronology (8). These new OSL ages refine the previous age determinations and support the interpretation that the polar bear jawbone is of last interglacial (Eemian) age (130–115 ky). The polar bear mandible is very well preserved. The only tooth remaining is the canine, which is pre-mortem worn at the apex, suggesting that it belonged to an adult bear.

Figure S2. Ancient polar bear jawbone (right) excavated at Poolepynten on Prins Karls Forland, a narrow strip of land on the far western edge of Svalbard, Norway (left; see map insert). Photo: Ó. Ingólfsson.

Genome sequencing

DNAs from all modern bear tissue and blood samples were extracted using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's recommendations. Extraction of DNA from the ancient polar bear tooth followed previously described methods (9).

A total of 1,674.44 Gb of DNA sequence was generated for 28 bear samples, including genomic DNA from the 130-110 ky old jawbone specimen (9). All the samples were sequenced to generate paired-end reads of length 101 bp using the Illumina HiSeq 2000 sequencing platform (Table S1). The sample PB7 was sequenced to a greater depth of coverage using multiple paired-end libraries with span sizes of 160 bp, 180 bp, and 300 bp, and a mate-pair library of 3 kb.

Genome assembly

We assembled the Illumina short reads for PB7 using SOAPdenovo (10), which has been used with similar data (11). We used the paired-end reads from the short insert size libraries (160 bp, 180 bp, 300 bp) to assemble the genome into contigs. The resulting assembly had an N50 contig length of 3,596 bp and spanned 2.56 Gb. We then aligned all the available sequence data to these contigs, and used the mate-pair information in the order of estimated insert-size (160 bp to 3 kbp) to generate the scaffolds. Scaffolding was followed by local reassembly of sequences in the intra-scaffold gaps, and the resulting draft assembly had an N50 contig length of 61 kbp and spanned 2.53 Gb.

To evaluate the accuracy of the draft sequence, we aligned all the paired-end reads from PB7 back to the assembly using BWA (12). The peak sequencing depth was 127, and more than 25 reads covered over 90% of the assembled sequence. We were able to align around 90% of the reads from the sample to the assembly, and around 75% of the pairs aligned as proper pairs. There are nine known mRNAs for polar bear genes in GenBank. All of these genes were found intact on the scaffolds of the draft assembly using SIM4 (13). Taken together, these data indicate that the draft assembly has good coverage at least in the uniquely mappable regions of the genome.

Identification of nuclear SNPs using the PB7 assembly

Reads from the 27 modern samples were aligned to the draft assembly using BWA version 0.5.9. We used the default parameters to align the sequences, with the exception of "-q 15" which was used to trim the low quality regions on the 3' end of the sequences prior to mapping. The potential PCR duplicates were flagged and collapsed using the MarkDuplicates tool of the Picard (http://picard.sourceforge.net/) software suite.

We identified the putative variant locations using SAMtools (14) version 0.1.16 with the option "-C 50" to reduce the mapping quality of the reads with multiple mismatches. The variant locations in the nuclear genome were filtered to keep positions where the total coverage was less than 750 reads and the RMS (root mean square) mapping quality was greater than 10. Once a list of putative variant locations was constructed, we then used the "pileup" command in SAMtools to call the genotypes for each sequenced individual.

The sequences from the ancient polar bear sample were pre-processed to remove any adapter sequences in the reads using LASTZ (15). The trimmed reads were then aligned to the draft assembly using the same parameters as used for the modern samples. The potential PCR

duplicates were removed using MarkDuplicates and then the alignments were queried using the SAMtools “pileup” command to report the read counts for the two alleles seen in the modern samples.

Identification of nuclear SNPs using the canFam2 (dog) genome assembly

Reads from all the modern samples were aligned to the dog (*Canis familiaris*) whole-genome shotgun (WGS) assembly v2.0 using LASTZ (15). To facilitate the mapping of reads, a "self-masking" process was first used to identify regions in the dog genome where reads should map uniquely. The dog genome was split into fragments and these were then aligned back to the genome. Fragments overlapped each other by half their length. Any reference position appearing in no more than two alignments was considered to be uniquely mappable, since we expect it to align only to the two fragments that include it. Fragments were 100 bp long (as compared to 101 bp for bear reads) and scoring parameters were the same as used in the read mapping stage. By this process, 26% of the dog genome was soft-masked.

Read mapping was performed by aligning bear reads to the soft-masked dog genome. Reads were required to have an alignment with at least 60 matched bases. Reads with more than one alignment were considered mappable only if the best alignment had at least 3 more matched bases than the second best. Alignment for self-masking and mapping were both performed using LASTZ. Scoring parameters were T=2 O=30 E=3 K=300 X=60 Y=60 with the following matrix for substitution scores:

```
    A    C    G    T
A  11  -22  -10  -28
C -22   15  -20  -10
G -10  -20   15  -22
T -28  -10  -22   11
```

These alignments were used as input to a reference-based assembler (16) that generates a consensus contig along with a BAM file of the alignments. We then identified the putative variant locations using SAMtools version 0.1.16, again using the option "-C 50" to reduce the mapping quality of the reads with multiple mismatches, and called the genotypes for each sequenced individual using the same procedures as described above. The sequences from the ancient polar bear sample were similarly processed as described above.

Mitochondrial genome assembly

We used BWA with default parameters to align the reads for all the bear samples (with the exception of the black bear) to a polar bear reference mitochondrial sequence (GenBank accession NC003428). For the black bear sample, the reads were aligned to the reference mitochondrion using LASTZ with the following parameters --yasra85short --coverage=70. 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. We were able to assemble complete mitochondrial genomes from all the 27 modern bear samples that were sequenced. Mitochondrial genome variant locations were called similarly to the nuclear genome, except the upper limit on depth of coverage was increased to 1,000,000.

Transcriptome sequencing

Total RNA from a polar bear (AK018) and a non-ABC brown bear (KEN001) was extracted from whole blood sampled in either tubes containing RNAlater (Life Technologies) or in PAXgene Blood RNA Tubes (PreAnalytiX) using the RiboPure-Blood Kit (Ambion/Life Technologies) or the PAXgene Blood RNA Kit (PreAnalytiX), respectively, following the manufacturer's recommendations. The total RNA was further purified using LiCl precipitation. First and second strand cDNA synthesis was performed with 3 ug of the total RNA using Message Amp-II kit (Ambion) following the manufacturer's recommendations. Each bear species was treated separately. Synthesized cDNA were amplified by in vitro transcription and the resulting 70-90 ug of antisense RNA (aRNA) was purified using Qiagen RNAeasy columns (Qiagen). A second round of cDNA synthesis was performed using the aRNA as template (20 ug). cDNA synthesis was performed as described above except that random primers (mostly hexamers) were used at the first strand synthesis. This procedure yielded about 10 ug of cDNA that was purified using the DNA Clear Kit for cDNA purification (Ambion). cDNA was then treated with Ampure magnetic particles (Agencourt, Beckman Coulter) to obtain fragments of 200 – 700 bp. cDNA samples were prepared with the Ion Xpress™ Template Kit (Life Technologies) according to the Ion Xpress™ Template Kit User Guide and sequenced on a Personal Genome Machine™ (PGM™) sequencer using six 3.18 semiconductor chips, with two chips for each sample. Approximately, a total of nine million reads were generated for each bear species, with an estimated average length of 190 bases (Table S2). Transcriptome reads were masked using the SeqClean software pipeline to eliminate sequence regions that would cause incorrect assembly. Targets for masking include: Ion Adapters, ends rich in Ns (undetermined bases), low complexity sequences, and short reads (< 100 bases). Additionally, natural and artificial duplicates in Ion reads were eliminated using the CD-HIT pipeline (17).

Using GS-Reference-Mapper software (Newbler v2.6), we mapped the masked reads from each bear against our polar bear genome assembly (see above). Around 5 to 7 millions of reads totaling ~0.7-1 Gbp aligned to the reference sequence. As expected, the reference showing the highest percentage of mapped reads was the polar bear (AK018), with 95.44% unique read coverage (Table S2, Fig. S3). Additionally, the masked reads were mapped to 22,291 dog (*Canis familiaris*) cDNA sequences (mRNA). 1,430,009 (24.28%) masked reads from AK018 aligned to 14,274 different *C. familiaris* cDNA sequences. Meanwhile, a comparable percentage of reads (~20%) from KEN001 were mapped to 15,782 dog genes, respectively. These numbers represent the most stringent underestimation of the minimal number of bear genes found expressed in the two species samples.

Table S2. Transcriptome read statistics.

	Polar bear (AK018)	**Brown bear (KEN001)**
No. of raw reads	8,571,870	9,786,592
Masked reads	6,137,359	8,024,099
No. of bases	1,166,774,629	1,538,917,893
Avg. length (nt)	190.11	191.79
Masked reads mapped to polar bear assembly:		
No. of reads	5,621,408 (95.44%)	7,187,145 (94.95%)
Bases	845,486,232	1,074,548,214
Masked reads mapped to Canis familiaris cDNA:		
No. of reads	1,430,009 (24.28%)	1,793,064 (23.69%)
Bases	158,695,718	197,231,281

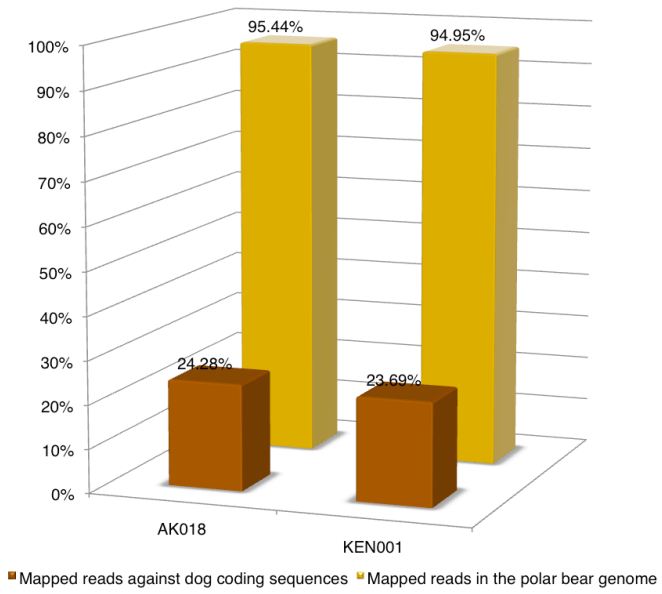


Figure S3. Transcriptome read mapping. Percentage of masked reads from a polar bear (AK018) and a brown bear (KEN001) mapped against dog (*Canis familiaris*) cDNA sequences (brown bars) and the polar bear (PB7) genome assembly (yellow bars).

Phylogenetic and dating analyses

Mitochondrial DNA genome analysis

Bear mitochondrial DNA genomes contain a variable number of tandem repeats (VNTR) in the control region. This repeat region was excluded from further analyses since the short nature of Illumina reads makes it difficult to accurately assemble. Diversity among our new mitochondrial genomes and publicly available bear genome sequences is depicted in Figure S4. Mitochondrial genome sequences were aligned using MUSCLE with the default parameters (www.ebi.ac.uk/Tools/msa/muscle/). The final aligned sequence length (excluding the VNTR) was 16,479 bp.

Phylogenetic analysis was conducted using both maximum likelihood and Bayesian inference. Maximum likelihood analyses were conducted using the RAxML Black Box (18). We used the general time reversible (GTR) model with rate heterogeneity modeled by a gamma distribution as well as a proportion of invariable sites. A total of 1000 bootstrap replicates were conducted to assess branch support. PAUP* was used to create a consensus of the 1000 bootstrap trees. Bayesian phylogenetic inference was conducted in BEAST v1.6.1 (19). The same substitution model was specified as above, in addition to the uncorrelated lognormal relaxed clock model. For this analysis, tip dates for ancient samples were set to 32 and 44 Ky (thousand years) for the cave bears and 120 Ky for the ancient polar bear. The coalescent constant tree prior was utilized, with other priors (largely uninformative) left at their default settings. Multiple independent runs were performed for 50 x 10^6 generations with states recorded every 5,000 generations, and the first 10% discarded as burn-in. Stationarity was assessed through plots of –lnL vs. generation using Tracer v1.5 and through investigation of effective sample sizes, all of which were > ~ 400. Convergence was examined through comparison of results from multiple independent runs. The maximum clade credibility tree was created using the program TreeAnnotator v.1.6.1. Genbank accession numbers for the individuals in Figure 2A (in the main text) are given in Table S3.

Figure S4 (*next page*). SNP map of mitochondrial genomes from 40 polar and brown bears. Single-nucleotide substitutions as compared with a reference polar bear genome (GenBank accession NC003428) are depicted with black bars. Uma: polar bear, Uar: brown bear.

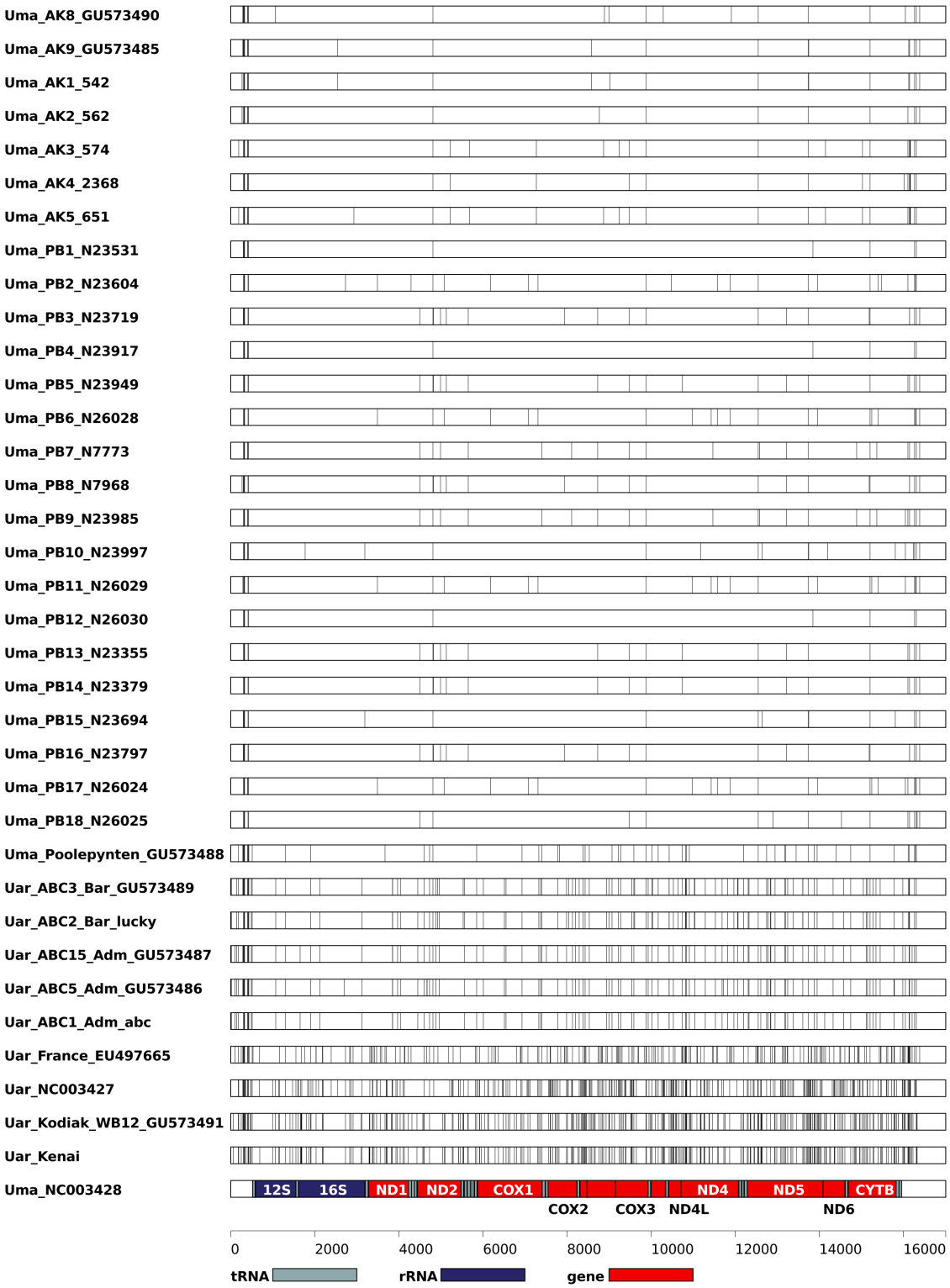


Table S3. Information for 40 mitochondrial genome sequences included in analyses. (* = previously published mitochondrial genomes obtained from GenBank.)

Species	Label	Internal ID	GenBank
American black bear	Black (BLK)	BLK014	JX196366
	Black 1*	-	NC003426
Cave bear	Cave 1*	-	NC011112
	Cave 2*	-	EU327344
Brown bear	Brown*	-	NC003427
	Kodiak*	WB12	GU573491
	Kenai (GRZ)	GRZ	JX196367
	France*	-	EU497665
	Baranof*	ABC 3	GU573489
	Baranof (ABC2)	ABC2	JX196368
	Admiralty 5*	ABC5	GU573486
	Admiralty 15*	ABC15	GU573487
	Admiralty (ABC1)	ABC1	JX196369
Polar bear	Ancient*	-	GU573488
	Alaska (AK1)	AK1	JX196370
	Alaska (AK2)	AK2	JX196371
	Alaska (AK3)	AK3	JX196372
	Alaska (AK4)	AK4	JX196373
	Alaska (AK5)	AK5	JX196374
	Alaska 8*	AK8	GU573490
	Alaska 9*	AK9	GU573485
	Polar*	-	NC003428
	Svalbard (PB1)	PB1	JX196375
	Svalbard (PB2)	PB2	JX196376
	Svalbard (PB3)	PB3	JX196377
	Svalbard (PB4)	PB4	JX196378
	Svalbard (PB5)	PB5	JX196379
	Svalbard (PB6)	PB6	JX196380
	Svalbard (PB7)	PB7	JX196381
	Svalbard (PB8)	PB8	JX196382
	Svalbard (PB9)	PB9	JX196383
	Svalbard (PB10)	PB10	JX196384
	Svalbard (PB11)	PB11	JX196385
	Svalbard (PB12)	PB12	JX196386
	Svalbard (PB13)	PB13	JX196387
	Svalbard (PB14)	PB14	JX196388
	Svalbard (PB15)	PB15	JX196389
	Svalbard (PB16)	PB16	JX196390
	Svalbard (PB17)	PB17	JX196391
	Svalbard (PB18)	PB18	JX196392

Molecular dating of mitochondrial genomes

We also performed additional analyses to estimate dates for the splits among different bear maternal lineages using the software BEAST (19). We conducted analyses with individual lognormal priors for each of the ancient samples, as well as analyses with and without an age dependent error rate for transitions (which may be due to artifacts caused by DNA degradation). In all analyses the median estimate for the date of the split between the ABC brown bears and the polar bears was between 157 and 162 Ky (the 95% highest posterior density [HPD] interval was approximately 107 – 240 Ky; Fig. S5), except for the analysis that included an age dependent error rate, for which the split was estimated to 148 Ky (95% HPD: 107 – 217 Ky). Bayes Factors estimated in Tracer showed no support for the model with error rate over the model without error rate (2lnBayes Factor = -2.17). The estimated dates are consistent with those found previously (9). See Table S4 for a summary of some published split time estimates based on nuclear and mtDNA sequence data for selected ursid lineages (Fig. S5, insert).

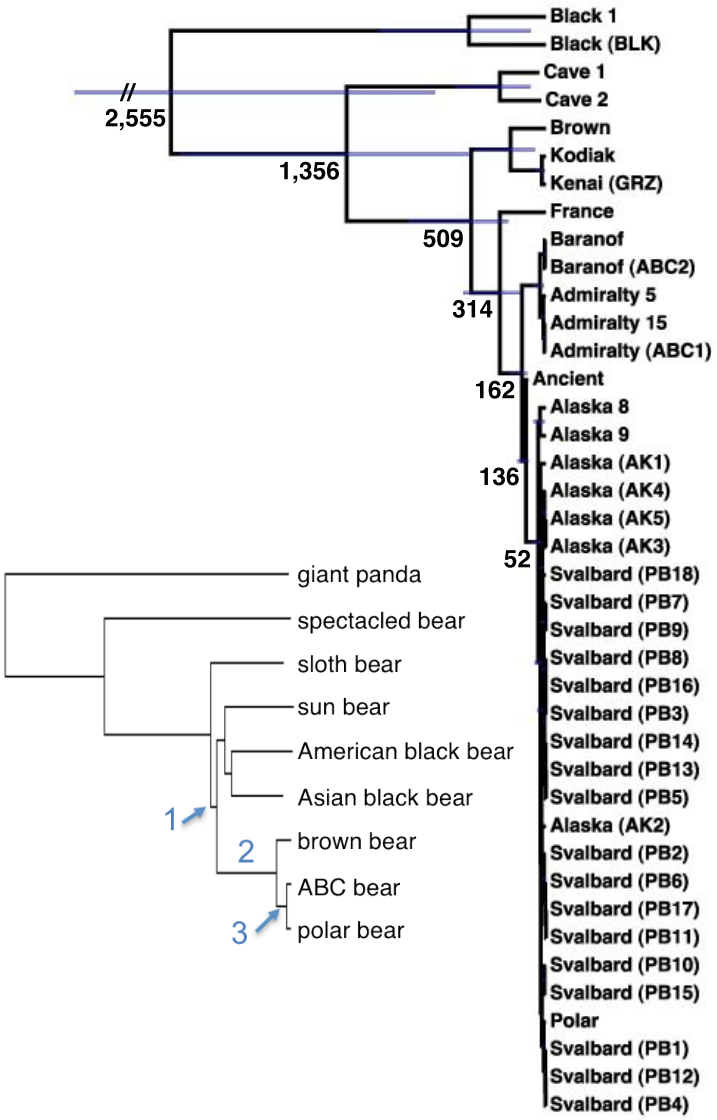


Figure S5. Maximum clade credibility tree from BEAST for 40 mitochondrial genomes (exluding the VNTR region). Blue bars indicate the 95% highest posterior density interval for estimates of divergence time, and numbers at nodes indicate the median estimate for the age in thousands of years (Ky). The credibility interval for the split between the lineages leading to black bears and cave/brown/polar bears has been truncated. New mitochondrial genomes are indicated with their ID in parenthesis. Insert: mtDNA maximum likelihood tree of the bear family (Ursidae) modified from (9). The blue numbers refer to the nodes, for which estimated split times are reported in Table S4.

Table S4. Some published split time estimates (Mya) for selected ursid lineages (see Fig. S5, insert)

Node, Lineage	Fossil data	Nuclear				mtDNA				
	Kurten 1968 (20) Wayne et al. 1991 (21)	Goldman et al. 1989 (22)	Yu et al. 2004 (23)	Edwards et al. 2011 (24)	Hailer et al. 2012 (25)	Talbot & Shields 1996 (26)	Yu et al. 2007 (27)	Bon et al. 2008 (28)	Krause et al. 2008 (29)	Lindqvist et al. 2010 (9); this study
1, **Black-brown**	2.0	~5	–	–	0.95	~6	–	2.8	5.05	2.6
2, **Brown-polar**	0.07-0.1	2-3	1-1.5	0.4-2	0.6	0.3-0.4	1.32	0.6	0.88	0.5
3, **ABC-polar**	–	–	–	–	–	–	–	–	–	0.15-0.16

–, no divergence time estimate reported; note that the "Black" lineage includes Asian black bear and sun bear. References (9, 27-29) are studies based on complete mitochondrial genomes.

Nuclear genome phylogenetic trees

The nuclear DNA tree was constructed via the neighbor joining algorithm from a distance matrix calculated using the "Phylogenetic Tree" tool on Galaxy (30-32). Galaxy computes the distance between two individuals by summing, over SNPs, the absolute value of the difference between estimated minor-allele frequencies (estimated by either read counts or genotype), then dividing by the number of SNPs where the estimations were obtained. The dataset of SNPs discovered from the polar bear assembly was selected, the minimum number of reads was set to 2, and distances were calculated using the sequence coverage to infer the genotype. Additional trees were constructed using higher minimum number of reads, different settings for minimum quality, and use of the SAMtools called genotype instead. All of these settings produced trees with identical topologies regarding non-ABC brown, ABC brown, and polar bear relationships, except that when the minimum number of reads was very high some modern polar bear individuals with lower sequencing depth were excluded from the analyses. Finally, we also reduced the data set of SNPs so that there was 0% missing data. This data set also produced an identical topology.

Test for positive selection on brown and polar bear mitochondrial genomes

We tested for evidence of positive selection using the dataset of 36 brown and polar bear mitochondrial genomes. Since the widely used dN/dS ratio (which compares the ratio of non-synonymous to synonymous substitutions) can fail to detect the signature of positive selection in highly conserved gene sequences, tests for selection were carried out using the program TreeSAAP (33, 34). This program implements the MM01 model and takes into account the physio-chemical properties of amino substitutions as well (e.g., ranging from 1 for a conservative change to 8 for a very radical change). A significant signal of positive selection is detected when the observed distribution of genetic changes differs from that expected under the null hypothesis of selective neutrality.

The phylogenetic tree utilized to test these hypotheses (Fig. S6) was constructed using RAxML (18) as described above. We tested for the signal of positive selection in the thirteen protein coding genes of the mitochondrial genome: ATP6, ATP8, Cox1, Cox2, Cox3, Cytb, ND1, ND2, ND3, ND4, ND4L, ND5, and ND6 (11,405 bp total). The best fit model of nucleotide substitution for each gene was investigated using the program jModelTest (35) and the Akaike Information Criterion. The HKY model was selected for ATP8, Cox2, Cox3, Cytb, ND1, ND2, ND5, and ND6; the TrN model was selected for ATP6, Cox1, ND3, and ND4L; and the GTR model was selected for ND4. Tests were conducted in TreeSAAP using the 31 amino acid properties available and a sliding window size of 20 codons. Non-synonymous substitutions were considered to be under positive selection if the change was relatively radical (i.e. levels 6-8), and if Z-scores were significant ($p<0.001$) both for the sliding windows containing the substitution and for the physiochemical change.

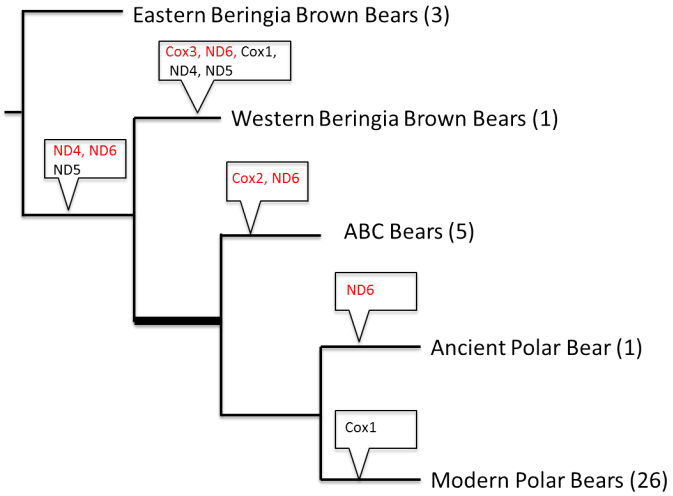


Figure S6. Phylogenetic tree constructed in RAxML from complete sequences of 36 brown and polar mitochondrial genomes (minus the VNTR). Numbers in parentheses indicate number of sequences. This highly supported topology was utilized for tests of positive selection in TreeSAAP. Genes in red have highly significant hits at $p < 0.001$, whereas genes in black have significant hits at $p < 0.01$. No evidence of selection was found on the branch (in bold) leading to ABC brown plus polar bears, nor the branch leading to the polar bear lineage.

Coalescence hidden Markov model of bear demographic history

To learn about bear demographic history and to make inferences about both split times and gene flow, we applied a coalescence hidden Markov model (CoalHMM) to four of our deep-coverage bear genomes: the polar bear (PB7), one ABC brown bear (ABC1), the non-ABC brown bear (GRZ), and the black bear (BLK). We first employed a simple isolation model (36), comparing pairs of genomes under the assumption of allopatric speciation. For all comparisons we obtained unexpected estimates of very recent split times and very large ancestral effective population sizes (N_e). If the true demographics involved were not simple splits but instead initial splits followed by prolonged periods with structured populations and gene flow, these findings would be consistent with misspecification of the demographic model. We therefore applied an extended model, isolation-with-migration, estimating an initial split time followed by a period of gene flow before a complete split.

Isolation-with-migration CoalHMM

To estimate split times, taking into account gene flow or admixture events following an initial split, we applied a coalescent hidden Markov model similar to Li and Durbin (37) and Mailund

et al. (36). CoalHMMs exploit changes in the local genealogy along a genome alignment for demographic inference; in this particular case the changing coalescence time between two haploid genomes.

The demographic model used in our CoalHMM assumes that at time τ2 in the past a panmictic mating population split into two subpopulations. These populations then continued to exchange genes until later time τ1 where gene flow ended, and the populations – now species – evolved independently afterward. The model is parameterized with five parameters (Fig. S7), the coalescence rate, C, (the inverse of the effective population size), the recombination rate, R, the migration rate between the populations, M, the initial split time, τ2, and the end of gene flow, τ1. Assuming a calibrated substitution rate, the split time parameters can be translated into years and the coalescence rate into an effective population size.

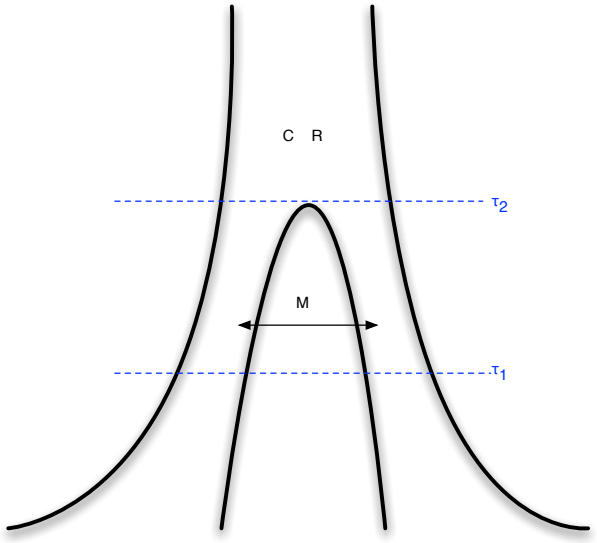


Figure S7. Isolation-with-migration model. Our isolation-with-migration model considers two separated populations (sub-species or species) derived from a shared ancestral population in the recent past. The model assumes that the ancestral population split into two populations in the past, at time τ_2, and that these two populations exchanged genes with migration rate M until a later time τ_1, where gene flow stopped. The coalescence process in this model is parameterized with a coalescence rate (inverse of the effective population size), C, and a recombination rate, R. The model is translated into a finite-state hidden Markov model by discretizing time into intervals with break points.

Through coalescence simulations we have tested that the model accurately estimates parameters, except for the recombination rate, which is underestimated (Fig. S8). The latter is a known consequence of the Markov assumption along the genomic sequence (36). While the mathematical framework of the CoalHMM does allow the coalescence rate to change over time and to be different in the two populations, there are identifiability issues that prevent us from estimating different coalescence rates. We have examined this in (36) and found that changes in the individual populations' coalescence rates by an order of magnitude, without changing the estimated rate, essentially captures only the ancestral species' coalescence rate.

The model compares haploid genomes, but since our genomes are diploid and unphased, we tested how picking random phases for heterozygotic sites would affect the parameter estimates (Fig. S9 and S10). Based on simulation results we expect a random phase to lead to an overestimate of the coalescence rate for the most closely related genomes, and a slight overestimate of the earliest split time if it is in the low hundreds of thousands, but otherwise the model will give correct estimates using a random phase.

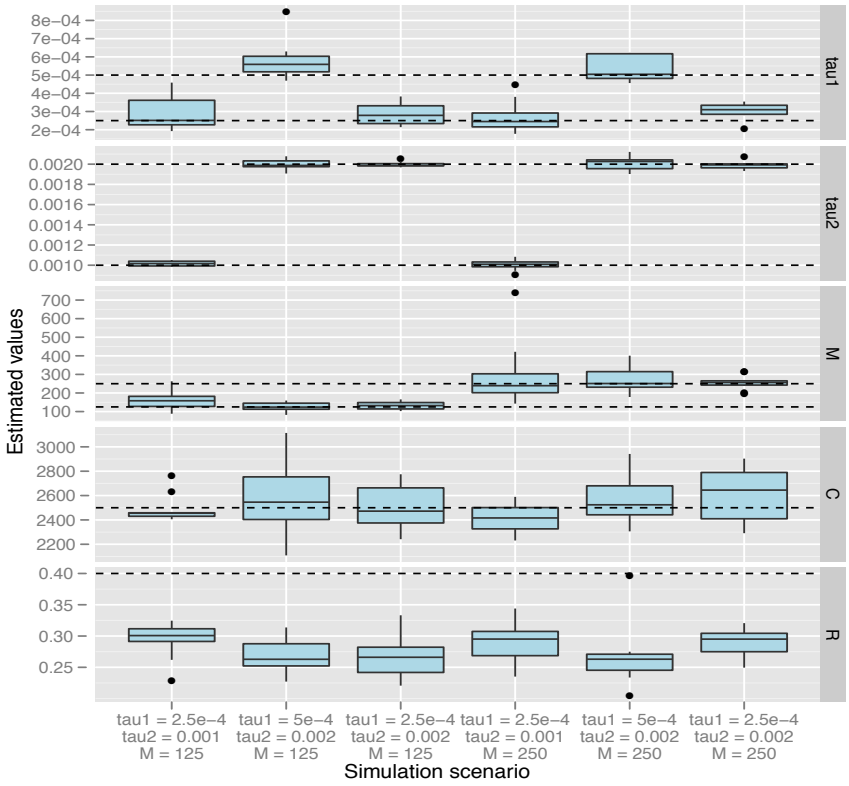


Figure S8. Validation of parameter estimation. Box plot showing the distribution of parameter estimates for six different simulation scenarios. In all scenarios the coalescence rate and the recombination rate parameters are kept fixed, while the split times and migration rate varies between scenarios. For each simulation scenario, 10 independent data sets were generated and analyzed. The dashed horizontal lines indicate the simulated values for the five parameters.

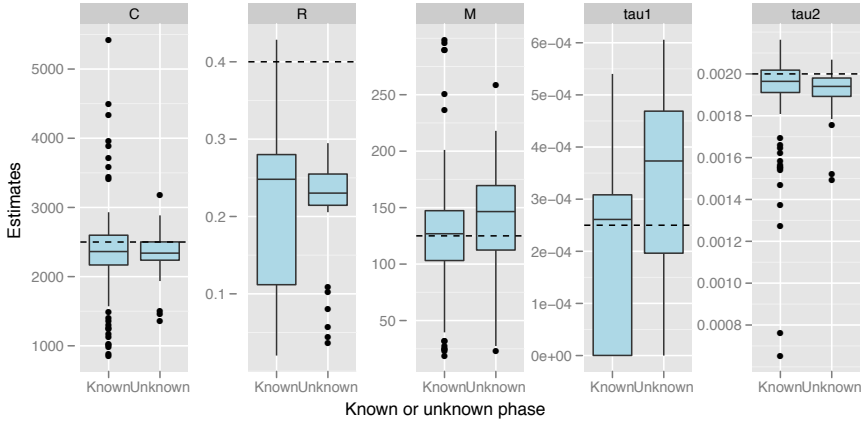


Figure S9. Effect of not knowing the genotype phase. We simulated the situation where the genotype phase is unknown by simulating two genomes and selecting a random allele for all heterozygotic sites. The plot shows the effect on parameter estimates of not knowing the phase.

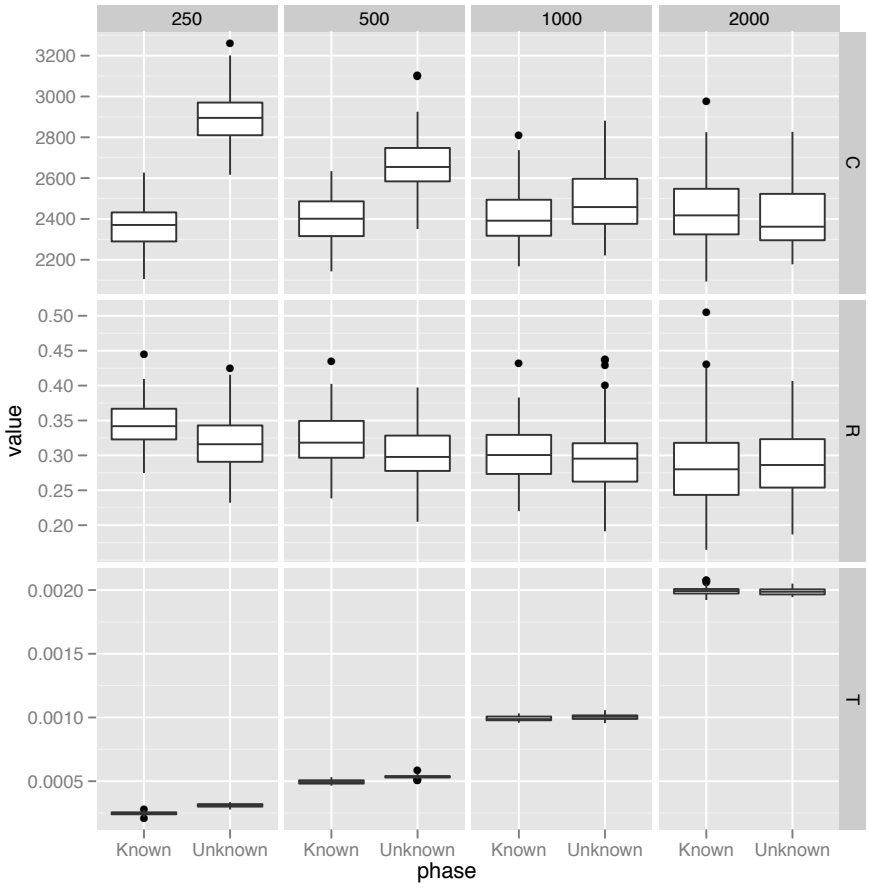


Figure S10. The effect of divergence time on parameter estimates using a random phase. Using the simpler isolation model (36) we explored the effect of not knowing the phase but picking a random phase for four different split times (x-axis). The main effect of incorrect phase is an overestimate of the coalescence rate and a slight overestimation of the split time for the most recent split time.

Data analysis pipeline

We split the genome into segments of ~10Mbp by selecting contigs until the total length exceeded 10 Mbp. We constructed pairwise alignments of the genomes by mapping variable sites onto the contigs – ignoring indel variation – and picking random alleles for heterozygotic sites. In order to detect potential problems with random phasing, we sampled 8 random-phased alignments for each 10Mbp segment and analyzed these independently. In general, we found very little variation between random phase samples compared to the variation between genomic segments, suggesting that the unknown phase is not a large concern.

We found a bimodal estimate of the initial divergence of brown bears and polar bears (T2; Fig. S11) The strongest correlation within the estimated parameters appear to be between the effective population size (the coalescence rate) and the split times where the recent T2 correlates strongly with larger effective population sizes (around 50,000-60,000 rather than 10,000). This is to be expected, since the model among other things needs to fit the mean genomic divergence time, a divergence that is determined by the species split time plus the mean coalescence time within the ancestral species (two times the effective population size). If the split time is decreased, then effective population size must be increased, and vice versa. Whether an effective population size around 50,000 is reasonable depends on how this parameter is interpreted. The effective population size is only partly determined by census population size; population structure has a large effect on the parameter. If, as we suggest, the model is picking up signals of

admixture events, then the small divergence time and large effective population size could both be a consequence of a complex population structure that the model is not capturing.

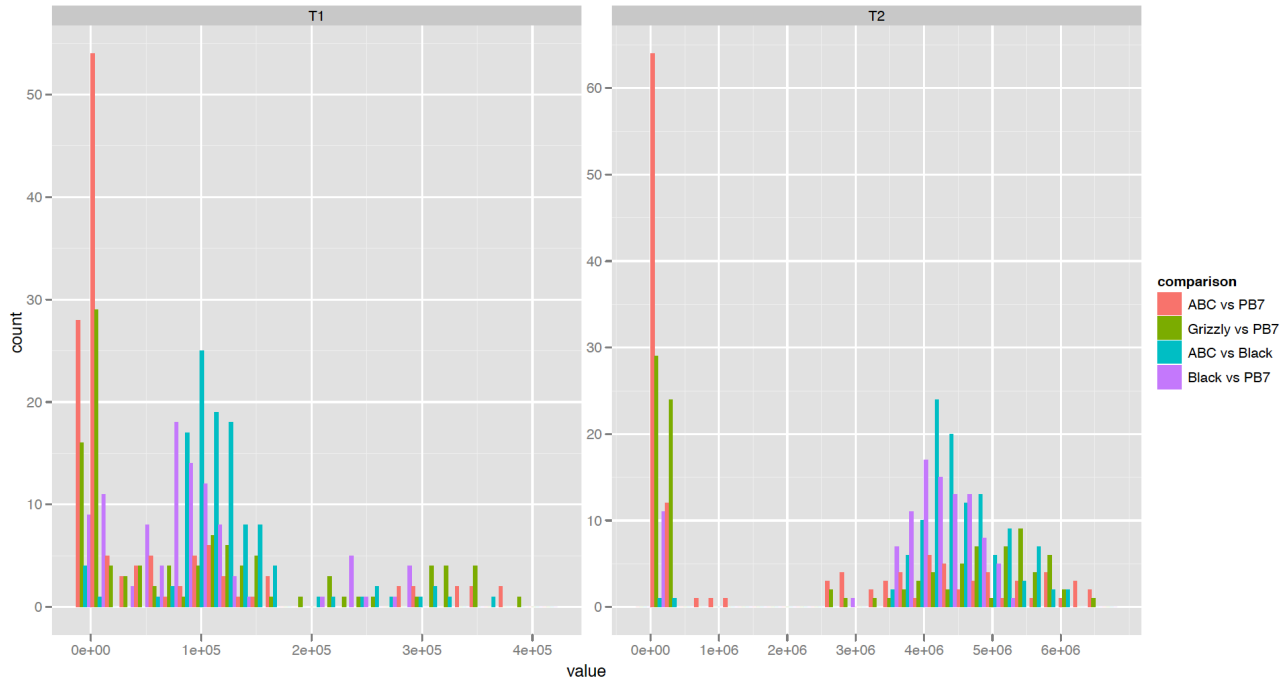


Figure S11. Histogram of split time estimates between members of pairs**.** T1 (left) indicates the estimated time for the end of gene flow, while T2 (right) indicates estimates of initial split time. The pairwise runs shown here are: ABC1 vs. PB7 (red), brown vs. PB7 (green), ABC1 vs. black (blue), and black vs. PB7 (purple). Note the marked bimodal time distribution in T2 for the comparisons between ABC1 and PB7 (in particular) and brown vs. PB7.

PCA-based "admixture fraction" estimations (Figure 4A in main text)

The underlying PCA-plot in Fig. 4A was computed by Galaxy commands, one of which runs the program smartpca (38) and estimates "admixture fractions". The particular commands are available as a "workflow" on the Galaxy website (usegalaxy.org). The PCA-based estimation of admixture fractions uses a method proposed by (39) and used by, e.g., (40). Assume that smartpca computes the eigenvalues e and f (for the first two principal components), and reports the coordinates (a, b) for individual ABC, (w, x) for GRZ and (y, z) for the centroid of the PBs. Then the fraction of the way along the segment from GRZ to the PBs *taken in the plane defined by the principal component vectors in the original space of genotype data* is d/c where:

$a' = e \times a$

$b' = f \times b$

$w' = e \times w$

$x' = f \times x$

$y' = e \times y$

$z' = f \times z$

$c = (y' - w')^2 + (z' - x')^2$

$d = (y' - w') \times (a' - w') + (z' - x') \times (b' - x')$

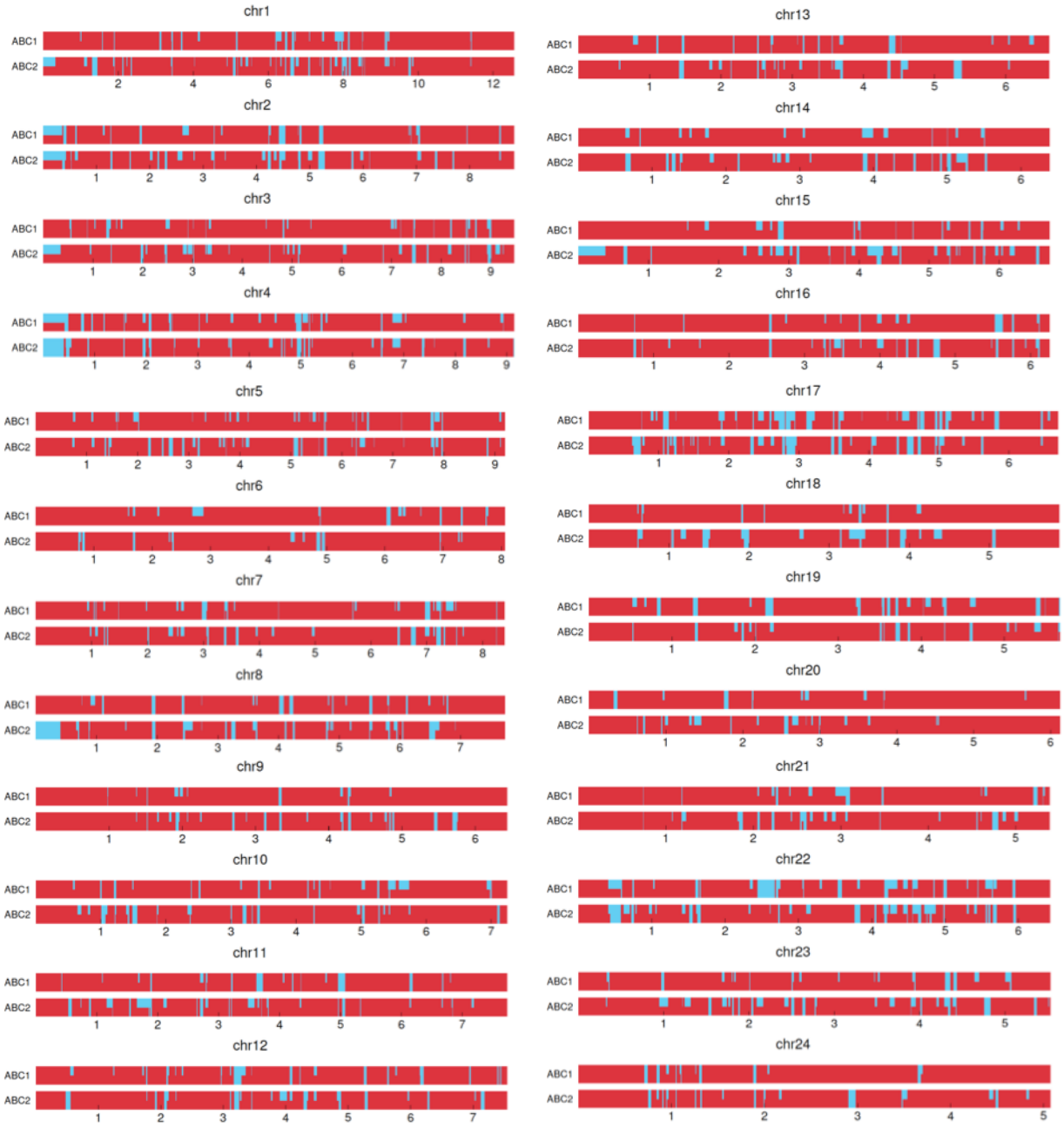
The first six commands, which multiply coordinates by the respective eigenvalues, in essence translate points to their coordinates with respect to unit vectors in the plane of the first two principal components in the space of genotypes. Then d/c gives the fraction of the way along the line segment from (w', x') to (y', z') that is closest to (a', b').

Because we employed smartpca as a "black box", without knowledge of its internal details, we used simulated data to evaluate the accuracy of our estimations of admixture fraction. The idea was: take two sets of allele frequencies from actual populations of bears, synthesize a set of genotypes by sampling a known fraction g from the second population and the remainder from the first population, perform our computation, and compare g with the computed estimation. We found that smartpca and our post-processor systematically underestimate the true g, typically producing a value of roughly $0.6\times g$.

The admixture map (Figure 4B in main text)

The admixture map illustrated in Fig. 4B (see the full map in Fig. S12) was created with Galaxy commands, starting with the assignment of F_{ST} values to SNPs. Galaxy provides both Sewall Wright's original definition of F_{ST} (41) and an unbiased estimator thereof (42). We used the original form and a set of "ancestry informative markers" consisting of the 2,240,636 autosomal SNPs where the F_{ST} between the polar bears (denoted PB) and the 2-chromosome "population" of autosomes from the sequenced non-ABC brown bear (called GRZ) satisfied $F_{ST} \geq 0.5$. For each of the two ABC brown bears, and each possible genotype, 0, 1, or 2, corresponding to the number of copies of the reference allele (inferred from the dog genome) at that position in that individual, we determined the probability of generating the genotype by random sampling from PB or GRZ. For instance, suppose the reference allele's frequencies for PB and GRZ are f and g, respectively, and we observe heterozygosity (genotype 1) at the SNP. Modeling this genomic region as two chromosomes randomly selected from PB, the genotype's probability is $2f \times (1–f)$, while modeling it as one chromosome from each of PB and GRZ, the probability is $f \times (1–g) + g \times (1–f)$. Continuing this line of reasoning, we see that for each of the three possible observed genotypes in a putatively admixed individual, knowing allele frequencies in PB an GRZ, we get a probability of generating that genotype from each of the three theoretical models for the admixture state at that genomic position. We combine these signals by a dynamic-programming method that is akin to Viterbi's method for hidden Markov models (HMMs). The process was very efficient, producing the genome-wide admixture map in approximately one minute. All the data and tools that we used for the admixture analysis can be found on Galaxy, including a "workflow" of the exact commands and their parameters that we used.

Figure S12. Admixture maps for two ABC brown bears in terms of polar bears and a non-ABC brown bear. The maps are drawn based on presumed orthology to chromosomes in the canFam2 assembly of the dog genome. Coordinates for chromosomes 1-25 are in units of 10 Mbp, with units of 1 Mbp for chromosomes 26-38. Red regions are modeled as having both chromosomes from non-ABC brown bears, blue regions are modeled as having both chromosomes from polar bears, and half-red-half-blue regions are modeled as having one chromosome of each type. The map was produced using the Galaxy web server with a "switch penalty" of 50. Exact coordinates can be found by re-running the workflow available at Galaxy. Note that many polar-bear-like genomic intervals are long (up to ~4 Mbp) and strongly divergent between the two individuals sequenced.



Such admixture maps can be mined for clues to regions under selection (43). The basic idea is that admixture events introduce genetic variants that may increase in frequency and be preferentially retained by selective forces. We inspected the 135 genomic intervals where both chromosomes in both ABC1 and ABC2 were more similar to polar bears than to the other brown bear. We hypothesized that there might be some selective advantage for acquisition (or retention) of particular variants of some genes that we predicted to lie in these intervals. Our approach was to ask which KEGG pathways have the most dependence on genes in that set. The glycine, serine and threonine pathway was identified as the most affected, by virtue of the necessity of the ALDH7A1 gene. This gene is involved in the synthesis of betaine from betaine-aldehyde, and interestingly has been found to protect mammal and plant cells against hyperosmotic stress (44, 45). A differential betaine accumulation might be expected from a differential expression/activity of ALDH7A1, and might be a molecular adaption to cold, salinity, and dehydration (44). We offer this example to illustrate some of the opportunities that our data and software make possible.

Genome-wide SNP diversity in modern polar bears

To better interpret the number of polymorphic positions observed in the modern polar bears (2,540,010), we created an analogous set of data for humans. We started with 12 published sets of Illumina sequences for five Europeans, four Africans, and three Asians, taking only between 4-fold and 8-fold coverage per individual. We applied the same alignment and SNP-calling software that we used for the bear sequences, and detected 4,952,040 variable positions among the Europeans, 6,011,697 among the Africans, and 3,491,384 among the Asians.

SNP genotyping and population structure in polar bears

For SNP genotyping and validation we utilized samples from 118 bears (Table S5, Fig. 5 in the main text). These samples included 25 modern brown and polar bears for which genomic sequencing was performed (Table S1). Other blood or tissue samples were obtained from collections held at the Alaska Science Center, USGS. In other cases, we received DNA aliquots from the Alaska Science Center, USGS, or the Natural History Museum, University of Oslo. DNAs from all modern bear tissue and blood samples were extracted using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's recommendations. DNA aliquots that were received from the Alaska Science Center were extracted via a "salting out" procedure using previously published protocols (46). Extraction of DNA from the ancient polar bear tooth followed previously described methods (9).

SNP selection, genotyping, and validation

From the SNPs discovered by genomic sequencing of 23 modern polar bears, a set of 100 high quality SNPs were identified based on the criteria that loci would be variable among the 23 polar bears and that both alleles in each locus were unambiguous in the ancient polar bear. SNP loci were selected on the autosomes, and could be in either coding or non-coding regions, with at most 4 – 5 per chromosome in the corresponding location in the dog reference genome (canFam2). Primers were designed using Galaxy to amplify products approximately 100 bp in length (Table S6).

Table S5. Geographical sampling for SNP genotyping among brown and polar bear populations.

Species/Lineage	Locality	N
Brown bear		**60**
Eastern Beringia	*Southeast Alaska*	*18*
	Kenai Peninsula	*5*
Western Beringia	*Seward Peninsula*	*8*
	Kodiak Peninsula	*6*
	Kamchatka Peninsula	*8*
Norway		*4*
Yellowstone NP		*2*
ABC	*Admiralty, Baranof*	*9*
Polar bear		**58**
	Chukchi Sea	*18*
	Barents Sea (Svalbard)	*18*
	Southern Beaufort Sea	*11*
	Southern Hudson Bay	*10*
	Ancient (Svalbard)	*1*
Total samples		**118**

We conducted multiplex polymerase chain reaction (PCR) amplification of the SNP loci with 12-13 loci amplified per well. PCR conditions were optimized over a range of annealing temperatures and using different concentrations of PCR additives. Final reaction mixtures, 100 uL total volume, contained 1X PCR Buffer II (Applied Biosystems), 3.0 mM MgCl, 0.2 mM each dNTP, 0.1 uM each forward and reverse primer, 0.04% BSA, 5% DMSO, 5.0 mM TMACl, 2 units AmpliTaq Gold DNA polymerase (Applied Biosystems), and 120 ng genomic DNA. Reactions were kept on ice and primers were added at the last step to prevent non-specific

binding. Reactions were conducted under the following thermocycling profile: denaturation at 94°C for 10m, 40 cycles of 94°C for 30s, 48°C for 30s, 72°C for 30s, and a final extension step of 7m at 72°C. Multiplex PCR products for each individual were pooled in equal volumes and cleaned using the QIAquick PCR purification spin column kit.

Barcoded Illumina TruSeq libraries were prepared for each sample. Briefly, Illumina TruSeq adapters, including one of 24 unique barcode sequences, were added to 1 ug of purified multiplex PCR product via the automated SPRIworks Fragment Library System I kit on the Beckman Coulter SPRI-TE robot. Libraries were selectively enriched via 7 cycles of PCR using primers corresponding to the Illumina adapter sequences. These PCR products were purified, and size selected (47) if necessary (to remove adapter dimer) using solid-phase reversible immobilization on Agencourt AMPure XP beads. Cleaned products were quantified using the PicoGreen fluorometric assay. Twenty-four libraries were pooled in equimolar amounts and sequenced in a single run on the Illumina MiSeq platform. As a validation of the sequencing platform, we also prepared non-barcoded libraries for four individuals and sequenced them individually on the Life Technologies Ion PGM™ Sequencer. Identical genotypes were obtained from the two platforms.

Table S6. Primers used for SNP genotyping.

Dog Chrom. Location	Primer1	Primer1 Sequence 5'-3'	Primer2	Primer2 Sequence 5'-3'
chr1_50889498_50890025_251	chr1_A1	ATGTGGCATGCTGGCATAGG	chr1_A2	TTCCCTCTGTCTAGCTCTCC
chr1_53899271_53899734_240	chr1_B1	GTGAGAACCTAATTAGTGAG	chr1_B2	TCCTAAACAGTCTAATCTC
chr2_25668872_25669387_221	chr2_A1	ACAGCTTCTCTCTCTCCAAG	chr2_A2	CCAGATGTTAAAGAGGGAGG
chr2_26508452_26509442_800	chr2_B1	TCTTCCAGTGGGATCATGAG	chr2_B2	TAGCTCTTCTTTGCTCCACC
chr2_30600984_30601320_102	chr2_C1	CCTTGAGGAAGATAGAAGGG	chr2_C2	GGGCAGAATTAACAGCAACC
chr2_45492860_45493357_199	chr2_D1	AGACAGCAAAAGAATGAGGG	chr2_D2	CGCATTTACTTTTTCTTCTG
chr2_48968373_48968769_213	chr2_E1	GAGTAAAAATGCACAATGTC	chr2_E2	GGGCAAAAATTTAGTGTCTG
chr3_39742510_39742903_177	chr3_A1	CCATGTTCTGAAAGAATGGC	chr3_A2	TACTATCTACCTTTCCGCAG
chr3_47815836_47816501_202	chr3_B1	AATTACAGAGGCCTTCCTCC	chr3_B2	AGTACTAGCACTTCCTCTCC
chr3_58243048_58244483_180	chr3_C1	GGCAGGAAAAACAGGAAGTC	chr3_C2	ACAGACTCATTGGTTGTGGG
chr3_65655190_65655659_136	chr3_D1	TTAAGGGACTCCGGATCTTC	chr3_D2	ACTTGCTAGCCACACTTCTG
chr3_73597992_73599106_206	chr3_E1	TTCTCAGTGGATGTGGAAGG	chr3_E2	TGAATCATCAGAGCGACAGG
chr4_10250203_10250908_454	chr4_A1	TGTCTTCGCTGACTGCTTTC	chr4_A2	ACAGTGACAACAGAGATCGC
chr4_20988607_20989694_964	chr4_B1	CCAGTGCATATCAAGTCCAG	chr4_B2	CTACAGTGTTAATGCACATC
chr4_7165448_7166004_261	chr4_C1	GTCAGTGATGATGCTGTCTC	chr4_C2	AACGTCTGTAGAGAGAGCTG
chr4_90446758_90447660_251	chr4_D1	GAACAAGTTCATGGCTGAGG	chr4_D2	CCCATCTCTGGTCTTACATC
chr5_11183480_11184080_270	chr5_A1	GGGTGCTGTCAGTCATTTAC	chr5_A2	AGCCTGCTTCCCTGTTTGAC
chr5_43027949_43028719_575	chr5_B1	AAAGGATGAGTGTGGAGAAC	chr5_B2	AACACACGGTGTGATTTCCC
chr5_47410799_47412080_199	chr5_C1	CTTCTTGTCAGAACTGAGGC	chr5_C2	GTGTTCTTGTCCACCAAAGG
chr5_77384567_77384880_197	chr5_D1	AGACCTCCACTTCTGTCAGC	chr5_D2	GTGATTGCAAGTGTCTAACC
chr6_19115724_19116021_165	chr6_A1	ACACAAATACGCGCAAAGCC	chr6_A2	CACCGCTGGTTTCTAGCAAA
chr6_24367345_24367856_277	chr6_B1	AGCTACTTTCAGAATGACCC	chr6_B2	GAGGTTGCAGTAATTACAAG
chr6_35052053_35052407_111	chr6_C1	CCACTTCATAGTATTCAACG	chr6_C2	GGGTTGTTCTAAGGGCTTAC
chr6_7105063_7105836_507	chr6_D1	TGCAGGTCTGAGGTCTTTGG	chr6_D2	GGCATAATTCCTTCCATCCC
chr6_72237775_72238360_239	chr6_E1	TCAGCCTGATAAAGCAAGGG	chr6_E2	AACCATCTGAACCCAGATCA
chr7_10614285_10615278_733	chr7_A1	CATTTCCTCTGATGGTTATG	chr7_A2	CCCAGGGAAAGGCAAGTTAA
chr7_18289297_18290534_148	chr7_B1	GGAGAAAGGTGCTTCTAAGG	chr7_B2	TTTGCCCATTGGCCACTTAG
chr7_54327155_54327594_248	chr7_C1	CTCACTTATGATATTTTACC	chr7_C2	GGTGTAGGGAAGACAAGTAG
chr7_55765923_55766540_184	chr7_D1	CCCATGCTGGTACATTTAAG	chr7_D2	ACCGAAAGGTCTCTCAGTGC
chr7_82168054_82168524_228	chr7_E1	AACCAAAAGACTGTTAGTG	chr7_E2	CCTCTTTGTTATGGAAGAGT
chr8_17778688_17779134_257	chr8_A1	TTCATGAAATCCAAAATCC	chr8_A2	GAACACATTGGACTTCCAAC
chr8_22644204_22644662_271	chr8_B1	CAATATGTTAGAGGCCACGG	chr8_B2	CATTAATGGCAACTGCCCTG
chr8_29889805_29890739_170	chr8_C1	GAATAAAACTAACTAGGTG	chr8_C2	CTTTTGTGTCGATGAAGAAC
chr8_33565554_33565974_212	chr8_D1	CACAGACACATACCCATTGC	chr8_D2	GACTTGGGTGTGCTCTTTTG

chr8_40713280_40713696_124	chr8_E1	TATGGCGTTTAAGGTGGGTG	chr8_E2	GTTTGCATACTTCTTCCCAC
chr8_48989842_48990186_227	chr8_F1	AGTTTGCAGAAACTCAGCCC	chr8_F2	GGTTGACAGGTTGAGTCATC
chr8_61574018_61575090_175	chr8_G1	AGCACTGACCAAAGAAAGGC	chr8_G2	TTGCTAATGGCCAAGCTGTC
chr9_12404856_12406260_1201	chr9_A1	TAGTGGTGAGAAGGTAGGTG	chr9_A2	GAGCCTTCTTCGTGAAATGG
chr9_30483261_30483925_186	chr9_B1	ATGGTGGTCTCTAGCTCCTC	chr9_B2	TTGGCCAAAAGGGCCCTTCA
chr9_53125479_53126232_629	chr9_C1	ATGTCAAGACACGGCATGAG	chr9_C2	GATCCTCCACAGTGCTGATG
chr10_41829952_41830377_217	chr10_A1	TCCCTAGGCAAATCAGGAAG	chr10_A2	TCCAAGATGTCACCAGAAGG
chr10_57183063_57183448_283	chr10_B1	CATAGTATGGGCAGACACAG	chr10_B2	GAAGCTCCTCATTCAGCAAC
chr10_57483043_57484111_203	chr10_C1	CATTCTTTCCCTAAGGTTGG	chr10_C2	CATCCTGATTTTAGAGTCTG
chr11_13525676_13526570_652	chr11_A1	GTTGGCAATTCCTAACAACG	chr11_A2	AGTGAGGTGAAATACAGCCC
chr11_22690810_22691533_542	chr11_B1	GGCTGACATGATGTGAGTAA	chr11_B2	CTCCTTCTACAACCCTAACC
chr11_22902045_22902621_240	chr11_C1	AAACATGCGAAGTCTTTAG	chr11_C2	TGTTAGTGTTATCTTTGTC
chr11_31779416_31781057_1431	chr11_D1	TGAAAGGCTTGGAGGTAATG	chr11_D2	CCTGGTTTTTTAGAGAAAAG
chr12_17990563_17991211_478	chr12_A1	ATGCAGTTTGTCCAGTTTCC	chr12_A2	TTAACAGAGTTTCCCCCAGC
chr12_35114483_35115051_163	chr12_B1	ATGGAGGAAATGATGCTGAG	chr12_B2	TTCCACTACCATTTCTCCCC
chr12_69781564_69781812_124	chr12_C1	TGCCTGTAGATTCCCAAAGC	chr12_C2	GAGCATGTTGACTTATTGCC
chr12_8739193_8739436_130	chr12_D1	CAGCAGACTTTGATGTTGAC	chr12_D2	TTGTCTAACCCACACAGCAG
chr13_22539267_22539819_180	chr13_A1	AAGTCCAGAGAAGCTAGACC	chr13_A2	AAACTGTTCTCTAAGGTCCC
chr13_42993400_42994106_350	chr13_B1	GCCTCTATGGAGAGAGAATG	chr13_B2	AAACACATACCTCTTCAGGG
chr14_38926692_38927399_341	chr14_A1	GAGTGTTTGGAGACCCAATG	chr14_A2	GCAATTCAGAGCTGTGCCAG
chr14_50866090_50866580_128	chr14_B1	GATGTACTCACATGTCTAAG	chr14_B2	AGGAAGCCAGGCATCATTTC
chr15_14621958_14622355_233	chr15_A1	GTCCCCTCCTCTGATATATG	chr15_A2	GTAGTAGGAGAGATGAAGAC
chr15_4974250_4974837_245	chr15_B1	GTCCAATGGCCTGGGTTTC	chr15_B2	TCCATAGATGAGAAGAGAGG
chr15_53455595_53456974_398	chr15_C1	ATACGGAGGATAAGAAGGGG	chr15_C2	TTCTCATTTTCAAAGGCTC
chr15_54786264_54786700_302	chr15_D1	CTTCAAACTGACCAGTAGGC	chr15_D2	AGTCCCCAAATGTCCTATGC
chr15_65095991_65097056_865	chr15_E1	GGTAACATTAGCAAGGCATC	chr15_E2	GGCTTATTTAGAGGCACTTC
chr16_26228651_26228990_189	chr16_A1	AGGAGTGCTGAATTTGGGTG	chr16_A2	GACACAACGCTTGCTTTTTG
chr16_47188288_47189423_237	chr16_B1	TGAGAACGTTCTGCACACTG	chr16_B2	CAGCTTTTCCTCTCTTGGAC
chr16_8363958_8364659_538	chr16_C1	TATCCGCTCAGATCCTTGAC	chr16_C2	TCTTGCAAAAGCTTGCTCAG
chr17_11314060_11315363_108	chr17_A1	CTTTGAAGCAAAAAACCTG	chr17_A2	CCAAATTCTACAGGCCTTAC
chr17_27768188_27768501_146	chr17_B1	CTCCTTCCAGATGAGAAGGC	chr17_B2	TGAGCCTGTCATGTCTGTGC
chr17_34124384_34125394_723	chr17_C1	TGGCTCTTTTACTTGCAGCG	chr17_C2	TTGCCAGACAGAACGTTGAG
chr17_57823635_57824195_229	chr17_D1	GAGTCAGAATCTTAGAGTCG	chr17_D2	CAACACTTGCCTACATGTTC
chr18_21054734_21055130_190	chr18_A1	TTTTCAAGGGAGTGGCCAAG	chr18_A2	GGATCACACAACACTGCAAG
chr19_37213514_37215472_166	chr19_A1	GGCTAATAGGAGTCTGCATC	chr19_A2	TCCTCTGACAGGTAAACCAC
chr20_10090373_10091758_1192	chr20_A1	CCTGCCCATTGGAAGTGTTG	chr20_A2	TAACACAAGGTGCCTGATGG
chr20_41235473_41236271_105	chr20_B1	TAGTTTGCAAGTGCTCTCCC	chr20_B2	AACCCATGGCCAAGCAATAG
chr20_6551604_6552270_376	chr20_C1	GCCATCTCACAAAAACCCTG	chr20_C2	GAGATGACTGAAGACCCTAC
chr21_11701292_11702628_1168	chr21_A1	AAAGGGCAATAGGTGAAGGG	chr21_A2	AGACCGGTCATGTTAATTGG
chr21_21053971_21054677_232	chr21_B1	AAGTAGCTGGGTGAAGGATG	chr21_B2	CTCATGCCAAGTTCTCTCAC
chr21_38312878_38313220_113	chr21_C1	CAAAGGCAGTAGATGTGTCG	chr21_C2	TCTCTGAGATTTGGTCTCCG
chr21_40019526_40021025_1397	chr21_D1	GGAAACTGAGATCCAGAGAG	chr21_D2	CCATATCTTGCTATCAGCCC
chr22_54501499_54502435_139	chr22_A1	CTGTTCCCCTGTTCAGAATC	chr22_A2	GGATACAAATCATATGGGAC
chr22_63387455_63388086_495	chr22_B1	TCCCACCGGAATTTAAATGC	chr22_B2	TGTTAACATACCACATCTG
chr23_27014074_27014525_210	chr23_A1	GCAGAGTGCCAAAAACACAG	chr23_A2	AACACTTCCTATAAGTCTC
chr23_52113961_52114359_189	chr23_B1	TCAGAGTACACTGATGGATG	chr23_B2	CCACCCATATGACCGATAAG
chr23_5599259_5600571_209	chr23_C1	CTTCTGGTGGATGATACAGG	chr23_C2	TTTCTGCAGTGTCAGCCTTG
chr24_15117794_15118317_324	chr24_A1	ATCAGATTCCGAGCCAGGAG	chr24_A2	GACTGAGGCATCTTAGTGTC
chr24_24887459_24887991_244	chr24_B1	AGACGGAGTGCCAGGGGAT	chr24_B2	AGGCCGGCAACGTGGCCTG
chr24_30694933_30695834_734	chr24_C1	GTTTCGGGTTTTTCACAATGC	chr24_C2	AACAGAAAAGGAACAGGCCC
chr25_20004778_20005455_457	chr25_A1	CAGCCCCTTATGAGGAATAA	chr25_A2	GTGCAGTCTACGTGATGATG
chr25_34736551_34736998_163	chr25_B1	CTCCCTCCATGCATCTTATC	chr25_B2	TTGTGGGTGTAAATAAGGAG
chr25_35491377_35491792_189	chr25_C1	GATTCAAAACACACGGTAGC	chr25_C2	CAGATCGGATTTAGAGTAAG
chr25_8131430_8131796_103	chr25_D1	GAATGTACACCATGAGTCTG	chr25_D2	CACACACAACTTTTGGTGAC
chr26_10898151_10898468_152	chr26_A1	TTGGTGTGGTGAACAGGTTG	chr26_A2	CCAGCTAGAAGTCTCACAAC
chr26_32088673_32089024_237	chr26_B1	GGCCTGGATTTTGAATCAGC	chr26_B2	ATATCAGCTCAGTCAGCCAG
chr28_5196874_5197534_108	chr28_A1	CAATTCAAGGATGGACAGGG	chr28_A2	TGTCAGTGACCAGCACAGG

chr29_6721610_6722868_1088	chr29_A1	TCACTTTTGTGAGTTTGCCC	chr29_A2	ACCTAGCAGTCAGCACTTTC
chr30_11429087_11429574_243	chr30_A1	CTCCTTTCCTCAAATGATGG	chr30_A2	CAAGGTGACTGTTGATGTCG
chr30_20922778_20923235_236	chr30_B1	ACAGTTCTACCCCTGCTTAC	chr30_B2	TGGAGGCCTCATCTAGGACA
chr31_28589851_28590396_120	chr31_A1	GGCACCAGTATTTTAATATG	chr31_A2	TTCACTTTTTTTTTCAATGC
chr34_22897353_22899005_647	chr34_A1	AACCAGCCTGATTCAGGGAG	chr34_A2	GCAGGAAAAAACTCAAGCCG
chr35_11629306_11629773_351	chr35_A1	TGTGATCATGAAAGTGTGCC	chr35_A2	TGTCAACTCACTTCAGGAGC
chr35_12571349_12571703_103	chr35_B1	CCTGTGATTCTATCAATGGC	chr35_B2	AAGAAGAGGATCCAATGCTG
chr37_13376977_13377309_194	chr37_A1	CACAAGTCTGTGGAATAGGC	chr37_A2	TAAGGACAGCCTTACTTGCC
chr38_13923113_13923743_120	chr38_A1	AGTTGGTACAACCACTTCCG	chr38_A2	AATCAGTAACCACTGTTGTC
chr38_19308726_19310013_1080	chr38_B1	CTGGTAAGAAGAGAGGGAAC	chr38_B2	GCCTTGGCTCCCATTTCAAC

Data analysis and results

Reads from each individual were demultiplexed and aligned to the original sequence using the program LastZ (15). The number of reads containing each base (e.g. A, C, G, T) was counted. A particular nucleotide, for example a T, was considered present if 20 x (number of reads supporting T) > (number of reads supporting A)+(number of reads supporting C)+(number of reads supporting G). If the alleles called matched those identified when the SNP was discovered, then the genotype was called, otherwise it was flagged for visual inspection. Of the 100 SNPs discovered, 99 were found to be variable in the dataset. For four individuals that were both sequenced and genotyped, the alleles obtained were identical within these individuals. For the remaining 21 individuals where genomic sequences were available, sequences were used to determine the genotype at each SNP locus. SNPs with >10% missing data were discarded from the data set, for a total of 88 loci retained. These SNPs had 1.2% missing data on average.

To summarize the data, expected and observed heterozygosities were calculated in GenALEx 6.41 (48). Preliminary analyses that also included genotypes for four black bears indicated that heterozygosity dropped rapidly as divergence from polar bears increased, although amplification failure did not increase correspondingly. In the 100-SNP dataset, expected heterozygosity for populations of non-ABC brown bears, ABC brown bears, and polar bears ranged from 0.037 to 0.296. Within polar bears, expected heterozygosity for each population ranged between 0.253 and 0.294. In general, sample sizes were relatively small for each of the populations included in these analyses.

To graphically visualize the relationship among polar, ABC brown, and non-ABC brown bears in the larger genotyping data set, Principal Component Analysis (PCA) was performed using GenAlEx. Codominant genetic distances were calculated according to Smouse and Peakall (49), the distance matrix was converted to a standardized covariance matrix, and PCA was performed. In the principal component analysis of the 100-SNP data set, the first axis explains 71.5% of the variation, while the second and third axes explain 8.1% and 6.2%, respectively, for a cumulative total of 85.8% of the variation explained by the first three axes combined.

GenAlEx was used to conduct an analysis of molecular variance (AMOVA). In an analysis of molecular variance among polar, brown, and ABC brown bears, 40% of the variation occurred between these three groups, while 14% and 45% occurred between individuals within a group and within individuals, respectively. Pairwise F_{ST} was 0.426 between polar and non-ABC brown bears, 0.276 between ABC and other brown bears, and 0.296 between ABC brown and polar bears ($p \leq 0.01$ after 100 permutations). In the analysis of molecular variance conducted for polar bears, global F_{ST} was low compared to that found between species ($F_{ST} = 0.054$, $p \leq 0.001$). Among population variation only represented 5%, while variation among individuals within a

population explained 17% and variation within individuals explained 78% of the total variation. Pairwise F_{ST} values between all populations, and associated *p*-values, are shown in Table S7.

Table S7. Pairwise F_{ST} between four populations of polar bears. F_{ST} below the diagonal, *p*-values above.

	Chukchi	**Barents**	**S. Beaufort**	**S. Hudson**
Chukchi	-	0.001	0.241	0.002
Barents	0.041	-	0.003	0.001
S. Beaufort	0.004	0.033	-	0.003
S. Hudson	0.077	0.105	0.082	-

The Bayesian clustering program *Structure* v2.3.3 (50) was used to investigate the presence of admixture between brown and polar bears. We conducted the analysis for K=2 using the admixture ancestry model. Since these SNPs were originally discovered in polar bears, we also allowed for independent allele frequencies with a separate lambda parameter inferred for each population. An uninformative prior was used (i.e., samples were grouped without using information about collection location or phylogeny), and analyses were performed for 1,000,000 generations after an initial burn-in period of 100,000 generations. Multiple analyses were performed to investigate the potential for variation between runs. See main text for results.

The Bayesian clustering analysis was also performed on the subset of samples exclusively representing the polar bears. Ten replicate analyses were performed for each K from 1 to 6 using the admixture model with correlated allele frequencies (51). Both an uninformative and an informative prior (e.g., collection location) were utilized, but results were similar in both cases. Analyses were performed for 1,000,000 generations each, after a burn-in period of 100,000 generations. With K=2, the population from S. Hudson Bay, Canada, appears to be distinct. With K=3, both the S. Hudson Bay and Barents Sea populations are genetically distinct from one another as well as from the genetic group containing polar bears from the S. Beaufort and Chukchi seas. Further analyses with increasing K fail to find additional distinct units and only increase the noise in the data set. Results from *Structure* were input into Structure Harvester (52) to calculate an ad-hoc delta K statistic (53). Under this criterion, the grouping of K=3 has the highest support (Fig. S13). In all cases the ancient polar bear groups with modern polar bears inhabiting the Barents Sea region.

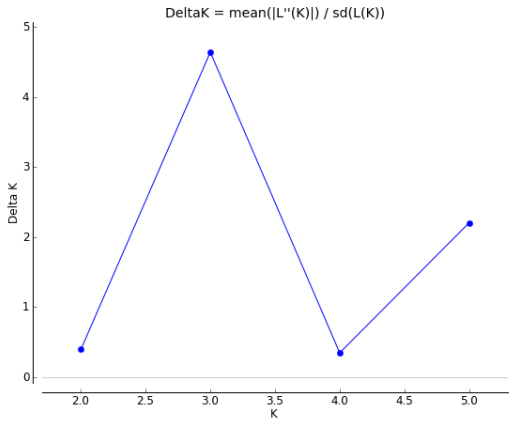


Figure S13. The delta K statistic proposed by (53) suggests highest support for the case where K = 3

Ancestral population sizes

In order to obtain a detailed population history and ancestral population sizes, we used the implementation of the pairwise sequentially Markovian coalescent model (PSMC; 37) to analyze our five deeply sequenced bear genomes: a polar bear (PB7), two ABC brown bears (ABC1 and ABC2), the non-ABC brown bear (GRZ) and the black bear (BLK), in addition to a lower-coverage polar bear (PB3). We first used SAMtools to generate the diploid consensus sequence for the various samples. The option "-C 50" was used to reduce the mapping qualities of reads with multiple mismatches, and only locations with coverage ≤ 100 were considered in these calls. The utility fq2psmcfa (provided with the PSMC software) was used to convert this diploid consensus sequence to the required input format. We then applied the PSMC model using 10 years for the generation time, which is commonly used (2, 54), and 10^{-9}/year/site for the mutation rate, which is the general rate in mammals (37, 55). Although we used the same generation time and mutation rate for all species and all loci, variations on which may influence the results, we were able to infer some general patterns that appear to correlate with key, ancient climate events on a large time scale. Bootstrap tests for each of the seven samples were run in PSMC by splitting genomic scaffolds into smaller segments using the splitfa utility and then randomly sample (with replacement) these segments applying the "-b" option in PSMC. A total of 100 bootstrap replicates were run for each genomic sample set (Fig. S14).

We tested the effect of initial sequence data coverage on the results by performing two runs of the polar bear PB7 sample based on two datasets from different pools of sequence data (PB7a was based on ~10-fold sequence coverage whereas PB7b was based on 127-fold coverage), as well as a second polar bear sample (PB3) that was sequenced to lower depth. For all three PSMC runs, 2Gb sequence data was used (Fig. S15). The deeper original sequence coverage for PB7b appeared to contain overall more heterozygote sites than PB7a, and resulted in an extended coalescence time and N_e compared to PB7a. However, although depth of sequence coverage seems to have an effect on the magnitude of effective population size at earlier time intervals, the overall pattern in population history is comparable among the three different analyses (Fig. S15).

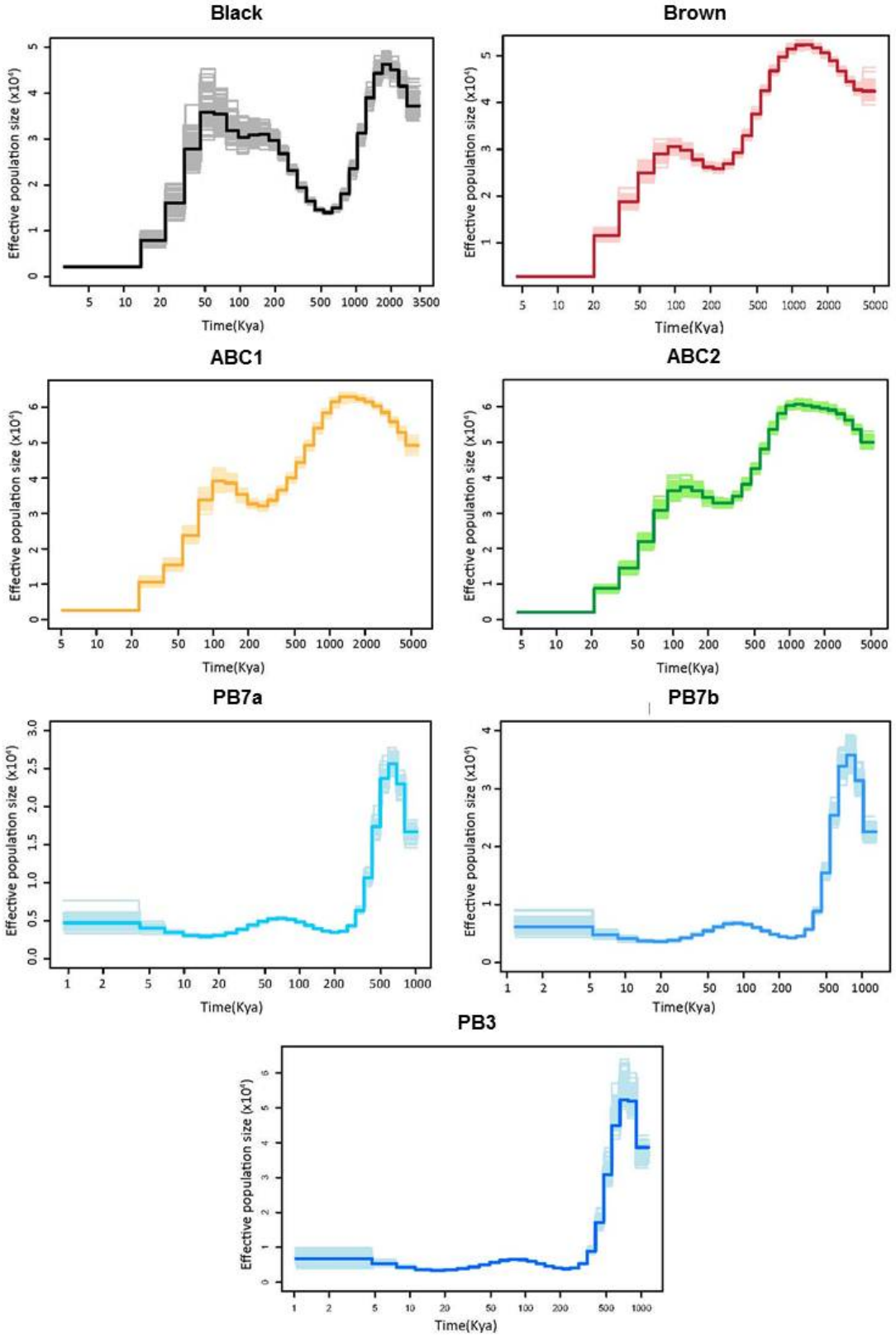


Figure S14. PSMC estimates of seven genomic datasets. The thin curves in each graph are the PSMC estimates for 100 bootstrap replicate samples from the original sequence data.

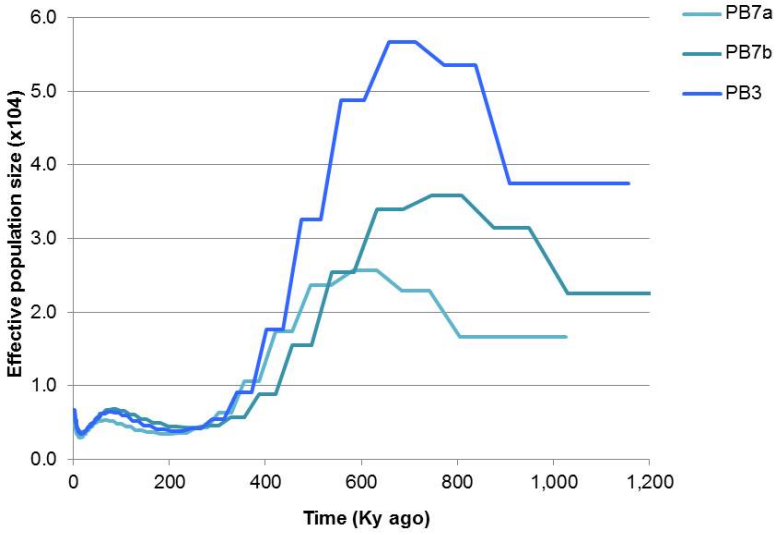


Figure S15. PSMC (37) estimates of polar bear effective population size history shown in a time span of 1.2 million years. The analysis is inferred from three sequence datasets from two Barents Sea polar bears: PB7 (PB7a and PB7b) and PB3.

Potential signals of positive selection

The search for genomic intervals possibly affected by a selective sweep, as described in the results of the main paper, was conducted using Galaxy commands as follows. We kept only the SNPs on autosomes that map to a dog chromosomal position at least 50 bases from any other SNP. That left us with 4,438,672 SNPs. We assigned an F_{ST} value (between 0.0 and 1.0) to each SNP using allele frequencies estimated from sequence coverage and with Wright's original definition. The polar bears were one population, while the other population consisted of the three sequenced brown bears. The Galaxy tool that looks for remarkable genomic regions asks the user to provide a "shift" value to be subtracted from the number assigned to each SNP, so as to make the shifted values mostly negative. We used the value 0.8. The tool then looks for genomic intervals where the sum of shifted values is maximal, i.e., where growing or shrinking the interval cannot increase the total score. The benefits of this approach include that no guess at a "window size" is made, and that the intervals can be computed very efficiently (56).

To help determine a minimum total score for retaining intervals, the Galaxy tool provides the option to randomly shuffle the shifted scores a specified number of times and use the largest obtained score as the threshold. With 100 shuffles, the largest "random" score was 2.247, above which we interpreted interval scores to have an empirical statistical significance $p < 0.01$. Note that the intervals in Table S8 have scores above 5.8. Galaxy provides additional tools to determine which dog genes intersect a set of intervals, as well as the other manipulations needed to produce Table S8. Here and for the following analysis of amino-acid polymorphisms, our gene set is the "canonical" Ensembl transcripts (57), which are often just the splice variant with longest coding region. All data and the complete set of commands (a so-called workflow) that we used are available at Galaxy.

Among high-scoring intervals are the genes *DAG1* (see main text), *KCNT2*, and *AKT1S1*. *KCNT2* is a potassium channel subfamily member associated with hyper-insulinemia (58) and *AKT1S1* has been shown to be involved in physiological insulin action (59). The fat-storing phase of the hibernation cycle is characterized by hyper-insulinemia and insulin-resistance in other hibernating mammals (60), and it is possible these genes are expressed as well in hibernating bears, although their function in polar bears are unknown.

Table S8. Highest-scoring intervals of putative selective sweeps. Intervals scoring over 5.8 in the scheme described above (the 58 highest) are shown, along with any "canonical" dog transcripts (Ensembl annotation) that they intersect.

Dog chr.	Start	End	Score	Transcript	Gene
chr13	44724323	44862421	17.5165	.	.
chr37	22115064	22207876	16.2495	ENSCAFT00000022535	XM_545629.2
chr20	42685068	42751070	12.0779	ENSCAFT00000017810	DAG1_CANFA
chr20	42685068	42751070	12.0779	ENSCAFT00000037806	ENSCAFG00000024492
chr38	5983540	6077001	11.2467	ENSCAFT00000016600	KCNT2
chr6	33430482	33530272	11.0451	.	.
chr11	47411709	47496828	11.0386	.	.
chr1	115597086	115749261	10.9234	ENSCAFT00000008119	AXL
chr1	115597086	115749261	10.9234	ENSCAFT00000008087	HNRNPUL1
chr1	115597086	115749261	10.9234	ENSCAFT00000037877	ENSCAFG00000024538
chr1	115597086	115749261	10.9234	ENSCAFT00000008128	CYP2S1
chr1	115597086	115749261	10.9234	ENSCAFT00000008133	CP2BB_CANFA
chr14	42071667	42118353	10.7971	.	.
chr38	7878499	7993446	10.7787	.	.
chr14	60664531	60761367	10.4379	.	.
chr7	52186416	52263594	9.7459	.	.
chr6	30952791	30984255	8.8333	ENSCAFT00000027269	MRP1_CANFA
chr8	53733070	53798705	8.6680	.	.
chr4	50186627	50223897	8.6168	.	.
chr28	13398703	13443835	8.5097	ENSCAFT00000014303	Q59I58_CANFA
chr8	61990618	62023554	8.2580	.	.
chr1	109539881	109600448	8.1414	ENSCAFT00000005539	IL4I1
chr1	109539881	109600448	8.1414	ENSCAFT00000005517	NUP62
chr1	109539881	109600448	8.1414	ENSCAFT00000005586	TBC1D17
chr1	109539881	109600448	8.1414	ENSCAFT00000005591	AKT1S1
chr1	109539881	109600448	8.1414	ENSCAFT00000005610	PNKP
chr1	109539881	109600448	8.1414	ENSCAFT00000005622	PTOV1
chr11	56793660	56959810	8.1144	ENSCAFT00000038713	ZBTB5
chr11	56793660	56959810	8.1144	ENSCAFT00000003730	POLR1E
chr11	56793660	56959810	8.1144	ENSCAFT00000003736	FBXO10
chr3	60100015	60185036	8.0963	ENSCAFT00000022144	FAH
chr34	17037850	17086568	8.0721	.	.
chr32	19395745	19451037	7.9943	ENSCAFT00000015920	XM_544985.2
chr12	29544545	29598480	7.7175	.	.
chr17	20772669	20848374	7.6462	.	.
chr1	93064015	93132734	7.5625	.	.
chr13	21062308	21080006	7.4909	.	.
chr27	31131133	31171160	7.4660	.	.
chr6	9662108	9704863	7.4143	ENSCAFT00000020676	TBL2
chr6	9662108	9704863	7.4143	ENSCAFT00000036670	BCL7B
chr17	15456585	15517925	7.3508	.	.
chr2	62878937	63072921	7.3278	.	.
chr8	53118819	53174039	7.2767	ENSCAFT00000027222	GSTZ1
chr8	53118819	53174039	7.2767	ENSCAFT00000027186	NGB_CANFA
chr8	53118819	53174039	7.2767	ENSCAFT00000027201	POMT2
chr10	6313923	6356938	7.2339	.	.
chr2	37612201	37659856	7.2223	ENSCAFT00000037926	ENSCAFG00000024585
chr2	37612201	37659856	7.2223	ENSCAFT00000036390	ENSCAFG00000023602
chr2	37612201	37659856	7.2223	ENSCAFT00000009187	DNAJC18
chr2	37612201	37659856	7.2223	ENSCAFT00000009197	TMEM173

chr20	42409383	42451017	7.1343	ENSCAFT00000017494	XM_533824.2
chr20	42409383	42451017	7.1343	ENSCAFT00000017607	TRAIP
chr20	42409383	42451017	7.1343	ENSCAFT00000017572	CAMKV
chr13	12948812	12981695	6.9961	ENSCAFT00000038299	ENSCAFG00000024786
chr36	5837886	5865349	6.9387	.	.
chr17	4236103	4327172	6.8634	ENSCAFT00000005892	ENSCAFG00000003669
chr32	20170985	20226017	6.7429	.	.
chr34	9981896	10043896	6.5275	ENSCAFT00000016392	PAPD7
chr38	21165269	21199492	6.4635	.	.
chr17	9866920	9898356	6.4539	ENSCAFT00000005431	TAF1B
chr10	24632379	24692004	6.4455	ENSCAFT00000001372	PARVB
chr1	117413377	117464708	6.2961	ENSCAFT00000009509	RYR1
chr25	30894517	30927058	6.2338	ENSCAFT00000012950	MSRA
chr20	49653782	49726505	6.2118	ENSCAFT00000025271	ENSCAFG00000015931
chr20	49653782	49726505	6.2118	ENSCAFT00000025291	CYP4F22
chr8	26854704	26886221	6.2000	ENSCAFT00000022485	MDGA2
chr18	13447808	13468886	6.2000	.	.
chr1	63621968	63698324	6.1591	.	.
chr29	42378009	42437593	6.1450	ENSCAFT00000014852	DPY19L4
chr8	56687464	56729884	6.0779	ENSCAFT00000027418	ENSCAFG00000017308
chr36	4522959	4566372	6.0671	.	.
chr34	34699284	34752121	6.0578	.	.
chr9	7970228	8012143	6.0344	ENSCAFT00000007782	LLGL2
chr9	7970228	8012143	6.0344	ENSCAFT00000007733	TSEN54
chr23	14498555	14544410	6.0280	.	.
chr36	32156366	32191387	5.9773	.	.
chr4	22051845	22159855	5.9101	ENSCAFT00000021088	ENSCAFG00000013277
chr27	29493379	29549555	5.8998	ENSCAFT00000019663	SLCO1C1
chr7	20985845	21058393	5.8216	ENSCAFT00000021393	FAM129A
chr25	45769165	45830806	5.8198	ENSCAFT00000017088	CAB39
chr35	27265512	27317333	5.8133	ENSCAFT00000017271	XM_545403.2
chr35	27265512	27317333	5.8133	ENSCAFT00000017299	BTN1A1

Investigation of amino acid polymorphisms among bear species

Among our discovered SNPs, 79,496 were traced to coding regions of putatively orthologous genes known from the dog genome (canFam2); 26,001 out of these SNPs were found to result in amino acid replacements (SAPs) (Fig. S16), and 7,014 out of these SAPs were found to be possibly or probably “damaging” (see below) by Polyphen-2 (61) (Fig. S17). Additionally, 122 resulted in truncated proteins. These 7,136 SAPs were located in 4,314 known genes, from which 391 high quality SAPs were found to be fixed in polar bears. We defined high quality SAPs as those being called with at least 2 sequences, and as many as 60 sequences, having a quality value higher than 48. To determine the possible role of these mutations in the evolution of polar bears, we traced them into metabolic pathways, and analyzed their possible role in specific adaptations of interest (Table S9).

Table S9. Genes with high quality "damaging" mutations putatively associated with polar bear adaptations.

Adaptation	Genes
Fatty acid metabolism (62, 63)	*ACSL6* (64), *ATRN* (65), *CPT1B* (66), *HSD17B12* (67), *IL1F5* (68), *PLTP* (69), *RABEP2* (70), *TBC1D1* (71, 72), *FTO* (73, 74)
Hibernation (62, 75)	*CPS1* (76, 77), *CPT1B* (78), *CRY2* (79, 80), *URIC* (81)

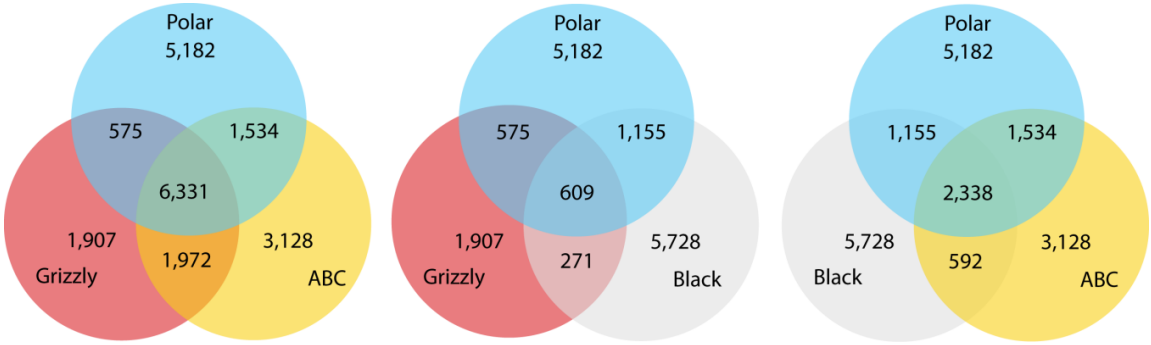


Figure S16. Number of high quality alleles in polymorphic loci resulting in amino acid replacement, and shared by different bear clades. High quality SAPs were called in regions with at least 2 and as many as 60 aligned sequences, having a quality value higher than 48.

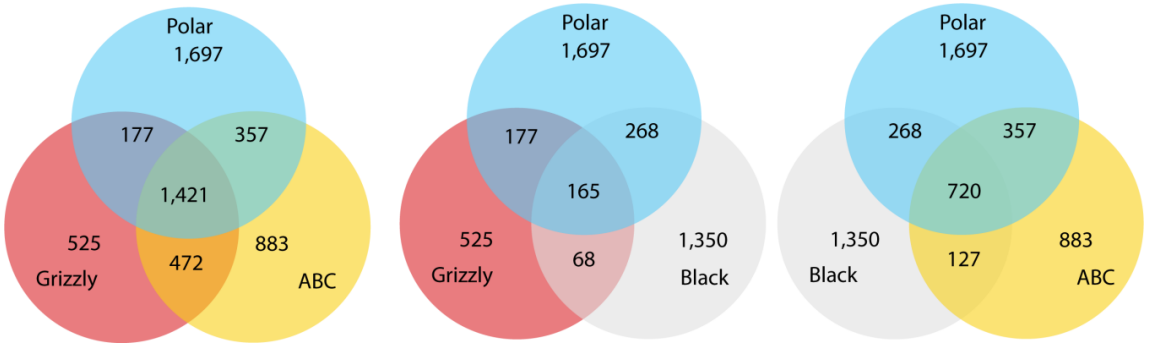


Figure S17. Number of high quality alleles in polymorphic loci resulting in "damaging" aminoacid replacement, and shared by different bear clades. High quality SAPs were called in regions with at least 2 and as many as 60 aligned sequences, having a quality value higher than 48.

We predicted the functional effect of each amino acid substitution, and then determined its global impact on metabolism. The functional effect of each SAP was predicted with PolyPhen 2 (61), which gives each SAP a classification of "possibly damaging", "probably damaging", or "benign". Using those SAPs classified as "damaging" (i.e., "possibly damaging" and "probably

damaging"), we ranked a set of KEGG pathways based on three different metrics: percentage of genes affected, change in the number of paths, and change in the length of paths. The goal of these measures was to estimate the perturbation produced by modifying SAPs in the production of final metabolites in the pathway (Table S10 and S11). This perturbation was measured using pathways as directed graphs.

Table S10. Number of genes with high quality "damaging" mutations in known pathways (the top 20 are shown).

Pathway	Number of genes with "damaging" mutations	Percentage of total genes in pathway with "damaging" mutations
Caffeine metabolism	1	0.25
Valine, leucine and isoleucine biosynthesis	2	0.2
Renin-angiotensin system	2	0.12
Sulfur relay system	1	0.091
Folate biosynthesis	1	0.091
Riboflavin metabolism	1	0.083
Carbohydrate digestion and absorption	3	0.081
Galactose metabolism	2	0.077
ECM-receptor interaction	5	0.063
Adipocytokine signaling pathway	4	0.061
Steroid hormone biosynthesis	2	0.061
Taste transduction	2	0.056
Arginine and proline metabolism	3	0.054
Biosynthesis of unsaturated fatty acids	1	0.053
Glycosaminoglycan degradation	1	0.053
Amoebiasis	5	0.047
Aminoacyl-tRNA biosynthesis	3	0.045
Fatty acid metabolism	2	0.045
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	1	0.045
Amino sugar and nucleotide sugar metabolism	2	0.043

Table S11. Change in the number and mean length of paths between initial reactants and final products in the metabolic pathways of bears. Change in pathways is expected to result from the modified enzymatic action of the mutated gene products. Results were calculated using high quality SAPs.

Metabolic pathway	Change in mean length of paths	Mean length of paths in black and brown bears[1]	Mean length of paths in polar bear[2]	Change in number of paths	Number of paths in black and brown bears[1]	Number of paths in polar bear[2]
Steroid hormone biosynthesis	1.54	6.31	4.77	6	32	26
Glycine, serine and threonine metabolism	3.64	8.44	4.8	4	9	5
Purine metabolism	1.6	18.5	16.9	3	12	9
Glycosaminoglycan degradation	0.4	4	3.6	1	6	5
Caffeine metabolism	-0.3	3.2	3.5	1	5	4
Folate biosynthesis	-3	7	10	1	2	1

Reference pathway [1]including the genes and [2] excluding the genes with predicted modifying mutations.

Polar bears are adapted to the extreme conditions of the Arctic. These adaptations include mechanisms that allow them to maintain homeostasis under low temperatures, and extend from evident morphological features to more subtle physiological traits (82). It has been suggested that all three bear species, black, brown and polar bear, share four biochemical and physiological stages: hibernation, walking hibernation, normal activity and hyperphagia (62). However, the physiological states in polar bears are quite dynamic; even in summer, the animals may stay in a biochemical state of hibernation while still walking (62, 75). In addition, only pregnant females must den in order to provide a suitable environment for neonates (75). With the goal of finding molecular signatures of phenotypic adaptations in polar bears, we investigated our sequenced genomes, representing major bear lineages (black, non-ABC brown, ABC brown, and polar bear), for fixed genes that might be involved in hibernation induction and fatty acid metabolism (Tables S9 and S12).

Table S12. Alleles fixed in genes of interest for each animal clade analyzed. These alleles represent "damaging" mutations in polar bears. Alleles in uppercase letters indicate a high quality SAP call (present in at least 2 and as many as 60 sequences, having a quality value higher than 48).

Gene	Position (nt/aa)	Dog (nt/aa)	Black bear (nt/aa)	ABC brown bear (nt/aa)	non-ABC brown bear (nt/aa)	Polar bear (nt/aa)
ATRN	605	C/Q	C/R	C/R	c/r	T/W
CPT1B	625	T/M	c/t	c/t	C/T	T/M
CRY2	207	G/D	G/D	G/D	g/d	A/N
FTO	42	G/V	G/V	g/v	G/V	A/M
HSD17B12	266	A/Q	a/q	A/Q	A/Q	G/R
IL1F5	142	G/S	G/S	G/S	G/S	T/I
PLTP	167	G/A	G/A	G/A	G/A	A/T
RABEP2	26	A/Q	t/h	t/h	T/H	A/Q
TBC1D1	769	G/A	g/a	G/A	G/A	A/T
URIC	84	C/I	C/I	C/I	C/I	G/M
ACSL6	288	G/V	G/V	G/V	G/V	A/M
CPS1	131	A/D	A/D	A/D	A/D	T/V
LAH2	275	A/Q	C/P	C-A/P-Q	C-A/P-Q	A/Q
TRPM1	423	G/R	G/R	G/R	G/R	G-T/R-I

Alleles possibly associated with fatty-acid metabolism and hibernation

Among the genes selected, we found a fixed modifying variant in polar bear of the fat mass and obesity-associated protein, encoded by *FTO*. Mutations in *FTO* have been previously correlated with severe obesity in humans (73, 74). In mice, *fto* -/- mutation results in postnatal growth retardation, a significant reduction in adipose tissue and lean body mass, and hyperphagia (74). This phenotype might occur as a consequence of an increased expenditure of energy, produced by a modified hypothalamic development and a concomitantly altered activity of the melanocortin system, and/or by an altered cycling of fatty acids and triglycerides (74). Polar bears are known to be energetically inefficient walkers (83, 84). We speculate that the fixed

variant in polar bears has a modified activity in comparison to brown and black bears, possibly resulting in higher energy expenditure in polar bears, although their basal metabolic rate may not differ from other bears.

In obese humans, increased activity of *PLTP* has been associated with hypertriglyceridemia, small, dense HDL particles, and low HDL-C levels (69). The product of this gene (phospholipid transfer protein) is involved in the synthesis of HDL from VLDL and LDL, and also in its catabolism (69, 85). It has been hypothesized that the reduction of *PLTP* activity and an increase of HDL particle size are important to convert the lipoprotein profiles of obese individuals into those of lean ones (69). In previous experiments, mice with the *PLTP* gene knockdown show a lower plasma concentration of HDL phospholipids, cholesterol, and apolipoprotein AI under a chow diet. Such mice presented a higher concentration of VLDL and LDL phospholipids, free cholesterol and cholesteryl ester under a high-fat diet (85). In wild polar bears, HDL constitutes about 80% of total phospholipids in plasma (86). In contrast, HDL of denning black bears, captive black bears and seals (the main source of fat for polar bears) contain about 38% (87), 62-74% (88), and 71% (89) of total phospholipids, respectively. We speculate that the mutation found in polar bears modifies the efficiency/activity of PLTP. This might result in a higher transference of phospholipids from VLDL and LDL to HDL, and/or a modified stability of HDL particles in comparison to brown and black bears. In turn, this might affect the availability of triglycerides, and weight gain/loss during hyperphagia and hibernation stages.

The *TBC1D1* gene encodes a Rab-GTPase-activating-related protein that regulates the traffic of *glut4* (71, 90, 91). *GLUT4* is mainly expressed in adipose tissue and muscles. The protein contributes to glucose uptake in insulin-responsive tissues, and to whole-body glucose homeostasis (90). Mutations in this gene have been associated with an increased risk for obesity in humans, and also with protection from diet-induced obesity in mice (71, 92). It has been hypothesized that leanness is produced as a result of unexpected “decreased respiratory quotient, increased fatty acid oxidation, and reduced glucose uptake in isolated skeletal muscle” (92). Also, it is possible that the altered expression of this gene is linked to a hypermetabolic state trough its interaction with *tardbp* (92). We speculate that this gene might be related to the glucose balance and constant physiological activity of polar bears, and perhaps to a modified respiratory quotient. This might represent an adaptation, as fewer calories would be lost in respiration.

Interestingly, two genes involved in acyl-CoA synthesis and catabolism were found to have modifying fixed variants in polar bear: *CPT1B* and *ACSL6*. Muscle carnitine palmitoyltransferase-1B (encoded by *CPT1B*) is involved in the regulation of skeletal muscle mitochondrial beta-oxidation of long-chain fatty acids (66, 78). Mutations in this gene have been associated with differences in fat accumulation of skeletal muscle in humans (66), and is overexpressed in hibernating arctic ground squirrels (78). On the other hand, the gene *ACSL6* encodes long-chain acyl-CoA synthetase 6. This enzyme is involved in the catabolism of long-chain fatty acids, de novo lipid synthesis, and remodeling of membranes (64). It has been suggested that storage of unsaturated and shorter chain fatty acids is important for thermal insolation in marine mammals due to lower melting points than saturated and longer chain lipids (93). However, this pattern has not been clearly found in polar bears, indicating that their blubber is not as efficient at insulation compared to seals (93). We speculate that the fixed variants in genes of polar bears might be related to the fatty acid profile of the animal.

The 17β-Hydroxysteroid dehydrogenase 12 (encoded by *HSD17B12*) is involved in the interconversion between 17-keto and 17β-hydroxysteroids, between estrone and estradiol, and

additionally in the elongation of very long chain fatty acids (VLCFA) (94). Previous experiments showed that *hsd17b12* -/- mutants are nonviable, probably due to an imbalance in the production of arachnoic acid (94). We speculate that the polar bear ortholog may have decreased activity/efficiency in comparison to brown bears, which in turn might decrease the proportion of VLCFA. Previous studies have reported an increase in arachnoic acid during hibernation in other mammals (95, 96), and it is possible that this pattern supports previous observations concerning the dynamic nature of hibernation in polar bears (62). On the other hand, this enzyme might play other roles in hibernation through its action on testosterone metabolism. It has been found that an increased concentration of testosterone can inhibit torpor in squirrels (97).

Other genes with "damaging" SAPs fixed in polar bears and possibly related to the characteristics of adipose tissue include *IL1F5*, *ATRN*, and *RABEP2*. The function of *RABEP2* is unknown. Nonetheless, copy number variations of a genomic region including this gene have been associated with severe obesity in humans (70). *IL1F5* was found to be a molecular marker that predicted response to different obesity treatments (68). In that study, a cluster of 12 genes (including *IL1F5*) identified two subtypes of obesity. Individuals of different subtypes showed different body weight losses, visceral adipose tissue masses, and decreased size of very large fat cells (68). Finally, polymorphisms in the attractin (*ATRN*) gene of pigs have been associated with a differential average weight daily gain, and back fat thickness (65).

The *URIC* gene encodes the enzyme urate oxidase, which catalyzes the oxidation of uric acid to allantoin in most mammals (except some primates) (98). Mutations in this gene have been associated with hyperuricemia in humans (98). In mice, it has been found that decrease of uricase-mediate degradation of urate accounts for its preservation in plasma (99). It has been suggested that urate might confer certain physiological benefits, including maintenance of blood pressure in salt-poor environments (99). Indeed, urate concentration in the brain is significantly increased during the hibernation of Artic ground squirrels (100). Interestingly, a previous study found different urea concentrations in plasma of black and polar bears (62). We hypothesize that *uric* fixed variants may be correlated with this difference, perhaps due to increases of enzymatic activity in polar bear.

CRY2 encodes the cryptochrome 2 protein, and is a central gene in the training of circadian rhythm (78, 101). A significantly increased expression of this gene has been observed during the torpor of Artic ground squirrels (78). This gene has been associated with delayed sleep phase syndrome in humans (101), and is responsible for altered activity and thermoregulation under different light intensities in *cry2* -/- mice (102). We hypothesize that this variant might be related to physiological and biochemical differences found between polar and brown/black bears, especially those related with the synchronization of the different stages through a year (62).

Alleles associated with coat color and density

In order to determine polymorphisms that might be associated with differences in hair and skin pigmentation among the different bears studied, we looked for SAPs in genes previously associated with skin and hair pigmentation. We found amino acid polymorphisms in eight out of 24 genes studied (independent of their quality, Fig. S18 and Table S13).

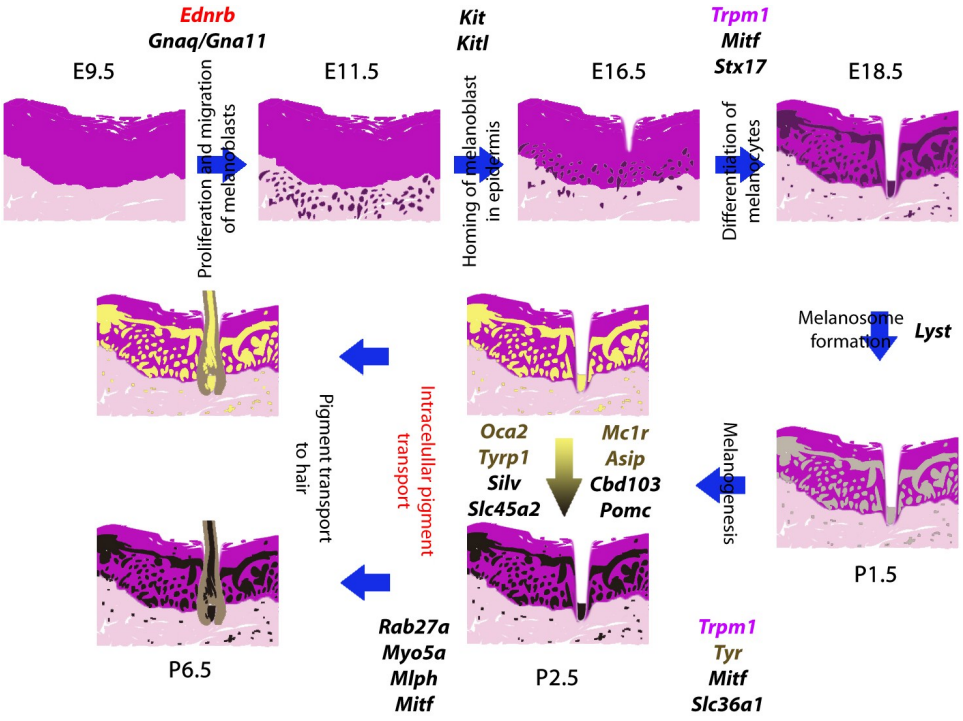


Figure S18. Model representing the genes and stages involved in skin and hair pigmentation. Stages correspond to those previously described for mice (103-105).

Only *EDNRB* presented a (low quality) variant that was fixed in polar bears. *EDNRB* encodes a G-protein-coupled receptor expressed during neural crest development, where it is required for the migration of melanoblasts in dermis and enteric neuroblasts (105). Previous experiments have found that some mutations in this gene result in different color patterns in different mammals (103, 105), while others have been associated with the Hirschsprung disease and the Waardenburg syndrome in humans (which also result in altered pigmentation) (106). It is possible that the *EDNRB* variant fixed in polar bears might be involved with the black pigmentation of their skin. A delay in the dermal migration of melanoblasts could result in their consequent late migration to the epidermis. This late migration might result in a higher accumulation of melanocytes in skin (Fig. S19). However, migration of melanoblasts to the hair shaft may be disrupted due to the anatomy of polar bear hair, resulting in the lack of hair pigmentation in polar bears.

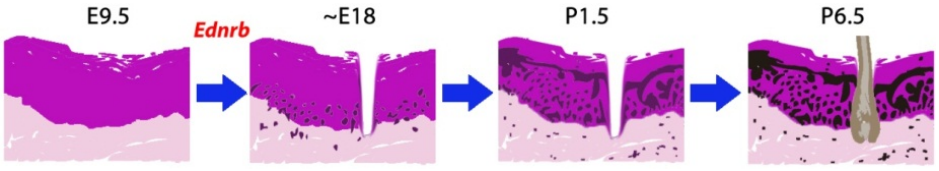


Figure S19. Hypothetical skin and hair pigmentation process in polar bears. Stages correspond to those previously described for mice (103-105).

Variants that differentiated black bears from the other ursids studied were found in genes involved in the differentiation of pheomelanine and eumelanine. For example, the "damaging" mutation found in *MC1R* (V313A) is in proximity to a previous mutation found to be responsible for the Kermode bear color (Y298C, Fig. S20) (107). The product of *MC1R* is activated by alpha-MSH, leading to the production of eumelanine (black pigment). The product of *ASIP* is an

antagonist of alpha-MSH and competes for binding to MC1R, leading to the production of pheomelanine (red/yellow pigment) (104). In the same way, the protein encoded by *OCA2* is thought to regulate eumelanine synthesis by affecting pH in melanosomes (108). Similarly, *TYRP1* and *DCT* are involved in eumelanine synthesis. The product of *DCT* catalyzes the synthesis of DHICA from dopachrome, while the exact role of *TYRP1* remains elusive (104, 109). All these genes have fixed variants that are different between black bears and the other bears studied (Table S13). Previous studies have shown how the interactions among these genes determine coat color patterns in mice, horses and dogs (108, 110-112).

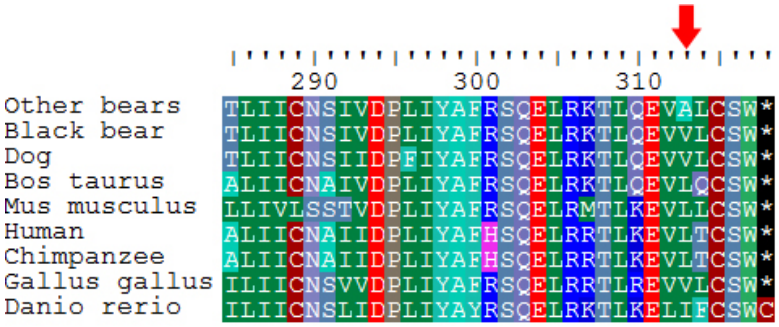


Figure S20. Sequence alignment of *MC1R* gene orthologs in different species. The charge of amino acid 313 is well conserved in the eight species analyzed. The polarity change in the amino acid of "other bears" might result in modified enzyme activity and later in a different eumelanin/pheomelanin ratio, which could result in a different skin and hair color.

We also found an interesting set of nine SAPs in the *TRPM1* (*MLSN1*) gene of the studied ursids (Table S13). The fixed alleles in polar bears are also present in brown bears; nevertheless, the latter have additional variants shared with ABC and/or non-ABC brown bears. The high number of SAPs found in this gene is unusual. This might be due to gene duplication in one or more of the bears studied. This holds particularly for non-ABC brown bears, which presented polymorphisms for most of the alleles in this gene. *TRPM1* might be involved in intracellular responses to phosphatidylinositol and protein kinase C by mediating Ca^{2+} entrance into hyperpolarized cells, and might be involved in the survival of melanoblasts (104, 113). Different coat color patterns in Appaloosa horses have been associated with mutations in this gene. This includes the so-called "fewspot" phenotype, which has a completely white coat in contrast to the brown color of wild horses (113). Recently, it has been shown that this gene is associated with the graying of hair in humans (114). In this regard, a previous study found that an independent effect on hair and skin color is apparent under conditions where survival of epidermal melanoblasts is compromised (103). The same study also suggests that dermis permits a permanent increase in melanoblasts while epidermis does not.

Another interesting gene is *LAH2* (*LIPH*), which encodes Lipase H. A mutation in this gene has been associated with hair follicle disintegration, hypotrichosis and sparse hair in humans (115). In rabbits, the deletion of *LAH2* results in the so called "rex" phenotype, where brown fur is the result of the atrophy of one kind of hair follicle and the proper development of the other two, all of which are functional in wild white rabbits (116). Mean density of hairs is higher in polar bears than in brown bears (117), and is more even than in other ursids (118). We hypothesize that the mutations in this gene might be related to the differences observed in pelage density, and perhaps to fur color as well. If so, the presence in ABC and non-ABC brown bears of alleles that are fixed in polar and black bears might indicate that the allele present in polar bears is recessive. This has been previously suggested for other fur related genes in Kermode bears (107).

Table S13. Genes related to coat color with different alleles in the bear clades studied. Uppercase letters indicate a high quality call (at least two and as many as 60 sequences, having a quality value higher than 48).

Gene	Position (aa)	Dog (nt/aa)	Black bear (nt/aa)	ABC brown bear (nt/aa)	non-ABC brown bear (nt/aa)	Polar bear (nt/aa)	Damaging mutation	Species (mutation) /Skin-hair phenotype
ASIP	33	G/R	A/K	G/R	G/R	G/R		Cow (316ter) / Braunvieh albino(119). Cat (G301R, G227W) / Siamese, Burmese(120). Horse (2Exon*del*) / Black coat color (110).
DCT	51	A/I	a/i	G/V	G/V	G/V	*	Mouse (R194Q, G486R) / Slaty, Slaty light (109).
DCT	55	A/Q	a/h	G/R	G/R	G/R	*	
DCT	144	A/E	C/A	A/E	A/E	A/E		
EDNRB	345	A/H	a/h	G/R	g/r	G/R	*	Mouse (*EDNRB* -/-) / White skin/coat color (121).
EDNRB	537	C/S	C/S	C/S	C/S	t/f	*	
MC1R	313	T/V	T/V	C/A	C/A	C/A		Black bear (Y298C) / White coat color of the "Kermode" bear (107). Dog (R306X) / Red, yellow, gold, apricot, and orange coat colors (112). Horse (C901T) / Chestnut color (110). Human (R151C, R160W, D294H) / Red hair and pale skin (122).
OCA2	55	A/M	G/V	A/M	A/M	A/M		Human (rs12913832) / Eye, skin, and hair color (123).
OCA2	142	C/Q	C/Q	C-G/Q-E	C-G/Q-E	C/Q		
OCA2	190	A/E	A/E	A-T/E-D	a-t/e-d	A/E		
TRPM1	1174	G/D	G/D	A/N	A/N	A/N		Appaloosa horse (Multiple) / Congenital Stationary Night Blindness and Coat spotting patterns (113, 124).
TRPM1	1284	C/L	A/I	A/I	A-C/I-L	C/L		
TRPM1	1455	A/S	a/s	A/S	A-C/S-R	C/R	*	
TRPM1	1457	A/I	A/I	A/I	A-G /I-V	G/V		
TRPM1	1509	G/E	A/K	G/E	G/E	G/E		
TRPM1	1527	C/A	C/A	T/V	C-T/A-V	C/A		
TRPM1	1595	A/D	A/D	C/A	C/A	C/A	*	
TRPM1	1600	A/T	A/T	A/T	A-T/T-S	T/S		
TRPM1	1602	A/T	G/A	A/T	A-G/T-A	G/A		
TYR	278	G/R	G/R	A/H	A-G/H-R	G/R	*	Cat (975*del*C) / albino (125).
TYRP1	10	G/G	G/G	A-G/E-G	A/E	A/E	*	Dog (331ter, 345*del*P, S41C) / Brown and white coat color (112). Horse (C189T-A1188G) / Different coat color -No obvious association- (110). Soay sheep (C290F) / White coat color and weight (126, 127).
TYRP1	37	G/R	C/T	G/R	G/R	G/R		
TYRP1	307	G/E	G/E	A-G/K-E	A/K	A/K		

References

1. Jackson JV, Talbot SL, & Farley S (2008) Genetic characterization of Kenai brown bears (*Ursus arctos*): microsatellite and mitochondrial DNA control region variation in brown bears of the Kenai Peninsula, south central Alaska. *Canadian J. Zool.* 86:756-764.
2. Cronin MA, Amstrup SC, Talbot SL, Sage GK, & Amstrup KS (2009) Genetic variation, relatedness, and effective population size of polar bears (*Ursus maritimus*) in the southern Beaufort Sea, Alaska. *J. Heredity* 100:681-690.
3. Longmire JL*, et al.* (1988) Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family Falconidae. *Genomics* 2:14-24.
4. Stirling I, Spencer C, & Andriashek D (1989) Immobilization of polar bears (*Ursus maritimus*) with Telazol in the Canadian Artic. *J. Wildlife Diseases* 25:159-168.
5. Ingólfsson Ó & Wiig Ø (2008) Late Pleistocene fossil find in Svalbard: the oldest remains of a polar bear (*Ursus maritimus* Phipps, 1744) ever discovered. *Polar Res.* 28:455-462.
6. Andersson T, Forman SL, Ingólfsson Ó, & Manley WF (1999) Late Quaternary environmental history of central Prins Karls Forland, western Svalbard. *Boreas* 28:292-307.
7. Bergsten H, Andersson T, & Ingólfsson Ó (1998) Foraminiferal stratigraphy of raised marine deposits representing isotope stage 5, Prins Karls Forland, western Svalbard. *Polar Res.* 17:81-91.
8. Alexanderson H, Ingólfsson Ó, Murray A, & Dudek J (In Press) An interglacial polar bear and an early Weichselian glaciation at Poolepynten, western Svalbard. *Boreas*.
9. Lindqvist C*, et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proc. Natl. Acad. Sci. U.S.A* 107:5053-5057.
10. Li R*, et al.* (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265-272.
11. Li R*, et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463:311-317.
12. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
13. Florea L, Hartzell G, Zhang Z, Rubin GM, & Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* 8:967-974.
14. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.
15. Harris RS (2007) Improved pairwise alignment of genomic DNA. Ph.D. dissertation (Pennsylvania State University).
16. Ratan A (2009) Assembly algorithms for next-generation sequence data. Ph.D dissertation (Pennsylvania State University).
17. Niu B, Fu L, Sun S, & Li W (2010) Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* 11:187.
18. Stamatakis A, Hoover P, & Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web-Servers. *Syst. Biol.* 75:758-771.
19. Drummond A & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
20. Kurtén B (1968) *The Pleistocene Mammals of Europe* (Aldine, Chicago).

21. Wayne RK, Van Valkenburgh B, & O'Brien SJ (1991) Molecular distance and divergence time in carnivores and primates. *Mol. Biol. Evol.* 8:297-319.
22. Goldman D, Giri PR, & O'Brien SJ (1989) Molecular genetic-distance estimates among the ursidae as indicated by one- and two-dimensional protein electrophoresis. *Evolution* 43:282-295.
23. Yu L, Li QW, Ryder OA, & Zhang YP (2004) Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Mol. Phyl. Evol.* 32:480-494.
24. Edwards CJ*, et al.* (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* 21:1251-1258.
25. Hailer F*, et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* 336:344-347.
26. Talbot SL & Shields GF (1996) Phylogeography of brown bears (*Ursus arctos*) of Alaska and paraphyly within the Ursidae. *Mol. Phyl. Evol.* 5:477-494.
27. Yu L, Li Y-W, Ryder O, & Zhang Y (2007) Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evol. Biol.* 7:198-209.
28. Bon C*, et al.* (2008) Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proc. Nat. Acad. Sci. U.S.A.* 105:17447-17452.
29. Krause J*, et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.* 8:220.
30. Blankenberg D*, et al.* (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19:Unit 19.10.11-21.
31. Giardine B*, et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15:1451-1455.
32. Goecks J, Nekrutenko A, Taylor J, & The Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in life sciences. *Genome Biol.* 11:R86.
33. Woolley S, Johnson J, Smith MJ, Crandall KA, & McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671-672.
34. McClellan DA*, et al.* (2005) Physiochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol Biol Evol* 22:437-455.
35. Posada D (2008) jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25:1253-1256.
36. Mailund T, Dutheil JY, Hobolth A, Lunter G, & Schierup MH (2011) Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS Genet.* 7:e1001319.
37. Li H & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493-496.
38. Patterson N PA, & Reich D (2006) Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
39. Paschou P*, et al.* (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet.* 3:1672-1686.
40. Bryc K*, et al.* (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Nat. Acad. Sci. U.S.A.* 107:786-791.
41. Wright S (1951) The genetical structure of populations. *Ann. Eugen.* 15:323-354.

42. Weir B & Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Genetics* 155:945-959.
43. Tang H*, et al.* (2007) Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* 81:626-633.
44. Brocker C*, et al.* (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J. Biol. Chem.* 285:18452-18463.
45. Kumar S, Dhingra A, & Daniell H (2004) Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots, and leaves confers enhanced salt tolerance. *Plant Physiol.* 136:2843-2854.
46. Sonsthagen SA, Talbot SL, Lanctot RB, Scribner KT, & McCracken KG (2009) Heirarchical spatial genetic structure of common eiders (*Somateria mollissima*) breeding along a migratory corridor. *Auk* 126:744-754.
47. Lennon NJ*, et al.* (2010) A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol.* 11:R15.
48. Peakall R & Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6:288-295.
49. Smouse PE & Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561-573.
50. Pritchard JK, Stephens M, & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
51. Falush D, Stephens M, & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.
52. Earl D & vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resources* 4:359-361.
53. Evanno G, Regnaut S, & Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol. Ecol.* 14:2611-2620.
54. Tallmon DA, Bellemain E, Swenson JE, & Taberlet P (2004) Genetic monitoring of Scandinavian brown bear effective population size and immigration. *J. Wildlife Management* 68:960-965.
55. Nachman MW & Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297-304.
56. Huang X, Pevzner P, & Miller W (1994) Parametric recomputing in alignment graphs. *Combinatorial Pattern Matching; Springer Lecture Notes in Computer Science, 807*, eds Crochemore D & Gusfield D (Springer-Verlag, Berlin), pp 87-101.
57. Hubbard TJ*, et al.* (2009) Ensembl 2009. *Nucleic Acid Research* 37:D690-697.
58. Lin Y-W, MacMullen C, Ganguly A, Stanley CA, & Shyng S-L (2006) A novel KCNJ11 mutation associated with congenital hyperinsulinism reduces the intrinsic open probability of beta-cell ATP-sensitive potassium channels. *J. Biol. Chem.* 281:3006-3012.
59. Nascimento EBM*, et al.* (2006) Insulin-mediated phosphorylation of the proline-rich Akt substrate PRAS40 is impaired in insulin target tissues of high-fat diet-fed rats. *Diabetes* 55:3221-3228.
60. Martin SL (2008) Mammalian hibernation: a naturally reversible model for insulin resistance in man? *Diab. Vascular Disease Res.* 5:76-81.

61. Adzhubei IA, *et al.* (2010) A method and server for predicting damaging missense mutations. *Nature Methods* 7:248-249.
62. Nelson RA, *et al.* (1983) Behavior, biochemistry, and hibernation in black, grizzly, and polar bears. *International Conf. Bear Res. Manage.* 5:284-290.
63. Thiemann GW, Iverson SJ, & Stirling I (2006) Seasonal, sexual and anatomical variability in the adipose tissue of polar bears (*Ursus maritimus*). *J. Zool.* 269:65-76.
64. Soupene E, Dinh N, Siliakus M, & Kuypers F (2010) Activity of the acyl-CoA synthetase ACSL6 isoforms: role of the fatty acid Gate-domains. *BMC Biochem.* 11:18.
65. Kim JB, *et al.* (2005) Association of mahogany/attractin gene (ATRN) with porcine growth and fat. *Asian Australasian J. Anim. Sci.* 18:1383-1386.
66. Miljkovic I, *et al.* (2009) Association of the CPT1B gene with skeletal muscle fat infiltration in Afro-Caribbean men. *Obesity* 17:1396-1401.
67. Nagasaki S, Miki Y, Akahira J-i, Suzuki T, & Sasano H (2009) Transcriptional regulation of 17beta-hydroxysteroid dehydrogenase type 12 by SREBP-1. *Mol. Cell. Endocr.* 307:163-168.
68. Wang S, *et al.* (2009) Subtyping obesity with microarrays: implications for the diagnosis and treatment of obesity. *Int. J. Obes. (Lond.)* 33:481-489.
69. Kaser S, *et al.* (2004) Effects of weight loss on PLTP activity and HDL particle size. *Int. J. Obes. Relat. Metab. Disord.* 28:1280-1282.
70. Bochukova EG, *et al.* (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 463:666-670.
71. Meyre D, *et al.* (2008) R125W coding variant in TBC1D1 confers risk for familial obesity and contributes to linkage on chromosome 4p14 in the French population. *Hum. Mol. Genet.* 17:1798-1802.
72. Stone S, *et al.* (2006) TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition. *Human Molecular Genetics* 15:2709-2720.
73. Dina C, *et al.* (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 39:724-726.
74. Fischer J, *et al.* (2009) Inactivation of the Fto gene protects from obesity. *Nature* 458:894-898.
75. Bruce DS, *et al.* (1990) Is the polar bear (*Ursus maritimus*) a hibernator?: continued studies on opioids and hibernation. *Pharmacol. Biochem. Behavior* 35:705-711.
76. Nakagawa T, Lomb DJ, Haigis MC, & Guarente L (2009) SIRT5 deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle. *Cell* 137:560-570.
77. Wang J, *et al.* (2011) Molecular characterization of CPS1 deletions by array CGH. *Mol. Genet. Metab.* 102:103-106.
78. Yan J, Barnes BM, Kohl F, & Marr TG (2008) Modulation of gene expression in hibernating arctic ground squirrels. *Physiol. Genomics* 32:170-181.
79. Nakamura TJ, Shinohara K, Funabashi T, & Kimura F (2001) Effect of estrogen on the expression of Cry1 and Cry2 mRNAs in the suprachiasmatic nucleus of female rats. *Neuroscience Research* 41:251-255.
80. van der Horst GT, *et al.* (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* 398:627-630.
81. Kanabrocki EL, *et al.* (2000) Circadian relationship of serum uric acid and nitric oxide. *JAMA* 283:2240-2241.

82. Øritsland NA (1970) Temperature regulation of the polar bear (*Thalarctos maritimus*). *Comp. Biochem. Physiol.* 37:225.
83. Hurst RJ, Leonard ML, Watts PD, Beckerton P, & Øritsland NA (1982) Polar bear locomotion: body temperature and energetic cost. *Can. J. Zool.* 60:40-44.
84. Øritsland NA, Jonkel C, & Ronald K (1976) A respiration chamber for exercising polar bears. *Norw. J. Zool.* 24:65-67.
85. Jiang XC*, et al.* (1999) Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *J. Clin. Invest.* 103:907-914.
86. Kaduce TL, Spector AA, & Folk GE (1981) Characterization of the plasma-lipids and lipoproteins of the polar bear. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 69:541-545.
87. Hissa R*, et al.* (1994) Seasonal patterns in the physiology of the European brown bear (Ursus arctos arctos) in Finland. *Comp. Biochem. Physiol. A. Physiol.* 109:781-791.
88. Frank N, Elliott SB, Allin SB, & Ramsay EC (2006) Blood lipid concentrations and lipoprotein patterns in captive and wild American black bears (*Ursus americanus*). *Am. J. Vet. Res.* 67:335-341.
89. Davis RW*, et al.* (1991) Lipoproteins in pinnipeds: analysis of a high molecular weight form of apolipoprotein E. *J. Lipid Res.* 32:1013-1023.
90. Sakamoto K & Holman GD (2008) Emerging role for AS160/TBC1D4 and TBC1D1 in the regulation of GLUT4 traffic. *Am. J. Physiol. Endocrinol. Metab.* 295:E29-E37.
91. Chiang P-M*, et al.* (2010) Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. *Proc. Nat. Acad. Sci. U.S.A.* 107:16320-16324.
92. Chadt A*, et al.* (2008) Tbc1d1 mutation in lean mouse strain confers leanness and protects from diet-induced obesity. *Nat Genet* 40:1354-1359.
93. Grahl-Nielsen O*, et al.* (2003) Fatty acid composition of the adipose tissue of polar bears and their prey: ringed seals, bearded seals and harp seals. *Mar. Ecol. prog. Ser.* 265:275.
94. Rantakari P*, et al.* (2010) Hydroxysteroid (17beta) dehydrogenase 12 is essential for mouse organogenesis and embryonic survival. *Endocrinology* 151:1893-1901.
95. Platner WS, Patnayak BC, & Musacchia XJ (1972) Seasonal changes in the fatty acid spectrum in the hibernating and non-hibernating ground squirrel *Citellus tridecemlineatus*. *Comp. Biochem. Physiol. A. Physiol.* 42:927-938.
96. Turchetto E, Gandolfi M, & Rondinini R (1963) The fatty acids of the deposit lipids in the hibernant. *Cell. Mol. Life Sci.* 19:403-404.
97. Lee TM, Pelz K, Licht P, & Zucker I (1990) Testosterone influences hibernation in golden-mantled ground squirrels. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 259:R760-R767.
98. Wu X*, et al.* (1994) Hyperuricemia and urate nephropathy in urate oxidase-deficient mice. *Proc. Nat. Acad. Sci. U.S.A.* 91:742-746.
99. Eraly SA*, et al.* (2008) Multiple organic anion transporters contribute to net renal excretion of uric acid. *Physiol. Genomics* 33:180-192.
100. Ma YL*, et al.* (2004) Ascorbate distribution during hibernation is independent of ascorbate redox state. *Free Radical Biol. Medicine* 37:511-520.
101. Hamet P & Tremblay J (2006) Genetics of the sleep-wake cycle and its disorders. *Metabolism* 55, Supplement 2:S7-S12.

102. Nagashima K*, et al.* (2005) The involvement of Cry1 and Cry2 genes in the regulation of the circadian body temperature rhythm in mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 288:R329-R335.
103. Van Raamsdonk CD, Fitch KR, Fuchs H, de Angelis MH, & Barsh GS (2004) Effects of G-protein mutations on skin color. *Nat Genet* 36:961-968.
104. Cieslak M, Reissmann M, Hofreiter M, & Ludwig A (2011) Colours of domestication. *Biol. Rev.* 86:885-899.
105. Shin MK, Levorse JM, Ingram RS, & Tilghman SM (1999) The temporal requirement for endothelin receptor-B signalling during neural crest development. *Nature* 402:496-501.
106. Edery P*, et al.* (1996) Mutation of the endothelin-3 gene in the Waardenburg-Hirschsprung disease (Shah-Waardenburg syndrome). *Nat Genet* 12:442-444.
107. Ritland K, Newton C, & Marshall HD (2001) Inheritance and population structure of the white-phased "Kermode" black bear. *Curr. Biol.* 11:1468-1472.
108. Hirobe T, Ito S, & Wakamatsu K (2010) The mouse pink-eyed dilution allele of the P-gene greatly inhibits eumelanin but not pheomelanin synthesis. *Pigment Cell Melanoma Res.* 24:241-246.
109. Costin GE*, et al.* (2005) Mutations in dopachrome tautomerase (Dct) affect eumelanin/pheomelanin synthesis, but do not affect intracellular trafficking of the mutant protein. *Biochem. J.* 391:249-259.
110. Rieder S, Taourit S, Mariat D, Langlois B, & Guérin G (2001) Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mammalian Genome* 12:450-455.
111. Schmutz SM, Berryere TG, Ellinwood NM, Kerns JA, & Barsh GS (2003) MC1R studies in dogs with melanistic mask or brindle patterns. *J. Hered.* 94:69-73.
112. Schmutz SM, Berryere TG, & Goldfinch AD (2002) TYRP1 and MC1R genotypes and their effects on coat color in dogs. *Mammalian Genome* 13:380-387.
113. Bellone RR*, et al.* (2008) Differential gene expression of TRPM1, the potential cause of congenital stationary night blindness and coat spotting patterns (LP) in the Appaloosa horse (*Equus caballus*). *Genetics* 179:1861-1870.
114. Lu S*, et al.* (2010) The correlation of TRPM1 (Melastatin) mRNA expression with microphthalmia-associated transcription factor (MITF) and other melanogenesis-related proteins in normal and pathological skin, hair follicles and melanocytic nevi. *J. Cutaneous Pathol.* 37:26-40.
115. Naz G*, et al.* (2009) Novel missense mutations in lipase H (LIPH) gene causing autosomal recessive hypotrichosis (LAH2). *J. Dermatolog. Sci.* 54:12-16.
116. Diribarne M*, et al.* (2011) A deletion in exon 9 of the LIPH gene is responsible for the rex hair coat phenotype in rabbits (Oryctolagus cuniculus). *PLoS ONE* 6:e19281.
117. Meyer W*, et al.* (2011) Basic structural and functional characteristics of the epidermal barrier in wild mammals living in different habitats and climates. *Eur. J. Wildlife Res.* 57:873-885.
118. Amstrup SC (2003) Polar bear - *Ursus maritimus*. *Wild Mammals of North America*, eds Feldhamer GA, Thompson BC, & Chapman JA (Marine Mammal Commission: Biology, Management, and Conservation, Washington, D.C.), Second Ed, pp 39-56.
119. Schmutz S*, et al.* (2004) A form of albinism in cattle is caused by a tyrosinase frameshift mutation. *Mammalian Genome* 15:62-67.

120. Schmidt-Küntzel A, Eizirik E, O'Brien SJ, & Menotti-Raymond M (2005) Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. *J. Hered.* 96:289-301.
121. Gariepy CE, Cass DT, & Yanagisawa M (1996) Null mutation of endothelin receptor type B gene in spotting lethal rats causes aganglionic megacolon and white coat color. *Proc. Nat. Acad. Sci. U.S.A.* 93:867-872.
122. Rees JL (2003) Genetics of hair and skin color. *Ann. Rev. Genet.* 37:67-90.
123. Han J*, et al.* (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* 4:e1000074.
124. Wade CM*, et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865-867.
125. Imes DL, Geary LA, Grahn RA, & Lyons LA (2006) Albinism in the domestic cat (Felis catus) is associated with a tyrosinase (TYR) mutation. *Anim. Genet.* 37:175-178.
126. Gratten J*, et al.* (2007) Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proc. Royal Soc. Lond. B: Biol. Sci.* 274:619-626.
127. Gratten J*, et al.* (2008) A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* 319:318-320.