Electronic supplementary information

"Ancient mitochondrial DNA reveals convergent evolution of giant short-faced bears (Tremarctinae) in North and South America."

Sample Information: Our *Arctotherium* femur specimen was a surface find from Cueva del Puma, Chile (No. 32104, Centro de Estudios del Hombre Austral, Instituto de la Patagonia, Universidad de Magallanes). The femur can be assigned to *Arctotherium*, but is insufficiently diagnostic to be confidently identified to species level. However, based on the age of the specimen (10,345 ± 75 ybp, [1] and the size, temporal range, and distribution of *Arctotherium* species [2], it is likely to represent one of the terminal Pleistocene species (e.g. *Arctotherium tarijense*).

DNA extraction: The surface of the bone sample was removed using a Dremel tool to remove surface contamination. The cleaned bone was then powdered using a Mikro-Dismembrator (Sartorius), and 0.55 g of bone powder was decalcified overnight in 10 ml of 0.5 M EDTA (pH 8) on a rotary mixer at room temperature. The decalcified material was collected by centrifugation and further digested overnight at 55 °C on a rotary mixer with the addition of 3 ml of 100 mM Tris-HCl (pH 8), 100 mM NaCl, 0.5 mg/ml proteinase K, 10 mg/ml DTT, and 1% SDS. Following digestion, extraction was performed twice with Tris-saturated phenol and once with chloroform. The final aqueous phase was desalted using an Amicon Ultra-4 Centrifugal Filter Unit (Millipore) and concentrated to a final volume of 130 µL.

Sequencing library preparation: An Illumina sequencing library was constructed from the extracted *Arctotherium* DNA following the protocol of Meyer and Kircher [3], but using truncated and uniquely barcoded Illumina adapters exactly as described in a previous publication [4]. The sequencing library was enriched for *Arctotherium* mitochondrial DNA using commercially synthesised biotinylated 80 mer RNA baits (MYcroarray, MI, USA), which were designed based on published mitochondrial genome sequences for a number of mammal species including *Arctodus simus* (Table S1). DNA-RNA hybridisation enrichment was performed according to manufacturer's recommendations (MYbaits protocol v1) with the exception of the incubation step, which we extended to 44 hr (3 hr at 60°C, 12 hr at 55 °C, 12 hr at 50 °C, 17 hr at 55 °C). Subsequently, a final round of PCR was performed with fusion primers to add full-length Illumina sequencing adapters to the enriched sequencing [as for 4].

DNA sequencing and data processing: The library was diluted to 2 nM and run on an Illumina MiSeq using 2 x 150 bp (paired-end) sequencing chemistry. Following sequencing, reads were demultiplexed according to the adapter barcodes using 'sabre' (https://github.com/najoshi/sabre) (default parameters: no mismatches allowed). Adapter sequences were removed and paired-end reads were merged using Adapter Removal v2.1.2 [5]. Low quality bases were trimmed (<Phred20 –-minquality 4) and merged reads shorter than 25 bp were discarded (--minlength 25). Read quality was visualized using fastQC v0.10.1 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) before and after trimming to make sure the trimming of adapters was efficient.

Mitochondrial genome assembly: Reads were mapped against the mitogenome sequence of the brown bear (AF303110) using BWA v0.7.8 [6] (aln -l 1024, -n 0.01, -o 2). Reads with a mapping quality Phred score >30 were selected using the SAMtools v1.4 [7] view command (-q 30), and duplicate reads were discarded using 'FilterUniqueSAMCons.py' [8]. A 50% (majority-rule) consensus sequence was generated from all mapped reads and used as the reference for a subsequent round of mapping. This process was repeated for 21 iterations, until no additional reads could be aligned to the reference. A final 75% consensus sequence was then generated and checked by eye in Geneious v8.1.6 (Biomatters; http://www.geneious.com). Mapping a total of 1,460,585 filtered reads resulted in 26,657 unique aligned reads, which spanned 100% of the resulting consensus (all sites covered by >1 read), contributing to a final sequence of length 16,532 bp (GenBank accession KU886001). Mean length of individual reads was 80.5 bp (standard deviation = 34.7), while mean read depth across the consensus was mean 129.7x (standard deviation = 39.8). We used MapDamage v.2.0.6 to ensure that damage patterns in our data were consistent with authentic ancient DNA (Figure S1). Subsequently, we aligned the *Arctotherium* mitochondrial genome sequence to a number of previously published sequences using the MUSCLE algorithm as implemented in Geneious v8.1.6 (Biomatters; http://www.geneious.com), resulting in a matrix containing data for 13 species (Table S2).

PartitionFinder analysis: We provided PartitionFinder v1.1.1 [9] with an input of 32 regions: first and second positions of each individual mitochondrial protein-coding gene; and stem and loop sites of 12S rRNA, 16S rRNA, and concatenated

tRNAs. Stem and loop positions of RNA genes were identified using RNAalifold [10]. Optimum partitioning schemes were chosen based on the Bayesian Information Criterion (Table S3, S4).

Phylogenetic analyses:

RAxML

Our partitioned RAxML analysis comprised a maximum likelihood search for the best-scoring tree from 1,000 bootstrap replicates (-f a -m MULTIGAMMA -# 1000).

BEAST

We used BEAST to simultaneously infer the phylogeny and estimate the timeline for tremarctine bear evolution. In our partitioned BEAST analyses, we implemented a Birth-Death tree prior, a single lognormal relaxed clock model (with a rate multiplier parameter for each partition), and constrained the age of three key nodes in accordance with the fossil record (see Supplementary Figure S2):

1) The divergence between ursids and pinnipeds was constrained to the Oligocene/Eocene (23 – 55.8 Ma) based on the presence of early fossil members of these clades (e.g. *Enaliarctos*, *Ursavus*) in the Late Oligocene/Early Miocene, and a lack of identifiable modern carnivoran lineages prior to the Eocene. This calibration was implemented as a uniform distribution with a hard minimum

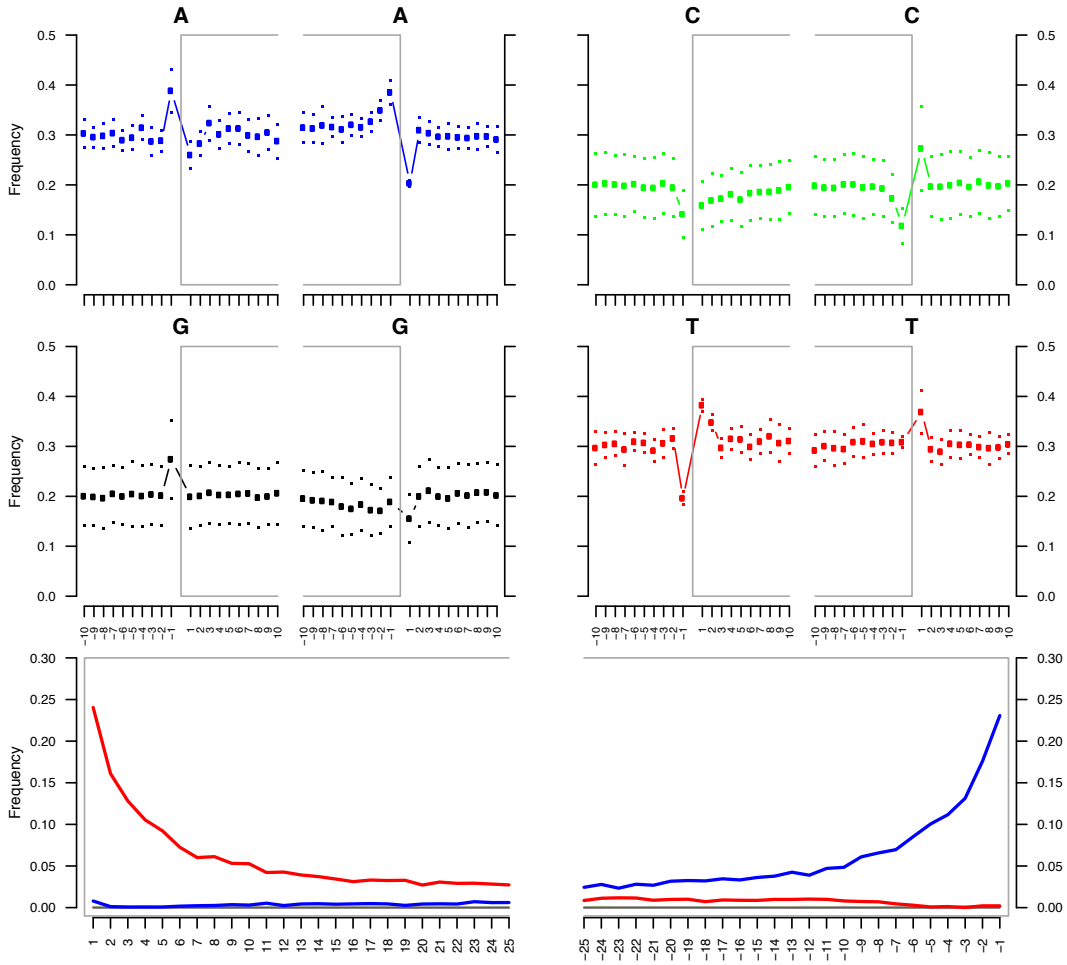
(23.0 Ma) and maximum (55.8 Ma) in order to reflect the uncertainty surrounding the timeframe for the divergence between bears and seals.

2) The crown-age of Ursidae (i.e. the divergence between the giant panda and the remaining extant/recent ursids) was constrained to the Miocene (11.6 – 23 Ma) based on the presence of a putative ailuropodine in the Middle Miocene (*Kretzoiarctos* [11]), and the fact that some Early Miocene representatives of *Ursavus* are likely ancestral to all modern and recent ursids [see 12]. This calibration was implemented as a lognormal distribution with a hard minimum at 11.6 Ma (offset = 11.6 Ma), a soft 95% maximum at 23.0 Ma, and a shape reflecting that the true divergence is unlikely to have occurred close to the minimum or maximum bound (standard deviation = 0.75, mean [in real space] = 4.4).

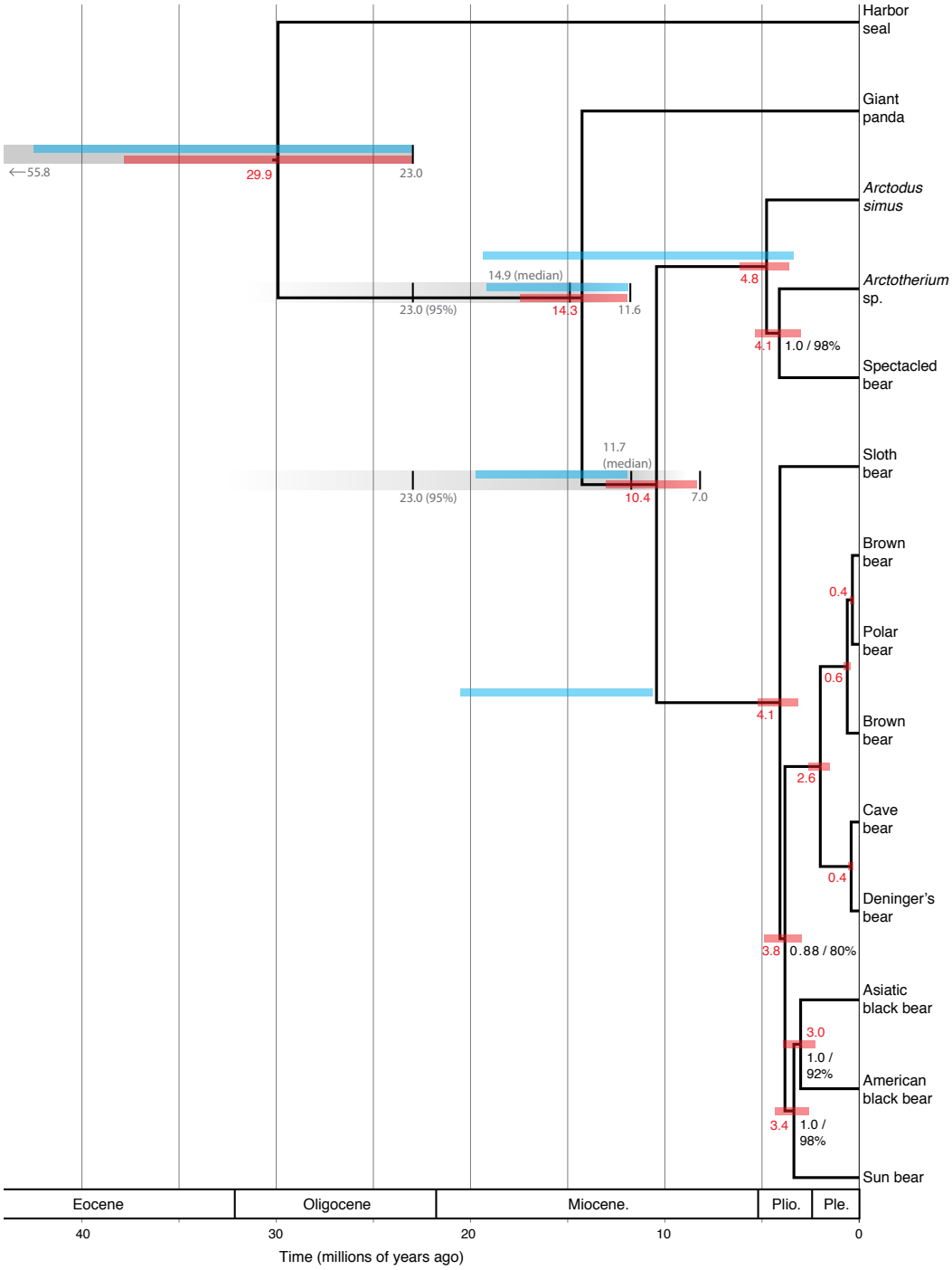
3) The divergence between Ursinae and Tremarctinae was constrained to the Miocene (7 – 23 Ma) based on the presence of putative ursine and tremarctine bears in the Late Miocene/Early Pliocene (e.g. *Plionarctos*), and the fact that some Early Miocene representatives of *Ursavus* are likely ancestral to both Ursinae and Tremarctinae [see 12]. This calibration was implemented as a lognormal distribution with a hard minimum at 7.0 Ma (offset = 7.0 Ma), a soft 95% maximum at 23.0 Ma, and a shape reflecting that the true divergence is unlikely to have occurred close to the minimum or maximum bound (standard deviation = 0.75, mean [in real space] = 6.175).

We ran three independent MCMC chains beginning from user-specified random starting trees generated from simulated data in Mesquite v3.04 [13]. Each chain was run for 10^8 generations, and we sampled trees and parameter values every 10^5 generations. The first 10% of samples were discarded as burnin. Parameter values were monitored and compared between the three independent chains in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure convergence and ESSs >200. We combined sampled trees and parameter values from each BEAST chain before summarising the results (Supplementary Figure S2).

We also ran our BEAST analysis without data (i.e. priors only) to determine the relative contribution of the data and the priors to the posterior node age estimates (Supplementary Figure S2). The posterior estimates for calibrated nodes 1 and 2 (see above) closely reflected the prior distributions, suggesting that much of the temporal signal was driven by these constraints (although the data still had a non-negligible impact on the posterior distributions). For most other nodes on the phylogeny posterior distributions for node age deviated considerably from the priors, demonstrating that most of the signal for the relative age of these divergences came from the data itself.



Supplementary Figure S1: MapDamage report for the final round of mapping to the *Arctotherium* mitochondrial genome consensus. The top panels show the characteristic high frequency of purines (A and G) immediately before the reads. The two lower panels show the accumulation of 5' C-to-T (red) and 3' G-to-A (blue) misincorporations characteristic of ancient DNA.



Supplementary Figure S2: BEAST maximum clade credibility tree. Bars associated with nodes are the calibrations specified in the BEAST xml file (grey; see main text for details), realised 95% age priors obtained by running the BEAST analysis without data (blue; displayed for select nodes only), and the 95% Highest Posterior Densities (red). Mean node ages are displayed in red text. Branch support values (BEAST posterior probability / RAxML bootstrap %) are given in black text for nodes that received less than maximum support (i.e. 1.0 / 100%).

Supplementary Table S1: Mitochondrial DNA sequences used to design RNA baits for hybridisation enrichment.

Species name	**Common name**	**Accession**
Arctodus simus	Giant short-faced bear	NC_011116
Bison bison	American bison	NC_012346
Bos taurus	Cattle	GU947008
Bradypus tridactylus	Three-toed sloth	AY960979
Bubalus bubalis	Water buffalo	NC_006295
Capra pyrenaica	Spanish ibex	FJ207528
Capricornis crispus	Japanese serow	AP003429
Dasypus novemcinctus	Nine-banded armadillo	NC_001821
Echinops telfairi	Lesser hedgehog tenrec	AB099484
Gazella dorcas	Dorcas gazelle	JN632637
Hippotragus niger	Sable antelope	JN632648
Kobus leche	Lechwe	JN632652
Lemur catta	Ring-tailed lemur	AJ421451
Macaca mulatta	Rhesus macaque	NC_005943
Mammuthus primigenius	Woolly mammoth	EU153453
Myotragus balearicus	Balearic Islands cave goat	AY380560
Oryctolagus cuniculus	European rabbit	NC_001913
Ovis aries	Sheep	HM236182
Panthera leo	Lion	KC834784
Pipistrellus abramus	Japanese pipistrelle	AB061528
Rhinolophus ferrumequinum	Greater horseshoe bat	NC_020326
Rupicapra rupicapra	Chamois	FJ207539
Sciurus vulgaris	Red squirrel	AJ238588
Ursus arctos	Brown bear	HQ685964

Supplementary Table S2: Mitochondrial genome sequences included in data matrix for phylogenetic analysis.

Species name	**Common name**	**Accession**
Ailuropoda melanoleuca	Giant panda	EF212882
Arctodus simus	-	NC_011116
Arctotherium sp.	-	KU886001
Helarctos malayanus	Malayan sun bear	EF196664
Melursus ursinus	Sloth bear	EF196662
Phoca vitulina	Harbor seal	NC_001325
Tremarctos ornatus	Spectacled bear	EF196665
Ursus americanus	American black bear	AF303109
Ursus arctos	Brown bear	EU497665
Ursus arctos	Brown bear	AF303110
Ursus deningeri	Deninger's bear	KF437625
Ursus maritimus	Polar bear	GU573488
Ursus spelaeus	Cave bear	EU327344
Ursus thibetanus	Asiatic black bear	EF196661

Supplementary Table S3: Best scoring partitioning scheme (according to the BIC) for RAxML analyses

Partition	Model	Positions
1	GTR+G	First codon positions of ATP6, CYTB, ND1, ND2, ND3, ND4L, ND4, ND5, and ND6, plus second codon positions of ND6
2	GTR+G	First codon positions of CO1, CO2, and CO3, plus RNA stems
3	GTR+G	First and second codon positions of ATP8, plus RNA loops
4	GTR+G	Second codon positions of ATP6, CO1, CO2, CO3, CYTB, ND1, ND2, ND3, ND4L, ND4, and ND5
5	GTR+G	Third codon positions of ATP8, CYTB, ND1, ND2, ND4, and ND6
6	GTR+G	Third codon positions of ATP6, CO1, CO2, CO3, ND3, ND4L, and ND5

Supplementary Table S4: Best scoring partitioning scheme (according to the BIC) for BEAST analyses

Partition	Model	Positions
1	TrNef+I+G	First codon positions of CO1, CO2, CO3, and ND1, plus RNA stems
2	TrN+I+G	First codon positions of ATP6, CYTB, ND2, ND3, ND4L, ND4, ND5, and ND6, plus second codon positions of ND6
3	HKY+G	First and second codon positions of ATP8, plus RNA loops
4	K81uf+I+G	Second codon positions of CO1, CO2, CO3, CYTB, ND1, ND4L, and ND4
5	HKY+I+G	Second codon positions of ATP6, ND2, ND3, and ND5
6	TrN+I+G	Third codon positions of CO3, CYTB, ND1, ND2, ND3, ND4L, ND4, ND5, and ND6
7	HKY+I	Third codon positions of ATP6, ATP8, CO1, and CO2

References

1. Martin F.M., Prieto A., Román M.S., Morello F., Prevosti F., Cárdenas P., Borrero L. 2004 Late-Pleistocene Megafauna at Cueva del Puma, Pali-Aike Lava Field, Chile. *Curr Res Pleistocene* (21), 101-103.

2. Soibelzon L.H., Tonni E.P., Bond M. 2005 The fossil record of South American short-faced bears (Ursidae, Tremarctinae). *J South Am Earth Sci* **20**(1-2), 105-113.

3. Meyer M., Kircher M. 2010 Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**(6), 1-10.

4. Mitchell K.J., Llamas B., Soubrier J., Rawlence N.J., Worthy T.H., Wood J., Lee M.S.Y., Cooper A. 2014 Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**(6186), 898-900.

5. Lindgreen S. 2012 AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Research Notes* **5**(1), 337.

6. Li H., Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14), 1754-1760.

7. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup G.P.D.P. 2009 The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* **25**(16), 2078-2079.

8. Kircher M. 2012 Analysis of High-Throughput Ancient DNA Sequencing Data. In *Ancient DNA: Methods and Protocols* (pp. 197-228).

9. Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012 PartitionFinder: combined selection of partitioning schemes and substitution. *Mol Biol Evol* **29**(6), 1537-1719.

10. Bernhart S., Hofacker I., Will S., Gruber A., Stadler P. 2008 RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**(1), 474.

11. Abella J., Alba D.M., Robles J.M., Valenciano A., Rotgers C., Carmona R., Montoya P., Morales J. 2012 Kretzoiarctos gen. nov., the Oldest Member of the Giant Panda Clade. *PLoS ONE* **7**(11), e48985.

12. McLellan B. 1994 A review of bear evolution. *International Conference Bear Research and Management* **9**(1), 85-96.

13. Maddison W.P., Maddison D.R. 2015 Mesquite: a modular system for evolutionary analysis. Version 3.04 http://mesquiteproject.org.