

Current Biology, Volume 21

## Supplemental Information

### Ancient Hybridization and an Irish Origin

#### for the Modern Polar Bear Matriline

Ceiridwen J. Edwards, Marc A. Suchard, Philippe Lemey, John J. Welch, Ian Barnes, Tara L. Fulton, Ross Barnett, Tamsin C. O'Connell, Peter Coxon, Nigel Monaghan, Cristina E. Valdiosera, Eline D. Lorenzen, Eske Willerslev, Gennady F. Baryshnikov, Andrew Rambaut, Mark G. Thomas, Daniel G. Bradley, and Beth Shapiro

#### Supplemental Inventory

**1. Supplemental Figures**

Figure S1, related to Figure 1 and Supplemental Experimental Procedures section C

Figure S2, related to Figure 2 and Supplemental Experimental Procedures section D

**2. Supplemental Tables**

Table S1, related to Figures 1 and 2 and Supplemental Experimental Procedures section A

Table S2, related to Figure 3 and Supplemental Experimental Procedures section A

**3. Supplemental Experimental Procedures**

A. Mitochondrial and Nuclear DNA Data Collection

B. Radiocarbon Dating and Isotope Analysis

C. Phylogeographic Analysis of mtDNA Control Region Sequences

D. Assessing the Sensitivity of Phylogeographic Inference to Potential Sampling Bias

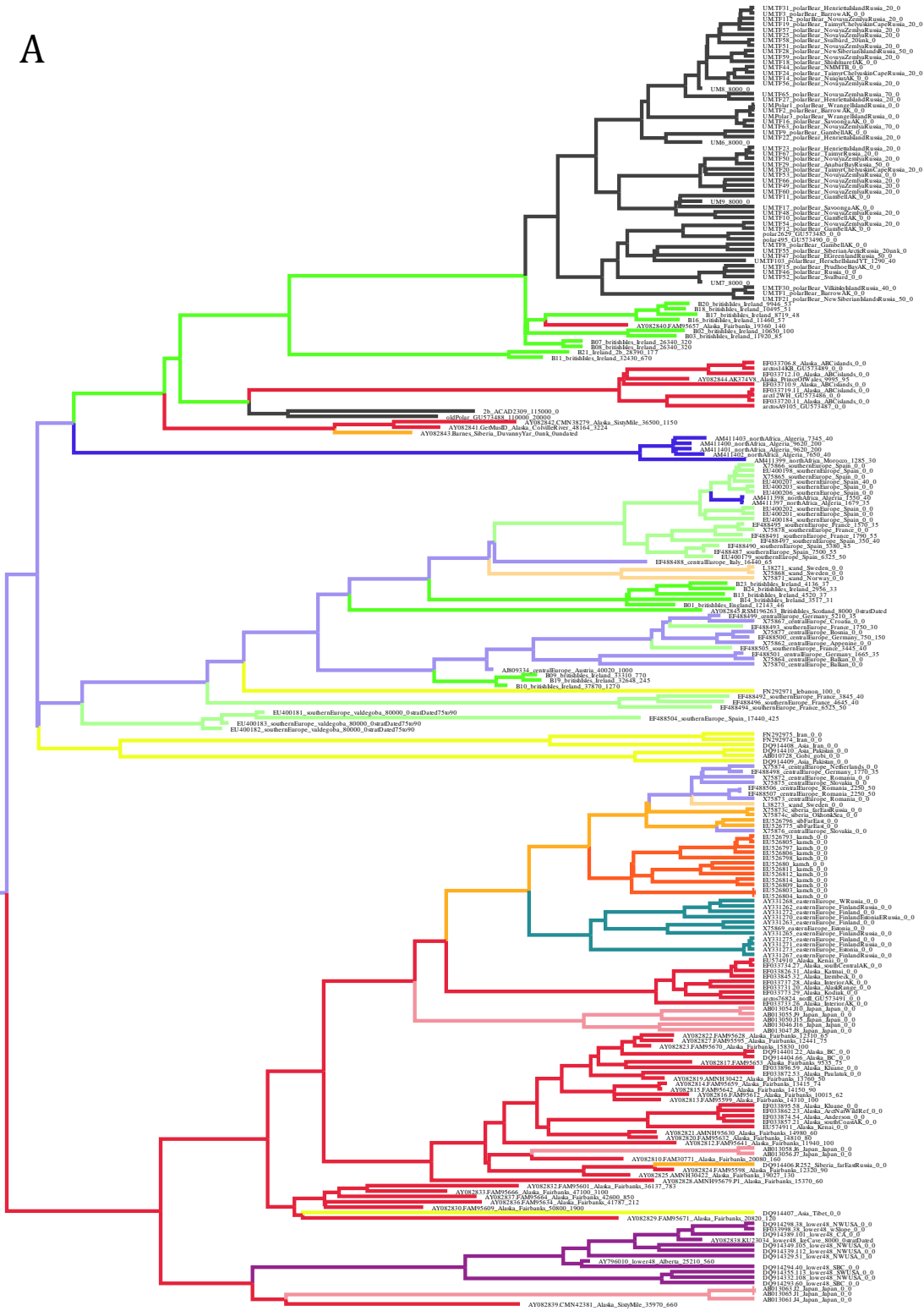
E. Assessing the Demographic Plausibility of the Recent MRCA of Polar Bear Matrilines

F. Evaluating the Molecular Rate

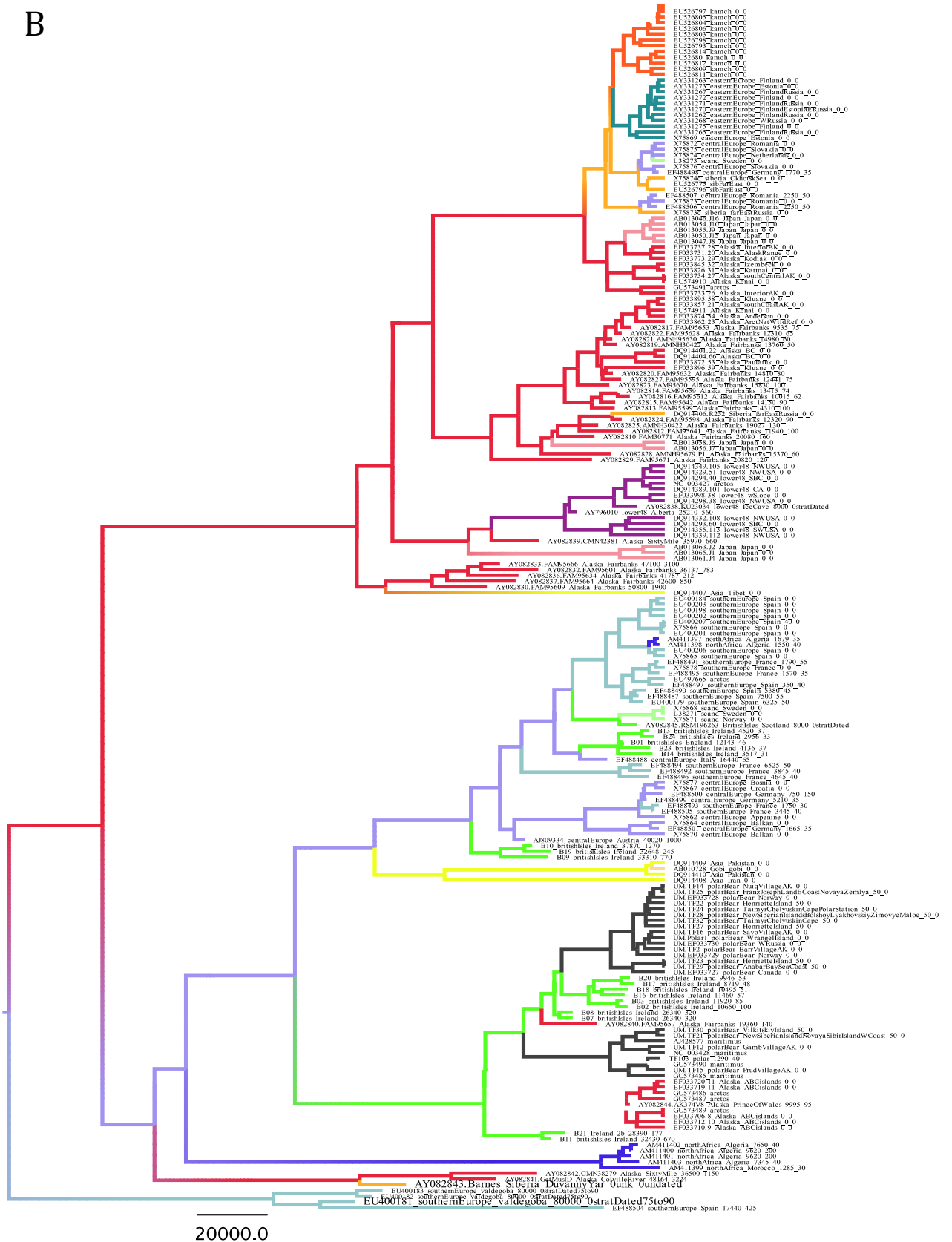
**4. Supplemental Acknowledgments**

**5. Supplemental References**

A

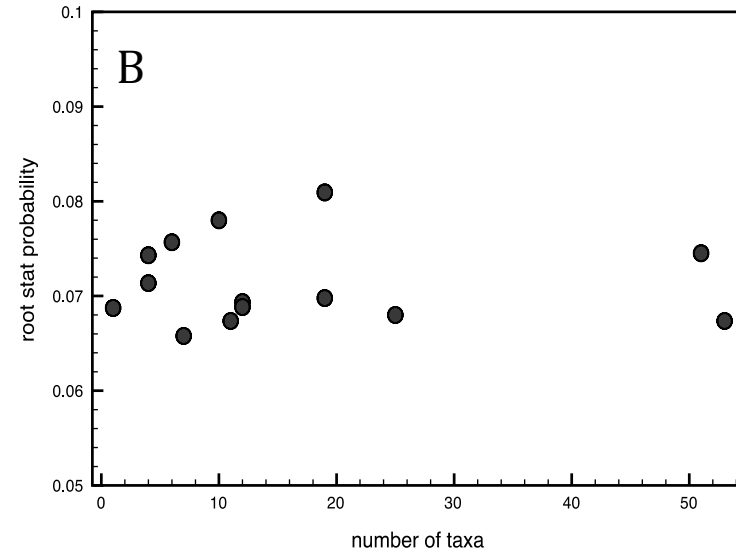
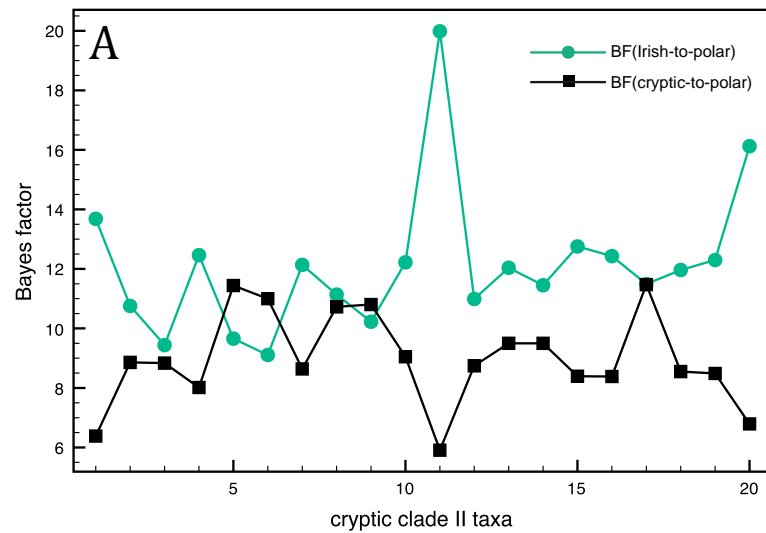


B



**Figure S1.**

The MCC genealogy depicted as Figure 1 in the main text, with sequence names (zoom in to read taxon labels) (A), and including ten complete mitochondrial genomes and the pruned data set of 15 polar bears (B). Colors along the branches are as in the main text, and indicate the most probable geographic location.



**Figure S2. Exploration of the Influence of the Sampling Scheme on the Phylogeographic Reconstruction Described in Figures 1 and 2**

(A) The impact of cryptic clade II taxa on phylogeographic inference of modern polar bear ancestry. Bayes factor support for an Irish (green) and cryptic (black) polar bear ancestry is plotted as function of an increasing number of cryptic taxa within clade II.

(B) Root state probability as a function of the number of taxa sampled from that particular location state. The root location estimates result from a MCMC analysis that randomizes location state assignment.

**Table S1.**

*(See separate Excel spreadsheet.)*

**Table S2. Supplemental Details about the Bear Specimens from Britain and Ireland Used in the Analyses Depicted in Figures 1 and 2**

| Lab code | Site information                      | Genbank Accession Number (CR) | Genbank Accession Number (Cytb) | NMING number | # of Dublin extractions | Sequencing D=Dublin; O=Oxford | # cloned replicates | Mean !13C | Mean !15N | Uncalibrated radiocarbon date (Lab number) | CR1              | CR2 | cytb | mtDNA clade assignment |
|----------|---------------------------------------|-------------------------------|---------------------------------|--------------|-------------------------|-------------------------------|---------------------|-----------|-----------|--|------------------|-----|------|------------------------|
| B01      | Carsington Pasture Cave, Derbyshire   | JF900158                      | JF900089                        | n/a          | 5                       | D/O                           | 3                   | -18.89    | 5.76      | 12143 +/- 46 bp (KIA26352)                 | yes              | yes | yes  | I                      |
| B02      | Red Cellar Cave, Co. Limerick         | JF900159                      | JF900090                        | F21104       | 3                       | D/O                           | 3                   | -21.22    | 7.01      | 10650 +/- 100 bp (OxA-3704)                | yes              | yes | yes  | II                     |
| B03      | Plunkett Cave, Kesh Corran, Co. Sligo | JF900160                      | JF900091                        | F21119       | 2                       | D                             | n/d                 | -17.80    | 6.72      | 11920 +/- 85 bp (OxA-3706)                 | 51bp             | yes | yes  | II                     |
| B04      | Donore Bog, Co. Laois                 | n/a                           | n/a                             | F16002       | 3                       | D                             | n/a                 | n/a       | n/a       | 8930 +/- 80 bp (OxA-3713)                  | no amplification |     |      | n/a                    |
| B05      | Donore Bog, Co. Laois                 | n/a                           | n/a                             | F16003       | 3                       | D                             | n/a                 | n/a       | n/a       | not dated                                  | no amplification |     |      | n/a                    |
| B06      | Foley Cave, Co. Cork                  | n/a                           | n/a                             | F20291       | 3                       | D                             | n/a                 | n/a       | n/a       | n/a  | no amplification |     |      | n/a                    |
| B07      | Foley Cave, Co. Cork                  | JF900161                      | JF900092                        | F20292       | 2                       | D                             | n/d                 | -16.10    | 9.63      | 26340 +/- 320 bp (OxA-3722)                | yes              | no  | yes  | II                     |
| B08      | Foley Cave, Co. Cork                  | JF900162                      | JF900093                        | F20293       | 2                       | D                             | n/d                 | -21.55    | 9.49      |  | yes              | yes | yes  | II                     |
| B09      | Castlepook Cave, Co. Cork             | JF900163                      | JF900094                        | F21141       | 2                       | D                             | 6                   | -19.66    | 5.95      | 33310 +/- 770 bp (OxA-4231)                | yes              | yes | yes  | I                      |
| B10      | Castlepook Cave, Co. Cork             | JF900164                      |                                 | F21148       | 3                       | D/O                           | n/d                 | -20.22    | 5.99      | 37870 +/- 1270 bp (OxA-4238)               | yes              | yes | no   | I                      |
| B11      | Shandon Cave, Co. Waterford           | JF900165                      | JF900095                        | F21160       | 5                       | D/O                           | n/d                 | -18.67    | 7.22      | 32430 +/- 670 bp (OxA-4245)                | yes              | yes | yes  | II                     |
| B12      | Ballynamindra Cave, Co. Waterford     | n/a                           | n/a                             | F21168       | 2                       | D                             | n/a                 | n/a       | n/a       | 35570 +/- 1100 bp (OxA-4252)               | no amplification |     |      | n/a                    |
| B13      | Poll na mBéar Cave, Co. Leitrim       | JF900166                      | JF900096                        | F21456/8     | 2                       | D                             | 16                  | -21.65    | 8.83      | 4520 +/- 37 bp (UB-6697)                   | yes              | yes | yes  | I                      |
| B14      | Poll na mBéar Cave, Co. Leitrim       | JF900167                      | JF900097                        | F21458       | 2                       | D                             | n/d                 | n/a       | n/a       | 3517 +/- 31 bp (OxA-15479)                 | yes              | yes | yes  | I                      |
| B15      | Lough Gur, Co. Limerick               | n/a                           | n/a                             | F21646/3     | 2                       | D                             | n/a                 | -23.89    | 3.98      | not dated                                  | no amplification |     |      | n/a                    |
| B16      | Plunkett Cave, Kesh Corran, Co. Sligo | JF900168                      | JF900098                        | F21748       | 2                       | D                             | 11                  | -17.88    | 6.53      | 11460 +/- 57 bp (UB-6698)                  | yes              | yes | yes  | II                     |
| B17      | Red Cellar Cave, Co. Limerick         | JF900169                      | JF900099                        | F21749       | 2                       | D                             | n/d                 | -21.39    | 6.81      | 8719 +/- 48 bp (UB-6699)                   | yes              | yes | yes  | II                     |
| B18      | Edenvale Cave, Co. Clare              | JF900170                      |                                 | F21750       | 2                       | D                             | n/d                 | -19.91    | 6.15      | 10495 +/- 51 bp (UB-6700)                  | yes              | yes | 66bp | II                     |
| B19      | Castlepook Cave, Co. Cork             | JF900171                      | JF900100                        | F21751       | 2                       | D                             | n/d                 | -19.98    | 5.48      | 32648 +/- 245 bp (UB-6701)                 | yes              | no  | yes  | I                      |
| B20      | Newhall Cave, Edenvale, Co. Clare     | JF900172                      | JF900101                        | F21752       | 2                       | D                             | 15                  | -18.49    | 9.41      | 9946 +/- 53 bp (UB-6702)                   | yes              | yes | yes  | II                     |
| B21      | Shandon Cave, Co. Waterford           | JF900173                      | JF900102                        | F21753       | 2                       | D                             | 15                  | -19.06    | 7.64      | 28390 +/- 177 bp (UB-6703)                 | yes              | yes | yes  | II                     |
| B22      | Derrykeel Bog, Co. Offaly             | n/a                           | n/a                             | F16011       | 2                       | D                             | n/a                 | n/a       | n/a       | not dated                                  | no amplification |     |      | n/a                    |
| B23      | Polldownin Cave, Co. Sligo            | JF900174                      | JF900103                        | F21439/85    | 2                       | D                             | n/d                 | -21.14    | 4.57      | 4136 +/- 37 bp (UB-6704)                   | yes              | yes | yes  | I                      |
| B24      | Poll na mBéar Cave, Co. Leitrim       | JF900175                      | JF900104                        | F21455/43    | 2                       | D                             | n/d                 | -21.71    | 4.63      | 2956 +/- 33 bp (UB-6705)                   | yes              | yes | yes  | I                      |

## Supplemental Experimental Procedures

### A. Mitochondrial and Nuclear DNA Data Collection

#### Ancient Irish and British Brown Bears

We collected skeletal elements of 23 *Ursus* sp. samples from 12 cave sites in Ireland and one individual from England for DNA analysis. Appropriate procedures were followed for the handling and analysis of ancient specimens during all steps of the DNA extraction and amplification procedure [46-48]. DNA was extracted in the specialized ancient DNA facilities at the Smurfit Institute of Genetics, Trinity College, Dublin, following the protocol described in Edwards et al [49]. To overlap with previously published data, two non-overlapping fragments of 100-bp and 176-bp of a hypervariable section of the mitochondrial control region (HVR) were amplified using primers and protocols described in Barnes et al [15]. Amplicons were directly sequenced in both directions, and a subset of PCR products from seven samples (Table S1) were cloned using the kit provided by NextGen Sciences Ltd., Cambridgeshire, according to manufacturer's instructions. Between three and 16 clones were sequenced to examine the extent of damage and to confirm the absence of nuclear copies of mitochondrial sequences (numts) [48]. Four of the 24 samples were independently replicated at the specialized Henry Wellcome Ancient Biomolecules Centre, Oxford, following the protocol described by Shapiro et al [50] and using the primers and conditions as above. Cloned consensus sequences and independent replication revealed minimal levels of DNA damage. In addition, no amplification products were observed either in extraction or amplification negative controls. In total, 18 out of 24 samples (75%) were successful for CR amplification.

To further confirm the authenticity of clade II bears in Ireland, an additional 165-bp fragment of the mitochondrial cytochrome *b* (*cytb*) gene was targeted. The primer pair L15425 (5'-CCC RTT CCA YCC ATA CTA TAC AA-3') and H15547 (5'-AAC CAR TCG GGT TTG ATG TGG-3') was designed for screening, as it amplifies a fragment of *cytb* spanning three of the five diagnostic positions differentiating clade II from all other known brown bear clades [13]. Amplification took place in 50  $\mu$ l reactions comprising: 10 mM Tris-HCl pH 8.8, 2.5 mM MgCl<sub>2</sub>, 50 mM KCl, 0.1 % Triton X-100, deoxyribonucleotides at 200  $\mu$ M each, 0.4  $\mu$ M each primer, 1.0 unit of PLATINUM taq DNA polymerase. Amplifications were performed in an M-J Research PTC-200 DNA Engine using a heated lid. A three-minute denaturation step at 94 °C was followed by 40 cycles of 30 seconds denaturation at 93 °C, 30 seconds annealing at 55 °C, 30 seconds extension at 72 °C and then a four-minute final extension step at 72 °C. The *cytb* fragment was amplified successfully from 17 of 18 bears for which the CR was amplified. One individual (B10) did not yield any amplification product for *cytb*. All other clade I individuals (B01, B09, B13, B14, B19, B23 and B24) were typed as GAC, whereas the remaining 10 bears (B02, B03, B07, B08, B11, B16, B17, B18, B20 and B21) had AGT.

#### Modern and Holocene Polar Bears

To augment the number of polar bear mtDNA lineages available for analysis, whole-genomic DNA was extracted and a short fragment of the mtDNA CR was amplified from 17 modern polar bears, 29 historic polar bears (ranging in collection year from 1883 to 1964 and sampled from the collection housed at the Paleontological Institute of the Russian

Academy of Science, St Petersburg), one polar bear from Herschel Island, Yukon Territory, Canada, with a radiocarbon age of  $1290 \pm 40$  years [51], and four polar bears from an archaeological site in Zhokov Island in the De-Long Archipelago of the Siberian High Arctic (New Siberian Islands), which have been dated to ca. 8000 years BP [52] (Table S1). Modern polar bears were extracted using the QiaGen DNEasy Tissue Kit (Qiagen, USA) according to manufacturers instructions. Historic and late Holocene bears were extracted using the protocol for ancient bone described in Rohland and Hofreiter [53]. The four early Holocene polar bears were extracted using the protocol described in [54]. To avoid potential contamination by modern DNA, DNA extraction and PCR set-up for ancient and specimens was performed in specialized ancient DNA facilities at either Penn State University or the University of Copenhagen, with downstream molecular biology work including PCR amplification occurring in the physically isolated modern DNA laboratory. Either a single 330 bp fragment or two fragments nested within the larger fragment were amplified using a combination of previously published primers [43, 55, 56] with variation in amplicon length determined by the preservation of the specimen. For historic specimens, samples were amplified for both the 330-bp product and one of the smaller, nested regions, and in no case were base mismatches observed.

All novel sequences have been submitted to GenBank, with accession numbers JF900098-JF900175 (Tables S1 and S2).

#### Assembling the Mitochondrial DNA Data Set

In addition to the newly isolated sequences described above, we compiled a data set of all published mtDNA CR sequences from brown bears and polar bears as of February 2010, including sequences from all seven extant European brown bear populations and populations in Pakistan, Tibet, Mongolia, Siberia, Kamchatka, Japan and North America [15, 16, 57-59]. Available complete mitochondrial genome sequences for brown bears and polar bears were also acquired and the short fragments were aligned to these. Ancient DNA sequences from extinct North African populations and Late Pleistocene populations from Europe, Siberia and North America were included [15, 16, 43, 58, 60, 61]. This large amount of data was pruned to a smaller data set consisting of 242 sequences representing as broad a geographic and temporal range as is currently possible. From each of the deeply sampled modern locations (Alaska, central North America, Japan, Eastern Europe, Kamchatka) we selected 10-15 modern mtDNA sequences chosen to best represent the geographic extent of brown bears from that region. As a sensitivity analysis, the number of polar bear sequences included in the analysis was also pruned from 58 to 15 sequences via random deletion of sequences, with no effect on the results of either the true-data analyses or the two sensitivity analyses described below.

As many fewer ancient than modern sequences were available, all ancient specimens, with the exception of two ancient Irish bears that were missing >40% of the total sequence coverage, were included in the analysis. To maximize the temporal and geographic coverage of the data set, analyses were performed on two mitochondrial data sets. First, the alignment was restricted to two short fragments of the mtDNA HVR that had been sequenced for the ancient specimens, separated by a removed region of polybasic runs, for a total of ~200 bp. While limited in length, this alignment has been shown previously to be sufficiently informative to identify local population structure in brown bears [15, 56, 60-62]. Second, the short fragment alignment was aligned to the eleven



available complete mitochondrial genome sequences from brown bears and polar bears and analyzed together using a data-partitioning approach (see below). Prior to analysis, all brown bears and polar bears were assigned to one of 14 geographic locations, which were selected based on: (i) the current distribution of bears; (ii) known locations of extinct populations; and (iii) the number of samples available from a particular region. GenBank accession numbers and geographic information are provided in Table S1. The geographic locations are also provided in Figure 1 in the main text.

#### Assembling the Nuclear DNA Data Set

Nuclear DNA alignments were constructed for 20 fragments where data were available from each of the eight species of Ursidae used in the complete mitochondrial analysis. These 20 fragments included genes from autosomal, X and Y chromosomes, and included both coding and non-coding regions [7, 22, 63]. Four data sets were constructed: (i) a data set comprising all 20 available nuclear DNA fragments (D11, D12, FES, SIS, SRY, TG, TRH, selenocysteine tRNA, TSH $\beta$ , vWF, ZFY, UBEY1, TTR, IRBP, ZFY final intron, ZFX final intron, PLP, SMCY, SMCX, ALAS2; 15,761 bp); (ii) a data set restricted to the five fragments located on the Y chromosome (4,694 bp); (iii) a data consisting of all autosomal (non-X, non-Y) nuclear loci (5,556 bp); and (iv) a subset of 12 nuclear fragments that were sequenced as part of an analysis by Pages *et al* [7] (7,698 bp). For the Pages *et al* data, each published brown bear and polar bear sequence is a consensus of several individuals, with intra-species polymorphisms coded as ambiguous bases. For brown bears, that consensus includes both clade IV and clade I bears, and therefore should represent the broadest possible extent of modern brown bear genetic diversity. By comparing this diversity with that of the mitochondrial DNA, we therefore provide a test of the potentially different nuclear and mitochondrial evolutionary histories for polar bears.

#### **B. Radiocarbon Dating and Isotope Analysis**

Eleven of the 23 Irish brown bears had been previously dated as part of The Irish Quaternary Fauna Project [64] (OxA dates; Table S2). A brown bear from Derbyshire (B01) was dated by the Leibniz-Laboratory for Radiometric Dating, Kiel (KIA26352; Table S1). AMS radiocarbon dates were generated for the nine additional Irish samples from which DNA was successfully amplified by the <sup>14</sup>Chrono Centre, Queen's University Belfast (UB dates; Table S2). Dates in Table S2 are uncalibrated radiocarbon years before present (BP). Included within these are the youngest bear remains as yet reported from Ireland: the Poll na mBéar sample, originating from County Leitrim, with an age of 2956  $\pm$  33 uncalibrated radiocarbon years before present (UB-6705; Table S2). This postdates the previously proposed extinction of brown bears in Ireland, which was thought to have occurred before or simultaneously with the arrival of humans on the island ca. 10,000 BP.

Stable isotope analysis was carried out at the McDonald Institute for Archaeological Research, and the Godwin Laboratory, University of Cambridge. Collagen was extracted from the samples using a modified version of the procedure described in O'Connell & Hedges [65]. 0.5-1.0 g pieces of bone were sand blasted then demineralized in 0.5 M hydrochloric acid at 4 °C for between one and three days, until all the mineral had been removed. The samples were then gelatinized in water of pH 3 for 48 hours at 75 °C, and the soluble collagen lyophilized. The dried collagen was analyzed in triplicate using continuous

flow isotope ratio mass spectrometry (Costech elemental analyzer coupled to a Finnigan MAT253 mass spectrometer). Carbon and nitrogen stable isotope ratios are given as  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values, relative to internationally defined scales (VPDB for carbon, and AIR for nitrogen), and are expressed in units of 'per mil' [66]. Based on replicate analyses of international and laboratory standards, measurement errors were less than  $\pm 0.2$  ‰ for  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ . Stable isotope values are shown in Table S2.

Isotopic analysis of our ancient brown bears from Britain and Ireland (Table S2) indicate a diet distinct from polar bears and similar to Late Pleistocene Alaskan brown bears [15].

### **C. Phylogeographic Analysis of mtDNA Control Region Sequences**

#### A Novel Approach to Bayesian Phylogeographic Analysis

To infer the spatial dynamics of Late Pleistocene and Holocene brown bears and polar bears, we provide two specific innovations to extend and apply a Bayesian approach to large-scale phylogeographic questions. Briefly, a recently developed approach [38] allowed rates of diffusion between predefined locations to be estimated according to a continuous-time Markov chain (CTMC) model. We extend this to allow an asymmetric transition matrix, that is, to estimate separate directional rates of diffusion between a pair of locations. This approach provides a much more realistic approximation of the true diffusion process, where organisms are not likely to move at the same rate both *to* and *from* a given location. Second, we adopt a Bayesian stochastic search variable selection (BSSVS) procedure, which makes it possible to identify and estimate only those rates that are most informative to reconstruct the phylogeographic history of the given data, and, most crucially, makes possible explicit tests of alternate phylogeographic hypotheses. We integrate these methods into the popular, freely available, Bayesian phylogenetic analysis package, BEAST [39], where they can be applied to many data sets and phylogeographic hypotheses.

#### Bayesian Stochastic Search Variable Selection

By integrating CTMC models for discretized diffusion in a statistical framework centered on time-scaled phylogenies, we infer spatial dynamics in real timescales. To enable more precise location state reconstructions when the amount of data is small, we adopt a BSSVS procedure to select among all possible migration pathways (here, rates of diffusion between locations) [38]. BSSVS identifies a parsimonious scenario of spatial connectivity that appropriately explains the phylogeographic diffusion processes. BSSVS accomplishes this task through two assumptions. First, the search places non-negligible probability on each of the rate parameters exactly equaling zero. When a rate equals zero, it is not possible to directly diffuse from one location to another. The second aspect assumes that the number of non-zero rates is quite small unless the data suggest otherwise. This assumption is formalized as a Poisson prior distribution on the number of non-zero rates (26). Because directional rates of diffusion accommodate a more restricted set of phylogeographic dispersal patterns than an equivalent number of reversible rates, we grant more variance to this Poisson prior by setting its hyperparameter to  $K-1$ , where  $K$  is the number of locations, without any particular offset. Although this may in theory result in

migration graphs that are not fully connected, stable computation is ensured as described below.

### Asymmetric Transition Matrices

The standard assumption of reversibility in CTMC models of diffusion leads to the selection of migration pathways that impose balanced transitions among locations. Such balance is unlikely observed in reality, and a major limitation of existing approaches. To address this shortcoming, we develop a novel CTMC model in phylogenetics that allows for possibly different rates of diffusion between locations depending on the direction traveled.

Mathematically, an  $n \times n$  rate matrix  $\Lambda$  characterizes a CTMC-based diffusion between  $n$  locations. Under our new model,  $\Lambda$  contains  $n(n-1)$  off-diagonal, non-negative rate parameters  $\lambda_{ij}$  for  $i, j = 1, \dots, n$ . Previously phylogenetic modeling [e.g. 38] assumes  $\lambda_{ij} = \lambda_{ji}$ , or at least that  $\Lambda$  is similar to a symmetric matrix, as in, for example, the HKY and GTR models.

For our general case, almost every rate matrix  $\Lambda$  remains complex diagonalizable. This means that it is possible to decompose  $\Lambda = V D V^{-1}$ , where matrix  $V$  is a set of real eigenvectors and matrix  $D$  is of block diagonal form with  $1 \times 1$  submatrices of real eigenvalues and  $2 \times 2$  submatrices of complex conjugate pairs of eigenvalues.

To draw inference under this model, we wish to compute the likelihood of the observed locations at the tips of the tree. This computation requires determining the CTMC finite-time transition probabilities  $\exp(\Lambda t)$  for real times  $t$  along branches in the tree. Determining these transition probabilities proceeds by first decomposing  $\Lambda$ , exponentiating  $D$  and finally pre- and post-multiplying by  $V$  and  $V^{-1}$ , respectively. We provide specialized computer code in BEAST to handle the matrix-rotation that the complex pairs induce in this calculation.

BSSVS in the general, not strictly reversible models adds an additional challenge to this diagonalization as the BSSVS process puts non-negligible probability on structural zeros in the rate matrix. Consequentially, the set of non-diagonalizable matrices no longer has measure zero. To handle these limited situations numerically and ensure stable computation, we further require that the condition number of matrix  $V$  returned from the diagonalizing code not exceed a large pre-specified constant (10,000; results insensitive to larger choices and within several orders of magnitude smaller). We assume that rates are *a priori* independent and gamma-distributed, and explore two different prior expectations [38]: (i) all expectations are equal and (ii) prior rate expectations are inversely proportional to the geographical distance between sampling locations. The latter allows the inference to draw upon geographical information that traditionally remains unused. We estimate probable migration pathways, diffusion rate parameters and gene genealogies simultaneously via the MCMC sampler implemented in BEAST [39].

### Phylogeographic Analysis of Brown and Polar Bear mtDNA CR

To infer the spatiotemporal evolutionary history of brown bear and polar bear matriline, we performed several Bayesian genealogical analyses using BEAST. Analyses were performed using the data set of 242 brown bears and polar bears representing the extent of their Late Pleistocene and Holocene distribution described above. Two alignments were used. First, we analyzed the ca. 200 bp data as described above. Second, we performed an analysis in which complete mitochondrial genomes are included for all modern bears for

which they are currently available. For the latter analysis, the smaller data set comprising only a subsample of the modern and historic polar bears were used, for a total of 210 taxa. In this analysis we allowed for different evolutionary rate and nucleotide substitution parameters between the short, variable fragment and the remaining mtDNA genome by considering as separate alignment partitions.

We assumed the HKY+G model of nucleotide substitution, as was selected using MODELTEST [67], and the flexible Bayesian skyride (Gaussian Markov random field) coalescent prior [41] for the short fragment analysis, and the Bayesian skyline plot [42] with four groups for the partitioned analysis. Mean AMS radiocarbon dates for ancient specimens were used to calibrate the molecular clock, and the tip ages for six specimens that were calibrated based on associated with stratigraphic layers were co-estimated with evolutionary and model parameters [21] (see below). Analyses of the short fragment data set were initially performed both with and without the post mortem damage (PMD) model [44], which makes it possible to assess the effect of potentially damaged sites on the phylogenetic analysis. Bayes factor (BF) comparison [18] via the relative marginal model likelihoods indicated that the addition of the PMD model did not improve the model fit, and it was excluded from further analyses.

For both alignments, we ran two Markov chain Monte Carlo (MCMC) simulations of 150 million iterations. After discarding the initial 10% of realizations as chain burn-in and sub-sampling every 10,000 iterations to decrease autocorrelation, we combined the remaining posteriors samples across chains. We inspected effective sample sizes and MCMC convergence visually using Tracer [68] and the maximum clade credibility (MCC) tree was summarized using TreeAnnotator. To provide a spatial projection, we convert the MCC tree with inferred posterior modal node location states and median node heights into a keyhole markup language (KML) file suitable for viewing with Google Earth. The resulting interactive projection is available online from <http://www.phylogeography.org/BEARS.html>.

We used a BF test [18] to determine which diffusion links were statistically significant ( $BF > 8$ ) based on standard BSSVS protocol. Further, the BFs reported in the main text supporting the geographic location of internal nodes are calculated as posterior odds/prior odds, where the prior odds derive from assigning each node to one of the thirteen specified locations with equal probability.

The majority of published ancient specimens associated with infinite radiocarbon dates were excluded from the global phylogeographic analysis; however six specimens originating from underrepresented geographic regions were included: one from Siberia [15], three from Valdegoba Cave, Spain [69], and the two fossil polar bears from Svalbard and northern Norway [10, 14]. For each of these, the age of the sequences was assumed *a priori* to arise from a lognormal distribution reflecting the most likely age of the specimen (Siberia: 90-40 kya; Valdegoba: 110-75 kya; polar bears: 130-45 kya). Tip ages were numerically integrated via sampling in the MCMC chain for each of these six specimens simultaneously with evolutionary and demographic parameters [21].

The MCC genealogy depicted as Figure 1 in the main text, but in which the taxon names are readable is provided as Figure S1A. The MCC genealogy resulting from the partitioned analysis including the ten complete modern bear mitochondrial genomes is provided as Figure S1B.

#### **D. Assessing the Sensitivity of Phylogeographic Inference to Potential Sampling Bias**

Our statistical phylogeographic method infers the locations of common ancestors by drawing from the set of observed locations of the available sequence data. The inference therefore relates only to the ancestry of the specific sequence sample. Although we are analyzing a comprehensive collection of bear mtDNA, the inference is inherently restricted by the availability of samples. In consideration of potential sampling biases, we investigate the sensitivity of our findings to taxon-sampling variability in two different ways.

##### Investigating the Impact of Unsourced Geographic Locations

Our results demonstrate that the modern polar bear mtDNA lineages are nested within the diversity of Late Pleistocene Irish brown bear mtDNA lineages. In establishing this close genetic relationship between modern polar bears and Irish bears from the data at hand, we cannot exclude, however, the possibility that a bear lineage even more closely related to modern polar bears was present at another, as of yet unsourced, location. To address this, we assess how our conclusion concerning the modern polar bear ancestry changes as an increasing number of ancient mtDNA lineages become available from a currently unsourced location. We design a simulation experiment that formalizes this scenario within our Bayesian phylogeographic inference by introducing an increasing number of ancient, unknown sequences from an undefined (unknown) location. To maximize the influence of these sequences on the results relating to polar bear ancestry, we restrict the simulated sequence diversity so that they fall within clade II.

Since we do not know the specific sequences these cryptic samples may yield, we integrate over all possible sequence realizations such that the samples are restricted to clade II with unknown sampling dates integrated over the range of 50 to 5 kya. We perform this integration prior to including geographic information about the unknown locations, as this information would unduly cluster the cryptic sequences together and test the sensitivity of our original conclusions. To accomplish this integration and then perform our phylogeographic inference on the observed data and cryptic samples assigned to a new unknown location with integrated sequences jointly, we employ a two-stage sequential MCMC analysis.

Figure S2A illustrates the sensitivity of our conclusions regarding the location or origin of the modern polar bear matriline to adding an increasing number of ancient mtDNA sequences originating from a currently unsourced (cryptic) location. We plot the BF support for the linkage between Irish/British brown bears and polar bears (green) and between bears from the unsourced location and polar bears (black) as a function of an increasing number of cryptic taxa. We observe BF support for the cryptic location as the location of origin for modern polar bear matriline, however Ireland/Britain remains the most likely location in most analyses, including analyses with a reasonably large number of simulated cryptic taxa. These results are consistent with our estimated genealogy, in which the modern polar bears are not reciprocally monophyletic with respect to Irish/British brown bears with high statistical support, but instead cluster together, frequently into several lineages within the greater diversity of Irish/British brown bears, suggesting a divergence that was too recent to be resolved using the available sequence data.

### Investigating the Influence of Tip Location Frequencies on Ancestral State Reconstruction

It is possible that the reconstruction of ancestral states is biased by the number of sequences that are sampled from each of the given locations. To investigate whether inferred ancestral locations are correlated with sampling frequencies in our phylogeographic inference, we consider the prior expectation for the root location frequencies given a particular sampling scheme. If we expect that the prior location distribution at the root were correlated with the location frequencies at the tips, we can also expect that ancestral reconstruction throughout the entire phylogeny will be influenced by this tip-location sampling frequency. To investigate this possibility, we perform an additional analysis in which we randomize the location assignments at the tips throughout the MCMC procedure. We achieve this by implementing a new MCMC transition kernel ('operator' in BEAST nomenclature) that swaps location states between two randomly selected taxa. Figure S2B plots the number of samples per location against the root location frequency resulting from this analysis. The empirical frequency converges quickly to the prior root location probability for a particular sampling scheme. This indicates that the sampling frequencies have little impact on the root location probability *a posteriori*, providing reassurance that the spatial structure in the data, rather than the sampling scheme, is informing our estimates. This is also consistent with previous findings of approximate independence [38].

### **E. Assessing the Demographic Plausibility of the Recent MRCA of Polar Bear Matrilines**

To investigate the plausibility of our estimate for the time to MRCA of the modern polar bear mitochondria, we calculated the time to coalescence that would be expected of a comparable sample of alleles under selective neutrality. This comparison does not make direct use of our genetic data, and so provides an effectively independent check on the plausibility of the date obtained.

From standard coalescent theory [e.g. 70], the timing of the first coalescent event occurring in a collection of  $n$  lineages is exponentially distributed, with an expected value of  $2N_e/(n(n-1))$  generations. From this result, it is trivial to simulate coalescence times for a sample of 51 alleles - the size of our sample from modern polar bears.

To translate the simulated values into real time, we require estimates of the generation time and effective population size of female polar bears. The effective size is particularly difficult to estimate, particularly in subdivided populations [71, 72]. However, estimates of both quantities have been presented by Cronin et al. [23]. We simulated coalescence times for an estimated female generation time of 9.7 years [23, 73], and global  $N_e$  estimates of 660, 1800, 2250 and 3304. Estimates of 1800 and 2250 were based on an estimated 18% of females reproducing in a total census size of 10,000-12,500 females [23, 73, 74]. More extreme estimates of 660 and 3304 were obtained from a stepwise mutation model estimator [23, 75], which is less likely to be affected by very recent demographic events. Simulated coalescence times assuming 51 alleles were calculated from  $10^6$  simulated values. Simulated coalescence times are as follows: for a female  $N_e$  estimate of 660, 51 alleles coalesce 10.88 ka (95% confidence intervals 4.22-30.38 ka); for a female  $N_e$  estimate of 1800: 29.72 ka (11.50-82.99 ka); for a female  $N_e$  estimate of 2250: 37.14 ka (14.36-103.66 ka); for a female  $N_e$  estimate of 3304: 54.58 ka (21.15-152.19 ka). Comparison of these coalescence times to estimates of the same event obtained from our Bayesian molecular

dating method suggests that our results are consistent with the timescale of evolution by genetic drift, as inferred from current knowledge of polar bear biology.

## **F. Evaluating the Molecular Rate**

Our estimate of the MRCA of brown bear and polar bear matriline assumes a molecular rate that is calibrated via the ages of the ancient bear sequences. Tip-calibrated rates are generally faster than those calibrated by constraining the ages at internal nodes in the tree, such as fossil ages [17, 76-78]. This is potentially a consequence of the time-dependency of molecular rate estimates, where more recent calibrations lead to faster evolutionary rates [17]. While the reasons for time-dependency remain controversial [77, 79, 80], this problem is clearly visible in brown bear data [17]. Recent evidence suggests that time-dependency may not significantly affect rate estimates over relatively short time-frames, including those situations where the calibration is within several hundred thousand years of the dates being estimated [25]. Several lines of evidence support the use of tip-calibration, as opposed to the fossil-calibration, for the inference of relatively recent divergence times in our brown bear control region analyses. First, three of our estimates as to the first appearance of brown bears in a particular region are supported by independent, paleontological evidence. Of these, two are very recent (the first appearance of brown bears in Eastern Europe ca. 22-5.5 kya and the first brown bears in Scandinavia 28-9 kya). The third (the establishment of clade III in Alaska 105-70 kya) coincides with anecdotal evidence from the small number of well-dated deposits in Eastern Beringia dating to MIS 5a or older, from which brown bears are not recovered [81, 82, pers. obs.]. Second, when a fossil-calibrated rate derived from an analysis of the same mtDNA control region fragment is applied to the control region analysis [17], not only do the molecular estimates disagree with the paleontological evidence, but they indicate that bears colonize Scandinavia while the Scandinavian Ice Sheet is known to be present, and no suitable habitat for bears is available. We argue therefore that, at least for the more recent divergence estimates, there is strong evidence in favor of using a tip-calibrated evolutionary rate.

### *3.E Comparing the Timing of the Nuclear and Mitochondrial MRCA of Brown Bears and Polar Bears*

To further investigate the hypothesis that polar bear matriline share a very recent common ancestor due to a hybridization event with female Irish brown bears, we performed additional BEAST analyses of 20 nuclear loci for eight species within the Ursidae. Four separate analyses were performed: i) all available nuclear loci; (ii) autosomal nuclear loci; (iii) twelve loci sequenced by Pages et al [7]; and (iv) only those loci that are found on the Y-chromosome. The most appropriate nucleotide substitution model for each gene or fragment was chosen using *Findmodel*, an online implementation of Modeltest [67] found at <http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html>. For each analysis, a Yule prior was used to inform the speciation rate, and both strict and uncorrelated lognormal relaxed clock analyses were performed. The molecular clock was calibrated via sampling the age of the Ursine radiation from a lognormal distribution with 95% of its probability distribution between 3 and 5 million years [45]. For each analysis, two MCMC chains of 100 million iterations were run, sub-sampling from the posterior every 10,000 iterations. The first 10% of iterations were discarded as burn-in and the

remainder from both analyses combined. Effective sample sizes and convergence of MCMC chains to stationarity were visually inspected using Tracer v1.4 [68].

The branching order within the ursine radiation is not well resolved [7, 8, 22, 63]. However, all four analyses of nuclear loci provided high statistical support for a close phylogenetic relationship between brown bears and polar bears. The small evolutionary distance between brown bears and polar bears combined with the stochastic nature by which mutations accumulate through time limited our ability to estimate precisely when the divergence occurred (particularly for the Y-chromosome data set), however the 95% HPD of sampled divergence dates recovered from all three nuclear analyses overlapped. Using the 95% HPDs from all four nuclear data sets, the MRCA of brown bears and polar bears spans the interval 2 Ma – 400 kya, with mean values ranging from 1.5 Ma – 700 kya, depending on the data set. The interpretation of these results is complicated, as the age of the common mitochondrial ancestor of modern brown bears may be more recent than the nuclear common ancestor for more than one reason, including a smaller mitochondrial effective population size, which may lead to shorter coalescence intervals and a more dynamic evolutionary history. Nonetheless, this estimate is significantly older than the MRCA of polar bear and brown bear matriline obtained from either the short fragment data set alone or from the partitioned analysis including complete mitochondrial genomes. More nuclear data will be necessary to determine precisely the age of the nuclear divergence between the two bear species.

As the existence of only few variable sites decreases the power of phylogenetic estimates of the MRCA using nuclear DNA, we use another approach to investigate the nuclear divergence between the two species. We focus on the analysis of 12 nuclear DNA fragments sequenced by Pages et al [7]. For each of these, the published sequences are a consensus of several individuals for each species. For brown bears, the consensus sequence was produced using a brown bear from southwestern Canada (clade IV) as well as several bears from western Europe (clade I), and therefore represents the most divergent modern lineages of brown bears. Polymorphisms within species were coded as nucleotide ambiguities. In the total of 7,698 bp, the polar bear consensus contains three ambiguous bases (polymorphic sites) and the brown bear consensus contains four. None of these seven polymorphic sites overlapped between brown bears and polar bears. In addition to these intra-specific polymorphisms, 17 nucleotide differences distinguish brown bears from polar bears: two synonymous differences in coding regions, four nonsynonymous differences, and 11 differences occurring in non-coding regions. These data suggest a much deeper divergence between brown and polar bears than is indicated by the mtDNA genomic analysis, and is consistent with a recent hybridization occurring between polar bears and brown bears.

### **Supplemental Acknowledgments**

Thanks also to Lauren Cadwallader, Department of Archaeology, Cambridge University, and Mike Hall and James Rolfe, Godwin Laboratory, Cambridge University, for assistance with the isotopic analyses.



## Supplemental References

46. Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. *Trends in Ecology and Evolution* *20*, 541-544.
47. Cooper, A., and Poinar, H.N. (2000). Ancient DNA: Do It Right or Not at All. *Science* *289*, 1139.
48. Greenwood, A.D., and Pääbo, S. (1999). Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Molecular Ecology* *8*, 133-137.
49. Edwards, C.J., MacHugh, D.W., Dobney, K.M., Martin, L., Russell, N., Horwitz, L.K., McIntosh, S.K., MacDonald, K.C., Helmer, D., Tresset, A., et al. (2004). Ancient DNA analysis of 101 cattle remains: limits and prospects. *Journal of Archaeological Science* *31*, 695-710.
50. Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., et al. (2004). Rise and fall of the Beringian steppe bison. *Science* *306*, 1561-1565.
51. Zazula, G.D., Hare, P.G., and Storer, J.E. (2009). New radiocarbon-dated vertebrate fossils from Herschel Island : implications for the palaeoenvironments and glacial chronology of the Beaufort Sea coastlands *Arctic* *62*, 273-280.
52. Pitul'ko, V.V. (1993). An early Holocene site in teh Siberian high arctic. *Arctic Anthropology* *30*, 13-21.
53. Rohland, N., and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nat Protoc* *2*, 1756-1762.
54. Yang, D.Y., Eng, B., Wayne, J.S., Dudar, J.C., and Saunders, S.R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol* *105*, 592-543.
55. Hänni, C., Laudet, V., Stehelin, D., and Taberlet, P. (1994). Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequencing. *Proceedings of the National Academy of Sciences of the U.S.A.* *91*, 12336-12340.
56. Taberlet, P., and Bouvet, J. (1994). Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proc Biol Sci* *255*, 195-200.
57. Calvignac, S., Hughes, S., and Hänni, C. (2009). Genetic diversity of endangered brown bear (*Ursus arctos*) populations at the crossroads of Europe, Asia and Africa. *Diversity and Distributions* *15*, 742-750.
58. Matheus, P., Burns, J., Weinstock, J., and Hofreiter, M. (2004). Pleistocene brown bears in the mid-continent of North America. *Science* *306*, 1150.
59. Kohn, M., Knauer, F., Stoffella, A., Schröder, W., and Pääbo, S. (1995). Conservation genetics of the European brown bear - a study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology* *4*, 95-104.
60. Hofreiter, M., Serre, D., Rohland, N., Rabeder, G., Nagel, D., Conard, N., Munzel, S., and Paabo, S. (2004). Lack of phylogeography in European mammals before the last glaciation. *Proc Natl Acad Sci U S A* *101*, 12963-12968.
61. Leonard, J.A., Wayne, R.K., and Cooper, A. (2000). Population Genetics of Ice Age Brown Bears. *Proceedings of the national Academy of Sciences U.S.A* *97*, 1651-1654.
62. Miller, C.R., Waits, L.P., and Joyce, P. (2006). Phylogeography and mitochondrial diversity of extirpated brown bear (*Ursus arctos*) populations in the contiguous United States and Mexico. *Mol Ecol* *15*, 4477-4485.

63. Yu, L., Li, Y.W., Ryder, O.A., and Zhang, Y.P. (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.
64. Woodman, P., McCarthy, M., and Monaghan, N. (1997). The Irish quaternary fauna project. *Quaternary Science Reviews* 16.
65. O'Connell, T.C., and Hedges, R.E.M. (1999). Isotopic comparison of hair and bone: archaeological analyses. *Journal of Archaeological Science* 26, 661-665.
66. Hoefs, J. (1997). *Stable Isotope Geochemistry*, (Springer).
67. Posada, D., and Crandall, K.A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.
68. Rambaut, A., and Drummond, A.J. (2007). Tracer v. 1.5, Available from <http://beast.bio.ed.ac.uk/>.
69. Valdiosera, C.E., Garcia-Garitagoitia, J.L., Garcia, N., Doadrio, I., Thomas, M.G., Hanni, C., Arsuaga, J.L., Barnes, I., Hofreiter, M., Orlando, L., et al. (2008). Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences of the United States of America* 105, 5123-5128.
70. Rodrigo, A. (2009). The coalescent: population genetic inference using genealogies. In *The Phylogenetic Handbook*, P. Lemey, M. Salemi and V. A.-M., eds. (Cambridge, UK: Cambridge University Press), pp. 549-563.
71. Rousset, F. (2004). *Genetic structure and selection in subdivided populations*, (Princeton, NJ: Princeton University Press).
72. Whitlock, M.C. (2003). Fixation probability and time in subdivided populations. *Genetics* 164, 767-779.
73. Aars, J., Lunn, N.J., and Derocher, A.E. (2006). *Polar Bears: Proceedings of the 14th Working Meeting of the IUCN/SSC Polar Bear Specialist Group*.
74. Regehr, E.V., Amstrup, S.C., and Stirling, I. (2006). Polar bear population status in the southern Beaufort Sea. U.S. Geological survey Open-File report. 2006-11337.
75. Ohta, T., and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res* 22, 201-204.
76. Ho, S.Y.W., Kolokotronis, S.O., and Allaby, R.G. (2007). Elevated substitution rates estimated from ancient DNA sequences. *Biology Letters* 3, 702-705.
77. Ho, S.Y.W., Shapiro, B., Phillips, M.J., Cooper, A., and Drummond, A.J. (2007). Evidence for time dependency of molecular rate estimates. *Systematic Biology* 56, 515-522.
78. Henn, B.M., Gignoux, C.R., Feldman, M.W., and Mountain, J.L. (2009). Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 26, 217-230.
79. Debruyne, R., and Poinar, H.N. (2009). Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? *Syst Biol* 58, 348-360.
80. Emerson, B.C. (2007). Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst Biol* 56, 337-345.
81. Harington, C.R. (1987). Stop 27: Pleistocene vertebrates of Old Crow Locality 11A. In *INQUA 87 Excurion Guide Book A-20 and I.U.f.Q. Research*, ed. (Ottawa, Canada).
82. Jopling, A.V., Irving, W.N., and Beebe, B.F. (1981). Stratigraphic, sedimentological and faunal evidence for the occurrence of Pre-Sangamonian artefacts in northern Yukon. *Arctic* 34, 3-33.